Phase Space Ray Tracing for Illumination Optics

Carmela Filosa

Cover art:
Photography:

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-3972-7

Phase Space Ray Tracing for Illumination Optics

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen

door

Carmela Filosa

geboren te Torre del greco, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr.
1e promotor:	prof.dr. W.L. IJzerman
copromotor:	dr. J.H.M. ten Thije Boonkkamp
leden:	

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Contents

List of symbols

t	time
Q	Total energy emitted from a light sorce or received by a target
Φ_r	Radiant flux
Φ	Luminous flux
λ	Wavelength
Ψ_r	Power per wavelength
$\bar{y}(\lambda)$	Luminousity function
E	Illuminance
$\mathrm{d}\Omega$	Solid angle
I	Intensity
L	Luminance
U	étendue
$\boldsymbol{\nu}$	Surface normal
n	Index of refraction of the medium in which a surface is immersed
θ	Angle between the direction of the solid angle and the normal $\boldsymbol{\nu}$
n_i	Index of refraction of the medium in which the incident ray travels
$n_\mathrm{r} = n_\mathrm{i}$	Index of refraction of the medium in which the reflected ray is located
n_t	Index of refraction of the medium in which the transmitted ray travels
$n_\mathrm{i,t}$	$\frac{n_\mathrm{i}}{n_\mathrm{t}}$
θ_i	Angle between the incident ray and the normal $\boldsymbol{\nu}$
θ_r	Angle between the reflected ray and the normal $\boldsymbol{\nu}$
θ_t	Angle between the transmitted ray and the normal $\boldsymbol{\nu}$
θ_c	Critical angle
$\boldsymbol{t}_\mathrm{i}$	Direction of the incident ray
$\boldsymbol{t}_\mathrm{r}$	Direction of the reflected ray
$\boldsymbol{t}_\mathrm{t}$	Direction of the transmitted ray
$\boldsymbol{\nu}_j$	Normal to the line j
t_j	Angle that the ray located on line j forms with respect to the optical axis
θ_j	Angle between the ray and the normal $\boldsymbol{\nu}_j$ to line j
n_j	Index of refraction of the medium in which line j is located

Chapter 1

Introduction

1.1 Motivation

1.2 Methods and results

1.3 Content of this thesis

Chapter 2

Illumination optics

This chapter provides some concepts of illumination optics used in this thesis. We start explaining the difference between radiometry and photometry. In particular, we focus on the photometric variables, defining them both in three and two dimensions. The reflection and refraction laws and the phenomenon of total internal reflection are explained next . The last paragraph of the chapter gives a brief introduction to Fresnel reflection.

2.1 Radiometric and photometric variables

Radiometry is concerned with the measurement of electromagnetic radiation across the entire electromagnetic spectrum. Photometry is the subfield of radiometry that takes into account only the portion of the electromagnetic spectrum corresponding to the visible light [1]. Radiometry deals with radiometric quantities. An important radiometric quantity is the radiant flux Φ_{r} (unit watt W) which is the total energy emitted from a source or received by a target per unit time:

$$\Phi_{\mathrm{r}} = \frac{\mathrm{d}Q}{\mathrm{d}t}\,, \tag{2.1.1}$$

where Q is the energy and t the time.

In illumination optics the measurement of light is given in terms of the impression that it gives on the human eye. Therefore, illumination optics deals with photometric variables rather than with radiometric variables. The most important photometric variables are defined in the following using the notation adopted by Chaves in [2]. The luminous flux Φ (unit lumen lm) is defined as the *perceived* power of light by the human eye. The radiant and the luminous flux are related by the luminous efficacy function y, unit lm/W, which defines how many lumen correspond to one Watt of power at a given wavelength. The luminous efficacy reaches its maximum at a wavelength of 555 nm where it is equal to 683 lm/W. We may normalize the luminous efficacy function with its maximum value of 683. The normalized function $\bar{y}(\lambda)$ is the luminous efficiency shown in Figure 2.1 where λ is the wavelength. It is a dimensionless quantity with a range of value between 0 and 1, [3].

The luminous flux corresponding to one Watt of radiation power at any wavelength

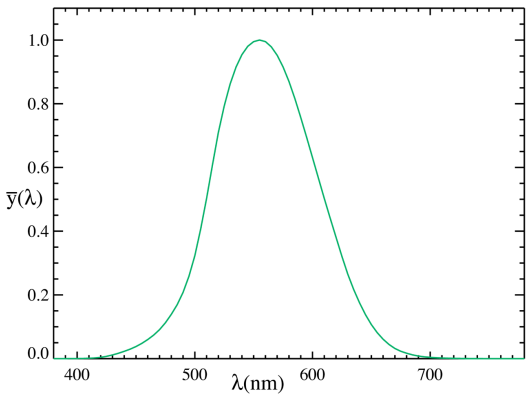


Figure 2.1: Luminosity function $\bar{y}(\lambda)$: relation between the eye's sensitivity and the wavelength of light. The luminosity function is dimensionless [4].

is given by the product of 683 lm/W and the luminosity function at the same wavelength, i.e. $683\,\bar{y}(\lambda)$. Hence, the total luminous flux Φ has unit lumen (lm) and it is defined as:

$$\Phi = 683 \int_0^\infty \Phi_{\mathrm{r}}(\lambda)\bar{y}(\lambda)\mathrm{d}\lambda, \tag{2.1.2}$$

where $\Phi_{\mathrm{r}}(\lambda)$ is the spectral radiant flux, i.e. the radiant flux per unit wavelength (unit W/m). The luminous emittance $M = M(x, \theta)$ is the total flux emitted in all direction from a unit area. It is measured in lumens pr square meters ($\mathrm{lm/m^2}$).

A beam of light can be described as a collection of parallel light rays, where a light ray can be interpreted as a path along which the energy travels. The luminous flux $\mathrm{d}\Phi$ incident on a surface is called illuminance E (unit $\mathrm{lm/m^2}$) and is defined as:

$$E = E(x) = \frac{\mathrm{d}\Phi}{\mathrm{d}A}\,, \tag{2.1.3}$$

where $\mathrm{d}A$ is an infinitesimal area receiving radiation. The density of light emitted by a point source in a given direction is determined by the solid angle.

The solid angle in a given direction is expressed by a cone of rays emitted in that particular direction by a point source located at the center of the unit sphere, [5]. Let $\mathrm{d}S$ be the area on the unit sphere subtended by the cone, the infinitesimal solid angle $\mathrm{d}\Omega$ is given by:

$$\mathrm{d}\Omega = \mathrm{d}S = \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi \tag{2.1.4}$$

where θ and ϕ are the polar and the azimuthal angle that the normal $\boldsymbol{\nu}$ to $\mathrm{d}A$ makes with the direction of the central line of $\mathrm{d}\Omega$, respectively (see Figure 2.2). The solid angle on the entire sphere is $\Omega = 4\pi$ and its unit is steradian sr, [6]. The luminous intensity I (unit candela cd = lm/sr) is defined as the luminous flux $\mathrm{d}\Phi$ per solid angle $\mathrm{d}\Omega$ and is given by:

$$I = I(\theta, \phi) = \frac{\mathrm{d}\Phi}{\mathrm{d}\Omega}\,. \tag{2.1.5}$$

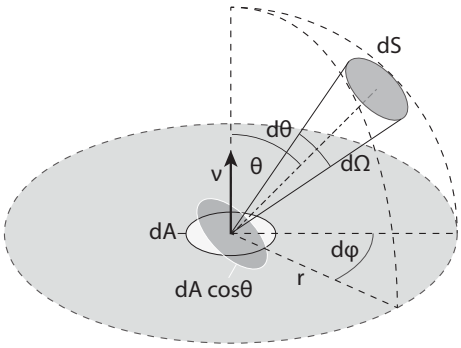


Figure 2.2: Solid angle $\mathrm{d}\Omega$ in a given direction θ with θ the angle that the central line forms with the normal to the area $\mathrm{d}A$.

Let us now consider a finite source $\mathrm{d}A$. The luminance $L = L(\boldsymbol{x}, \theta)$ (unit $\mathrm{cd/m^2}$) depends both on the position and the direction, it is the luminous flux per unit solid angle $\mathrm{d}\Omega$ and per unit projected area $\cos\theta \mathrm{d}A$. L is given by:

$$L = L(\boldsymbol{x}, \theta) = \frac{\mathrm{d}\Phi}{\cos\theta \mathrm{d}A \mathrm{d}\Omega}\,. \tag{2.1.6}$$

Note that from (2.1.5) and (2.1.6) we can derive a relation between the intensity and the luminance. The intensity I emitted by the infinitesimal area $\mathrm{d}A$ is given by:

$$I = \frac{\mathrm{d}\Phi}{\mathrm{d}\Omega} = L(\boldsymbol{x}, \theta)\cos\theta \mathrm{d}A\,. \tag{2.1.7}$$

When the luminance is uniform over a finite area A, the luminous intensity emitted in the direction θ is:

$$I(\boldsymbol{x}, \theta) = I(\theta) = L(\theta) A \cos\theta\,. \tag{2.1.8}$$

Thus, when $L(\boldsymbol{x}, \theta)$ does not depend on the position and the direction (i.e. $L(\boldsymbol{x}, \theta) = L$), we obtain Lambert's cosine law:

$$I(\theta) = I_0 \cos\theta\,. \tag{2.1.9}$$

where $I_0 = I(\theta = 0) = LA$.

Finally, the étendue U (unit $\mathrm{m^2 sr}$) describes the ability of a source to emitt light or the capability of an optical system to receive light [7]. The quantity $\mathrm{d}U$ of a source is defined as:

$$\mathrm{d}U = n^2 \frac{1}{L} \mathrm{d}\Phi = n^2 \cos\theta \mathrm{d}A \mathrm{d}\Omega, \tag{2.1.10}$$

where n is the index of refraction of the medium in which the surface $\mathrm{d}A$ is immersed. In optics the étendue is considered to be a volume in phase space (or an area for two-dimensional systems). This concept will be clarified in Chapter 4 in which we treat the phase space in more detail. An important property of the étendue is that it is conserved within an optical system in absence of absorption. We now show, using the approach of Chaves in [2], how conservation of this quantity can be derived. Consider

a light ray emitted from an infinitesimal area $\mathrm{d}A_1$ to the area $\mathrm{d}A_2$. Suppose that the centers of $\mathrm{d}A_1$ and $\mathrm{d}A_2$ are located at a distance d from each other, see Figure 2.3. Indicating with $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ the normals to the surfaces $\mathrm{d}A_1$ and $\mathrm{d}A_2$, respectively and

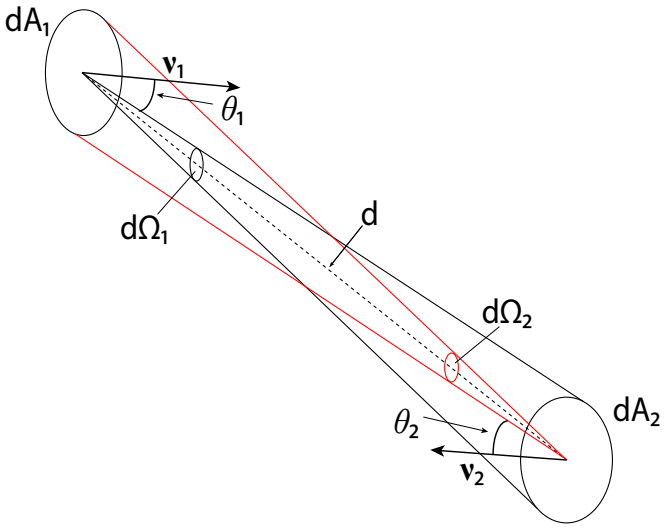


Figure 2.3: $\mathrm{d}A_1$ and $\mathrm{d}A_2$ are two surfaces with normals ν_1 and ν_2, respectively. Their centers are located at a distance d. θ_1 and θ_2 are the angles made by the central ray with the normals ν_1 and ν_2, respectively.

with θ_1 and θ_2 the angles that the ray connecting the centers of $\mathrm{d}A_1$ and $\mathrm{d}A_2$ forms with $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, respectively, the flux $\mathrm{d}\Phi_1$ passing through $\mathrm{d}A_2$ emitted from $\mathrm{d}A_1$ and the corresponding solid angle $\mathrm{d}\Omega_1$ are defined as:

$$\begin{aligned} \mathrm{d}\Omega_1 &= \frac{\mathrm{d}A_2 \cos(\theta_2)}{d^2}\,, \\ \mathrm{d}\Phi_1 &= L \cos\theta_1 \mathrm{d}A_1 \mathrm{d}\Omega_1. \end{aligned} \tag{2.1.11}$$

Similarly, the flux $\mathrm{d}\Phi_2$ passing through $\mathrm{d}A_1$ emitted from $\mathrm{d}A_2$ is equal to:

$$\begin{aligned} \mathrm{d}\Omega_2 &= \frac{\mathrm{d}A_1 \cos\theta_1}{d^2}\,, \\ \mathrm{d}\Phi_2 &= L \cos\theta_2 \mathrm{d}A_2 \mathrm{d}\Omega_2. \end{aligned} \tag{2.1.12}$$

Then from Eq. (2.1.10) we obtain the following relations:

$$\begin{aligned} \mathrm{d}U_1 = n^2 \cos\theta_1 \mathrm{d}A_1 \mathrm{d}\Omega_1 &= \frac{n^2 \cos\theta_1 \mathrm{d}A_1 \mathrm{d}A_2 \cos\theta_2}{d^2}, \\ \mathrm{d}U_2 = n^2 \cos\theta_2 \mathrm{d}A_2 \mathrm{d}\Omega_2 &= \frac{n^2 \cos\theta_2 \mathrm{d}A_2 \mathrm{d}A_1 \cos\theta_1}{d^2}. \end{aligned} \tag{2.1.13}$$

for $\mathrm{d}A_1$ and $\mathrm{d}A_2$, respectively. From the previous equations we can conclude that $\mathrm{d}U_1 = \mathrm{d}U_2$ and therefore the étendue $\mathrm{d}U$ is conserved along a beam of light. Since

also the flux through the areas $\mathrm{d}A_1$ and $\mathrm{d}A_2$ is conserved, the following relation holds:

$$L := n^2 \frac{\mathrm{d}\Phi}{\mathrm{d}U} = constant\,. \tag{2.1.14}$$

In the optical systems we will consider in this work, the source and the target are located in the same medium (air) with $n = 1$, so the luminance L equals the basic luminance $L^* = L/n^2$ at the source and the target of the system.

In this thesis we consider two-dimensional optical systems. Hence, the definitions of the photometric parameters have to be given in two dimensions. An infinitesimal line segment of length $\mathrm{d}a$ that emits a light beam and the ray that makes an angle θ with the normal $\boldsymbol{\nu}$ are considered, see Fig. 2.4. The two-dimensional illuminance

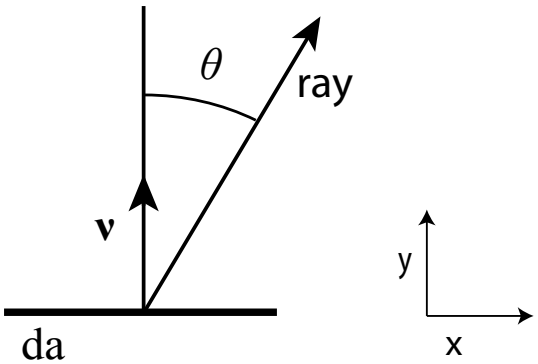


Figure 2.4: Ray emitted by an infinitesimal line segment $\mathrm{d}a$ that makes an angle θ with respect to the line normal $\boldsymbol{\nu}$.

$\big(\text{unit } \big[\mathrm{lm/m}\big]\big)$ denotes the luminous flux falling on an infinitesimal line segment of length $\mathrm{d}a$ and it is given by:

$$E = \frac{\mathrm{d}\Phi}{\mathrm{d}a}\,. \tag{2.1.15}$$

The luminous intensity $\big(\text{unit } \big[\mathrm{lm/rad}\big]\big)$ is the luminous flux per angle $\mathrm{d}\theta$:

$$I = \frac{\mathrm{d}\Phi}{\mathrm{d}\theta}\,. \tag{2.1.16}$$

The two-dimensional luminance $\big(\text{unit } \big[\mathrm{lm/(rad \cdot m)}\big]\big)$ is given by:

$$L = \frac{\mathrm{d}\Phi}{\cos\theta\,\mathrm{d}a\,\mathrm{d}\theta}. \tag{2.1.17}$$

Thus the following relation holds:

$$I = L(x,\theta)\cos\theta\,\mathrm{d}a \tag{2.1.18}$$

where x is a certain position at the light source $\mathrm{d}a$. Finally, the étendue $\mathrm{d}U$ (unit $[\mathrm{m \cdot rad}]$) in two dimensions is given by:

$$\mathrm{d}U = n\cos\theta \mathrm{d}a\,\mathrm{d}\theta. \tag{2.1.19}$$

In order to determine the light distribution on a surface and to compute the photometric variables on that surface, we need to understand how the light emitted from the source propagates. In the field of geometric optics the light propagation is described by light rays. The propagation of a light ray traveling through different media is determined by the reflection and refraction law. In the following we introduce these two laws and we explain the total internal reflection phenomenon.

2.2 Reflection and refraction law

A light ray is described by a position vector $\boldsymbol{x}$ on a surface and a direction vector $\boldsymbol{t}$ and can be parameterized by the arc length s. Light rays travel in a homogeneous medium along straight lines, once they hit a reflective surface their direction changes. Denoting with $\boldsymbol{t}_\mathrm{i}$ the direction of the incident ray and with $\boldsymbol{\nu}$ the unit normal to the surface at the location of incidence, the direction $\boldsymbol{t}_\mathrm{r}$ of the reflected ray is given by:

$$\boldsymbol{t}_\mathrm{r} = \boldsymbol{t}_\mathrm{i} - 2(\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu})\boldsymbol{\nu}\,, \tag{2.2.1}$$

where the vectors $\boldsymbol{t}_\mathrm{i}$ and $\boldsymbol{\nu}$ are unit vectors and $\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu}$ indicates the scalar product between $\boldsymbol{t}_\mathrm{i}$ and $\boldsymbol{\nu}$. From Eq. (2.2.1) it follows that the vector $\boldsymbol{t}_\mathrm{r}$ is a unit vector too, indeed considering the scalar product $(\boldsymbol{t}_\mathrm{r}, \boldsymbol{t}_\mathrm{r})$ we conclude:

$$\boldsymbol{t}_\mathrm{r} \cdot \boldsymbol{t}_\mathrm{r} = \boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{t}_\mathrm{i} - 4(\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu})(\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu}) + 4(\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu})^2(\boldsymbol{\nu} \cdot \boldsymbol{\nu}) = 1. \tag{2.2.2}$$

The vectors $\boldsymbol{t}_\mathrm{i}$, $\boldsymbol{t}_\mathrm{r}$ and $\boldsymbol{\nu}$ live all in the same plane. Defining the incident angle θ_i and the reflective angle θ_r such that θ_i, $\theta_\mathrm{r} \in [0, \pi/2)$. the reflection law states that $\theta_\mathrm{i} = \theta_\mathrm{r}$, see Fig. 2.5.
When a ray propagates through two different media, its direction changes according to the law of refraction. Indicating with n_i the index of refraction of the medium in which the incident ray travels and with n_t the index of refraction of the medium of the transmitted ray, the direction $\boldsymbol{t}_\mathrm{t}$ of the transmitted ray is given by:

$$\boldsymbol{t}_\mathrm{t} = n_{\mathrm{i,t}}\, \boldsymbol{t}_\mathrm{i} + \left[\sqrt{1 - n_{\mathrm{i,t}}^2 + n_{\mathrm{i,t}}^2(\boldsymbol{\nu} \cdot \boldsymbol{t}_\mathrm{i})^2} - n_{\mathrm{i,t}}(\boldsymbol{\nu} \cdot \boldsymbol{t}_\mathrm{i})\right]\boldsymbol{\nu}\,, \tag{2.2.3}$$

where $n_{\mathrm{i,t}} = n_\mathrm{i}/n_\mathrm{t}$, [2]. Note that in Eq. (2.2.1) the direction of the normal $\boldsymbol{\nu}$ to the surface is not relevant for the computation of the direction of the reflective ray, since:

$$\boldsymbol{t}_\mathrm{r} = \boldsymbol{t}_\mathrm{i} - 2(\boldsymbol{t}_\mathrm{i} \cdot \boldsymbol{\nu})\boldsymbol{\nu} = \boldsymbol{t}_\mathrm{i} - 2(\boldsymbol{t}_\mathrm{i} \cdot -\boldsymbol{\nu})(-\boldsymbol{\nu}), \tag{2.2.4}$$

however, this is not the case for Eq. (2.2.3), therefore in the latter case we need to specify the direction of $\boldsymbol{\nu}$ which is usually chosen in such a way that the angle that it forms with the incident ray $\boldsymbol{t}_\mathrm{i}$ is smaller than or equal to $\pi/2$. Hence, if $(\boldsymbol{t}_\mathrm{i}, \boldsymbol{\nu}) \leq 0$ the normal $\boldsymbol{\nu}$ directed inside the same medium in which travels the incident ray is taken as in Fig. 2.5, otherwise the normal $-\boldsymbol{\nu}$ directed inside the same medium in which the transmitted ray will travel has to be considered.

Eq. (2.2.3) is only valid for

$$1 - n_{\mathrm{i,t}}^2 + n_{\mathrm{i,t}}^2(\boldsymbol{\nu} \cdot \boldsymbol{t}_\mathrm{i})^2 \geq 0 \tag{2.2.5}$$

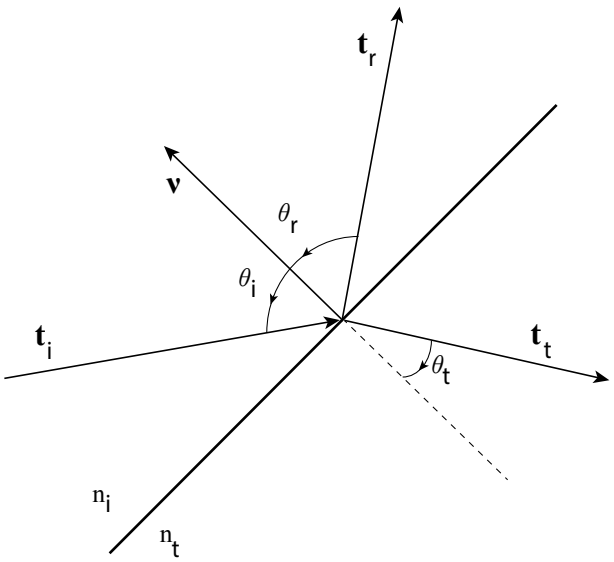


Figure 2.5: Propagation of a ray through two different media with index of refraction n_i and n_t.

which implies that

$$\frac{n_\mathrm{t}}{n_\mathrm{i}} \geq \sqrt{1-(\boldsymbol{\nu}\cdot\boldsymbol{t}_\mathrm{i})^2} \tag{2.2.6}$$

from which we obtain:

$$n_\mathrm{t} \geq n_\mathrm{i}\sin\theta_\mathrm{i}\,. \tag{2.2.7}$$

The angle θ_c for which the equality holds is

$$\theta_\mathrm{c} = \arcsin\left(\frac{n_\mathrm{t}}{n_\mathrm{i}}\right) \tag{2.2.8}$$

and it is called the critical angle, [2]. When the incident angle θ_i is exactly equal to the critical angle θ_c, the square root in Eq. (2.2.3) is zero and the inner product $(\boldsymbol{t}_\mathrm{t},\boldsymbol{\nu}) = 0$, hence the transmitted ray propagates parallel to the refractive surface. When $\theta_\mathrm{i} > \theta_\mathrm{c}$ the light ray is no longer refracted but is only reflected by the surface. This phenomenon is called total internal reflection (TIR). When TIR occurs, 100% of light is reflected and there is no loss of energy. Therefore, optical systems designed such that rays are reflected by TIR are very efficient. Lght that hits an ordinary refractive surface can be reflected and refracted. The energy that is reflected and refracted is determined by the Fresnel's coefficients. In the next paragraph an overview of the Fresnel coefficients is given.

2.3 Fresnel's equations

In order to derive Fresnel's equations we need to describe light as an electromagnetic wave. It is therefore useful to study the light propagation from the perspective

of electromagnetic theory which gives information about the incident, reflected and transmitted radiant flux density that are denoted with E_i, E_r and E_t, respectively.

onent of the electric field $\boldsymbol{\mathcal{E}}$ can be written as

$$\boldsymbol{\mathcal{E}}(\boldsymbol{x},t) = \boldsymbol{\mathcal{E}}_0(\boldsymbol{x})e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \tag{2.3.1}$$

where $\boldsymbol{x}$ is the position vector and T is the time. The amplitude $\boldsymbol{\mathcal{E}}_0(\boldsymbol{x})$ is constant in time and $\omega = \frac{ck}{n}$ is the value of the angular frequency with c the velocity of light and n the index of refraction in which the wave is traveling, which is the ratio of the speed of light c in vacuum and the speed of light v in the material. Note that the angular frequency can be also written as $\omega = vk$, in particular when a wave travels in vacuum $n = 1$ and $\omega = ck$. The vector $\boldsymbol{k}$ has the same direction of the wave and its absolute value $|\boldsymbol{k}| = k = \frac{2\pi}{\lambda}$ is the wave number in vacuum, with λ the wavelength. Similarly, the magnetic field has the form:

$$\boldsymbol{\mathcal{B}}(\boldsymbol{x},t) = \boldsymbol{\mathcal{B}}_0(\boldsymbol{x})e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\,. \tag{2.3.2}$$

Light can be seen as an electromagnetic wave, that is an oscillating electric field $\boldsymbol{\mathcal{E}}$ and an oscillating magnetic filed $\boldsymbol{\mathcal{B}}$ which propagates always perpendicular to $\boldsymbol{\mathcal{E}}$. Th eelectric field oscillates perpendicular to the wave propagation. Light is said to be polarized if the direction of the electic field is well defined. When the electric field propagates in different directions we talk about unpolarized light. By convention, we refer to the light's polarization as the direction of the electric field $\boldsymbol{\mathcal{E}}$, [8] with respect to the incident plane that is defined by the incident and reflected rays as is shown in Fig. 2.6.

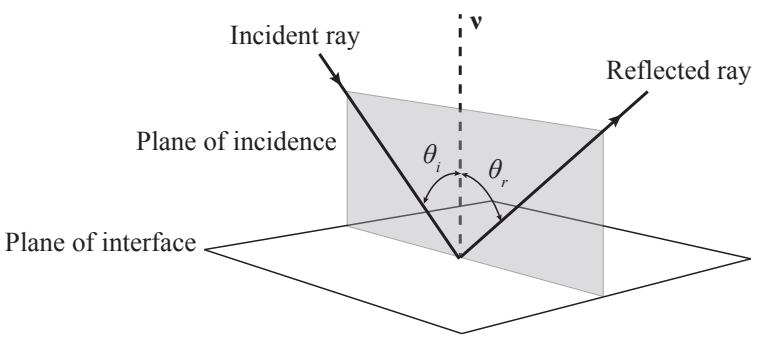


Figure 2.6: Light ray that hits a mirror located on the reflecting plane. The incident and the reflected ray leave in the same plane of the normal to the mirror that is called plane of incident.

In order to derive the Fresnel's coefficients the polarization of light must to be taken into account. Those coefficients are obtained considering Maxwell's equations and the boundary conditions due to the conservation of energy. The details of Fresnel's equations are widely explained in the literature. In the following we provide Fresnel coefficients and we briefly explain their physical interpretation. We refer the reader to [9, 10] for more details. Fresnel's coefficients can also be derived using a different

approach that does not involve Maxwell's equations, this method is explained in [11]. The following particular cases of light's polarization need are considered.

1. $\boldsymbol{\mathcal{E}}$ is perpendicular to the plane of incidence (see Fig. 2.7). In this case light is said to be s-polarized.

2. $\boldsymbol{\mathcal{E}}$ is parallel to the plane of incidence (see Fig. 2.8). In this case light is said to be p-polarized.

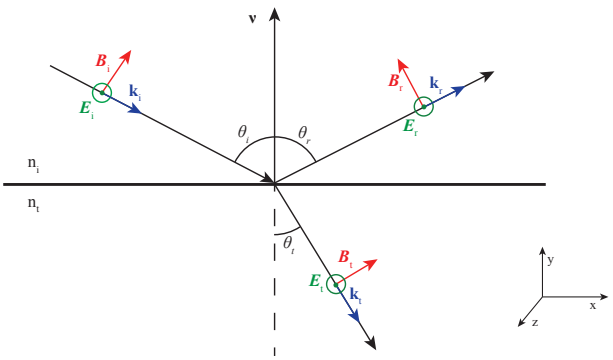


Figure 2.7: Propagation of an electromagnetic wave where $\boldsymbol{\mathcal{E}}$ is perpendicular to the incident plane. The components of $\boldsymbol{\mathcal{E}}$ are is indicated with the green circles. The components of $\boldsymbol{\mathcal{B}}$ are indicated with red arrows.

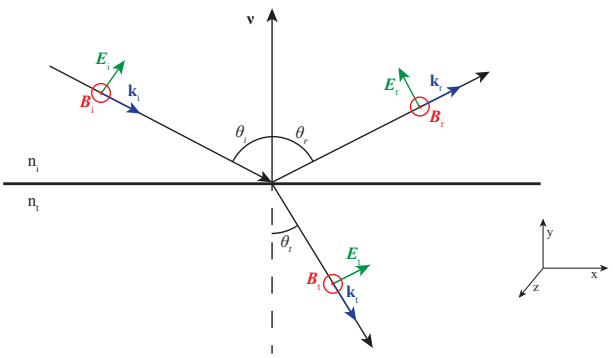


Figure 2.8: Propagation of an electromagnetic wave where $\boldsymbol{\mathcal{E}}$ is parallel to the incident plane. The components of $\boldsymbol{\mathcal{B}}$ are indicated with the red circle. The components of $\boldsymbol{\mathcal{E}}$ are indicated with greed arrows.

Energy conservation gives the boundary conditions of the electromagnetic field at the plane of the interface (which is perpendicular to the incident plane). In the following we derive Fresnel's coefficients for case 1. Similarly, the Fresnel's coefficients can be derived for the second case.

For s-polarized light the tangential components of $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}/\mu$ across the boundary between the two different media must be continuous. The continuity of the tangential component of $\boldsymbol{\mathcal{E}}$ leads to:

$$|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}| + |\boldsymbol{\mathcal{E}}_{0\mathrm{r}}| = |\boldsymbol{\mathcal{E}}_{0\mathrm{t}}|, \tag{2.3.3}$$

while the continuity of the tangential component of $\boldsymbol{\mathcal{B}}/\mu$ gives:

$$-\frac{|\boldsymbol{\mathcal{B}}_{0,\mathrm{i}}|}{\mu_\mathrm{i}}\cos\theta_\mathrm{i} + \frac{|\boldsymbol{\mathcal{B}}_{0,\mathrm{r}}|}{\mu_\mathrm{r}}\cos\theta_\mathrm{r} = -\frac{|\boldsymbol{\mathcal{B}}_{0,\mathrm{t}}|}{\mu_\mathrm{t}}\cos\theta_\mathrm{t}, \tag{2.3.4}$$

where the negative sign in front of $|\boldsymbol{\mathcal{B}}_{0,\mathrm{i}}|$ and $|\boldsymbol{\mathcal{B}}_{0,\mathrm{t}}|$ is due to the convention that a positive direction is considered with increasing x. Since $\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{E}}/v$, Eq. (2.3.4) can be written as

$$\frac{1}{\mu_\mathrm{i} v_\mathrm{i}}(|\boldsymbol{\mathcal{E}}_{0,\mathrm{i}}| - |\boldsymbol{\mathcal{E}}_{0,\mathrm{r}}|)\cos\theta_\mathrm{i} = \frac{1}{\mu_\mathrm{t} v_\mathrm{t}}|\boldsymbol{\mathcal{E}}_{0,\mathrm{t}}|\cos\theta_\mathrm{t}, \tag{2.3.5}$$

where we employed the fact that $v_\mathrm{i} = v_\mathrm{r}$, and $\theta_\mathrm{i} = \theta_\mathrm{r}$. Using Eq. (2.3.1) and $n = c/v$, the previous equation becomes:

$$\frac{n_\mathrm{i}}{\mu_\mathrm{i}}(|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}| - |\boldsymbol{\mathcal{E}}_{0\mathrm{r}}|)\cos\theta_\mathrm{i} = \frac{n_\mathrm{t}}{\mu_\mathrm{i}}|\boldsymbol{\mathcal{E}}_{0\mathrm{t}}|\cos\theta_\mathrm{t} \tag{2.3.6}$$

Finally, assuming that $\mu_\mathrm{i} = \mu_\mathrm{t} = \mu_0$ and employing Eq. (2.3.3) we obtain:

$$\begin{aligned} r_s &= \frac{|\boldsymbol{\mathcal{E}}_{0\mathrm{r}}|_s}{|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}|_s} = \frac{n_\mathrm{i}\cos\theta_\mathrm{i} - n_\mathrm{t}\cos\theta_\mathrm{t}}{n_\mathrm{i}\cos\theta_\mathrm{i} + n_\mathrm{t}\cos\theta_\mathrm{t}}, \\ t_s &= \frac{|\boldsymbol{\mathcal{E}}_{0\mathrm{t}}|_s}{|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}|_s} = \frac{2n_\mathrm{i}\cos\theta_\mathrm{i}}{n_\mathrm{i}\cos\theta_\mathrm{i} + n_\mathrm{t}\cos\theta_\mathrm{t}}. \end{aligned} \tag{2.3.7}$$

The coefficients r_s and t_s are amplitude coefficients for the reflected and transmitted light. They are the perpendicular components of r and t for s-polarized light. Using Snell's law, that is $n_\mathrm{i}\sin\theta_\mathrm{i} = n_\mathrm{t}\sin\theta_\mathrm{t}$, the relations for r_s and t_s are simplified as follows:

$$\begin{aligned} r_s &= -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \\ t_s &= -\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)}. \end{aligned} \tag{2.3.8}$$

A similar argument for the p-polarized light leads to the calculation of the parallel components r_p and t_p of r and t. In case $\boldsymbol{\mathcal{E}}$ is parallel to the plane of incidence the amplitude coefficients are:

$$\begin{aligned} r_p &= \frac{n_t\cos\theta_i - n_i\cos\theta_t}{n_i\cos\theta_t + n_t\cos\theta_i}, \\ t_p &= \frac{2n_i\cos\theta_i}{n_i\cos\theta_t + n_t\cos\theta_i}, \end{aligned} \tag{2.3.9}$$

and their simplified relations are:

$$\begin{aligned} r_p &= \frac{\tan(\theta_i - \theta_t)}{\theta_i + \theta_t}, \\ t_p &= \frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t)}. \end{aligned} \tag{2.3.10}$$

Furthermore, it can be checked that

$$\begin{aligned} t_s - r_s &= 1, \\ t_p + r_p &= 1. \end{aligned} \tag{2.3.11}$$

The amplitude coefficients are shown in Fig. 2.9 for the case in which light travels from a less dense to a more dense medium ($n_i < n_t$), that is external reflection. In Fig. 2.10 the reflection coefficients are shown for the case in which $n_i > n_t$, that is internal reflection. Note from Fig. 2.9 that r_p approaches to 0 when θ_i approaches to θ_p and it gradually decreases reaching -1 for an incident angle $\theta_i = 90°$. The angle θ_p is called Brewster's angle or polarization angle as only the component perpendicular to the incident plane is reflected at that angle and therefore light is perfectly polarized. Similarly, Fig. 2.10 shows that $r_p = 0$ for $\theta_i = \theta_{p\prime}$. It can be show that $\theta_p + \theta_{p\prime} = 90°$. Both r_p and r_s reach 1 when $\theta_i = \theta_c$. θ_c is called the critical angle. Light that hits the incident plane with an incident angle equal to or greater than the critical angle is totally reflected back and no transmitted light is observed. This phenomenon is called total internal reflection.

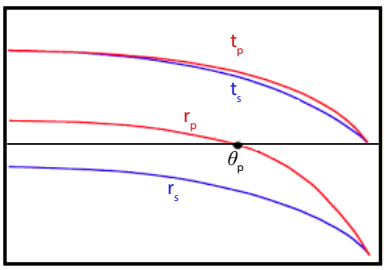


Figure 2.9: Amplitude coefficients of reflection and transmission as a function of the incident angle θ_i in the case of external reflection, i.e. $n_t < n_i$ ($n_t = 1$ and $n_i = 1.5$). θ_p is the polarization angle, [10].

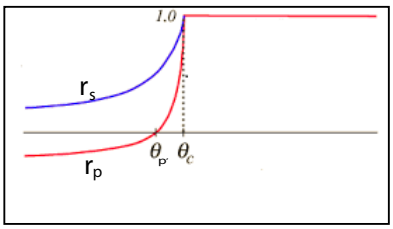


Figure 2.10: Reflection coefficients as a function of the incident angle θ_i in the case of internal reflection, i.e. $n_t > n_i$ ($n_t = 1.5$ and $n_i = 1$). $\theta_{p\prime}$ is the polarization angle and θ_c is the critical angle, [10].

The we introduce the Poynting vector $\boldsymbol{P}$ that defines the energy flux of an electromagnetic field. It is measured in [W/m^2], and it is given by:

$$\boldsymbol{P} = \frac{1}{\mu}\Big(\boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{B}}\Big), \tag{2.3.12}$$

where $\mu = \frac{1}{\varepsilon v^2}$ is the permeability and ε the permittivity of the medium. In the following, the parameters for vacuum are indicated with the subscript 0. All quantities defined in the media of the incident, reflective and transmitted light are indicated with the subscripts i, r and t, respectively. Optical rays are perpendicular to the wave front of an electromagnetic wave and parallel to the Poynting vector, [12]. The irradiance E is defined as the average energy that crosses in unit time a unit area A perpedicular to the direction of the energy flow. Therefore, defining the average of the vector $\boldsymbol{P}$

over the time as:

$$\langle \boldsymbol{P} \rangle_T = \frac{1}{T} \int_0^T \boldsymbol{P} \mathrm{d}T \tag{2.3.13}$$

we can write the irradiance E as:

$$\boldsymbol{E} = \langle \boldsymbol{P} \rangle_t = v\varepsilon |\boldsymbol{\mathcal{E}}|^2 \,. \tag{2.3.14}$$

Considering a beam of light that hits a surface such that an area A is illuminated, the incident, reflected and transmitted beams are $\boldsymbol{E}_{\mathrm{i}} A \cos\theta_{\mathrm{i}}$ $\boldsymbol{E}_{\mathrm{r}} A \cos\theta_{\mathrm{r}}$ and $\boldsymbol{E}_{\mathrm{t}} A \cos\theta_{\mathrm{t}}$, respectively. The reflectance $\mathcal{R}$ is the ratio of the reflected power to the incident power:

$$\mathcal{R} = \frac{|\boldsymbol{E}_{\mathrm{r}}| \cos\theta_r}{|\boldsymbol{E}_{\mathrm{i}}| \cos\theta_{\mathrm{i}}} = \frac{|\boldsymbol{\mathcal{E}}_{0\mathrm{r}}|^2}{|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}|^2} = r^2 \tag{2.3.15}$$

where the second equality holds because $v_{\mathrm{i}} = v_{\mathrm{t}}$, $\varepsilon_{\mathrm{i}} = \varepsilon_{\mathrm{t}}$ and $\theta_{\mathrm{i}} = \theta_{\mathrm{t}}$. Similarly, the transmittance $\mathcal{T}$ is the ratio between the transmitted to the incident power:

$$\mathcal{T} = \frac{|\boldsymbol{E}_{\mathrm{t}}| \cos\theta_{\mathrm{t}}}{|\boldsymbol{E}_{\mathrm{r}}| \cos\theta_{\mathrm{r}}} = \frac{n_{\mathrm{t}} \cos\theta_{\mathrm{t}}}{n_{\mathrm{t}} \cos\theta_{\mathrm{i}}} \frac{|\boldsymbol{\mathcal{E}}_{0\mathrm{t}}|^2}{|\boldsymbol{\mathcal{E}}_{0\mathrm{i}}|^2} = \frac{n_{\mathrm{t}} \cos\theta_{\mathrm{t}}}{n_{\mathrm{t}} \cos\theta_{\mathrm{i}}} t^2 \,. \tag{2.3.16}$$

Employing total energy conservation, that is:

$$\boldsymbol{E}_{\mathrm{i}} A \cos\theta_{\mathrm{i}} = \boldsymbol{E}_{\mathrm{r}} A \cos\theta_{\mathrm{r}} + \boldsymbol{E}_{\mathrm{t}} A \cos\theta_{\mathrm{t}}, \tag{2.3.17}$$

we can easily prove that:

$$\mathcal{R} + \mathcal{T} = 1 \,. \tag{2.3.18}$$

The parallel and perpendicular components of $\mathcal{R}$ and $\mathcal{T}$ are:

$$\begin{aligned} \mathcal{R}_p &= r_p^2, \\ \mathcal{T}_p &= \frac{n_t \cos\theta_t}{n_t \cos\theta_i} t_p^2, \\ \mathcal{R}_s &= r_s^2, \\ \mathcal{T}_s &= \frac{n_t \cos\theta_t}{n_t \cos\theta_i} t_s^2. \end{aligned} \tag{2.3.19}$$

it can be show that

$$\begin{aligned} \mathcal{R}_s + \mathcal{R}_p &= 1, \\ \mathcal{T}_s + \mathcal{T}_p &= 1 \,. \end{aligned} \tag{2.3.20}$$

For normal incidence, i.e. $\theta_i = 0$, there is no polarization and Eqs. (2.3.19) lead to:

$$\begin{aligned} \mathcal{R} = \mathcal{R}_p = \mathcal{R}_s &= \left(\frac{n_i - n_t}{n_t + n_i} \right)^2, \\ \mathcal{T} = \mathcal{T}_p = \mathcal{T}_s &= \frac{4 n_i n_t}{(n_t + n_i)^2} \,. \end{aligned} \tag{2.3.21}$$

Many common light sources such as sunlight, halogen lighting, LED spotlights, and incandescent bulbs produce unpolarized light. In case of unpolarized light the

amount of reflected and transmitted light is given by the average of reflectance $\mathcal{R}$ and transmittance $\mathcal{T}$ calculated considering first p-polarized light and then s-polarization, that is:

$$\begin{aligned} \mathcal{R} &= \frac{\mathcal{R}_p + \mathcal{R}_s}{2}, \\ \mathcal{T} &= \frac{\mathcal{T}_p + \mathcal{T}_s}{2}, \end{aligned} \tag{2.3.22}$$

where $\mathcal{R}_p$, $\mathcal{R}_s$, $\mathcal{T}_p$ and $\mathcal{T}_s$ are given in Eqs. (2.3.19).

With this overview we conclude this chapter. The notions given in Section 2.1 will be used in the entire thesis as our goal is to study the distribution of light at the target of some optical systems. In particular we will focus on the computation of the output intensity distribution. The reflection and refraction laws explained in Section 2.2 are needed to determine how the optical system changes the ray's direction every time that it hits a surfaces (or a line in the two-dimensional case). In Chapters 3, 4, 6, 7 and 8 only systems where the reflection and refraction laws play a role are considered. Systems with Fresnel reflection are treated in the last chapter. The amount of reflected and transmitted light is calculated using the Fresnel's equation (introduced in the last paragraph of this chapter). Since, we restrict ourselves to two-dimensional systems, the value of reflectance and transmittance will be computed using Eqs. (2.3.22).

Chapter 3

Ray tracing

Optical ray tracing is a tool to calculate the transport of light within optical systems. Given an optical system and a set of rays at the source, ray tracing relates the emitted light with its output distribution. The influence of diffraction on the transport of a ray is neglected.

Although the method can be implemented for two or more dimensions and for any optical system, here we consider the two-dimensional case only. From now on, we will thus refer to optical lines instead of optical surfaces. The two-dimensional case has limitations. For example, it may not identify skew rays that are turned back by the system, with the consequence that a 2D analysis cannot guarantee a proper treatment of non meridional rays in 3D. Nevertheless, the two-dimensional case is particularly relevant because it is a good test case to demonstrate the performance of new methods. Optical designers often start with 2D systems, where only the meridional plane is taken into account because it gives a good prediction of the target distribution of the rays (see [13], chapter 4, p.$50-65$).

3.1 Ray tracing for two-dimensional optical systems

Light rays are straight lines and they are reflected or refracted by the optical components. Every ray emitted from the source is followed until it reaches the target. The ray tracing procedure is constructed such that the position and the direction of the rays are calculated on every optical line that they hit.

Given a Cartesian coordinate system (x, z), a two-dimensional optical system symmetric with respect to the z-axis is defined. Hence, usually the optical axis coincides with the z-axis. The optical system is formed by a source S, a target T and some optical components labeled with indexes j where $j \in \{2, \cdots, \mathrm{Nl}-1\}$ and Nl indicates the number of lines that form the system. S and T are indicated with the indexes 1 and Nl, respectively. The index of refraction of the medium in which line j is located is indicated with n_j. Every ray emitted by S (line 1) can hit some optical components $j \in \{2, \cdots, \mathrm{Nl}-1\}$ before reaching T (line Nl). The intersection point of the rays with line j are $(x_j, z_j)_{j=1,\cdots,\mathrm{Nl}}$ and, $\boldsymbol{s}_j = (-\sin t_j, \cos t_j)$ indicates the direction vector of the rays that leave j, with t_j the angle that the ray forms with respect to the optical axis measured counterclockwise. As we consider only forward rays, the angles

$t_j \in (-\pi/2, \pi/2)$. Therefore, a ray segment between (x_j, z_j) and (x_k, z_k) with $k \neq j$ is parameterized in real space by:

$$\boldsymbol{r}(s) = \begin{pmatrix} x_j - s\sin t_j \\ z_j + s\cos t_j \end{pmatrix} \qquad 0 < s \leq s_{\max}\,, \tag{3.1.1}$$

where s denotes the arc-length and $s_{\max}$ is the maximum value that it can assume. Fig. 3.1 shows an example where a single ray is traced inside a very simple optical system, the so-called two-faceted cup. The light source $\mathsf{S}= [-a, a]$ (line 1) and the

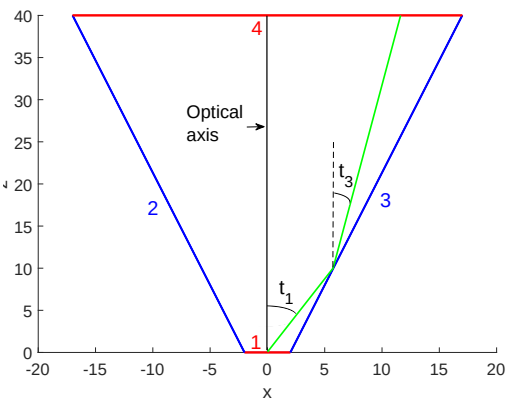


Figure 3.1: Shape of the two-faceted cup. Each line of the system is labeled with a number. The source $\mathsf{S}= [-2, 2]$ (line number 1) is located on the x-axis. The target $\mathsf{T}= [-17, 17]$ (line 4) is parallel to the source and is located at a height $z = 40$. The left and right reflectors (line 2 and 3) connect the source with the target.

target $\mathsf{T} = [-b, b]$ (line 4) are two segments normal to the z-axis, where $a = 2$ and $b = 17$. The left and right reflectors (line 2 and 3) are oblique segments that connect the source and the target. All the optical lines $j \in \{1, \cdots, 4\}$ are located in air, thus the refractive index is $n_j = 1$ for every j.

In order to compute the target photometric variables, we need to know how the optical system influences the direction of the rays when they hit an optical line. Ray tracing relates the position coordinates (x_1, z_1) and the direction vector $\boldsymbol{s}_1$ of every ray at the source S with the corresponding position $(x_{\mathrm{Nl}}, z_{\mathrm{Nl}})$ and direction $\boldsymbol{s}_{\mathrm{Nl}}$ at the target T. In the following we will often use the target coordinates of the rays thus, to simplify the notation, we do not write the subscript Nl for the target coordinates. Hence, we write (x, z) instead of $(x_{\mathrm{Nl}}, z_{\mathrm{Nl}})$, t instead of t_{Nl} and $\boldsymbol{s}$ instead of $\boldsymbol{s}_{\mathrm{Nl}}$ for the target coordinates. The ray tracing algorithm can be outlined as follows:

1. Given a ray that leaves S with initial position (x_1, z_1) and initial direction $\boldsymbol{s}_1 = (-\sin t_1, \cos t_1)$, use Eq. (3.1.1) to implement the ray parametrization $\boldsymbol{r}(s_1)$;

2. Compute the coordinates $(x_k, z_k)_{k=1\cdots,\mathrm{Nl}}$ of the intersection point of the parameterized ray $\boldsymbol{r}(s)$ with all the lines that it crosses

   a) if the shape of the lines is described by an analytical equation, the intersection points are determined analytically;

   b) if there is no analytic description for the optical lines, the intersections need to be determined using iterative methods;

3. Determine the closest line j that the forward ray encounters;
4. If $j = \mathrm{Nl}$ stop the procedure, the target ray's coordinates (x, z) and $\boldsymbol{s}$ are found.
5. Calculate the normal $\boldsymbol{\nu}_i$ to line i at the point (x_i, z_i);
6. Compute the new ray direction $\boldsymbol{s}_j$ of the ray that leaves line j at the point (x_i, z_i):
   a) if the incident line is a reflective line, $\boldsymbol{s}_j$ is given by Eq. (2.2.1);
   b) if the incident line is a refractive line, $\boldsymbol{s}_j$ is given by Eq. (2.2.3);
7. Restart the procedure from 1. for the ray that leaves line j instead of S. Consider as initial ray position coordinates (x_i, z_i) instead of (x_1, z_1) and as initial ray direction $\boldsymbol{s}_j = (-\sin t_j, \cos t_j)$ instead of $\boldsymbol{s}_1$.

The procedure explained above is repeated for every ray traced through the system, [14]. Once the target position and the direction of every ray traced are computed, the target photometric variables can be calculated using the definitions explained in the previous chapter, see section 2.1.

There are different ways to implement the ray tracing procedure. The efficiency of the ray tracing can be related with the distribution of the rays at the source. If the initial position and direction of the rays are chosen randomly we have Monte Carlo (MC) ray tracing. This is a very common method in non-imaging optics as it is very powerful and easy to implement. MC ray tracing will be explained in details in the next paragraph. If the rays are chosen from a so-called low discrepancy sequence we have the Quasi-Monte Carlo (QMC) ray tracing. This approach is discussed in Section 3.3.

3.2 Monte Carlo ray tracing

Before explain MC ray tracing we give a general introduction to the MC methods approximate computation of integrals. Given an interval $D = [\boldsymbol{a}, \boldsymbol{b}]$ with $\boldsymbol{a} = (a_1, \cdots, a_d)$ and $\boldsymbol{b} = (b_1, \cdots, b_d)$ elements of $\mathbb{R}^d$ such that $[\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times \cdots \times [a_d, b_d]$, a function $f : [\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^d \mapsto \mathbb{R}$ and a random variable $\boldsymbol{y} \in D$ with probability density function $\rho(\boldsymbol{y})$, the expected value of f with respect of ρ is

$$\mathbb{E}[f] = \int_D f(\boldsymbol{y})\rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y}. \tag{3.2.1}$$

If ρ is a uniform probability density function,

$$\mathbb{E}[f] = \int_D f(\boldsymbol{y})\rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \frac{1}{(\boldsymbol{b} - \boldsymbol{a})}\int_D f(\boldsymbol{y})\mathrm{d}\boldsymbol{y}. \tag{3.2.2}$$

Monte Carlo approximates Eq. (3.2.2) by

$$S_{\mathrm{N}}(f) = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i) \tag{3.2.3}$$

$\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}} \in D$ are indipendent samples of the density function ρ, [15]. According to the strong law of large numbers,

$$\Pr\Big(\lim_{\mathrm{N}\to\infty} \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i) = \mathbb{E}[f(\boldsymbol{y})]\Big) = 1. \tag{3.2.4}$$

Therefore,

$$\mathbb{E}[f] = \int_D f(\boldsymbol{y})\rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \approx \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i). \tag{3.2.5}$$

From the linearity of the expected values, it follows the obvious relation

$$\mathbb{E}[S_{\mathrm{N}}(f)] = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \mathbb{E}[f] = \mathbb{E}[f], \tag{3.2.6}$$

while the Bienaymé formula leads to

$$\mathrm{Var}(S_{\mathrm{N}}) = \mathrm{Var}\left(\frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i)\right) = \frac{1}{\mathrm{N}^2} \sum_{i=1}^{\mathrm{N}} \mathrm{Var}\big(f(\boldsymbol{y}_i)\big) \tag{3.2.7}$$

which can be applied because the variables $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}}$ are independent, [16], Chap. 6. Suppose that f has variance $\mathrm{Var}[f] = \mathbb{E}[(f - \mathbb{E}(f))]^2 = \sigma^2[f]$, Eqs. (3.2.6) and (3.2.7) give

$$\mathrm{Var}[S_{\mathrm{N}}(f)] = \mathbb{E}[(S_{\mathrm{N}}(f) - \mathbb{E}[S_{\mathrm{N}}(f)])^2] = \mathbb{E}[(S_{\mathrm{N}}(f) - \mathbb{E}[f])^2] = \sigma[f]^2/\mathrm{N}. \tag{3.2.8}$$

Let us denote the integration error with:

$$\mathrm{err}(f, S_{\mathrm{N}}) = \int_D f(\boldsymbol{y})\rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - S_{\mathrm{N}}(f) = \mathbb{E}[f] - S_{\mathrm{N}}(f), \tag{3.2.9}$$

then

$$\mathbb{E}\big[|\mathrm{err}(f, S_{\mathrm{N}})|\big] \leq \sqrt{\mathbb{E}\big[\mathrm{err}(f, S_{\mathrm{N}})^2\big]} = \frac{\sigma[f]}{\sqrt{\mathrm{N}}}, \tag{3.2.10}$$

where the inequality is true because

$$\begin{aligned} \mathbb{E}\big[|\mathrm{err}(f, S_{\mathrm{N}})|\big] &= \frac{1}{\mathrm{N}}\sqrt{\left(\sum_{\mathit{1}=1}^{\mathrm{N}} |\mathrm{err}(f, S_{\mathrm{N}})|\right)^2} \leq \frac{1}{\mathrm{N}}\sqrt{\mathrm{N}\sum_{\mathit{1}=1}^{\mathrm{N}} \big(\mathrm{err}(f, S_{\mathrm{N}})\big)^2} \\ &= \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathit{1}=1}^{\mathrm{N}} \big(\mathrm{err}(f, S_{\mathrm{N}})\big)^2} = \sqrt{\mathbb{E}\big[\mathrm{err}(f, S_{\mathrm{N}})^2\big]} \end{aligned} \tag{3.2.11}$$

and the equality follows from Eqs (3.2.8) and (5.3.5). Hence, the absolute value of the integration error is, on average, bounded by $\sigma[f]/\sqrt{\mathrm{N}}$, where $\sigma[f]$ is the standard deviation of f, [17]. It is very important to note that $\mathrm{err}(f, S_{\mathrm{N}})$ does not depend on the dimension d of f.

MC technique can be combined with the ray tracing procedure in order to compute

the light distribution at the target of an optical system. In MC ray tracing the position and the direction of every ray at the source are chosen randomly. In the two-dimensional case (d=2), for every ray we need to choose one position coordinate x_1 at the source and one angular coordinate t_1 at the target, while the z_1 coordinate of every ray at the source is always given (for instance, for the two-faceted cup in Fig. 3.1, $z_1 = 0$ for every ray). Therefore, given a set of random variables $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\} \in [\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^2$, the initial position coordinate x_1 of the k-th ray corresponds to the first component of the k-th random variable $\boldsymbol{y}_k$ and, the starting angular coordinate t_1 of the k-th ray corresponds to the second component of the k-th random variable $\boldsymbol{y}_k$. Next, rays with those random coordinates at S are traced from S to T and, a probabilistic interpretation of the output photometric variables is provided. In particular, we are interested in the total target intensity I which is computed as a function of the angular coordinate t. The MC intensity is calculated dividing the target into intervals of equal length, the so-called bins. A partitioning $P_1 : -\pi/2 = t_0 < t_1 < \cdots < t_{\mathrm{Nb}} = \pi/2$ of the interval $[-\pi/2, \pi/2]$ is defined where Nb is the number of bins in P_1. We remark that, with a slight abuse of notation, we indicated the angular coordinates of the rays at the target (line Nl) with t_j instead of $t_{\mathrm{Nl},j}$ for every $j \in \{0, \cdots, \mathrm{Nb}\}$.

The normalized approximated intensity $I_{\mathrm{MC}}(t)$ is a piecewise constant function, whose value over the j-th bin is the ratio between the number of rays that fall into that bin $\mathrm{Nr}[t_{j-1}, t_j)$ and the total number of rays traced $\mathrm{Nr}[-\pi/2, \pi/2)$. Hence, I_{MC} is defined by

$$I_{\mathrm{MC}}(t) = \frac{\mathrm{Nr}[t_{j-1}, t_j)}{\mathrm{Nr}[-\pi/2, \pi/2)} \qquad \text{for } t \in [t_{j-1}, t_j). \tag{3.2.12}$$

The output intensity is computed from the value of the intensity $I_{\mathrm{MC}}(t_{j-1/2})$ along the direction $t_{j-1/2} = (t_{j-1} + t_j)/2$ for every bin $[t_{j-1}, t_j)_{j=1,\cdots,\mathrm{Nb}}$. The intensity $I_{\mathrm{MC}}(t_{j-1/2})$ gives an estimate of the probability that a ray reaches the target with an angle in the j-th interval $[t_{j-1}, t_j)$ of the partitioning P_1. This probability $\mathrm{P}_{j,\Delta t}$ is given by

$$\mathrm{P}_{j,\Delta t} = \Pr(t_{j-1} \le t < t_j) = \frac{\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t}{\int_{-\pi/2}^{\pi/2} I(t)\mathrm{d}t}\,, \tag{3.2.13}$$

where $I(t)$ is the output intensity (not normalized). Note that $\sum_{j=1}^{\mathrm{Nb}} \mathrm{P}_{j,\Delta t} = 1$. From the mean value theorem for the function $I(t)$, continuous in $[t_{j-1}, t_j]$, there exists a value $t_k \in [t_{j-1}, t_j]$ for which the integral at the numerator of the previous equation can be written as

$$\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t = \Delta t\; I(t_k). \tag{3.2.14}$$

Hence, $\mathrm{P}_{j,\Delta t}$ is proportional to the size $\Delta t = (t_{\mathrm{Nb}} - t_0)/\mathrm{Nb}$ of the intervals and to $I(t_k)$. Although t_k does depend on the number of bins Nb, $I(t_k)$ is constant as it is the value of the intensity on a given direction, so Eq. (3.2.14) proves that $\mathrm{P}_{j,\Delta t}$ is inversely proportional to the number of bins Nb of the partitioning P_1. Indicating with $\Phi = \int_{-\pi/2}^{\pi/2} I(t)\mathrm{d}t$ the total flux (measured in lumen [lm]), the error between the

intensity $I(t_{j-1/2})$ and the averaged MC intensity $\Phi I_{\mathrm{MC}}(t_{j-1/2})/\Delta t$ is given by

$$\begin{aligned}\left|I(t_{j-1/2}) - \frac{\Phi}{\Delta t} I_{\mathrm{MC}}(t_{j-1/2})\right| \leq \\ \left|I(t_{j-1/2}) - \frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t\right| + \\ \frac{1}{\Delta t}\left|\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t - \Phi\, I_{\mathrm{MC}}(t_{j-1/2})\right|.\end{aligned} \tag{3.2.15}$$

The first term of the right hand side of inequality (3.2.15) gives an estimate of how much the averaged intensity $\frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t$ differs from the exact intensity $I(t_{j-1/2})$. This term is due to the discretization of the target and therefore it depends on the number of bins Nb considered. Substituting $I(t)$ with its Taylor expansion around the point $t_{j-1/2}$ we obtain that this term is proportional to the square of the size of the bins. Therefore,

$$\left|I(t_{j-1/2}) - \frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t\right| = \mathrm{C}_1/\mathrm{Nb}^2 \tag{3.2.16}$$

with $C_1 > 0$ a certain constant.

The second part of the right hand side of inequality (3.2.15) gives an estimate of the MC error and therefore it depends also on the number of rays traced. In order to show how this term decreases as a function of the number of rays traced, we define the random variable $X_j(t)$ as the variable that is equal to 1 if the ray with angular coordinate t is inside the interval $[t_{j-1}, t_j)$ and equal to 0 otherwise:

$$X_j(t) = \begin{cases} 1 & \text{if} \quad t \in [t_{j-1}, t_j), \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.17}$$

The Bernoulli trial X_j follows a binomial distribution $B(1, \mathrm{P}_{j,\Delta t})$. Considering a sample of Nr rays, the variable $Y_j = \sum_{k=1}^{\mathrm{Nr}} X_j(t_k)$ follows a binomial distribution $B(\mathrm{Nr}, \mathrm{P}_{j,\Delta t})$, where t_k is the angle that the k-th ray forms with the optical axis. Then, using the de Moivre-Laplace theorem, we conclude that the variable Y_j is approximated by a normal distribution with mean value $E[Y_j] = \mathrm{NrP}_{j,\Delta t}$ and variance $\sigma^2[Y_j] = \mathrm{NrP}_{j,\Delta t}(1 - \mathrm{P}_{j,\Delta t})$ when a large number of rays is considered, see [18, 19]. Thus, the normalized intensity along the direction $t_{j-1/2}$ is

$$I_{\mathrm{MC}}(t_{j-1/2}) = \sum_{k=1}^{\mathrm{Nr}} X_j(t_k)/\mathrm{Nr}. \tag{3.2.18}$$

The mean value $E[I_{\mathrm{MC}}(t_{j-1/2})] = \mathrm{P}_{j,\Delta t}$ and the variance $\sigma^2[I_{\mathrm{MC}}(t_{j-1/2})] = \mathrm{P}_{j,\Delta t}(1 - \mathrm{P}_{j,\Delta t})/\mathrm{Nr}$. Note that the standard deviation $\sigma_j := \sigma[I_{\mathrm{MC}}(t_{j-1/2})]$ is approximated by

$$\sigma_j = \sqrt{\mathrm{P}_{j,\Delta t}(1 - \mathrm{P}_{j,\Delta t})/\mathrm{Nr}} \approx \frac{\mathrm{C}_2}{\sqrt{\mathrm{NbNr}}}, \tag{3.2.19}$$

for some $\mathrm{C}_2 > 0$. σ_j can be used to give an estimate of the difference between the intensity $I_{\mathrm{MC}}(t_{j-1/2})$ and its mean value $\mathrm{P}_{j,\Delta t}$. Therefore, the second term of the

right hand side of relation (3.2.15) becomes

$$\begin{aligned}\frac{1}{\Delta t}\Big|\int_{t_{j-1}}^{t_j} I(t)\mathrm{d}t - \Phi\, I_{\mathrm{MC}}(t_{j-1/2})\Big| = \\ \frac{\Phi}{\Delta t}\Big|\mathrm{P}_{j,\Delta t} - I_{\mathrm{MC}}(t_{j-1/2})\Big| \approx \\ \frac{\Phi}{\Delta t}\sigma_j[I_{\mathrm{MC}}(t_{j-1/2})] \approx \mathrm{C}_3\frac{\mathrm{Nb}}{\sqrt{\mathrm{NbNr}}} = \mathrm{C}_3\sqrt{\frac{\mathrm{Nb}}{\mathrm{Nr}}}\,,\end{aligned} \tag{3.2.20}$$

for some $\mathrm{C}_3 > 0$, where the approximation holds because σ_j gives a measure for the error between $I_{\mathrm{MC}}(t_{j-1/2})$ and the probability $\mathrm{P}_{j,\Delta t}$, [20]. The second approximation follows from (3.2.19). The MC error over the j-th bin is estimated by

$$\Big|I(t_{j-1/2}) - \frac{\Phi}{\Delta t}I_{\mathrm{MC}}(t_{j-1/2})\Big| = \frac{\mathrm{C}_1}{\mathrm{Nb}^2} + \mathrm{C}_4\sqrt{\frac{\mathrm{Nb}}{\mathrm{Nr}}}, \tag{3.2.21}$$

for $\mathrm{C}_4 > 0$. Considering a fixed number of bins, we obtain that the minimal error is reached when $\mathrm{Nr} \approx \mathrm{Nb}^5$. Hence, if we double the number of bins we need to trace 2^5 times more rays.

We conclude this chapter implementing MC ray tracing for the two-faceted cup the profile of which is depicted in Fig. 3.1. Considering a set of $\mathrm{Nr} = 10^3$ random rays at the source, we obtain an example of the rays distribution on the (x,t)-plane shown in Fig. 4.9a. Since the rays are chosen randomly, the distribution at the source could be different from the one shown in that figure.

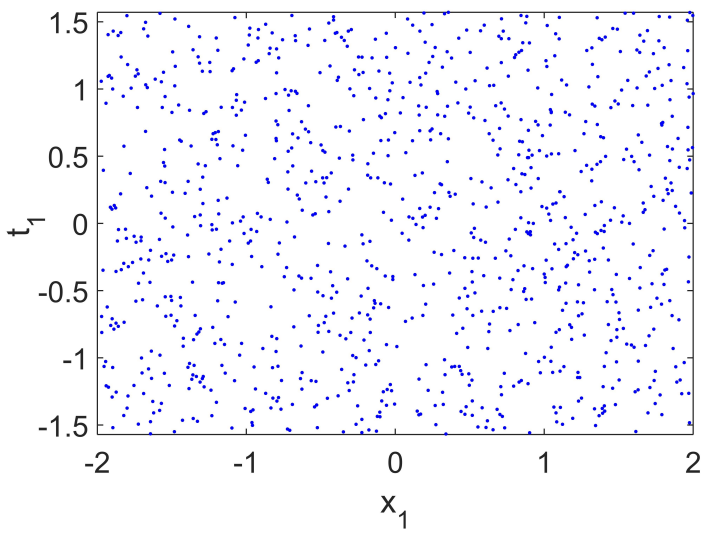


Figure 3.2: Rays at the source of the two-faceted cup with random position coordinate x and random angular coordinates t. 10^3 rays are depicted in this figure.

Then, every sample ray is traced inside the system using the ray tracing procedure.

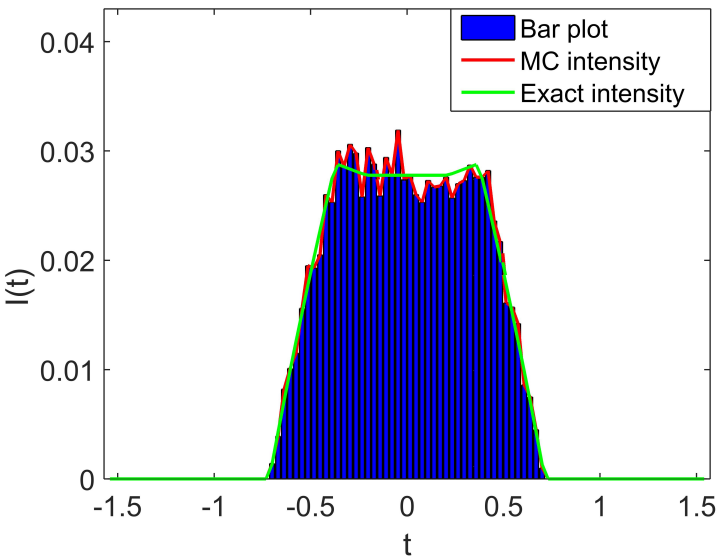


Figure 3.3: Comparison between the averaged normalized MC intensity and the normalized exact intensity.

The target $\mathsf{T} = [-b, b]$ is divided into Nb $= 100$ bins. Using Eq. (3.2.12), the normalized intensity I_{MC} is computed. I_{MC} is a piecewise constant function, therefore the averaged normalized intensity $\hat{I}_{\mathrm{MC}}(t_{j-1/2})$ is given considering the values that the intensity I_{MC} assumes on the middle point $(t_{j-1/2})_{j=0,\cdots,\mathrm{Nb}}$ of every bin of the partitioning P_1. The profile of $\hat{I}_{\mathrm{MC}}$ is depicted in Fig. 3.3 with the red line. The exact intensity (analytic intensity) is shown with the green line in the same figure. MC ray tracing has the advantages of being very easy to implement and it does not require too much regularity of the function that has to be approximate. Furthermore, the error convergence does not depend on the dimension of the domain in which the function is defined. On the other hand, MC method is time consuming as the error, for a fixed number of bins, has a speed of convergence of order $O(1/\sqrt{\mathrm{Nr}})$. Thus, to decrease the error of a factor 10 we need to increase the number of rays of a factor 100. As, MC ray tracing is a binning procedure, the error depends also on the number of bins in which the target is divided. It is a statistical procedure and the error bound is only a *probabilistic* error as shown in Eq. (3.2.10). This means that, to calculate the value of the error, several simulations have to be repeated and the average of the errors obtained in every simulation has to be calculated.

Instead of considering random variables, the sample of rays can be defined in such a way that they have a regular distribution on the domain $D \subseteq \mathbb{R}^d$ of the function f of which we want to compute the integral. Methods based on this deterministic approach are called Quasi Monte Carlo (QMC) methods. They can be seen as an improvement of MC method.

3.3 Quasi-Monte Carlo ray tracing

Quasi-Monte Carlo (QMC) methods were proposed for the first time in the 1950s in order to speed up MC. Likewise MC methods, QMC procedures can be used to approximate the integral of a function.

This chapter provides basic notions about uniform distributed theory, it follows Chapter 2 of [17]. It is useful to restrict ourselves to intervals of the form $[\boldsymbol{a}, \boldsymbol{b}) \subseteq [0,1)^d$ and introduce the concept of sequences uniformly distributed modulo one.

Definition 3.3.1. An infinite sequence $\{y_n\}_{n\in\mathbb{N}_0} \in [0,1)^d$ is said to be *uniformly distributed modulo one* (or equidistributed), if for every interval $[\boldsymbol{a}, \boldsymbol{b}) \subseteq [0,1)^d$ it holds

$$\lim_{\mathrm{N}\to\infty} \frac{\mathrm{card}(A([\boldsymbol{a}, \boldsymbol{b}), \mathrm{N}))}{\mathrm{N}} = \lambda_d([\boldsymbol{a}, \boldsymbol{b})) \tag{3.3.1}$$

where $\mathrm{card}(A([\boldsymbol{a}, \boldsymbol{b}), \mathrm{N}))$ is the cardinality of the following set

$$A([\boldsymbol{a}, \boldsymbol{b}), N) = \{n \in \mathbb{N}_0 : 0 \leq n \leq \mathrm{N} - 1 \text{ and } y_n \in [\boldsymbol{a}, \boldsymbol{b})\}, \tag{3.3.2}$$

and $\lambda_d([\boldsymbol{a}, \boldsymbol{b})) = \prod_{j=1}^{d}(b_j - a_j)$ is the d-dimensional Lesbegue measure of the interval $[\boldsymbol{a}, \boldsymbol{b})$.

Given a sequence $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}} \in [0,1)^d$ uniformly distributed modulo one and a Riemann integrable function $f : [0,1]^d \mapsto \mathbb{R}$, the integral of f can be approximate as the average of the values that f assumes on $\{\boldsymbol{y}_i\}$ for every $j = \{1, \cdots, \mathrm{N}\}$, that is:

$$\lim_{\mathrm{N}\to\infty} \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} f(y_i) = \int_{[0,1]^d} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}. \tag{3.3.3}$$

The idea of QMC methods is to generate the set of points in $[\boldsymbol{a}, \boldsymbol{b}]$ such that they are not randomly distributed but also not exactly uniformly distributed. To measure how much the distribution of these points differs from a uniform distribution, the concept of discrepancy was introduced. Intuitively, discrepancy measures how much the samples differ from a uniform distribution. Therefore, random sequences have a very high discrepancy, while uniform distributed sequences have zero discrepancy. The definition of discrepancy in more mathematical terms is provided below.

Definition 3.3.2. Given a set $\mathcal{S} = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\mathrm{N}\}$ of N points in $[0,1)^d$. The discrepancy $D_\mathrm{N}(\mathcal{S})$ of $\mathcal{S}$ is defined as

$$D_\mathrm{N}(\mathcal{S}) = \sup_{\boldsymbol{a},\boldsymbol{b}\in[0,1)^d} \left| \frac{\mathrm{card}(A([\boldsymbol{a}, \boldsymbol{b}), \mathrm{N}))}{\mathrm{N}} - \lambda_d([\boldsymbol{a}, \boldsymbol{b})) \right| \tag{3.3.4}$$

where $\lambda_d([\boldsymbol{a}, \boldsymbol{b})) = \prod_{j=1}^{d}(b_j - a_j)$ is the d-dimensional Lesbegue measure of the interval $[\boldsymbol{a}, \boldsymbol{b})$.

Often, it is enough to consider the discrepancy in the intervals $[0, \boldsymbol{a}) \subseteq [0,1)^d$, in that case we talk about star discrepancy.

Definition 3.3.3. Let $\mathcal{S} = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{\mathrm{N}}\}$ be a set of N points in $[0,1)^d$. The star discrepancy $D^*_{\mathrm{N}}(\mathcal{S})$ of $\mathcal{S}$ is defined as:

$$D^*_{\mathrm{N}}(\mathcal{S}) = \sup_{\boldsymbol{a}\in[0,1)^d} \left| \frac{\mathrm{card}(A([0,\boldsymbol{b}),\mathrm{N}))}{\mathrm{N}} - \lambda_d([0,\boldsymbol{a})) \right| \tag{3.3.5}$$

Sequences constructed such that the corresponding star discrepancy has an order of $O(\log(\mathrm{N})^d/\mathrm{N})$ are called *low-discrepancy sequences*, [15]. An important results shows that, using a low-discrepancy sequence $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}}$, the absolute error of a QMC algorithm:

$$\mathrm{err}(f, S_{\mathrm{N}}) = \left| \int_{[0,1)^d} f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i) \right| \tag{3.3.6}$$

can be bounded by the product of a term that depends on f and another term that depends on the discrepancy of the set $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}}$. This is the result provided by the Koksma-Hlawka inequality which gives the following estimation of the error:

$$\left| \int_{[0,1)^d} f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} f(\boldsymbol{y}_i) \right| \leq V(f) D^*_{\mathrm{N}}(\mathcal{S}) \tag{3.3.7}$$

where $V(f)$ is the so-called variation function of f in the sense of Hardy-Krause (see [21, 22] for details). From the definition of low-discrepancy sequences and from the Koksma-Hlawka inequality we can state that:

$$\mathrm{err}(f, S_{\mathrm{N}}) < C\frac{\log(\mathrm{N})^d}{\mathrm{N}}. \tag{3.3.8}$$

For small dimensions d, QMC performs much better than MC methods, while for large dimension d the factor $\log(\mathrm{N})$ could be very big. The convergence of QMC method depends on the of low-discrepancy sequence that is used.

There are many ways to generate low-discrepancy sequences. The most common QMC approach uses the so-called Sobol' sequence. The algorithm for generating Sobol' sequences is widely explained in the literature, (see for instance , [23]). In appendix A we give an overview of how these kind of sequences can be constructed.

Based on QMC methods, QMC ray tracing considers as position and angular coordinates of the rays at the source, the coordinates of the corresponding points of a low-discrepancy sequence. Therefore, to implement QMC ray tracing in two-dimensions we need to construct a low-discrepancy sequence in two-dimensions. Given, for instance, a Sobol' sequence $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}}$ with $\boldsymbol{y}_i \in [0,1)^2$ for every $i = 1, \cdots, \mathrm{N}$, the two dimensional QMC ray tracing consider the position coordinate of the i-th ray at the source equal to the first component of the i-th point $\boldsymbol{y}_i$ of the Sobol' sequence $\{\boldsymbol{y}_i\}_{i=1,\cdots,\mathrm{N}}$ and, the direction coordinate of the i-th ray at the source equal to the second component of the i-th point $\boldsymbol{y}_i$ of the same sequence. A set of $\mathrm{Nr} = \mathrm{N}$ rays with these initial coordinates is traced within the system and, once the target coordinates of all the rays traced are computed, the output intensity is calculated using the same approach used for MC ray tracing, see Eqs 3.2.12 and 3.2.15. The difference between MC and QMC ray tracing consists only on the choice of the initial ray set.

In Fig. 4.9b we show the distribution of the position and direction coordinates of

the rays at the source of the two-faceted cup in Fig. 3.1. A set of 10^3 rays generated from a $2D$ Sobol sequence is considered, the coordinates (x_1, t_1) of every ray at the source are depicted with blue dots. We note that the rays have a regular distribution on the (x, t)-plane. We need to remark that, for the system in Fig. 3.1, $x_1 \in [-2, 2]$ and the angular coordinates $t_1 \in [-\pi/2, \pi/2]$. Since Sobol' sequences are defined inside intervals of the $[\boldsymbol{a}, \boldsymbol{b}) \subseteq [0, 1)^2$, we scaled the points of the sequence $\boldsymbol{y}_i$ in order to take all the possible positions and directions that the rays tan assume at the source.

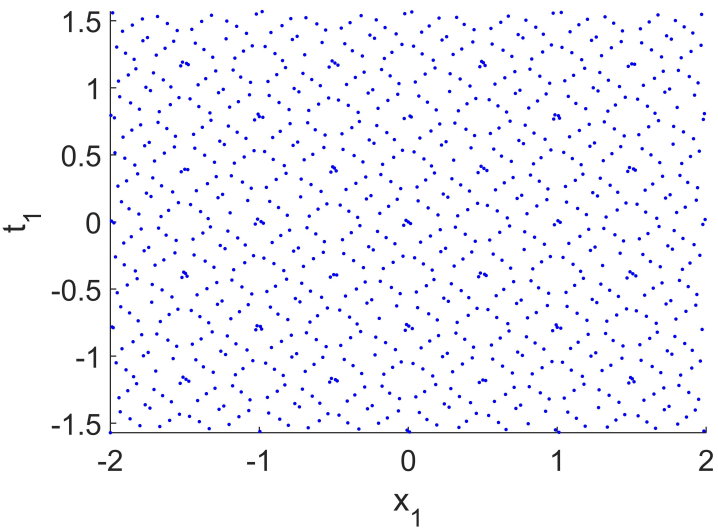


Figure 3.4: 10^3 rays at the source of the two-faceted cup with position x_1 and angular t_1 coordinates with a regular distributions. They are distributed as the points of a Sobol' sequence in two-dimensions.

Dividing the target into Nb = 100 bins, we computed the target intensity. In Fig. 3.5 we show the profile of the output intensity at the target of the two-faceted cup computed using QMC ray tracing with 10^4 rays. The QMC intensity is depicted with the red line. It is compared to the analytic intensity shown in the same figure with the green dotted line. A comparison between Fig. 3.3 and 3.5 gives the insight that for the two-faceted cup and for a set of Nr = 10^4 rays, QMC ray tracing performs better than MC ray tracing. In order to show the accuracy obtained using MC and QMC methods, we calculate the target intensity gradually increasing the number of rays traced inside the two-faceted cup. The error between the approximates intensity and the analytic intensity is calculated for every sample of rays. The speed of convergence for MC is shown in Fig. 3.6 with the red, while the behavior of QMC ray tracing is depicted in the same picture with the blue line. The results shown for a simple optical system are indeed consistent with what we expected from the theoretical analysis. Although QMC ray tracing is an improvement of MC ray tracing for small dimensions, it has two main disadvantages. First, its convergence is strongly related with the dimension in which it is implemented. Second, likewise MC ray tracing, QMC ray

tracing is a binning procedure, therefore the error still depends on the number of bins in which the target is divided and only the averaged value of the intensity over every bin is provided.

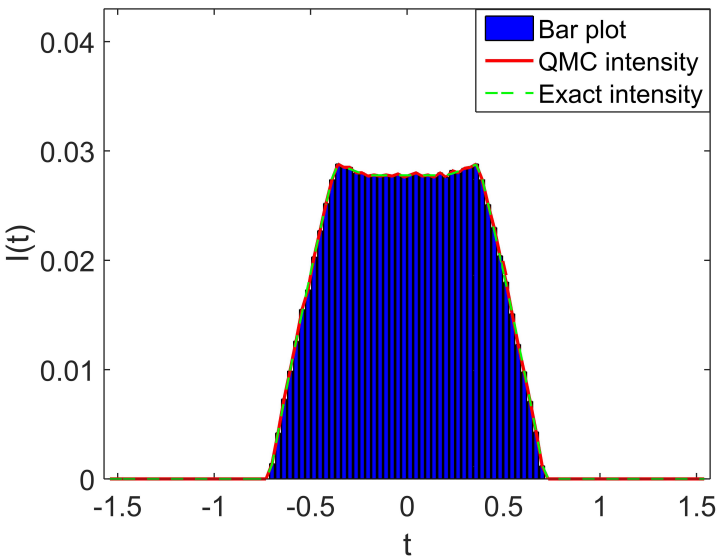


Figure 3.5: QMC intensity for the two-faceted cup obtained tracing Nr = 10^4 rays and dividing the target into Nb = 100 bins.

From the results provided in this chapter we can conclude that the choice of the initial ray set can make a big impact on the performance of the ray tracing procedure. Based on the idea of taking a smart choice of the initial ray set, we develop a new ray tracing method which is based on phase space. The phase space (PS) concept will be introduced in the next chapter. The new ray tracing method employs the PS of the source and the target of the optical systems. We will show in this thesis that phase space ray tracing allows to trace only few rays inside the system to obtain the desired accuracy of the target intensity.

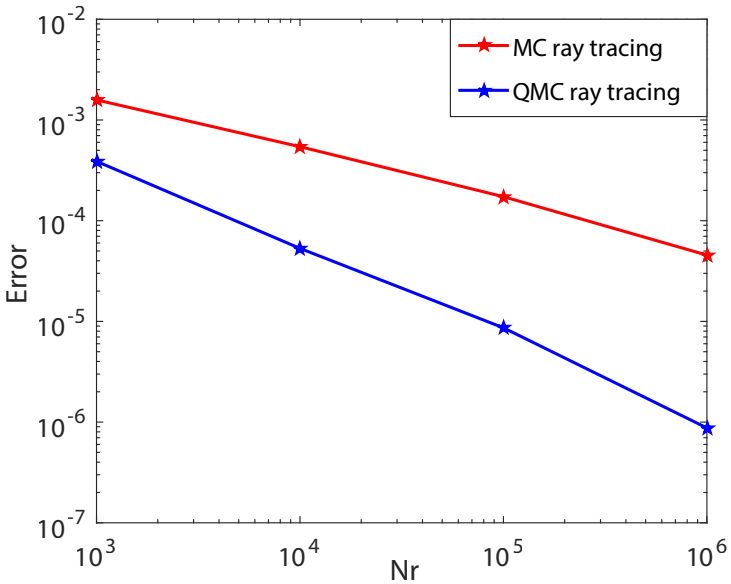


Figure 3.6: Error as function of the number of rays traced in a logarithmic scale for fixed number of bins Nb = 100. MC ray tracing convergence is of the order $O(1/\sqrt{\mathrm{Nr}})$ and it is shown with the red line. QMC ray tracing convergence is of the order $O(1/\mathrm{Nr})$ and it is depicted with the blue line.

Chapter 4

Ray tracing on phase space

Ray tracing on phase space is a method which employs the phase space (PS) of the source and the target of the optical system. Moreover, it takes into account the trajectory that every ray follows during its propagation. Before explaining the method, we introduce the PS concept.

4.1 Phase space

The PS of a three-dimensional systems is a four-dimensional space, indeed every ray is described by two position coordinates and two direction coordinates. The two position coordinates are given by two of the coordinates of the intersection point of the ray with the surface, while the two direction coordinates are the momentum coordinates of the vector tangent to the ray projected on the optical surface, [24].

For two-dimensional systems every ray in PS is given by a point in a two-dimensional space. Given an optical line j, the ray position coordinate on PS is the x-coordinate of the intersection point between the ray and the line j. The direction coordinate is the sine of the angle that the ray forms with respect to the normal $\boldsymbol{\nu}$ of line j which is always directed inside the same medium in which the incident ray travels, multiplied by the index of refraction n. We indicate the PS with $S{=}Q{\times}P$, where Q is the set of the position coordinates q and P is the set of the direction coordinates $p = n\sin\theta$, with $\theta \in [-\pi/2, \pi/2]$ the angle between the ray segment inside the system and the normal measured counterclockwise. In the following, the phase space is considered only for the source S and the target T and for no other line of the optical system. The coordinates of every ray on S and T are indicated with (q_1, p_1) and (q, p), respectively.

As an example, in Figures 5.4 and 5.6 we show the source and target PS of the two-faceted cup (in Figure 3.1), respectively, sampled with 10^4 randomly traced rays. The coordinates of every point in Figure 5.4 correspond to the position and direction coordinates of a ray at the source, while the coordinates of every point in Figure 5.6 correspond to the position and direction coordinates of a ray at the target, which are calculated using the ray tracing procedure. Furthermore, we store the path that every ray follows, where we refer to a path as the sequence of the lines encountered by the ray. In Figures 5.4 and 5.6 a color is associated to every path, hence all the rays that follow the same path are depicted with the same color. We note that the

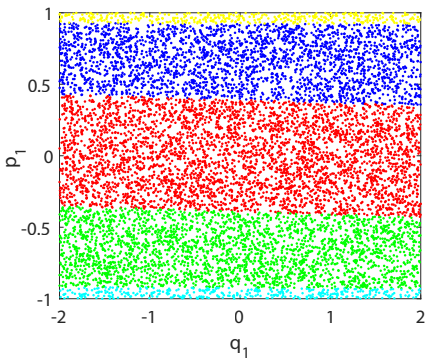


Figure 4.1: Source PS of the two-faceted cup. Five different paths can occur.

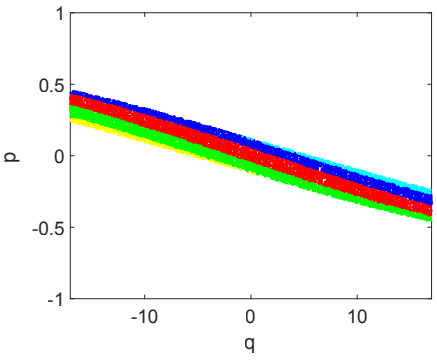


Figure 4.2: Target PS of the two-faceted cup. Five different paths can occur

source and target phase spaces are partitioned into different regions according to the path Π followed by the rays. Given a path Π, the corresponding regions are indicated with $R_1(\Pi)$ and $R(\Pi)$ at the source and the target PS, respectively. Rays that propagate through the two-faceted cup can follow 5 different paths. Some rays are emitted from the source and arrive at the target without hitting any other line, they follow path $\Pi_1 = (1, 4)$. These rays are depicted in red in the PS pictures. Some other rays can hit the left or the right reflector (line 2 and 3, respectively) once, their corresponding paths are $\Pi_2 = (1, 2, 4)$ and $\Pi_3 = (1, 3, 4)$, respectively. These rays are the blue and green dots in PS. Finally, there is the possibility that the rays have two reflections before hitting the target. They follow either path $\Pi_4 = (1, 2, 3, 4)$ or path $\Pi_5 = (1, 3, 2, 4)$ and they are depicted with the yellow and cyan points.

For the two-faceted cup all light emitted by the source arrives at the target. In order to derive the photometric variables at the target we need to understand where light ends up, i.e. which parts of the target PS are illuminated by the source. Indeed, while the source PS is completely covered by rays, some parts of the target PS are not reached by any ray at all, that is

$$\begin{aligned} S &= \bigcup_{\Pi} R_1(\Pi), \\ T &\supset \bigcup_{\Pi} R(\Pi), \end{aligned} \tag{4.1.1}$$

where the union is over all the possible paths. This means that, while the luminance at the source PS is positive for any possible position and direction, the luminance at the target PS is positive only inside the regions $R(\Pi)$, for every path Π, and it is equal to 0 outside those regions. For this reason, from now on we will refer to $R(\Pi)$ as the regions with positive luminance.

It is very important to remark that, although S and T have a different ray distribution, the area covered by the rays is conserved. This follows from étendue conservation. From (2.1.19), using $\mathrm{d}p = n\cos(\theta)\mathrm{d}\theta$, we rewrite the two-dimensional étendue

in source PS as:

$$U = \int_Q \int_P \mathrm{d}q\,\mathrm{d}p. \tag{4.1.2}$$

Therefore, in two dimensions, étendue can be seen as an area in PS. Etendue conservation leads to the conservation of the areas of regions with positive luminance.

For the two-faceted cup in Figure 3.1 we have that the étendue at the source is $U = 8$ (see Figure 5.4). Computing the total area covered by the regions with positive luminance at the target using the tétendue trapezoidal rule, we proved that it is equal to $U = 8$. From this follows a fundamental principle in non-imaging optics which is referred to as "the edge-ray principle". A literature overview of this principle is given in the next paragraph.

4.2 The edge-ray principle

The goal in non-imaging optics is to transfer all light from the source aperture to the output aperture. Systems that satisfy this property are referred to *ideal optical systems*. Several methods to design ideal optical systems are based on the edge-ray principle, [25, 26]. Basically it states that all the light rays exiting the edges of the source will end at the edges of the target. This guarantees that all light emitted from the source will arrive at the receiver, see Figure 4.3.

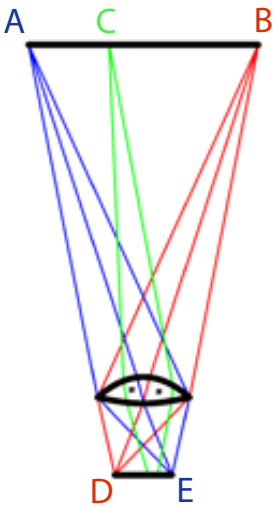


Figure 4.3: A lens that receives light from a source AB and redirects it to the receiver DE. Rays that leave the edges of the source hit the edges of the target (blue and red rays). Rays coming from the interior of the source will end at the interior of the target (green rays), [27].

In 1985 Miñano proved the principle by using the PS of the source and the target of an optical system, [28, 29]. He proved the principle for systems in inhomogeneous media, where the index of refraction is a continuous function, so the map that connects

the source and target phase spaces is a continuous map. Indicating with $\mathrm{M}(P)$ the optical map of a point P, Miñano showed that if $\mathrm{M}(\partial S) = \mathrm{M}(\partial T)$ then $\mathrm{M}(S) = \mathrm{M}(T)$ and vice versa. Note that the trajectory of two rays in PS cannot cross. The first version of the edge-ray principle [29] can be enunciated in two-dimensions as follows:

Lemma 4.2.1. *Edge-ray principle (version1)*
Suppose that:

a) There are two regions R_1 *and* R *in PS with the same area such that*

$$\mathrm{M}(\partial \mathrm{R}_1) = \mathrm{M}(\partial \mathrm{R});$$

b) The refractive-index distribution n is a continuous function;

Then, the following relation holds:

$$\mathrm{M}(\mathrm{R}_1) = \mathrm{M}(\mathrm{R}).$$

The second assumption in the previous lemma implies that the optical map is continuous in PS. However, for some optical systems, as for instance the compound parabolic concentrator (CPC), the ray mapping in phase space is not continuous. This is due to multiple reflections that rays can encounter with the reflectors. This implies that some rays at the edge of the source could not be mapped into rays at the edges of the target [30].

In 1994 Ries and Rabl reformulated the edge-ray principle such that it is valid for all systems even if the ray map in PS is not continuous [31]. Suppose that $R_1(\Pi)$ and $R(\Pi)$ are the regions, corresponding to path Π, at the source and the target PS, respectively. They showed that, for a given path Π, if the boundaries $\partial R_1(\Pi)$ are mapped into the boundaries $\partial R(\Pi)$, then also the regions $R_1(\Pi)$ are mapped into the regions $R(\Pi)$. Then, to map S to T it is necessary and sufficient that the first version of the edge ray principle is observed for all part of S and T defined by the number of reflections [31].

Lemma 4.2.2. *Edge-ray principle (generalized version)*
Let indicate every possible path with $(\Pi_j)_{j=1,\cdots,Np}$*, where Np is the number of all possible paths. Every possible path correspond to a certain number of reflections. Let denote with* $\mathrm{R}_1(\Pi_j)$ *and* $\mathrm{R}(\Pi_j)$ *the regions at* S *and associated to path* Π_i *such that they are a partition of* S *and* T*, that is:*

$$\mathrm{S} = \bigcup_{j=1}^{Np} \mathrm{R}_1(\Pi_i), \ \textit{with } \mathrm{R}_1(\Pi_j) \cap \mathrm{R}_1(\Pi_i) = \emptyset \textit{ for } i \neq j$$

$$\mathrm{T} = \bigcup_{j=1}^{Np} \mathrm{R}(\Pi_i), \ \textit{with } \mathrm{R}(\Pi_j) \cap \mathrm{R}(\Pi_i) = \emptyset \textit{ for } j \neq i.$$

Then, to map a source region into a target, it is necessary and sufficient that the first version of the edge ray principle is observed for all parts of S *and* T *defined by the number of reflections:*

$$\mathrm{M}\big(\partial \mathrm{R}_1(\Pi_j)\big) = \partial \mathrm{R}(\Pi_j), \quad \forall j \in \{1, \cdots, Np\}.$$

Hence, the edge-ray principle constitutes a tool for designing ideal systems and, to this purpose, it is sufficient that the rays of $\partial R_1(\Pi)$ are transformed to the rays of $\partial R(\Pi)$ for every path Π [32].

Using the PS concept and the edge-ray principle we develop a new ray tracing method. A non-uniform distribution of the rays is provided by developing a triangulation refinement at the source PS which is explained in the next section. The triangulation refinement provides more rays close to the boundaries of the regions $R_1(\Pi)$ each of them is formed by the rays that follow the same path Π. Next, the boundaries $\partial R_1(\Pi)$ are approximated by using two different approaches explained in Chapter 5. For every path Π, the boundaries at the target $\partial R(\Pi)$ are obtained by mapping their corresponding boundaries ∂R_1 at the source.

4.3 Phase space ray tracing

PS ray tracing takes advantage of the fact that there exists an optical map $\mathrm{M} : S \mapsto T$ such that

$$\mathrm{M}(q_1, p_1) = (q, p), \tag{4.3.1}$$

for every $(q_1, p_1) \in S$. For very simple systems, like the two-faceted cup, it is possible to determine an analytic expression for M (as explained in Appendix B). This is not the case for most of the optical systems we deal with. In these cases it is necessary to implement ray tracing to calculate how light is distributed at the target. As mentioned in the previous paragraph, for some optical systems M is not even continuous. Nevertheless, given a path Π, the restriction of M to R_1 $\mathrm{M}(\Pi)$: $R_1(\Pi) \mapsto R(\Pi)$ which maps the region $R_1(\Pi)$ in S into the region $R(\Pi)$ in T is a continuous and bijective map. The edge ray principle guarantees that $\mathrm{M}(\Pi)$ maps $R_1(\Pi)$ onto $R(\Pi)$ preserving topological features. In particular, the boundary $\partial R_1(\Pi)$ is mapped onto the boundary $\partial R(\Pi)$. Employing the maps $\mathrm{M}(\Pi)$ for all the possible paths Π, the output light distribution is determined. Therefore, the photometric variables at the target can be calculated.

The luminance $L(q, p)$ at the target PS is given by:

$$\begin{aligned} &L(q, p) > 0 \text{ for } (q, p) \in R(\Pi) \text{ for some path } \Pi, \\ &L(q, p) = 0 \text{ otherwise.} \end{aligned} \tag{4.3.2}$$

Since the luminance is conserved along a ray and a Lambertian source is considered, it is constant inside the regions $R(\Pi)$. The target intensity along a given direction $p = \text{const}$ is computed through an integration of the target luminance $L(q, p)$ and it is defined in T by:

$$I_{\mathrm{PS}}(p) = \int_Q L(q, p) \mathrm{d}q. \tag{4.3.3}$$

Note that, while in the real space the intensity is defined as a function of the angular coordinate θ (see Chapter 2), in PS the intensity is defined as a function of the direction coordinate $p = n\sin(\theta)$. The previous equation implies that, assuming a Lambertian source, the problem of computing the target intensity is reduced to the problem of calculating the boundaries $\partial R(\Pi)$ for all possible paths Π. Hence, the intensity along the direction $p = const.$ is given by the sum of the lengths intervals formed by

the support of the luminance and line $p = const.$. For example, if two intersection points between line $p = const.$ and the boundary $\partial R(\Pi)$ are found, indicating their position coordinates with $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$, where $q^{\min}(\Pi, p) < q^{\max}(\Pi, p)$, and using Eq. (4.3.2), we obtain that Eq. (4.3.3) reduces to:

$$I_{\mathrm{PS}}(p) = \sum_{\Pi} \int_{q^{\min}(\Pi,p)}^{q^{\max}(\Pi,p)} L(q,p)\mathrm{d}q = \sum_{\Pi} \left(q^{\max}(\Pi, p) - q^{\min}(\Pi, p) \right), \tag{4.3.4}$$

where the sum is over all the possible paths and the second equation holds as we assume $L = 1$ in $R(\Pi)$. In case more then two intersection points occur, a generalized equation needs to be used for calculating the intensity. Note that for every single ray only one path is possible as we are assuming that all the lines are reflective lines. Because of this, the regions $R(\Pi)$ do not overlap, i.e.

$$\bigcap_{\Pi} R(\Pi) = \emptyset, \tag{4.3.5}$$

where the intersection is over all possible paths.

From Eq. (4.3.4) we note that, using the PS structure, we could trace less rays inside the system to obtain the target intensity profile. The aim is to construct a ray tracing procedure that allows us tracing less rays overall and more rays close to the discontinuity of the luminance, i.e. close to the boundaries $\partial R(\Pi)$. To this purpose, we start from a triangulation made by only two triangle, then a triangulation refinement at S is defined as explained in the following.

The regions $R(\Pi)$ can be determined only when some rays are traced. The procedure starts with coordinates $(q_1^k, p_1^k)_{k=1,\cdots,4}$ of the four corner points of S. For each of them, the corresponding path $(\Pi^k)_{k=1,\cdots,4}$ is calculated. Next, the grid is divided into two equal triangles joining two opposite vertices (in our simulation we always trace the diagonal north-west to define the new triangles). For each triangle the rays located at its corners are traced. If not all the paths corresponding to those rays are the same, one or more boundaries $\partial R(\Pi)$ are expected to cross the triangle. In that case, the middle points $(q_1^k, p_1^k)_{k=5,6,7}$ of each side of the triangle are added and the three corresponding rays are traced (unless they were already traced in the previous steps). Each refinement step leads to four new triangles (see Figure 4.4).

When all the rays corresponding to the corners of each triangle have the same path, it is not necessary to refine the triangles anymore. The triangles very close to the boundaries have always two vertices whose corresponding rays follow the same path and one vertex whose corresponding ray follows another path. Since they are crossed by a boundaries, two different paths are found for the rays at the vertices of those triangles. Because of this, the procedure could continue infinitely, therefore, two parameters $\varepsilon_q^{\min}$ and $\varepsilon_p^{\min}$ are introduced to defined a stopping criterion. The algorithm stops when the length of the sides of the triangles is smaller than $\varepsilon_q^{\min}$ and $\varepsilon_p^{\min}$ in q and p direction. We indicate all the possible paths with $(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ where Np is the maximum number of paths[1] (Np= 5 for the two-faceted cup). If the size of the triangles is too big, it can happen that a region formed by rays that follow a path

[1]We indicate with $\Pi^k = \Pi(q_1^k, p_1^k)$ the path followed by rays with coordinates (q_1^k, p_1^k) in Source PS, can happen $\Pi^k = \Pi^h$ for $k \neq h$. With $(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ we indicate all the possible Np paths that can occur, therefore $\Pi_i \neq \Pi_j$ if $i \neq j$.

Π_j is located completely inside a triangle whose vertices are related to another path Π_i with $j \neq i$, see Figure 4.5. To avoid this, two parameters $\varepsilon_q^{\max}$ and $\varepsilon_p^{\max}$ are defined for the q_1-axis and the p_1-axis, respectively. When the length of the sides of the triangle are greater than these parameters, a new triangle is defined even if its vertices correspond to the same path. The values of the parameters $\varepsilon_q^{\min}$, $\varepsilon_p^{\min}$, $\varepsilon_q^{\max}$ and $\varepsilon_p^{\max}$ determine the number of rays traced. Thus, on the one hand, decreasing $\varepsilon_q^{\min}$ and $\varepsilon_p^{\min}$ more rays close to the boundaries are traced; on the other hand, decreasing the values of $\varepsilon_q^{\max}$ and $\varepsilon_p^{\max}$ more rays in the interior of the regions are traced.

The triangulation refinement is provided by Algorithm 1 which uses the two recursive functions LEFT TRIANGLE and RIGHT TRIANGLE. The function LEFT TRIANGLE is defined in Algorithm 2 (see Figure 4.6). A similar procedure gives the function RIGHT TRIANGLE (see Figure 4.7).

Algorithm 1 Triangulation refinement algorithm

Initialize $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}$, and $\varepsilon_p^{\max}$, Ray = [empty];
▷ $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}$, and $\varepsilon_p^{\max}$ are fixed parameters needed to stop the procedure
▷ Ray: structure that contains all the information about the rays traced[2].

1: $(q_1^1, p_1^1) \leftarrow$ left bottom corner of source PS: (-a, -1)
2: $(q_1^2, p_1^2) \leftarrow$ right bottom corner of source PS: (a, -1)
3: $(q_1^3, p_1^3) \leftarrow$ right upper corner of source PS: (a, 1)
4: $(q_1^4, p_1^4) \leftarrow$ left upper corner of source PS: (-a, 1)
5: **for** $k = 1 \rightarrow 4$ **do**
6: Trace the ray with initial coordinates (q_1^k, p_1^k) in S;
7: Calculate the corresponding path Π^k;
▷ Store the information found in the structure Ray;
8: Ray.$q \leftarrow$ [Ray.q, q_1^k];
9: Ray.$p \leftarrow$ [Ray.p, p_1^k];
10: Ray.$\Pi \leftarrow$ [Ray.Π, Π^k];
11: **end for**
12: VL $\leftarrow [1, 2, 4]$ ▷ VL vertices of the left triangle
13: VR $\leftarrow [2, 3, 4]$ ▷ VR vertices of the right triangle
14: LEFT TRIANGLE(VL, Ray, $\varepsilon_{q_1}^{\min}, \varepsilon_{q_1}^{\max}, \varepsilon_{p_1}^{\min}, \varepsilon_{p_1}^{\max}$) ▷ Refine the left triangle
15: RIGHT TRIANGLE(VR, Ray, $\varepsilon_{q_1}^{\min}, \varepsilon_{q_1}^{\max}, \varepsilon_{p_1}^{\min}, \varepsilon_{p_1}^{\max}$) ▷ Refine the right triangle
16: **return** Ray;

Figure 4.8 shows an example of a triangulation refinement at the source PS of the two-faceted cup in Figure 3.1. For this optical system, the width of the q_1-axis in source phase space is two times the width of the p_1-axis. Thus, our choice is $\varepsilon_p^{\min} = \frac{1}{2}\varepsilon_q^{\min}$ and $\varepsilon_p^{\max} = \frac{1}{2}\varepsilon_q^{\max}$ with $\varepsilon_q^{\min} = 0.1$ and $\varepsilon_q^{\max} = 1$. Using the triangulation refinement, all the possible paths $(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ are found and their corresponding regions $R_1(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ are determined. Using the edge-ray principle, we conclude that also the regions $R(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ at the target are determined and only the rays close to the boundaries ∂R_1 need to be considered to obtain the target ray distribution.

4.4 Conclusions

In this chapter we introduced the phase space concept. We explained a new ray tracing method based on the source and the target PS representation. In PS every point corresponds to a unique ray. The coordinates of every point correspond to the initial ray position q_1 and the initial ray direction $p_1 = \sin(\theta_1)$ (expressed with respect to the normal of the source). The method also takes into account the paths followed by every ray traced. Considering only reflection, every single ray follows only one path and, therefore, the PS regions do not overlap.

As an example, we provided the source and the target PS representation of the two-faceted cup. The edge-ray principle guarantees that all the rays that follow the same path are located in the same regions in PS. If we know these regions at the source we can determine the corresponding regions at the target. It is sufficient to map the boundaries at the source $\partial R_1(\Pi)$ to obtain their corresponding target boundaries $\partial R(\Pi)$.

The boundaries $\partial R(\Pi)$ are particularly relevant because there the luminance jumps from 0 to a positive value. Assuming a Lambertian source, only the rays at the boundaries are needed to compute the target intensity. Based on this idea, a triangulation in S is constructed such that the rays closest to $\partial R_1(\Pi)$ are selected and more rays in their vicinity are created to get progressively better estimates of the boundaries.

In Figures 4.9 we show three different ray distributions on the source PS of the two-faceted cup. In Figure 4.9a, 10^3 random points are shown. MC ray tracing is based on this random distribution of the initial rays set. In Figure 4.9b, 10^3 points of a two-dimensional Sobol sequence are shown. Since Sobol sequences are defined in a unit square, we scaled it such that all the source PS $S = [-2, 2] \times [-1, 1]$ is covered by rays. QMC ray tracing considers as initial set, rays distributed as the points of a Sobol sequence. Such regular distribution can lead to several advantages for the computation of the target intensity, see Section 3.3. Finally, Figure 4.9c shows a non-uniform distribution of rays at the source PS. Such distribution is obtained from the triangulation refinement explained in the previous section. The procedure requires tracing more rays close to the boundaries $\partial R_1(\Pi)$ and only few rays in their interior.of the regions in source PS. From the edge ray-principle, we obtain that these rays will be located close to the boundaries $\partial R(\Pi)$ of the regions at the target PS. PS ray tracing is based on this kind of initial ray distribution at the source.

The target PS intensity is calculated using only the rays that are located at the boundaries $\partial R(\Pi)$. Thus, in order to obtain the intensity profile at the target, the boundaries $\partial R(\Pi)$ need to be determined.

In the next chapter we provide two different approaches to find the boundaries $\partial R(\Pi)$ using a set of rays given by the triangulation refinement.

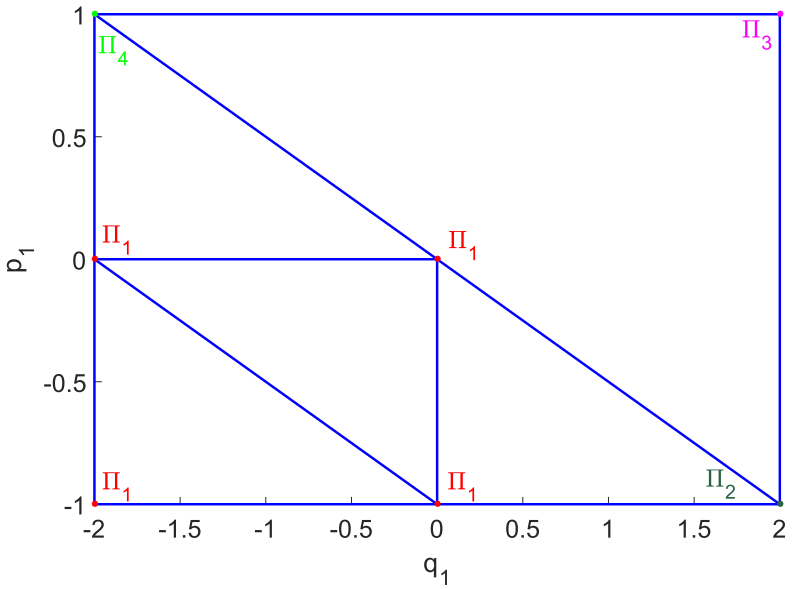


Figure 4.4: Triangulation refinement: when the rays related to the vertices of the triangles follow a different path a new refinement step is required. Each refinement step leads to four new triangles.

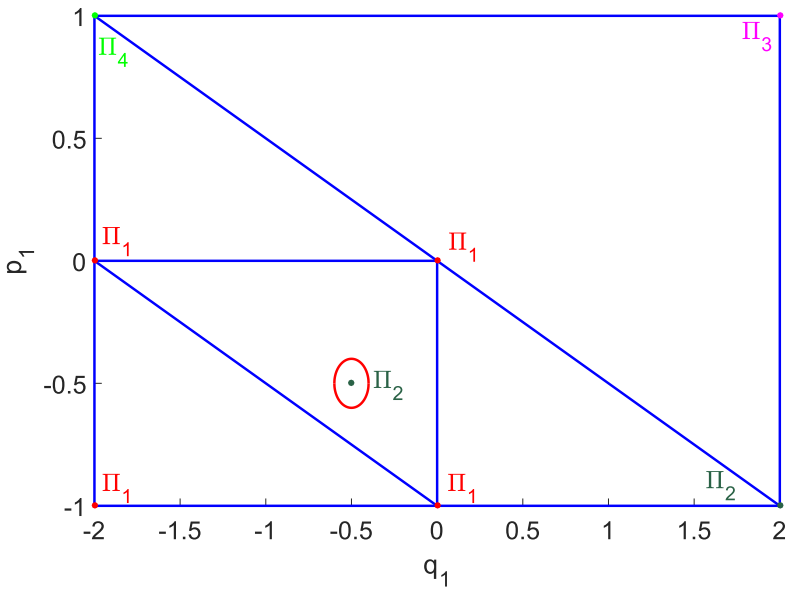


Figure 4.5: The red line encloses a region of rays that follow the path Π_2 and is completely located inside a triangle. The algorithm is not able to detect that region and, a further refinement is required.

Algorithm 2 Algorithm for the refinement of the left triangles

procedure LEFT TRIANGLE(VL, Ray, $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}, \varepsilon_p^{\max}$)
  VL $\leftarrow [1, 2, 4]$
  $q_1^1 \leftarrow$ Ray.q(VL(1)), $p_1^1 \leftarrow$ *Ray*.p(VL(1))
  $q_1^2 \leftarrow$ Ray.q(VL(2)), $p_1^2 \leftarrow$ *Ray*.p(VL(2))
  $q_1^3 \leftarrow$ Ray.q(VL(3)), $p_1^3 \leftarrow$ *Ray*.p(VL(4))
  $\text{dist}_q \leftarrow |q_1^2 - q_1^1|$
  $\text{dist}_p \leftarrow |p_1^3 - p_1^1|$
  RefineTriangle $\leftarrow$ false;
  DifferentPath $\leftarrow$ false;
  if $\text{dist}_q > \varepsilon_q^{\max}$ or $\text{dist}_p > \varepsilon_p^{\max}$ **then**
    RefineTriangle $\leftarrow$ true;
  end if
  for $k = 1 \rightarrow 2$ **do**
    if $\Pi^k \neq \Pi^{k+1}$ **then**
      DifferentPath $\leftarrow$ true;
    end if
  end for
  if $\text{dist}_q > \varepsilon_q^{\min}$ or $\text{dist}_p > \varepsilon_p^{\min}$ **then**
    RefineTriangle $\leftarrow$ DifferentPath;
  else
    if (DifferentPath is true) **then**
      Ray(VL).boundary $\leftarrow$ true; ▷ A boundary crosses the triangle
    end if
  end if
  if (RefineTriangle is true) **then**
    Define the points at the middle of each side of the triangle
    $(q_1^5, p_1^5) = ((q_1^1 + q_1^2)/2, p_1^1)$
    $(q_1^6, p_1^6) = (q_1^5, (p_1^1 + p_1^2)/2)$
    $(q_1^7, p_1^7) = (q_1^1, p_1^6)$
    for $k = 5 \rightarrow 7$ **do**
      if The ray with coordinates (q_1^k, p_1^k) is not traced yet **then**
        Trace the ray with initial coordinates: (q_1^k, p_1^k) in PS;
        Compute the corresponding path Π^k;
        Store the ray's coordinates Ray.$q \leftarrow$ [Ray.q, q_1^k];
        Store the ray path Ray.$\Pi \leftarrow$ [Ray.Π, Π^k];
      end if
    end for
    return LEFT TRIANGLE($[\text{VL}(1), 5, 7]$, Ray, $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}, \varepsilon_p^{\max}$);
    return LEFT TRIANGLE($[5, \text{VL}(2), 6]$, Ray, $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}, \varepsilon_p^{\max}$);
    return LEFT TRIANGLE($[7, 6, \text{VL}(3)]$, Ray, $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}, \varepsilon_p^{\max}$);
    return RIGHT TRIANGLE($[5, 6, 7]$, Ray, $\varepsilon_q^{\min}, \varepsilon_q^{\max}, \varepsilon_p^{\min}, \varepsilon_p^{\max}$);
  end if
return Ray;
end procedure

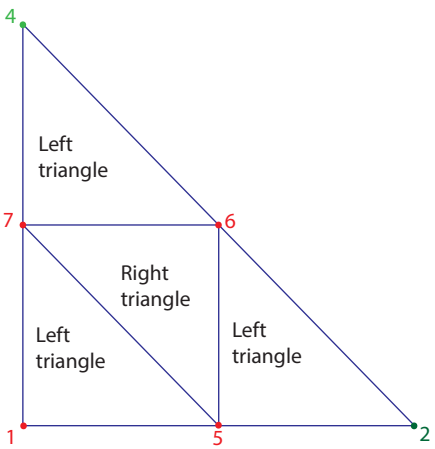


Figure 4.6: Left triangulation refinement algorithm (recursive function LEFT TRIANGLE).

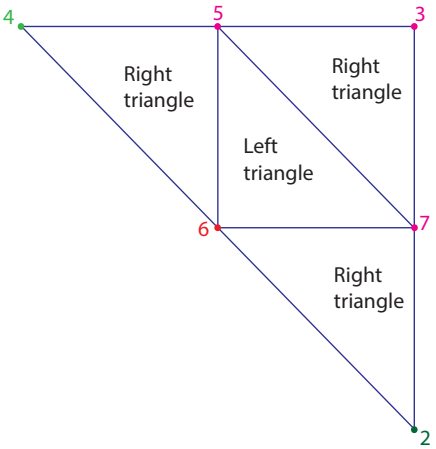


Figure 4.7: Right triangulation refinement algorithm (recursive function RIGHT TRIANGLE).

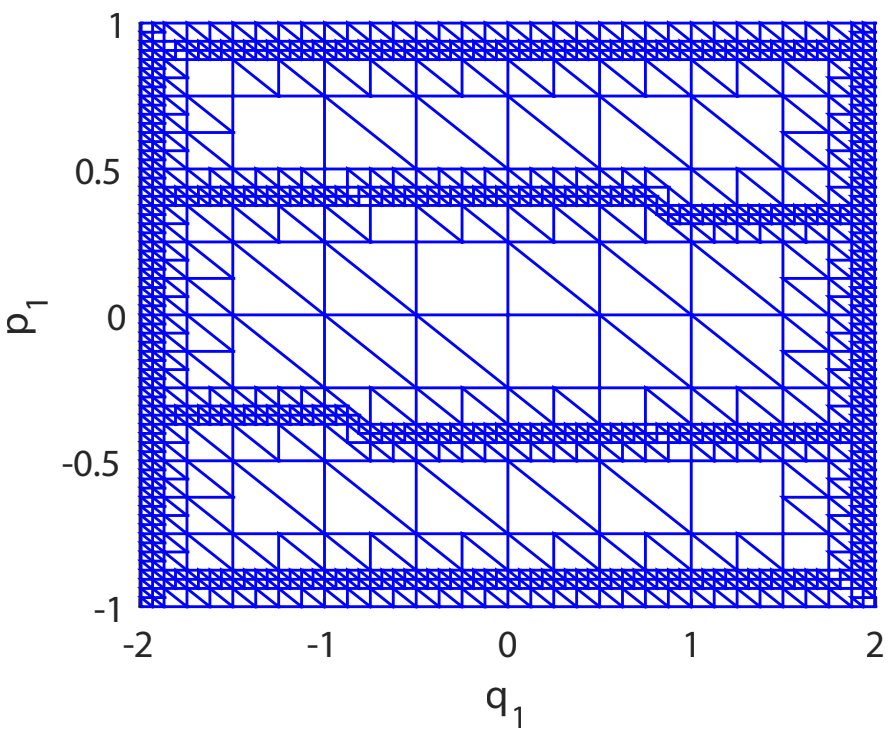
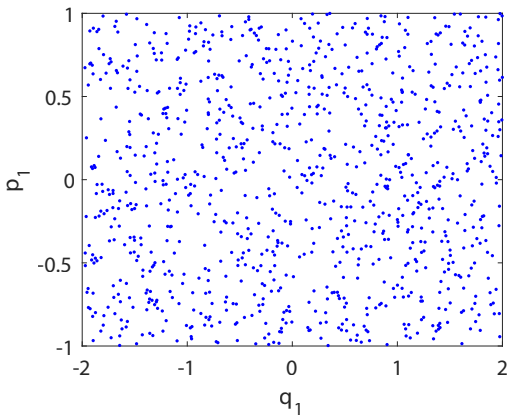
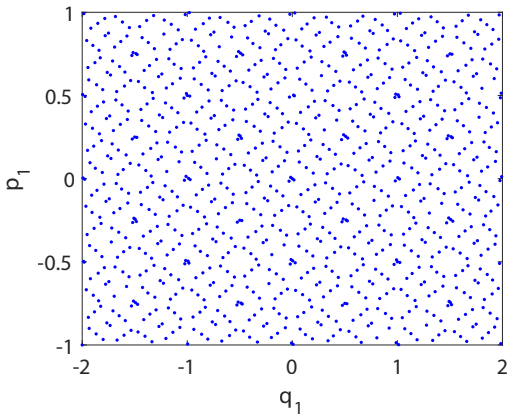


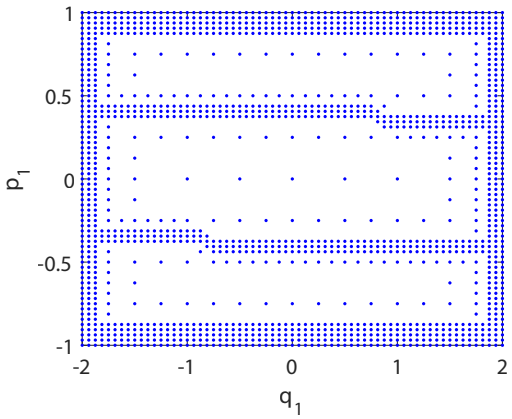
Figure 4.8: Triangulation refinement of source phase space: near the boundaries more rays are traced. The values of the parameters are $\varepsilon_{q_1^{\min}} = 0.1$ and $\varepsilon_{q_1^{\max}} = 1$.



(a) 10^3 random rays at the source PS (MC ray tracing).



(b) 10^3 rays at the source PS distributed as the point of a Sobol sequence (QMC ray tracing).



(c) $1.5 \cdot 10^3$ rays distributed using the triangulation refinement (PS ray tracing).

Figure 4.9: Three different ray distributions at the source of the two-faceted cup.

Chapter 5

The α-shapes approach

In the previous chapter we presented a new ray tracing approach based on PS. We explained that, in order to compute the target intensity, it is necessary to know the boundaries of the regions in target PS with positive luminance. Ray tracing in PS requires tracing only the rays close to these boundaries. The rays traced can be seen as a point cloud in PS. To detect the shape formed by those rays, the α-shapes approach is employed [33].

Methods based on α-shapes are widely used to reconstruct an unknown shape formed by a set of finite data points [34]. α-shapes are a very powerful tool to construct the shape of a point cloud. As α varies, we can obtain different α-shapes from the point set itself to the convex hull [35]. The disadvantage of such methods is that it can be very hard to choose the appropriate value of the parameter α and, in most cases it can be selected only by numerical simulations.

We develop a technique based on α-shape that gives a criterion to determine the value of the parameter α, for which the boundaries are approximated well [36].

This chapter is organized as follows. An overview of the-state-of-the-art about α-shape methods is provided in Section 5.1; the technique used for computing the α value is explained in Section 5.2; the results for two different kind of total internal reflection (TIR)-collimators are given in Section 5.3. Discussions are conclusions are provided in the last paragraph of this chapter.

5.1 α-shapes theory

Given a finite set $V = \{v_1, \cdots, v_N\} \subset \mathbb{R}^2$ of points, α-shapes are geometrical objects that give us an approximation of the shape formed by the point cloud. For now we do not further specify the notion of shape. A more precise definition will be provided later.

Before giving a formal definition, we explain an intuitive and nice interpretation of α-shapes [37]. Let us think of a stracciatella ice-cream[1]. If we desire to know the shape formed by the chocolate pieces we can start eating the ice cream using a spoon with a spherical scoop and try not to remove any piece of chocolate. We

[1]Stracciatella ice cream is made with milk-based ice-cream and fine pieces of chocolate, [38].

will obtain a shape formed by arcs and points (see Figure **??** for the two-dimensional case). Straightening the arcs to line segments we obtain broken lines which constitute the boundary of the so-called α-shape of the point set V. A very small spoon will allow us to eat the entire ice cream without eating any piece of chocolate, while with a larger spoon we are not able to eat any chunk of the ice cream without chocolate pieces. In this example, the chocolates peaces are the points of V and, the parameter α determines the radius of the carving spoon (the spherical spoon in two-dimension is simply a circle).

The formal definition of α-shape was first given by Edelsbrunner, Kirkpatrick and Seidel in 1983, [39]. They describe α-shape as a generalization of the convex hull of a finite set of points in the plane. Let α be a non negative number $0 \leq \alpha < \infty$. If $\alpha = 0$ the shape degenerates to the point set V. On the other hand, when $\alpha \to \infty$ the α-shape is simply the convex hull of V. If $0 < \alpha < \infty$ the α-shape is a polytope of V [40]. The construction of the α-shape is closely related to the Delaunay triangulation of V [41]. Therefore, a formal definition of triangulation and Delanauy triangulation is now required.

Given a set V of not all aligned points, let us consider the set E of all the straight-line segments whose endpoints are in V. A triangulation T of V is the subset of E with the maximum number of segments such that all the line segments of T intersect only at their endpoints [42]. A more formal definition of triangulation is provided in the following

Let $\Omega \subset \mathbb{R}^2$ be the convex hull of V and $T_h = \{K_1, \cdots, K_h\}$ be a partition of Ω into closed triangles, that is triangles that included their boundary. Suppose that the following properties are verified:

(*a*) $\Omega = \bigcup_{i=1}^{h} K_i$;

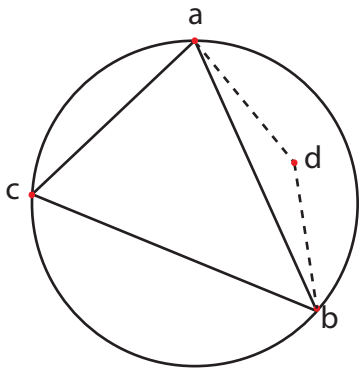
b) For $K_i, K_j \in T_h$, $K_i \neq K_j$, and

$$\text{int}(K_i) \cap \text{int}(K_j) = \emptyset,$$

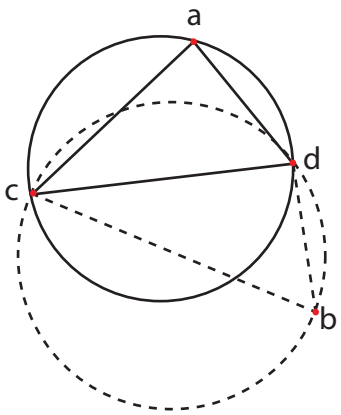
where $\text{int}(K) = K - \partial K$.

Then T_h is called a *triangulation* of Ω [43].

The Delaunay triangulation T' of the points set V has the property that the circle circumcircle of any triangle of T does not contain any point of V. This is called the Delaunay property. A very commonly used algorithm to construct such triangulation is explained in the following. T' is constructed by modifying a general triangulation T such that every point satisfies the Delaunay property. Therefore, every triangle (or tetrahedron in three dimensions) that does not satisfy such property is flipped such that the new edge is part of the triangulation (see Figure 5.1). Given, for example, an arbitrary triangulation T in two-dimensions, for each edge $\overline{\text{ab}}$ in T which is not on the boundary of the convex hull the two triangles Δ_{abc} and Δ_{abd} with the common edge $\overline{\text{ab}}$ are found. Then, if either the circumcircle of triangle Δ_{abc} contains point d or the circumcircle of triangle Δ_{abd} contains point c the edge $\overline{ab}$ cannot be included in the Delaunay triangulation and, therefore, it is flipped such that the other two possible triangles Δ_{acd} and Δ_{bcd} are found. The new edge $\overline{\text{cd}}$ locally satisfies the Delaunay property and the triangles Δ_{acd} and Δ_{bcd} are added to the Delaunay triangulation T', see Figure 5.1.



(a) The point d is inside the circle circumscribing the triangle Δ_{abc}, therefore the edge $\overline{\text{ab}}$ cannot be included in the Delaunay triangulation.



(b) The flipped triangle Δ_{acd} satisfies the Delaunay property, thus it is included in the Delaunay triangulation.

Figure 5.1: Construction of the Delaunay triangulation in 2D.

Several other algorithms have been developed to construct a Delaunay triangulation, see for example [44, 45]. Given a point set V and a triangulation T, it can be proved that the corresponding Delaunay triangulation T' is unique. Moreover, it has the property to have the largest minimum angle among all possible triangulations of a point set V [46].

Alternatively the Delaunay triangulation can be constructed as the dual of the Voronoi diagram [47].

For *almost*[2] every point $x \in \mathbb{R}^2$, there is a point which is the closest point to x. The Voronoi cell of a point $v_i \in V$ contains all points in $\mathbb{R}^2$ which are closer to v_i, see Figure 5.2. The Voronoi diagram of $V \subset \mathbb{R}^2$ is defined as the set of all Voronoi cells, [48]. A more formal definition of the Voronoi diagram is given in the following.

Definition 5.1.1. Let $V = \{v_1, \cdots, v_N\}$ be a set of point in $\mathbb{R}^2$. The Voronoi cell V_i associated to point v_i is defined as:

$$V_i = \{x \in \mathbb{R}^2 \mid |x - v_i| < |x - v_j| \quad \forall j \neq i\}, \tag{5.1.1}$$

The Voronoi diagram U is defined as

$$U = \bigcup_{i=1}^{N} V_i \tag{5.1.2}$$

where $V_i \cap V_j = \emptyset$ for $i \neq j$.

For the definition of Voronoi diagram in higher dimensions see [49].

The Delaunay triangulation triangulates the convex hull of V and, therefore it does

[2]It is needed to specify the word *almost* because some points can have the same distance with two or more points of V.

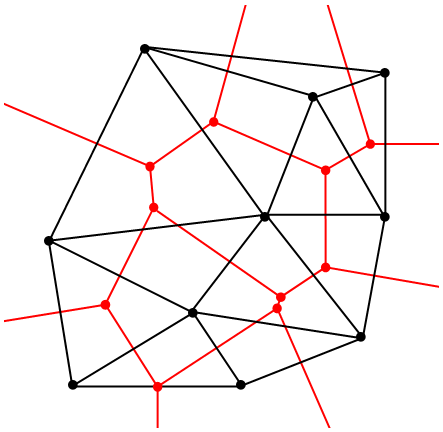


Figure 5.2: Relationship between the Delaunay triangulation (in black) and the Voronoi Diagram (in red), [50]

not constitute a suitable method for reconstructing the contour formed by a point cloud. Therefore, α-shape methods were developed to solve such problem [51, 34]. Starting from the Delaunay triangulation T' of a point set V, the corresponding α-shape of V is formed by the only triangles of T' that satisfy the so-called "α-test" which is now briefly explained. For each triangle we calculate the circumradius, i.e. the radius of the circumcircle. If the radius is larger than α the triangle is removed from the shape. The choice of the parameter α is highly significant in the α-shapes procedure and, it has to be selected such that the desired approximation of the shape formed by the points of V is obtained. Therefore, α is closely related to the radius of the circumcircles. A possible strategy is to find the radius of the greater empty circumcircle. Thus α can be selected according to the density δ of the point sets V with C a constant and δ:

$$\delta = \frac{N}{A} \,, \tag{5.1.3}$$

where N is the number of points in V and A is the area of the convex hull of V. The value of α can be chosen, for instance, inversely proportional to $\sqrt{\delta}$:

$$\alpha = C\frac{1}{\sqrt{\delta}} \,, \tag{5.1.4}$$

Note that, while δ is given for a fixed point set V, the value of C needs to be determined by numerical simulations.

To summarize, the α-shape construction can be outlined as follows:

1. Construct a Delaunay triangulation[3] T' of the point cloud V;
2. For every triangle $T'(i) \in T'$ calculate its circumradius $r(i)$;
3. If $r(i) \leq \alpha$ keep the triangle $T'(i)$ in the triangulation;

[3]In the simulations we present in this chapter the Matlab function *Delaunay* is used.

4. If $r(i) > \alpha$ remove the triangle from the triangulation;

5. For every triangle return the edges belong to only one triangle of T', the so-called *free boundary* edges[4]. By definition, the free boundary edges are not a common edge of two triangles.

α-shapes provide a nice mathematical definition of the *shape* of a set of points. In two dimensions, α-shapes gives the contour of the point cloud which is approximated by a family of broken curved lines. Although they are a powerful tool for determining the shape of a point cloud, there exist shapes that are not described well by classical α-shapes. Indeed for some surfaces there is no value of α that includes all desired triangles and deletes all undesired triangles. If the parameter α is determined according to the density of the point cloud, it can be difficult to obtain a good approximation of a shape formed by a non-uniform points set. Furthermore, the α-shape method does not work well when the shape we need to approximate has a sharp turn or a joint. To overcome these issues, Teichmann and Capps presented alternatives approaches to establish the value of α, [52]. Their anisotropic density-scaled α-shapes method constitutes an improvement of classical α-shapes.

There are several ways to determine the value of α [53]; in the next section we provide a technique that exploits the conservation of étendue in PS.

5.2 Determination of α using étendue conservation

As mentioned in Section 4.1, in two-dimensions étendue can be seen as an area in PS. Therefore, given an optical system with a line segment source $\mathsf{S} = [-a, a]$, the étendue at the source coincides with the area of source PS S, and it is given by:

$$U = 4n_1\, a \sin(\theta_1^{\max})\,, \tag{5.2.1}$$

where a is the half length of the source, n_1 the index of refraction of the medium in which the S is located and $\theta_1^{\max}$ is the maximum value of the angle that the rays make with the normal $\boldsymbol{\nu}_1$ of the source.

For some optical systems, all the rays emitted by the source arrive at the target, for some others there are also rays that can end at other detectors which are located outside the system. Indicating with $R_1(\Pi)$ the regions in source PS formed by the rays that reach the target following path Π and with $R(\Pi)$ the corresponding regions at the target, the étendue U_1 at source related to the only rays that arrive to the target is given by:

$$U_1 = \sum_{\Pi} U\big(R_1(\Pi)\big), \tag{5.2.2}$$

where the sum is over all possible paths Π and $U\big(R_1(\Pi)\big)$ is the contribution to the étendue given by the rays inside $R_1(\Pi)$ in source PS, obtained by:

$$U_1\big(R_1(\Pi)\big) = \iint_{R_1(\Pi)} \mathrm{d}q\,\mathrm{d}p. \tag{5.2.3}$$

[4]In the simulations we present in this chapter the Matlab function *freeBoundary* is used.

Similarly the étendue at the target of the rays emitted by the source is:

$$U_{\mathrm{t}} = \sum_{\Pi} U\big(R(\Pi)\big), \tag{5.2.4}$$

with

$$U\big(R(\Pi)\big) = \iint_{R(\Pi)} \mathrm{d}q\,\mathrm{d}p. \tag{5.2.5}$$

In order to determine the value of α in the α-shape procedure that approximates the boundaries $\partial R(\Pi)$ accurately, we use étendue conservation ($U_{\mathrm{t}} = U_1$). The α-shapes method is applied to every region $R(\Pi)$ for a range of values of α; for each value an approximation of the boundaries $\partial R(\Pi)$ is obtained and the intersection points $q^{\max}(\Pi, p)$ and $q^{\min}(\Pi, p)$ between $\partial R(\Pi)$ and the horizontal lines $p = const$, with $p \in [-1, 1]$, are computed for every path Π. Therefore Equation (5.2.5) becomes:

$$U_{\mathrm{t}}\big(R(\Pi)\big) = \int_{-1}^{1} \Big(q^{\max}(\Pi, p) - q^{\min}(\Pi, p)\Big)\mathrm{d}q\,\mathrm{d}p. \tag{5.2.6}$$

In case more than two intersection points between line $p = const$ and $\partial R(\Pi)$ occur, the previous equation needs to be generalized. Suppose that r intersection points $\big(q^{i}(\Pi, p), p\big)_{i=1,\cdots,r}$ are found. Ordering their q coordinates in ascending order, the target étendue is calculated by:

$$U_{\mathrm{t}}\big(R(\Pi)\big) = \sum_{i=1}^{m} \int_{-1}^{1} (q^{2i}(\Pi, p) - q^{2i-1}(\Pi, p))\mathrm{d}q\,\mathrm{d}p\,, \tag{5.2.7}$$

where m is the integer part of $r/2$. The integrals in Equations (5.2.6) and (5.2.7) are calculated discretizing the interval $[-1, 1]$ into $\mathrm{Nb} = 100$ sub-intervals of equal length, the so-called bins, and using the trapezoidal rule.

Matching the value of the étendue at the source U_1 with the value of the étendue at the target U_{t}, a unique value α_c of α is determined. Implementing the α-shapes procedure with $\alpha = \alpha_c$, the best approximation of the boundaries $\partial R(\Pi)$ is found and the intensity at the target can be calculated.

If two intersection points between $p = const$ and $\partial R(\Pi)$ are found the target intensity is calculated using Equation (4.3.4). If more than two-intersection points are found we use the generalized equation:

$$I_{\mathrm{PS}}(p) = \sum_{\Pi,i} \int_{q^{2i-1}(\Pi,p)}^{q^{2i}(\Pi,p)} L(q, p)\mathrm{d}q = \sum_{\Pi,i} \left(q^{2i}(\Pi, p) - q^{2i-1}(\Pi, p)\right), \tag{5.2.8}$$

where $q^{2i}(\Pi, p) > q^{2i-1}(\Pi, p)$, the summation over Π is for all the paths Π for which the intersection $p = const$ and $R(\Pi)$ is not empty, and the summation over i is for $i = 1, 2, \cdots, m$. The second equation holds as we assume $L(q, p) = 1$.

To clarify our idea we apply the method to two different optical systems, the results are presented next.

5.3 Results for a TIR-collimator

We apply the α-shapes method to the set of points in target PS obtained by using PS ray tracing. In this chapter the procedure is applied to two different kinds of total internal reflection (TIR)-collimators.

Let us first describe the TIR-collimator depicted in Figure 5.3. It is an optical system symmetric with respect to the z-axis, it consists of a lens (central curve), two broken lines adjacent to the lens, two curved lines on each side and a top formed by a horizontal segment. The lens (line 2) and the broken lines, formed by a collection of three segments (lines $3, 4,$ and 5 and $9, 10$ and 11), are refractive line segments while the curved lines (labeled with 6 and 8) are designed in such a way that light is totally internal reflected (which explains the name TIR). The light source $\mathcal{S}$ (line 1) and the target $\mathcal{T}$ (line 12) are two straight line segments normal to the optical axis. The source $\mathsf{S} = [-2, 2]$ is located at a height $z_1 = 0.3$ from the x-axis. The target $\mathsf{T} = [-9.7, 9.7]$ is parallel to the source and is located at a height $z = 8.2$. Both S and T are located in air ($n_1 = 1$). The volume inside the collimator is filled with a material with index of refraction $n_2 = 1.5$ (e.g. glass). The collimator is surrounded by two vertical lines (lines 13 and 15) and two horizontal lines (12 and 14) that receive the light emitted from the source that do not arrive at the target; among these the one at the top (line 12) is assumed to be the target, and it is located at a small distance from the top (line 7).

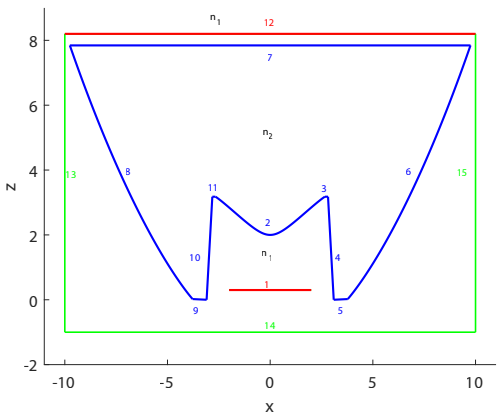


Figure 5.3: **Shape of the TIR-collimator.** Each surface of the system is labeled with a number. The shape of the collimator is shown with a blue line. Three detectors depicted with green lines (surfaces 13, 14, and 15) are located at the left, the right and the bottom of the optical system. The source (line1) and the target (line 12) are depicted in red. The sagitta of the lens is approximately 1.17.

Using PS ray tracing explained in Section 4.3 with parameters $\varepsilon_{q_1}^{\max} = 0.1$, $\varepsilon_{p_1}^{\max} = 5 \cdot 10^{-2}$, $\varepsilon_{q_1}^{\min} = 9 \cdot 10^{-3}$ and $\varepsilon_{p_1}^{\min} = 4.5 \cdot 10^{-3}$, around $1.9 \cdot 10^4$ rays are traced. We discard rays with direction parallel to the source, therefore $p_1 \in [-0.99, 0.99]$. The ray distribution at the source PS S is shown in Figure 5.4, where we depicted the rays that

follow the same path with the same color. Seven different paths are found. The yellow rays follow path $\Pi_1 = (1, 2, 7, 12)$; the red rays follow path $\Pi_2 = (1, 4, 6, 7, 12)$; the green rays follow path $\Pi_3 = (1, 10, 8, 7, 12)$; the blue rays follow path $\Pi_4 = (1, 3, 7, 12)$ and the magenta rays follow path $\Pi_5 = (1, 11, 7, 12)$. The rays located inside the white areas correspond to rays that do not reach the target, they follow either path $\Pi_6 = (1, 4, 7, 6, 13)$ or path $\Pi_7 = (1, 10, 7, 8, 15)$ and they do not give any contribution to the target intensity. Note that, given two adjacent paths the regions $R_1(\Pi)$ in S

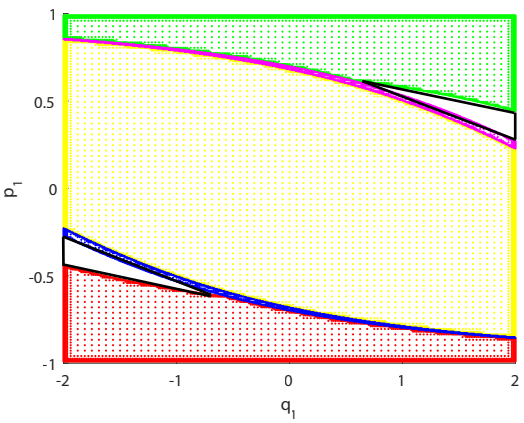


Figure 5.4: **Distribution of the rays on S**. Around $1.67 \cdot 10^4$ rays are traced using the triangulation refinement with parameters: $\varepsilon_q^{\max} = 0.1$, $\varepsilon_p^{\max} = 5 \cdot 10^{-2}$, $\varepsilon_q^{\min} = 9 \cdot 10^{-3}, \varepsilon_p^{\min} = 4.5 \cdot 10^{-3}$. Rays that belong to the same region are depicted with the same color. The rays located inside the white areas do not reach the target. The boundaries of the two white regions are approximated by triangles depicted with black lines.

have usually a common boundary. Since for this system not all the rays emitted by the source arrive at the target, U_{t} needs to be compared with the étendue U_1 at the source given by only those rays that reach the target (the rays that follow paths $\Pi_6 = (1, 10, 8, 7, 12)$ and $\Pi_7 = (1, 4, 7, 6, 15)$ are discarded). To this purpose U_1 is calculated by removing from the total area U of S those areas occupied by the regions formed by the rays that hit the left or the right detector (white regions in Figure 5.3). For the TIR collimator in Figure 5.3, the total source étendue obtained from Equation (6.1.2) is

$$U = 4 \cdot 1 \cdot 2 \cdot 0.99 = 7.92. \tag{5.3.1}$$

Indicating with A_T the area of each white region in Figure 5.3, U_1 can be approximated by:

$$U_1 = 7.92 - 2A_T \approx 7.77, \tag{5.3.2}$$

where A_T is the approximated area of the triangles shown in Fig. 5.4 with black lines.

U_{t} is calculated several times from Equation (5.2.6) where every time the boundaries $\partial R(\Pi)$ are obtained by using α-shapes for a different value of α. An accurate approximation of $\partial R(\Pi)$ gives a value of U_{t} close to the exact étendue. Matching U_1

with all the approximations of U_t we find the best value α_c of α that approximates $\partial R(\Pi)$ and, therefore, U_t.

In Figure 5.5 we represent the approximated value of the source étendue $U_1 \approx 7.77$ with the red line. Different approximations of the target étendue U_t are calculated using Equation (5.2.7) where every time the boundaries $\partial R(\Pi)$ are found using α-shapes for a different value of α. In Figure 5.5, we show how the étendue changes by increasing the value of α. This graph shows that using PS ray tracing with $1.67 \cdot 10^4$ rays, the best approximation of the boundaries $\partial R(\Pi)$ is given considering $\alpha = \alpha_c = 0.139$ in the α-shapes procedure. Applying α-shapes with $\alpha = \alpha_c$, a good approximation of

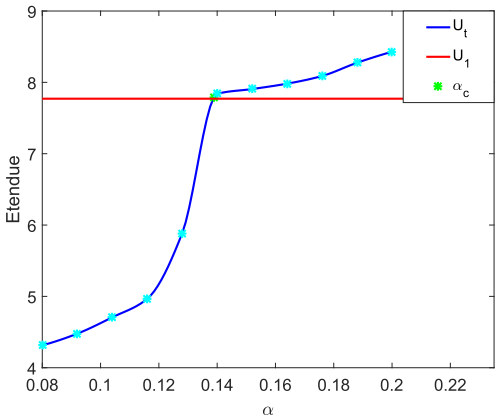


Figure 5.5: **Etendue for the TIR-.collimator** U_t is computed for a range of values for α. $U_1 \approx 7.77$ The green dot indicates the value of $\alpha_c = 0.139$ which gives the best approximation of the boundaries $\partial R(\Pi)$ at the target. Around $1.67 \cdot 10^4$ rays have been traced using PS ray tracing.

$\partial R(\Pi)$ is found. In Figure 5.6 we show the boundaries $\partial R(\Pi)$ in target PS T with $\alpha_c = 0.139$ and tracing $1.67 \cdot 10^4$ rays. Once the boundaries are computed, the target intensity $I_{\mathrm{PS}}(p)$ for every $p \in [-1, 1]$ is obtained from Equation (4.3.4).

To validate our method we compare the PS intensity with the MC intensity. To this purpose a partitioning $P_2 : -1 = p_0 < p_1 < \cdots < p_{\mathrm{Nb}} = 1$ of the interval $[-1, 1]$ into $\mathrm{Nb} = 100$ bins is considered. The averaged and normalized PS intensity $\hat{I}_{\mathrm{PS}}$ is calculated for every $\left(p^{h+1/2} = \frac{1}{2}(p^{h+1} + p^h)\right)_{h=0,\cdots,\mathrm{Nb}-1}$ dividing the PS averaged intensity by the total étendue:

$$\hat{I}_{\mathrm{PS}}(p^{h+1/2}) = \frac{1}{U_t} \int_{p_h}^{p_{h+1}} I_{\mathrm{PS}}(p)\mathrm{d}p. \tag{5.3.3}$$

The averaged and normalized MC intensity $\left(\hat{I}_{\mathrm{MC}}(p^{h+1/2})\right)_{h=0,\cdots,\mathrm{Nb}-1}$ intensity is given by

$$\hat{I}_{\mathrm{MC}}(p^{h+1/2}) = \frac{\mathrm{Nr}[p^h, p^{h+1})}{\mathrm{Nr}[-1, 1)} \qquad \text{for } p \in [p^h, p^{h+1}). \tag{5.3.4}$$

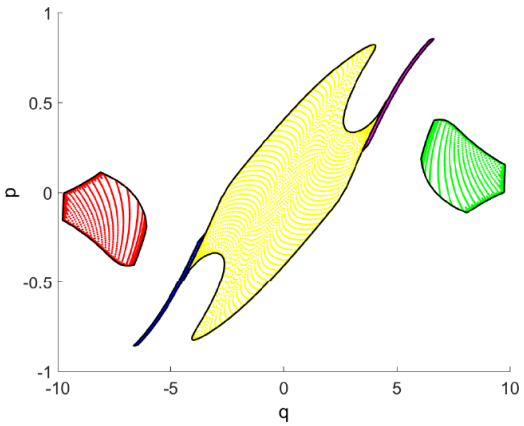


Figure 5.6: **Target PS representation.** A set of $1.67 \cdot 10^4$ rays are traced. Rays that follow the same path are depicted with the same color. The choice of the colors is constistent with Figure 5.4. The boundaries $\partial R(\Pi)$ are computed through the α-shapes method with $\alpha = \alpha_c = 0.139$.

Both approximate intensities $\hat{I}_{\mathrm{A}}(\mathrm{A} = \mathrm{PS}, \mathrm{MC})$ are compared with an intensity $\hat{I}_{\mathrm{ref}}$ taken as a reference. For some optical systems, there is an explicit solution for the target intensity but this is not the case of the TIR-collimator. Therefore, a MC simulation with $1.7 \cdot 10^8$ rays is run to obtain the averaged normalized intensity $\hat{I}_{\mathrm{ref}}$. The intensity profile $\hat{I}_{\mathrm{PS}}$ obtained using PS ray tracing with 66 855 rays and $\alpha = \alpha_c = 0.02$ is depicted in Fig. 5.7 with a red line. $\hat{I}_{\mathrm{PS}}$ is hardly distinguishable from $\hat{I}_{\mathrm{ref}}$ (dashed and blue line in Figure 5.7).

Finally, we calculate the error between $\hat{I}_{\mathrm{A}}$ and $\hat{I}_{\mathrm{ref}}$, defined as:

$$\text{error} = \frac{\sum_{h=1}^{\mathrm{Nb}} |\hat{I}_A(p^{h+1/2}) - \hat{I}_{\mathrm{ref}}(p^{h+1/2})|}{\mathrm{Nb}}. \tag{5.3.5}$$

The MC and PS intensities are calculated several times increasing the number of rays to improve the accuracy. Table 5.1 and 5.2 describe how the number of rays traced affects the error. In Table 5.1 the correlation between α_c and the number of rays is evident. Note that increasing the number of rays the value of α_c and the corresponding error decrease.

In Table 5.2 the numerical results of MC ray tracing are reported. Increasing the number of rays traced in MC ray tracing, the error gradually decreases. In Figure 5.8, the results listed in Table 5.1 and Table 5.2 are shown. The red line depicts the convergence of the PS error and the blue line indicates the MC error. Note from Figure 5.8 that the error for the MC method decreases as $\frac{1}{\sqrt{\mathrm{Nr}}}$, while for the PS simulation the speed of convergence is much higher.

We need to emphasize that the PS ray tracing convergence may change according to the design of the optical system. This is because the approximation of the boundaries in PS depends on the accuracy of the α-shapes method. The α-shapes procedure is unable to properly detect the boundaries of regions with a sharp turn if not enough

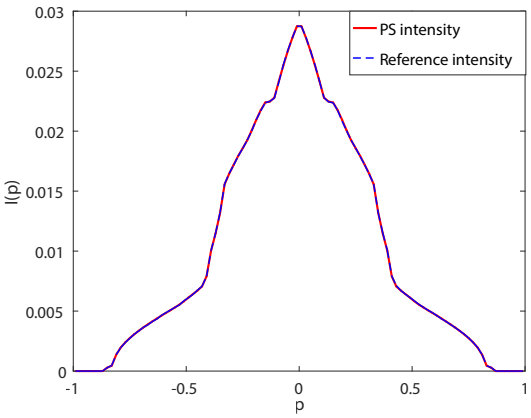


Figure 5.7: **Target intensity profile.** The exact intensity is computed using the MC method for a set of $1.7 \cdot 10^8$ rays. For the PS intensity a set of $6.6 \cdot 10^4$ rays is considered and $\alpha_c = 0.02$ is chosen to compute the boundaries $\partial R(\Pi)$.

Table 5.1: **Errors of the PS intensity**

Number of rays	$\varepsilon_q^{\max}$	$\varepsilon_q^{\min}$	$\varepsilon_p^{\max}$	$\varepsilon_p^{\min}$	α_c	PS error
3 363	0.9	0.1	0.50	0.025	0.119	$1.20 \cdot 10^{-3}$
6 949	0.5	0.050	0.25	0.020	0.098	$2.50 \cdot 10^{-4}$
15 870	0.4	0.025	0.02	0.001	0.050	$5.49 \cdot 10^{-5}$
37 455	0.2	0.020	0.10	0.005	0.037	$2.00 \cdot 10^{-5}$
66 855	0.1	0.009	0.05	0.004	0.020	$1.00 \cdot 10^{-5}$

points are given [52]. Indeed, on the one hand a low density requires a large value of α to accept the triangles in a region, on the other hand, choosing α large, the shape of the region could be destroyed (some triangles inside the regions could be taken into account). Figure 5.9 clarifies this concept showing that the region formed by rays that hit the lens is hard to approximate when there is a small number of rays inside the region. Consequently either a region bigger than the area covered by the rays is considered or some triangles which are not part of the boundaries are considered in the triangulation. This results in an inaccurate calculation of the intensity (either too high or to low). To obtain a good approximation of the boundaries of these kind of patches more rays have to be traced. The PS error decreases very fast increasing the number of rays (see Table 5.1 and Figure 5.8).

To show how the error plot changes according to the regularity of the shape of the regions $\partial R(\Pi)$, we consider another example of a TIR-collimator. Figure 5.9 shows that the hardest region to approximate is given by those rays that follow path $\Pi_1 = (1, 2, 7, 12)$. We therefore consider a TIR-collimator with a flatter lens and with the target located at a closer distance from the top (see Figure 5.10). The source $\mathcal{S} = [-2, 2]$ (surface number 1) is located in air at a height $z_1 = 0.3$ from the x-axis. The target $\mathcal{T} = [-9.7, 9.7]$ (surface 12) is parallel to the source and is located in air

Table 5.2: **Error values of the MC intensity**

Number of rays	MC error
972	$2.10 \cdot 10^{-3}$
9 714	$6.69 \cdot 10^{-4}$
97 103	$2.08 \cdot 10^{-4}$
971 627	$7.00 \cdot 10^{-5}$
9 716 519	$2.00 \cdot 10^{-5}$

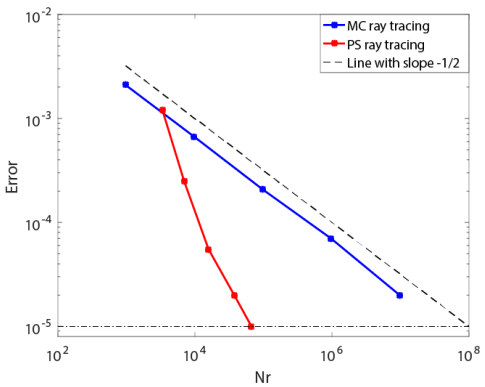


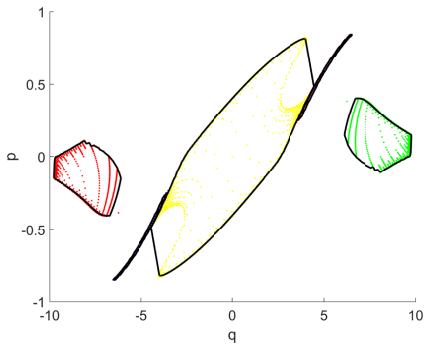
Figure 5.8: **PS and MC errors as a function of the number of rays** The horizontal dotted line shows that an error equal to $2.00 \cdot 10^{-5}$ can be obtained tracing at least 10^2 times fewer rays in phase space.

at a height $z = 7.85$. The shape of the collimator is shown as a blue line. Three detectors depicted with green lines (surfaces 13, 14, and 15) are located at the left, the right and the bottom of the optical system.

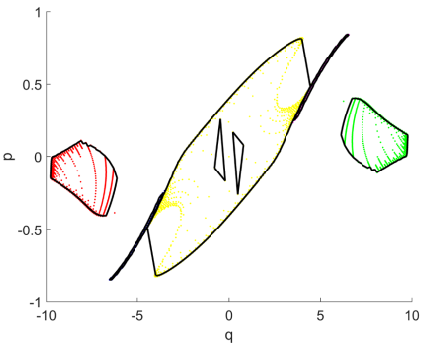
Tracing 3281 rays using PS ray tracing, we obtain the target rays distribution shown in Figure 5.11. Compared it with the distribution in Figure **??**, we note that a flatter lens removes one of the two spikes of the region formed by the rays that hit the lens. Moreover a target located very close to the top makes the shape of that region less stretched along the q-axis. Therefore, it is expected that α-shapes method performs well, even for a small number of rays.

PS and MC ray tracing are implemented for the TIR-collimator in Figure 5.11. The approximated intensities $\hat{I}_{\mathrm{A}}$ ($\mathrm{A} = \mathrm{PS}, \mathrm{MC}$) are compared with the reference intensity $\hat{I}_{\mathrm{ref}}$ (MC ray tracing with 10^7 rays). PS error is depicted with the red line and, MC error is depicted with the blue line.

For the TIR-collimator in Figure 5.10 we obtained a speed of convergence of the order of $O\left(\frac{1}{\mathrm{Nr}}\right)$ for PS ray tracing versus an order of convergence of $O\left(\frac{1}{\sqrt{\mathrm{Nr}}}\right)$ for MC ray tracing.



(a) Boundaries approximation obtained using the α-shapes method with $\alpha_c = 0.3$ (black lines).



(b) Boundaries approximation obtained using the α-shapes method with $\alpha_c = 0.31$ (black lines).

Figure 5.9: **Approximated boundaries at the target PS.** Tracing 3339 rays and using α-shapes, the boundaries cannot be approximated well. A small change of the parameter α leads to a completely different approximation of the boundaries.

5.4 Conclusion

The aim of this chapter was using α-shapes to detect the boundaries of the regions formed by the rays traced.

First, we reported some theory about α-shapes which are commonly used to approximate the shape formed by a point cloud. These methods depend on a parameter α that in most cases can be determined only by several simulations.

Using étendue conservation, we developed a new approach to detect the value of α that better approximates the boundaries in target PS. We applied α-shapes to two different kinds of TIR-collimators. The target PS intensity was computed for both systems several times increasing every time the number of rays traced. Finally, the corresponding errors between the intensities found and a reference intensity was calculated. We observed that PS ray tracing leads to trace far less rays compared to MC ray tracing. Numerical results show that using PS ray tracing the desired accuracy can be achieved reducing significantly the number of rays traced..

However, we observed that the error convergence for PS ray tracing strongly depends on the design of the optical system (shapes of the region in target PS). Indeed, the intensity accuracy is related to the precision of the α-shape, that is, to the choice of the parameter value of α. For more complicated shapes in PS, more rays need to be traced for a good boundaries reconstruction.

In order to remove the dependence of PS ray tracing on the parameter α, we will construct another procedure to detect the boundaries of the regions in target PS. The new technique is based on the triangulation refinement explained in Section 4.3. The details are explained in the next chapter and numerical results are reported for several optical systems.

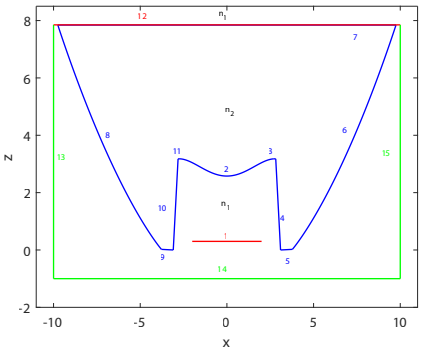


Figure 5.10: **Shape of the TIR-collimator.** Each surface of the system is labeled with a number. $n_1 = 1$ is the refraction index of the medium (air) where the source and the target are located, and $n_2 = 1.5$ the refraction index of the medium (glass) inside the optical system. The sagitta of the lens is equal to 0.6.

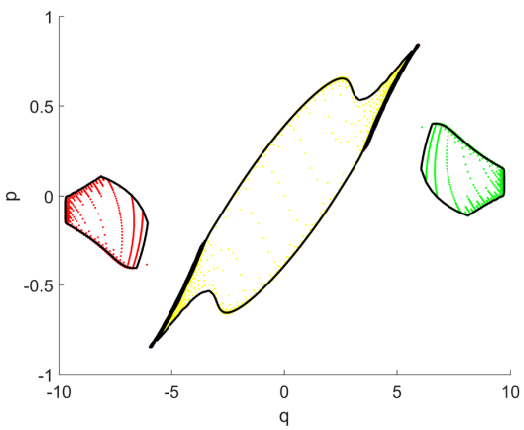


Figure 5.11: **Target phase space for the TIR-collimator depicted in Figure 5.10.** The black line depicts the best approximation of $\partial R(\Pi)$ for 3281 rays. The α-shapes method gives an accurate approximation of the boundaries for $\alpha_c = 0.9$.

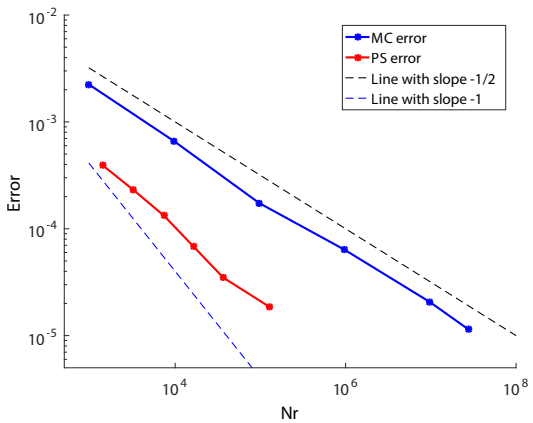


Figure 5.12: **PS and MC errors.** The errors are in a logarithmic scale.

Chapter 6

Boundaries reconstruction based on the triangulation refinement

The purpose of this chapter is to provide an alternative approach to the α-shapes methods for determining the boundaries $\partial R(\Pi)$ of the regions with positive luminance in target PS. The idea of the method is based on the triangulation refinement of the source PS explained in Chapter 4. The boundaries $\partial R(\Pi)$ are approximated by connecting those vertices of the triangles that follow path Π. Numerical results are provided for three optical systems. First we test the method for a very simple system formed by straight lines all located in the same medium (air). Next, we present the result for a system formed also by curved line and located in two different media: the TIR-collimator. Finally, we show the results for a more complicated system: a parabolic reflector. This system is formed by parabolic reflectors along which rays can have multiple reflections before arriving at the target.

The PS method is compared to both MC and QMC ray tracing. Discussion and results are provided in the last section of this chapter.

6.1 Reconstruction of the boundaries

In Chapter 4, we have seen that, using the triangulation refinement, more rays close to the boundaries are traced selecting increasingly smaller values for the parameters $\varepsilon_{q_1}^{\min}$ and $\varepsilon_{p_1}^{\min}$. Once the algorithm stops, only the triangles that are expected to be crossed by a boundary $\partial R(\Pi)$ are taken into account. From now on we call these triangles the *boundary triangles*. Two triangles are neighbors if they have a side in common. For each boundary triangle its neighbor is found so that an ordered sequence of triangles is construct. Given a path Π, the corresponding boundary $\partial R_1(\Pi)$ on S is approximated by the vertices of the boundary triangles which correspond to rays following path Π. The edge-ray principle is employed in order to define the corresponding boundaries

$\partial R(\Pi)$ at the target. Thus, $\partial R(\Pi)$ at the target are given by

$$\mathrm{M}(\partial R_1(\Pi)) : \partial R_1(\Pi) \to \partial \mathrm{R}(\Pi), \tag{6.1.1}$$

where M is defined in Equation (4.3.1) and $\mathrm{M}(\partial R_1(\Pi))$ is the restriction of M to $\partial R(\Pi)$ for every path Π.

In this chapter we develop a criterion to establish the value of the parameters $\varepsilon_{q_1}^{\min}$, $\varepsilon_{q_1}^{\max}$, $\varepsilon_{p_1}^{\min}$ and $\varepsilon_{p_1}^{\max}$ which gives a good approximation of $\partial R(\Pi)$. Similar to the selection of α in the α-shapes procedure, to select the triangulation parameters we use étendue conservation, i.e. conservation of area in PS. The core of our approach is the following.

The étendue U_1 at the source PS S related to the rays that arrive at the target is calculated. If all the rays emitted by the source are received by the target, U_1 can be easily determined by

$$U = 4 n_1 a \sin(\theta_1^{\max}) \,, \tag{6.1.2}$$

where a is the half length of the source, n_1 the index of refraction of the medium in which the S is located and $\theta_1^{\max}$ is the maximum value of the angle that the rays make with the normal $\boldsymbol{\nu}_1$ of the source. In case some rays do not arrive at the target and rather reach other detectors, we use

$$U_1 = \sum_{\Pi} U\big(R_1(\Pi)\big), \tag{6.1.3}$$

where the sum is over all possible paths Π and $U\big(R_1(\Pi)\big)$ is the contribution to the étendue given by the rays inside.

The étendue U_{t} at the target PS T is calculated using Equations (5.2.5) and (5.2.7). To calculate the integral in Equation (5.2.7), the triangulation refinement method is applied to the regions $R(\Pi)$ for a range of values of $\varepsilon_{q_1}^{\max}$ and for a fixed value of $\varepsilon_{q_1}^{\min}$. The parameters along the p-axis are scaled as

$$\begin{aligned} w &= \frac{q_1^{\max} - q_1^{\min}}{p_1^{\max} - p_1^{\min}}, \\ \varepsilon_{p_1}^{\min} &= \frac{\varepsilon_{q_1}^{\min}}{w}, \\ \varepsilon_{p_1}^{\max} &= \frac{\varepsilon_{q_1}^{\max}}{w}. \end{aligned} \tag{6.1.4}$$

where $p_1^{\min}$ and $p_1^{\max}$ are the minimum and the maximum p-coordinate in S, respectively, and $q_1^{\min}$ and $q_1^{\max}$ are the minimum and the maximum q-coordinate in S, respectively. Every set of parameters gives a certain triangulation, for each of them an approximation of the boundaries $\partial R(\Pi)$ is obtained. Next, the intersection points $(q^i(\Pi, p), \Pi)_{i=1,\cdots,r}$ between $\partial R(\Pi)$ and the horizontal line $p = \textit{constant}$ are calculated for every path Π, and for $p \in [-1, 1]$. Ordering their q-coordinates in ascending order, the integral in Equation (5.2.7) is computed. Changing the values of the parameters, different approximations of $\partial R(\Pi)$ are found and, consequently, different values of U_{t}. By construction, U_{t} is always underestimated ($U_{\mathrm{t}} < U_1$) because the approximated boundaries are found joining the vertices of the *boundaries*

triangles which are *inside* the regions $R(\Pi)$.

To use the parameters that give a good accuracy of the target photometric variables, the difference $\Delta U = U_1 - U_\mathrm{t}$ is calculated for every value of U_t found. The values of the parameters that give a small ΔU provide a triangulation refinement from which a good approximation of the target photometric variables can be computed.

A similar method as described here is presented by Moore, [54]. In Moore's method each ray leaves the source at the same position while the angle coordinate changes. The path followed by the rays is taken into account and an interpolation is required to finalize the illumination pattern. This interpolation can affect the efficiency of the method. Our method employs the distribution of the rays at the target PS and avoids using any interpolation. Moreover, a criterion to stop our algorithm is provided in such a way that no more rays than necessary are traced. This makes ray tracing in PS more accurate compared to Moore's procedure.

The triangulation refinement method is tested for several optical systems. The results are presented next.

6.2 The two-faceted cup

In this paragraph we apply the triangulation refinement in PS to the two-faceted cup described in Chapter 3 and depicted in Figure 3.1. We start tracing rays inside the system using PS ray tracing as explained in Chapter 4. To avoid rays parallel to the source and rays emitted from the endpoints, we consider their initial position q_1 and initial direction p_1 such that

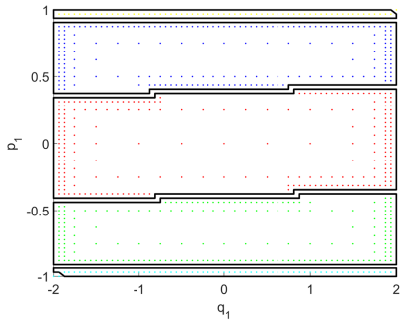
$$\begin{aligned} p_1 &\in [-1 + 10^{-6}, 1 - 10^{-6}] = [p_1^{\min}, p_1^{\max}], \\ q_1 &\in [-2 + 10^{-12}, 2 - 10^{-12}] = [q_1^{\min}, q_1^{\max}]. \end{aligned}$$

A stopping criterion for the triangulation is defined using étendue conservation. Since the two-faceted cup is formed by only reflective lines and its target is adjacent to the left and the right reflector (it is located exactly at the top of the system), all the rays emitted by the source arrive at the target. Thus,

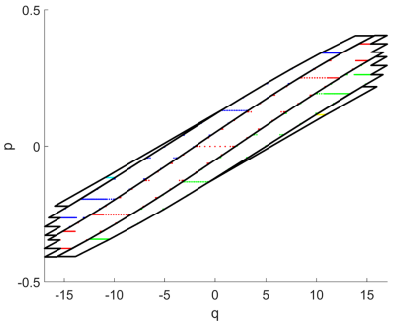
$$U_1 = U = 4(2 - 10^{-12})(1 - 10^{-6}) \approx 8, \tag{6.2.1}$$

where the second equality follows from Equation (6.1.2) with $n_1 \sin(\theta_1^{\max}) = p_1^{\max}$ and $a = q_1^{\max}$. To establish the number of rays needed to achieve a good accuracy of the target intensity, we compare the approximated U_t, obtained from a given number of rays, to the exact étendue $U = U_1$. To this purpose, ray tracing in PS is implemented by varying the parameter $\varepsilon_{q_1}^{\min}$, and fixing $\varepsilon_{q_1}^{\max}$ (we choose $\varepsilon_{q_1}^{\max} = 1$), while the other two parameters are given by Equation (6.1.4). Every set of parameters give a different triangulation a the source PS S. The approximated boundaries are computed for several triangulation joining the vertices of the triangles crossed by a boundary that follow the same path Π (from now on we refer to these triangle as the *boundary triangles*).

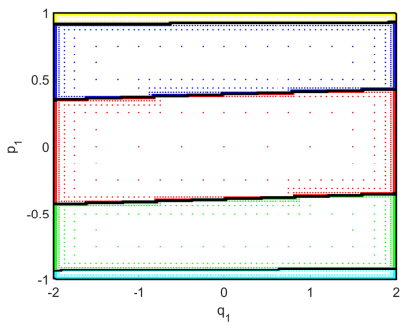
For example, if we consider $\varepsilon_{q_1}^{\min} = 0.1$, $\varepsilon_{q_1}^{\max} = 1$ and the parameters for the p-axis given by Equation (6.1.4), a triangulation with around 1500 rays (vertices of the triangles) is found. The boundaries $\partial R_1(\Pi)$ and $\partial R(\Pi)$ are calculated using this



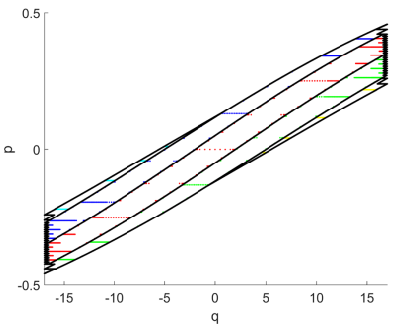
(a) The black lines are the boundaries at S. 1500 rays are traced using the triangulation refinement with $\varepsilon_{q_1}^{\min} = 0.1$.



(b) The black lines are the boundaries at T. 1500 rays are traced using the triangulation refinement with $\varepsilon_{q_1}^{\min} = 0.1$.



(c) The black lines are the boundaries at S. 7500 rays are traced using the triangulation refinement with $\varepsilon_{q_1}^{\min} = 0.025$.



(d) The black lines are the boundaries at T. 7500 rays are traced using the triangulation refinement with $\varepsilon_{q_1}^{\min} = 0.025$.

Figure 6.1: **Boundaries at S and T of the two-faceted cup.** The approximated boundaries are computed using the triangulation refinement with two different values of $\varepsilon_{q_1}^{\max}$.

triangulation refinement which are depicted in black in Figures 6.1a and 6.1b, respectively. For this set of rays we found $\Delta U \approx 0.53$. Next, we can decrease $\varepsilon_{q_1}^{\min}$ to obtain a more precise approximation of U_{t}. Choosing $\varepsilon_{q_1}^{\min} = 0.025$ and $\varepsilon_{q_1}^{\max} = 1$, a triangulation formed by around 7500 rays is obtained. The approximated boundaries $\partial R_1(\Pi)$ and $\partial R(\Pi)$ are depicted with black lines in Figures 6.1c and 6.1d, respectively. The approximation of the target étendue gives $\Delta U \approx 0.13$. Obviously, the boundaries computation obtained using $\varepsilon_{q_1}^{\min} = 0.025$ is more accurate. Note that, decreasing $\varepsilon_{q_1}^{\min}$, the number of rays increases.

In Figure 6.2 we show with the blue line how the target étendue varies as a function of the parameter $\varepsilon_{q_1}^{\min}$. The exact étendue $U = 8$ is depicted with the red line and it is computed using Equation (5.2.7). By decreasing $\varepsilon_{q_1}^{\min}$ an increase of U_{t} is observed. Referring as Figure 6.2, the best approximation of U_{t} is obtained tracing around $1.2 \cdot 10^5$ rays using PS ray tracing with parameters $\varepsilon_{q_1}^{\min} = 0.8 \cdot 10^{-4}$, $\varepsilon_{q_1}^{\max} = 1$,

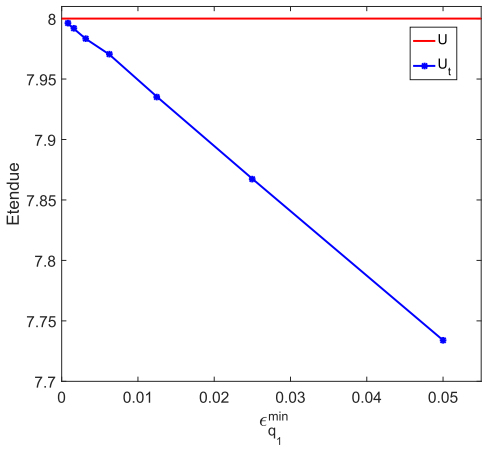


Figure 6.2: **Etendue for the two-faceted cup.** The total étendue as an area in PS is depicted with the red line. The approximated étendue for a range of values of $\varepsilon_{q_1}^{\min}$ is shown with the blue line.

$\varepsilon_{p_1}^{\min} = \varepsilon_{q_1}^{\min}/2$ and $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2$.

The PS intensity $\hat{I}_{PS}$ with $1.2 \cdot 10^5$ rays is calculated from Equation (4.3.4). The intensity profile is shown in Figure 6.3 with the red line. In the same graph we show the reference intensity $\hat{I}_{\text{ref}}$ with the dotted blue line. For the two-faceted cup the reference intensity is actually the exact intensity ($\hat{I}_{\text{ref}} = \hat{I}_{\text{exact}}$).

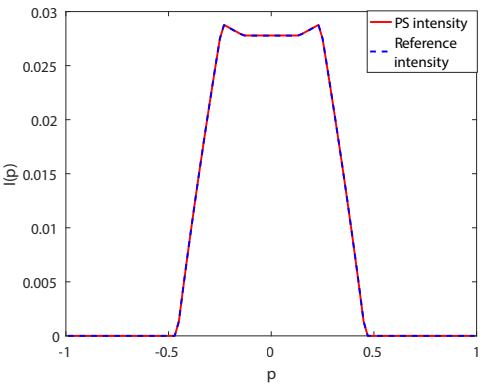


Figure 6.3: **Intensity profile at the target of the two-faceted cup.** The reference intensity is the exact intensity. The PS intensity is computed using the triangulation refinement with $\varepsilon_{q_1}^{\min} = 0.8 \cdot 10^{-4}$, $\varepsilon_{q_1}^{\max} = 1$, $\varepsilon_{p_1}^{\min} = \varepsilon_{q_1}^{\min}/2$ and $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2$. Around $1.2 \cdot 10^5$ rays are traced.

Finally, we compare PS ray tracing with both MC and QMC ray tracing by computing the error between the approximated intensities $\hat{I}_{\text{A}}(\text{A} = \text{MC}, \text{QMC}, \text{PS})$ and the

exact intensity $\hat{I}_{\text{ref}}$. For the error calculation we use Equation (5.3.5) with Nb = 100. The results are shown in Figure 6.4 where MC, QMC and PS intensity are depicted with the green, blue and red line, respectively. The graph shows that using PS and

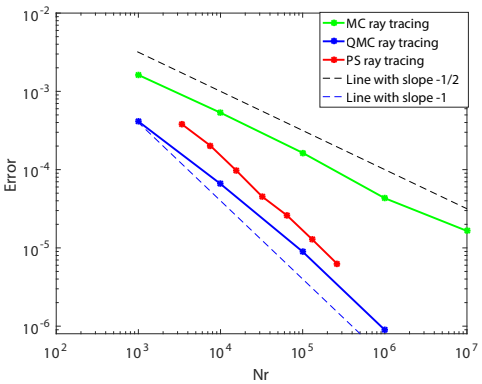


Figure 6.4: **Error plot for the two-faceted cup.** The errors between the approximated intensities $\hat{I}_{\text{A}}(\text{A} = \text{MC}, \text{QMC}, \text{PS})$ and the exact intensity $\hat{I}_{\text{exact}}$ are shown in a logarithmic scale. Using PS ray tracing far less rays need to be traced compared to MC ray tracing but still more rays than for QMC ray tracing.

the triangulation refinement for the boundaries computation admits tracing far less rays compared to MC ray tracing. On the other hand, a comparison between QMC ray tracing and PS ray tracing shows that more rays are needed in PS. Indeed, although the shapes of all the regions $R(\Pi)$ are very smooth, their boundaries at the edge of the target phase space T are difficult to approximate by triangles. With the triangulation refinement, the vertical and straight lines at the edge of T are always approximated by a broken line. On the other hand, since the two-faceted cup is a very simple system, QMC ray tracing does not requires a large number of rays to obtain the desired accuracy.

Nevertheless, PS ray tracing has a big advantage compared to QMC ray tracing. Indeed, as we have seen in Chapter 3, MC and QMC ray tracing are binning procedure. Therefore, the MC and QMC intensities are given by the average over every bin and the error also depends on the number of bins. PS ray tracing gives a pointwise intensity along all possible directions. In the simulations shown in this thesis we always compute the average PS intensity. This is needed to give a fair comparison of PS ray tracing versus MC and QMC ray tracing. It is very important to observe that no error related to the number of bins is involved in the PS procedure.

To investigate in more details the performance of PS ray tracing, we test the method for more complicated systems. In the next paragraph we present the results for a TIR-collimator.

6.3 A TIR-collimator

In this section we provide the results of PS ray tracing for a TIR-collimator, using the triangulation refinement to compute the boundaries $\partial R(\Pi)$ in target PS. In particular, we consider the TIR-collimator depicted in Figure 5.10. Since this system is located in two different media (air and glass), also the refraction law plays a role in the ray tracing procedure. We run PS ray tracing for the TIR-collimator several times gradually increasing the number of rays, i.e. gradually decreasing the values of the parameters $\varepsilon_{q_1}^{\min}, \varepsilon_{q_1}^{\max}, \varepsilon_{p_1}^{\min}$ and $\varepsilon_{p_1}^{\max}$ in the triangulation. In order to trace more rays close to the boundaries, we decide to vary only the value of $\varepsilon_{q_1}^{\min}$ and $\varepsilon_{p_1}^{\min}$ that determine the number of triangles close to the , fixing the values of $\varepsilon_{q_1}^{\max}$ and $\varepsilon_{p_1}^{\max}$ as they are responsible of the number of rays inside the regions $R(\Pi)$. Every ray traced has initial position coordinate $q_1 \in [-a, a]$ with $a = 2$ and the initial direction coordinate $p_1 = [-1, 1]$. Therefore, the source PS of the TIR-collimator is the rectangular S= $[-2, 2] \times [-1, 1]$. The parameters $\varepsilon_{p_1}^{\min}$ and $\varepsilon_{p_1}^{\max}$ are scaled as in Equation (6.1.4).

To determine the triangulation refinement that gives a good approximation of the target intensity we compare U_1 (source étendue) to U_{t} (target étendue) and use étendue conservation. In this case, not all light emitted by the source of the TIR-collimator arrives at the target. Indeed, using PS ray tracing, $\mathrm{Np} = 7$ different paths $(\Pi_j)_{j=1,\cdots,\mathrm{Np}}$ are found but only five of them are paths from the source (line 1) to the target (line 12), see also Section 5.3. Thus, we need to remove from the total area of S those parts occupied by the rays that arrive at some others detectors and not at the target. Indicating with A_T the area of each of these parts (see Section 5.3), the source PS is therefore given by:

$$U_1 = 8 - 2A_T \approx 7.77. \tag{6.3.1}$$

The target étendue U_{t} is obtained from Equation (5.2.6) for a range of values of $\varepsilon_{q_1}^{\min}$ and for $\varepsilon_{q_1}^{\max} = 1$ fixed. The boundaries $\partial R(\Pi)$ are found for every value of $\varepsilon_{q_1}^{\min}$, and U_{t} is calculated for each of these boundaries. The results shown in Figure 6.5 give the étendue plot as a function of $\varepsilon_{q_1^{\min}}$.

The best approximation of U_{t} shown in the previous graph is obtained using $\varepsilon_{q_1}^{\min} = 1.6 \cdot 10^{-3}$ in the triangulation refinement, tracing around $1.62 \cdot 10^5$ rays. The boundaries $\partial R_1(\Pi_j)$ and $\left(\partial R(\Pi_j)\right)_{j=1,\cdots,5}$ of the regions formed by these rays are shown in Figure 6.6 with the red lines.

The target PS intensity $\hat{I}_{\mathrm{PS}}$ is found and it is compared with a reference intensity $\hat{I}_{\mathrm{ref}}$ which is given by QMC ray tracing with 10^7 rays (as the exact intensity for the TIR-collimator is unknown). The profile of the two intensities is given in Figure 6.7. The difference between the two intensity profiles is not discernible by eye.

To validate our method, PS ray tracing is compared to both MC and QMC ray tracing. The error between the approximated intensities $\hat{I}_{\mathrm{A}}(\mathrm{A} = \mathrm{QMC}, \mathrm{MC}, \mathrm{PS})$ and the reference intensity $\hat{I}_{\mathrm{ref}}$ as a function of the number of rays traced is calculated. The error plot is shown in a logarithmic scale in Figure 6.8 where the MC, QMC and PS convergences are shown with the green, the blue and the red line, respectively. The black dotted line is a line with slope $-\frac{1}{2}$, the blue dotted line has slope -1.The graph shows that MC ray tracing converges, for $\mathrm{Nr} \to \infty$, with an order of $\mathcal{O}(\frac{1}{\sqrt{\mathrm{Nr}}})$, while both PS and QMC ray tracing have a speed of convergence of the order $\mathcal{O}(\frac{1}{\mathrm{Nr}})$. Note

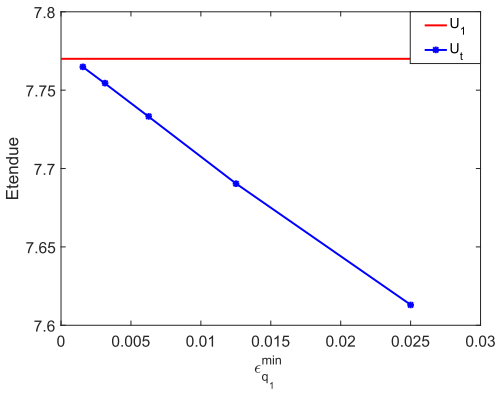


Figure 6.5: **Etendue of the TIR-collimator.** A comparison between U_1 and U_t shows that by decreasing the value of $\varepsilon_{q_1^{\min}}$, $\Delta U = U_1 - U_t$ decreases.

Table 6.1: **Errors of the PS intensity for the TIR-collimator**

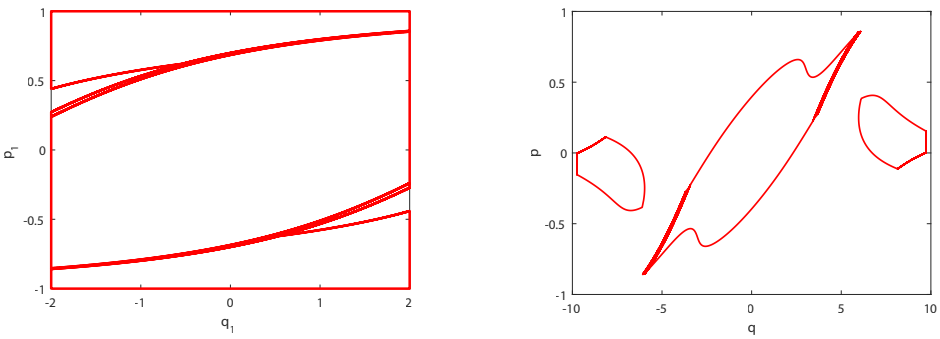
$\varepsilon_q^{\max}$	Nr	Etendue	PS error	PS CPU-time (sec.)
0.05	3 547	7.50	$1.75 \cdot 10^{-4}$	1.98
0.025	8 055	7.61	$1.49 \cdot 10^{-4}$	4.69
0.125	17 300	7.69	$8.68 \cdot 10^{-5}$	10.61
$6.3 \cdot 10^{-3}$	38 300	7.73	$4.43 \cdot 10^{-5}$	26.56
$3.1 \cdot 10^{-3}$	79 600	7.75	$2.27 \cdot 10^{-5}$	83, 21
$1.6 \cdot 10^{-3}$	162 300	7.76	$1.20 \cdot 10^{-5}$	240.53

that PS ray tracing allows tracing 10^2 times less rays compared to MC ray tracing and around 10 times less rays compared to QMC ray tracing.

Finally, in order to show the advantages of PS ray tracing in terms of the computational time, in Figure 6.9 we provide an error convergence as a function of the CPU-time for all the three methods (MC, QMC and PS raytracing). The choice of the colours is consistent with Figure 6.7.

The results shown in Figures 6.8 and 6.9 are reported in Tables 6.1 and 6.2. We observe that PS ray tracing outperforms both MC and QMC ray tracing. Using PS ray tracing an error equal to 10^{-4} is achieved tracing almost 10 time less compared to QMC ray tracing and 100 times less rays compared to QMC ray tracing. THis makes PS ray tracing 10 times faster than QMC ray tracing and 100 times faster than MC ray tracing.

Next we show the result for a system where more than 5 paths are possible. In particular we present the results for an optical system for which multiple reflections between rays and the mirrors can occur.



(a) Boundaries at the source PS.

(b) Boundaries at the target PS.

Figure 6.6: **Boundaries at S and T of the TIR-collimator.** The red lines show the boundaries found using $\varepsilon_{q_1}^{\min} = 1.6 \cdot 10^{-3}, \varepsilon_{q_1}^{\max} = 1, \varepsilon_{p_1}^{\min} = \varepsilon_{q_1}^{\min}/2$ and $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2$ in the triangulation refinement of the source PS.

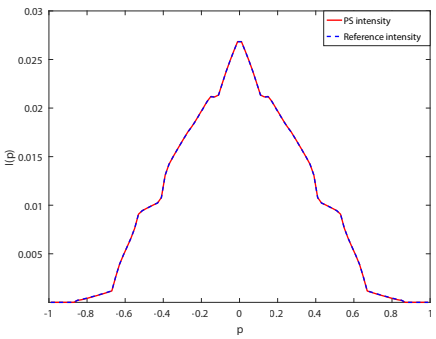


Figure 6.7: **Target intensity for the TIR-collimator.** The PS intensity $\hat{I}_{\text{PS}}$ is computed using PS ray tracing with around $1.62 \cdot 10^5$ rays. The reference intensity $\hat{I}_{\text{ref}}$ is obtained by QMC ray tracing with 10^7 rays.

6.4 A Parabolic reflector

In this section we show an example of a parabolic reflector the design of which is depicted in Figure 6.10. It consists of a source S (line 1), a target T (line 4) parallel to S and two reflectors (lines 2 and 3) which are arcs of the same parabola. The minimum of the parabola is located at the point with x-coordinate equal to 0. S = [-a, a] (with a = 2) and $\mathsf{T} = [-\text{b}, \text{b}]$ (with b = 17) are lines perpendicular to the optical axis (z-axis) and are located at $z = 0$ and $z = 40$, respectively. All the optical lines are located in air, therefore the index of refraction $n = 1$ for every line. The optical axis of the system in Figure 6.10 corresponds to the z-axis. We trace rays in PS with source direction coordinates $p_1 = [-1, 1]$ and source position coordinates $q_1 \in [-a + \varepsilon, a - \varepsilon]$ where $\varepsilon > 0$ is a small number. In particular we take $\varepsilon = 10^{-12}$.

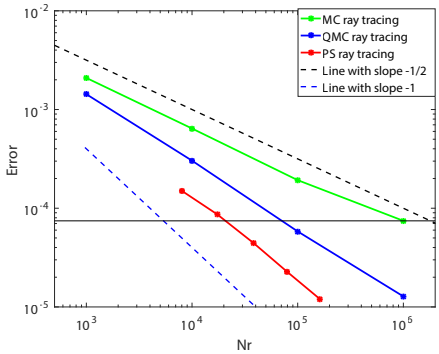


Figure 6.8: **Error as a function of the number of rays for the TIR-collimator.** The reference intensity $\hat{I}_{\text{ref}}$ is obtained by QMC ray tracing with 10^7 rays.

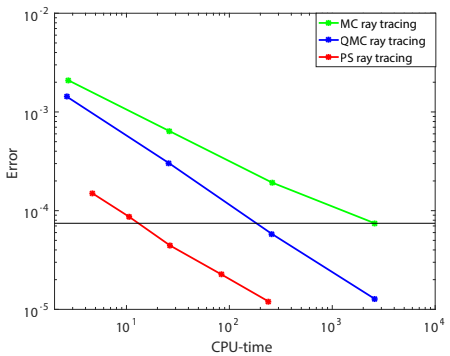


Figure 6.9: **Error as a function of the CPU-time for the TIR-collimator.** The reference intensity $\hat{I}_{\text{ref}}$ is obtained by QMC ray tracing with 10^7 rays.

As an example we show the triangulation refinement obtained for the parameters

$$\varepsilon_{q_1}^{\min} = 0.025,\ \varepsilon_{q_1}^{\max} = 0.25,\ \varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\min}/2,\ \text{and}\ \varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2,$$

for which around 7700 rays are traced in PS. Their distribution at S and T is shown in Figures 6.11a and 6.11b, respectively. The distribution of the rays in PS gives information about the path they follow. We note that for the parabolic reflector many paths are found. Every path corresponds to a given number of reflections. Rays can have multiple reflections at lines 2 and 3 before arriving at the target. These are the rays that leaves the end points of the source with direction tangential to the mirrors 2 and 3. The parameters used in the triangulation refinement establish not only the number of rays traced but also the number of paths detected. For instance, for the values of the parameters defined above, the triangulation refinement is able to detect 17 different paths. This means that until 8 multiple reflections can occur between the rays and the two mirrors. Counting the number of rays that follow a given path II, we can calculate the frequency of every path. For example, tracing around 7700

Table 6.2: **Errors of the MC and QMC intensities for the TIR-collimator**

Nr	MC error	MC CPU-time (sec.)	QMC error	QMC CPU-time (sec.)
10^3	$2.09 \cdot 10^{-3}$	2.73	$1.43 \cdot 10^{-3}$	2.63
10^4	$6.42 \cdot 10^{-4}$	25.98	$3.03 \cdot 10^{-4}$	25.84
10^5	$1.92 \cdot 10^{-4}$	259.92	$5.82 \cdot 10^{-5}$	258.28
10^6	$7.45 \cdot 10^{-5}$	2585.83	$1.28 \cdot 10^{-5}$	2482.67

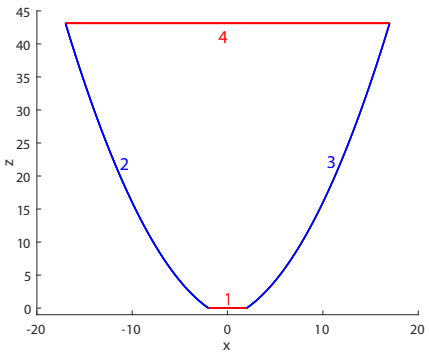


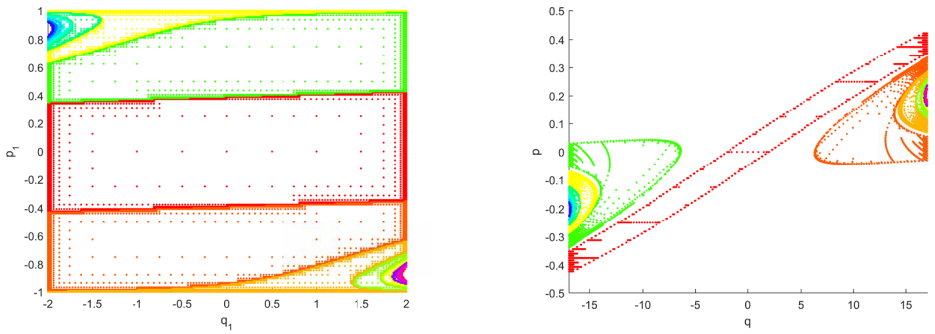
Figure 6.10: A parabolic reflector. Each line of the system is labeled with a number. The source $\mathsf{S}= [2, 2]$ (line 1) is located on the x-axis. The target $\mathsf{T}= [-17, 17]$ (line 4) is parallel to the source and is located at a height $z = 40$. The left and right reflectors (lines 2 and 3) are arcs of the same parabola.

rays, the frequency of the rays that have 8 multiple reflections along one of the two reflectors is around 0.13%. Rays that reflect many times before reaching the target do not give a significant contribution to the target intensity. Decreasing the value of the parameter $\varepsilon_{q_1}^{\min}$, more paths can be found. In order to find *all* the possible paths, also the parameter $\varepsilon_{q_1}^{\max}$ needs to be decreased. In fact, very small regions cannot be detected if $\varepsilon_{q_1}^{\max}$ is too large, see the example shown in Figure 4.5. Increasing the number of reflections considered, the corresponding regions in PS become smaller and smaller, see Figure 6.11.

Like for the optical systems considered in the previous sections, a stopping criterion of the triangulation refinement is determined for the parabolic reflector. Etendue conservation is used in order to find the values of the parameters that give the best approximation of the boundaries of the regions with positive luminance in PS. For the parabolic reflector in Figure 6.10 all the rays that leave the source arrive at the target. Indeed, line 2 and 3 can only reflect rays (refraction law is not involved) and the target coincides with the source. From Equation (6.1.2) we obtain:

$$U = U_1 = 4(a - \varepsilon) \approx 8. \tag{6.4.1}$$

A range of values of $\varepsilon_{q_1}^{\min}$ and $\varepsilon_{q_1}^{\max}$ is considered (the triangulation parameters for the p-axis depend on the q-axis parameters according to Equation (6.1.4)). For each couple of values, an approximation of the boundaries $\partial R(\Pi)$ is found for every path Π as explained. U_{t} is calculated for the approximated boundaries using Equation



(a) Source PS $S = [-a + \varepsilon, a - \varepsilon] \times [-1, 1]$ (b) Target PS $T = [-b, b] \times [-1, 1]$

Figure 6.11: **Rays distribution at S and T of the parabolic reflector.** Around 7700 rays are traced using PS ray tracing with parameters $\varepsilon_{q_1}^{\min} = 0.025$, $\varepsilon_{q_1}^{\max} = 0.25$, $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\min}/2$, and $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2$. 17 different paths are found, each of them correspond to a certain number of reflections.

Table 6.3: **Results of the triangulation refinement.**

$\varepsilon_q^{\min}$	$\varepsilon_q^{\max}$	Nr	Np	Etendue
0.2	1	240	9	5.71
0.125	1	625	11	6.70
0.1	1	1 500	15	7.23
0.025	0.5	7 700	17	7.65
0.025/2	0.5/2	17 400	17	7.82
0.025/8	0.5/8	81 500	17	7.96
0.025/16	0.5/16	176 215	17	7.98

(5.2.6). In Table 6.3 we show how the number of rays traced, the paths found and the value for the target étendue depend on the triangulation parameters. We observe that, decreasing $\varepsilon_{q_1}^{\min}$ and $\varepsilon_{q_1}^{\max}$ the number of both the rays traced and the paths found increases. A maximum of 17 different paths are detected. Furthermore, the value of U_{t} gets closer and closer to the exact étendue U.

Using the triangulation refinement that gives the best étendue approximation, we calculate the target PS intensity $\hat{I}_{\mathrm{PS}}$ from Equation (4.3.3). The intensity profile is shown with the red line in Figure 6.12. The PS intensity is compared to a reference intensity $\hat{I}_{\mathrm{ref}}$, computed using QMC ray tracing with 10^7 rays. $\hat{I}_{\mathrm{ref}}$ is depicted Figure 6.12 with the dotted blue line. The graph shows that the two intensities coincide.

Now, our method is compared to both MC and QMC ray tracing. The error between the approximate intensities $\hat{I}_{\mathrm{A}}(\mathrm{A} = \mathrm{PS}, \mathrm{MC}, \mathrm{QMC})$ and the reference intensity is calculated. In Figure 6.13 the error as a function of the number of rays traced is shown for the three methods. The green line represents MC error, the blue line pictures the QMC error and the red line depicts PS error. The errors are shown in a logarithmic scale, the dotted black line has slope $-\frac{1}{2}$, the dotted blue line has slope -1. Likewise for the other systems, MC ray tracing converges proportionally to $\frac{1}{\sqrt{\mathrm{Nr}}}$.

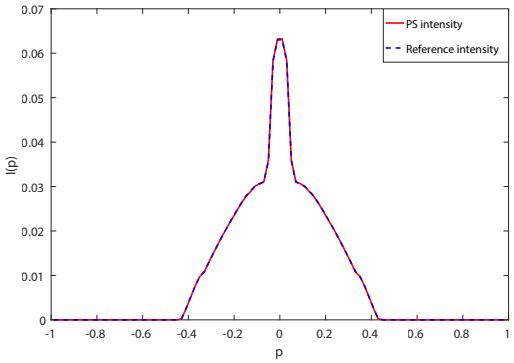


Figure 6.12: **Target intensity of the parabolic reflector.** For the PS intensity the parameters $\varepsilon_{q_1}^{\min} = 1.56 \cdot 10^{-3}$, $\varepsilon_{q_1}^{\max} = 3.13 \cdot 10^{-2}$, $\varepsilon_{p_1}^{\min} = \varepsilon_{q_1}^{\min}/2$, and $\varepsilon_{p_1}^{\max} = \varepsilon_{q_1}^{\max}/2$ are used. Around $8.15 \cdot 10^4$ rays are traced in PS. For the reference intensity QMC ray tracing with 10^7 rays is implemented.

Table 6.4: **Errors of the PS intensity for the parabolic reflector**

Nr	PS error	CPU-time (sec.)
241	$1.10 \cdot 10^{-3}$	0.14
625	$8.60 \cdot 10^{-4}$	0.20
1 500	$5.16 \cdot 10^{-4}$	0.44
7 700	$2.70 \cdot 10^{-4}$	2.07
17 400	$1.27 \cdot 10^{-4}$	5.78
81 500	$2.45 \cdot 10^{-5}$	60.35
176 215	$1.10 \cdot 10^{-5}$	241.79

QMC convergence is proportional to $\frac{1}{\mathrm{Nr}}$. The PS convergence behaviour is similar to MC ray tracing until all the possible paths are found. Once all the paths are detected, PS ray tracing converges proportional to $\frac{1}{\mathrm{Nr}}$. Less rays compared to both MC and QMC ray tracing are needed to trace in PS. In particular, to achive an error of 10^{-4}, around 10 times less rays in PS are needed compared to MC and few rays less than QMC.

Finally, the error as a function of the CPU-time is shown in Figure 6.14. The MC, QMC and PS errors are depicted with the green, blue and red line, respectively. We observe that PS ray tracing is 10 times faster than MC ray tracing. Note that PS ray tracing is faster than QMC ray tracing if an error of around 10^{-4} is desired, for a better accuracy, PS ray tracing becomes slightly slower than QMC ray tracing. The detailed results of the numerical simulations are reported in Tables 6.5 and **??**.

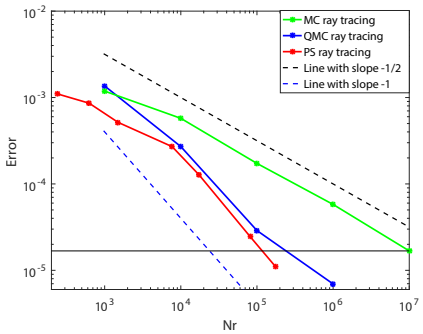


Figure 6.13: **Error as a function of the number of rays traced.** Less rays are needed using PS ray tracing compared to both MC and QMC ray tracing.

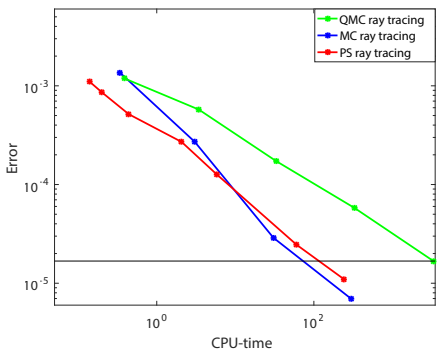


Figure 6.14: **Error as a function of the CPU-time.** PS ray tracing has significant advantages in terms of the CPU-time compared to MC ray tracing. For the parabolic reflector the computational time is comparable with QMC ray tracing.

6.5 Discussion and conclusions

In this chapter we presented a method to calculate the boundaries of the regions with positive luminance in PS. This method does not depend on a parameter α needed for the α-shape method presented in Chapter 5. Indeed, given a triangulation at the source PS, the boundaries are computed connecting the vertices of the boundary triangles, i.e. triangles crossed by a boundary, that follow the same path. Employing étendue conservation, a stopping criterion for the triangulation refinement was developed. We applied the method to three different optical systems: the two-faceted cup, a TIR-collimator, and a parabolic reflector. Numerical results show that PS ray tracing is faster and more accurate than MC ray tracing. Compared to QMC ray tracing we observed accuracy and speed advantages of an order of magnitude with our method for the TIR-collimator. For the two-faceted cup, PS ray tracing has a slower convergence compared to QMC ray tracing. For the parabolic reflector PS and QMC ray tracing display similar convergence. The conclusion is that QMC ray

Table 6.5: **Errors of the MC intensity for the parabolic reflector**

Nr	MC error	CPU-time (sec.)
10^3	$1.18 \cdot 10^{-3}$	0.39
10^4	$5.74 \cdot 10^{-4}$	3.43
10^5	$1.73 \cdot 10^{-4}$	33.13
10^6	$5.79 \cdot 10^{-5}$	328.96

Table 6.6: **Errors of the QMC intensity for the parabolic reflector**

Nr	QMC error	CPU-time (sec.)
10^3	$1.36 \cdot 10^{-3}$	0.34
10^4	$2.71 \cdot 10^{-4}$	3.05
10^5	$2.89 \cdot 10^{-5}$	30.58
10^6	$6.96 \cdot 10^{-6}$	298.22

tracing performs better than PS ray tracing for very simple optical systems, but the PS approach is more suitable for more complicated optical systems.

In order to further improve PS ray tracing we develop a new method which employs the PS of *all* the optical lines. The approach is explained in the next chapter.

Chapter 7

Inverse ray mapping: analytic approach

In the previous chapter we have seen that PS ray tracing based on the source and the target PS constitutes an improvement of MC and QMC ray tracing. Now, a method that employs not only the source and the target PS but also the PS of *all* the other lines that constitute the optical system is introduced. In this chapter, we consider systems formed only by straight and reflective line segments. All lines can be modeled as detectors of the incident light and emitters of the reflected light, they constitute the target for incident rays and the source for reflected rays. Moreover, we assume that the source can only emit light and the target can only receive light. Therefore, two different phase spaces are considered for the reflectors and one PS for S and T. All these phase spaces are connected through a map which relates the ray coordinates on every PS. This map can be written as the concatenation of many maps which can be classified as two different kind of maps, i.e., the map that connects the source and the target PS of two *different* lines and the map that connects the target and the source PS of the *same* line. Employing the inverses of these maps we are able to detect the parts of target PS illuminated by the source. All the PS considered are divided into regions, the boundaries of which can be determined exactly for systems formed by straight lines. We make the assumption of a Lambertian source. As a consequence, the output intensity along a given direction is given by the total width of all the patches with positive luminance, measured along that direction.

In this chapter, two different optical systems are investigated: the two-faceted cup and the so-called multi-faceted cup. Next, the details of the procedure are explained for the two-faceted-cup.

7.1 Phase spaces of the two-faceted cup

A two-faceted cup is introduced in Chapter 3 and depicted in Figure 3.1. It is formed by a source S, a target T and two reflectors which are straight lines segments. We use the same notation of Chapter 4 indicating with $S = Q \times P$ the PS and with (q, p) the rays coordinates in S.

Let's now introduce some new notation. The source and the target PS of a line j are indicated with S_j and T_j, respectively. The coordinates of every ray that reaches the line $j \in \{2,3,4\}$ are indicated with $(q_{\mathrm{t},j}, p_{\mathrm{t},j})$ on T_j. In the following, to simplify the notation, we indicate the target coordinates of the rays on T_4 with $(q,\, p)$ instead of $(q_{\mathrm{t},4}, p_{\mathrm{t},4})$. After reflection, the ray leaves line $j \in \{2,3,4\}$ at the same position but with a different direction, the new ray coordinates are indicated with $(q_{\mathrm{s},j}, p_{\mathrm{s},j})$ on S_j. Note that $q_{\mathrm{s},j} = q_{\mathrm{t},j}$, while $p_{\mathrm{s},j}$ is obtained applying the reflection law to the direction coordinate $p_{\mathrm{t},j}$ of the incident ray. The phase spaces S_j and T_j of each line j are partitioned into different regions, $(S_{j,k})_{k=2,3,4}$ and $(T_{j,l})_{l=1,2,3}$, respectively, where $k \neq j$ is the index of the line that is illuminated by j and $l \neq j$ is the index of the line that illuminates j. Hence, $S_{j,k} \subset S_j$ is the part of S_j corresponding to rays that illuminate line k, and $T_{j,l} \subset T_j$ is the part of T_j corresponding to rays originating from the line l. Note that, due to the fact that the source only emits light, we do not define its target PS T_1. Similarly, since the target only receives light, its source PS S_4 is not defined. For the two-faceted cup, six different phase spaces need to be considered which are given by the following expressions:

$$\begin{aligned} S_1 &= S_{1,2} \cup S_{1,3} \cup S_{1,4}, \\ S_2 &= S_{2,3} \cup S_{2,4}, \\ S_3 &= S_{3,2} \cup S_{3,4}, \\ T_2 &= T_{2,1} \cup T_{2,3}, \\ T_3 &= T_{3,1} \cup T_{3,2}, \\ T_4 &= T_{4,1} \cup T_{4,2} \cup T_{4,3}. \end{aligned} \tag{7.1.1}$$

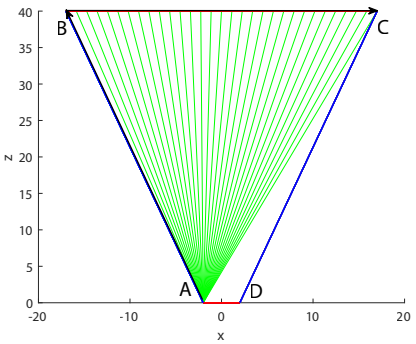
We need to note that, as the source cannot receive light and the target cannot emit light, the regions $(S_{j,1})_{j=2,3}$ and $(T_{j,4})_{j=2,3}$ are not considered.

The boundaries $\partial S_{j,k}$ are mapped into the boundaries $\partial T_{k,j}$ for every $j = \{1,2,3\}$ and $k = \{2,3,4\}$ with $k \neq j$ (edge-ray principle). For the two-faceted cup, and for all systems formed by straight line segments, these boundaries are determined analytically. Given two lines j and k with $j \neq k$, the boundaries S_j and T_k are determined as follows. Let $(x_{j,\ell}, z_{j,\ell})$ and $(x_{j,\mathrm{r}}, z_{j,\mathrm{r}})$ be the coordinates of the points located at the left and the right extreme of line j, respectively. Similarly, $(x_{k,\ell}, z_{k,\ell})$ and $(x_{k,\mathrm{r}}, z_{k,\mathrm{r}})$ are the coordinates of the points located at the left and the right extreme of line k, respectively. The boundaries $\partial S_{j,l}$ and $\partial T_{l,j}$ are obtained considering all the rays that leave the end points of line j and all the rays that reach the end points of k. They are formed by four different curves, two of them are given by all the rays that leave the end points of line j and hit line l and, the others two are given by the rays that leave the interior of line j and hit the end points of line l. The boundaries $\partial S_{j,l}$ and $\partial T_{l,j}$ are given by:

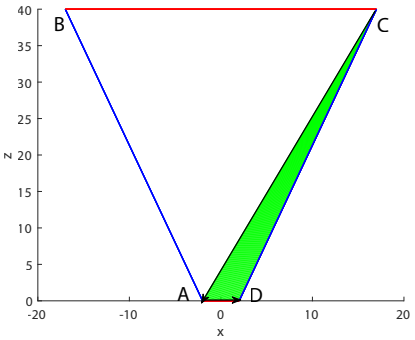
$$\begin{aligned} \partial S_{j,l} &= \partial S^1_{j,l} \cup \partial S^2_{j,l} \cup \partial S^3_{j,l} \cup \partial S^4_{j,l}, \\ \partial T_{l,j} &= \partial T^1_{l,j} \cup \partial T^2_{l,j} \cup \partial T^3_{l,j} \cup \partial T^4_{l,j}. \end{aligned} \tag{7.1.2}$$

In the following we explain in more details the case for $j = 1$ and $l = 4$, see Figure 7.1.

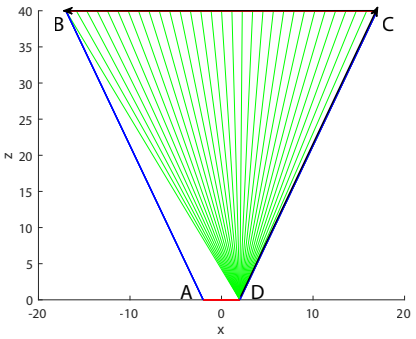
The boundaries $\partial S_{1,4}$ and $\partial T_{4,1}$ are given in Figs. 7.2 and 7.3, respectively. $\partial S^1_{1,4}$ and $\partial T^1_{4,1}$ are obtained tracing out line 4 from $q_\ell = -b$ to $q_{\mathrm{r}} = b$ by rays leaving $q_{1,\ell} = -a$



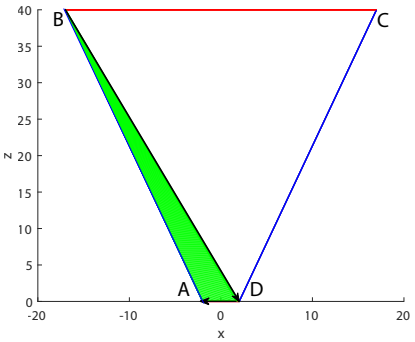
(a) Rays that leave the left end point of the source (line 1) and trace out the target (line 4).



(b) Rays that trace out the source (line 1) and hit the right end point of the target (line 4).



(c) Rays that leave the right end point of the source (line 1) and trace out the target (line 4).



(d) Rays that trace out the source (line 1) and hit the left end point of the target (line 4).

Figure 7.1: **Rays located on the boundaries of the regions $\partial \boldsymbol{S}_{1,4}$ and $\partial \boldsymbol{T}_{4,1}$.** $A = (x_{1\ell}, z_{1,\ell}) = (-2, 0)$ and $D = (x_{1,\mathrm{r}}, z_{1,\mathrm{r}}) = (2, 0)$ are the left and right corner points (or end points) of S (line 1). $B = (x_{4,\ell}, z_{4,\ell}) = (-17, 40)$ and $C = (x_{4,\mathrm{r}}, z_{4,\mathrm{r}}) = (17, 40)$, are the left and right corner points of T (line 4).

with varying p_1, these rays are shown in Fig. 7.1a, and the boundary segments $\partial S^1_{1,4}$ and $\partial T^1_{4,1}$ are the orange line segments labeled with c. $\partial S^2_{1,4}$ and $\partial T^2_{4,1}$ are given tracing out line 1 from $q_{1,\ell} = -a$ to $q_{1,\mathrm{r}} = a$ with varying p_1, such that all rays hit $q_{\mathrm{r}} = b$, these rays are shown in Fig. 7.1b, the boundary segments $\partial S^2_{1,4}$ and $\partial T^2_{4,1}$ are depicted in blue (lines segments labeled with d). Likewise, $\partial S^3_{1,4}$ and $\partial T^3_{4,1}$ are obtained tracing out line 4 from $q_{\mathrm{r}} = b$ to $q_\ell = -b$ by rays leaving $q_{1,\mathrm{r}} = x_{1,\mathrm{r}} = a$ with varying p_1. These rays are shown in Fig. 7.1c, $\partial S^3_{1,4}$ and $\partial T^3_{4,1}$ are the red line segments labeled with e. Finally, $\partial S^4_{1,4}$ and $\partial T^4_{4,1}$ are given tracing out line 1 from $q_{1,\mathrm{r}} = a$ to $q_{1,\ell} = -a$ with varying p_1, such that all rays hit $q_\ell = -b$, these rays are shown in Fig. 7.1d, $\partial S^4_{1,4}$ and $\partial T^4_{4,1}$ are the green lines segments labeled with f. We remind the reader that we use the notation (x, z) for the Cartesian coordinates of the optical system, while PS has (q, p) coordinates. It is worth noting that $q_{1,\ell} = x_{1,\ell}$, $q_{1,\mathrm{r}} = x_{1,\mathrm{r}}$, $q_\ell = x_{4,\ell}$ and $q_{\mathrm{r}} = x_{4,\mathrm{r}}$.

For the two-faceted cup there is an analytic expression for every line segment $\partial S^m_{j,k}$ and $\partial T^m_{k,j}$ in Eq. (7.1.2) with $m \in \{1, \cdots, 4\}$. For instance, the rays on the boundaries $\partial S^1_{j,k}$ and $\partial T^1_{k,j}$ are parameterized in the $(x,\, z)$-plane by

$$\boldsymbol{r}_{j,k}(t) = \begin{pmatrix} x_{k,\ell} - x_{j,\ell} + t(x_{k,\mathrm{r}} - x_{k,\ell}) \\ z_{k,\ell} - z_{j,\ell} + t(z_{k,\mathrm{r}} - z_{k,\ell}) \end{pmatrix} \qquad 0 \le t \le 1\,. \tag{7.1.3}$$

These rays are located on a vertical line segment in S_j as only the p_j-coordinate changes and on a curved line in T_k as both the target position and direction vary. The analytic expressions for $\partial S^1_{j,k}$ and $\partial T^1_{k,j}$ are

$$\partial S^1_{j,k}(t) = \left\{ (q_j, p_j) = \Big(q_{j,\ell}, |\boldsymbol{\nu}_j \times \hat{\boldsymbol{r}}_{j,k}(t)| \Big) \right\}, \tag{7.1.4}$$

$$\partial T^1_{k,j}(t) = \left\{ (q_k, p_k) = \Big(q_{k,\ell} - q_{j,\ell} + t(q_{k,\mathrm{r}} - q_{k,\ell}), |\boldsymbol{\nu}_k \times \hat{\boldsymbol{r}}_{j,k}(t)| \Big) \right\}, \tag{7.1.5}$$

where we have indicated with $\hat{\boldsymbol{r}}_{j,k}(t)$ the normalization of the ray in Eq. (7.1.3) and $\boldsymbol{\nu}_j$ and $\boldsymbol{\nu}_k$ are the normalized inward normals to lines j and k, respectively. Note that $\sin\tau_j = |\nu_j \times \hat{\boldsymbol{r}}_{j,k}(t)|$ and $\sin\tau_k = |\nu_k \times \hat{\boldsymbol{r}}_{j,k}(t)|$. Likewise, the boundaries $\partial S^m_{j,l}$ and $\partial T^m_{l,j}$ are calculated for every $m \in \{2, 3, 4\}$. Finally, $\partial S_{j,l}$ and $\partial T_{l,j}$ are found using Eq. (7.1.2).

In Figs. $7.4 - 7.9$, $(\partial S_{j,k})_{j\neq k=2,3,4}$ and $(\partial T_{j,l})_{j\neq l=1,2,3}$ are depicted in blue and red, respectively. The source and target PS of lines 2 and 3 have some empty regions. These parts correspond to the regions formed by the rays that either go back to the source or are emitted from the target. These regions are not taken into account, see Eq. (7.1.2). We observe that, because of the symmetry of the optical system, S_3 is the mirror image of S_2 after reflection in the central point $(q, p) = (-9.5, 0)$ followed by a translation $(q, p) \to (q + 19, p)$. Likewise T_3 is the mirror image of S_2 after the same reflection and translation.

7.1.1 Computation of the target photometric variables

In this section we explain how to compute the target photometric variables in PS. The intensity I along a given direction $p \in [-1, 1]$ in target phase space T_4 is a function

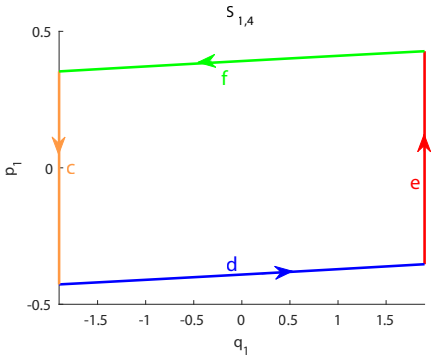


Figure 7.2: Source phase space of line 1. Boundary of the region $S_{1,4}$.

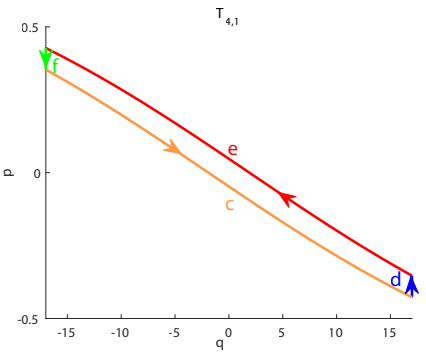


Figure 7.3: Target phase space of line 4. Boundary of the region $T_{4,1}$.

of the luminance $L(q,p)$ defined as in Equation (4.3.3). For the two-faceted cup, it becomes:

$$I_{\mathrm{PS}}(p) = \int_{-b}^{b} L(q,p)\mathrm{d}q\,. \tag{7.1.6}$$

The parts of T_4 that are illuminated by S_1 correspond to parts with positive luminance, for the other parts the luminance is equal to 0. Assuming positive luminance on S, the following relations hold:

$$\begin{aligned} &L(q,p) > 0 \qquad && \forall (q,p) \in T_{4,1}, \\ &L(q,p) \geq 0 \qquad && \forall (q,p) \in (T_{4,j})_{j=2,3}. \end{aligned} \tag{7.1.7}$$

Once a ray leaves the source S it can hit the reflectors several times before hitting the target T. To relate S and T, a map $\mathrm{M}_{1,4}$: $S_1 \to T_4$ is introduced such that $\mathrm{M}_{1,4}(q_{\mathrm{s},1}, p_{\mathrm{s},1}) = (q,p)$. As not all parts of T_4 are illuminated by the source S, the map $\mathrm{M}_{1,4}$ is not surjective. Therefore, we need to determine the subsets of T_4 illuminated by S corresponding to the regions where the luminance is positive. To this purpose, we consider two different kinds of maps. The first map relates the coordinates of the source and the target PS of two *different* lines, we call it the *propagation map*. The second map relates the coordinates of the target and the source PS of the *same* line, we call it the *reflection map*. In particular, given two lines j and k with $j \neq k$, the propagation map $\mathrm{P}_{j,k} : S_{j,k} \mapsto T_{k,j}$ relates $S_{j,k}$ with $T_{k,j}$ and, it is defined as follows:

$$\mathrm{P}_{j,k}(q_{\mathrm{s},j}, p_{\mathrm{s},j}) = (q_{\mathrm{t},k}, p_{\mathrm{t},k}), \tag{7.1.8}$$

where $q_{\mathrm{t},k}$ is given by the x-coordinate of the intersection point between the ray and line k, and $p_{\mathrm{t},k}$ is computed considering the direction of the incident ray with respect to the normal of line k. For one single line k, the reflection map $\mathrm{R}_{k,l,h}$: $T_{k,l} \mapsto S_{k,h}$ relates the regions $T_{k,l} \subset T_k$ and $S_{k,h} \subset S_k$. To simplify the notation, from now on we omit the dependence of $\mathrm{R}_{k,l,h}$ from l and h, i.e., $\mathrm{R}_{k,l,h} = \mathrm{R}_k$. The reflection map is defined as:

$$\mathrm{R}_k(q_{\mathrm{t},k}, p_{\mathrm{t},k}) = (q_{\mathrm{s},k}, p_{\mathrm{s},k}), \tag{7.1.9}$$

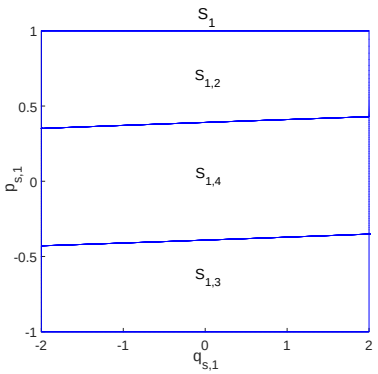


Figure 7.4: The PS S_1 of line 1 is partitioned into regions $(S_{1,k})_{k=2,3,4}$ formed by rays that leave line 1 and hit line k.

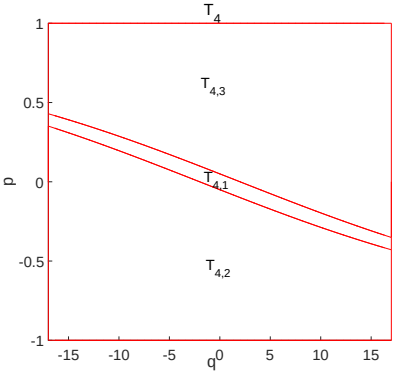


Figure 7.5: The PS T_4 of line 4 is partitioned into regions $(T_{4,l})_{l=1,2,3}$ formed by rays that leave line l and hit line 4.

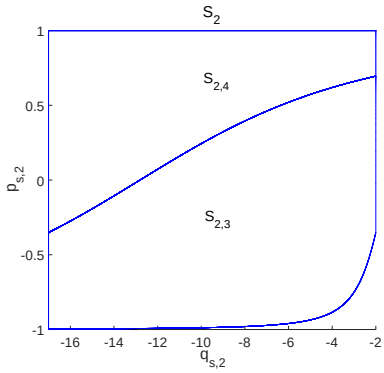


Figure 7.6: The PS S_2 of line 2 is partitioned into regions $(S_{2,k})_{k=3,4}$ formed by rays that leave line 2 and hit line k.

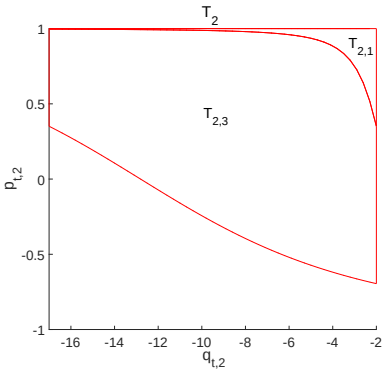


Figure 7.7: The PS T_2 of line 2 is partitioned into regions $(T_{2,l})_{l=1,3}$ formed by rays that leave line l and hit line 2.

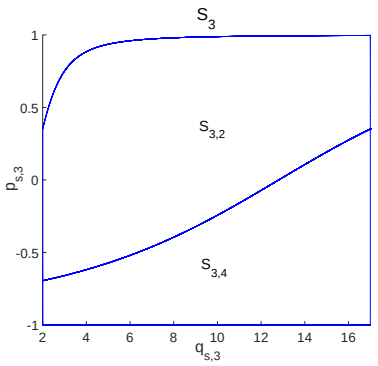


Figure 7.8: The PS S_3 of line 3 is partitioned into regions $(S_{3,k})_{k=2,4}$ formed by rays that leave line 3 and hit line k.

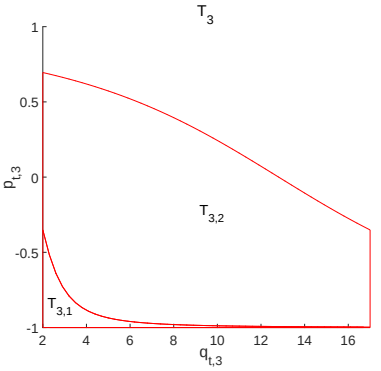


Figure 7.9: The PS T_3 of line 3 is partitioned into regions $(T_{3,l})_{l=1,2}$ formed by rays that leave line l and hit line 3.

where $p_{\mathrm{t},k}$ changes according to the reflection law and $q_{\mathrm{t},k} = q_{\mathrm{s},k}$ as R_k maps the target PS into the source PS of the same line k, that is S_k into T_k. Using a procedure similar to the ray transport matrices approach (see [10], Chapter 6), the map $\mathrm{M}_{1,4}$ is described by the composition of $\mathrm{P}_{j,k}$ and R_k defined in Eqs. (7.1.8) and (7.1.9). This composition depends on the path Π followed by the rays. We indicate with $\mathrm{M}_{1,4}(\Pi)$ the map $\mathrm{M}_{1,4}$ restricted to path Π and with $R(\Pi) \subset T_4$ the regions on T_4 formed by the rays that follow path Π. Considering all the possible paths Π from S to T, all the regions $R(\Pi)$ with positive luminance on T_4 can be determined.

To clarify this concept, we provide the following example. Consider a ray that is emitted from the source (line 1), hits the left reflector (line 2) and finally reaches the target (line 4). The path Π followed by this ray is defined as $\Pi = (1, 2, 4)$ and the corresponding map $\mathrm{M}_{1,4}(\Pi) : S_1 \mapsto R(\Pi)$ that describes the propagation of all rays that follow path Π is defined by:

$$\mathrm{M}_{1,4}(\Pi) : S_{1,2} \mapsto T_{2,1} \mapsto S_{2,4} \mapsto T_{4,2}, \tag{7.1.10}$$

which can be written as:

$$\mathrm{M}_{1,4}(\Pi) = \mathrm{P}_{2,4} \circ \mathrm{R}_2 \circ \mathrm{P}_{1,2}\,. \tag{7.1.11}$$

In general, to construct the map $\mathrm{M}_{1,4}(\Pi)$ we need to know its corresponding path Π. To determine all possible paths Π, instead of tracing the rays from S to T, we start considering the rays in T_4. In particular, along a given direction $p \in [-1, 1]$ we consider the intersection points between the line $p = \text{const}$ and $(\partial T_{4,j})_{j=1,2,3}$. These points are traced back to line j from which they are emitted and their corresponding coordinates on S_j and T_j are computed. This is done applying sequentially the maps $\mathrm{P}_{j,4}^{-1} : T_{4,j} \mapsto S_{j,4}$ and $\mathrm{R}_j^{-1} : S_j \mapsto T_j$. Then the same procedure is repeated considering these new coordinates on T_j. The computation stops either when the points found are emitted from the source, that is when they are located on S_1, or when they reach again the target, that is when they are located on T_4. If a ray reaches S_1, then a path Π from S to T is found. If a ray reaches again the target T_4, then we conclude that it is not emitted by S and therefore, it is located inside the parts of T_4 with luminance equal to 0.

Finally, the inverse $\mathrm{M}_{1,4}^{-1}(\Pi)$ of the map $\mathrm{M}_{1,4}(\Pi)$ is constructed for every possible path Π. The map $\mathrm{M}_{1,4}^{-1}(\Pi)$ is the composition of the inverses of the propagation and the reflection maps in reverse order according to the path Π. For instance, for path $\Pi = (1, 2, 4)$, $\mathrm{M}_{1,4}^{-1}(\Pi)$ is given by:

$$\mathrm{M}_{1,4}^{-1}(\Pi) = \mathrm{P}_{1,2}^{-1} \circ \mathrm{R}_2^{-1} \circ \mathrm{P}_{2,4}^{-1}. \tag{7.1.12}$$

The steps of the procedure are shown in Figure 7.10 where the map in Eq. (7.1.12) is written in red.

Using the procedure explained above, given a ray with coordinates $(q, p) \in T_4$ we can establish whether it is located inside one of the regions $R(\Pi)$ with positive luminance or not. In case the ray is inside a region $R(\Pi)$, its corresponding coordinates $(q_{\mathrm{s},1}, p_{\mathrm{s},1}) \in S_1$ are obtained using $\mathrm{M}_{1,4}^{-1}(\Pi)$, where Π is the path followed by this ray. The luminance in Equation (7.1.7) is, therefore, defined as in Equation (4.3.2), for some path Π connecting S and T. The target intensity is calculated from Equation

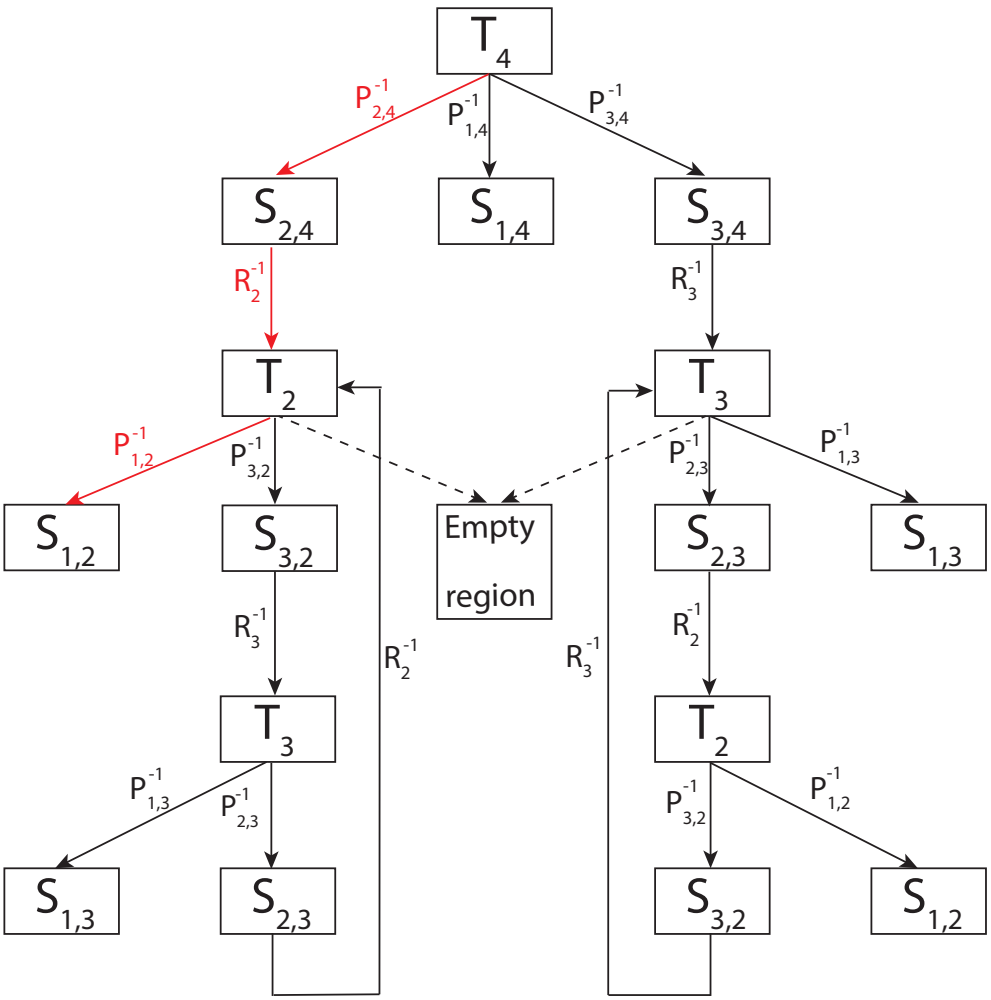


Figure 7.10: Tree that describes how to detect all the possible paths from S to T.

(7.1.6). Indicating with $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$ the minimum and maximum position coordinates of the intersection points between the boundaries $\partial R(\Pi)$ and the line $p = \text{const}$, Equation (7.1.6) reduces to Equation (4.3.4), if only two intersection points are found, and to Equation (5.2.8) in case more than two intersection points occur. For the two-faceted cup there are only two intersection points between line $p = \text{const}$ and $\partial R(\Pi)$, hence, in this chapter we use Equation (4.3.4). We remark that, for a given ray with corresponding coordinates (q, p) on T_4, only one path is possible as we are assuming that all lines are reflective. Because of this, the regions $R(\Pi)$ do not overlap. Next, the details of the procedure to compute the coordinates $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$ are explained.

7.1.2 The structure of the inverse ray mapping algorithm

The goal is to determine the target intensity along a given direction $p = \text{const}$. Also in the ray mapping method, we assume a Lambertian source, therefore the intensity is equal to the sum of the lengths of the line segments given by the intersection of the line $p = \text{const}$ and the support of L (see Equation (4.3.4)). To determine these line segments, a recursive procedure is developed. The procedure starts on T_4 with a given direction $p = \text{const}$ and with the parallel rays corresponding to the end points $(q_\ell, p) = (-b, p)$ and $(q_r, p) = (b, p)$. We set the initial intensity $I(p) = 0$ along direction $p = \text{const}$. Considering the intersection between the line $p = \text{const}$ and the boundaries $(\partial T_{4,j})_{j=1,2,3}$ three intervals are found. Each interval corresponds to rays emitted by line j ($j = \{1, 2, 3\}$). The rays corresponding to the end points of these intervals are traced back from T_4 to T_j where j is the line from which they are emitted. Then, another interval of parallel rays along the corresponding direction in T_j has to be considered and the intersection points between the line $p = p_{\mathrm{t},j}$ and $\partial T_{j,k}$ (with $j \neq 4$ and $j \neq k$) are calculated, where $p_{\mathrm{t},j}$ is the new direction of the rays traced back. The procedure continues recursively until the source is found.

Before explaining the details, let us introduce some new notation. The role of the variables we introduce will become clear later on. The coordinates in T_k of the rays traced back from a line $j \neq k$ to line k are indicated with $(q^1_{\mathrm{t},k}, p_{\mathrm{t},k})$ and $(q^2_{\mathrm{t},k}, p_{\mathrm{t},k})$. The minimum and the maximum position coordinates are $q^{\min}_{\mathrm{t},k} = \min\{q^1_{\mathrm{t},k}, q^2_{\mathrm{t},k}\}$ and $q^{\max}_{\mathrm{t},k} = \max\{q^1_{\mathrm{t},k}, q^2_{\mathrm{t},k}\}$, respectively. The coordinates of the intersection points of $p = p_{\mathrm{t},k}$ with boundaries $\partial T_{k,j}$ need to be determined for every $j = \{1, 2, 3\}$ and $k = \{2, 3, 4\}$ with $k \neq j$. They are indicated with $(u^{\min}_{k,j}, p_{\mathrm{t},k})$ and $(u^{\max}_{k,j}, p_{\mathrm{t},k})$ where $u^{\min}_{k,j} < u^{\max}_{k,j}$. Since not all the rays whose corresponding coordinates are located inside the segment $[q^{\min}_{\mathrm{t},k}, q^{\max}_{\mathrm{t},k}]$ with direction $p = p_{\mathrm{t},k}$ follow the same path, the intersection segment $[v^{\min}_{k,j}, v^{\max}_{k,j}] = [q^{\min}_{\mathrm{t},k}, q^{\max}_{\mathrm{t},k}] \cap [u^{\min}_{k,j}, u^{\max}_{k,j}]$ needs to be calculated. $(v^{\min}_{k,j}, p_{\mathrm{t},k})$ and $(v^{\max}_{k,j}, p_{\mathrm{t},k})$ are the coordinates of the rays that need to be traced back from line k to line j.

The method can be outlined as follows.

1. Calculate the intersection points $(u^{\min}_{4,j}, p)$ and $(u^{\max}_{4,j}, p)$ between line $p = \text{const}$. and $\partial T_{4,j}$ for every $j = \{1, 2, 3\}$, where $u^{\min}_{4,j} < u^{\max}_{4,j}$. This can be done analytically because the exact expression of the boundaries $\partial T_{4,j}$ is found as explained in Section 7.1.

2. Calculate the intersection segment

$$[v_{4,j}^{\min}, v_{4,j}^{\max}] = [u_{4,j}^{\min}, u_{4,j}^{\max}] \cap [q^{\min}, q^{\max}].$$

3. If $j = 1$, the coordinates $(v_{4,1}^{\min}, p)$ and $(v_{4,1}^{\max}, p)$ equal the coordinates $(q^{\min}(\Pi, p), p)$ and $(q^{\max}(\Pi, p), p)$ of the rays located on the boundary $\partial R(\Pi)$ with $\Pi = (1, 4)$. All the parallel rays with direction coordinate p and q-position coordinate $u_{4,1}^{\min} \leq q \leq u_{4,1}^{\max}$ are emitted by the source and they directly hit the target. Update the intensity using Eq. (4.3.4)

$$I(p) = I(p) + q^{\max}(\Pi, p) - q^{\min}(\Pi, p).$$

4. If $j \neq 1$, continue with the following steps

5. Trace back $(v_{4,j}^{\min}, p)$ and $(v_{4,j}^{\max}, p)$ from line 4 to line j to find their corresponding coordinates on T_j

$$(q_{\mathrm{t},j}^1, p_{\mathrm{t},j}) = \mathrm{R}_j^{-1} \circ \mathrm{P}_{j,4}^{-1}(v_{4,j}^{\min}, p),$$
$$(q_{\mathrm{t},j}^2, p_{\mathrm{t},j}) = \mathrm{R}_j^{-1} \circ \mathrm{P}_{j,4}^{-1}(u_{4,j}^{\max}, p).$$

6. Update the path $\Pi = (j, 4)$

7. Determine $q_{\mathrm{t},j}^{\min} = \min\{q_{\mathrm{t},j}^1, q_{\mathrm{t},j}^2\}$ and $q_{\mathrm{t},j}^{\max} = \max\{q_{\mathrm{t},j}^1, q_{\mathrm{t},j}^2\}$

8. Calculate the intersection points $(u_{j,k}^{\min}, p)$ and $(u_{j,k}^{\max}, p)$ between the line $p = p_{\mathrm{t},j}$ and $\partial T_{j,k}$ for every $k = \{1, 2, 3\}$ with $k \neq j$.

9. Since not all rays whose corresponding coordinates are located inside the segment $[q_{\mathrm{t},j}^{\min}, q_{\mathrm{t},j}^{\max}]$ follow the same path, compute the intersection segment

$$[v_{j,k}^{\min}, v_{j,k}^{\max}] = [u_{j,k}^{\min}, u_{j,k}^{\max}] \cap [q_{\mathrm{t},j}^{\min}, q_{\mathrm{t},j}^{\max}].$$

10. If $k \neq 1$

    a) Trace back $(v_{j,k}^{\min}, p_{\mathrm{t},j})$ and $(v_{j,k}^{\max}, p_{\mathrm{t},j})$ from j to k

$$(q_{\mathrm{t},k}^1, p_{\mathrm{t},k}) = \mathrm{R}_k^{-1} \circ \mathrm{P}_{k,j}^{-1}(v_{j,k}^{\min}, p_{\mathrm{t},j}),$$
$$(q_{\mathrm{t},k}^2, p_{\mathrm{t},k}) = \mathrm{R}_k^{-1} \circ \mathrm{P}_{k,j}^{-1}(v_{j,k}^{\max}, p_{\mathrm{t},j}).$$

    b) Update the path $\Pi = (k, \Pi)$

    c) Put $j = k$ and repeat the procedure from point 7.

11. Else if $k = 1$, the rays reached the source and a possible path $\Pi = (1, \cdots, 4)$ is found.

    a) Trace back to source

$$(q_{\mathrm{s},1}^1, p_{\mathrm{s},1}) = \mathrm{P}_{1,j}^{-1}(v_{j,1}^{\min}, p_{\mathrm{t},j}),$$
$$(q_{\mathrm{s},1}^2, p_{\mathrm{s},1}) = \mathrm{P}_{1,j}^{-1}(v_{j,1}^{\max}, p_{\mathrm{t},j}).$$

b) Apply the forward map

$$(q^1(\Pi, p), p) = \mathrm{M}_{1,4}(\Pi)(q^1_{\mathrm{s},1}, p_{\mathrm{s},1}),$$
$$(q^2(\Pi, p), p) = \mathrm{M}_{1,4}(\Pi)(q^2_{\mathrm{s},1}, p_{\mathrm{s},1}).$$

c) Update intensity

$$I(p) = I(p) + q^{\max}(\Pi, p) - q^{\min}(\Pi, p)$$

where $q^{\min} = \min\{q^1(\Pi, p), q^2(\Pi, p)\}$ and $q^{\max} = \max\{q^1(\Pi, p), q^2(\Pi, p)\}$.

To clarify the technique, we give an example that describes how the target intensity along direction $p = -0.2$ is calculated. From Fig. 7.11 to Fig. 7.18 the steps used in this example are shown. A detailed description of those figures is given in the following.

The procedure starts with the rays with direction $p = 0.2$ on T_4, where $q_\ell = -b$ and $q_\mathrm{r} = b$ are the left and the right end points of the target T, respectively. The intersection points $(u^{\min}_{4,j}, p)$ and $(u^{\max}_{4,j}, p)$ of the line $p = -0.2$ with boundaries $\partial T_{4,j}$ are computed for every $j \neq 4$.

We start from $j = 1$. Therefore the coordinates $(u^{\min}_{4,1}, p)$ and $(u^{\max}_{4,1}, p)$ of the intersection points between line $p = -0.2$ and the boundary $\partial T_{4,1}$ are computed and these points are depicted in Fig. 7.11. The source is now reached because $j = 1$ and, one possible path is found. The points $(u^{\min}_{4,1}, p)$ and $(u^{\max}_{4,1}, p)$ are located on the boundaries of the region formed by the rays that leave the source and directly hit the target, that is the rays located on $\partial R(\Pi_1)$ with $\Pi_1 = (1, 4)$. Therefore, the contribution to the intensity formed by the rays that follow the path $\Pi_1 = (1, 4)$ is given by $u^{\max}_{4,1} - u^{\min}_{4,1}$.

We continue with $j = 2$. The boundary $\partial T_{4,2}$ is considered in order to find other paths. The intersection points $(u^{\min}_{4,2}, p)$ and $(u^{\max}_{4,2}, p)$ of line $p = -0.2$ with the boundary $\partial T_{4,2}$ are calculated. They are depicted in Fig. 7.12 with the magenta dots. Also the intersection segment

$$[v^{\min}_{4,2}, v^{\max}_{4,2}] = [u^{\min}_{4,2}, u^{\max}_{4,2}] \cap [q^{\min}, q^{\max}] \tag{7.1.13}$$

is calculated. In T_4 $v^{\min}_{4,2} = u^{\min}_{4,2}$ and $v^{\max}_{4,2} = u^{\max}_{4,2}$ because $q^{\min} = -b$ and $q^{\max} = b$ always coincide with the end points of T_4. Their corresponding position coordinates $q^1_{\mathrm{s},2}$ and $q^2_{\mathrm{s},2}$ on S_2 are obtained from:

$$\begin{aligned} \mathrm{P}^{-1}_{2,4}(v^{\min}_{4,2}, p) &= (q^1_{\mathrm{s},2}, p^1_{\mathrm{s},2}), \\ \mathrm{P}^{-1}_{2,4}(v^{\max}_{4,2}, p) &= (q^2_{\mathrm{s},2}, p^2_{\mathrm{s},2}). \end{aligned} \tag{7.1.14}$$

The directions $p^1_{\mathrm{s},2}$ and $p^2_{\mathrm{s},2}$ on S_2 are given considering the direction $p_{\mathrm{t},2} = p$ with respect to the normal $\boldsymbol{\nu}_2$ of line 2. Note that $p^1_{\mathrm{s},2} = p^2_{\mathrm{s},2}$ because all the lines are straight lines, their normals do not depend on the position at which it is computed. Thus, in the following we will omit the subscripts for the direction coordinates. Then, the corresponding direction $p^1_{\mathrm{t},2} = p^2_{\mathrm{t},2}$ on T_2 is calculated from:

$$\begin{aligned} \mathrm{R}^{-1}_2(q^1_{\mathrm{s},2}, p_{\mathrm{s},2}) &= (q^1_{\mathrm{t},2}, p_{\mathrm{t},2}), \\ \mathrm{R}^{-1}_2(q^2_{\mathrm{s},2}, p_{\mathrm{s},2}) &= (q^2_{\mathrm{t},2}, p_{\mathrm{t},2}). \end{aligned} \tag{7.1.15}$$

Note that $q^1_{\mathrm{s},2} = q^1_{\mathrm{t},2}$ and $q^2_{\mathrm{s},2} = q^2_{\mathrm{t},2}$ since the reflection map does not change the position coordinates. Eqs. (7.1.14) and (7.1.15) lead to:

$$\begin{aligned} \mathrm{R}_2^{-1} \circ \mathrm{P}_{2,4}^{-1}(v^{\min}_{4,2}, p) &= (q^1_{\mathrm{t},2}, p_{\mathrm{t},2}), \\ \mathrm{R}_2^{-1} \circ \mathrm{P}_{2,4}^{-1}(v^{\max}_{24,}, p) &= (q^2_{\mathrm{t},2}, p_{\mathrm{t},2}). \end{aligned} \tag{7.1.16}$$

The map $\mathrm{R}_2^{-1} \circ \mathrm{P}_{2,4}^{-1}$ is depicted in red in Fig. 7.10. The minimum $q^{\min}_{\mathrm{t},2} = \min\{q^1_{\mathrm{t},2}, q^2_{\mathrm{t},2}\}$ and maximum $q^{\max}_{\mathrm{t},2} = \max\{q^1_{\mathrm{t},2}, q^2_{\mathrm{t},2}\}$ are calculated. The points with coordinates $(q^{\min}_{\mathrm{t},2}, p_{\mathrm{t},2})$ and $(q^{\max}_{\mathrm{t},2}, p_{\mathrm{t},2})$ are depicted in blue Fig. 7.13, where $p_{\mathrm{t},2} = 0.82$. To understand whether the corresponding rays are illuminated or not by the source, the procedure used for T_4 is now applied to T_2 along direction $p_{\mathrm{t},2} = 0.82$.

Next, the intersection points $(u^{\min}_{2,j}, p_{\mathrm{t},2})$ and $(u^{\max}_{2,j}, p_{\mathrm{t},2})$ of line $p_{\mathrm{t},2} = 0.82$ with boundaries $\partial T_{2,j}$ are computed for every $j \in \{1, 3\}$. We start from the boundary $\partial T_{2,1}$ obtaining the points $(u^{\min}_{2,1}, p_{\mathrm{t},2})$ and $(u^{\max}_{2,1}, p_{\mathrm{t},2})$ shown in Fig. 7.13. Now, the position coordinates $v^{\min}_{2,1} = \max\{q^{\min}_{\mathrm{t},2}, u^{\min}_{2,1}\}$ and $v^{\max}_{2,1} = \min\{q^{\max}_{\mathrm{t},2}, u^{\max}_{2,1}\}$ need to be determined. All the rays located inside the segment $[v^{\min}_{2,1}, v^{\max}_{2,1}]$ in T_2 and with direction $p_{\mathrm{t},2}$ follow path $\Pi_2 = (1, 2, 4)$. In particular, the rays corresponding to the coordinates $(v^{\min}_{2,1}, p_{\mathrm{t},2})$ and $(v^{\max}_{2,1}, p_{\mathrm{t},2})$ are located on the boundaries of the region $R(\Pi_2)$ on T_4 formed by all the rays that follow path Π_2. Their corresponding coordinates $(q^1(\Pi_2, p), p)$ and $(q^2(\Pi_2, p), p)$ on T_4 are obtained from:[1]

$$\begin{aligned} \mathrm{P}_{2,4} \circ \mathrm{R}_2(v^{\min}_{2,1}, p_{\mathrm{t},2}) &= (q^1, p), \\ \mathrm{P}_{2,4} \circ \mathrm{R}_2(v^{\max}_{2,1}, p_{\mathrm{t},2}) &= (q^2, p). \end{aligned} \tag{7.1.17}$$

The rays corresponding to the coordinates (q^1, p) and (q^2, p) are located on the boundary $\partial R(\Pi_2)$ along direction $p = -0.2$. Indicating with $q^{\min} = \min\{q^1, q^2\}$ and $q^{\max} = \max\{q^1, q^2\}$, the distance $q^{\max} - q^{\min}$ gives the contribution to the intensity $I(p)$ of the rays located in $R(\Pi_2)$ where $p = -0.2$.

T_2 can also be illuminated by line 3, therefore the intersection points $(u^{\min}_{2,3}, p_{\mathrm{t},2})$ and $(u^{\max}_{2,3}, p_{\mathrm{t},2})$ of line $p_{\mathrm{t},2} = 0.82$ and $\partial T_{2,3}$ are calculated, these points are depicted in Fig. 7.14. The coordinates $(v^{\min}_{2,3}, p_{\mathrm{t},2})$ and $(v^{\max}_{2,3}, p_{\mathrm{t},2})$ are shown in the same figure. As the source is not reached yet ($j = 3$), the rays corresponding to $(v^{\min}_{2,3}, p_{\mathrm{t},2})$ and $(v^{\max}_{2,3}, p_{\mathrm{t},2})$ are followed back using the inverses of the propagation and the reflection maps. The coordinates on T_3 are shown with blue circles in Figure 7.15 and they are obtained from:

$$\begin{aligned} \mathrm{R}_3^{-1} \circ \mathrm{P}_{3,2}^{-1}(v^{\min}_{2,3}, p_{\mathrm{t},2}) &= (q^1_{\mathrm{t},3}, p_{\mathrm{t},3}), \\ \mathrm{R}_3^{-1} \circ \mathrm{P}_{2,3}^{-1}(v^{\max}_{2,3}, p_{\mathrm{t},2}) &= (q^2_{\mathrm{t},3}, p_{\mathrm{t},3}). \end{aligned} \tag{7.1.18}$$

The minimum and the maximum position coordinates are $q^{\min}_{\mathrm{t},3} = \min\{q^1_{\mathrm{t},3}, q^2_{\mathrm{t},3}\}$ and $q^{\max}_{\mathrm{t},3} = \max\{q^1_{\mathrm{t},3}, q^2_{\mathrm{t},3}\}$, respectively. We found that $v^{\max}_{3,2} \neq u^{\max}_{3,2}$ because $[q^{\min}_{\mathrm{t},3}, q^{\max}_{\mathrm{t},3}] \subset [u^{\min}_{3,2}, u^{\max}_{3,2}]$, this means that the rays with corresponding position coordinates inside the interval $[q^{\max}_{\mathrm{t},3}, u^{\max}_{3,2}]$ will follow a different path. The procedure continues recursively. It stops either when the ray encounters the source, i.e., when $j = 1$, or when no intersection points between the direction $p = p_{\mathrm{t},k}$ and the boundaries $\partial T_{k,j}$ are found for any $j = 1, 2, 3$ with $j \neq k$.

[1] With a slight abuse of notation we indicate $q^1(\Pi, p)$ with q^1 and $q^2(\Pi, p)$ with q^2.

If the source is reached, then a valid path $\Pi = (1, 3, 2, 4)$ is found. Using the inverse of the propagation map, we compute

$$\begin{aligned} \mathrm{P}_{1,3}^{-1}(q_{\mathrm{t},3}^{\min}, p_{\mathrm{t},3}) &= (q_{\mathrm{s},1}^{1}, p_{\mathrm{s},1}), \\ \mathrm{P}_{1,3}^{-1}(q_{\mathrm{t},3}^{\max}, p_{\mathrm{t},3}) &= (q_{\mathrm{s},1}^{2}, p_{\mathrm{s},1}). \end{aligned} \tag{7.1.19}$$

The direct map $\mathrm{M}_{1,4}(\Pi)$:$S_1 \mapsto R(\Pi)$ restricted to path $\Pi = (1, 3, 2, 4)$, i.e.

$$\mathrm{M}_{1,4} = \mathrm{P}_{2,4} \circ \mathrm{R}_2 \circ \mathrm{P}_{3,2} \circ \mathrm{R}_3 \circ \mathrm{P}_{1,3} \tag{7.1.20}$$

is applied to the coordinates $(q_{\mathrm{s},1}^{1}, p_{\mathrm{s},1})$ and $(q_{\mathrm{s},1}^{1}, p_{\mathrm{s},1})$:

$$\begin{aligned} \mathrm{M}_{1,4}(q_{\mathrm{s},1}^{1}, p_{\mathrm{s},1}) &= (q^1(\Pi, p), p), \\ \mathrm{M}_{1,4}(q_{\mathrm{s},1}^{2}, p_{\mathrm{s},1}) &= (q^2(\Pi, p), p). \end{aligned} \tag{7.1.21}$$

The coordinates $(q^1(\Pi, p), p)$ and $(q^2(\Pi, p), p)$ located on $\partial R(\Pi)$ in T_4 are found. Indicating with $q^{\min} = \min\{q^1, q^2\}$ and $q^{\max} = \max\{q^1, q^2\}$, the contribution to the intensity due to the rays that follow path Π is given by:

$$I(p) = I(p) + q^{\max}(\Pi, p) - q^{\min}(\Pi, p). \tag{7.1.22}$$

If no intersection points are found, then the rays traced are not emitted by the source, therefore no contribution to the intensity needs to be added. This is, for instance, the case of rays with coordinates $(v_{2,3}^{\min}, 0.82)$ and $(v_{2,3}^{\max}, 0.82)$ on T_2 in Fig. 7.14. Below we explain this case in detail.

In Fig. 7.15, the coordinates $(q_{\mathrm{t},3}^{\min}, p_{\mathrm{t},3})$ and $(q_{\mathrm{t},3}^{\max}, p_{\mathrm{t},3})$ in T_3 with $p_{\mathrm{t},3} = -0.29$ are shown. They are obtained from:

$$\begin{aligned} \mathrm{R}_3^{-1} \circ \mathrm{P}_{3,2}^{-1}(v_{2,3}^{\min}, 0.82) &= (q_{\mathrm{t},3}^{1}, p_{\mathrm{t},3}), \\ \mathrm{R}_3^{-1} \circ \mathrm{P}_{3,2}^{-1}(v_{2,3}^{\max}, 0.82) &= (q_{\mathrm{t},3}^{2}, p_{\mathrm{t},3}). \end{aligned} \tag{7.1.23}$$

From Fig. 7.15 we note that there are no intersection points of the line $p_{\mathrm{t},3} = -0.29$ with $\partial T_{3,1}$. So, only the coordinates of the intersections $(u_{3,2}^{\min}, -0.29)$ and $(u_{3,2}^{\max}, -0.29)$ between line $p_{\mathrm{t},3} = -0.29$ and $\partial T_{3,2}$ are calculated. Next, the intersection interval

$$[v_{3,2}^{\min}, v_{3,2}^{\max}] = [u_{3,2}^{\min}, u_{3,2}^{\max}] \cap [q_{\mathrm{t},3}^{\min}, q_{\mathrm{t},3}^{\max}], \tag{7.1.24}$$

formed by parallel rays with direction $p_{\mathrm{t},3} = -0.29$, is considered. Using:

$$\begin{aligned} \mathrm{R}_2^{-1} \circ \mathrm{P}_{2,3}^{-1}(v_{3,2}^{\min}, -0.29) &= (q_{\mathrm{t},2}^{\min}, p_{\mathrm{t},2}), \\ \mathrm{R}_2^{-1} \circ \mathrm{P}_{2,3}^{-1}(v_{3,2}^{\max}, -0.29) &= (q_{\mathrm{t},2}^{\max}, p_{\mathrm{t},2}), \end{aligned} \tag{7.1.25}$$

the corresponding coordinates $(q_{\mathrm{t},2}^{\max}, p_{\mathrm{t},2})$ and $(q_{\mathrm{t},2}^{\min}, p_{\mathrm{t},2})$ on T_2 are found (see Figure 7.16) with $p_{\mathrm{t},2} = -0.41$. Now the procedure is repeated again for T_2 along the direction $p_{\mathrm{t},2}$. No intersection points between the line $p_{\mathrm{t},2} = -0.41$ and $\partial T_{2,1}$ occur. Only, the intersection points $(u_{2,3}^{\min}, p_{\mathrm{t},2})$ and $(u_{2,3}^{\max}, p_{\mathrm{t},2})$ of line $p_{\mathrm{t},2} = -0.41$ and $\partial T_{2,3}$ are found. The intersection segment

$$[v_{2,3}^{\min}, v_{2,3}^{\max}] = [u_{2,3}^{\min}, u_{2,3}^{\max}] \cap [q_{\mathrm{t},2}^{\min}, q_{\mathrm{t},2}^{\max}] \tag{7.1.26}$$

is calculated. The coordinates on T_3 corresponding to the end points of the intersection interval are found using:

$$\begin{aligned} \mathrm{R}_3^{-1} \circ \mathrm{P}_{3,2}^{-1}(v_{2,3}^{\min}, p_{\mathrm{t},2}) &= (q_{\mathrm{t},3}^{\min}, p_{\mathrm{t},3}), \\ \mathrm{R}_3^{-1} \circ \mathrm{P}_{3,2}^{-1}(v_{2,3}^{\max}, p_{\mathrm{t},2}) &= (q_{\mathrm{t},3}^{\max}, p_{\mathrm{t},3}), \end{aligned} \tag{7.1.27}$$

where $p_{\mathrm{t},3} = 0.91$ (see Fig. 7.17).

Considering the PS T_3 and the direction $p_{\mathrm{t},3} = 0.91$, we note that there are no intersection points of line $p_{\mathrm{t},3} = 0.91$ with both $\partial T_{3,1}$ and $\partial T_{3,2}$. Indeed, the whole segment $[q_{\mathrm{t},3}^{\min}, q_{\mathrm{t},3}^{\max}]$ is outside both $T_{3,2}$ and $T_{3,1}$. Because of this, all the rays with q-coordinates inside the interval $[q_{\mathrm{t},3}^{\min}, q_{\mathrm{t},3}^{\max}]$ and with direction $p = p_{\mathrm{t},3}$ are not illuminated by the source and no new real path is found.

Finally, the recursive procedure is applied to $T_{4,3}$. The first step is depicted in Fig. 7.18. We decided not to show all the steps for $T_{4,3}$ as they are similar to those used for $T_{4,2}$ and explained above.

Finally, to compute the intensity along another direction $p^h \in [-1, 1]$ on T_4, the procedure explained for $p = -0.2$ is repeated for $p = p^h$. In this way we find all the possible paths Π and the regions $R(\Pi)$ with positive luminance on T_4. Furthermore, considering every time the coordinates located on the boundaries of the regions $T_{j,k}$ for every k, also the boundaries $\partial R(\Pi)$ are determined for a given path Π as well as the coordinates $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$ for every $p \in [-1, 1]$. In Algorithm 3, the main steps to calculate the intensity $I(p)$ along a direction $p = p^h$ in T_4 are given, where for the first step we take $k = 4$.

In the next section we provide the numerical results for the two-faceted cup.

7.2 Numerical results for the two-faceted cup

To demonstrate the accuracy of the method, a comparison with MC and QMC ray tracing is provided. The MC and QMC intensities are computed as explained in Chapter 3. We consider here the same partitioning $P : -1 = p^0 < \cdots < p^{\mathrm{Nb}} = 1$ of the interval $[-1, 1]$ used for all the simulation presented in the previous chapters. The profile of the QMC intensity is obtained tracing 10^7 rays and taking $\mathrm{Nb} = 100$.

The PS intensity is obtained from Eq. (4.3.4) where the rays on the boundaries are obtained applying the inverse ray mapping. We observe that the method is suitable for detecting all the possible paths Π that a ray can follow during the propagation through the system. According to the results obtained with PS ray tracing, 5 different paths are found for the two-faceted cup. Given a path Π, the coordinates $(q^{\min}(\Pi, p^h), p^h)$ and $(q^{\max}(\Pi, p^h), p^h)$ of the corresponding rays located on $\partial R(\Pi)$ are determined for every $p = (p^h)_{h=0,\cdots,\mathrm{Nb}}$ where the values p^h are given from the partitioning P used for QMC ray tracing. These rays are depicted in Figure 7.19, where all the rays that follow the same path are shown with the same color. For the two-faced cup, given a direction p^h and a path Π, only two rays are located on the boundary $\partial R(\Pi)$ of the corresponding region along that direction. As a consequence, at most $2\,\mathrm{Np}\,\mathrm{Nb}$ rays need to be traced from the target to the source, where $\mathrm{Np} = 5$ is the number of paths. The averaged normalized PS intensity is given by Equation (5.3.3) where the integrals are calculated using the trapezoidal rule. The approximated intensities $\hat{I}_{\mathrm{A}}(\mathrm{A} = \mathrm{PS}, \mathrm{QMC})$ are compared to the reference intensity $\hat{I}_{\mathrm{ref}}$ which

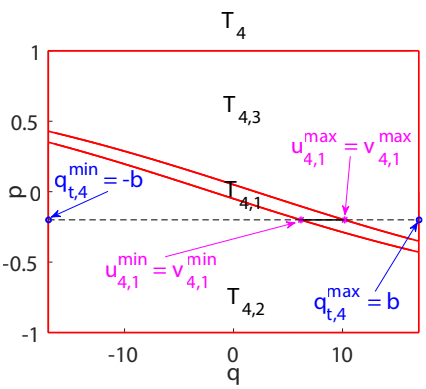


Figure 7.11: Target phase space of line 4. $q^{\min}_{\mathrm{t},4}$ and $q^{\max}_{\mathrm{t},4}$ are the x-coordinates of the end points of line 4. The intersection points between the line $p = -0.2$ and $\partial T_{4,1}$ are $(u^{\min}_{4,1}, p)$ and $(u^{\max}_{4,1}, p)$. $v^{\min}_{4,1} = \max\{q^{\min}_{\mathrm{t},4}, u^{\min}_{4,1}\}$ and $v^{\max}_{4,1} = \min\{q^{\max}_{\mathrm{t},4}, u^{\max}_{4,1}\}$.

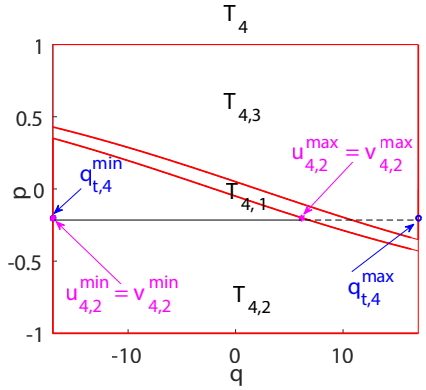


Figure 7.12: Target phase space of line 4. The intersection points between the line $p = -0.2$ and $\partial T_{4,2}$ are $(u^{\min}_{4,2}, p)$ and $(u^{\max}_{4,2}, p)$. $v^{\min}_{4,2} = \max\{q^{\min}_{\mathrm{t},4}, u^{\min}_{4,2}\}$ and $v^{\max}_{4,2} = \min\{q^{\max}_{\mathrm{t},4}, u^{\max}_{4,2}\}$.

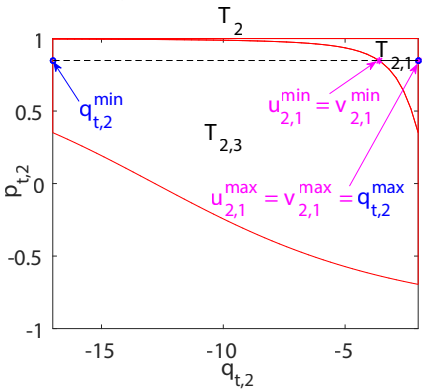


Figure 7.13: Target phase space of line 2. The coordinates of the intersection points between the line $p_{\mathrm{t},2} = 0.82$ and $\partial T_{2,1}$ are $(u^{\min}_{2,1}, p_{\mathrm{t},2})$ and $(u^{\max}_{2,1}, p_{\mathrm{t},2})$. $v^{\min}_{2,1} = \max\{q^{\min}_{\mathrm{t},2}, u^{\min}_{2,1}\}$ and $v^{\max}_{2,1} = \min\{q^{\max}_{\mathrm{t},2}, u^{\max}_{2,1}\}$.

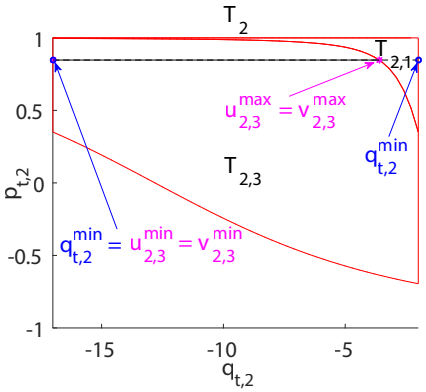


Figure 7.14: Target phase space of line 2. The coordinates of the intersection points between the line $p_{\mathrm{t},2} = 0.82$ and $\partial T_{2,3}$ are $(u^{\min}_{2,3}, 0.82)$ and $(u^{\max}_{2,3}, 0.82)$. $v^{\min}_{2,3} = \max\{u^{\min}_{2,3}, q^{\min}_{\mathrm{t},2}\}$ and $v^{\max}_{2,3} = \min\{u^{\max}_{2,3}, q^{\max}_{\mathrm{t},3}\}$.

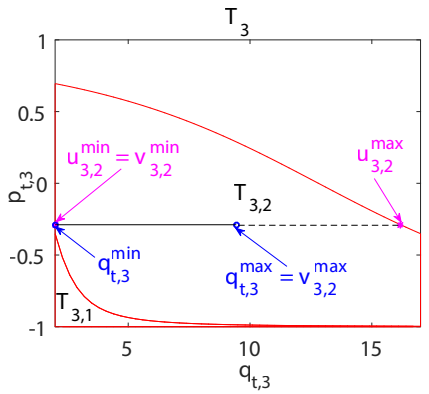


Figure 7.15: Target phase space of line 3. The position coordinates of the intersection points between the line $p_{\mathrm{t},3} = -0.29$ and $\partial T_{3,2}$ are $u_{3,2}^{\min}$ and $u_{3,2}^{\max}$. $v_{3,2}^{\min} = \max\{u_{3,2}^{\min}, q_{\mathrm{t},3}^{\min}\}$ and $v_{3,2}^{\max} = \min\{u_{3,2}^{\max}, q_{\mathrm{t},3}^{\max}\}$.

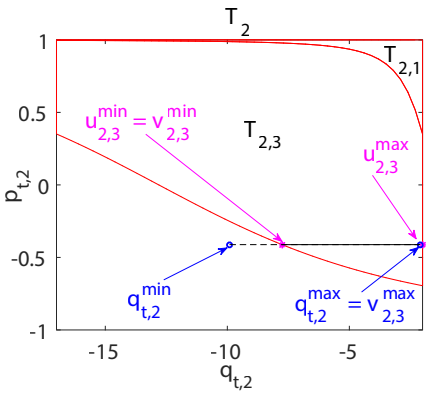


Figure 7.16: Target phase phase of line 2. The intersection points between the line $p_{\mathrm{t}} = p_{\mathrm{t},2}$ and $\partial T_{2,3}$ are $(u_{2,3}^{\min}, p_{\mathrm{t},2})$ and $(u_{2,3}^{\max}, p_{\mathrm{t},2})$. $v_{2,3}^{\min} = \max\{u_{2,3}^{\min}, q_{\mathrm{t},2}^{\min}\}$ and $v_{2,3}^{\max} = \min\{u_{2,3}^{\max}, q_{\mathrm{t},2}^{\max}\}$.

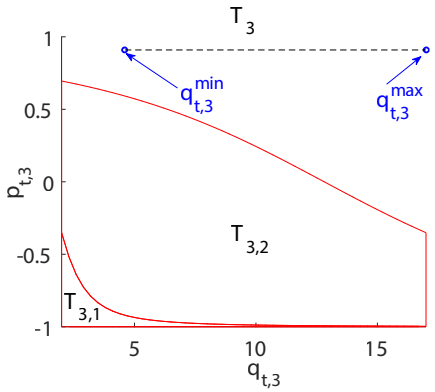


Figure 7.17: Target phase space of line 3. There are no intersection points of line $p_{3,2} = 0.91$ with the boundaries $\partial T_{3,2}$ and $\partial T_{3,1}$. The rays with coordinates inside the dotted segment hit again line 4 after some reflections and, therefore, are not emitted by the source.

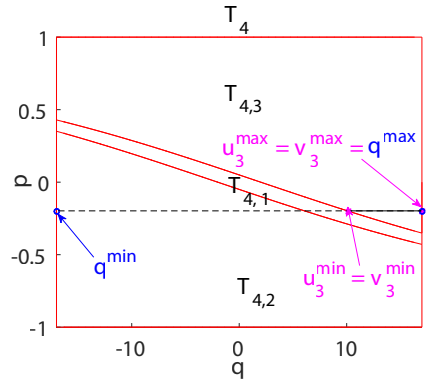


Figure 7.18: Target phase space of line 4. $q_{\mathrm{t},4}^{\min} = -b$ and $q_{\mathrm{t},4}^{\max} = b$. The intersection points between the line $p = -0.2$ and $\partial T_{4,3}$ are $(u_{4,3}^{\min}, p)$ and $(u_{4,3}^{\max}, p)$. $v_{4,3}^{\min} = \max\{u_{4,3}^{\min}, q_{\mathrm{t},4}^{\min}\}$ and $v_{4,3}^{\max} = \min\{u_{4,3}^{\max}, q_{\mathrm{t},4}^{\max}\}$.

Algorithm 3 Recursive procedure for the intensity calculation

Initialize $k=4$, $q^{\min}_{\mathrm{t},4} = q^{\min} = -b$, $q^{\max}_{\mathrm{t},4} = q^{\max} = b$, $p_{\mathrm{t},4} = p = \text{const}$, $\Pi = (4)$.

procedure INTENSITY COMPUTATION(k, $q^{\min}_{\mathrm{t},k}$, $q^{\max}_{\mathrm{t},k}$, $p_{\mathrm{t},k}$, Π)
  for $j = 1, 2, 3$ **do**
    if $j \neq k$ **then**
      Compute the intersection points $(u^{\min}_{k,j}, p_{\mathrm{t},k})$ and $(u^{\max}_{k,j}, p_{\mathrm{t},k})$
      $\Pi \leftarrow (j, \Pi)$
      Compute $[v^{\min}_{k,j}, v^{\max}_{k,j}] = [u^{\min}_{k,j}, u^{\max}_{k,j}] \cap [q^{\min}_{\mathrm{t},k}, q^{\min}_{\mathrm{t},k}]$
      if $(j \neq 1)$ & $(j \neq 4)$ **then**
        Trace back from T_k to T_j

$$(q^1_{\mathrm{t},j}, p_{\mathrm{t},j}) = \mathrm{R}^{-1}_j \circ \mathrm{P}^{-1}_{j,k}(v^{\min}_{k,j}, p_{\mathrm{t},k})$$
$$(q^2_{\mathrm{t},j}, p_{\mathrm{t},j}) = \mathrm{R}^{-1}_j \circ \mathrm{P}^{-1}_{j,k}(v^{\max}_{k,j}, p_{\mathrm{t},k})$$

        Determine

$$q^{\min}_{\mathrm{t},j} = \min\{q^1_{\mathrm{t},j}, q^2_{\mathrm{t},j}\} \text{ and } q^{\max}_{\mathrm{t},j} = \max\{q^1_{\mathrm{t},j}, q^2_{\mathrm{t},j}\}$$

        return INTENSITY COMPUTATION(j, $q^{\min}_{\mathrm{t},j}$, $q^{\max}_{\mathrm{t},j}$, $p_{\mathrm{t},\,j}$, Π)
      else
        if j=1 **then**
          if $k \neq 4$ **then**
            Trace back from T_k to S_1, next apply the direct map $\mathrm{M}_{1,4}(\Pi)$

$$(q^1_{\mathrm{s},1}, p_{\mathrm{s},1}) = \mathrm{P}^{-1}_{1,k}(v^{\min}_{k,1}, p_{\mathrm{t},k})$$
$$(q^2_{\mathrm{s},1}, p_{\mathrm{s},1}) = \mathrm{P}^{-1}_{1,k}(v^{\max}_{k,1}, p_{\mathrm{t},k})$$
$$(q^1(\Pi, p), p) = \mathrm{M}_{1,4}(\Pi)(q^1_{\mathrm{s},1}, p_{\mathrm{s},1})$$
$$(q^2(\Pi, p), p) = \mathrm{M}_{1,4}(\Pi)(q^2_{\mathrm{s},1}, p_{\mathrm{s},1})$$

            Calculate

$$q^{\min}(\Pi, p) = \min\{q^1, q^2\},$$
$$q^{\max}(\Pi, p) = \max\{q^1, q^2\},$$

            where $q^1 := q^1(\Pi, p)$ and $q^2 := q^2(\Pi, p)$.
            return $I(p) = I(p) + q^{\max}(\Pi, p) - q^{\min}(\Pi, p)$.
          else

$$q^{\min}(\Pi, p) = v^{\min}_{4,k} \text{ and } q^{\max}(\Pi, p) = v^{\max}_{4,k}$$

            return $I(p) = I(p) + q^{\max}(\Pi, p) - q^{\min}(\Pi, p)$.
          end if
        else
          return $I(p)$
        end if
      end if
    end if
  end for
end procedure

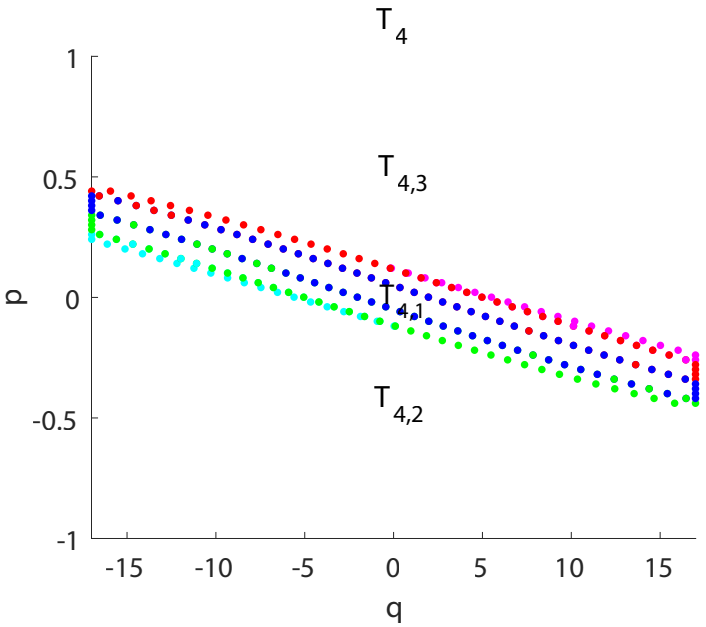


Figure 7.19: **Target phase space of the two-faceted cup divided into** 100 **bins.** Five different paths are found. The rays with coordinates $(q^{\min}, p)$ and $(q^{\max}, p)$ in T_4 that are located at the boundaries $\partial R(\Pi)$ are depicted with dots, the color of the dots depends on the path Π followed by the rays. Using the ray mapping method, only these rays need to be traced from S to T for the intensity computation.

in this case in the exact intensity ($\hat{I}_{\text{ref}} = \hat{I}_{\text{exact}}$). The results in Figure 7.20 show that our method computes the intensity correctly.

To compare the speed of convergence of the two methods, we consider the error between the approximate intensities $\hat{I}_{\text{A}}$ ($\text{A} = \text{MC}, \text{QMC}, \text{PS}$) and the exact intensity $\hat{I}_{\text{exact}} = \hat{I}_{\text{ref}}$. The three errors as a function of the CPU-time are depicted in a logarithmic scales in Figure 7.21. Numerical results show that MC ray tracing converges proportionally to the inverse of the square root of the number of number of rays traced, QMC error is proportional to the inverse of the number of rays, the PS ray mapping method is able to compute the output intensity of the two-faceted cup exactly. Also, it is much faster than MC ray tracing when an error smaller than 10^{-4} is required and it is faster than QMC ray tracing if an error smaller than around 10^{-5} is desired.

7.3 Extension of the method for the multi-faceted cup

The method can be generalized to more complicated optical systems. In particular, it can be used for all systems formed by straight line segments. The goal of this section

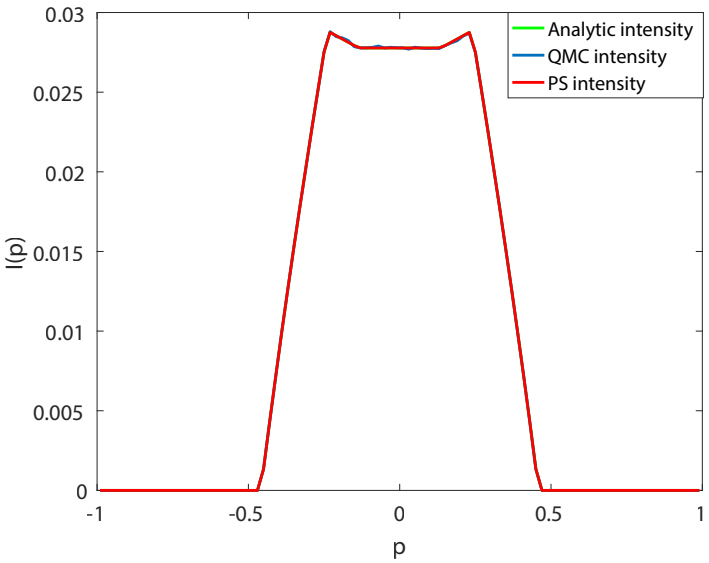


Figure 7.20: **Intensities for the two-faceted cup.** The intensities found with three different approaches are shown.

is to show the generalization of the method to the multi-faced cup which is a system with many left and right segments as reflectors. The design of this system is explained below.

A multi-faceted cup is an optical system formed by a source, a target and $\mathrm{Nl}-2$ reflectors, where Nl is the number of optical line segments that form the system. Defining a Cartesian coordinate system (x, z), the multi-faceted cup is symmetric with respect to the optical axis (z-axis). An example of this system is depicted in Fig. 7.22 where all the lines are labeled with numbers. The source $\mathsf{S} = [-a, a]$ (line 1) and the target $\mathsf{T} = [-b, b]$ (line 22) are two segments both perpendicular to the optical axis, with $a = 2$ and $b = 17$. S is located at the height $z = 0$ while T has a height $z = 40$. Both sides of the system are divided into 10 segments which connect S with T. The ten adjacent segments at the left of the system (lines $2, \cdots, 11$) connect the left extreme of the source with the left extreme of the target. Similarly, ten adjacent segments at the right of the system (lines $12, \cdots, 21$) connect the right extreme of the source with the right extreme of the target. These segments are designed as follows. The intervals $[-b, -a]$ and $[a, b]$ are divided into ten sub-intervals of the same length $(b-a)/10$. The x-coordinates of the end points of the line segments $12, \cdots, 21$ are equal to the x-coordinates of the sub-intervals of $[a, b]$, while the x-coordinates of the end points of the line segments $2, \cdots, 11$ are equal to the x-coordinates of the sub-intervals of $[-a, -b]$. The z-coordinates of every end point of the line segments $2, \cdots, 21$ are given substituting their x-coordinates into the equation of the parabola whose symmetry axis is equal to the z-axis and that passes through the points $(-17, 40)$ and $(17, 40)$. The 20-faceted cup is now well defined and can be seen as an approximation of a

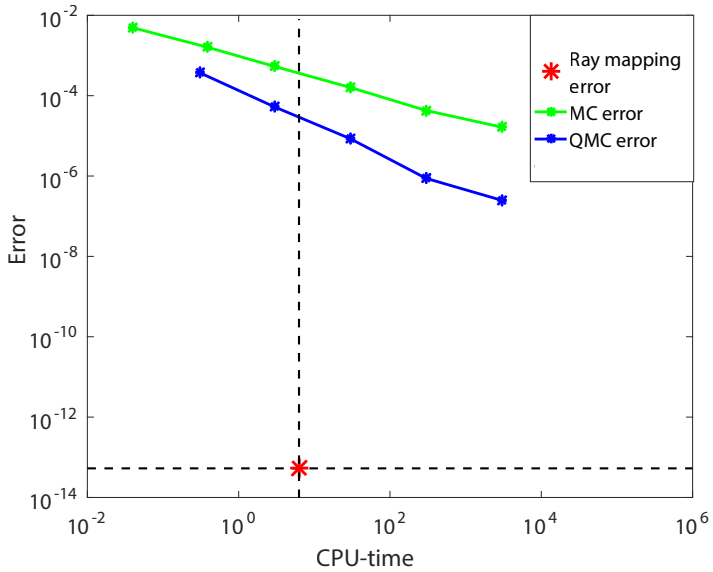


Figure 7.21: **Errors for the two-faceted cup.** The errors are depicted as a function of the CPU time (in seconds).

parabolic reflector.

Similarly to the two-faceted cup, also for the multi-faceted cup we define the phase spaces of all the lines $j \in \{1, \cdots, \mathrm{Nl}\}$ (for the 20-faceted cup $\mathrm{Nl} = 22$ which is also the index of the target). For the system in Figure 7.22, 42 different phase spaces need to be considered. In general, for a system formed by Nl straight line segments, $2\mathrm{Nl} - 2$ phase spaces are considered. For all the systems formed by straight line segments, the boundaries $(\partial S_{j,k})_{j \neq k = 2, \cdots, \mathrm{Nl}}$ and $(\partial T_{j,l})_{j \neq l = 1, \cdots, \mathrm{Nl}-1}$ of the regions that form every PS are determined.

The boundaries $(\partial T_{Nl,l})_{l=1,\cdots,\mathrm{Nl}-1}$ for the 20-faceted cup are depicted in Fig. 7.23 with red lines. All the possible paths that the rays can follow when propagating within the 20-faceted cup are determined using the same algorithm developed for the two-faceted cup and explained in Section 7.1.2. As the number of optical lines increases, the number of possible paths increases as well. Therefore, we have to construct a more complicated tree than the one in Fig. 7.10. Despite this, the algorithm explained in the previous section still works fine and, also for the multi-faceted cup we are able to determine all the possible paths Π and all the regions $R(\Pi)$ with positive luminance at target PS T_{Nl}. For a given direction $p = \text{const}$ the position coordinates $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$ of the intersection points between the boundaries $\partial R(\Pi)$ and the line $p = \text{const}$ are calculated for every possible path Π. Finally, the target intensity $\hat{I}_{\mathrm{PS}}(p)$ along the direction p is obtained. Numerical results for a 20-faceted cup are given in the next section.

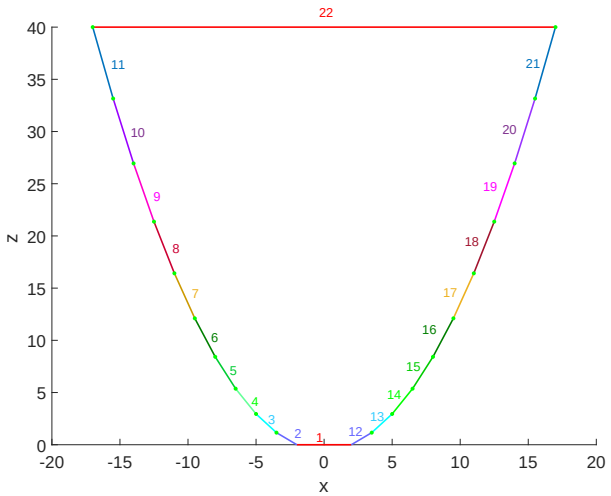


Figure 7.22: **The 20-faceted cup.** The system is formed by 22 different line segments: the source S, the target T, ten left reflectors and ten right reflectors. $\mathsf{S} = [-2, 2]$ is located at $z = 0$. $\mathsf{T} = [-17, 17]$ is parallel to the source and it is located at a height $z = 40$. All the lines are located in air.

7.4 Numerical results for the 20-faceted cup

In this section the results for the 20-faceted cup are presented. We compute the target intensity both with the inverse ray mapping method and MC ray tracing. The same partitioning P of the interval $[-1, 1]$ used for the two-faceted cup is considered. A comparison between the reference intensity $\hat{I}_{\text{ref}}$ and the ray mapping intensity $\hat{I}_{\text{PS}}$ is shown in Figure 7.24, where $\hat{I}_{\text{ref}}$ is obtained using QMC ray tracing with 10^8 rays.

Note that the intensity profile in Fig. 7.24 is more concentrated around the direction $p = 0$ than the intensity of the two-faceted cup (see Fig. 7.20). In particular, increasing the number of left and right reflectors the intensity profile becomes more and more peaked around the center approaching the profile of a parabolic reflector, (see Chapter 6). The error between the approximate intensities $\hat{I}_{\text{A}}$ $(\text{A} = \text{QMC}, \text{PS})$ and the reference intensity $\hat{I}_{\text{ref}}$ is shown in Figure 7.25. The PS intensity is calculated using Equation 5.3.3 where the integral is approximated using the trapezoidal rule. Increasing the number of intervals in the trapezoidal rule, the PS error decreases. We remark that the PS method gives the value of the intensity pointwise, therefore we can compute the PS intensity without numerical integration. Nevertheless, we calculate the averaged intensity because we want to compare it with the QMC intensity $\hat{I}_{\text{QMC}}$.

The error convergence is depicted in Fig. 7.25 with the red line. Since all the boundaries of the regions in PS are calculated exactly, our expectation is that the PS intensity is the exact intensity. From Fig. 7.25 we observe that the minimum ray mapping error has an order of magnitude of 10^{-7}. This is due to the fact that for the 20-faceted cup the intensity cannot be computed exactly. Therefore, we took as reference intensity $\hat{I}_{\text{ref}}$ an intensity computed with QMC ray tracing using 10^8 rays which is not the exact intensity. The error between the normalized exact intensity

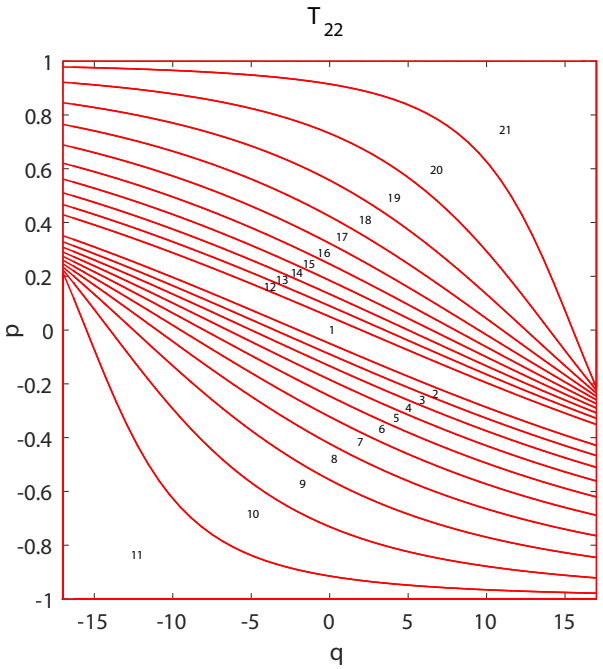


Figure 7.23: **Target PS of the 20-faceted cup.** The red lines are the boundaries $(\partial T_{22,l})_{l=1,\cdots,21}$ which are determined analytically. The numbers inside the regions $T_{22,l}$ indicate the value of the index l.

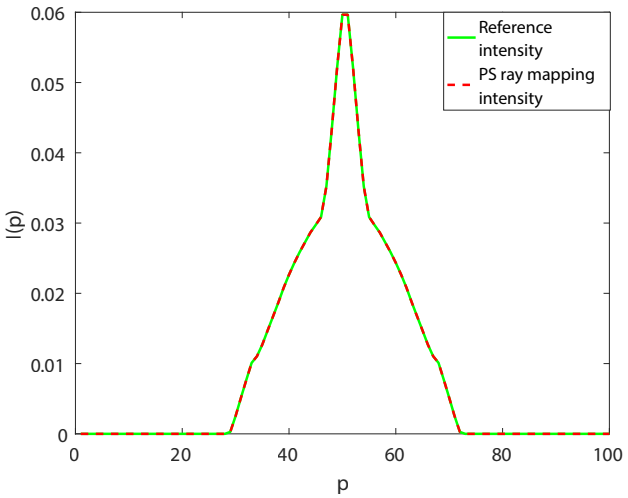


Figure 7.24: **Intensity for the 20-faceted cup**. Comparison between the reference intensity (QMC ray tracing with 10^8 rays) and the ray mapping intensity.

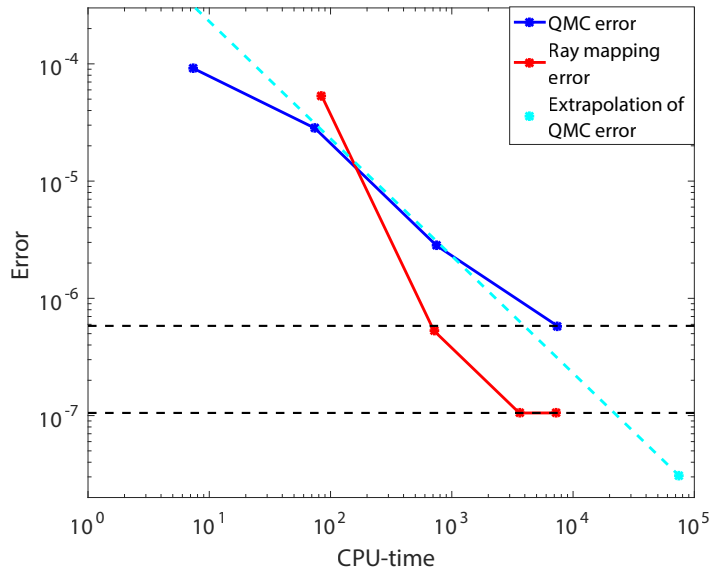


Figure 7.25: **Errors for the 20-faceted cup as a function of the CPU.** The ray mapping method is more accurate than QMC ray tracing and it is faster in case an error smaller than 10^{-5} is desired.

$\hat{I}_{\text{exact}}$ and the normalized approximate intensity $\hat{I}_{\text{A}}$ is given by:

$$\frac{1}{\text{Nb}}\sum_{h=1}^{\text{Nb}}|\hat{I}_{\text{exact}}(p^h)-\hat{I}_{\text{A}}(p^h)| \leq \frac{1}{\text{Nb}}\left(\sum_{h=1}^{\text{Nb}}\left|\hat{I}_{\text{exact}}(p^h)-\hat{I}_{\text{ref}}(p^h)\right|+\sum_{h=1}^{\text{Nb}}\left|\hat{I}_{\text{ref}}(p^h)-\hat{I}_{\text{A}}(p^h)\right|\right). \tag{7.4.1}$$

Assuming that QMC error converges proportional to $\frac{C}{CPU-time}$, we can calculate the constant C and extrapolate the QMC error. From this extrapolation we obtain an approximation of the error between the reference intensity and the exact intensity, this error is depicted in Figure 7.25 with the cyan dot, the dotted cyan line shows the linear extrapolation. From numerical simulation we obtain the difference between the extrapolated value and the exact intensity

$$\sum_{h=1}^{\text{Nb}}|\hat{I}_{\text{exact}}(p^h)-\hat{I}_{\text{ref}}(p^h)|/\text{Nb} \approx 3*10^{-8}.$$

The results show that

$$\sum_{h=1}^{\text{Nb}}|\hat{I}_{\text{exact}}(p^h)-\hat{I}_{\text{ref}}(p^h)|/\text{Nb} \approx \sum_{h=1}^{\text{Nb}}|\hat{I}_{\text{ref}}(p^h)-\hat{I}_{\text{PS}}(p^h)|/\text{Nb}.$$

Therefore, we claim that the error found with the inverse ray mapping method is also due to the QMC error. We can conclude that the inverse ray mapping method performs very well also for more complicated systems. Compared to QMC ray tracing the new method is not only faster but also much more accurate.

7.5 Discussions

In this chapter, we presented an inverse ray mapping method to compute the target intensity of a given optical system. The method employs the PS of *all* the lines that form the system. All these phase spaces are related to each other through two different kind of maps. A concatenation of these two maps gives a map that connects the coordinates of the rays at the source with those at the target. Employing the inverse of the concatenated map, all the possible paths that rays can follow during their propagation are found. Only the rays located on the boundaries of the regions with positive luminance are traced, where every region is formed by rays that follow the same path during their propagation. From those rays the output intensity is calculated.

We presented numerical results for two optical systems: the two-faceted cup and the 20-faceted cup. The boundaries of the regions that form every PS are determined exactly. Numerical results shown that the exact output intensity is obtained. We compared our method with MC and QMC ray tracing showing significant advantages in terms of the accuracy and the computational time. We conclude that the ray mapping method applied to systems formed by straight line segments calculate the *exact* intensity.

In the next chapter we present the method extended to systems formed by curved lines.

Chapter 8

Inverse ray mapping: numerical approach

In Chapter 7 we introduced an inverse method based on ray mapping reconstruction in PS. The goal was to calculate the intensity distribution at the target of optical systems. The idea was to construct a map from the target T to the source S using the PS of all the optical lines which are divided into several regions. The method developed in the previous chapter requires that the boundaries of these regions can be determined exactly in every PS. This is possible for systems formed by straight line segments. The results found show that the procedure allows tracing only the rays located exactly on the boundaries of the regions with positive luminance. From these rays, the *exact* intensity was determined.

In this chapter we extend the method to systems formed by curved lines. In this case, the boundaries of the regions that form the PS cannot be determined exactly. Because of this, we need to apply a numerical procedure. In particular, we develop a method that employs only the PS of the target of the system. The boundaries are detected applying a bisection procedure in target PS in combination with the inverse ray tracing.

In this chapter we test the method for two optical systems: the TIR-collimator and a parabolic reflector. The results are presented in Section 8.2 and 8.3, respectively.

8.1 Bisection method and inverse ray tracing

The purpose of this section is to present the generalized inverse ray mapping method valid for systems formed by curved lines. Given a partition $-1 = p^0 < p^1 < \cdots < p^{\mathrm{Nb}} = 1$ of the interval $[-1, 1]$ with Nb the number of bins in the partitioning, the intensity in target PS is given by Equation (4.3.4) for every $p \in P$. Therefore, the problem reduces to calculating the coordinates $q^{\min}(\Pi, p)$ and $q^{\max}(\Pi, p)$ of the rays on $\partial R(\Pi)$ for every path Π where we used the same notation of the previous chapter for the rays coordinates.

We indicate with $(q^{\mathrm{a}}, p) = (-b, p)$ and $(q^{\mathrm{b}}, p) = (b, p)$ the coordinates of the end points of T along direction p and with $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ the parametrization of rays

corresponding to these two points. The coordinates of each ray on line $j \neq \mathrm{Nl}$ are indicated with (q_j, p_j) where q_j is the position x-coordinate of the intersection point between the ray and line j and p is the direction of the ray with respect to the optical axis. Note that $q_j = q_{\mathrm{t},j}$ while $p_j \neq p_{\mathrm{t},j}$ where $(q_{\mathrm{t},j}, p_{\mathrm{t},j})$ are the coordinates of the rays in target PS T_j introduced in the previous chapter where the directions coordinates where expressed with respect to the normal of j. It is useful here to specify with a subscript to which ray every couple of coordinates belongs. For instance, $(q_j^{\mathrm{a}}, p_j^{\mathrm{a}})$ are the coordinates of ray $\boldsymbol{r}_{\mathrm{a}}$ on line j.

The procedure starts considering an intensity $I(p) = 0$ and the end points $(q^{\mathrm{a}}, p) = (-b, p)$ and $(q^{\mathrm{b}}, p) = (b, p)$. Now, since the boundaries of the the regions of all the phase spaces are unknown, to understand from which line $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are emitted we apply the inverse ray tracing. We denote with Nl the index of the target (line Nl) and with $k \in \{1, \cdots, \mathrm{Nl}-1\}$ and $l \in \{1, \cdots, \mathrm{Nl}-1\}$ the lines from which $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are emitted. $\Pi^{\mathrm{a}} = (k, \mathrm{Nl})$ and $\Pi^{\mathrm{b}} = (l, \mathrm{Nl})$ are the last part of the paths followed by the two rays $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$, respectively. In particular, $(q^{\mathrm{a}}, p) \in \partial R(\Pi^{\mathrm{a}})$ and $(q^{\mathrm{b}}, p) \in \partial R(\Pi^{\mathrm{b}})$. At this stage we know whether the two rays are emitted from the same line or not.

First, assume $k = l$, then $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ hit the same line before reaching the target. In case $k = 1$ a possible path from the source to the target is found and the intensity is updated according to:

$$I(p) = I(p) + q^{\max}(\Pi^{\mathrm{a}}, p) - q^{\min}(\Pi^{\mathrm{a}}, p), \tag{8.1.1}$$

where $q^{\min} = \min\{q^{\mathrm{a}}, q^{\mathrm{b}}\}$ and $q^{\max} = \max\{q^{\mathrm{a}}, q^{\mathrm{b}}\}$. In case $k \neq 1$ the two rays $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are traced back further using inverse ray tracing.

Next, if $k \neq l$ the rays $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are emitted by two-different lines, hence $\Pi^{\mathrm{a}} \neq \Pi^{\mathrm{b}}$ and belong to different regions $R(\Pi^{\mathrm{a}})$ and $R(\Pi^{\mathrm{b}})$ in T. To determine the other coordinates of the rays on the boundary $\partial R(\Pi^{\mathrm{a}})$, the bisection method is applied to the interval $[q^{\mathrm{a}}(\Pi^{\mathrm{a}}, p), q^{\mathrm{b}}(\Pi^{\mathrm{b}}, p)]$ in target PS T along direction p. Thus, this interval is repeatedly halved until the position coordinate in target PS of the ray that follows the same path Π^{a} of $\boldsymbol{r}_{\mathrm{a}}$ is found (the corresponding direction coordinate p is fixed).

The bisection procedure continues until the length of the segment considered becomes smaller than a fixed tolerance. Giving as input the coordinates $q^{\mathrm{a}}(\Pi^{\mathrm{a}}, p)$ and $q^{\mathrm{b}}(\Pi^{\mathrm{b}}, p)$ of $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$, the path Π^{a} and the tolerance $\mathrm{tol} = 10^{-12}$, the bisection method is implemented as in Algorithm 4. Similarly the others paths will be find later using the same procedure applied to another interval in target PS.

Once bisection stops, two points with coordinates (q^{c}, p) and (q^{d}, p) in T are found. The corresponding rays $\boldsymbol{r}_{\mathrm{c}}$ and $\boldsymbol{r}_{\mathrm{d}}$ follow path $\Pi^{\mathrm{c}} = \Pi^{\mathrm{a}}$ and $\Pi^{\mathrm{d}} \neq \Pi^{\mathrm{a}}$. All the rays with target coordinates (q, p) and $q^{\mathrm{a}} \leq q \leq q^{\mathrm{c}}$ follow path Π^{a}, while the rays with target coordinates (q, p) with $q^{\mathrm{d}} \leq q \leq q^{\mathrm{b}}$ follow another path $\Pi \neq \Pi^{\mathrm{a}}$ (see Figure 8.1).

Now, if $k \neq 1$ the procedure applied to the interval $[q^{\mathrm{a}}(p), q^{\mathrm{b}}(p)]$ needs to be applied to $[q^{\mathrm{a}}(p), q^{\mathrm{c}}(p)]$ until the source is reached, i.e., until $k = 1$. If $k = 1$, the source is reached by the rays traced back from the target. This means that a possible path Π^{a} from S to T is found and the position coordinates $q^{\min}(\Pi^{\mathrm{a}}, p)$ and $q^{\max}(\Pi^{\mathrm{a}}, p)$ on T of the rays located at the boundaries $\partial R(\Pi^{\mathrm{a}})$ of the rays that follow that path are determined. To determine all the possible paths and the rays on the corresponding boundaries, the bisection procedure needs to be applied again to the interval $[q^{\mathrm{d}}(p), q^{\mathrm{b}}(p)]$. First, the rays on the boundary $\partial R(\Pi^{\mathrm{d}})$ will be determined.

Algorithm 4 Bisection

Initialize step $= 0$ $j =$ Nl

while $|q^{\mathrm{a}} - q^{\mathrm{b}}| >$ tol **do**
  $\Pi^{\mathrm{m}} =$ (Nl),
  $p^{\mathrm{m}} = p$,
  $\boldsymbol{r}_{\mathrm{m}} = q^{\mathrm{m}} + sp$ with $s > 0$ the arc-length,
  while step $<$ length$(\Pi^{\mathrm{a}}) - 1$ **do**
    Trace back $\boldsymbol{r}_{\mathrm{m}}$ from j
    Find the line k that the ray $\boldsymbol{r}_{\mathrm{m}}$ hits
    Find the coordinates $(q_k^{\mathrm{m}}, p_k^{\mathrm{m}})$ in T_k corresponding to $\boldsymbol{r}_{\mathrm{m}}$
    Calculate the direction p^{m} with respect to the optical axis
    $(q^{\mathrm{m}}, p^{\mathrm{m}}) = (q_k^{\mathrm{m}}, p_k^{\mathrm{m}})$
    $\Pi^{\mathrm{m}} = (k, \Pi^{\mathrm{m}})$.
    if $k = 1$ or $k =$ Nl **then**
      step $=$ length(Π^{a}) ▷ If the source or the target are reached
                                ▷ then exit from the while loop.
    else
      step $=$ step $+ 1$
    end if
  end while
  if $\Pi^{\mathrm{a}} = \Pi^{\mathrm{m}}$ **then**
    $(q^{\mathrm{a}}, p) = (q^{\mathrm{m}}, p)$
    $\boldsymbol{r}_{\mathrm{a}} = \boldsymbol{r}_{\mathrm{m}}$
  else
    $(q^{\mathrm{b}}, p) = (q^{\mathrm{m}}, p)$
    $\Pi^{\mathrm{b}} = \Pi^{\mathrm{m}}$
  end if
end while
$(q^{\mathrm{c}}, p) = (q^{\mathrm{a}}, p), \Pi^{\mathrm{c}} = \Pi^{\mathrm{a}}$.
$(q^{\mathrm{d}}, p) = (q^{\mathrm{b}}, p), \Pi^{\mathrm{d}} = \Pi^{\mathrm{b}}$.
return (q^{c}, p), (q^{d}, p), Π^{c} and Π^{d}.

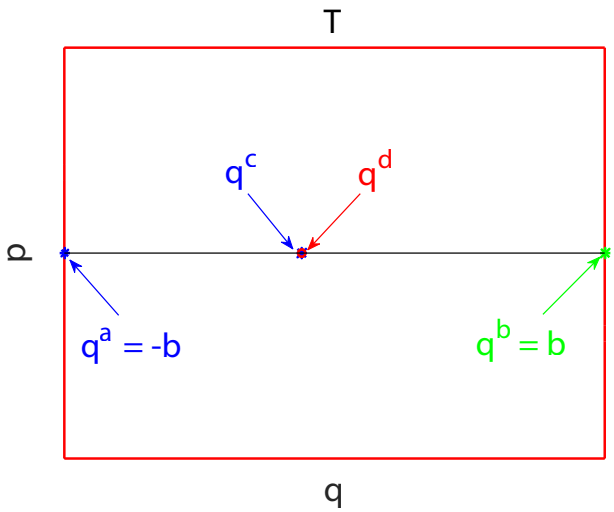


Figure 8.1: **Bisection in target PS *T*.** Algorithm 4 is run for the interval $[q^{\mathrm{a}}, q^{\mathrm{b}}]$ along direction $p = 0$. The coordinates q^{c} and q^{d} are found such that $|q^{\mathrm{c}} - q^{\mathrm{d}}| < \mathrm{tol}$. $\Pi^{\mathrm{c}} = \Pi^{\mathrm{a}}$ and $\Pi^{\mathrm{d}} \neq \Pi^{\mathrm{a}}$.

Finally, to detect all the possible paths that can occur along direction p the procedure explained above is applied also to the interval $[q^{\mathrm{d}}(p), q^{\mathrm{b}}(p)]$ along direction p continuing until all the interval $[q^{\mathrm{a}}(p), q^{\mathrm{b}}(p)]$ is investigated. The main steps of the method are outlined in the following.

1. Given a direction p, consider the end points $(q^{\mathrm{a}}, p^{\mathrm{a}})$ and $(q^{\mathrm{b}}, p^{\mathrm{b}})$ of the target PS T, where $q^{\mathrm{a}} = -b$, $q^{\mathrm{b}} = b$ and $p^{\mathrm{a}} = p^{\mathrm{b}} = p$. Start from $j = \mathrm{Nl}$
2. Using inverse ray tracing, trace back the rays from line j $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ corresponding to the coordinates $(q_j^{\mathrm{a}}, p_j^{\mathrm{a}})$ and $(q_j^{\mathrm{b}}, p_j^{\mathrm{b}})$.
3. Determine indices $k \neq j$ and $l \neq j$ of the lines from which $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are originated.
4. Update the paths Π^{a} and Π^{b}, $\Pi^{\mathrm{a}} = (k, \Pi^{\mathrm{a}})$ and $\Pi^{\mathrm{b}} = (l, \Pi^{\mathrm{b}})$,
5. If $k = l \neq 1$ and $k = l \neq \mathrm{Nl}$
   - $j = k$
   - Restart the procedure from point 2
6. If $k = l = 1$
   - A relevant path $\Pi^{\mathrm{a}} = \Pi^{\mathrm{b}}$ is found.
   - Determine
   $$q^{\min}(\Pi^{\mathrm{a}}, p) = \min\{q^{\mathrm{a}}(\Pi^{\mathrm{a}}, p), q^{\mathrm{b}}(\Pi^{\mathrm{b}}, p)\}$$
   $$q^{\max}(\Pi^{\mathrm{a}}, p) = \max\{q^{\mathrm{a}}(\Pi^{\mathrm{a}}, p), q^{\mathrm{b}}(\Pi^{\mathrm{b}}, p)\}.$$

- Update the intensity

$$I(p) = I(p) + q^{\max}(\Pi^{\mathrm{a}}, p) - q^{\min}(\Pi^{\mathrm{a}}, p)$$

7. If $k \neq l$

   - Apply the bisection method to the interval $[q^{\mathrm{a}}, q^{\mathrm{b}}]$ along direction p.
   - Find the points with coordinates (q^{c}, p) and (q^{d}, p) in target PS T where $|q^{\mathrm{c}} - q^{\mathrm{d}}| < \mathrm{tol}$.
   - If $k \neq \mathrm{Nl}$
     - $(q^{\mathrm{b}}, p) = (q^{\mathrm{c}}, p)$,
     - $(q_k^{\mathrm{b}}, p_k^{\mathrm{b}}) = (q_k^{\mathrm{c}}, p_k^{\mathrm{b}})$
     - $j = k$
     - Restart from 2 with the updated coordinates,
   - Update $(q^{\mathrm{a}}, p) = (q^{\mathrm{d}}, p)$,
   - $\Pi^{\mathrm{a}} = \Pi^{\mathrm{c}}$
   - Update $(q_k^{\mathrm{a}}, p_k^{\mathrm{a}}) = (q_k^{\mathrm{d}}, p_k^{\mathrm{d}})$
   - Restart from 2

Giving as input $I(p) = 0$ for every direction p and the tolerance $\mathrm{tol} = 10^{-12}$, the method is defined in the recursive Algorithm 5.

The explained procedure is able to determine all the possible paths that the rays can follow during their propagation from S to T. Also, the rays located on the boundaries of the regions with positive luminance on target PS T are found. The method is summarized by the flowchart in Figure 8.2

Next, the method is applied to two optical systems formed by curved lines. In the next section we show the results for the TIR-collimator.

8.2 Results for the TIR-collimator

In this section we apply the extended inverse ray mapping to the TIR-collimator presented in Chapter 5 and depicted in Figure 5.10. The target PS of this system is the rectangular $T = [-b, b] \times [-1, 1]$ with $b = 9.7$. The aim is to detect all the possible path Π and the rays located on the boundaries $\partial R(\Pi)$ of the corresponding regions in target PS.

In Chapter 5, we found five different paths for the TIR-collimator (see Figure (5.11). The boundaries of the corresponding regions in target PS T are in general difficult to approximate. Furthermore, along one direction p more than two points can be located on the boundary $\partial R(\Pi)$ of the region $R(\Pi)$ corresponding to a certain path Π. To determine properly all the boundaries $\partial R(\Pi)$, we need to divide the interval $[-b, b]$ in T into intervals of the same length (bins). Hence, we consider a partitioning $Q = -b = q^0 < q^1 < \cdots < q^{\mathrm{Ni}} = b$ of $[-b, b]$ where Ni is the total number of bins along the q-axis. For each direction $p \in [-1, 1]$ the procedure explained in Section 8.1 is repeated for every sub-interval $[q^{\mathrm{k}}(p), q^{\mathrm{k}+1}(p)] \subset [q^{\mathrm{a}}(p), q^{\mathrm{b}}(p)]$ with $\mathrm{k} = 0, \cdots, \mathrm{Ni} - 1$ and $q^{\mathrm{a}}(p) = -b$ and $q^{\mathrm{b}}(p) = b$.

Algorithm 5 Recursive function for the numerical inverse ray mapping

Initialize $j = \mathrm{Nl}$, $q^{\mathrm{a}} = q_j^{\mathrm{a}} = -b$, $q^{\mathrm{b}} = q_j^{\mathrm{b}} = b$, $p = p_j^{\mathrm{a}} = p_j^{\mathrm{b}} = \mathrm{const}$, $\Pi^{\mathrm{a}} = (\mathrm{Nl})$.

procedure INTENSITY COMPUTATION(q^{a}, q^{b}, q_j^{a}, q_j^{b}, p, p_j^{a}, p_j^{b}, Π^{a}, j)
    Apply inverse ray tracing to the coordinates $(q_j^{\mathrm{a}}, p_j^{\mathrm{a}})$ and $(q_j^{\mathrm{b}}, p_j^{\mathrm{b}})$ on j of rays $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$.
    Determine the lines $k \neq j$ and $l \neq j$ from which $\boldsymbol{r}_{\mathrm{a}}$ and $\boldsymbol{r}_{\mathrm{b}}$ are emitted.
    Update path $\Pi^{\mathrm{a}} \leftarrow (k, \Pi^{\mathrm{a}})$
    Calculate the coordinates $(q_k^{\mathrm{a}}, p_k^{\mathrm{a}})$ and $(q_k^{\mathrm{b}}, p_k^{\mathrm{b}})$ on k
    if $k = l$ **then**
        if $k \neq 1$ **then**
            return INTENSITY COMPUTATION(q^{a}, q^{b}, q_k^{a}, q_k^{b}, p, p_k^{a}, p_k^{b}, Π^{a}, k)
        else
            Calculate $q^{\min} = \min\{q^{\mathrm{a}}, q^{\mathrm{b}}\}$ and $q^{\max} = \max\{q^{\mathrm{a}}, q^{\mathrm{b}}\}$
            $I(p) = I(p) + q^{\max}(\Pi^{\mathrm{a}}, p) - q^{\min}(\Pi^{\mathrm{a}}, p)$,
        end if
    else
        Apply bisection to the segment $[q^{\mathrm{a}}(\Pi^{\mathrm{a}}, p), q^{\mathrm{b}}(\Pi^{\mathrm{b}}, p)]$
        Find the target coordinates (q^{c}, p) and (q^{d}, p) of the rays $\boldsymbol{r}_{\mathrm{c}}$ and $\boldsymbol{r}_{\mathrm{d}}$, where

$$|q^{\mathrm{c}} - q^{\mathrm{d}}| < \mathrm{tol}$$

        if $k \neq \mathrm{Nl}$ **then**
            return INTENSITY COMPUTATION(q^{a}, q^{c}, q_k^{a}, q_k^{c}, p, p_k^{a}, p_k^{c}, Π^{a}, k)
        end if
        return INTENSITY COMPUTATION(q^{d}, q^{b}, q_k^{d}, q_k^{b}, p, p_k^{d}, p_k^{b}, Π^{d}, k)
    end if
end procedure

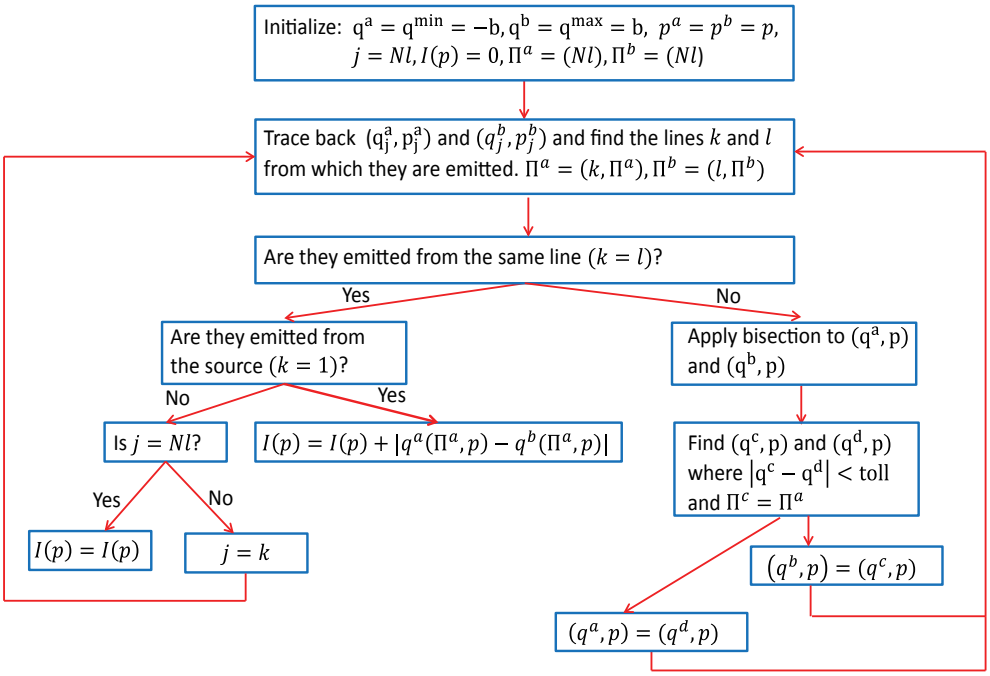


Figure 8.2: **Main steps of the numerical inverse ray mapping extended to systems with curved lines.**

To establish in how many bins Ni we need to divide the target, we exploit étendue conservation. We use the same idea applied to determine the value of α for α-shapes methods and to provide a stopping criterion for the triangulation refinement (see Chapters 5 and 6). The source étendue U_1 is calculated from Equation (5.3.2), obtaining $U_1 \approx 7.7$. The target étendue U_t is given by Equation (5.2.7). U_t is calculated several times considering every time a different partitioning Q for the q-axis of the target PS. Next, the absolute value of the difference between the source and target étendue is obtained from

$$\Delta U = \left| U_1 - U_t \right|. \tag{8.2.1}$$

If a small value of ΔU is obtained, then a good approximation of U_t is found and therefore, the partition Q used for the computation of U_t is suitable for detecting correctly the boundaries $\partial R(\Pi)$. In Figure 8.3 we show how ΔU decreases by increasing the number of bins Ni in the partitioning Q. In Figure 8.4 we show the distribution of the rays traced using the inverse ray mapping method with Ni = 30 and Nb = 100. In this case, inverse ray mapping detects 11 different paths from the source to the target. We observe that only 5 of them are the paths that we expected from PS ray tracing which are

$$\begin{array}{llll} \Pi_1 & = (1,2,7,12), & & \\ \Pi_2 & = (1,4,6,7,12), & \Pi_2 & = (1,10,8,7,12), \\ \Pi_4 & = (1,3,7,12), & \Pi_5 & = (1,11,7,12). \end{array} \tag{8.2.2}$$

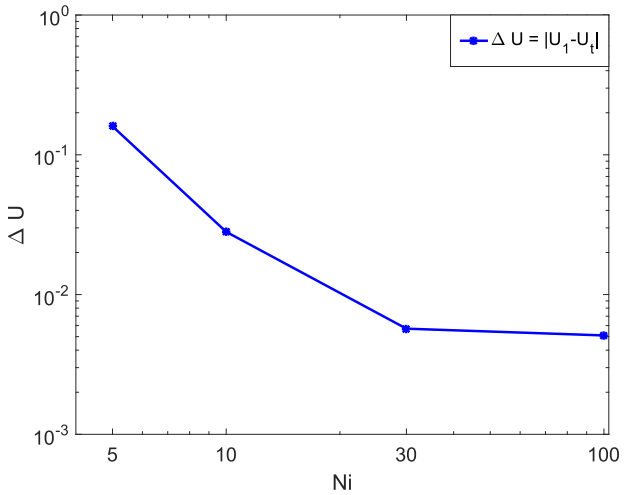


Figure 8.3: **Difference between the source and the target étendue for the TIR-collimator.** U_1 is calculated from Equation (5.3.2). U_t is computed four times increasing every time the number of bins Ni where $\mathrm{Ni} = \{5, 10, 30, 100\}$.

The inverse ray mapping also detect the spurious paths:

$$\begin{array}{llll} \Pi_6 & = (1,2,9,8,7,12), & \Pi_7 & = (1,2,5,6,7,12), \\ \Pi_8 & = (1,2,2,7,12), & \Pi_9 & = (1,7,12), \\ \Pi_{10} & = (1,2,4,6,7,12), & \Pi_{11} & = (1,2,10,8,7,12). \end{array} \tag{8.2.3}$$

Those paths are due to numerical error, the rays that follow these paths are rimmed in blue in Figure 8.4. The numerical error can be related to the precision of the bisection method and inverse ray tracing. We remark that in inverse ray tracing, the intersection between the ray and the lens (line 2) is computed using the Newton-Raphson procedure. To detect only the boundaries of the regions formed by the rays that follow a *real* path Π_j, with $j \in \{1, \cdots, 5\}$, we check the index of refraction that every ray has once it arrives at the source. If this is equal to the same index of S ($n = 1$ for the TIR-collimator), then the ray follows a physical path, otherwise it follows one of the paths in (8.2.3) and, therefore, it is not considered for the intensity calculation. This gives the rays distribution at the target PS shown in Figure 8.5 where $\mathrm{Ni} = 30$. We observe that, discarding those rays, 5 different paths are found. These are the same paths we obtained using PS ray tracing. Figure 8.5 shows that the rays on the boundaries $\partial R(\Pi_j)$ are determined for every path Π_j with $j \in \{1, \cdots, 5\}$. The ability of the inverse ray mapping method to recognize spurious paths makes it suitable to detect ghost stray light that is unwanted light which can reduce the performance of optical systems [55]. Optical designers are interested in developing methods for minimizing stray light intensity [56]. Inverse ray mapping could constitute an alternative approach for such purpose.

Note that some rays in the interior of the regions are still traced. This is related to the fact that we divide the target PS along the q-axis into $\mathrm{Ni} = 30$ bins. As a

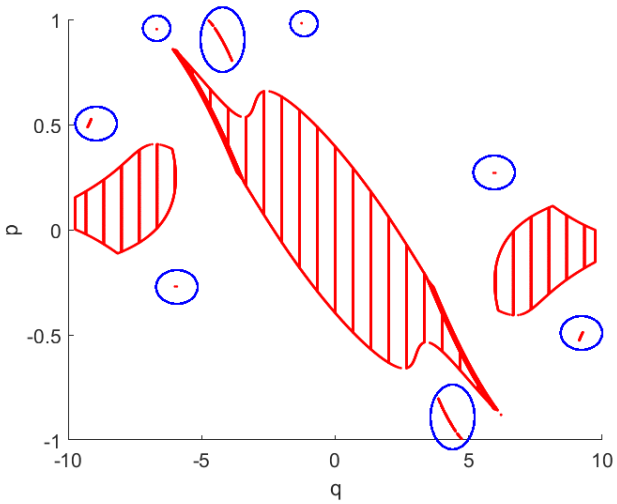


Figure 8.4: **Rays distribution at target PS of the TIR-collimator.** The q-axis is divided into Ni $= 30$ bins, the p-axis is divided into Nb $= 100$ bins. Around $6 * 10^3$ rays are traced from the target to the source (red dots). Because of numerical errors, few rays outside the region with positive luminance are found (rays rimmed in blue).

consequence, also the rays located at the end points of every bin are traced.

The target ray mapping intensity $\hat{I}_{\mathrm{RM}}$ is calculated using Equation 5.2.8. The profile of $\hat{I}_{\mathrm{RM}}$ with Ni $= 30$ is depicted in Figure 8.6 with the red line. It is compared with the reference intensity (blue line) that is given by QMC ray tracing with 10^7 rays. The picture shows that the inverse ray mapping method calculates the intensity correctly.

Finally, we compare inverse ray mapping with QMC ray tracing. The errors are obtained from Equation (5.3.5) and are shown in a logarithmic scale in Figure 8.7 as a function of the CPU-time. The approximation of the RM intensity $\hat{I}_{\mathrm{RM}}$ is improved by increasing the number of bins Ni in the partitioning Q. The approximated QMC intensity $\hat{I}_{\mathrm{QMC}}$ is calculated several times gradually increasing the number of rays Nr. Both intensities are computed using the same number of bins, Nb $= 100$ in the partitioning P of the p-axis. The minimum ray mapping error is obtained with Ni $= 100$ bins, while the minimum QMC error is achieved tracing 10^6 rays. We observe that the minimal error obtained using the inverse ray mapping is of an order of magnitude of 10^{-6}, while the minimum QMC error is of the order of 10^{-5}. Furthermore, an extrapolation of the QMC error shows that the inverse ray mapping is more than 1000 times faster compared to QMC. In Tables 8.1 the numerical results obtained for the inverse ray mapping are reported. The error values for QMC ray tracing were already reported in Chapter 6 (Table 6.2).

In the next section we present the method for a parabolic reflector.

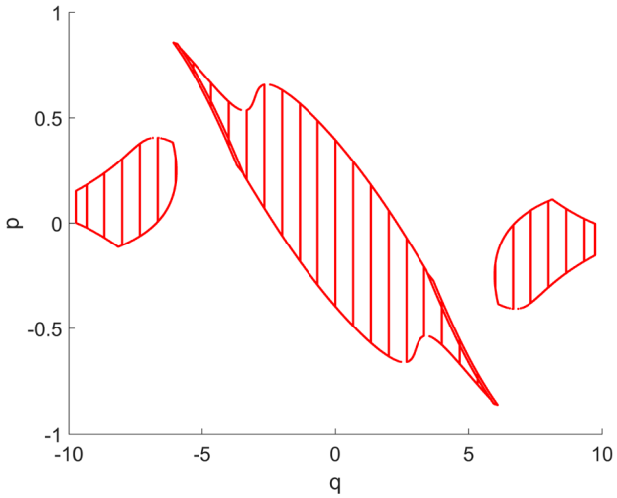


Figure 8.5: **Rays distribution at target PS of the TIR-collimator.** The q-axis is divided into Ni = 30 bins, the p-axis is divided into Nb = 100 bins. Considering only the rays that arrive to the source with the correct index of refraction (Nl = 1), the regions with positive luminance are computed correctly.

Table 8.1: **Errors of the PS intensity for the TIR-collimator**

Ni	$\lvert\Delta U\rvert$	RM error	CPU-time (sec.)
5	$1.6 \cdot 10^{-1}$	$5.19 \cdot 10^{-4}$	269
10	$2.8 \cdot 10^{-2}$	$2.09 \cdot 10^{-4}$	284
30	$5.7 \cdot 10^{-3}$	$3.15 \cdot 10^{-6}$	313
100	$5.1 \cdot 10^{-3}$	$2.52 \cdot 10^{-6}$	359

8.3 Results for the parabolic reflector

In this section we provide the results for the parabolic reflector in Figure 6.10. This is a very challenging example of an optical system. Indeed, the rays that propagate through such a system can reflect many times along the left and the right mirror. As we have seen using PS ray tracing, this leads to many different paths. Every path corresponds to a given number of reflections with one of the two reflectors. In Chapter 6 we found 17 different paths for this parabolic reflector. Here, we apply the inverse ray mapping method to detect all this paths.

The target PS of the parabolic reflector is the rectangular $T = [-b, b] \times [-1, 1]$ where $b = 17$. Like for the TIR-collimator, we divide the interval $[-b, b]$ in target PS into sub-intervals of the same length (bins). We remind the reader that we indicate with Ni the number of bins in which we divide the interval $[-b, b]$ along the q-axis of T. Considering a partitioning $Q = -b = q^0 < q^1 < \cdots < q^{\mathrm{Ni}} = b$ of $[-b, b]$ and a direction $p \in [-1, 1]$, the inverse ray mapping explained in Section 8.1 is applied to every sub-interval $[q^{\mathrm{k}}, q^{\mathrm{k}+1}] \subset [-b, b]$ with $\mathrm{k} = 0, \cdots, \mathrm{Ni} - 1$ and for every direction $p \in [-1, 1]$.

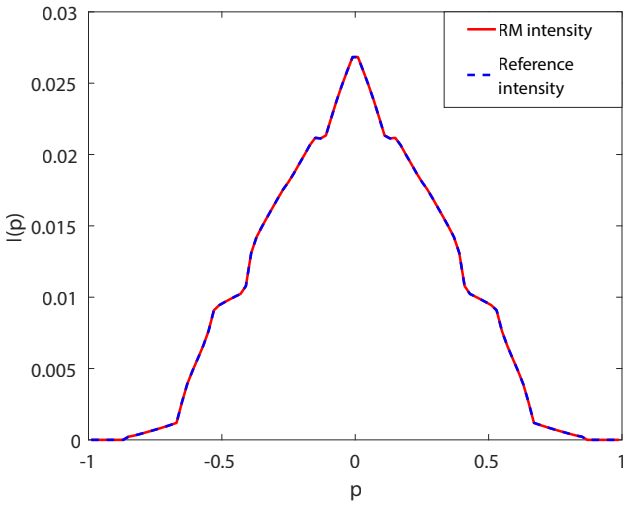


Figure 8.6: **Profile of the intensity for the TIR-collimator.** The ray mapping intensity is calculated dividing the q-axis into Ni $= 30$ bins. The reference intensity is obtained from QMC ray tracing with 10^7 rays.

To determine how many bins Ni are needed for a good calculation of the target photometric variables, we employ the same idea of the TIR-collimator. The source étendue U_1 is compared to several approximations of the étendue at the target U_t each of them is given by a different partitioning Q of $[-b, b]$. For the parabolic reflector all the rays emitted from the source arrive at the target. Therefore, the exact étendue as an area in PS is computed from Equation (6.1.2) obtaining $U = U_1 = 8$. The approximated target étendue U_t is given by Equation (5.2.5). In Figure 8.8 we show the comparison between U_1 and several approximations of U_t by gradually increasing the number of bins Ni in the partitioning Q while fixing the maximum number of multiple reflections to 30. Increasing the number of bins Ni, U_t increases approaching the exact value $U_1 = 8$. After the division into Ni $= 4$ bins the improvement is slightly visible. Therefore, we conclude that Ni $= 4$ bins are enough to detect 30 multiple reflections.

In Figure 8.9 we show the ray distribution at the target PS obtained using inverse ray mapping with Ni $= 4$ and at most 30 multiple reflections. The rays traced from the target to the source are depicted with the red dots. Most of the rays traced are located on the boundaries $\partial R(\Pi)$ of the regions with positive luminance. Only few rays are traced inside those regions. These are the rays located at the end points of every bin $[q^\mathrm{k}, q^{\mathrm{k}+1}]$ with $\mathrm{k} = 0, \cdots, \mathrm{Ni} - 1$.

The inverse ray mapping method is able to detect 61 different paths. Indeed up to 30 multiple reflections can occur at the left reflector and the right reflector. The path that goes directly from the source to the target (no reflections with the reflectors) has to be added. For PS ray tracing we found at most 17 paths for the same parabolic reflector. Hence, we claim that ray mapping is much more accurate

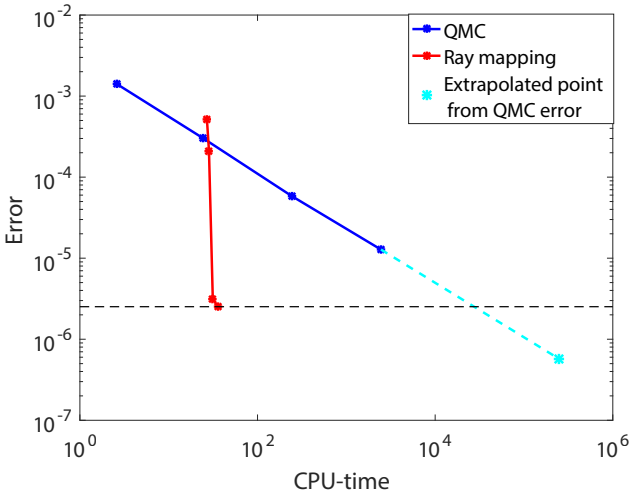


Figure 8.7: **Errors of ray mapping and QMC for the TIR-collimator.** The extended inverse ray mapping method is faster and more accurate than QMC ray tracing.

than PS ray tracing. Also, we observe the procedure can be stopped when even more than 30 multiple reflections are reached. The more reflections are considered better the accuracy obtained. Again, to define a stopping criterion we use étendue conservation. Fixing the number of bins Ni = 4 and Nb = 100 and gradually increasing the number of multiple reflections we see that the approximated target étendue U_{t} changes as the blue line in Figure 8.10. The horizontal red line represents the exact intensity $U = U_1 = 8$. We note that the more reflections are considered smaller the value of $\Delta U = |U_1 - U_{\mathrm{t}}|$ is (see also Table 8.2). We observe that after around 30 multiple reflections there is no significant improvement in the computation of U_{t}. This is due to the fact that only few rays follow multiple reflections and therefore, they do not give a significant contribution to the total étendue, the regions in PS formed by those rays are very small compared to the entire PS. From this we conclude that that inverse ray mapping has a good accuracy when around 30 multiple reflections and Ni = 4 bins are taken into account. Once a stopping criterion is established, ray mapping is run and the rays on the boundaries are determined. Finally, the target intensity is calculated from Equation (4.3.4).

In Figure 8.11 both the ray mapping intensity (red line) and the reference intensity (dotted blue line) are shown. The ray mapping intensity $\hat{I}_{\mathrm{RM}}$ is obtained considering at most 30 multiple reflections, Ni = 4 and Nb = 100. The reference intensity $\hat{I}_{\mathrm{ref}}$ is given by QMC ray tracing with 10^8 rays and Nb = 100. The two intensities coincide.

To conclude we calculate the errors between the approximated intensity $\hat{I}_{\mathrm{A}}(\mathrm{A} = \mathrm{RM}, \mathrm{QMC})$ and the reference intensity $\hat{I}_{\mathrm{ref}}$. From the results in Figure 8.12 we observe that the RM error (red line) converges faster than QMC error (blues line) as long as

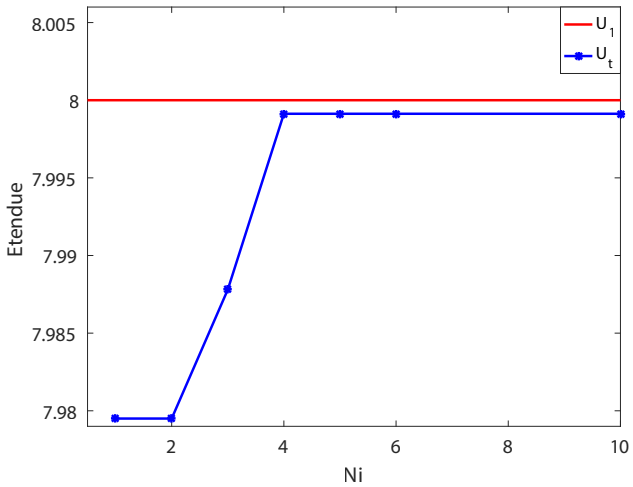


Figure 8.8: **Comparison between the exact étendue and the approximated target étendue by increasing Ni.** At most 30 multiple reflections are considered. Increasing Ni, the target étendue gets closer to the exact value $U_1 = 8$.

an error of an order of 10^{-6} is desired. Ray mapping results to be around 1000 times faster than QMC ray tracing. Furthermore, it is much more accurate than QMC. Our method is able to detect *all* the possible paths that can occur. The procedure is stopped when 200 multiple reflections are reached. Our expectation is that, increasing the number of multiple reflections and the number of bins Ni, the accuracy can be improved even more. In Tables 8.2 and 8.3 these numerical results are reported.

8.4 Conclusions

In this chapter we extended the analytic ray mapping method to systems formed by curved lines. Employing inverse ray tracing and a bisection procedure in target PS, an inverse map from the target to the source was constructed such that all the possible paths that the rays can follow are determined. The numerical inverse ray mapping method is able to detect the rays located on the boundaries of the regions formed by rays that follow the same path. From these rays the target intensity is calculated.

We presented to examples of optical systems: the TIR-collimator and the parabolic reflector. In both cases the target PS is divided into bins and the procedure is applied to each bin. A stopping criterion based on étendue conservation is developed to determine the number of bins needed to obtain a good accuracy. For the TIR-collimator we noticed that the method is able to detect rays that follow a spurious path due to numerical error. This gives the expectation that the inverse ray mapping can be used also for detecting ghost stray light. For the parabolic reflector, many paths can occur along the reflectors. Etendue conservation is used again to determine the number of multiple reflections to be considered. The target intensity is computed

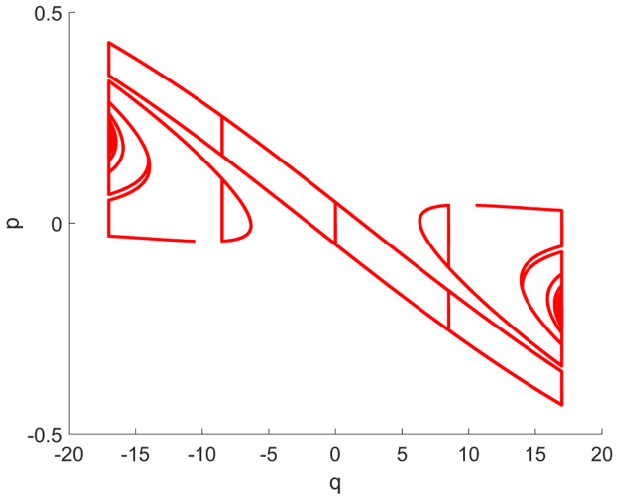


Figure 8.9: **Rays on the boundaries of the regions with positive luminance in target PS.** At most 30 multiple reflections are considered and Ni = 4 and Nb = 100. Only the rays on the boundaries and on the end points of each interval are traced.

for both systems and is compared with a reference intensity given by QMC ray tracing with a large number of rays. The results shown that the method is able to detect all the possible paths tracing a relatively small number of rays, typically around 10^3. Comparing our method to QMC ray tracing, significant advantages are observed in accuracy and computational time for both optical systems.

In the next chapter we explain how to apply the method to systems with Fresnel reflection. We present the method for a system formed by the source, an ideal lens and the target.

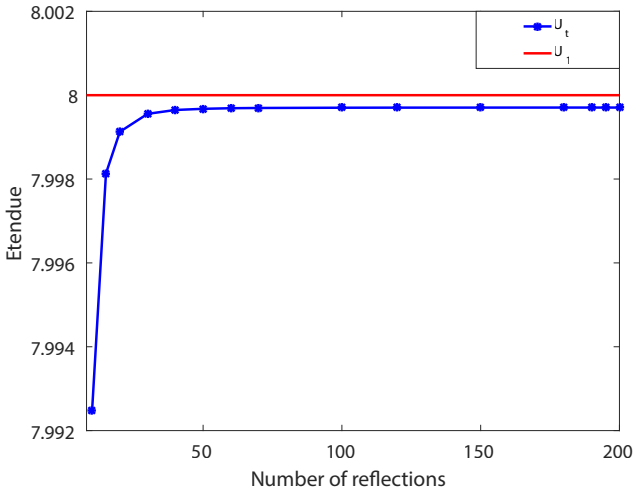


Figure 8.10: **Comparison between the exact étendue and the approximated target étendue by increasing the number of multiple reflections.** Fixing the number of bins along the q-axis Ni $= 4$ and increasing the number of reflections considered, the étendue increases approaching to the exact value $U_1 = 8$.

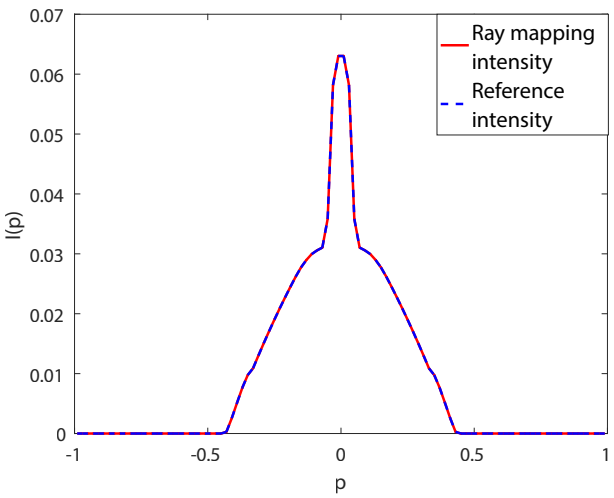


Figure 8.11: **Ray mapping intensity compared to a reference intensity.** The ray mapping error is calculated considering Ni $= 4$ and at most 30 multiple reflections. The reference intensity is obtained by running QMC ray tracing with 10^8 rays.

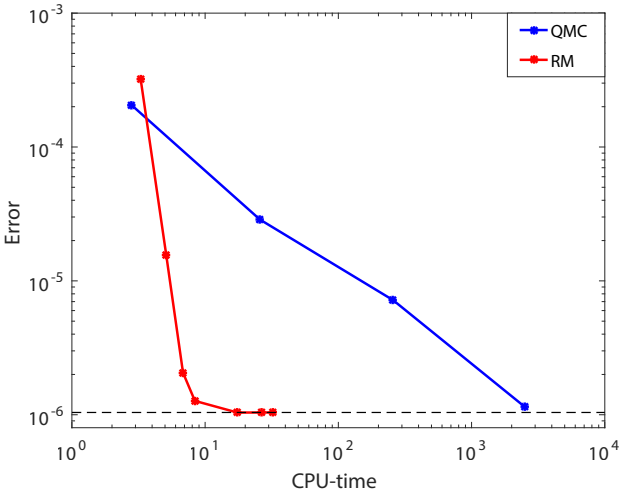


Figure 8.12: **Errors of ray mapping and QMC for the parabolic reflector.** The ray mapping error decreases by increasing the number of reflections considered. The QMC error reduces by tracing more rays. The extended inverse ray mapping method is faster and more accurate than QMC ray tracing.

Table 8.2: **Errors of the ray mapping intensity for the parabolic reflector**

Number of reflections	$\lvert\Delta U\rvert$	RM error	CPU-time (sec.)
5	$1.74 \cdot 10^{-1}$	$3.222 \cdot 10^{-4}$	3.28
10	$7.52 \cdot 10^{-3}$	$1.555 \cdot 10^{-5}$	5.11
20	$9.00 \cdot 10^{-4}$	$2.059 \cdot 10^{-6}$	6.83
30	$5.00 \cdot 10^{-4}$	$1.269 \cdot 10^{-6}$	8.38
100	$2.99 \cdot 10^{-4}$	$1.038 \cdot 10^{-6}$	17.49
150	$2.96 \cdot 10^{-4}$	$1.039 \cdot 10^{-6}$	26.38
200	$2.95 \cdot 10^{-4}$	$1.039 \cdot 10^{-6}$	32.21

Table 8.3: **Errors of the QMC intensity for the parabolic reflector**

Nr	QMC error	CPU-time (sec.)
10^4	$2.05 \cdot 10^{-4}$	2.81
10^5	$2.87 \cdot 10^{-5}$	25.81
10^6	$7.18 \cdot 10^{-6}$	257.54
10^7	$1.15 \cdot 10^{-6}$	2491.32

Chapter 9

Inverse ray mapping for systems with Fresnel reflection

In this chapter we present the inverse ray mapping method for systems formed by optical lines for which Fresnel reflection can occur. In particular, we are interested in explain how to calculate the boundaries of the regions with positive luminance in target PS. We provide numerical results for a simple system formed by the source, a Fresnel lens and the target. The chapter begins with a general introduction of the method, then the details are explained for a simple system, finally we show the results obtained for this system.

9.1 Introduction

Including Fresnel reflections, every time that a ray hits an optical line, a part of it is reflected and the other part is refracted. Hence, a ray emitted from the source generates two rays after interactions with the optical lines (the reflected ray and the transmitted one) each of them has the percentage of the original power given by the reflectance $\mathcal{R}$ and the transmittance $\mathcal{T}$ obtained from Fresnel equations (see Section 2.3). At the next intersection of the two rays with a line, each incident ray is split again. This process is called ray splitting. The ray emitted from the source is called the parent rays. Once it hits a line one of the two rays keeps the parent status while the other one is called child ray. The convention is to assign the parent status to the ray with the most power. This results tracing a rapidly increasing number of rays, including those with a very little power (at every interaction the power of the ray decreases). Many different paths can occur and the number of paths is determined considering all the possible combinations that a ray can have. The number of possibilities increases with the number of optical lines that form the systems. For example, for a system formed by a source, a Fresnel line and a target, two possibilities can occur since a ray emitted from the source can either refract reaching the target or reflect going back to the source. Note that no scattering phenomena are considering in this

thesis. If a system is formed by a source, two Fresnel lines and a target much more paths are possible since the ray can reflect forward and backward between the two optical lines before arriving either at the target or at the source again. For systems where multiple reflections can occur the number of paths can increases dramatically. Ray splitting is a powerful method to calculate *all* physical paths but it generates more and more rays many of which are not significant for computing the target intensity. This makes ray splitting methods time consuming. There exist methods to limit the number of child rays generated. Ray generation can be controlled, for example, neglecting rays with a small flux or limiting the total number of interaction that a ray can undergo during its propagation. A different approach could be to decide for each ray whether its reflected or refracted part has to be considered every time that an optical line is encountered. Such methods are called *one ray in one ray out* because the number of rays emitted from the source equals the number of rays traced through the system.

MC and QMC ray tracing are examples of these kind of processes. They consider a single ray every time that it hits an optical line. At every intersection between the ray and the line the transmittance $\mathcal{T}$ and reflectance $\mathcal{R}$ are determined. Next, a randomly generated number between 0 and 1 establishes which path the rays will continue to follow. Usually the assumption is to consider the reflected ray when $\mathcal{R}$ is greater than the random number, and the transmitted one otherwise. Therefore, for each ray traced a unique path is possible and the number of rays emitted from the source corresponds to the total number of rays traced. Compared to methods that consider all possible paths, MC and QMC are easy to implement and the computational time required to achieve a good accuracy is reduced. However it has some disadvantages. First, part of the energy is lost, therefore the luminance at the target will be always less than the luminance at the source of the system. To achieve a very accurate luminance or intensity distribution at the target, more rays need to be traced such that more paths will be considered. Second, it can beneficial in proper modeling of an optical system to consider ray splitting (see [5]).

To improve the exiting methods we here apply the numerical inverse ray mapping to systems formed by Fresnel lines. For every incident ray we take into account of both the reflected and the refracted ray. Therefore, for a given position and direction of a ray emitted by the source at least two possible paths are allowed. Although not all of these paths are *physical* paths, that is paths from the source to the target, it can happen that starting from the same ray coordinates, more than one physical path is permitted. A point in source PS can be associated to two or more points in target PS. Therefore, considering Fresnel reflections, the regions in target PS formed by the rays that follow the same path overlap. This would not make the inverse ray mapping suitable for this systems.

To overcome this issue we modify the inverse ray mapping by considering the physical paths one by one. Every time that the inverse ray tracing is applied, we consider either the reflected of the refracted ray depending on the path of which we want to determine the boundary. This allows determining the boundaries of the corresponding region in target PS. Furthermore, the power of every ray at each intersection with a line is calculated. Hence, once the rays on the boundaries are determined, their corresponding luminance at the target is also calculated. Note that the output luminance cannot be constant along a given direction and it depends on both the position

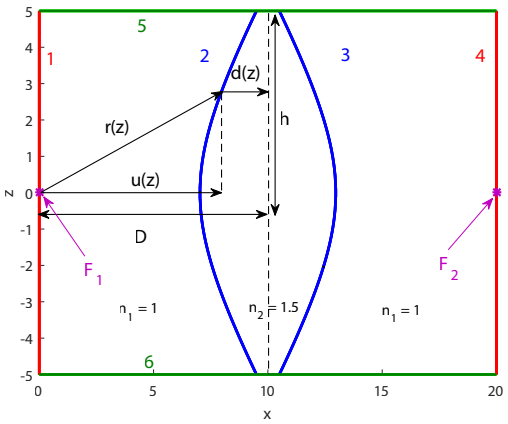


Figure 9.1: **A Fresnel lens** The source (line 1) and the target (line 4) are located in air ($n = 1$), the material between lines 2 and 3 has index of refraction $n = 1.5$. Two reflectors (lines 5 and 6) are located at the top and the bottom of the lens to detect light exiting from the system. $\mathsf{F}_1 = (0, 0)$, $\mathsf{F}_2 = (20, 0)$, $D = 10$, $h = 5$, $d(h) = 0.5$.

q and the direction p in target PS. To determine the luminance related to the rays that follow a certain path, an interpolation between the rays on the corresponding boundary in PS is done along every direction. The procedure is repeated for every path (or for all the paths needed to obtain a good accuracy). The partial luminance related to every path is computed. The total luminance is given by the sum of all the partial luminance calculated. Finally, the intensity is given by an integration of the luminance along all the possible positions.

The method is presented in detail in the next section for a simple system where Fresnel reflection plays a role.

9.2 Inverse ray mapping for a Fresnel lens

In this section we present the numerical inverse ray mapping for a simple optical system formed by the source, the target and a Fresnel lens. First we describe the geometry of the system, then we explain how the idea of the method for this system.

9.2.1 The geometry of the system

Let us consider the optical system depicted in Figure 9.1. It is formed by the source (line 1), a lens formed by two optical lines (line 2 and 3), the target (line 4) and two detectors (line 5 and 6) at the top and the bottom of the system. The source S is a vertical line segment located at $x = 0$ and with $z \in [-h, h]$ where $h = 5$. The target T is a line segment of the same length of the source S, parallel to S and located at a distance $x = a$ from S with $a = 20$. The lens is formed by two refractive lines (simple lens), and it is convex as it is thicker at its middle point than at its edges. The lines 2

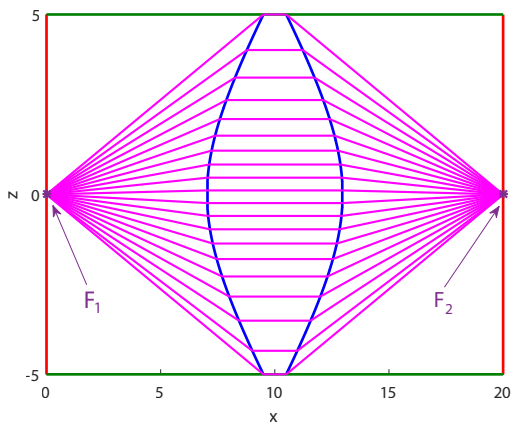


Figure 9.2: **A Fresnel lens.** All the rays exiting from F_1 arrive at F_2.

and 3 that form the lens are symmetric to each other with respect to the axis $x = 10$. The optical axis is the line $z = 0$. The point $\mathsf{F}_1 = (0, 0)$ and $F_2 = (a, 0)$ with $a = 20$ are the focal points of line 2 and 3 respectively.

To derive the equations of the curved lines we impose that the optical path length is constant. In the following we explain how the equation for the left lens (line 2) is obtained. We indicate with $d(z)$ half of the lens thickness at high z, with $\boldsymbol{r}(z)$ the parametrization of the ray emitted from F_1 and that reaches line 2 at high z, $u(z)$ is the projection of $\boldsymbol{r}(z)$ on the optical axis and $D = 10$ is the distance between the F_1 and the axis $x = 10$. The following relations hold:

$$\begin{aligned} \boldsymbol{r}^2(z) &= (u(z))^2 + z^2, \\ u(z) &= (D - d(z))^2, \end{aligned} \tag{9.2.1}$$

The optical path length is given by:

$$|\boldsymbol{r}(z)| + n_2|d(z)| = \tilde{C}, \tag{9.2.2}$$

where $\tilde{C}$ is a constant greater than 0. Substituting Equations (9.2.1) in the previous equation, we obtain for every $z \in [-h, h]$ the expression for line 2 given by:

$$Au(z)^2 + B(z) + C + z^2 = 0 \tag{9.2.3}$$

where

$$\begin{aligned} A &= (1 - n_2^2), \\ B &= 2n_2(n_2 D - \tilde{C}), \\ C &= -\tilde{C}^2 + 2n_2 D\tilde{C} - n_2^2 D^2. \end{aligned} \tag{9.2.4}$$

The right lens is given by a reflection with respect to the axis $x = 0$ and a translation $(x, z) \to (x + 20, z)$.

The lens constructed in this way has the property that every ray that hits each

curved line is bent towards the optical axis. The rays passing through it connects the conjugate points F_1 and F_2 (see Figure 9.2). Therefore, all the rays diverging from F_1 converge to F_2. In the next section we explain how to apply the numerical inverse ray mapping to this system.

9.2.2 The explanation of the method

Our goal is to show that, using the numerical inverse ray mapping extended to system with Fresnel reflection, we can detect *all* the possible paths limiting the number of rays traced. Furthermore, we are able to determine the rays that are located at the boundaries of the regions formed by the rays that follow a physical path. First, let us analyze the inverse ray tracing process for the system in Figure 9.1 introduced in the previous section. Every ray is emitted from the target T with an angle $\theta_1 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, that is with a direction $p_1 \in [-1, 1]$. Every time that a ray hits a Fresnel lens, i.e. either line 2 or 3, the ray is split in two rays, the reflected ray and the refracted one. The reflectance $\mathcal{R}$ and the transmittance $\mathcal{T}$ are calculated. The reflected ray will continue to propagate within the system with the percentage of power given by $\mathcal{R}$. Similarly the transmitted ray will travel inside the system with the power obtained by $\mathcal{T}$. The trajectory of each ray is stopped either when it arrives at the source S or when it reaches again the target T. This leads to many different paths for a single ray with a given initial position and direction. Note that at every interaction with a Fresnel line the ray loses some energy due to the fact that it is divided into two more rays.

We start considering the target PS of the system in Figure 9.1. Since, for the system in Figure 9.1 the source and the target are vertical lines (the x coordinates are fixed while varying the z coordinates), the PS coordinates are (q, p), where p is the ray direction coordinate defined as for the previous systems while q is the z coordinate of the x coordinate. To have an idea of the rays distribution at the target we represent the target coordinates on PS of 10^5 rays traced using MC ray tracing. MC ray tracing decides randomly which part of the ray has to be considered at every interaction with a line. Then the ray propagates with the corresponding energy. Running MC ray tracing for the lens in Figure 9.1 with 10^5 rays and storing the corresponding paths, the following three paths are found:

$$\begin{aligned} \Pi_1 &= (1, 2, 3, 4), \\ \Pi_2 &= (1, 2, 3, 2, 3, 4), \\ \Pi_3 &= (1, 2, 3, 2, 3, 2, 3, 4). \end{aligned} \tag{9.2.5}$$

In Figure 9.3 we provide the target PS of the Fresnel lens where we depicted with the same color the rays that follow the same path. The rays in magenta are rays that follow path Π_1, the cyan rays have two reflections before reaching the target following path Π_2, the black rays have three reflections inside the lens, these rays follow path Π_3 (see Equation (9.2.5)). Note that the rays with more reflections have a less frequency inside the system. Furthermore, some areas inside the region $R(\Pi_1)$ are not covered by any ray (white parts). Therefore, rays that leave two close positions coordinates at the source with a close direction coordinates can follow different paths during their propagation. This can be due to either TIR or a different choice of rays (reflected or refracted) at the intersections with Fresnel lines. In the first case, the rays will

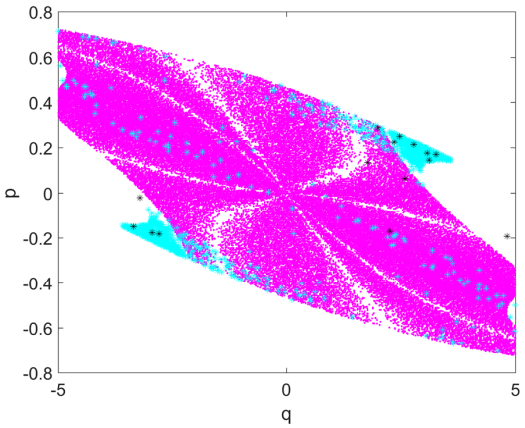


Figure 9.3: **Target PS of the Fresnel lens** The magenta rays follow path Π_1, the cyan rays follow path Π_2 and the black rays follow path Π_3.

be located outside the region $R(\Pi_1)$, otherwise they can reach the interior of $R(\Pi_1)$. The latter case implies that, given two different paths Π_i and Π_j with $i \neq j$, the corresponding regions $R(\Pi_i)$ and $R(\Pi_j)$ can overlap in target PS. Thus:

$$\bigcap_{\Pi} R(\Pi) \neq \emptyset \tag{9.2.6}$$

where the intersection is over all the possible paths. Because of this, we need to slightly modify the full inverse ray mapping explained in the previous chapter to detect all the possible paths.

The idea is to apply the bisection method combined with the inverse ray tracing to the entire PS for every single path Π separately. All the possible paths from the target to the source can be visualized in the tree in Figure 9.4. R and T indicate whether the reflected or the transmitted part of the ray is considered at every intersection point. Many possibilities can occur, we indicate with C the sequence of the choices that is made at every intersection with the Fresnel lens. For example, the path Π_1 in the reverse order that is $\Pi_1' = (4, 3, 2, 1)$ corresponds to the choice $C_1 = (T, T)$ from the target to the source (red path in Figure 9.4). This means that the rays that follow path Π_1 are transmitted on both line 2 and 3. To find the boundary $\partial R(\Pi_1)$ of the region corresponding to that path Π_1 along a given a direction $p = \text{const}$, we start considering the rays with corresponding coordinates (q^{a}, p) and (q^{b}, p) located at the end points of the target PS, where $q^{\text{a}} = -b$ and $q^{\text{b}} = b$ and $b = 5$. Then the bisection method and the inverse ray tracing is applied as explained in the previous chapter with the difference that now, every time that the ray is traced back either the reflected or the refracted ray is considered according to the choice $C_1 = (T, T)$. This allows us to determine the rays located on the boundaries $\partial R(\Pi_1)$. We indicate the coordinates of these rays with $(q^{\min}(\Pi_1, p), p)$ and $(q^{\max}(\Pi_1, p), p)$.

Since including Fresnel reflection the luminance at the target is not constant, an interpolation between the rays on the boundary $\partial R(\Pi_1)$ is needed. Some rays with

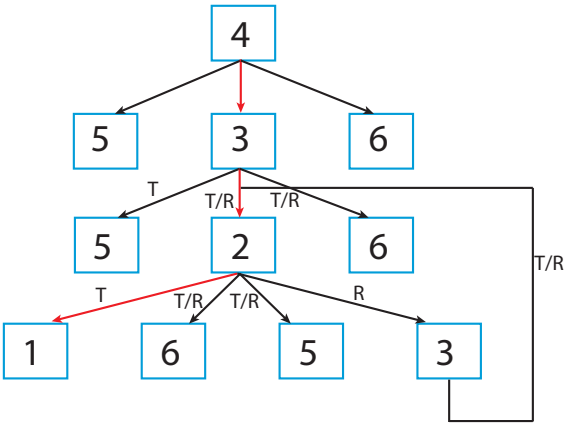


Figure 9.4: **Tree with all the possible paths.** Considering Fresnel reflections the rays can reflect many times between line 2 and 3 before reaching either the target or one of the two detectors or the source again. The inverse of path $\Pi_1 = (1, 2, 3, 4)$ that is $\Pi_1' = (4, 3, 2, 1)$ is depicted in red. It corresponds to the choice $C_1 = (T, T)$.

corresponding coordinates $(q(\Pi_1, p), p)$ where $q \in [q^{\min}(\Pi_1, p), q^{\max}(\Pi_1, p)]$ are traced back from the target to the source according to the choice C_1 at every intersection. We indicate with $L_{\Pi_1}(q, p)$ the luminance corresponding to path Π_1. The interpolation between $(q^{\min}(\Pi_1, p), p)$ and $(q^{\max}(\Pi_1, p), p)$ allows us to calculate the profile of $L_{\Pi_1}(q, p)$ for every $q \in [q^{\min}, q^{\max}]$ and $p = \text{const}$. Repeating the procedure for all the possible directions $p \in [-1, 1]$, the profile of the partial luminance L_{Π_1} given by all the rays that follow path Π_1 is computed for every $(q, p) \in T$.

Next another path Π and the corresponding choice C are considered. We remark that every path is associated to a unique choice but the vice versa is not always true. For example, the path $\Pi_2 = (1, 2, 3, 2, 3, 4)$ considered in reverse order from the T to S, i.e. $\Pi_2' = (4, 3, 2, 3, 2, 1)$ can only be associate to the choice $C_2 = (T, R, R, T)$ which leads also to the two inverse paths $\Pi_5' = (4, 3, 2, 3, 2, 5)$ and $\Pi_6' = (4, 3, 2, 3, 2, 6)$. The last two paths Π_5' and Π_6' are not paths from the target to the source, therefore they are not considered for the intensity calculation. Given a choice C many paths are possible but only one is a physical path. In the flowchart in 9.5 we show all the steps needed to calculate the partial luminance $L_{\Pi_1}(q, p)$ along a given direction p and related to path $\Pi_1 = (1, 2, 3, 4)$ which corresponds to a certain choice $C = (T, T)$. The procedure to determine the other paths is the same with the only difference that a different choice has to be considered. Considering *all* the possible choices, *all* the physical paths are determined. The corresponding partial luminance $L_\Pi(q, p)$ along every direction $p \in [-1, 1]$ and for every position coordinate $q \in [q^{\min}(\Pi, p), q^{\max}(\Pi, p)]$ is calculated for every path Π.

Finally, the total luminance $L(q, p)$ is given by:

$$L(q, p) = \sum_{\Pi} L_\Pi(q, p), \tag{9.2.7}$$

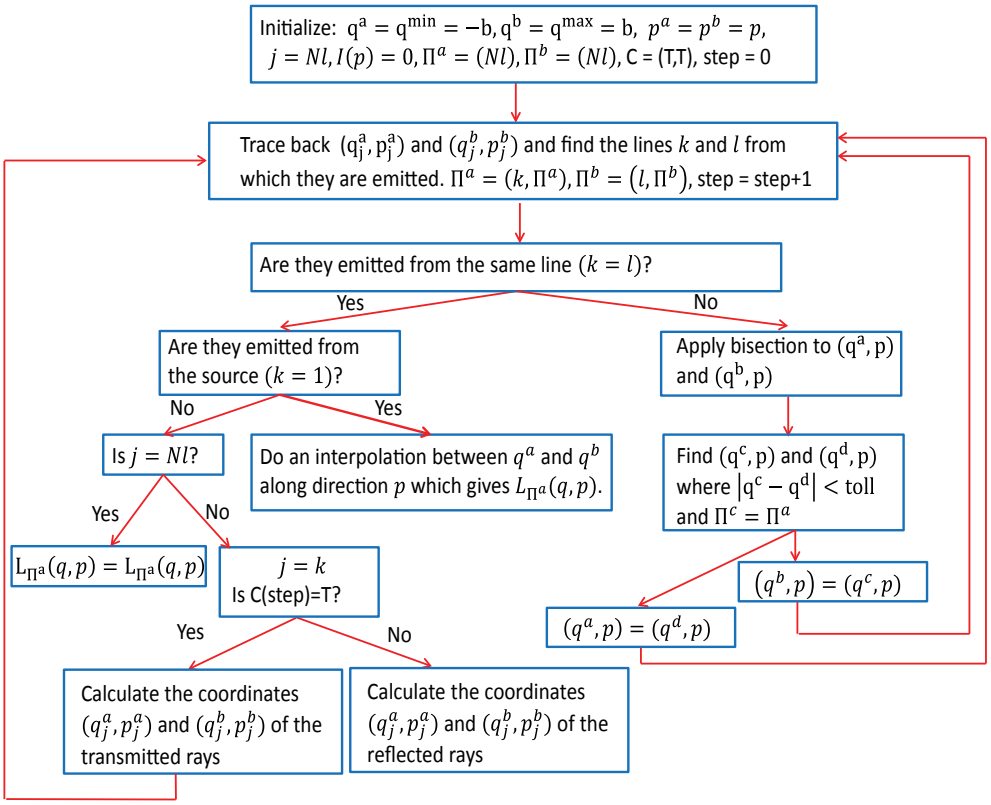


Figure 9.5: **Main steps of the inverse ray mapping extended to systems with Fresnel reflection.**

where the summation is over all the possible paths. From an integration of $L(q,p)$ the intensity at the target is given by:

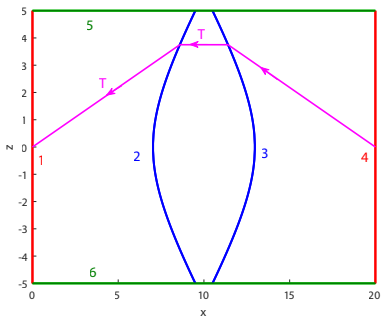
$$I(q,p) = \int_Q L(q,p)\mathrm{d}q. \tag{9.2.8}$$

In the next section numerical results are shown.

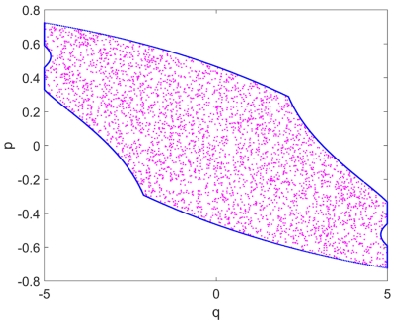
9.3 Numerical results

In this section the numerical results for the lens are presented. We detect the boundaries of the regions formed by the rays that follow the path for every physical path. To this purpose we decide a priori the choice that has to be done at every intersection of the ray with a Fresnel lens. The choice establishes whether the reflected or the transmitted part of the incident ray has to be considered when the rays traced back hit a Fresnel line.

We start considering the choice $C_1 = (T,T)$ which corresponds to the physical



(a) **Ray traced back from the target to the source**. The rays follows path $\Pi_1 = (1, 2, 3, 4)$ corresponding to the choice $C_1 = (T, T)$ from the target to the source. The percentage of power of the ray at the source is around the 55% of the initial power.



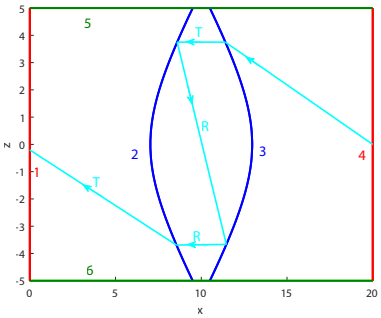
(b) **Boundary $\partial \boldsymbol{R}(\Pi_1)$.** The boundary is depicted with the blue line, the red rays are traced using MC ray tracing and considering always the transmitted ray.

Figure 9.6: **Computation of the boundary $\partial \boldsymbol{R}(\Pi_1)$.**

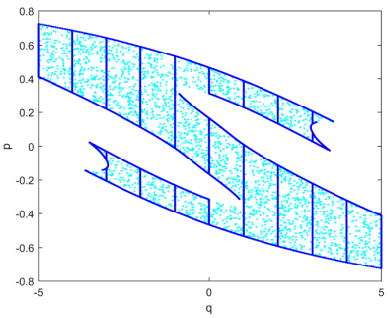
path $\Pi_1 = (1, 2, 3, 4)$. Indeed, every ray traced back hit the right curved line (line 3) and is split in two rays, at this point the reflectance $\mathcal{R}$ and transmittance $\mathcal{T}$ are calculated and, according to the first component of C_1, the transmitted ray is considered and traced back further with the corresponding power $\mathcal{T}$. This ray continues to propagate through the system until it hits either the detectors (line 5 and 6) or the curved line 2. In case it hit the detectors the procedure is stopped, otherwise the ray is split again into two more rays. Now, according to the second component of C_1, the transmitted ray is considered and it continues its trajectory reaching either the detectors or the source (line 1). For example, in Figure 9.6a we draw in red a ray that follows path $\Pi_1 = (1, 2, 3, 4)$. It is traced back from the target and has target PS coordinates $(q, p) = (0, -0.4)$. It arrives at the source with around the 55% of the initial power.

Using the inverse numerical ray mapping and taking into account the choice C_1 every time that the inverse ray tracing is applied and the rays hit a Fresnel line, the boundary $\partial R(\Pi_1)$ is found which is the blue line in Figure 9.6b. To verify whether this boundary is correct we trace around 10^4 rays using MC ray tracing imposing that at every interaction of the ray with lines 2 and 3 always the transmitted part of is considered. These rays are traced in red in Figure 9.6b. The picture shows that all the rays traced are inside the boundary $\partial R(\Pi_1)$ which is calculated correctly.

To detect the other paths we need to consider all the possible choice. We continue considering the choice $C_2 = (T, R, R, T)$. This means that once the ray is traced back from the target, if it hits the line 3, the transmitted part is considered according to the first component of C_2 and it is traced back further with its corresponding power given by $\mathcal{T}$. If the transmitted ray does not hit the detectors, it hits line 2 where it is split again. Now the reflected part of the ray is considered according to the second component of C_2. The power of the reflected ray is given by the reflectance $\mathcal{R}$ multiplied



(a) **Ray traced back from the target to the source**. The rays follows path $\Pi_2 = (1, 2, 3, 2, 3, 4)$ corresponding to the choice $C_2 = (T, R, R, T)$ from the target to the source. The percentage of power of the ray at the source is the 7.2% of the initial energy.

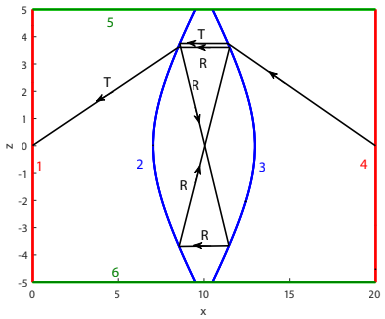


(b) **Boundary $\partial \boldsymbol{R}(\Pi_2)$.** The boundary is depicted with the blue line, the cyan rays are traced using MC ray tracing with 10^4 and considering two reflections between the lenses.

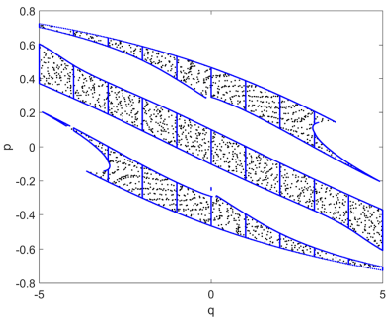
Figure 9.7: **Computation of the boundary $\partial \boldsymbol{R}(\Pi_2)$.**

by the power of the incident ray. The ray continues to propagates inside the system, if it hits line 3 it is reflected according to the third component of C_2 and, in case it hit line 2 next, it is finally transmitted according to the last component of C_2. If the ray reach finally the source, it follows the physical path $\Pi_2 = (1, 2, 3, 2, 3, 4)$. In Figure 9.7a we show in cyan a ray traced back from the target to the source that follows path Π_2. The ray with target PS coordinates $(q, p) = (0, -0.37)$ is traced back and after two reflections inside the lens it arrives at the source with the 7.2% of the initial energy. The numerical inverse ray mapping applied considering the choice C_2 every time that the inverse ray tracing is employed provides the boundary of the region $\partial R(\Pi_2)$ which is depicted in blue in Figure 9.7b. To properly detect the boundary we divided the target PS into $\mathrm{Ni} = 10$ bins as explained in Section 8.2. Likewise for the boundary $\partial R(\Pi_1)$, to prove that the method computes the boundary correctly we traced 10^4 rays using MC ray tracing and using the decision C_2 at every intersection with line 2 and 3. These rays are traced in cyan in Figure 9.7b. We observe that all the rays traced are located inside the blue line, therefore the boundary $\partial R(\Pi_2)$ is calculated correctly.

Since more than two reflection can occur between line 2 and 3, the procedure continues considering the choice $C_3 = (T, R, R, R, R, T)$ that leads to four reflections between line 2 and 3. Every ray traced back from the source is first transmitted, then reflected between 2 and 3 four times and finally transmitted again. If the ray hits the source then it follows the physical path $\Pi_3 = (1, 2, 3, 2, 3, 2, 3, 4)$ corresponding to the choice C_3 from the target to the source. An example of a ray that follows path Π_3 is depicted in black in Figure 9.8a. The ray is traced back from the target with target PS coordinates $(q, p) = (0, -0.39)$. It arrives at the source with a power equal to the 0.0073% of the initial energy. The inverse ray mapping with the choice C_3 gives the boundary $\partial R(\Pi_3)$ depicted in blue in Figure 9.8b. This boundaries is obtained



(a) **Ray traced back from the target to the source**. The rays follows path $\Pi_3 = (1, 2, 3, 2, 3, 2, 3, 4)$ corresponding to the choice $C_3 = (T, R, R, R, R, T)$ from the target to the source. The percentage of power of the ray at the source is the 0.0073% of the initial energy.



(b) **Boundary $\partial \boldsymbol{R}(\Pi_3)$.** The boundary is depicted with the blue line, the black rays are traced using MC ray tracing with 10^4 and considering four reflections between the lenses.

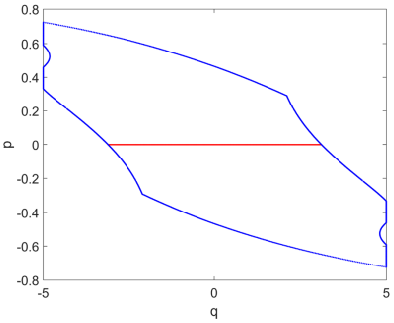
Figure 9.8: **Computation of the boundary $\partial \boldsymbol{R}(\Pi_3)$.**

dividing the target PS into Ni = 10 bins and applying the inverse ray mapping to each bin. The black dots correspond to the rays in target PS obtained using QMC ray tracing with 10^4 rays where at every intersection with the Fresnel lines either the reflected or the transmitted ray are selected according to choice C_3. Also in this case we found that the boundary computation is correct.

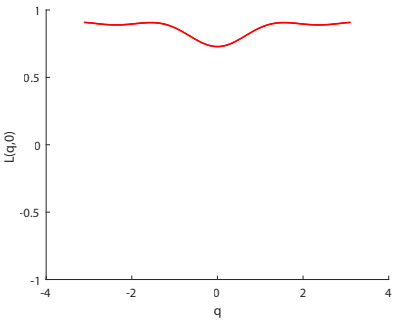
The inverse ray mapping is able to detect *all* the boundaries of the regions with positive luminance in target PS. Since rays with multiple reflections do not bring a significant amount of power, we decide do not to consider the rays with more than four reflections. The procedure can be stopped accordingly to the desired accuracy. The more reflections are considered, the better accuracy is obtained.

We remark now that the luminance at the target cannot be constant because every ray has a certain amount of energy that depends on the Fresnel coefficients. Because of this, an interpolation between the rays on the boundaries is needed to compute the profile of the luminance at the target. In Figure 9.9a we show the procedure to compute the luminance related to path Π_1 along direction $p = 0$. The rays corresponding to the coordinates $(q(\Pi_1, p), p)$ with $q \in [q^{\min}(\Pi_1, p), q^{\max}(\Pi_1, p)]$ and $p = 0$ are traced from the source to the target using the inverse ray tracing and taking into account of the choice C_1 at every intersection with the Fresnel lines. The rays traced are depicted in red. The luminance $L_{\Pi_1}(q, p)$ is calculated for every $q \in [q^{\min}(\Pi_1, p), q^{\max}(\Pi_1, p)]$ and $p = 0$. Its profile is depicted in Figure 9.9b.

Repeating the interpolation along all the possible directions $p \in [-1, 1]$, the luminance $L_{\Pi_1}(q, p)$ related to the rays that follow path Π_1 is found for every $p \in [-1, 1]$ and $q \in [q^{\min}(\Pi_1, p), q^{\max}(\Pi_1, p)]$. The same procedure is applied to all the paths considered and the luminance corresponding to each path is calculated. The total luminance $L(q, p)$ is given by the sum in Equation (9.2.7). Finally, the intensity is computed using Equation (9.2.8).



(a) Interpolation between the rays on $\partial R(\Pi_1)$ along direction $p = 0$.



(b) Luminance $L_{\Pi_1}(q, 0)$ related to path Π_1 along direction $p = 0$ and with $q \in [q^{\min}(\Pi_1, p), q^{\max}(\Pi_1, p)]$.

Figure 9.9: **Determination of the partial luminance.** $L_{\Pi_1}(q, 0)$ is related to path Π_1 along direction $p = 0$.

We do not show here the numerical results of the total luminance and intensity because some work is still in progress. Our expectation is that the inverse ray mapping is suitable also for systems with Fresnel reflection. Indeed, the results show that the boundaries of the regions with positive luminance are calculated correctly. Once the boundaries are found also the luminance and the intensity can be computed applying an interpolation between the rays on the boundaries along every possible direction. Furthermore we expect that the method is much more accurate and faster than both MC and QMC ray tracing. This is due to the fact that we can analyse every single path independently from the others.

9.4 Conclusion and outlook

In this chapter we introduced the numerical inverse ray mapping extended to optical systems where Fresnel reflection plays a role. We observed that, a ray emitted from the source can follow many different paths as it is split into two rays every time that it hits a Fresnel line. Thus, a unique point at the source PS can correspond to several points in target PS. This results in an overlap of the regions with positive luminance in target PS. The purpose of the method is to detect the boundaries of *all* these regions. This can be done fixing a priori which part of the ray has to be considered at every intersection of the ray with a Fresnel line. The inverse ray mapping explained in Chapter 8 is extended in this chapter such that only one boundary is computed at each run. The procedure is run as many times as the number of paths we want to consider. We presented the numerical results for a system formed by the source, the target and a Fresnel lens formed by two curved lines. We shown that the method is able to determine the boundaries of all the regions with positive luminance correctly. We noticed that including Fresnel reflections the luminance cannot be constant. An interpolation between the rays on the boundaries is needed to obtain the luminance profile. We provided an example of such interpolation for a given path and along a

certain direction. Furthermore, we explained the theory for the luminance and the intensity computation.

Future work might regard the calculation of the luminance and the intensity and a comparison with both MC and QMC ray tracing. Numerical results about the intensity computation should be provided.

Chapter 10

Conclusions

In this thesis we investigated new methods based on *phase space ray tracing for illumination optics*. The aim was to understand how light propagates through non imaging optical systems in order to calculate the target photometric variables, e.g. luminance and intensity. The core of this work is to use the *phase space* which provided a full description of geometric optics. Phase space consists of all the rays positions and directions. Thus, each ray is described by a unique point in phase space. In this thesis we restrict ourselves to two-dimensional optical systems the phase space of which is a two-dimensional space.

The phase space representation of the optical system shows that for every ray traced inside the system its path, that is the sequence of the optical lines that it encounters, can be considered. All the rays that follow the same path are located inside the same patches in phase space which is therefore divided into closed regions each of them is formed by all rays that follow the same path. The edge-ray principle guaranties that all the rays located at the boundaries of the regions in source phase space are located on the boundaries also in target phase space, [31]. our idea is therefore to determine the boundaries of those regions to obtain all the information needed to calculate the photometric variables. Based on this idea, we developed two methods: the phase space ray tracing and the inverse ray mapping in phase space. The purpose of both is to trace only the rays close to the boundaries reducing to the total number of rays traced compared to the exiting methods used for the intensity calculation as, for example, Monte Carlo and Quasi-Monte Carlo ray tracing.

Phase space ray tracing exploits the phase space of the source and the target of the optical system. In Chapter 4 we introduced a procedure to construct a triangulation on the source phase space which allows increasingly tracing rays close the boundaries and only few rays inside the regions with positive luminance. The boundaries of those regions were approximated using two different approaches: the α-shapes method and a technique based on the triangulation refinement. The first one relies on a parameter α which establishes which triangles have to be kept in the PS triangulation and which once have to be removed from it to approximate the boundaries correctly. In Chapter 5 numerical results are provided for two different kinds of TIR-collimators. They showed that phase space ray tracing using α-shapes is

much faster and more accurate than Monte Carlo ray tracing. We observed that the speed of convergence depends on the smoothness of the shape of the regions in target phase space and, therefore, on the optical system. To eliminate the parameter α from the calculation of the boundaries, we developed a new approach for the boundaries computation based on the triangulation refinement explained in Chapter 6. This technique is able to determine the boundaries triangles of the triangulation that is those that are crossed by a boundary. Connecting the vertices of the boundaries triangles corresponding to the rays that follow the same path, a good approximation of all the boundaries is obtained. Tracing more rays leads to construct smaller triangles resulting in a better accuracy of the boundaries computation. The method was applied to several optical systems with reflected and refractive optical lines. The results show that the boundaries of all the regions with positive luminance in target PS are calculated correctly even for complicated systems such as the parabolic reflector for which multiple reflection of the rays with the mirrors can occur. The intensity was computed assuming a Lambertian source. The intensity profile obtained using phase space ray tracing based on the triangulation refinement is compared to the two intensities found with Monte Carlo (MC) and Quasi-Monte Carlo (QMC) ray tracing. Significant advantages in terms of the accuracy and the computational time were observed with our method. Phase space ray tracing allows tracing far less rays compared to MC ray tracing resulting in a significant reduction of the computational time. Moreover, the results showed that for some systems, as for example the TIR-collimator, phase space ray tracing outperforms also QMC ray tracing and it has a behavior similar to QMC for more complicated systems as for the parabolic reflector. In order to further improve the phase space ray tracing we developed a second method that allows trace only the rays located exactly on the boundaries of the regions with positive luminance.

The inverse ray mapping employs the phase space of all the lines that form the system. For the lines that can receive and emit light two phase spaces are considered: the phase space of the line as a source (which contains the information of the incident rays) and the phase space of the line as a target (which contains the information of the reflected or refracted rays). For the source and the target one phase space was considered (the source and the target phase space, respectively). The idea was to construct an inverse map from the target to the source that connects the coordinates of the rays on the phase space of each optical line that it encounters.

In Chapter 7 we presented the analytic method for systems formed by straight line segments. We showed that the boundaries of the regions that form every phase space considered can be calculated *analytically*. Because of this, the inverse ray mapping reconstruction calculates the intensity *exactly*. Compared to QMC ray tracing the method is much more accurate (the exact intensity is found) and also faster. In Chapter 8 we introduced the numerical inverse ray mapping which is an extension of the analytic method to systems formed by curved lines. In this case the boundaries of all the phase spaces cannot be calculated analytically, therefore a bisection procedure combined with the inverse ray tracing is developed for determining the boundaries of the regions with positive luminance in phase space. As a consequence the *exact* target intensity cannot be provided for systems formed by curved line. Nevertheless, the method remains very accurate and numerical results show that it is able to detect numerical noise. Compared to QMC ray tracing the inverse ray mapping method res-

ults more accurate and faster also for complicated systems as the parabolic reflector.

Finally, in Chapter 9 we investigated systems were also Fresnel reflection was involved. Fresnel systems lead to multiple paths due to the fact that, at every interaction with a Fresnel line, each ray is splitted in two more rays (reflected and the transmitted). The inverse ray mapping is able to detect *all* the possible paths that can occur. Moreover, we shown that only the rays located on the boundaries of the regions in target phase space related to the physical paths are traced from the target to the source. To validate our method we traced forward a set of rays using MC ray tracing and we showed that the boundaries found with the inverse ray mapping encompass all the rays traced. The power of energy associated to each ray on the boundary is calculated. We explained how to compute the luminance for Fresnel systems where it is not constant.

In this thesis we showed that phase space is a powerful concept that fully characterize the optical systems. We presented two new methods based on phase space which allow traced far less rays than existing procedures to obtain the desired accuracy. This results in a significant reduction of the computational time. We evaluated the method for several optical systems in two-dimensions. It might be useful to investigate few more things in two-dimensions. Regarding phase space ray tracing could be interested to analyze systems with a non Lambertian source. Our insight is to calculate the boundaries as explained for a Lambertian source and then compute the profile of the luminance using an interpolation between the rays on the boundaries. The intensity can be obtained by merely integrating the luminance along all the possible positions. Regarding the numerical inverse ray mapping we are interested in provide simulations for calculating the intensity profile also for systems with Fresnel reflection. The results shown in Chapter 9 give the expectation that the inverse ray mapping method is suitable for systems with Fresnel reflection and that it is not only much more precise but also faster than both MC and QMC ray tracing. We believe that the two methods developed in two dimensions are ready to be extended to three dimensional systems. We believe that they can constitutes design tools for optical designers and that the time to design the optical systems would be greatly reduced. The numerical inverse ray mapping could be used also to detect and reduce ghost stray light.

Appendix A

Implementation of Sobol' sequences

A.1 Van der Corput sequences

In the following we show a particular construction of a low-discrepancy sequence for $d = 1$ that was introduced the first time by Van der Corput in 1935. This kind of sequences, called *van der Corput* sequences, are particular interesting not only because they give an intuition of how to construct low discrepancy sequences but also because many other kind of sequences in higher dimensions are based on this one-dimensional case. Before introducing these sequences we need to give the concept of radical inverse function. Let $\mathrm{b} \geq 2$ be an integer base. Any natural number $n \in \mathbb{N}_0$ can be decomposed in base b as follows:

$$n = \sum_{i=0}^{\infty} d_i \mathrm{b}^i \tag{A.1.1}$$

where $d_i \in \{0, 1, \cdots, \mathrm{b}-1\}$ are the digit numbers. The radical inverse function $\phi_{\mathrm{b}} : \mathbb{N}_0 \mapsto [0, 1)$ in base b is defined as:

$$\phi_{\mathrm{b}}(n) = \sum_{i=1}^{\infty} \frac{d_{i-1}}{\mathrm{b}^i}. \tag{A.1.2}$$

As an example we provide in the following the radical inverse function $\phi_{\mathrm{b}}(5)$ in base $\mathrm{b} = 2$. The digit expansion in base b of $n = 5$ is:

$$5 = 1 \cdot 2^0 + 1 \cdot 2^2. \tag{A.1.3}$$

Therefore, $d_0 = 1$, $d_1 = 0$ and $d_2 = 1$. The radical inverse function $\phi_2(5)$ is:

$$\phi_2(5) = \frac{1}{2} + \frac{1}{8} = \frac{5}{8}. \tag{A.1.4}$$

Definition A.1.1. The Van der Corput sequence in base b is defined as $\{\phi_{\mathrm{b}}(n)\}_{n \in \mathbb{N}_0}$.

For example, suppose we have the finite sequence of numbers $n \in \{0, 1, \cdots, 8\}$ the corresponding Van der Corput sequence $\{\phi_{\rm b}(n)\}_{n\in\{0,1,\cdots,8\}}$ in base $b = 2$ is:

$$\left\{\phi_2(n)\right\}_{n\in\{0,1,\cdots,8\}} = \left\{0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}\right\}. \tag{A.1.5}$$

It can be proved that the Van der Corput sequence in base b is uniformly distributed modulo one, [17]. The van der Corput sequence has been extended to higher dimensions. The most common QMC approach uses Sobol sequence which can be seen as an extended Van der Corput sequence in base $b = 2$. Sobol' sequence uses the same base $b = 2$ for all the dimensions $d \geq 2$.

A.2 Sobol' sequences

The aim is to generate a low-discrepancy sequence in the ipercube $[0, 1]^d$. Let us start from the simplest case of one dimension, i.e. $d = 1$. First, we need to chose a primitive polynomial P_j of degree s_j of the form

$$P_j : x^{\mathrm{s}_j} + a_{1,j}x^{\mathrm{s}_j-1} + \cdots + a_{\mathrm{s}_j-1}x + 1 \tag{A.2.1}$$

where the coefficients $\{a_{i,j}\}_{i=1,\cdots,\mathrm{s}_j-1}$ are either 0 or 1. Then a sequence $\{m_1, m_2, \cdots\}$ is defined such that:

$$m_{k,j} := 2a_{1,j}m_{k-1,j} \oplus 2^2 a_{2,j}m_{k-2,j} \oplus \cdots \oplus 2^{\mathrm{s}-1}a_{k-1,j}m_{k-\mathrm{s}+1,j} \oplus 2^{\mathrm{s}}m_{k-1,j} \oplus m_{k-\mathrm{s},j}, \tag{A.2.2}$$

where we have indicated with $\oplus$ the bit by bit exclusive or operator which operates on two bit patterns and operates on each pair of the corresponding bins giving as result 1 if one of the two bits is 1 and 0 if both bits are equal either to 0 or 1. The values $m_{k,j}$, $1 \leq k \leq d$, are chosen such that they are odd and positive numbers less than 2^k. Now, the so-called direction numbers are defined by:

$$v_{k,j} = \frac{m_{k,j}}{2^k}. \tag{A.2.3}$$

Then, the sequence $\{x_{i,j}\}$ is given by

$$x_{i,j} = i_1 v_1 \oplus i_2 v_2 \oplus \cdots \tag{A.2.4}$$

for every i, where i_k is the k-th digit from the right when i is written in binary $i = (\cdots i_3 i_2 i_1)_2$, [57]. We provide in the following an example.

Given the primitive polynomial $x^3 + x^2 + 1$ of degree $\mathrm{s}_j = 3$, the first three coefficients $m_{1,j} = 1$, $m_{2,j} = 3$, and $m_{3,j} = 7$ lead to the following direction numbers

$$v_{1,j} = \frac{1}{2}, \qquad v_{2,j} = \frac{3}{4}, \qquad v_{3,j} = \frac{7}{8}, \tag{A.2.5}$$

that in binary notation are:

$$v_{1,j} = (0.1)_2 \qquad v_{1,j} = (0.11)_2, \qquad v_{1,j} = (0.111)_2. \tag{A.2.6}$$

From Eq. (A.2.2) we can derive the others coefficients $m_{4,j} = 5$, $m_{5,j} = 7$, etc. with he corresponding direction vectors:

$$v_{4,j} = \frac{5}{16} = (0.0101)_2 \qquad v_{5,j} = \frac{7}{32} = (0.00111)_2 \tag{A.2.7}$$

From Eq. (A.2.4) we finally find the sequence

(A.2.8)

The generalization of Sobol's sequence to higher dimensions $d > 1$ is calculated considering a sequence where the i-th point has the form:

$$q_i = (x_{i,1}, x_{i,2} \cdots, x_{i,d}), \tag{A.2.9}$$

where the second index of the variables $x_{i,j}$ it refers to the polynomial P_j (with corresponding degree s_j) which is considered to calculate the direction numbers. Therefore, d different sets of direction numbers are generated from a given polynomial P_j using Eq. A.2.3 and each component $x_{i,j}$ is computed using the corresponding direction vector.

Appendix B

Calculation of the boundaries at the target PS

B.1 Analytical method to find the boundaries of the different regions in phase space

In this section, we present an analytical method to find the boundaries of the regions formed by rays that follow the same path. Furthermore, we will represent those regions on source and target phase space.

It is possible to determine the maximum number of times that a ray reflects into the two-faceted cup as follows. Rotating the entire cup we can think of the path as a straight line that hits one of the rotated targets. The idea to rotate the cup comes from the fact that in this way we consider the paths as straight lines, hence it is sufficient to find only one intersection point between the ray and one line segment (also in the case where we have more than one reflection) and finally rotate back the intersection point to find the point on the target. Next we want to explain this procedure in more detail. Our optical system is defined as in the previous section, see Figure 3.1. Let B be defined by:

$$\begin{aligned} B &= \left(h + \tfrac{a}{\tan(\gamma)}\right)\tfrac{1}{\cos(\gamma)} - \tfrac{a}{\tan(\gamma)} \\ &= \tfrac{h}{\cos(\gamma)} + a\tan\left(\tfrac{1}{2}\gamma\right), \end{aligned} \tag{B.1.1}$$

and $P : (0, B)$ is the rotation point. We define B_k as the clockwise ($k < 0$) or counterclockwise ($k \geq 0$) rotations of the point $P : (0, B)$ over an angle $\alpha_k = (2k+1)\gamma$, with γ the angle that the normal to the source forms with the reflectors of the cup and $k \in \mathbb{Z}$. The x and z-coordinates of B_k are indicated with $b_{k,x}$ and $b_{k,z}$, respectively, Figure $B.1$ is illustrative. The position vector for the points B_k is given by $\mathbf{b}_k = \begin{pmatrix} b_{k,x} \\ b_{k,z} \end{pmatrix}$ where

$$\mathbf{b}_k + \begin{pmatrix} 0 \\ \frac{a}{\tan(\gamma)} \end{pmatrix} = \begin{pmatrix} \cos(\alpha_k) & -\sin(\alpha_k) \\ \sin(\alpha_k) & \cos(\alpha_k) \end{pmatrix} \begin{pmatrix} 0 \\ B + \frac{a}{\tan(\gamma)} \end{pmatrix}. \tag{B.1.2}$$

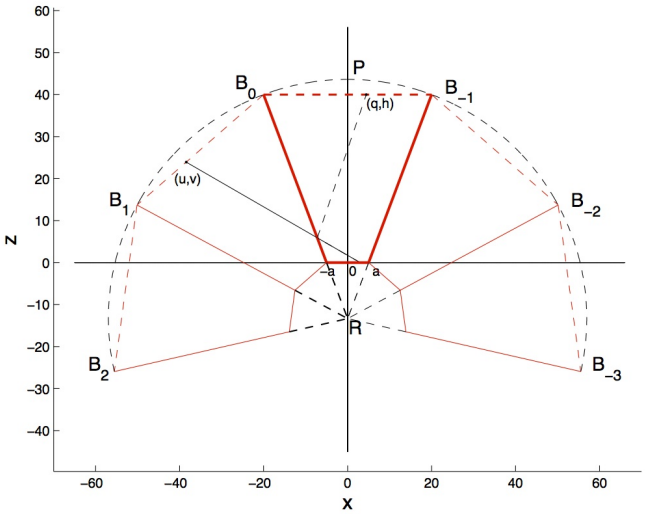


Figure B.1: The two-faceted cup rotated to both sides. The line segment $B_{k-1}B_k$ is the $|k|$ times rotated target. The point (u, v) of the intersection between a ray and the segment B_0B_1 corresponds to the point (q, h) on the target $B_{-1}B_0$. $P(0, B)$ is the point to rotate around the point $R = \left(0, -\frac{a}{\tan\gamma}\right)$. The length of the segments RB_k is equal to the radius of the dashed circle.

Then the maximum number of reflections r is:

$$r = \max\{k \in \mathbb{N} \mid b_{k-1,z} \geq 0\}\,. \tag{B.1.3}$$

This method of rotating the cup instead of reflecting the ray inside the system can also be applied to find the boundaries of the regions $M_{\mathrm{s},k}$ and $M_{\mathrm{t},k}$. In the following sections we will illustrate how this is done.

B.1.1 Source phase space

We observe that the set of rays that form the boundary of the regions $M_{\mathrm{s},k}$ only consists of rays that either leave the extremes of the source or hit one of the points B_k. In Figure $B.1$ is shown a ray that on the target phase space is located inside the region $M_{\mathrm{t},1}$, it does not constitute a point on any boundary. Furthermore, we note that the rays emitted from the corner points of the source form vertical lines in $\mathcal{P}_{\mathrm{s}}$, since $x = const$. On the other hand, rays that hit B_k form vertical lines in $\mathcal{P}_{\mathrm{t}}$, since $q = const$. Hence for the representation on the source phase space we have to choose rays that hit B_k, their directions are given by the relation

$$\tan t = \frac{x - b_{k,x}}{b_{k,z}}\,. \tag{B.1.4}$$

This is exactly what we did in the algorithm named *'Source'* (see Appendix **??** for details).

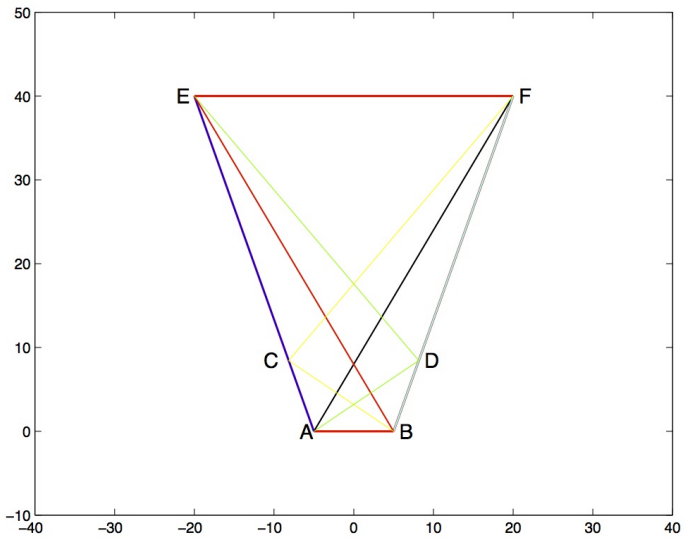


Figure B.2: Rays that leave the corner points of the source. The rays AF, BE, ACE, BDF are rays that do not hit the reflectors of the system. They constitute rays on the boundaries of the regions $M_{\mathrm{s},0}$, $M_{\mathrm{s},1}$ and $M_{\mathrm{s},-1}$. The rays ADE and BCF are rays that hit once the reflectors of the system. They constitute rays on the boundaries of the regions $M_{\mathrm{s},-1}$, $M_{\mathrm{s},-2}$, and $M_{\mathrm{s},1}$ or $M_{\mathrm{s},2}$, respectively.

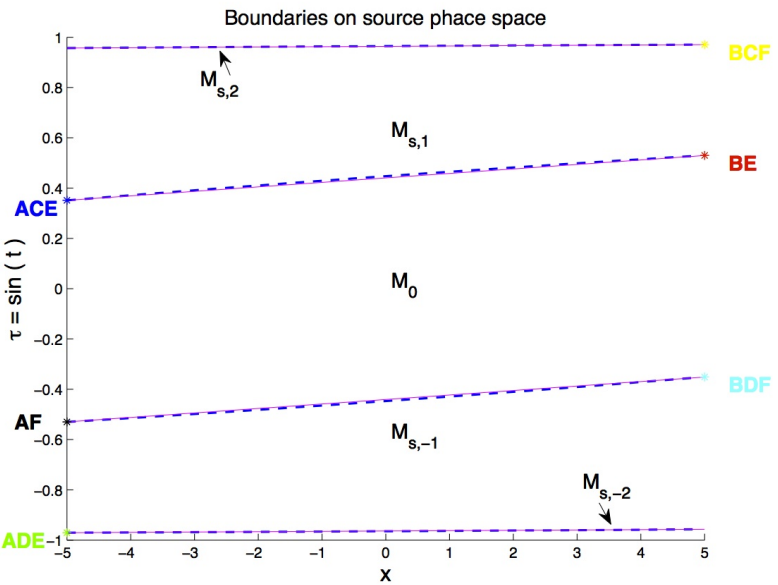


Figure B.3: Regions $M_{s,k}$ of rays that reflect $|k|$ times, with $(x,\tau) \in \mathcal{P}_s$. The parameter values are: $a = 5$, $b = 20$ and $h = 40$. The continuous lines are the boundaries of the regions M_s calculated considering rays that leave the source and hit the points B_k at the target. The dashed blue lines are the boundaries calculated using $(B.1.4)$

In Figure $B.2$ are shown some rays that compose the boundaries of $M_{s,k}$ which coordinates are:

$$ADE = \left(-a, \arctan\left(\frac{-a+b_{-1,x}}{b_{-1,z}}\right)\right),\ ACE = \big(-a, \sin(\gamma)\big),\ AF = \big(-a, -\sin(\delta)\big),$$
$$BCF = \left(a, \arctan\left(\frac{a-b_{1,x}}{b_{1,z}}\right)\right), BDF = \big(a, -\sin(\gamma)\big) \text{ and } BE = \big(a, \sin(\delta)\big).$$

The rays are represented by points in phase space. So we choose a proper number of rays that leave the source to obtain an accurate representation of the boundaries of $M_{s,k}$ regions. The final result is shown in Figure $B.3$. In addition, we derive the exact equation for the map $\mathcal{M}$. From equation $(B.1.4)$ we find the value of the angle for each ray at the source (depending on the ray position). Thus the boundaries are simply straight lines in the $(x, \tan(t))$-plane. The subdivision of phase space into regions is shown in Figure $B.3$, where we can also see the comparison between the two different methods to calculate the boundaries. Note that in this specific case the boundaries appear straight lines also in the $(x, sin(t))$-plane.

B.1.2 Target phase space

In this section we derive an exact expression for the map $\mathcal{M}$ in such a way that it is possible to determine the boundaries of the regions $M_{t,k}$ simply by finding the images

of some points on $\partial M_{\mathrm{s},k}$. Given a ray parameterization we are able to calculate the intersections point (u, v) between the ray and the line segment $B_{k-1}B_k$ as we did in *'Target'* (See Appendix **??** for the procedure). The corresponding point (q, h) on the target can be found by rotating or reflecting the point (u, v) back for k even or odd, respectively. Therefore we have the following expression for the point (q, h) on the target:

$$\begin{pmatrix} q \\ h \end{pmatrix} = \begin{pmatrix} (-1)^k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(-2k\gamma) & -\sin(-2k\gamma) \\ \sin(-2k\gamma) & \cos(-2k\gamma) \end{pmatrix} \begin{pmatrix} u \\ v + \frac{a}{\tan(\gamma)} \end{pmatrix} - \begin{pmatrix} 0 \\ \frac{a}{\tan(\gamma)} \end{pmatrix}. \tag{B.1.5}$$

We observe that the sign depends on the parity of k. When $k = 0$, i.e. the ray does not reflect, the first and the second matrices become the identity matrix and the cup is not rotated nor reflected. When k is even, the determinant of the product between the first and the second matrixes at the right hand of equation $(B.1.5)$ is equal to 1 and we obtained a rotation matrix, while when k is odd the determinant of the matrix given by the product between the first and the second matrix is equal to -1 and we have a reflection matrix. Also the angle on the target is calculated. It is an addition of an angle and a change of sign depending on k:

$$\theta = (-1)^k(t - 2k\gamma). \tag{B.1.6}$$

For every k, the mapping $(x, t) \mapsto (q, \theta)$ is now well determined and also the regions $M_{\mathrm{s},k}$ of rays that reflect k times are mapped to $M_{\mathrm{t},k}$. We observe that the lines shown if Figure $B.3$ are mapped to vertical lines in target phase space by the map $\mathcal{M}$ (see Figure B.4). Hence, to obtain the boundaries of the target,we will choose rays that are emitted from points close to the boundary of the source. According to what we said so far, the case of the target requires some good calculation to determine where a ray exits the cup. We can obtain those points analytically for a suitable number of rays, as we did in *'Target'*, and then we can drawn those points on the phase space as is shown in Figure $B.4$.
The coordinates of the rays traced in Figure $B.2$ at the target are given by:

$$ADE = \big(-b, -(t_1 + 2\gamma)\big),\ ACE = \big(-b, \sin(\gamma)\big),\ AF = \big(-b, -\sin(\delta)\big),$$
$$BCF = \big(b, -(t_2 - 2\gamma)\big), BDF = \big(b, -\sin(\gamma)\big) \text{ and } BE = \big(b, \sin(\delta)\big).$$

where $t_1 = \arctan(\frac{(}{-}a + b_{-1,x}b_{-1}))$ and $t_2 = \arctan(\frac{(}{a} - b_{-1,x}b_{-1}))$.

Figure $B.3$ and $B.4$ show also the symmetry of the regions $M_{\mathrm{s},k}$ and $M_{\mathrm{t},k}$. Finally we note that, since $k = 1$ is odd, the position of the regions $M_{\mathrm{t},1}$ and $M_{\mathrm{t},-1}$ are exchanged with respect to the position of $M_{\mathrm{s},1}$ and $M_{\mathrm{s},-1}$.

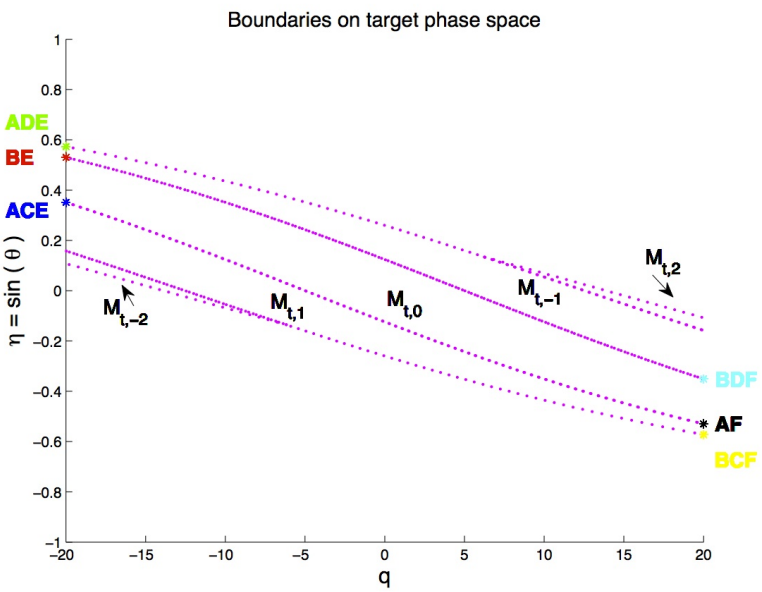


Figure B.4: Regions $M_{t,k}$ of rays that reflect $|k|$ times, for the two-faceted cup. The parameter values are: $a = 5$, $b = 20$ and $h = 40$.

Bibliography

[1] E. F. Zalewski, "Radiometry and photometry," *Handbook of optics*, vol. 2, pp. 24–1, 1995.

[2] J. Chaves, *Introduction to nonimaging optics*. CRC press, 2015.

[3] E. F. Schubert, T. Gessmann, and J. K. Kim, *Light emitting diodes*. Wiley Online Library, 2005.

[4] "Luminous efficacy-wikipedia the free encyclopedia," *https://commons.wikimedia.org/wiki/File:CIE 1931 Luminosity.png media/File:CIE 1931 Luminosity.png*.

[5] R. J. Koshel, *Illumination Engineering: design with nonimaging optics*. John Wiley & Sons, 2012.

[6] A. V. Arecchi, R. J. Koshel, and T. Messadi, "Field guide to illumination," SPIE, 2007.

[7] H. Zhu and P. Blackborow, "Etendue and optical throughput calculations," *Energetiq Technology, Inc., Woburn, MA*, 2011.

[8] R. P. Feynman, "Feynman lectures on physics. volume 2: Mainly electromagnetism and matter," *Reading, Ma.: Addison-Wesley, 1964, edited by Feynman, Richard P.; Leighton, Robert B.; Sands, Matthew*, 1964.

[9] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[10] E. Hecht, *Optics*. Parson Addison-Wesley, 2002.

[11] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, vol. 1. Basic books, 2011.

[12] P. H. Jones, O. M. Maragò, and G. Volpe, *Optical tweezers: Principles and applications*. Cambridge University Press, 2015.

[13] R. Winston, J. C. Miñano, and P. Benitez, *Nonimaging optics*. Academic Press, 2005.

[14] H. Gross, *Handbook of the Optical Systems*, vol. 1. Wiley-VCH, 2005.

[15] A. B. Owen, "Quasi-monte carlo sampling," *Monte Carlo Ray Tracing: Siggraph*, vol. 1, pp. 69–88, 2003.

[16] C. M. Grinstead and J. L. Snell, *Introduction to probability*. American Mathematical Soc., 2012.

[17] G. Leobacher and F. Pillichshammer, *Introduction to quasi-Monte Carlo integration and applications*. Springer, 2014.

[18] V. M. Zolotarev, *Modern theory of summation of random variables*. Walter de Gruyter, 1997.

[19] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*, vol. 10. John Wiley & Sons, 2016.

[20] D. M. Diez, C. D. Barr, and M. Cetinkaya-Rundel, *OpenIntro statistics*. CreateSpace, 2012.

[21] L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini, "A koksma–hlawka inequality for simplices," in *Trends in harmonic analysis*, pp. 33–46, Springer, 2013.

[22] A. B. Owen, "Multidimensional variation for quasi-monte carlo," in *International Conference on Statistics in honour of Professor Kai-Tai FangâĂŹs 65th birthday*, pp. 49–74, 2005.

[23] P. Bratley and B. L. Fox, "Algorithm 659: Implementing sobol's quasirandom sequence generator," *ACM Transactions on Mathematical Software (TOMS)*, vol. 14, no. 1, pp. 88–100, 1988.

[24] K. B. Wolf, *Geometric optics on phase space*. Springer Science & Business Media, 2004.

[25] W. Welford and R. Winston, "On the problem of ideal flux concentrators," *JOSA*, vol. 68, no. 4, pp. 531–534, 1978.

[26] P. Benitez, R. Mohedano, J. C. Miñano, R. Garcia, and J. C. Gonzalez, "Design of cpc-like reflectors within the simultaneous multiple surface design method," in *Proc. SPIE*, vol. 3139, pp. 19–28, 1997.

[27] "Edge ray-principle-wikipedia the free encyclopedia," *https://en.wikipedia.org/wiki/Nonimaging optics Edge ray principle*.

[28] J. C. Miñano, "Two-dimensional nonimaging concentrators with inhomogeneous media: a new look," *JOSA A*, vol. 2, no. 11, pp. 1826–1831, 1985.

[29] J. C. Miñano, "Design of three-dimensional nonimaging concentrators with inhomogeneous media," *JOSA A*, vol. 3, no. 9, pp. 1345–1353, 1986.

[30] P. Davies, "Edge-ray principle of nonimaging optics," *JOSA A*, vol. 11, no. 4, pp. 1256–1259, 1994.

[31] H. Ries and A. Rabl, "Edge-ray principle of nonimaging optics," *JOSA A*, vol. 11, no. 10, pp. 2627–2632, 1994.

[32] J. C. Miñano and J. C. Gonzalez, "New method of design of nonimaging concentrators," *Applied optics*, vol. 31, no. 16, pp. 3051–3060, 1992.

[33] J. Portegies and P. Lighting, "Fast ray tracing in phase space for optical design," 2013.

[34] B. Guo, J. Menon, and B. Willette, "Surface reconstruction using alpha shapes," in *Computer Graphics Forum*, vol. 16, pp. 177–190, Wiley Online Library, 1997.

[35] X. Xu and K. Harada, "Automatic surface reconstruction with alpha-shape method," *The visual computer*, vol. 19, no. 7, pp. 431–443, 2003.

[36] C. Filosa, J. ten Thije Boonkkamp, and W. IJzerman, "A new ray tracing method in phase space using alpha-shapes," 2015.

[37] A. Lucieer and M.-J. Kraak, "Alpha-shapes for visualizing irregular-shaped class clusters in 3 d feature space for classification of remotely sensed imagery," in *Proceedings of SPIE*, vol. 5295, pp. 201–211, 2004.

[38] "Stracciatella (ice cream)," *https://en.wikipedia.org/wiki/Stracciatella (ice cream)*.

[39] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on information theory*, vol. 29, no. 4, pp. 551–559, 1983.

[40] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Transactions on Graphics (TOG)*, vol. 13, no. 1, pp. 43–72, 1994.

[41] E. P. Mücke, "Shapes and implementations in three-dimensional geometry," 1993.

[42] E. L. Lloyd, "On triangulations of a set of points in the plane," in *Foundations of Computer Science, 1977., 18th Annual Symposium on*, pp. 228–240, IEEE, 1977.

[43] P. Knabner and L. Angermann, *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*. Springer Science & Business Media, 2003.

[44] D.-T. Lee and B. J. Schachter, "Two algorithms for constructing a delaunay triangulation," *International Journal of Computer & Information Sciences*, vol. 9, no. 3, pp. 219–242, 1980.

[45] R. J. Renka, "Algorithm 772: Stripack: Delaunay triangulation and voronoi diagram on the surface of a sphere," *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 3, pp. 416–434, 1997.

[46] W. H. Press, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[47] S. Fortune, "Voronoi diagrams and delaunay triangulations," *Computing in Euclidean geometry*, vol. 1, no. 193-233, p. 2, 1992.

[48] F. Cazals, J. Giesen, M. Pauly, and A. Zomorodian, "Conformal alpha shapes," in *Point-Based Graphics, 2005. Eurographics/IEEE VGTC Symposium Proceedings*, pp. 55–61, IEEE, 2005.

[49] K. Q. Brown, "Voronoi diagrams from convex hulls," *Information Processing Letters*, vol. 9, no. 5, pp. 223–228, 1979.

[50] "Delaunay voronoi.png," *https://commons.wikimedia.org/wiki/File:Delaunay Voronoi.png*.

[51] H. Edelsbrunner, "Alpha shapes - a survey," *Tessellations in the Sciences*, vol. 27, pp. 1–25, 2010.

[52] M. Teichmann and M. Capps, "Surface reconstruction with anisotropic density-scaled alpha shapes," in *Visualization'98. Proceedings*, pp. 67–72, IEEE, 1998.

[53] D. P. Mandal and C. Murthy, "Selection of alpha for alpha-hull in $\mathbb{R}^2$," *Pattern Recognition*, vol. 30, no. 10, pp. 1759–1767, 1997.

[54] K. Moore, "Methods and associated systems for simulating illumination patterns," Dec.8 8 Dec. 2015. US Patent 9,208,603.

[55] R. P. Breault, "Control of stray light," *Handbook of Optics*, vol. 1, pp. 38–1, 1995.

[56] S. Grabarnik, "Optical design method for minimization of ghost stray light intensity," *Applied optics*, vol. 54, no. 10, pp. 3083–3089, 2015.

[57] S. Joe and F. Y. Kuo, "Notes on generating sobol sequences," *ACM Transactions on Mathematical Software (TOMS), 29 (1), 49*, vol. 57, 2008.

Description of the research

In this thesis we studied the light propagation within optical systems. Optical engineers are interested in design systems in such a way the desired output distribution is obtained. The goal in illumination optics is to obtain the desired output distribution of light. To this purpose the ray tracing procedure is widely used. Ray tracing is a forward method where a set of rays is traced within the system from the source to the target. The propagation of light is determined computing the position and the direction of every ray for all the optical surfaces that it encounters. There are many ways to implement the ray tracing process. Monte Carlo (MC) ray tracing is often used in non-imaging optics. Rays are randomly traced from the source to the target and each time that a ray hits an optical surface the coordinates of the intersection point of the ray with the surface and the new ray direction are calculated. The output variables are computed dividing the target into intervals, the so-called bins, and counting the rays that fall into each bin. To obtain the desired accuracy, millions of rays are required, therefore the method is extremely computationally expensive and it converges as the inverse of the square root of the number of rays traced.

MC ray tracing can be improved using as sample of points a low discrepancy sequence instead of random points. Discrepancy can be interpreted as a measure of how much the sample distribution differs from a uniformly distributed sample. The discrepancy is therefore zero for uniformly distributed points. A low discrepancy sequence gives a sample of points which are regularly distributed but not exactly uniformly distributed. Quasi Monte Carlo (QMC) method considers these kind of sequences as sample of points. Therefore, QMC ray tracing is implemented tracing a set of rays whose position and direction are given by the coordinates of a low discrepancy sequence of points. The main advantage of QMC method is its rate of convergence, it is faster then MC for low dimensional problems. Nevertheless, it has some disadvantages. First, it is not easy to give an error estimation for QMC method. Second, for high dimensional spaces the QMC can become very slow. Third, it is still a binning procedure. Hence, the accuracy depends both on the number of rays traced and on the umber of bins.

In order to improve the existing methods, the phase space (PS) of the optical system is considered in this thesis. The PS of an optical surface gives information about the position and the direction of every ray on that surface where the direction is expressed with respect to the normal of the surface. In PS, the ray's direction is given by the sine of the angle that the ray forms with respect to the normal of the surface multiplied by the index of refraction of the medium in which the ray is located. In two dimensions, the PS is a two-dimensional space where the coordinates of every

ray are specified by one position coordinate and one angular coordinate. For three dimensional systems the PS is a four dimensional space because every ray is specified by two position and two angular coordinates. Our idea is to use the structure of PS to trace only the rays close to the discontinuities of the luminance at the target PS. Two new approaches based on PS are presented in this work. They are tested for two-dimensional systems.

The first method is called ray tracing on PS and it is based on the source and the target PS representation of the optical system. It takes into account the sequence of optical lines that each ray hits when it propagates inside the system, that is the ray path. We note that the source and target phase spaces are partitioned into different regions each of them is formed by the rays that follow the same path. The idea is to use the edge-ray principle proved by Ries and Rabl (1994) which states that the area of these regions is conserved: all rays that are neighbors at the source PS remain close to each other at the target PS. To this purpose, a nonuniform triangulation of the source PS is constructed in such a way that new triangles are added to the triangulation only where boundaries occur. Assuming constant brightness, we only need to compute the boundaries of the regions in target PS to obtain the output photometric variables. We test the method for optical systems where both reflection and refraction laws are involved. Numerical results show that ray tracing on PS is faster and more accurate compared to MC ray tracing.

The second method employs not only the source and the target PS, but also the PS of *all* the other lines that constitute the system. All lines can be modeled as detectors of the incident light and emitters of the reflected light. Moreover, we assume that the source can only emit light and the target can only receive light. Therefore, one PS is taken into account for the source and one for the target. For the other surfaces both the source and target PS are considered. Furthermore, instead of starting from the source, the new method starts tracing back rays from target PS. In order to determine the coordinates of these rays, an inverse map from the target to the source PS is constructed as a concatenation of the maps that relate the PS of two different lines. Employing this map we are able to detect the rays that in target PS are located on the boundaries of the regions with positive luminance. First, we implement the method for systems formed by straight and reflective lines. In this particular case, the boundaries of the regions that form every PS can be computed analytically. This allows us to obtain an analytic target intensity distribution. The results are shown for a two-faceted cup and a multi-faceted cup. In both cases we note significant advantages both in terms of the accuracy and the computational time. Second, the method is developed for systems formed by curved lines. In this case the boundaries cannot be determined analytically and therefore a numerical procedure is involved. In particular, we apply a bisection method on target PS. Also in this case we compare our method to MC ray tracing and we observe significant advantages using the PS method. Finally, the ray mapping method in PS is applied to systems where also Fresnel reflection is taken into account. We obtain relevant results also in the last case.

Curriculum Vitae

Carmela Filosa was born on November 28, 1985 in Torre del greco, Italy. She finished the high school in 2003 at Liceo Scientifico Statale "G. Marconi", Colleferro. She obtained a bachelor (2008) and Master (2013) degree in Mathematics at the University of Rome "La Sapienza", Italy. In March 2014, she moved in Eindhoven (the Netherlands) to start a PhD project at the Eindhoven University of Technology in the department of Mathematics and Computer Science. The PhD project was under the supervision of Wilbert IJzerman and Jan ten Thije Boonkkamp. The research conducted in her doctoral studies was funded by Technologiestichting STW and, the daily work took place at the Centre for Analysis, Scientific computing and Applications (CASA) of TU/e and at the department of Philips Lighting of the High Tech Campus in Eindhoven. The results of her research are presented in this thesis.

Acknowledgments