Geometrical and Technical Optics

Lecture about the principles of geometrical and technical optics

Norbert Lindlein

Institute of Optics, Information and Photonics
University of Erlangen–Nürnberg,
Staudtstr. 7/B2, D–91058 Erlangen
norbert.lindlein@physik.uni-erlangen.de

First edition: June 2006
Extended edition: October 2007
Introduction

This lecture shall discuss the basics and the applications of geometrical optical methods in modern optics. Geometrical optics has a long tradition and some ideas are many centuries old. Nevertheless, the invention of modern personal computers which can perform several millions of floating-point operations in a second revolutionized also the methods of geometrical optics and so several analytical methods lost importance whereas numerical methods such as ray tracing became very important. Therefore, the emphasis in this lecture is also on modern numerical methods such as ray tracing and some other systematic methods such as the paraxial matrix theory.

We will start with a chapter showing the transition from wave optics to geometrical optics and the resulting limitations of the validity of geometrical optics. Then, the paraxial matrix theory will be used to introduce the traditional parameters such as the focal length and the principal points of an imaging optical system. Also, an extension of the paraxial matrix theory to optical systems with non-centered elements will be briefly discussed. After a chapter about stops and pupils the next chapter will treat ray tracing and several extensions to analyze imaging and non–imaging optical systems. A chapter about aberrations of optical systems will give a more vivid insight into this matter than a systematic treatment. Nearly at the end, a chapter about the most important optical elements/instruments generally described with geometrical optics will be given. These are amongst others the diffractive lens, the achromatic lens, the camera, the human eye, the telescope and the microscope. The final chapter will treat the basic concept of radiometry and photometry which is important in optical illumination systems.

For more information about the basics of geometrical optics we refer to text books such as [1],[8],[13],[14],[31],[32],[34],[35].
Notes to this lecture script

The lecture **Geometrical and Technical Optics** (Grundkurs Optik I: Geometrische und Technische Optik) is the first course in optics at the University of Erlangen–Nürnberg. It will be followed by the second course about **Wave Optics** (Grundkurs Optik II: Wellenoptik). So, only basic knowledge of optics like it is given in the introductory physics lectures is needed or at least useful to understand this lecture. Besides this, basic knowledge of electromagnetism is very useful. In mathematics, basic knowledge of analysis, vector calculus, and linear algebra are expected. So, in general this lecture should be attended during the advanced study period after having passed the "Vordiplom" (for diploma students) or at the end of the bachelor phase in the 5. semester or at the beginning of the master phase (for bachelor/master students).

The lecture itself has two hours per week accompanied by an exercise course of also two hours per week. The exercises are performed partly as analytic calculations and partly as direct computer training using optical simulation/design software. In order to get a certificate the lecture and the exercises have to be attended on a regular base and it is expected that every student performs from time to time one of the exercises at the blackboard or on the computer. At the end, there will be a written examination to get the certificate.
## Contents

1 The basics and limitations of geometrical optics  
1.1 The eikonal equation ............................................. 1  
1.2 The orthogonality condition of geometrical optics ................. 4  
1.3 The ray equation ................................................... 4  
1.4 Limitations of the eikonal equation ............................... 5  
1.5 Energy conservation in geometrical optics .......................... 8  
1.6 Law of refraction ................................................... 9  
1.7 Law of reflection ................................................... 10

2 Paraxial geometrical optics  
2.1 Paraxial rays in homogeneous materials ............................ 12  
2.1.1 Some basic definitions ........................................ 12  
2.1.2 Optical imaging ................................................. 13  
2.1.3 A note to the validity of the paraxial approximation ............ 15  
2.1.4 Definition of a paraxial ray .................................... 15  
2.1.5 Transfer equation ............................................... 16  
2.2 Refraction in the paraxial case .................................... 17  
2.2.1 Paraxial law of refraction ..................................... 17  
2.2.2 Refraction at a plane surface .................................. 17  
2.2.3 Refraction at a plane parallel plate ............................. 17  
2.2.4 Some notes to sign conventions ................................. 19  
2.2.5 Refraction at a spherical surface ............................... 20  
2.3 The cardinal points of an optical system .......................... 21  
2.3.1 The principal points ........................................... 22  
2.3.2 The nodal points ............................................... 22  
2.3.3 The focal points ................................................ 22  
2.3.4 Calculation of the cardinal points of a general optical system ... 23  
2.3.4.1 Principal points ........................................... 24  
2.3.4.2 Nodal points ............................................... 25  
2.3.4.3 Focal points and focal lengths ............................... 25  
2.3.5 Relation between the focal lengths in object and image space .... 26  
2.3.6 The cardinal points of an optical system with identical surrounding refractive indices .............................. 27  
2.3.7 The cardinal points of a spherical refracting surface ............ 28  
2.4 The imaging equations of geometrical optics ....................... 29
2.4.1 The "lens equation" ............................................. 29
2.4.2 Newton equation ............................................. 31
2.4.3 Relation between lateral and longitudinal magnification .......................... 32
2.4.4 Some notes to the graphical image and ray construction in paraxial optics 32
2.4.5 The Smith–Helmholtz invariant ................................ 34

2.5 The thin lens .................................................. 35

2.6 The thick lens .................................................. 37
2.6.1 Thick lens in air ............................................. 39
2.6.2 Special cases of thick lenses in air ................................ 39
2.6.2.1 Ball lens .................................................. 39
2.6.2.2 The meniscus lens of Hoegh .................................. 40
2.6.2.3 Plane–convex or plane–concave lenses .................. 41

2.7 Reflecting optical surfaces .................................... 41
2.7.1 A plane reflecting surface ................................... 42
2.7.2 A spherical reflecting surface ................................ 42

2.8 Extension of the paraxial matrix theory to 3x3 matrices .................. 44
2.8.1 Paraxial ray tracing at a diffraction grating ................... 44
2.8.2 Tilted refractive plane surface ............................ 46
2.8.3 Thin prism .................................................. 47
2.8.4 The transformation matrices .................................. 47

3 Stops and pupils ................................................. 51
3.1 The aperture stop ............................................. 51
3.2 The field stop ................................................. 55

4 Ray tracing ...................................................... 56
4.1 Principle ..................................................... 56
4.2 Mathematical description of a ray ................................ 58
4.3 Determination of the point of intersection with a surface .................. 59
4.3.1 Plane surface ............................................. 59
4.3.2 Spherical surface .......................................... 60
4.3.3 General surface $z=f(x,y)$ .................................. 61
4.3.4 Coordinate transformation ................................... 61
4.4 Calculation of the optical path length .......................... 63
4.5 Determination of the surface normal ................................ 63
4.5.1 Plane surface ............................................. 63
4.5.2 Spherical surface .......................................... 63
4.5.3 General surface $z=f(x,y)$ .................................. 63
4.6 Law of refraction .............................................. 64
4.7 Law of reflection .............................................. 65
4.8 Non–sequential ray tracing and other types of ray tracing .............. 66

5 Aberrations ...................................................... 68
5.1 Calculation of the wave aberrations ................................ 68
5.2 The ray aberrations and the spot diagram ......................... 70
5.3 The Seidel terms and the Zernike polynomials ..................... 71
5.3.1 Spherical aberration ............................................. 72
5.3.2 Coma ................................................................. 72
5.3.3 Astigmatism .......................................................... 73
5.3.4 Curvature of field .................................................... 74
5.3.5 Distortion .............................................................. 74
5.3.6 The Zernike polynomials ........................................... 75
5.4 Chromatic aberrations ................................................ 75

6 Some important optical instruments .................................. 77
6.1 The diffractive lens .................................................... 77
   6.1.1 Formation of a Fresnel zone lens from a full lens ........ 77
   6.1.2 Grating model of a Fresnel zone lens ....................... 78
   6.1.3 Short glossary of DOEs ........................................ 80
   6.1.4 Phase function of a diffractive optical element ........... 82
   6.1.5 Example of designing the phase function ................... 83
6.2 The aplanatic meniscus .............................................. 84
   6.2.1 The aplanatic points of a sphere ............................. 84
   6.2.2 Properties of an aplanatic meniscus ......................... 85
6.3 The achromatic lens .................................................. 86
   6.3.1 Examples of designing achromatic doublets ............... 88
6.4 The camera ........................................................... 90
   6.4.1 The depth of field .............................................. 91
6.5 The human eye ....................................................... 95
6.6 The Telescope ........................................................ 96
   6.6.1 Telescope as beam expander and imaging system for far–distant objects 98
   6.6.2 Imaging property of a telescope for finite distant objects .... 98
   6.6.3 The astronomical and the Galilean telescope ............ 100
       6.6.3.1 Astronomical telescope .................................. 100
       6.6.3.2 Galilean telescope ...................................... 102
6.7 The Microscope ..................................................... 102
   6.7.1 The magnifier ................................................... 102
   6.7.2 The two–stage microscope .................................... 103

7 Radiometry and Photometry ........................................... 106
7.1 Definition of radiometric and photometric parameters ......... 106
   7.1.1 Radiometric parameters .................................... 106
   7.1.2 Photometric parameters (related to human eye) ....... 107
   7.1.3 Some illustrating examples .................................. 108
7.2 Imaging of light sources ............................................ 109
   7.2.1 Small (point–like) light source ............................ 109
   7.2.2 Behavior of the luminous intensity in the case of imaging 110
   7.2.3 Practical example: Slide–projector .......................... 111
7.3 Transition from a point source to an extended light source .... 111
   7.3.1 Radiator types according to Straubel ..................... 111
   7.3.2 Luminous flux in a 2π solid angle ........................ 113
   7.3.3 Illuminance on a plane screen for free space propagation 113
7.3.4 Behavior of the radiator types in the case of imaging .......................... 114
7.3.5 Illuminance in the image ........................................................................ 116
7.3.6 Difference between the imaging of an extended (sun) and a point–like 
(star) light source ...................................................................................... 118
7.3.6.1 Contrast improvement in a telescope .............................................. 119
7.3.7 Imaging with a telescope ...................................................................... 119
7.4 Photometric units ..................................................................................... 121
7.4.1 The candela ......................................................................................... 121
7.4.2 Luminance $\bar{L}$ of the black body radiation at 2045 K ........................ 122
7.4.3 Luminous flux $\Phi$ ............................................................................. 122
7.5 Generalization of the radiation formulas .................................................. 122
7.5.1 Applications ......................................................................................... 123
7.5.1.1 Surface elements within a sphere ................................................. 123
7.5.1.2 Decrease of the illuminance by imaging a screen ........................ 124
7.6 Invariance of the luminance by imaging ................................................ 125
7.6.1 More general treatment with different refractive indices ..................... 126
7.7 Etendue .................................................................................................. 127
7.8 Scatterer plate in the ray path .................................................................. 128
## List of Figures

1.1 Optical ray .................................................. 4  
1.2 Curved light ray in a general (inhomogeneous) material ................. 5  
1.3 Light tube .................................................... 8  
1.4 Snell’s law ................................................... 9  

2.1 Optical axis of a lens ......................................... 12  
2.2 Meridional plane of an optical system .......................... 13  
2.3 Schematic display of three different situations in optical imaging ......... 14  
2.4 Paraxial ray tracing between two parallel planes ....................... 16  
2.5 Paraxial refraction at a plane surface ................................ 17  
2.6 Application of the paraxial matrix theory at a plane parallel plate ........ 18  
2.7 Sign conventions in the paraxial matrix theory ........................ 19  
2.8 Parameters for the refraction at a spherical surface ..................... 20  
2.9 Cardinal points of an optical system ................................ 21  
2.10 Scheme for calculating the cardinal points of a general optical system consisting of refractive surfaces ........................................ 24  
2.11 Cardinal points and parameters of a spherical refracting surface ........ 28  
2.12 Parameters for explaining the imaging of an object point to an image point by a general optical system ........................................ 30  
2.13 Graphical explanation of the Newton equation .......................... 32  
2.14 Graphical construction of the image ................................ 33  
2.15 Graphical construction of a ray path ................................ 33  
2.16 Parameters of a thin lens ....................................... 35  
2.17 Different types of lenses ....................................... 37  
2.18 Parameters of a thick lens ...................................... 37  
2.19 The principal planes of a thick lens ................................ 38  
2.20 Parameters of a ball lens ....................................... 39  
2.21 The meniscus of Hoegh .......................................... 40  
2.22 Principal planes of a plane–convex or plane–concave lens ............... 41  
2.23 Reflection at a plane surface .................................... 42  
2.24 Reflection at a spherical surface .................................. 43  
2.25 Cardinal points of a convex mirror ................................ 44  
2.26 Diffraction at a grating ........................................ 45  
2.27 The refraction at a tilted plane surface ................................ 46  
2.28 Refraction at a thin prism ........................................ 47
2.29 Paraxial transformation between two relatively tilted and shifted coordinate systems 48
3.1 Aperture stop, entrance and exit pupil of an optical system 52
3.2 Imaging with a single lens and a telecentric system 53
4.1 Propagation of some light rays in a typical microscopic objective 57
4.2 Mathematical representation of a light ray 58
4.3 Mathematical description of a plane surface 59
4.4 Mathematical representation of a spherical surface 60
4.5 Coordinate transformation of a ray 62
4.6 Parameters for refraction of a ray at a surface 64
4.7 Parameters for the reflection of a ray at a surface 65
5.1 Explanation of the wave aberrations and the ray aberrations 69
5.2 Wave aberrations for the on-axis point of the microscopic objective of figure 4.1 70
5.3 Spot diagram for the on-axis point of the microscopic objective of fig. 4.1 71
5.4 Wave aberrations for the on-axis object point of the misadjusted microscopic objective of figure 4.1 72
5.5 Wave aberrations for an off-axis object point of the microscopic objective of figure 4.1 73
5.6 Curvature of field 74
5.7 Effect of distortion 75
6.1 Fresnel zone lens 78
6.2 Explanation of the grating equation 79
6.3 Different implementation types of holograms 80
6.4 Example for designing a DOE 83
6.5 Aplanatic points of a sphere 84
6.6 Aplanatic meniscus lens 85
6.7 Schemes of a refractive and of a hybrid achromatic doublet 87
6.8 The focal length of different achromatic doublets as function of the wavelength 89
6.9 The focal length of two single lenses as function of the wavelength 90
6.10 Principle of a camera 91
6.11 Calculation of the depth of field for the case of a camera with a thin lens 92
6.12 Ray tracing of an eye model as it is used in ophthalmology 95
6.13 Principle set-ups of telescopes 97
6.14 Parameters to calculate the paraxial matrix for imaging with the help of a telescope 99
6.15 The aperture stop and the field stop for imaging with an astronomical telescope 100
6.16 Principle of a magnifier 103
6.17 Principle of a microscope 104
7.1 Inverse-square law of photometry 109
7.2 Imaging of the luminous intensity 110
7.3 Principle of a slide-projector 111
7.4 Luminous intensity of different radiator types 112
7.5 Illuminance on a plane screen for free space propagation 114
7.6 Imaging of the different radiator types 115
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.7</td>
<td>Illuminance in the image</td>
<td>116</td>
</tr>
<tr>
<td>7.8</td>
<td>Scheme of an astronomical telescope/camera</td>
<td>118</td>
</tr>
<tr>
<td>7.9</td>
<td>Imaging with a telescope</td>
<td>120</td>
</tr>
<tr>
<td>7.10</td>
<td>Illustration of the general radiation formula</td>
<td>122</td>
</tr>
<tr>
<td>7.11</td>
<td>Radiation within a sphere</td>
<td>123</td>
</tr>
<tr>
<td>7.12</td>
<td>Imaging of a screen by a lens</td>
<td>124</td>
</tr>
<tr>
<td>7.13</td>
<td>Invariance of luminance by imaging</td>
<td>125</td>
</tr>
<tr>
<td>7.14</td>
<td>Illustration of the etendue</td>
<td>127</td>
</tr>
<tr>
<td>7.15</td>
<td>Luminance of a scatterer plate</td>
<td>128</td>
</tr>
</tbody>
</table>
Chapter 1

The basics and limitations of geometrical optics

1.1 The eikonal equation

Geometrical optics is normally defined to be the limiting case of wave optics for very small wavelengths $\lambda \to 0$. In fact it is well–known that electromagnetic waves with a large wavelength $\lambda$ such as radio waves cannot generally be treated with geometrical optical methods. X–rays and gamma radiation on the other hand propagate nearly like rays. They can generally be described quite well with geometrical optical methods provided the size of the optical elements (especially stops) is at least several hundred wavelengths. The accuracy of a geometrical optical calculation increases if the size of the optical element increases compared to the wavelength of the used light.

The basic equations of geometrical optics [1],[34] are derived directly from Maxwell equations. The restriction here is that only linear and isotropic materials are considered. Additionally, the electric charge density $\rho$ is assumed to be zero.

In this case the four Maxwell equations are:

\begin{align}
\nabla \times \mathbf{E}(r,t) &= - \frac{\partial \mathbf{B}(r,t)}{\partial t} & (1.1.1) \\
\nabla \times \mathbf{H}(r,t) &= \frac{\partial \mathbf{D}(r,t)}{\partial t} + j(r,t) & (1.1.2) \\
\nabla \cdot \mathbf{B}(r,t) &= 0 & (1.1.3) \\
\nabla \cdot \mathbf{D}(r,t) &= 0 & (1.1.4)
\end{align}

where the following quantities of the electromagnetic field are involved:

- electric vector $\mathbf{E}$ (dt.: Vektor der elektrischen Feldstärke)
- magnetic vector $\mathbf{H}$ (dt.: Vektor der magnetischen Feldstärke)
- electric displacement $\mathbf{D}$ (dt.: elektrische Verschiebungsichte)
- magnetic induction $\mathbf{B}$ (dt.: magnetische Induktion/Flußichte)
- electric current density $j$ (dt.: elektrische Stromdichte)
The arguments illustrate that all quantities are in the general case functions of the spatial coordinates $x, y, z$ with position vector $\mathbf{r} = (x, y, z)$ and of the time $t$. The so called Nabla operator

$$\nabla = \left( \begin{array}{c} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{array} \right)$$

(1.1.5)

is used and the symbol ”$\times$” indicates the vector product of two vectors whereas ”$\cdot$” indicates the scalar product of two vectors. It should be mentioned that the terms of above using Nabla operators are also well–known in some text books with some other names:

- $\nabla \times \mathbf{f} = \text{rot} \mathbf{f} = \text{curl} \mathbf{f}$ is the so called curl (dt.: Rotation) of a vector function $\mathbf{f}$
- $\nabla \cdot \mathbf{f} = \text{div} \mathbf{f}$ is the so called divergence (dt.: Divergenz) of a vector function $\mathbf{f}$
- $\nabla \varphi = \text{grad} \varphi$ is the gradient (dt.: Gradient) of a scalar function $\varphi$

The material equations in the case of linear and isotropic materials link the electromagnetic quantities with each other:

$$D(\mathbf{r}, t) = \varepsilon(\mathbf{r}) \varepsilon_0 E(\mathbf{r}, t)$$

(1.1.6)

$$B(\mathbf{r}, t) = \mu(\mathbf{r}) \mu_0 H(\mathbf{r}, t)$$

(1.1.7)

$$j(\mathbf{r}, t) = \sigma(\mathbf{r}) E(\mathbf{r}, t)$$

(1.1.8)

The function $\varepsilon$ is the dielectric function, $\mu$ is the magnetic permeability and $\sigma$ is the specific conductivity. The constants $\varepsilon_0$ and $\mu_0$ are the dielectric constant of vacuum and the magnetic permeability of vacuum, respectively. A quite general approach for stationary monochromatic waves is used to describe the electric and the magnetic field:

$$E(\mathbf{r}, t) = e(\mathbf{r}) e^{ik_0 L(\mathbf{r})} e^{-i\omega t}$$

(1.1.9)

$$H(\mathbf{r}, t) = h(\mathbf{r}) e^{i k_0 L(\mathbf{r})} e^{-i\omega t}$$

(1.1.10)

The real function $L$ is the optical path length (dt.: optische Weglänge) and the vectors $\mathbf{e}$ and $\mathbf{h}$ are in the general case complex valued to be able to represent all polarization states. The surfaces with constant optical path length $L$ are the wave fronts and the term $\Phi(\mathbf{r}) = k_0 L(\mathbf{r})$ is the phase of the wave. $\mathbf{e}$ and $\mathbf{h}$ are slowly varying functions of the position $\mathbf{r}$ whereas the term $\exp(ik_0 L)$ varies rapidly because the constant $k_0$ is defined as $k_0 = 2\pi/\lambda$ with the vacuum wavelength $\lambda$. The angular frequency $\omega$ of the wave is linked to $\lambda$ by $\omega = 2\pi c/\lambda = ck_0$, where $c$ is the speed of light in vacuum.

By applying these equations to the Maxwell equations the so called time–independent Maxwell equations result:

$$\nabla \times \left( e(\mathbf{r}) e^{i k_0 L(\mathbf{r})} \right) = i\omega \mu(\mathbf{r}) \mu_0 h(\mathbf{r}) e^{i k_0 L(\mathbf{r})}$$

(1.1.11)

$$\nabla \times \left( h(\mathbf{r}) e^{i k_0 L(\mathbf{r})} \right) = [-i\omega e(\mathbf{r}) \varepsilon_0 + \sigma(\mathbf{r})] e(\mathbf{r}) e^{i k_0 L(\mathbf{r})}$$

(1.1.12)

$$\nabla \cdot \left( \mu(\mathbf{r}) \mu_0 h(\mathbf{r}) e^{i k_0 L(\mathbf{r})} \right) = 0$$

(1.1.13)

$$\nabla \cdot \left( \varepsilon(\mathbf{r}) \varepsilon_0 e(\mathbf{r}) e^{i k_0 L(\mathbf{r})} \right) = 0$$

(1.1.14)
Equation (1.1.13) is not independent of equation (1.1.11) because it is well known that the quantity $\nabla \cdot (\nabla \times f)$ of an arbitrary vector function $f$ is always zero [2]. Therefore, if equation (1.1.11) is fulfilled, equation (1.1.13) will also be fulfilled. In the case of nonconducting materials, i.e. $\sigma = 0$, the same is valid for the relation between equation (1.1.12) and (1.1.14). In the more general case $\sigma \neq 0$ equation (1.1.12) and (1.1.14) require that:

$$\nabla \cdot \left( \sigma(r) e(r) e^{ik_0 L(r)} \right) = 0$$  \hspace{1cm} (1.1.15)

Using the rules of the Nabla calculus the left hand sides of equations (1.1.11) and (1.1.12) can be transformed to:

$$\nabla \times \left( e(r) e^{ik_0 L(r)} \right) = \nabla \left( e^{ik_0 L(r)} \right) \times e(r) + e^{ik_0 L(r)} \nabla \times e(r) = [ik_0 \nabla L(r) \times e(r) + \nabla \times e(r)] e^{ik_0 L(r)}$$  \hspace{1cm} (1.1.16)

$$\nabla \times \left( h(r) e^{ik_0 L(r)} \right) = [ik_0 \nabla L(r) \times h(r) + \nabla \times h(r)] e^{ik_0 L(r)}$$  \hspace{1cm} (1.1.17)

So, equations (1.1.11) and (1.1.12) give:

$$\nabla L(r) \times e(r) - c\mu(r) \mu_0 h(r) = \frac{i}{k_0} \nabla \times e(r)$$  \hspace{1cm} (1.1.18)

$$\nabla L(r) \times h(r) + c\epsilon(r) \epsilon_0 e(r) = \frac{i}{k_0} [\nabla \times h(r) - \sigma(r) e(r)]$$  \hspace{1cm} (1.1.19)

For the limiting case $\lambda \to 0$, i.e. $k_0 \to \infty$, the right sides of both equations become zero:

$$\nabla L(r) \times e(r) - c\mu(r) \mu_0 h(r) = 0$$  \hspace{1cm} (1.1.20)

$$\nabla L(r) \times h(r) + c\epsilon(r) \epsilon_0 e(r) = 0$$  \hspace{1cm} (1.1.21)

Now, equation (1.1.20) is inserted into equation (1.1.21) and the calculus for a double vector product is applied:

$$\nabla \cdot \left( (\nabla L(r) \times e(r)) \nabla L(r) - (\nabla L(r))^2 e(r) + n^2(r) e(r) \right) = 0$$  \hspace{1cm} (1.1.22)

Here, $\mu_0 \epsilon_0 = 1/c^2$ and $\mu \epsilon = n^2$ are used, where $n$ is the refractive index (dt.: Brechzahl) of the material.

Equation (1.1.21) shows that the scalar product $\nabla L \cdot e$ is zero and the final result is the well–known eikonal equation:

$$(\nabla L(r))^2 = n^2(r)$$  \hspace{1cm} (1.1.23)

This is the basic equation of geometrical optics which provides e.g. the basis for the concept of optical rays. A ray (dt.: Strahl) is defined as that trajectory which is always perpendicular to the wave fronts which are the surfaces of equal optical path length $L$ (see figure 1.1). Therefore, a ray points in the direction of $\nabla L$. Equation (1.1.23) has the name eikonal equation because the optical path length $L$ is for historical reasons sometimes called the eikonal [1].
1.2 The orthogonality condition of geometrical optics

Equations (1.1.20) and (1.1.21) can be solved for $e$ and $h$:

$$h(r) = \frac{1}{c\mu(r)\mu_0} \nabla L(r) \times e(r) \quad (1.2.1)$$

$$e(r) = -\frac{1}{ce(r)\epsilon_0} \nabla L(r) \times h(r) \quad (1.2.2)$$

This shows on the one hand that $h$ is perpendicular to $e$ as well as $\nabla L$ and on the other hand that $e$ is perpendicular to $h$ as well as $\nabla L$. Therefore, in the limiting case $\lambda \to 0$ $\nabla L$, $e$ and $h$ have to form an orthogonal triad of vectors. This confirms the well-known fact that electromagnetic waves are transversal waves.

At the end of the last section a light ray has been defined as being parallel to $\nabla L$ and in section 4 the important method of ray tracing will be explained. An extended method of ray tracing is polarization ray tracing where the polarization state of a ray which locally represents a wave is transported along with each ray [7],[47]. Using the results of this section it is clear that the vector $e$ indicating the polarization (and amplitude) of the ray has to be perpendicular to the ray direction $\nabla L$.

1.3 The ray equation

A surface with constant values $L$ is a surface of equal optical path length. Now, a ray is defined as that trajectory starting from a certain point in space which is perpendicular to the surfaces of equal optical path length. Therefore, $\nabla L$ points in the direction of the ray. We use the arc length $s$ along the curve which is defined by the ray (see fig. 1.2). Then, if $r$ describes now the position vector of a point on the ray, $dr/ds$ is a unit vector which is tangential to the ray curve and the eikonal equation (1.1.23) delivers:

$$\nabla L = n \frac{dr}{ds} \quad (1.3.1)$$

Here and in the following $L$ and $n$ are not explicitly indicated as functions of the position to tighten the notation. From equation (1.3.1) a differential equation for the ray can be derived by using again equation (1.1.23) and the definition of $d\nabla L/ds$ as being the directional derivative of $\nabla L$ along $dr/ds$:

$$\frac{d}{ds} \left( n \frac{dr}{ds} \right) = \frac{d}{ds} \nabla L = \frac{dr}{ds} \nabla (\nabla L) =$$
1.4 Limitations of the Eikonal Equation

Besides using directly the Maxwell equations the eikonal equation can also be derived from the wave equation and in the case of a monochromatic wave from the Helmholtz equation. This
will be done in the following for a homogeneous, isotropic and linear dielectric material, i.e. \( n \) is constant and \( \sigma = 0 \). Moreover, it is assumed that the scalar case is valid, i.e. that polarization effects can be neglected and only one component \( u(r) \) of the electric or magnetic vector has to be considered. In this limiting case it is easier to start directly with the scalar Helmholtz equation \([1]\) than to start like in section 1.1 with the Maxwell equations and then to make the transition to the scalar case.

The scalar Helmholtz equation is:

\[
\left( \nabla \cdot \nabla + (nk_0)^2 \right) u(r) = 0 \tag{1.4.1}
\]

Analogous to equations (1.1.9) or (1.1.10) the following approach for \( u \) is used

\[
u(r) = A(r) e^{ik_0 L(r)} \tag{1.4.2}
\]

where the amplitude \( A \) and the optical path length \( L \) are both real functions of the position and \( A \) varies only slowly with the position.

Then, by omitting the arguments of the functions we can write

\[
\nabla u = \nabla \left[ \frac{A e^{ik_0 L}}{A} \right] = e^{ik_0 L} \nabla A + ik_0 A e^{ik_0 L} \nabla L = \left( \frac{\nabla A}{A} + ik_0 \nabla L \right) u
\]

\[
\triangle u = \nabla \cdot \left[ \left( \frac{\nabla A}{A} + ik_0 \nabla L \right) u \right] = \left( \frac{\nabla A}{A} + ik_0 \nabla L \right)^2 u + \left( \frac{\triangle A}{A} - \frac{(\nabla A)^2}{A^2} + ik_0 \triangle L \right) u = \left( \frac{\triangle A}{A} - k_0^2 (\nabla L)^2 + 2ik_0 \frac{\nabla A \cdot \nabla L}{A} + ik_0 \triangle L \right) u
\]

Here, \( \nabla \cdot \nabla \) is the Laplace operator or Laplacian. So, by inserting the expression for \( \triangle u \) into the Helmholtz equation and dividing it by \( u \) the result is:

\[
\frac{\triangle A}{A} - k_0^2 (\nabla L)^2 + n^2 k_0^2 + 2ik_0 \frac{\nabla A \cdot \nabla L}{A} + ik_0 \triangle L = 0 \tag{1.4.3}
\]

Since \( A, L, k_0 \) and \( n \) are all real quantities the real and the imaginary part of this equation can be simply separated and both have to be zero.

To obtain the eikonal equation only the real part is considered:

\[
\frac{\triangle A}{A} - k_0^2 (\nabla L)^2 + n^2 k_0^2 = 0
\]

\[
\Rightarrow (\nabla L)^2 = n^2 + \frac{1}{k_0^2} \frac{\triangle A}{A} =: \gamma \tag{1.4.4}
\]
In the limiting case $\lambda \to 0 \Rightarrow k_0 \to \infty$ the term $\gamma$ can be neglected and again the eikonal equation (1.1.23) is obtained:

$$(\nabla L)^2 = n^2$$

But, equation (1.4.4) shows that also for a finite value of $\lambda$ the eikonal equation can be fulfilled with good approximation as long as the term $\gamma$ is much smaller than 1 because the order of magnitude of $n^2$ is typically between 1 (vacuum) and 12 (silicon for infrared light). Therefore, the condition is:

$$\gamma \ll 1 \Rightarrow \frac{\lambda^2}{4\pi^2} \frac{\Delta A}{A} \ll 1 \quad (1.4.5)$$

It is fulfilled with good approximation if $A$ is a slowly varying function of the position, i.e. if the relative curvature of $A$ over the distance of a wavelength is very small. If the term $\gamma$ is not very small the right hand side of equation (1.4.4) depends on the position (because $A$ depends generally on $r$) even though $n$ is constant. Formally this is equivalent to an eikonal equation with position dependent refractive index $n$ so that light rays would formally be bent in regions of a rapidly changing amplitude like e.g. in the focus. Therefore, the results of ray tracing calculations (see section 4) which assume rectilinear rays in a homogeneous material are not correct in the neighborhood of the focus where the amplitude changes very fast. If aberrations are present the variation of the amplitude in the focal region will be less severe and the accuracy of geometrical optical calculations improves with increasing aberrations. In practice, a rule of thumb is that the focal region of an aberrated wave calculated with ray tracing approximates the actual focus very good if the result of the ray tracing calculation gives a focus which has several times the size of the corresponding diffraction limited focus (Airy disc) which can be easily estimated (see the lecture about wave optics, term PSF).

There are also scalar waves which fulfill exactly the eikonal equation so that the term $\gamma$ is exactly zero. One example is a plane wave with $u(r) = u_0 \exp(ink_0 a \cdot r)$. $a$ is a constant unit vector in the direction of propagation and $u_0$ is also a constant. So, we have

$$A = u_0 \Rightarrow \Delta A = 0 \Rightarrow \gamma = 0$$

$$L = n a \cdot r \Rightarrow \nabla L = n a \Rightarrow (\nabla L)^2 = n^2$$

Of course, a plane wave is also a solution of the Maxwell equations.

A second example is a spherical wave which is only a solution of the scalar Helmholtz equation but not of the Maxwell equations themselves because the orthogonality conditions (1.2.1) and (1.2.2) cannot be fulfilled for a spherical wave in the whole space. Nevertheless, a spherical wave $u(r) = u_0 \exp(ink_0 r)/r$ with $r = |r|$ is a very important approximation in many cases and a dipole radiation behaves in the far field in a plane perpendicular to the dipole axis like a spherical wave. For the spherical wave we obtain:

$$A = \frac{u_0}{r} \Rightarrow \Delta A = u_0 \nabla \cdot \left( \frac{r}{r^3} \right) = -\frac{3u_0}{r^3} + \frac{3u_0 r \cdot r}{r^5} = 0 \Rightarrow \gamma = 0$$

$$L = nr \Rightarrow \nabla L = n \frac{r}{r} \Rightarrow (\nabla L)^2 = n^2$$

Here, the coordinate system has been chosen in such a way that the center of curvature of the spherical wave is at the origin. Of course, it is quite straightforward to formulate the spherical wave with an arbitrary center of curvature $r_0$ by replacing $r$ with $|r - r_0|$. So, plane waves and spherical waves, which are very important in geometrical optics, fulfill both the eikonal equation (1.1.23) not only in the limiting case $\lambda \to 0$ but also for finite wavelengths $\lambda$. 


1.5 Energy conservation in geometrical optics

The imaginary part of equation (1.4.3) gives information about the intensity of the transported amount of light:

\[ \Delta L + 2 \nabla L \cdot \frac{\nabla A}{A} = 0 \quad (1.5.1) \]

Since the intensity \( I \) of a light wave is proportional to the square of the amplitude \( A^2 \) the following equality holds:

\[ \frac{\nabla I}{I} = \frac{\nabla (A^2)}{A^2} = \frac{2 A \nabla A}{A^2} = 2 \frac{\nabla A}{A} \]

Therefore, equation (1.5.1) delivers

\[ I \Delta L + \nabla L \cdot \nabla I = 0 \]

or

\[ \nabla \cdot (I \nabla L) = 0 \quad (1.5.2) \]

Now, the integral theorem of Gauss can be applied

\[ \int_V \nabla \cdot (I \nabla L) \, dV = \oint_S I \nabla L \cdot dS = 0 \quad (1.5.3) \]

where the left integral symbolizes a volume integral over a volume \( V \) and the right integral a surface integral over the closed surface \( S \) which confines the volume \( V \).

A light tube (dt.: Lichtröhre) (see figure 1.3) is a tube-like entity (simple forms are e.g. a cylinder or a cone) where light rays form the mantle surface. Therefore, on the mantle surface the vectors \( \nabla L \) (ray direction) and \( dS \) (surface normal) are perpendicular to each other and therefore \( \nabla L \cdot dS = 0 \). At the two face surfaces of the light tube (refractive index \( n \)) with surface vectors \( dS_1 \) and \( dS_2 \) which are assumed to have an infinitesimally small size the electromagnetic power flux \( P_1 \) and \( P_2 \) is

\[ P_j = \frac{I_j}{n} |\nabla L_j \cdot dS_j|; \quad j \in \{1, 2\} \]
1.6 Law of refraction

Let us consider the interface between two materials with refractive index $n_1$ on the one side and $n_2$ on the other. This interface is assumed to be replaced by a very thin layer in which the refractive index varies quite rapidly but continuously from $n_1$ to $n_2$. An infinitesimally small rectangular closed loop $C$ is then constructed at the interface in such a way that two of the edges of the loop are parallel to the interface and the other two edges are parallel to the surface normal $N$ ($||N|| = 1$) of the interface where $N$ points from material 2 to material 1 (see figure 1.4). Since the direction vectors $a$ of the light rays can be expressed as the gradient of a scalar function (see equation (1.3.1)) the following identity is valid:

$$\nabla \times \left( n \frac{dr}{ds} \right) = \nabla \times \nabla L = 0$$  \hspace{1cm} (1.6.1)

The ray direction vector is written in the following as $a = dr/ds$.

Using the integral theorem of Stokes equation (1.6.1) delivers

$$\int_S \nabla \times (na) \cdot dS = \oint_C na \cdot dr = 0,$$
where the left integral is a surface integral over the infinitesimally small rectangular surface $S$ which is bounded by the closed loop $C$. The right integral is a line integral over the closed loop $C$.

If now the length of the side lines of the loop $C$ parallel to $N$ tends to zero the line integral is:

$$0 = lt \cdot (n_2a_2 - n_1a_1) \quad (1.6.2)$$

with $l$ being the length of a side line of the loop parallel to the interface and $t$ being a unit vector parallel to the interface. Another unit vector $b$ is defined as being perpendicular to both $N$ and $t$ and therefore also perpendicular to the surface $S$. This means that $N$, $t$ and $b$ form an orthogonal triad of unit vectors with $t = b \times N$ and therefore it holds:

$$(b \times N) \cdot (n_2a_2 - n_1a_1) = 0 \Rightarrow (N \times (n_2a_2 - n_1a_1)) \cdot b = 0$$

But, the rectangular integration area can be rotated about $N$ serving as axis. Therefore, the direction of $b$ can be chosen arbitrarily as long as it is perpendicular to $N$. By fulfilling the upper equation for an arbitrary vector $b$ we obtain the following equation which is the vectorial formulation of the law of refraction (dt.: Brechungsgesetz):

$$N \times (n_2a_2 - n_1a_1) = 0 \quad (1.6.3)$$

This means that $n_2a_2 - n_1a_1$ is parallel to $N$ (or $n_2a_2 - n_1a_1 = 0$ what is only possible for the trivial case $n_1 = n_2$) and therefore all three vectors $a_1$, $a_2$ and $N$ have to lie in the same plane. This means particularly that the refracted ray with direction vector $a_2$ lies in the plane of incidence formed by $N$ and $a_1$.

By defining the acute angles $\theta_j$ between the rays $a_j$ ($j \in \{1, 2\}$) and the surface normal $N$ (see figure 1.4) the modulus of equation (1.6.3) results in:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (1.6.4)$$

This is the well–known Snell’s law.

If $n_2$ is bigger than $n_1$ there is always a solution $\theta_2$ for a given angle $\theta_1$. But, if $n_2$ is smaller than $n_1$ there is the so–called critical angle of total internal reflection $\theta_{1, \text{critical}}$ (dt.: Grenzwinkel der Totalreflexion) for which the refracted ray grazes parallel to the interface, i.e. $\theta_2 = \pi/2$:

$$n_1 \sin \theta_{1, \text{critical}} = n_2 \sin \theta_2 = n_2 \Rightarrow \theta_{1, \text{critical}} = \arcsin \frac{n_2}{n_1} \quad (1.6.5)$$

For angles $\theta_1 > \theta_{1, \text{critical}}$ there exists no refracted ray because the sine function $\sin \theta_2$ cannot be larger than 1. Then, all light is reflected at the interface and only a reflected ray exists.

### 1.7 Law of reflection

If a plane wave enters the interface between two materials there is besides a refracted wave also a reflected wave and in the case of total internal reflection there is only a reflected wave. A ray represents locally a plane wave and the law of reflection is formally obtained from equation (1.6.3) by setting $n_2 = n_1$. Of course, the algebraic signs of the scalar products $N \cdot a_1$ and $N \cdot a_2$ have to be different in order to obtain a reflected ray whereas they have to be identical
to obtain a refracted ray. This will be discussed later in section 4 by finding explicit solutions of equation (1.6.3).

After having now discussed the basics of geometrical optics the next section will treat the paraxial ray tracing through an optical system by using a matrix theory [3],[11],[22],[42].
Chapter 2

Paraxial geometrical optics

2.1 Paraxial rays in homogeneous materials

2.1.1 Some basic definitions

In a homogeneous material the refractive index $n$ is constant and therefore, according to the ray equation (1.3.4), a ray propagates rectilinear. This means that a ray is definitely described by the position vector $p$ of one point on the ray and the ray direction vector $a$. So, 6 scalar parameters (each vector with 3 components) are necessary. In principle one component of $a$ is redundant (apart from the algebraic sign of this component) because $a$ is a unit vector ($a_x^2 + a_y^2 + a_z^2 = 1$). Another component can be saved if a reference plane is defined, e.g. the x–y plane at $z = 0$. Then, the x– and y–components of the point of intersection of the ray with this plane are sufficient. However, in the case of non–paraxial ray tracing (see section 4) all components of the two vectors $p$ and $a$ are stored and used since not all rays start in the same plane and the algebraic sign of each component of $a$ is needed. Moreover, it is often more efficient to store a redundant parameter instead of calculating it from other parameters.

Most convenient optical systems consist of a sequence of rotationally symmetric centered refractive and reflective components. The rotation axis is called the optical axis of the system. For a simple lens with two spherical surfaces the optical axis is defined by the two centers of curvature $C_1$ and $C_2$ of the spherical surfaces (see figure 2.1). Using equation (1.6.3) it has been shown that a refracted ray (and also a reflected ray) remains in the plane of incidence. Therefore, it is useful to define the meridional plane which is a plane containing an object point $P$ and the optical axis (see figure 2.2). All rays which come from the object point $P$ and lie in the

![Figure 2.1: Optical axis of a lens.](image-url)
meridional plane are called **meridional rays**. A plane perpendicular to the meridional plane which contains a special reference ray, mostly the chief ray (see section 3), is called the **sagittal plane** and rays lying in it are so called **sagittal rays**. In this section only meridional rays are discussed and moreover only so called **paraxial rays** are considered. Paraxial rays are rays which fulfill the following conditions:

- The distance $x$ of the ray from the optical axis is small compared to the focal length of each optical element of the system.
- The angle $\phi$ between the optical axis and the ray is small, i.e. $\phi \ll 1$. The same has to be valid for other angles, e.g. for the refraction angles at a lens.

For the angles this means that the following approximations have to be valid:

$$\sin \varphi \approx \tan \varphi \approx \varphi \quad \text{and} \quad \cos \varphi \approx 1$$

The most important optical systems consist of optical elements (refractive, reflective or diffractive elements) which are embedded into piecewise homogeneous materials. Therefore, the ray tracing (in the paraxial as well as non–paraxial case) through an optical system consists of the alternating sequence of propagation in a homogeneous material and refraction (or reflection or diffraction) at an element.

### 2.1.2 Optical imaging

At this point some words to the term **optical imaging** have to be said. An object point which is either illuminated by external light or self–illuminating emits a ray fan, i.e. in geometrical optics an object point is the source of a ray fan. On the other side an image point is the drain of a ray fan and in the ideal case all rays of the fan should intersect each other in the image point (see figure 2.3a)). Therefore, the image point can be defined in the ideal case by the point of intersection of only two rays. However, this is only useful in the case of **paraxial ray tracing** where all aberrations of the optical system are neglected. If the aberrations of an optical system have also to be taken into account the non–paraxial ray tracing (see section 4), simply called **ray tracing**, has to be used. Then, there are several definitions of an image point because there is in general no longer a single point of intersection of all rays of the ray fan coming from the object point (see figure 2.3b)). The lateral deviation of the actual point of intersection of a ray with the image plane from the ideal image point is called **ray aberration** (dt.: Strahlaberration).
Figure 2.3: Schematic display of three different situations in optical imaging: a) ideal imaging, b) image point showing ray aberrations (and of course also wave aberrations), c) image point showing no ray aberrations but nevertheless wave aberrations.
A more advanced definition of optical imaging has of course to take into account interference effects between the different rays coming from the object point since the image point is a multiple beam interference phenomenon. A typical example where the simple ray–based model fails would be an ideal spherical wave where a half–wave plate is introduced in half of the aperture (see figure 2.3c)). Then, the ray directions are unchanged and an ideal point of intersection of all rays exists, i.e. there are no ray aberrations. But, the image point would be massively disturbed because there is destructive interference in the center of the image point due to the different optical path lengths of the rays. Therefore, a more advanced ray–based model calculates additionally the optical path length along each ray. The deviation in the optical path length of a ray from the ideal optical path length is called wave aberration (dt.: Wellenaberration). But, in this section we will treat the very simple model of paraxial ray tracing which neither takes into account ray aberrations nor wave aberrations. Aberrations will be taken into account in section 4 about non–paraxial ray tracing.

2.1.3 A note to the validity of the paraxial approximation

The approximation of \( \sin \varphi \) by \( \varphi \) means that the next term of the Taylor series \(-\varphi^3/6\) and all higher order terms are neglected. In the case of \( \tan \varphi \) the next term of the Taylor series which is neglected is \(+\varphi^3/3\). So, the equivalence of \( \sin \varphi \) and \( \tan \varphi \) is only valid if the difference of both third order terms \( \varphi^3/2 \) is so small that it can be neglected. This is the case if the alternation of the optical path length from the object point to the image point by neglecting this term is smaller than the Rayleigh criterion of \( \lambda/4 \), where \( \lambda \) is again the wavelength. In the case of two rays with an optical path difference of \( \lambda/4 \) the phase difference is \( \Delta \Phi = \pi/2 \), i.e. the rays are in phase quadrature and the intensities have to be added because the interference term \( \cos(\Delta \Phi) \) is then zero. If the optical path difference is \( \lambda/2 \) the phase difference is \( \Delta \Phi = \pi \) and the amplitudes of both rays cancel each other (if the amplitudes have equal modulus). Then, the image point is strongly aberrated. So, the validity of the Rayleigh criterion is useful to define the limitations of the paraxial approximation.

In practice, the paraxial theory is quite important because it allows the definition of such important parameters as the focus, the focal length or the principal points of a lens or optical system. An optical designer [18],[19],[33],[36] will always first design an optical system by using the paraxial matrix theory (or another paraxial method) so that the paraxial parameters are right. Afterwards, he will try to optimize the non–paraxial parameters using ray tracing in order to correct aberrations of the system.

2.1.4 Definition of a paraxial ray

In the paraxial theory only rays in the meridional plane, which is here defined as the x–z–plane, are regarded. Then, the y–component of the ray direction vector \( \mathbf{a} \) and the y–component of the starting point \( \mathbf{p} \) of the ray are both zero: \( a_y = 0 \) and \( p_y = 0 \). We define for the x–component of the ray direction vector \( a_x = \sin \varphi \approx \tan \varphi \approx \varphi \). The z–component of the ray direction vector is then in the paraxial approximation \( a_z = \cos \varphi \approx 1 \). Therefore, a meridional paraxial ray at a certain z–position \( z \) can be described by the angle \( \varphi \) with the optical axis and the ray height \( x \) which is indeed the x–component \( p_x \) of the starting point \( \mathbf{p} \) of the ray. The z–component \( p_z \) of a ray is noted in the paraxial matrix theory externally because in many cases several rays starting at the same z–position \( z = p_z \) but having different values \( x \) and \( \varphi \) are considered.
Figure 2.4: Scheme showing the parameters of the paraxial ray tracing for the transfer between two parallel planes with distance $d$.

So, in total a paraxial ray is described by $x$ and $\varphi$. Since matrix methods play an important role in optics [3],[11] these two parameters are noted as the components of a vector

$$\begin{pmatrix} x \\ \varphi \end{pmatrix}$$

so that the optical operations which we will discuss now can be represented as 2 x 2 matrices.

2.1.5 Transfer equation

The paraxial ray tracing between two planes with distance $d$ which are perpendicular to the optical axis is one of the basic operations. Here, only the lines of intersection of these two planes with the meridional plane are regarded (see figure 2.4) even though we speak further on a little bit incorrectly of ”planes”. The ray parameters in the first plane shall be $x$ and $\varphi$ and those in the second plane $x'$ and $\varphi'$. Then, the transfer from the first plane to the second plane with distance $d$ is done by (see fig. 2.4):

$$\begin{pmatrix} x' \\ \varphi' \end{pmatrix} = \begin{pmatrix} x + \varphi d \\ \varphi \end{pmatrix}$$

(2.1.1)

This means that ray directions are not changed during the propagation of paraxial rays in a homogeneous material. Equation (2.1.1) can also be written by using a two times two matrix [3],[11],[22],[42]:

$$\begin{pmatrix} x' \\ \varphi' \end{pmatrix} = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \varphi \end{pmatrix} = M_T \begin{pmatrix} x \\ \varphi \end{pmatrix}$$

(2.1.2)

The matrix $M_T$ is called the paraxial transfer matrix in a homogeneous material.
2.2 Refraction in the paraxial case

2.2.1 Paraxial law of refraction

The law of refraction connects the angle $i$ between the incident ray and the surface normal with the angle $i'$ between the refracted ray and the surface normal (see figure 2.5). The law of refraction (see equation (1.6.4)) is in the paraxial formulation

$$ni = n'i'$$

(2.2.1)

where $n$ and $n'$ are the refractive indices of the two homogeneous materials in front of the surface and behind the surface.

2.2.2 Refraction at a plane surface

A paraxial ray with parameters $x$ and $\varphi$ hits a plane surface which is perpendicular to the optical axis (see fig. 2.5). The refractive index is $n$ in front of the surface and $n'$ behind the surface. Then, the ray height $x$ remains unchanged and only the ray parameter $\varphi$ changes according to the paraxial law of refraction (see equation (2.2.1)):

$$\begin{pmatrix} x' \\ \varphi' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{n'} \end{pmatrix} \begin{pmatrix} x \\ \varphi \end{pmatrix} = M_R \begin{pmatrix} x \\ \varphi \end{pmatrix}$$

(2.2.2)

The matrix $M_R$ is the paraxial matrix for the refraction at a plane surface.

2.2.3 Refraction at a plane parallel plate

The plane parallel plate is the simplest case for a sequence of several surfaces and can be used to demonstrate the principle of tracing paraxial rays through an optical system by using the paraxial matrix theory. It is well-known that the order of two matrices $A$ and $B$ is very important if two matrices have to be multiplied, i.e.:

$$AB \neq BA$$
Therefore, the matrix for the first operation has to be positioned immediately left to the vector \((x, \varphi)\) of the paraxial ray which has to be traced through the system. The matrix of the next operation has then to be multiplied from the left side and so on for all other matrices. Using the notations of fig. 2.6 the parameters of the paraxial ray at the right side of the plane parallel plate with thickness \(d\) and refractive index \(n\) left to the plane parallel plate and \(n'\) right to the plane parallel plate are:

\[
\begin{bmatrix}
    x' \\
    \varphi'
\end{bmatrix} = \begin{bmatrix}
    1 & 0 \\
    0 & \frac{n'}{n'}
\end{bmatrix} \begin{bmatrix}
    1 & d \\
    0 & 1
\end{bmatrix} \begin{bmatrix}
    1 & 0 \\
    0 & \frac{n}{n'}
\end{bmatrix} \begin{bmatrix}
    x \\
    \varphi
\end{bmatrix} = \begin{bmatrix}
    x' \\
    \varphi'
\end{bmatrix} = M_\text{P} \begin{bmatrix}
    x \\
    \varphi
\end{bmatrix}
\]

Here, \(n\) is the refractive index left to the plane parallel plate and \(n'\) the refractive index right to the plane parallel plate.

In total the parameters \(x'\) and \(\varphi'\) of a paraxial ray immediately behind the plane parallel plate are obtained from the parameters \(x\) and \(\varphi\) of the incident ray immediately in front of the plane parallel plate by multiplying them with the matrix \(M_\text{P}\) of a plane parallel plate:

\[
\begin{bmatrix}
    x' \\
    \varphi'
\end{bmatrix} = \begin{bmatrix}
    1 & d \frac{n}{n'} \\
    0 & \frac{n}{n'}
\end{bmatrix} \begin{bmatrix}
    x \\
    \varphi
\end{bmatrix} = M_\text{P} \begin{bmatrix}
    x \\
    \varphi
\end{bmatrix} \tag{2.2.3}
\]

The most important practical case is a plane parallel plate in air \((n = n' = 1)\). Then it holds:

\[
\begin{bmatrix}
    x' \\
    \varphi'
\end{bmatrix} = \begin{bmatrix}
    1 & d \\
    0 & 1
\end{bmatrix} \begin{bmatrix}
    x \\
    \varphi
\end{bmatrix} \tag{2.2.4}
\]
This means that the matrix of a plane parallel plate is identical to the transfer matrix in a homogeneous material by substituting the transfer distance $d$ in the homogeneous material with the term $d/n_P$. A lateral ray shift $\Delta x = x_{\text{without plate}} - x_{\text{with plate}}$ results at a plane parallel plate with thickness $d$ compared to the propagation in air by a distance $d$:

$$\Delta x = x + \varphi d - \left( x + \varphi \frac{d}{n_P} \right) = \varphi d \frac{n_P - 1}{n_P}$$

For normal glass with $n_P \approx 1.5$ the lateral ray shift is: $\Delta x = \varphi d/3$. This effect is used in optical systems to introduce a lateral shift where the size of the shift increases with the ray angle $\varphi$. So, in practice the plane parallel plate is tilted by an angle $\varphi$ with respect to the optical axis of the system to introduce such a lateral shift. However, a plane parallel plate can also introduce aberrations if the incident wave is not a plane wave. Therefore, the introduction of a lateral shift by using a plane parallel plate has to be used with care.

### 2.2.4 Some notes to sign conventions

Up to now no sign conventions were made for the paraxial matrix theory. This will be made up for now and illustrated graphically in fig. 2.7

- Ray angles $\varphi$ are positive if the acute angle between the optical axis and the ray is mathematically positive.
- Refraction angles are all treated as positive angles.
- All angles are acute angles.
- Light rays are always travelling from left to right for positive propagation distances $d$. A negative propagation distance $d$ means that the light is travelling from right to left and is only used for virtual rays.
- Ray heights $x$ are upwards positive.
- Radii of curvature $R$ are positive if the center of curvature is right to the vertex of the surface.
2.2.5 Refraction at a spherical surface

A spherical surface with radius of curvature \( R \) and refractive indices \( n \) in front of and \( n' \) behind the surface is hit by a paraxial ray with parameters \( x \) and \( \varphi \). In the paraxial approximation the ray height \( x \) at the point of intersection of the ray with the spherical surface is the same as in the vertex plane since the radius of curvature \( R \) is assumed to be large compared to \( x \) ! According to figure 2.8 and the paraxial law of refraction (equation (2.2.1)) the following relations are valid:

\[
\varphi' + \alpha = i' \\
\varphi + \alpha = i \\
\Rightarrow \varphi' = i' - \alpha = \frac{n}{n'}(\varphi + \alpha) - \alpha
\]

Additionally, the angle \( \alpha \) between the optical axis and the line connecting the center of curvature of the spherical surface and the point of intersection of the ray with the spherical surface is in the paraxial approximation defined as:

\[
\sin \alpha = \frac{x}{R} \Rightarrow \alpha = \frac{x}{R}
\]

Totally, this allows to express the ray angle \( \varphi' \) of the refracted paraxial ray as a function of the parameters of the incident ray and the spherical surface:

\[
\varphi' = \frac{n}{n'}\varphi - \frac{n' - n}{n'} \frac{x}{R} \tag{2.2.5}
\]

The ray height \( x \) itself remains constant in the case of refraction. Therefore, the matrix \( M_S \) for refraction at a spherical surface is defined as:

\[
\begin{pmatrix}
  x' \\
  \varphi'
\end{pmatrix} =
\begin{pmatrix}
  \frac{1}{n' - n} & \frac{n}{n'} \\
  -\frac{n' - n}{n' R} & \frac{n}{n'}
\end{pmatrix}
\begin{pmatrix}
  x \\
  \varphi
\end{pmatrix} = M_S \begin{pmatrix}
  x \\
  \varphi
\end{pmatrix} \tag{2.2.6}
\]

The validity of the sign conventions can be shown by regarding some concrete cases:

- For \( \varphi = 0, n' > n > 1 \) and \( R > 0 \) (convex surface) a positive ray height \( x \) of the incident ray results in a negative ray angle \( \varphi' \) of the refracted ray. This means that the convex spherical surface with lower refractive index on the left side has a positive refractive power and focusses a plane wave.
Figure 2.9: Cardinal points of an optical system: F and F’ are the foci in the object space and the image space, respectively. N and N’ are the nodal points in the object and image space and U and U’ are the unit or principal points in the object and image space.

- For \( \varphi = 0 \), \( n' > n > 1 \) and \( R < 0 \) (concave surface) the angle \( \varphi' \) of the refracted ray is positive if the ray height \( x \) of the incident ray is also positive. This means that two rays would only intersect virtually in front of the lens. Therefore, the concave spherical surface with lower refractive index on the left side has a negative focal power (dt.: negative Brechkraft).

2.3 The cardinal points of an optical system

An optical imaging system has several cardinal points (dt.: Kardinalpunkte) [1],[13] and by knowing these values the paraxial properties of the optical system are determined definitely. The cardinal points are the principal points (dt.: Hauptpunkte), the nodal points (dt.: Knotenpunkte) and the focal points (dt.: Brennpunkte). All these points are situated on the optical axis. In order to define them some additional definitions have to be made.

The cardinal points will be calculated in this section for a general optical system using the paraxial matrix theory[32]. At the end of this section the cardinal points of the simplest optical system, a refracting spherical surface, will be calculated explicitly to demonstrate the method.

Assume a general optical imaging system like symbolized in fig. 2.9. An object point \( P_O \) with a lateral distance \( x_O \) from the optical axis, which is called object height (dt.: Objekthöhe), is imaged by the optical system to an image point \( P_I \) with the lateral distance \( x_I \), called image height (dt.: Bildhöhe). The refractive indices are \( n \) in the object space and \( n' \) in the image space.

The lateral magnification \( \beta \) (dt.: Abbildungsmaßstab) of an imaging system is defined as the ratio of the image height \( x_I \) and the object height \( x_O \):

\[
\beta := \frac{x_I}{x_O} \tag{2.3.1}
\]
CHAPTER 2. PARAXIAL GEOMETRICAL OPTICS

According to our sign convention the lateral magnification in fig. 2.9 is negative since \( x_O \) is positive and \( x_I \) negative.

2.3.1 The principal points

The principal plane \( \mathcal{U} \) (dt.: Hauptebeene) or unit plane in the object space is that plane perpendicular to the optical axis which has the property that an object point in this principal plane is imaged to a point in the principal plane \( \mathcal{U}' \) of the image space with a lateral magnification \( \beta = +1 \). The points of intersection of the principal planes in object and image space with the optical axis are called the principal or unit points \( U \) and \( U' \), respectively. So, \( U' \) is the image of \( U \).

An important practical property of the principal planes following from the definition is that a ray which intersects the principal plane \( \mathcal{U} \) in the object space at a height \( x \) is transferred to the principal plane \( \mathcal{U}' \) of the image space at the same height (see fig. 2.9). This property is e.g. used to construct graphically the path of a paraxial ray.

2.3.2 The nodal points

The second cardinal points of an optical system are the nodal points \( N \) (in the object space) and \( N' \) (in the image space). A ray in the object space which intersects the optical axis in the nodal point \( N \) by the angle \( \varphi \) intersects the optical axis in the image space in the nodal point \( N' \) by the same angle \( \varphi' = \varphi \). Therefore, the angular magnification \( \gamma \) (dt.: Winkelvergrößerung) defined as

\[
\gamma := \frac{\varphi'}{\varphi}
\]

is \( \gamma = 1 \) for rays going through the nodal points. Additionally, since this has to be valid for arbitrary angles \( \varphi \) the nodal point \( N' \) is the image of the nodal point \( N \).

2.3.3 The focal points

The focal points \( F \) (in the object space) and \( F' \) (in the image space), also called principal foci or shortly foci, have the following properties. A ray starting from the focus \( F \) in the object space is transformed into a ray parallel to the optical axis in the image space. Vice versa, a ray which is parallel to the optical axis in the object space intersects the focus \( F' \) in the image space. The planes perpendicular to the optical axis which intersect the optical axis in the focal points are called the focal planes (dt.: Brennebenen). The distance between the principal point \( U \) and the focus \( F \) is called the focal length \( f \) in the object space (dt.: objektseitige Brennweite) and the distance between the principal point \( U' \) and the focal point \( F' \) is called the focal length \( f' \) in the image space (dt.: bildseitige Brennweite). In geometrical optics the sign convention for the focal length is usually that it is positive if the focus is right to the principal point. In figure 2.9 this means e.g. that the focal length \( f \) in the object space is negative whereas the focal length \( f' \) in the image space is positive.

A more general property of the focal planes is that rays starting from a point with object height \( x_O \) in the focal plane of the object space form in the image space a bundle of parallel rays making the angle \( \varphi' = -x_O/f' \) with the optical axis. The relation for \( \varphi' \) can be easily understood by the fact that a ray starting from the object point parallel to the optical axis is transferred at the
principal planes from $U$ to $U'$ with the same height $x_O$ and passes then after the distance $f'$ the focal point $F'$ in the image space. The negative sign has to be taken due to the sign convention.

### 2.3.4 Calculation of the cardinal points of a general optical system

Assume to have a general optical system which is formed by an arbitrary combination of refracting spherical and plane surfaces which are all situated on a common optical axis (see fig. 2.10). Then, the system can be described by a 2x2 matrix $M$ which is the product of a sequence of matrices $M_T$, $M_R$ and $M_S$ (or further matrices for other optical elements). The matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = M_{S,m} \cdot M_{T,m-1} \cdot M_{S,m-1} \cdot \ldots \cdot M_{T,2} \cdot M_{S,2} \cdot M_{T,1} \cdot M_{S,1}$$  \tag{2.3.3}$$

describes the propagation of a ray from a plane immediately in front of the vertex of the first surface (surface 1) to a plane immediately behind the vertex of the last surface (surface $m$). Here, only matrices $M_S$ of refractive spherical surfaces are taken because a plane surface with matrix $M_R$ can be represented as a spherical surface with radius of curvature $R = \infty$. Additionally, behind each surface (apart from the last surface) the transfer to the next surface is described by using a matrix $M_{T,i}$. In the special case of a thin lens (which does not exist in reality but which is an important idealization in geometrical optics) the propagation distance can just be set to zero so that the transfer matrix is identical to the unit matrix.

The restriction to spherical surfaces is not stringent because in paraxial optics an aspheric surface is identical to a spherical surface if the radius of curvature of the aspheric surface at the vertex is identical to the radius of curvature of the spherical surface. From a mathematical point of view the determination of the radius of curvature in the paraxial regime just means that in both cases the parabolic terms are taken. Moreover, also cylindrical surfaces can be calculated with this method if the radius of curvature in the selected $x$–$z$–plane is taken. For a plane which contains the cylinder axis the cylindrical surface behaves like a plane surface whereas the cylindrical surface behaves like a spherical surface if the cylinder axis is perpendicular to the regarded $x$–$z$–plane.

A ray starts in front of the optical system in a material with refractive index $n := n_1$ and ends behind the system in a material with refractive index $n' := n'_m$ (see fig. 2.10). $n_i$ and $n'_i$ with $i \in \{1,2,\ldots,m\}$ are the refractive indices in front of and behind each refracting surface which is described by the matrix $M_{S,i}$. Of course, there is the relation

$$n_i = n'_{i-1} \quad \text{for} \quad i \in \{2,3,\ldots,m\}$$  \tag{2.3.4}$$

Now, a matrix $M'$ is calculated which describes the propagation of a ray from a plane $\mathcal{P}$ through the optical system to a plane $\mathcal{P}'$. The plane $\mathcal{P}$ has the distance $d$ to the vertex of the first surface of the optical system, whereas the vertex of the last surface has the distance $d'$ to the plane $\mathcal{P}'$. Using the paraxial sign conventions $d$ is positive if $\mathcal{P}$ is in front of (i.e. left to) the first surface. Similar $d'$ is positive if $\mathcal{P}'$ is behind (i.e. right to) the last surface. It is very important to remember that $d$ is measured from the plane $\mathcal{P}$ to the vertex of the first surface whereas $d'$ is measured from the vertex of the last surface to the plane $\mathcal{P}'$! For these quantities the usual sign conventions are valid, i.e. they are positive if the propagation is from left to right and negative if the propagation is in the opposite direction.
Figure 2.10: Distances between the cardinal points in the object space and the vertex of the first surface (quantities without apostrophe) and the vertex of the last surface and the cardinal points in the image space (quantities marked by an apostrophe) of a general optical system consisting of refractive surfaces. \( d_U \) and \( d'_U \) are negative in the scheme, \( d_F \) and \( d'_F \) are positive and \( d_N \) and \( d'_N \) are again negative. \( n \) and \( n' \) are the refractive indices in front of and behind the whole system, whereas \( n_i \) and \( n'_i \) are the refractive indices in front of and behind the single refracting surface number \( i \).

By using equation (2.1.2) the matrix \( M' \) is:

\[
M' = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = M_T' M M_T = \\
\begin{pmatrix} 1 & d' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} = \\
\begin{pmatrix} A + C d' \\ C \end{pmatrix} \begin{pmatrix} A d + B + C d d' + D d' \\ C d + D \end{pmatrix}
\]

(2.3.5)

### 2.3.4.1 Principal points

To calculate the principal planes \( \mathcal{U} \) and \( \mathcal{U}' \) of the system the definition is used. If \( \mathcal{P} \) is identical to the principal plane \( \mathcal{U} \) and \( \mathcal{P}' \) identical to \( \mathcal{U}' \) an object point in \( \mathcal{P} \) has to be imaged to \( \mathcal{P}' \) with the lateral magnification \( \beta = +1 \). Imaging means that all rays with arbitrary ray angles \( \varphi \) starting from the object point with height \( x \) have the same height \( x' \) in \( \mathcal{P}' \) independent of \( \varphi \). Since the relation

\[
\begin{pmatrix} x' \\ \varphi' \end{pmatrix} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} \begin{pmatrix} x \\ \varphi \end{pmatrix} = \begin{pmatrix} A' x + B' \varphi \\ C' x + D' \varphi \end{pmatrix}
\]

(2.3.6)

holds, this means that the matrix element \( B' \) has to be zero in order to have imaging between the planes \( \mathcal{P} \) and \( \mathcal{P}' \). Therefore, we have as first condition:

\[
B' = A d + B + C d d' + D d' = 0
\]

(2.3.7)
2.3. THE CARDINAL POINTS OF AN OPTICAL SYSTEM

The second condition $\beta = 1$ means by using $B' = 0$:

$$x' = A'x + B'\varphi = (A + Cd')x = x \Rightarrow A + Cd' = 1 \Rightarrow d'_{U'} = \frac{1 - A}{C} \quad (2.3.8)$$

Here, we use the name $d'_{U'} := d'$ (see fig. 2.10) to indicate that it is the distance from the vertex of the last surface of the optical system to the principal point $U'$. From the first condition we obtain then the distance $d_{U} := d$ between the principal point $U$ and the vertex of the first surface:

$$B' = Ad + B + (1 - A)d + D\frac{1 - A}{C} = 0 \Rightarrow d_{U} = \frac{D}{C}(A - 1) - B \quad (2.3.9)$$

It has to be mentioned that in the case of optical imaging the coefficient $A'$ of the matrix $M'$ has a concrete meaning. It is:

$$x' = A'x \Rightarrow \beta = \frac{x'}{x} = A' \quad (2.3.10)$$

Therefore, the coefficient $A'$ is identical to the lateral magnification $\beta$ defined by equation (2.3.1).

2.3.4.2 Nodal points

If $P$ contains the nodal point $N$ and $P'$ the nodal point $N'$ the conditions for the ray parameters are $x = x' = 0$ and $\varphi' = \varphi$. Using equation (2.3.6) this gives:

$$0 = x' = A'x + B'\varphi = B'\varphi \Rightarrow B' = Ad + B + Cdd' + Dd' = 0 \quad (2.3.11)$$

Then, the distances $d_{N} := d$ between the nodal point $N$ and the vertex of the first surface of the optical system on the one hand and $d'_{N'} := d'$ between the vertex of the last surface of the optical system and the nodal point $N'$ on the other hand are:

$$d_{N} = 1 - \frac{D}{C} \quad (2.3.12)$$

$$d'_{N'} = \frac{A}{C}(D - 1) - B \quad (2.3.13)$$

2.3.4.3 Focal points and focal lengths

For the calculation of the focus $F$ in the object space it is assumed that $F$ is in the plane $\mathcal{P}$. Then, all rays starting from the height $x = 0$ have to be in the image space rays parallel to the optical axis, i.e. $\varphi' = 0$. Since this has to be valid in all planes in the image space the distance $d'$ in equation (2.3.5) is set to zero. So, the condition for the distance $d_{F} := d$ between the focus $F$ and the vertex of the first surface of the optical system is:

$$\varphi' = C'd'x + D'd'\varphi = D'd'\varphi = 0 \Rightarrow D' = Cd + D = 0 \Rightarrow d_{F} = -\frac{D}{C} \quad (2.3.14)$$

The focal length $f$ is defined as the distance between the principal point $U$ and the focus $F$, where the sign convention in geometrical optics is that $f$ is positive if $F$ is right to $U$. Therefore, by using the sign conventions for $d_{U}$ and $d_{F}$ it is:

$$f = d_{U} - d_{F} = \frac{D}{C}(A - 1) - B + \frac{D}{C} = \frac{AD}{C} - B = \frac{AD - BC}{C} \quad (2.3.15)$$
The focus $F'$ in the image space can be calculated analogous. There, rays parallel to the optical axis (i.e. $\varphi = 0$) in front of the optical system in an arbitrary plane $P$, e.g. at $d = 0$, have to focus in the image space in the focus $F'$ at $x' = 0$. If $F'$ is in the plane $P'$ this means for the distance $d_{F'}' := d'$ between the vertex of the last surface of the optical system and the focus $F'$ by using equation (2.3.6):

$$x' = A'x + B'\varphi = A'x = 0 \Rightarrow A' = A + Cd' = 0 \Rightarrow d_{F'}' = -\frac{A}{C} \tag{2.3.16}$$

Analogous, the focal length $f'$, which is positive if $F'$ is right to $U'$, can be calculated by:

$$f' = d_{F'}' - d_{U'} = -\frac{A}{C} - \frac{1 - A}{C} = -\frac{1}{C} \Rightarrow C = -\frac{1}{f'} \tag{2.3.17}$$

Now, the concrete meaning of the matrix coefficient $C$ as the negative reciprocal value of the focal length $f'$ in the image space becomes clear. $1/f'$ is also called the optical power (dt. Brechkraft) of the optical system, so that $C$ is the negative value of the optical power.

By summarizing equations (2.3.8), (2.3.9), (2.3.12), (2.3.13), (2.3.14) and (2.3.16) the distances $d_U, d_N$ and $d_F$ between the cardinal points in the object space and the vertex of the first surface of the optical system as well as the distances $d_{U'}', d_{N'}'$ and $d_{F'}'$ between the vertex of the last surface of the optical system and the cardinal points in image space are:

$$d_U = \frac{D}{C}(A - 1) - B$$
$$d_N = \frac{1 - D}{C}$$
$$d_F = -\frac{D}{C}$$
$$d_{U'}' = \frac{1 - A}{C}$$
$$d_{N'}' = \frac{A}{C}(D - 1) - B$$
$$d_{F'}' = -\frac{A}{C} \tag{2.3.18}$$

Also, the focal lengths can now be expressed as functions of the coefficients $A, B, C$ and $D$ of the matrix $M$ by summarizing equations (2.3.15) and (2.3.17):

$$f = \frac{AD - BC}{C} = \frac{\text{Det}(M)}{C} \tag{2.3.19}$$
$$f' = -\frac{1}{C}$$

### 2.3.5 Relation between the focal lengths in object and image space

There is a very interesting relation between the focal length $f$ in the object space and the focal length $f'$ in the image space. To derive it the ratio $f'/f$ is calculated by using equations (2.3.15) and (2.3.17):

$$\frac{f'}{f} = \frac{-1/C}{(AD - BC)/C} = -\frac{1}{AD - BC} = -\frac{1}{\text{Det}(M)} \tag{2.3.20}$$
Here, the determinant $\det(M)$ of the matrix $M$, defined by equation (2.3.3), has been used. According to the calculus of linear algebra the determinant of the product of several matrices is equal to the product of the determinants of these matrices. Therefore, it holds:

$$
\det(M) = \det(M_{S,m}) \cdot \det(M_{T,m-1}) \cdot \det(M_{S,m-1}) \cdot \ldots \cdot \det(M_{T,1}) \cdot \det(M_{S,1}) \quad (2.3.21)
$$

So, we have first to calculate the determinants of the two elementary matrices of equations (2.1.2) and (2.2.6):

$$
M_{T,i} = \begin{pmatrix} 1 & d_i \\ 0 & 1 \end{pmatrix} \Rightarrow \det(M_{T,i}) = 1 \quad (2.3.22)
$$

$$
M_{S,i} = \begin{pmatrix} 1 & 0 \\ -n'_i & n'_i \end{pmatrix} \Rightarrow \det(M_{S,i}) = \frac{n_i}{n'_i} \quad (2.3.23)
$$

Again, $n_i$ and $n'_i$ are the refractive indices in front of and behind the respective surface. $d_i$ is the distance between surface $i$ and $i+1$ ($i \in \{1, 2, \ldots, m-1\}$) and $R_i$ is the radius of curvature of surface $i$.

Now, we define again the refractive index in front of the first surface as $n := n_1$ and the refractive index behind the last surface of the optical system as $n' := n'_m$. Since the determinants of the transfer matrices $M_{T,i}$ are one the determinant of $M$ is:

$$
\det(M) = \prod_{i=1}^{m+1} \det(M_{S,m+1-i}) = \prod_{i=1}^{m} \frac{n_{m+1-i}}{n'_{m+1-i}} = \frac{n_m \cdot n_{m-1} \cdot \ldots \cdot n_2 \cdot n_1}{n'_m \cdot n'_{m-1} \cdot \ldots \cdot n'_2 \cdot n'_1} = \frac{n'_{m-1} \cdot n'_{m-2} \cdot \ldots \cdot n'_2 \cdot n'_1}{n'_{m-1} \cdot n'_{m-2} \cdot \ldots \cdot n'_2 \cdot n'_1} = \frac{n}{n'} = n 
$$

Here, relation (2.3.4) for the refractive indices of neighbored surfaces has been used. Therefore, the ratio of the focal length $f'$ and $f$ is according to equation (2.3.20):

$$
\frac{f'}{f} = -\frac{1}{\det(M)} = -\frac{n'}{n} \quad \text{or} \quad \frac{f'}{n'} = -\frac{f}{n} \quad (2.3.25)
$$

### 2.3.6 The cardinal points of an optical system with identical surrounding refractive indices

An interesting special case is that the refractive indices $n$ in front of the first surface of the optical system and $n'$ behind the last surface of the optical system are identical: $n = n'$. Then the determinant of the matrix $M$ is according to equation (2.3.24) $\det(M) = 1$. Therefore, the focal lengths in object and image space have due to equation (2.3.25) equal absolute value but different signs (due to the sign conventions of geometrical optics):

$$
f' = -f \quad (2.3.26)
$$

A second quite interesting property of an optical system with identical refractive indices in front of the first surface and behind the last surface is that the principal points and the nodal points
2.3.7 The cardinal points of a spherical refracting surface

The simplest optical imaging system is a single spherical refracting surface. As an application of the equations (2.3.18) and (2.3.19) the cardinal points of a spherical refracting surface shall be determined.

In this special case the matrix $M$ is according to equation (2.2.6):

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} := M_S = \begin{pmatrix} 1 & -\frac{n'-n}{n'R} \\ 0 & \frac{n}{n'} \end{pmatrix}$$ (2.3.29)

Then, according to the equations (2.3.18) the result is:

$$d_U = \frac{D}{C}(A - 1) - B = 0$$

$$d_N = \frac{1 - D}{C} = -R$$

$$d_F = \frac{-D}{C} = \frac{nR}{n' - n}$$

$$d_{U'} = \frac{1 - A}{C} = 0$$

$$d_{N'} = \frac{A}{C}(D - 1) - B = R$$

$$d_{F'} = -\frac{A}{C} = \frac{n' R}{n' - n}$$ (2.3.30)

This means (see fig. 2.11):
1. Both principal points coincide with the vertex of the spherical surface \((d_U = d'_U = 0)\).

2. Both nodal points coincide with the center of curvature of the spherical surface \((-d_N = d'_N = R)\). To understand this, the sign conventions have to be noticed: \(d_N\) is positive if the vertex of the surface is right to the nodal point \(N\), but \(d'_N\) is positive if the vertex of the surface is left to the nodal point \(N'\)!

3. For a convex surface \((R > 0)\) and \(n' > n\) the surface has a positive optical power and the focus \(F\) is in front of the surface and \(F'\) behind the surface. For a concave surface \((R < 0)\) but still \(n' > n\) the surface has a negative optical power and the foci change their positions, i.e. \(F\) is right to the vertex of the surface and \(F'\) is left to it.

Similar the focal lengths are calculated using the two equations (2.3.19):

\[
\begin{align*}
    f &= \frac{AD - BC}{C} = \frac{\text{Det}(M)}{C} = -\frac{nR}{n' - n} \\
    f' &= -\frac{1}{C} = \frac{n'R}{n' - n}
\end{align*}
\]  

Since, the principal points coincide with the vertex of the surface the focal length \(f\) is of course identical to \(f = -d_F\) and the focal length \(f'\) is \(f' = d'_F\). The general equation (2.3.25) \(f'/n' = -f/n\) is of course also valid.

2.4 The imaging equations of geometrical optics

2.4.1 The "lens equation"

On page 24 it has already been shown what imaging means. A point \(P_O\) lying in the plane \(P\) with a distance \(d\) in front of the vertex of the first surface of an optical system with the matrix \(M\) (see equation (2.3.3)) is imaged onto a point \(P_I\) in a plane \(P'\) with a distance \(d'\) to the vertex of the last surface of the system. This is only the case if the matrix element \(B'\) of the matrix \(M'\) (see equation (2.3.5)), which describes the complete ray propagation from \(P\) to \(P'\), is zero:

\[
B' = Ad + B + Cdd' + Dd' = 0
\]  

Then, all rays starting from the object point \(P_O\) intersect in the image point \(P_I\). The distance of the object point to the principal plane \(\mathcal{U}\) in the object space will be named \(d_O\) (dt.: Gegenstandsweite) and the distance of the principal plane \(\mathcal{U}'\) in the image space to the image point \(P_I\) will be \(d_I\) (dt.: Bildweite) (see figure 2.12). According to the sign conventions of geometrical optics \(d_O\) is positive if the object point is right to \(\mathcal{U}\) (i.e. \(d_O\) is negative in fig. 2.12) and \(d_I\) is positive if the image point is also right to \(\mathcal{U}'\) (i.e. \(d_I\) is positive in fig. 2.12).

Then, the relations between \(d\), \(d_O\) and \(d_U\) (distance between \(\mathcal{U}\) and vertex of first surface) on the one hand and \(d'\), \(d_I\) and \(d'_U\) (distance between vertex of last surface and \(\mathcal{U}'\)) on the other hand are:

\[
\begin{align*}
    d_O &= d_U - d \\
    d_I &= d' - d'_U
\end{align*}
\]
Figure 2.12: Parameters for explaining the imaging of an object point \( P_O \) to an image point \( P_I \) by a general optical system. The optical system is characterized by the vertices of the first and last surface and its cardinal points (without nodal points). The sign conventions mentioned in the text mean for the "classical" geometrical optical parameters: \( x_O > 0, x_I < 0, d_O < 0, d_I > 0, Z < 0, Z' > 0, f < 0, f' > 0 \). But for the other parameters which are only used in the paraxial matrix theory we have: \( d > 0, d' > 0, d_F > 0, d_{F'} > 0, d_U < 0, d_{U'} < 0 \).

Here, the different sign conventions for \( d_O \) (\( d_I \)) on the one hand and \( d_U \) (\( d_{U'} \)) and \( d' \) on the other are taken into account. By substituting equations (2.4.2) and (2.4.3) into equation (2.4.1) the following equation is obtained:

\[
A (d_U - d_O) + B + C (d_U - d_O) (d_I + d_{U'}) + D (d_I + d_{U'}) = 0 \quad (2.4.4)
\]

By using equations (2.3.8) and (2.3.9) to express \( d_U \) and \( d_{U'} \) as functions of the matrix elements \( A, B, C, D \) of \( M \) and some calculations the result is:

\[
A \left( \frac{D}{C} (A - 1) - B - d_O \right) + B + \left[ C \left( \frac{D}{C} (A - 1) - B - d_O \right) + D \right] \cdot \left( d_I + \frac{1 - A}{C} \right) =
\]

\[
= \frac{AD}{C} (A - 1) - AB - Ad_O + B + [AD - BC - Cd_O] \left( d_I + \frac{1 - A}{C} \right) =
\]

\[
= -Ad_O + B + ADd_I - BCd_I - Cd_Od_I - B - (1 - A)d_O =
\]

\[
= (AD - BC) d_I - Cd_Od_I - d_O =
\]

\[
= \text{Det}(M)d_I - Cd_Od_I - d_O = 0 \quad (2.4.5)
\]

The determinant of \( M \) is according to equation (2.3.24) \( \text{Det}(M) = n/n' \), where \( n \) is the refractive index in the object space and \( n' \) the refractive index in the image space. Additionally, according to equation (2.3.17) it is \( C = -1/f' \) with the focal length \( f' \) in the image space. So, the final
result is:

\[ d_O - \frac{n}{n'} d_I = \frac{d_O d_I}{f'} \quad \text{or} \quad \frac{d_O}{n} - \frac{d_I}{n'} = \frac{d_O d_I}{n f'} \]  \tag{2.4.6}

An equivalent formulation of this equation is the well-known imaging equation of geometrical optics (dt.: Abbildungsgleichung) which is often called the lens equation although it is valid for quite complex optical imaging systems:

\[ \frac{n'}{d_I} - \frac{n}{d_O} = \frac{n'}{f'} = \frac{n}{f} \]  \tag{2.4.7}

At the right side the equation (2.3.25) has been used.
If the refractive indices \( n \) and \( n' \) are identical the equation is:

\[ \frac{1}{d_I} - \frac{1}{d_O} = \frac{1}{f'} \]  \tag{2.4.8}

As defined above, the object distance \( d_O \) and the image distance \( d_I \) are measured in the lens equation relative to the principal planes.

### 2.4.2 Newton equation

Another formulation of the imaging equation is the Newton equation where the object distance and the image distance are measured relative to the focal points. Therefore, we define the distance between the focal point \( F \) in the object space and the object point \( P_O \) as \( Z \). Analogous, \( Z' \) is the distance from the focal point \( F' \) in the image space to the image point \( P_I \). Both quantities are again positive if the object/image point is right to the focus \( F/F' \). In fig. 2.12 \( Z \) is negative and \( Z' \) positive. From this figure, using the sign conventions and equations (2.4.2) and (2.4.3) it is clear that the following relations are valid:

\[ Z = d_F - d = d_F + d_O - d_U \quad \Rightarrow \quad d_O = Z + d_U - d_F = Z + f \]  \tag{2.4.9}

\[ Z' = d'_F - d' = d_I + d'_U - d'_F \quad \Rightarrow \quad d_I = Z' - d'_U + d'_F = Z' + f' \]  \tag{2.4.10}

where the equations (2.3.15) and (2.3.17) were used.
Substituting these equations into the lens equation (2.4.6) and using equation (2.3.25) gives

\[ \frac{Z + f}{n} - \frac{Z' + f'}{n'} = \frac{(Z + f)(Z' + f')}{n f'} \Rightarrow \]

\[ \frac{Z}{n} + \frac{f}{n} - \frac{Z'}{n'} - \frac{f'}{n'} = \frac{Z Z'}{n f'} + \frac{Z f}{n f'} + \frac{f}{n} \Rightarrow \]

\[ -\frac{f'}{n'} = \frac{Z Z'}{n f'} \]

and finally:

\[ ZZ' = f f' \]  \tag{2.4.11}

This is the well-known Newton equation for the imaging of an object point into an image point. The advantage of the Newton equation is its quite simple and symmetric form which does not explicitly depend on \( n \) and \( n' \). Of course, the dependence on the refractive indices in object and image space is hidden in the ratio of \( f \) and \( f' \).
2.4.3 Relation between lateral and longitudinal magnification

The Newton equation can also be easily explained by looking at figure 2.13. Due to the similar triangles the following relations are valid in the object space and the image space where the signs have to be noticed:

\[
\frac{x_O}{-Z} = \frac{x_I}{f} \Rightarrow \beta = \frac{x_I}{x_O} = -\frac{f}{Z} \\
-\frac{x_I}{Z'} = \frac{x_O}{f'} \Rightarrow \beta = \frac{x_I}{x_O} = -\frac{Z'}{f'}
\]

\[ZZ' = ff' \quad (2.4.12)\]

Here, the lateral magnification \( \beta \) defined by equation (2.3.1) has been used.

The longitudinal magnification (dt.: Tiefenmaßstab) is defined as \( \frac{dZ'}{dZ} \), i.e. the ratio of the axial (longitudinal) shift of the image plane to an axial shift of the object plane. According to the Newton equation (2.4.11) and the relation (2.3.25) between \( f \) and \( f' \) it holds:

\[
Z' = \frac{ff'}{Z} \Rightarrow \frac{dZ'}{dZ} = -\frac{ff'}{Z^2} = \frac{n'}{n} \left( \frac{f}{Z} \right) = \frac{n'}{n} \beta^2
\]

\[ (2.4.13) \]

This means that the longitudinal magnification is proportional to the square of the lateral magnification.

2.4.4 Some notes to the graphical image and ray construction in paraxial optics

In the last sections the rules for the graphical construction of the image of an object point and the graphical construction of a ray path were implicitly used in some figures (e.g. in fig. 2.12). Here, these well–known rules will be summarized again.
2.4. THE IMAGING EQUATIONS OF GEOMETRICAL OPTICS

Figure 2.14: Graphical construction of the image $P_I$ of an object point $P_O$ by using two rays. The upper figure shows the imaging at a positive lens, whereas the lower figure shows the imaging at a negative lens.

Figure 2.15: Graphical construction of the path of a ray which is refracted at an optical system by using the method of Listing.
Assume, that a paraxial optical system is characterized by its principal planes $U$ and $U'$, and by its focal points $F$ and $F'$. In the case of a positive lens (see the upper part of figure 2.14) the focal point $F$ is left to the principal plane $U$ and $F'$ is right to $U'$. For a negative lens (see the lower part of figure 2.14) $F$ is right to $U$ and $F'$ is left to $U'$. For the construction of the image point $P_I$ of an object point $P_O$ there are only two rays necessary: (i) a ray starting from $P_O$ parallel to the optical axis (dt.: Parallelstrahl), and (ii) a ray starting from $P_O$ which intersects the focal point $F$ in the object space (dt.: Brennpunktstrahl). The refracted ray (i) will then go through the focal point $F'$ in the image space, and the ray (ii) will be parallel to the optical axis in the image space. The transfer of the rays between the principal planes is of course horizontally because of the lateral magnification $+1$ between $U$ and $U'$. The point of intersection of the refracted rays (i) and (ii) is the image point $P_I$. These rules can be applied to the imaging with the help of a positive or a negative lens as it is shown in fig. 2.14. For the negative lens the image is virtual, i.e. at the image point there is not a real point of intersection of light rays, but the light seems to come from this point.

Another quite useful graphical method is the construction of the ray path of an arbitrary paraxial ray by the method of Listing (dt.: Listing'sche Strahlkonstruktion). In fig. 2.15 a ray is incident onto an optical system which is again characterized by the principal planes $U$ and $U'$, and the focal points $F$ and $F'$. To construct the ray path behind the optical system an auxiliary ray parallel to the original ray is drawn in such a way that it intersects the focal point $F$ in the object space. Then, we know that this auxiliary ray will become a ray parallel to the optical axis in the image space. Furthermore, we know that a bundle of rays which are parallel to each other in the object space will intersect in the focal plane in the image space. So, by drawing the point of intersection of the auxiliary ray with the focal plane in the image space, we also have a point on the path of our desired ray in the image space and we can construct it.

### 2.4.5 The Smith–Helmholtz invariant

The Smith–Helmholtz invariant, which has in the German literature the name Helmholtz–Lagrange invariant, connects the object height $x_O$, the aperture angle $\varphi_O$ and the refractive index $n$ in the object space with the respective values $x_I$, $\varphi_I$ and $n'$ in the image space of an optical system.

For the derivation figure 2.13 is used. The two rays starting from $P_O$ parallel to the optical axis and to the focal point $F$ form a pencil of rays with the aperture angle $\varphi_O$. These two rays form then the image point $P_I$ and include there the aperture angle $\varphi_I$. Due to our sign conventions for angles $\varphi_O$ and $\varphi_I$ have different signs in fig. 2.13 because they are defined as differences between the angles of the two spanning rays. So, taking into account the signs we see from the triangle formed by the ray through $F$ and the optical axis in the object space:

$$\tan \varphi_O = -\frac{x_I}{f}$$

From the triangle with the ray through $F'$ in the image space we obtain:

$$\tan \varphi_I = \frac{x_O}{f'}$$

By taking the paraxial approximation $\tan \varphi \approx \varphi$, combining both equations and using equation (2.3.25) we finally obtain the relation:

$$f = -\frac{x_I}{\varphi_O} = -\frac{n}{n'} f' = -\frac{n}{n'} \frac{x_O}{\varphi_I} \quad \Rightarrow \quad n x_O \varphi_O = n' x_I \varphi_I \quad \text{(2.4.14)}$$
35

\section*{2.5. THE THIN LENS}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{thin_lens.png}
\caption{Parameters of a thin lens.}
\end{figure}

So, it can be seen that the product of the object (or image) height, the aperture angle in the object (or image) space, and the refractive index in the object (or image) space is an invariant of the optical imaging system and this invariant is called the Smith–Helmholtz or Helmholtz–Lagrange invariant. If the system is in air \((n = n' = 1)\) it means for example that in the case of a scaling factor \(\beta = x_1/x_O\) the aperture angle in the image space has the value \(\varphi_I = \varphi_O/\beta\).

Later, we will see that the Smith–Helmholtz invariant is a paraxial approximation of the sine condition. It is also a paraxial approximation of the Herschel condition \([1]\) which will not be treated in this lecture.

\subsection*{2.5 The thin lens}

A quite important element in the paraxial theory is a so called \textbf{thin lens} \([8],[13],[14],[20]\). This means that the transfer from the first to the second surface is neglected (thickness \(d_1\) of the lens is assumed to be zero) and a paraxial ray which intersects the first surface at the ray height \(x\) has also immediately behind the second surface the same ray height \(x'\):

\[ x' = x \]

The refractive indices are \(n = n_1\) in front of the first surface, \(n_L := n'_1 = n_2\) between the two surfaces and \(n' = n'_2\) behind the second surface (see fig. 2.16). The radii of curvature of both spherical surfaces are \(R_1\) for the first surface and \(R_2\) for the second surface. The thin lens, as defined here, does of course not exist in reality but it is a good approximation for lenses which are "thin" compared to their focal length.

The matrix \(M_{L0}\) of a thin lens is obtained from equation (2.3.3) for \(m = 2\) and \(d_1 = 0\) (\(\Rightarrow M_{T,1}\) is the unit matrix) by multiplying the two matrices \(M_{S,2}\) and \(M_{S,1}\) for refraction at the two spherical surfaces in the correct order.

\[
M_{L0} = M_{S,2}M_{S,1} = \begin{pmatrix}
1 & 0 \\
-n'_{1}n'_{L} & n_{L}
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
-n_{1}n'_{L} & n_{L}
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
-n'_{1}n'_{L} & n_{L}
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
-n_{1}n'_{L} & n_{L}
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
-n'_{1}n'_{L} & n_{L}
\end{pmatrix}
\ \ \ \ \ (2.5.1)
\]
In the following the important case of identical external materials of the lens, i.e. $n' = n$, is considered. Then, the matrix is:

$$M_{L0} = \begin{pmatrix} \frac{n-n_L}{nR_2} & 0 \\ nL-n & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -nL-n & \frac{1}{R_1} - \frac{1}{R_2} \end{pmatrix}$$

$$\Rightarrow M_{L0} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f'} & 1 \end{pmatrix} \quad (2.5.2)$$

with

$$\frac{1}{f'} = \frac{n_L - n}{n} \left( \frac{1}{R_1} - \frac{1}{R_2} \right) \quad (2.5.3)$$

Here, the focal length $f'$ of a thin lens in the image space has been defined according to equation (2.3.19) and the focal length $f$ in the object space is of course $f = -f'$. By using equations (2.3.18) ($d_U = d'_{U'} = 0$) it can be seen that the principal points $U$ and $U'$ of the thin lens coincide with the vertices of the two surfaces which themselves coincide. Of course, the nodal points coincide also with the vertices because the nodal points coincide with the principal points due to $n = n'$.

In total the ray parameters $x', \varphi'$ immediately behind the thin lens are connected with the ray parameters $x, \varphi$ in front of the lens by

$$x' = x \quad (2.5.4)$$

$$\varphi' = \varphi - \frac{x}{f'} \quad (2.5.5)$$

For a lens with a positive focal power the focal length $f'$ is also positive and parallel incident rays intersect behind the lens in a real focus. For a lens with a negative focal power the focal length $f'$ is negative and this means that rays, which are originally parallel to the optical axis, would intersect in a so called virtual focus in front of the lens. Of course, a virtual focus has its name because there is in reality no focus at this position in front of the lens but the rays behind the lens seem to come from the virtual focus.

There are several different types of lenses depending on their radii of curvature:

- biconvex: $R_1 > 0$ and $R_2 < 0$
- plane–convex: $R_1 > 0$ and $R_2 = \infty$ (or $R_1 = \infty$, $R_2 < 0$)
- convex–concave (meniscus lens): $R_1 > 0$ and $R_2 > 0$ (or both negative)
- plane–concave: $R_1 < 0$ and $R_2 = \infty$ (or $R_1 = \infty$, $R_2 > 0$)
- biconcave: $R_1 < 0$ and $R_2 > 0$

These lenses have different focal powers. For the case $n_L > n$ (e.g. for a lens made of glass which is used in air) biconvex and plane–convex lenses have generally positive focal lengths, i.e. they are positive lenses. On the other side, biconcave and plane–concave lenses have negative focal lengths, i.e. they are negative lenses. Meniscus lenses can be either positive (if the convex surface has the smaller radius of curvature) or negative (if the convex surface has the larger radius of curvature). Pay attention to the fact that in the case $n_L < n$ (which can be realized e.g. by a hollow lens made of thin plastic which is filled with air and used in water) the properties of the different types of lenses are reverse. In this case a biconvex lens has e.g. a negative focal length.
2.6 The thick lens

In the case of a thick lens the ray transfer with thickness $d := d_1$ between the two spherical surfaces is taken into account. Of course, the radii of curvature of the two spherical surfaces are still assumed to be so large that the point of intersection of a paraxial ray with the surface is in the same plane as the vertex of the surface. The matrix $M_{Ld}$ of a thick lens is the product of three single matrices: matrix $M_{S,1}$ for refraction at the first spherical surface with radius of curvature $R_1$, matrix $M_{T,1}$ for the transfer between the two surfaces by the distance $d$ and matrix $M_{S,2}$ for the refraction at the second spherical surface with the radius of curvature $R_2$. The refractive indices in front of, in and behind the lens are $n = n_1$, $n_L := n'_1 = n_2$ and $n' = n'_2$, respectively. Then the matrix $M_{Ld}$ of a thick lens is:

$$M_{Ld} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
-n'_{L}n_{L} & n_{L} & 0 & 0 \\
-n'_{L}n_{L} & n_{L} & 0 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix} = \begin{pmatrix}
1 - \frac{n_L - n}{n_L R_1} & 0 & 0 & 0 \\
0 & 1 - \frac{n_L - n}{n_L R_1} & 0 & 0 \\
0 & 0 & 1 - \frac{n_L - n}{n_L R_1} & 0 \\
0 & 0 & 0 & 1 - \frac{n_L - n}{n_L R_1}
\end{pmatrix}$$

(2.6.1)
In the most important case of identical external materials \( n' = n \) equation (2.6.1) reduces to:

\[
M_{Ld} = \begin{pmatrix}
1 - \frac{n_L - n}{n_L R_1} d & \frac{n}{n_L} d \\
-\frac{n_L - n}{n} \left[ \frac{1}{R_1} - \frac{1}{R_2} + \frac{n_L - n}{n_L} \frac{d}{R_1 R_2} \right] & 1 + \frac{n_L - n}{n_L R_2} d
\end{pmatrix}
\]

(2.6.2)

The matrix element \( C \) in the first column of the second row is according to equation (2.3.19) defined as \(-1/f'\), where \( f' \) is the focal length of the thick lens in the image space:

\[
f' = \frac{n n_L R_1 R_2}{(n_L - n) [n_L (R_2 - R_1) + (n_L - n) d]} \tag{2.6.3}
\]

Because of \( n = n' \) the focal length \( f \) in the object space is \( f = -f' \) and the nodal points and the principal points coincide. So, it is now necessary to calculate the positions of the principal points \( U \) and \( U' \) (see fig. 2.19). By using the equations (2.3.18) it is obtained:

\[
d_U = \frac{D}{C} (A - 1) - B = \frac{AD - D - BC}{C} = \frac{1 - D}{C} = \]

\[
= \left( -\frac{n_L - n}{n} \left[ \frac{1}{R_1} - \frac{1}{R_2} + \frac{n_L - n}{n_L} \frac{d}{R_1 R_2} \right] \right. = \]

\[
= \frac{ndR_1}{n_L (R_2 - R_1) + (n_L - n) d} \tag{2.6.4}
\]

\[
d_{U'} = \frac{1 - A}{C} = \frac{n_L - n}{n_L R_1} \frac{d}{n_L R_1} \left[ \frac{1}{R_1} - \frac{1}{R_2} + \frac{n_L - n}{n_L} \frac{d}{R_1 R_2} \right] = \]

\[
= \frac{ndR_2}{n_L (R_2 - R_1) + (n_L - n) d}
\]
2.6. THE THICK LENS

The distance $d_{UU'}$ between the two principal planes, which is positive if $U'$ is right to $U$, is:

$$d_{UU'} = d + d_U + d_U' = d \left( 1 - \frac{n(R_2 - R_1)}{n_L(R_2 - R_1) + (n_L - n)d} \right)$$  \hspace{1cm} (2.6.5)

2.6.1 Thick lens in air

Since the special case of a thick lens in air ($n = 1$) is the most important in practice the equations (2.6.2) for $1/f'$, (2.6.4) for $d_U$ and $d_U'$, and (2.6.5) for $d_{UU'}$ shall be repeated for this case:

$$\frac{1}{f'} = (n_L - 1) \left[ \left( \frac{1}{R_1} - \frac{1}{R_2} \right) + \frac{n_L - 1}{n_L} \frac{d}{R_1 R_2} \right]$$ \hspace{1cm} (2.6.6)

$$d_U = \frac{d R_1}{n_L (R_2 - R_1) + (n_L - 1)d}$$ \hspace{1cm} (2.6.7)

$$d_U' = -\frac{d R_2}{n_L (R_2 - R_1) + (n_L - 1)d}$$ \hspace{1cm} (2.6.8)

$$d_{UU'} = d \left( 1 - \frac{R_2 - R_1}{n_L (R_2 - R_1) + (n_L - 1)d} \right)$$ \hspace{1cm} (2.6.9)

In the following three important cases of thick lenses in air will be described to illustrate the optical parameters of lenses.

2.6.2 Special cases of thick lenses in air

2.6.2.1 Ball lens

For a ball lens (dt.: Kugellinse) with radius of curvature $R > 0$ and refractive index $n_L$ the lens parameters are (see figure 2.20):

$$R_1 = R, \quad R_2 = -R, \quad d = 2R$$
This means according to equations (2.6.6)–(2.6.9) for the parameters in air:

\[
\begin{align*}
\frac{1}{f'} &= 2\frac{(n_L - 1)}{Rn_L} \Rightarrow f' = \frac{n_LR}{2(n_L - 1)} \\
\frac{1}{f'} &= \frac{2(n_L - 1)}{Rn_L} \Rightarrow f' = \frac{n_LR}{2(n_L - 1)} \\
d_U &= -R \\
d'_{U'} &= -R \\
d_{UU'} &= 0
\end{align*}
\]

(2.6.10)

This means that the principal points coincide and are at the center of curvature of the ball lens. For the special case \( n_L = 2 \) the focal length would be equal to the radius of curvature \( f' = R \) so that the focus in the image space would be on the backside of the sphere. For \( n_L < 2 \) (e.g. nearly all glasses) the focus is outside of the sphere, whereas for \( n_L > 2 \) (e.g. a silicon ball lens illuminated with infrared light) the focus would be inside of the sphere.

### 2.6.2.2 The meniscus lens of Hoegh

For the meniscus lens of Hoegh (dt.: Hoeghscher Meniskus) (see fig. 2.21) with refractive index \( n_L \) and thickness \( d \) the two radii of curvature are identical, i.e. \( R_1 = R_2 = R \). Then, equations (2.6.6)–(2.6.9) deliver:

\[
\begin{align*}
\frac{1}{f'} &= \frac{(n_L - 1)^2 d}{n_L R^2} \\
\frac{1}{f'} &= \frac{(n_L - 1)^2 d}{n_L R^2} \\
d_U &= \frac{R}{n_L - 1} \\
d'_{U'} &= -\frac{R}{n_L - 1} \\
d_{UU'} &= d
\end{align*}
\]

(2.6.11)

A thin meniscus with identical radii of curvature would have no optical effect. Contrary to this the thick meniscus of Hoegh has a positive optical power. At least one of the principal points
2.7. REFLECTING OPTICAL SURFACES

2.6.2.3 Plane–convex or plane–concave lenses

We assume now that the first surface of the thick lens with refractive index \( n_L \) and thickness \( d \) is curved (either convex, i.e. \( R_1 > 0 \), or concave) and the second is plane \( (R_2 = \infty) \). The equations (2.6.6)–(2.6.9) are in this case (see fig. 2.22):

\[
\begin{align*}
\frac{1}{f'} &= \frac{n_L - 1}{R_1} \\
d_U &= 0 \\
d_{U'} &= -d \frac{1}{n_L} \\
d_{UU'} &= d \left( 1 - \frac{1}{n_L} \right) = \frac{(n_L - 1)d}{n_L}
\end{align*}
\]

This means that the first principal point coincides with the vertex of the curved surface. Moreover, the focal length of a lens with one plane surface is calculated like the focal length of a thin lens. This is not astonishing since the plane–convex/plane–concave lens can be interpreted as a combination of a thin lens with focal length \( f' \) and a plane–parallel plate with thickness \( d \) and refractive index \( n_L \). This can be easily shown by calculating the matrix \( M = M_PM_L0 \) and comparing it with \( M_{Ld} \) of equation (2.6.2) for \( R_2 = \infty \).

2.7 Reflecting optical surfaces

Up to now only refracting surfaces have been treated which form lenses and complete objectives. But, there are of course also reflecting surfaces which are e.g. very important in astronomical telescopes [39] or which will be very important in the near future for optical lithography systems [26] in the EUV (extreme ultraviolet at a wavelength of 13 nm). However, a reflecting surface
can be easily included in a paraxial design by calculating its paraxial 2x2 matrix and including it instead of the surface of a refracting surface in equation (2.3.3). We will see that the determinant of the matrix of a reflecting surface is one so that our general discussions concerning the relation between the focal lengths \( f \) and \( f' \) are valid.

### 2.7.1 A plane reflecting surface

The reflection at a plane surface, which is perpendicular to the optical axis, is shown in fig. 2.23. The law of reflection means that the angle \( i' \) of the reflected ray with the surface normal is identical to the angle \( i \) of the incident ray, i.e. \( i = i' \). In the paraxial theory it is common practice not to take the reflected ray since then the light would travel from ”right to left”. Instead, the unfolded ray path is taken which is obtained by mirroring the reflected ray at the reflecting surface. By doing this the dashed ray in fig. 2.23 is obtained and there is no change of the paraxial ray parameters \( x \) and \( \varphi \). So, the paraxial ray matrix \( M_{RP} \) of a reflecting plane surface is just the unit matrix:

\[
M_{RP} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]  

(2.7.1)

Its determinant is of course one.

### 2.7.2 A spherical reflecting surface

The reflection at a spherical surface is treated analogously like in the case of a plane surface and is shown for a convex mirror in fig. 2.24. The ray which is reflected at the spherical surface is mirrored at a plane which goes through the vertex of the surface and is perpendicular to the optical axis. So, the dashed ray in figure 2.24 results. All angles in fig. 2.24 are positive so that the following relations are valid:

\[
\begin{align*}
    i &= \varphi + \alpha \\
    \alpha + i' &= \varphi' \\
    i &= i' \\
    \alpha &= x/R
\end{align*}
\]

\[
\Rightarrow \varphi' = \varphi + 2\alpha = \varphi + 2\frac{x}{R}
\]

(2.7.2)

Since the ray height \( x \) remains constant during reflection the paraxial ray matrix \( M_{RS} \) is:

\[
M_{RS} = \begin{pmatrix} \frac{1}{R} & 0 \\ 0 & 1 \end{pmatrix}
\]

(2.7.3)
Again the determinant is one. The matrix (2.7.3) is also valid for a concave mirror. There, the radius of curvature $R$ is negative so that the angle $\varphi'$ is smaller than the angle $\varphi$ for a positive ray height $x$. This is just the effect of a concave mirror with a positive optical power.

As an exercise the cardinal points of a spherical mirror shall be calculated by using equations (2.3.18) and (2.3.19):

$$
\begin{align*}
 d_U &= \frac{D}{C}(A-1) - B = 0 \\
 d_N &= \frac{1-D}{C} = 0 \\
 d_F &= \frac{D}{C} = -\frac{R}{2} \\
 d_{U'} &= \frac{1-A}{C} = 0 \\
 d_{N'} &= \frac{A}{C}(D-1) - B = 0 \\
 d_{F'} &= -\frac{A}{C} = -\frac{R}{2} \\
 f &= \frac{AD-BC}{C} = \frac{\text{Det}(M)}{C} = \frac{R}{2} \\
 f' &= -\frac{1}{C} = -\frac{R}{2}
\end{align*}
$$

(2.7.4)

So, the principal points $U$, $U'$ and the nodal points $N$, $N'$ all coincide with the vertex of the spherical mirror (see figure 2.25). The focus $F$ in the object space is at half the distance between the center of curvature of the spherical surface and the vertex. On the other side, the focus $F'$ in the image space would coincide with $F$ for the real reflected ray. But, since the unfolded ray...
CHAPTER 2. PARAXIAL GEOMETRICAL OPTICS

Figure 2.25: Cardinal points of a convex mirror. A ray coming from the left parallel to the optical axis has to go in the image space (virtually) through the focus $F'_{\text{mirrored}}$. $F'_{\text{mirrored}}$ is the focus of the unfolded ray path which is mirrored at the vertex plane. The real reflected ray would virtually go through the focus $F'$ which then coincides with $F$.

A tilted refractive plane surface or a diffraction grating which both introduce a global tilt of all rays can e.g. not be included in the 2x2 matrix theory. But there is an extension of this method by using 3x3 matrices [42]. This will be described in the following.

2.8 Extension of the paraxial matrix theory to 3x3 matrices

The paraxial 2x2 matrix theory can only be used as long as all elements are centered around the optical axis and symmetric with respect to the optical axis. A tilted refractive plane surface or a diffraction grating which both introduce a global tilt of all rays can e.g. not be included in the 2x2 matrix theory. But there is an extension of this method by using 3x3 matrices [42]. This will be described in the following.

2.8.1 Paraxial ray tracing at a diffraction grating

A ray representing a plane wave with wavelength $\lambda$ which hits a diffraction grating with a period $\Lambda$ is diffracted according to the well–known grating equation [1] (see fig. 2.26):

$$\sin \varphi' = \sin \varphi + m \frac{\lambda}{\Lambda}$$  \hspace{1cm} (2.8.1)

Here, the integer $m$ is the diffraction order of the grating and depending on the type of the grating there may be only one efficient order (e.g. for blazed gratings or volume holograms) or
many orders with non–vanishing efficiency (e.g. for binary phase elements or amplitude gratings) [15],[23],[43]. In the case of many diffraction orders each order has to be calculated separately. The angles \( \varphi \) and \( \varphi' \) are the angles of the incident and diffracted ray, respectively.

In the paraxial approximation the sine of the angles is replaced by the angle itself so that the grating equation is:

\[
\varphi' = \varphi + m \frac{\lambda}{\Lambda} \tag{2.8.2}
\]

Together with the equation for the ray height \( x \) \( (x' = x) \), which does not change in the case of diffraction at a grating, there are two equations relating the ray parameters before and after diffraction at the grating. But, it is no longer possible to write these two equations in a pure 2x2 matrix notation since it would be:

\[
\begin{pmatrix}
  x' \\
  \varphi'
\end{pmatrix} = \begin{pmatrix}
  1 & 0 \\
  0 & 1
\end{pmatrix} \begin{pmatrix}
  x \\
  \varphi
\end{pmatrix} + \begin{pmatrix}
  0 \\
  \frac{m\lambda}{\Lambda}
\end{pmatrix} \tag{2.8.3}
\]

So, the constant additive vector at the end would be necessary and the calculation of one 2x2 matrix for a complete optical system containing one or more diffraction gratings and several other optical elements would be impossible. But, there is a possibility to change this by using 3x3 matrices instead of 2x2 matrices and a paraxial ray vector with three components instead of two, where the third component is always 1. The 3x3 matrices and the paraxial ray vectors are of the form

\[
M_{3x3} = \begin{pmatrix}
  A & B & \Delta x \\
  C & D & \Delta \varphi \\
  0 & 0 & 1
\end{pmatrix} \Rightarrow \begin{pmatrix}
  x' \\
  \varphi'
\end{pmatrix} = M_{3x3} \begin{pmatrix}
  x \\
  \varphi
\end{pmatrix} = \begin{pmatrix}
  M \begin{pmatrix}
  x \\
  \varphi
\end{pmatrix} + \begin{pmatrix}
  \Delta x \\
  \Delta \varphi
\end{pmatrix}
\end{pmatrix} \tag{2.8.4}
\]

where \( M \) is the normal paraxial 2x2 matrix with the coefficients \( A, B, C \) and \( D \). The coefficients \( \Delta x \) and \( \Delta \varphi \) are constant values which symbolize a lateral shift or a tilt which is exerted on the
incident paraxial ray by the element. To obtain the 3x3 matrix appendant to a pure paraxial
2x2 matrix the coefficients $\Delta x$ and $\Delta \varphi$ just have to be set to zero.
The solution of our original example to define the paraxial 3x3 matrix $M_{G,3x3}$ of a (non–tilted)
diffraction grating is now quite easy:

$$M_{G,3x3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & m \chi \\ 0 & 0 & 1 \end{pmatrix}$$  \hspace{1cm} (2.8.5)

### 2.8.2 Tilted refractive plane surface

A refractive plane surface shall have a normal vector that is tilted by a small angle $\alpha$ with
respect to the optical axis. The surface with refractive indices $n$ in front of and $n'$ behind the
surface is hit by a paraxial ray with ray parameters $x$ and $\varphi$ (see figure 2.27). Since, the tilt
angle $\alpha$ has to be small and the ray heights $x$ also, the variation of the $z$–coordinates at the
points of intersection of the tilted surface and rays with different heights $x$ can be neglected:

$$\Delta z = x \tan \alpha \approx x \alpha \approx 0,$$

i.e. it is of second order and only first order terms are taken into account in the paraxial
approximation.

Additionally, the ray height $x$ remains constant for refraction. The ray angles, tilt angles and
refraction angles depend on each other by the following equations:

$$\varphi' = i' + \alpha$$
$$\varphi = i + \alpha$$
$$ni = n'i'$$

$$\Rightarrow \varphi' = \frac{n}{n'} i + \alpha = \frac{n}{n'} (\varphi - \alpha) + \alpha = \frac{n}{n'} \varphi + \frac{n' - n}{n'} \alpha$$  \hspace{1cm} (2.8.7)
2.8. EXTENSION OF THE PARAXIAL MATRIX THEORY TO 3X3 MATRICES

Figure 2.28: The refraction at a thin prism with prism angle $\gamma$. The incident ray is deflected by an angle $\delta$.

So, the 3x3 matrix $M_{R,\alpha,3x3}$ for refraction at a tilted plane surface is:

$$
M_{R,\alpha,3x3} = \begin{pmatrix}
1 & 0 & 0 \\
0 & \frac{n'}{n} & \frac{n' - n}{n} \alpha \\
0 & 0 & 1
\end{pmatrix}
$$

(2.8.8)

As an application and to see how the matrix of a complete system is determined the matrix of a thin prism will be calculated in the next paragraph.

2.8.3 Thin prism

A thin prism consists of two tilted refractive surfaces and we assume that the prism is made of a material with refractive index $n'$ and the refractive index outside of the prism is at both sides $n$. Since the prism is assumed to be thin the propagation between the two refractive surfaces is neglected and the total matrix $M_{\text{Prism},3x3}$ of the prism is obtained by just multiplying the 3x3 matrices of the single surfaces. The tilt angles of the two surfaces are $\alpha_1$ and $\alpha_2$ so that we have:

$$
M_{\text{Prism},3x3} = M_{R,\alpha_2,3x3} M_{R,\alpha_1,3x3} = \begin{pmatrix}
1 & 0 & 0 \\
0 & \frac{n'}{n} & \frac{n' - n}{n} \alpha_2 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 & 0 \\
0 & \frac{n'}{n} & \frac{n' - n}{n} \alpha_1 \\
0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & \frac{n' - n}{n} (\alpha_1 - \alpha_2) \\
0 & 0 & 1
\end{pmatrix}
$$

(2.8.9)

By defining the prism angle $\gamma := \alpha_1 - \alpha_2$ the total deflection angle $\delta$ of a thin prism with prism angle $\gamma$ is $\delta = (n' - n) \gamma / n$. For the most important case of a thin prism in air ($n = 1$) it is $\delta = (n' - 1) \gamma$.

2.8.4 The transformation matrices

The matrix of a tilted plane surface or other tilted and also laterally shifted surfaces can be calculated quite formally by introducing the paraxial transformation matrix between two coordi-
nate systems. The first coordinate system with the axes x and z will be named global coordinate system. The second coordinate system with axes x’ and z’ is called local coordinate system because in this coordinate system the surface will have a "simple" form, i.e. it is non–tilted and non–shifted in this local coordinate system. The local coordinate system is obtained from the global one by shifting a copy laterally in x–direction by the small distance $\Delta x$ and rotating it by an angle $\Delta \varphi$ (see figure 2.29). So, a paraxial ray with the ray parameters $(x, \varphi)$ in the global coordinate system has the ray parameters $(x', \varphi')$ in the local coordinate system and the following relations are valid:

\[
x' = (x - \Delta x) \cos(\Delta \varphi) \approx x - \Delta x \quad (2.8.10)
\]
\[
z' = (x - \Delta x) \sin(\Delta \varphi) \approx (x - \Delta x) \Delta \varphi \approx 0 \quad (2.8.11)
\]
\[
\varphi' \approx \varphi - \Delta \varphi \quad (2.8.12)
\]

Here, the paraxial approximations are used and since $\Delta x, \Delta \varphi, x$ and $\varphi$ are all paraxial (i.e. small) quantities, only terms in the first order are taken into account whereas terms of second order such as $(x - \Delta x)\Delta \varphi$ are set to zero. So, the $z'$–coordinate remains zero if the ray has a global coordinate $z = 0$ what is always the case by choosing the global coordinate system accordingly.

Therefore, the matrix $M_{G \rightarrow L,3x3}$ for the transformation of a paraxial ray from the global coordinate system to a local coordinate system is:

\[
M_{G \rightarrow L,3x3} = \begin{pmatrix} 1 & 0 & -\Delta x \\ 0 & 1 & -\Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} \quad (2.8.13)
\]

Vice versa, the matrix $M_{L \rightarrow G,3x3}$ for the transformation of a paraxial ray from the local coordi-
nate system to the global coordinate system is the inverse matrix to $M_{G \rightarrow L, 3 \times 3}$:

$$M_{L \rightarrow G, 3 \times 3} = M_{G \rightarrow L, 3 \times 3}^{-1} = \begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix}$$ (2.8.14)

It is also important to notice that in the paraxial approximation with small shifts $\Delta x$ and small angles $\Delta \varphi$ the order of shifting and tilting is arbitrary whereas this is not the case for finite quantities. Mathematically, this can be proved by calculating that the two matrices for a pure shift (i.e. $\Delta \varphi = 0$) and for a pure tilt (i.e. $\Delta x = 0$) permute:

$$\begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ (2.8.15)

This means, that in the paraxial approximation it is identical if the coordinate system is first tilted and afterwards shifted or if it is first shifted and then tilted. So, we can use one matrix for the whole transformation without taking care of the order of the single transformations.

As an application of the transformation matrices the 3x3 matrix for refraction at a tilted and laterally shifted spherical surface with a radius of curvature $R$ shall be calculated. The refractive indices are again $n$ in front of the surface and $n'$ behind it. The vertex of the spherical surface is laterally shifted by a distance $\Delta x$ with respect to the optical axis (global coordinate system) and the surface is rotated around an axis perpendicular to the meridional plane by an angle $\Delta \varphi$. The local coordinate system is of course that system in which the surface is neither tilted nor rotated.

Then, a ray in the local coordinate system can be calculated by multiplying the incident ray (in the global coordinate system) by the transformation matrix $M_{G \rightarrow L, 3 \times 3}$. In the local coordinate system, the ray is multiplied with the matrix of a normal non–tilted and non–shifted spherical surface $M_{S, 3 \times 3}$. Afterwards, the ray in the local coordinate system is transformed back into the global system by multiplying it with $M_{L \rightarrow G, 3 \times 3}$. So, the matrix $M_{S, \Delta x, \Delta \varphi, 3 \times 3}$ for refraction at a tilted and shifted spherical surface in the global coordinate system is just the product of the three matrices:

$$M_{S, \Delta x, \Delta \varphi, 3 \times 3} = M_{L \rightarrow G, 3 \times 3}M_{S, 3 \times 3}M_{G \rightarrow L, 3 \times 3} =$$

$$\begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{n'-n}{n' R} & \frac{n}{n'} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\Delta x \\ 0 & 1 & -\Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\Delta x \\ -\frac{n'-n}{n' R} & \frac{n}{n'} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \frac{n'-n}{n' R} & \frac{n}{n'} & \Delta x - \frac{n}{n'} \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ -\frac{n'-n}{n' R} & \frac{n}{n'} & \frac{n'-n}{n' R} \Delta x + \frac{n'-n}{n' R} \Delta \varphi \\ 0 & 0 & 1 \end{pmatrix}$$ (2.8.16)
The result shows that the ray height $x$ remains, as expected, unchanged by refraction at the surface ($x' = x$) and that there is for the ray angle $\varphi'$ besides the usual term of a spherical surface an additional term which does not depend on the angle of incidence but on the shift $\Delta x$ and the tilt $\Delta \varphi$. It can also be seen that this additional term is zero if the condition $\Delta x/R = -\Delta \varphi$ is fulfilled. This is the well-known fact that a lateral shift of a spherical surface can be compensated by tilting it.

A special case is $R \to \infty$ so that the spherical surface becomes a plane surface. In this case the matrix of equation (2.8.16) becomes:

$$M_{S,\Delta x,\Delta \varphi,3x3} \xrightarrow{R \to \infty} \begin{pmatrix} 1 & 0 & 0 \\ 0 & n' & n'-n \Delta \varphi/n' \\ 0 & 0 & 1 \end{pmatrix} = M_{R,\Delta \varphi,3x3} \quad (2.8.17)$$

This is of course the same result as for the matrix $M_{R,\alpha,3x3}$ for refraction at a tilted plane surface with $\Delta \varphi = \alpha$ which we obtained in equation (2.8.8) by deriving it directly from figure 2.27.
Chapter 3

Stops and pupils

In the preceding section about paraxial optics only rays and object points in the neighborhood of the optical axis have been considered. So, in the paraxial calculations stops have no influence. But this changes dramatically for the case of non–paraxial optics. There, stops are quite important optical elements which determine the light–gathering power of an optical system, its resolution, the amount of aberrations, its field and so on. In the following only some elementary definitions about stops and pupils can be given. For more information we refer to the literature [1],[13],[20],[35]. There are two especially important stops, the aperture stop and the field stop.

3.1 The aperture stop

Assume first of all a light emitting object point which radiates in all directions. Then, the aperture stop (dt.: Aperturblende oder Öffnungsblende) is that physical stop which limits the cross–section of the image–forming pencil of rays. To determine the aperture stop the size and position of the images of all stops (e.g. lens apertures or real stops) in the system by that part of the system which precedes the respective stop have to be calculated. To do this the paraxial matrix theory of the last section can for example be used. If the distance of the image of stop $i$ from the object point is $l_i$ and the diameter of the stop image is $d_i$, then the aperture angle $\varphi_i$ which can pass that stop is:

$$\tan \varphi_i = \frac{d_i}{2l_i} \quad (3.1.1)$$

The aperture stop is now that stop number $i$ which provides the minimum value $\varphi_O$ of $\varphi_i$. The image of the aperture stop made by that part of the optical system which precedes the aperture stop is called the entrance pupil (dt.: Eintrittspupille) and the image of the aperture stop made by that part of the optical system which follows the aperture stop is called the exit pupil (dt.: Austrittspupille). The full aperture angle $2\varphi_O$ is called the angular aperture on the object side (dt.: objektseitiger Aperturwinkel) and the corresponding quantity $2\varphi_I$ on the image side is called the angular aperture on the image side (dt.: bildseitiger Aperturwinkel). $\varphi_I$ can be determined by calculating the diameter $d_I$ of the exit pupil and the distance $l_I$ between the exit pupil and the image point and using again an equation like (3.1.1) by replacing $d_i$ with $d_I$ and $l_i$ with $l_I$.

If the aperture stop is in front of the optical system the aperture stop and the entrance pupil will be identical. Contrary, if the aperture stop is behind the whole optical system the aperture
Figure 3.1: Illustration of the aperture stop, entrance pupil, and exit pupil of an optical system calculated by ray tracing. The solid lines represent the real rays, whereas the dashed lines represent virtual extensions of the incident rays or of the rays leaving the optical system.
Figure 3.2: Difference between imaging with a normal single lens (top) and a telecentric imaging system (bottom). In this example, the telescopic system is telecentric in the object and image space. It can be seen that the two images with different object distances but identical object heights are at different lateral heights in the case of the single lens, whereas they are at the same height in the case of the telecentric system. Besides this, the chief rays are parallel to the optical axis in the case of the telecentric system.
stop and the exit pupil are identical. In the general case, where the aperture stop is somewhere in the optical system the entrance pupil and the exit pupil can also be somewhere and they can be real or virtual images of the aperture stop.

Another quite important definition of geometrical optics is the so called chief ray or principal ray (dt.: Hauptstrahl). This is that ray coming from the object point (which can of course be off–axis) which passes the center of the aperture stop. Since the entrance pupil and the exit pupil are both images of the aperture stop the chief ray also passes through the centers of entrance pupil and exit pupil (see fig. 3.1). If there are strong aberrations in the system this may not be exactly the case for object points which are strongly off–axis.

Fig. 3.1 shows the aperture stop, entrance pupil, and exit pupil of a concrete optical system which was calculated by ray tracing (see chapter 4). The entrance pupil and the aperture stop are nearly at the same position, but they have different sizes. The object point is off–axis so that it can be seen quite clearly that the grid of incident rays would hit the entrance pupil in a regular and centered grid. Similarly, the rays leaving the whole system seem to come from the exit pupil. The real rays which are refracted at the lenses hit the aperture stop in a regular and centered grid. Particularly, the chief ray hits the center of the aperture stop, of the entrance pupil, and of the exit pupil.

If an optical system consists of only one (thin) single lens the aperture stop, entrance pupil and exit pupil are of course all identical to the aperture of the lens itself. Another interesting case is e.g. an optical system where the aperture stop is in the back focal plane of the preceding part of the optical system. Then the entrance pupil is at infinity and the system is called telecentric on the object side (dt.: objektseitig telezentrisch). In this case all chief rays on the object side are parallel to the optical axis. Similarly, if the aperture stop is in the front focal plane of the part of the optical system which follows the aperture stop the exit pupil will be at infinity and the system is called telecentric on the image side (dt.: bildseitig telezentrisch). Optical systems which are telecentric on both sides are quite important in the optical metrology because in this case object points in different object planes have the same lateral magnification because the chief rays in object and image space are both parallel to the optical axis (see fig. 3.2). Therefore, the measured size of the object will be correct in a given image plane even though the object may be out of the object plane which is imaged sharply.

A quite important quantity to characterize an optical system is the numerical aperture \( NA \) (dt.: numerische Apertur). The numerical aperture \( NA_O \) on the object side is defined as

\[
NA_O = n_O \sin \varphi_O
\]  

and the numerical aperture \( NA_I \) on the image side is

\[
NA_I = n_I \sin \varphi_I
\]

where \( n_O \) and \( n_I \) are the refractive indices in the object and image space, respectively. It is an elementary property of optical imaging systems that \( NA_O \) and \( NA_I \) are connected by the lateral magnification \( \beta \) (see equation (2.3.1)) of the optical system if the sine condition is fulfilled:

\[
NA_I = \frac{NA_O}{\beta}
\]

In fact, by replacing \( \beta \) by the ratio \( x_I/x_O \) of the image size and the object size this equation can be written as

\[
x_I n_I \sin \varphi_I = x_O n_O \sin \varphi_O
\]
which is the usual formulation of the sine condition [1]. For the paraxial case this invariant reduces to the Smith–Helmholtz invariant (see equation (2.4.14)):

\[ x_{In} \phi_I = x_{On} \phi_O \] (3.1.6)

The numerical aperture determines how many light the optical system can gather from the object. It also determines (in the case of no aberrations) the resolution of the system due to diffraction. We will see in section 5.3 that many aberrations depend on the numerical aperture. The position of the aperture stop in an optical system also influences the aberrations [9].

3.2 The field stop

The second quite important stop is the field stop (dt.: Feldblende) which limits the diameter of the object field which can be imaged by an optical system. To find the field stop we calculate again the images of all stops (including possible stops in the object or image plane!) by that part of the optical system which precedes the respective stop. Let us assume that the image of stop number \( i \) has then again a diameter \( d_i \) and that the distance between the image of the stop and the entrance pupil of the system is \( L_i \). The field stop is then that stop which has the smallest value \( \phi_O \) of all values \( \phi_i \) with

\[ \tan \phi_i = \frac{d_i}{2L_i} \] (3.2.1)

The value 2\( \phi_O \) is called the field angle (dt.: objektseitiger Feldwinkel). The image of the field stop by that part of the optical system which precedes the field stop is called the entrance window (dt.: Eintrittsluke) and the image by that part of the optical system which follows the field stop is called the exit window (dt.: Austrittsluke).

If the line connecting an (off–axis) object point and the center of the entrance pupil is outside of the entrance window the chief ray cannot pass the field stop and so this object point cannot be imaged in most cases. However, there are cases where other rays coming from the object point can pass anyway and then there is no sharp border of the object field but the outer parts of the object field are imaged with lower intensity. This phenomenon is called vignetting (dt.: Vignettierung).

Two examples for the different positions of aperture stop and field stop in the case of a telescopic system are shown in figure 6.15 on page 100. There, also vignetting can be seen. In other examples, the field stop will be identical with the sensor (e.g. CCD chip or photographic plate).
Chapter 4

Ray tracing

It has been shown in section 1.3 that light can be described by rays as long as the approximation of geometrical optics is valid. The propagation of such rays through an optical system is a very important tool to develop optical systems and to calculate their expected quality. The propagation of light rays through an optical system is called ray tracing [45],[46] and it is the basic tool of optical design, i.e. the design and optimization of optical systems concerning their imaging quality or other properties (e.g. tolerance of a system against misalignments or fabrication errors of components). In this section the principle of ray tracing and some applications will be described. There is of course no room to discuss the basics of optical design itself. For this we refer to the literature [18],[19],[33],[36].

4.1 Principle

According to equation (1.3.4) a light ray propagates rectilinear in a homogeneous and isotropic material. At an interface to another material the ray is partially refracted and partially reflected depending on the property of the interface. If a material is inhomogeneous (e.g. in GRIN lenses or in air films with different temperatures) the light ray is curved during the propagation and the path of the ray has to be calculated by solving equation (1.3.2) in most cases numerically [40],[41]. However, in this section it is assumed that the optical system consists of different homogeneous materials which are separated by refracting or reflecting interfaces.

Ray tracing means that the path of a bundle of rays, which are e.g. emitted by an object point or form a plane wave (i.e. object point with infinite distance), is determined in an optical system (see e.g. figure 4.1 for tracing rays through an microscopic objective). In the approximation of geometrical optics the calculation is in this case exact and no other approximations, like e.g. paraxial approximations, are made. Since ray tracing can be easily automated with the help of computers it is nowadays the most important tool for designing lenses, telescopes and complete optical systems [18],[19],[33],[36]. For complex optical systems it is even today with the help of modern computers not possible to replace ray tracing by pure wave–optical methods. Moreover, for most macroscopic optical systems ray tracing in combination with wave–optical evaluation methods like the calculation of the point spread function [33], assuming that only the exit pupil of the system introduces diffraction, is a sufficiently accurate method to analyze imaging systems. Another quite modern application of ray tracing is the analysis of illumination systems with incoherent light. This will be discussed later shortly in section 4.8 about non–sequential
4.1. PRINCIPLE

Figure 4.1: Propagation of some light rays in a typical microscopic objective (NA=0.4, magnification 20x, focal length \( f' = 11.5 \text{ mm} \)) calculated with our internal software RAYTRACE. In this case the microscopic objective is used in the reverse order, i.e. to focus light.

Ray tracing.

A precondition for ray tracing is that the optical system is known very well. It is not sufficient to know some paraxial parameters but it is necessary to know the following data of the surfaces as well as the materials:

- Type of the surface like e.g. plane, spherical, parabolic, cylindrical, toric or other aspheric surface.

- Characteristic data of the surface itself like e.g. the radius of curvature in the case of a spherical surface or the aspheric coefficients in the case of an aspheric surface.

- Shape and size of the boundary of the surface like e.g. circular with a certain radius, rectangular with two side lengths or annular with an interior and an outer radius.

- Position and orientation of the surface in all three directions of space.

- Refractive indices of all materials in dependence on the wavelength.

The tracing of a given ray through an optical system has the following structure:

\( \text{a)} \) Determine the point of intersection of the ray with the following optical surface. If there is no point of intersection or if the hit surface is absorbing mark the ray as invalid and finish the tracing of this ray. Depending on the type of ray tracing it may also be necessary in this case to leave the ray unchanged and to go to d). If there is a point of intersection go to b).
b) Calculate the surface normal in the point of intersection.

c) Apply the law of refraction or reflection (or another law e.g. in the case of diffractive optical elements [44],[49]). Then, the new direction of the deflected ray is known and the point of intersection with the surface is the new starting point of the ray.

d) If there is another surface in the optical system go back to a) or if not then finish the tracing of this ray.

In the case of item a) the "following surface" can either be the physically next surface of the optical system which will be really hit by the ray (i.e. non-sequential ray tracing) or just the next surface in the computer list of surfaces where the order of the surfaces has been determined by the user (i.e. sequential ray tracing).

In the next sections the mathematical realization of the different steps of ray tracing will be described.

4.2 Mathematical description of a ray

A light ray (in a homogeneous material) can be described mathematically as a straight line with a starting point $p$ and a direction vector $a$ parallel to the ray. Here, $a$ is a unit vector, i.e. $|a| = 1$. According to equation (1.3.4) an arbitrary point on the ray with position vector $r$ is described by the equation:

$$ r = p + sa $$

The scalar parameter $s$ is the arc length on the ray, i.e. in this case of rectilinear rays it is just the distance between $r$ and $p$. The virtual part of the ray is described by $s < 0$ whereas that part where there is really light has $s \geq 0$. In practice, there is also a maximum value $s_{\text{max}}$ if the ray hits a surface where it is deflected.
4.3 Determination of the point of intersection with a surface

The determination of the point of intersection of a light ray described by equation (4.2.1) with a surface requires of course a mathematical description of a surface. It is well-known from mathematics that a surface can be described in an implicit form with a function $F$ fulfilling the equation

$$F(r) = 0 \quad (4.3.1)$$

Concrete examples will be given later. By combining equations (4.2.1) and (4.3.1) the determination of the point of intersection is mathematically equivalent to determine the roots of a function $G$ with the variable $s$:

$$G(s_0) := F(p + s_0 a) = 0 \quad (4.3.2)$$

After having determined the value $s_0$ at the root of $G$ the position vector $r_0$ of the point of intersection itself is obtained by applying $s_0$ to equation (4.2.1).

In many cases there can be several roots of $G$ and it is also necessary to check whether the point of intersection is in the valid part of the surface which is in practice limited by a boundary. Then, that root with the smallest positive value of $s_0$ lying in the valid part of the surface has to be taken. These queries can be quite complex in a computer program.

For general aspheric surfaces the solution of equation (4.3.2) will only be possible numerically. But for some simple cases the analytic solutions will be given in the following.

4.3.1 Plane surface

A plane surface can be described by the position vector $C$ of a point on the surface (typically this point is in the center of the plane surface) and by the surface normal $n_z$. Then, each point $r$ of the surface fulfills the equation

$$F(r) = (r - C) \cdot n_z = 0 \quad (4.3.3)$$

The solution of equation (4.3.2) is in this case:

$$(p - C) \cdot n_z + s_0 a \cdot n_z = 0 \quad \Rightarrow \quad s_0 = \frac{(C - p) \cdot n_z}{a \cdot n_z} \quad (4.3.4)$$
In the case \( \mathbf{a} \cdot \mathbf{n}_z = 0 \) there is no definite point of intersection with the surface.

Of course, equation (4.3.3) describes an unlimited surface whereas the surfaces of an optical system are limited. Therefore, it has to be checked whether the point of intersection is in the valid area of the surface.

For a circular surface with radius \( R \) and center \( \mathbf{C} \) this means e.g. that the point of intersection \( \mathbf{r}_0 \) has to fulfill the condition \(|\mathbf{r}_0 - \mathbf{C}| \leq R\).

For a rectangular surface a second vector \( \mathbf{n}_x \) (with \( \mathbf{n}_x \cdot \mathbf{n}_z = 0 \) and \(|\mathbf{n}_x| = 1\)) along one of the sides (the vector \( \mathbf{n}_y \) along the second side is then just \( \mathbf{n}_y := \mathbf{n}_z \times \mathbf{n}_x \)) and the side lengths \( l_x \) and \( l_y \) of the rectangle have to be defined, additionally. Then, it has to be checked whether the conditions \(|(\mathbf{r}_0 - \mathbf{C}) \cdot \mathbf{n}_x| \leq l_x/2 \) and \(|(\mathbf{r}_0 - \mathbf{C}) \cdot \mathbf{n}_y| \leq l_y/2 \) are fulfilled.

### 4.3.2 Spherical surface

A sphere with the position vector \( \mathbf{C} \) of the center of curvature and the radius \( R \) is described by the equation

\[
F(\mathbf{r}) = |\mathbf{r} - \mathbf{C}|^2 - R^2 = 0
\]

Therefore, equation (4.3.2) results in a quadratic equation for \( s_0 \):

\[
s_0^2 + 2s_0 (\mathbf{p} - \mathbf{C}) \cdot \mathbf{a} + |\mathbf{p} - \mathbf{C}|^2 - R^2 = 0
\]

The two solutions are

\[
s_0^{1,2} = (\mathbf{C} - \mathbf{p}) \cdot \mathbf{a} \pm \sqrt{[(\mathbf{C} - \mathbf{p}) \cdot \mathbf{a}]^2 - |\mathbf{C} - \mathbf{p}|^2 + R^2}
\]

where the superscript 1, 2 is an index marking the two solutions. Depending on the argument of the square root there exist no (if the argument is negative), one (if the argument is zero) or two (if the argument is positive) solutions.
4.3. DETERMINATION OF THE POINT OF INTERSECTION WITH A SURFACE

After having determined the points of intersection with the full sphere it has to be checked whether the points of intersection are in the valid part of the spherical surface. To do this an additional vector \( \mathbf{n}_z \) (\(|\mathbf{n}_z| = 1\)) along the local optical axis and the lateral diameter \( D \) of the surface have to be defined (see fig. 4.4). The radius of curvature \( R \) is positive if the vector \( \mathbf{n}_z \) points from the vertex to the center of curvature. In fig. 4.4 \( R \) is for example positive. By using some trigonometric relations it is easy to see that the condition

\[
\frac{(C - r_0) \cdot \mathbf{n}_z}{R} \geq \sqrt{1 - \frac{D^2}{4R^2}}
\]

has to be fulfilled by the point of intersection \( r_0 \) if it lies on the valid spherical surface.

4.3.3 General surface \( z = f(x,y) \)

There are many important surfaces in optics, e.g. aspheric surfaces, which are described by a function \( f \) and the equation \( z = f(x,y) \). The implicit formulation with the function \( F \) is then

\[
F(r) = z - f(x,y) = 0,
\]

with \( r = (x,y,z) \).

For a general function \( f \) the points of intersection of such a surface with a ray cannot be calculated analytically. But there are numerical methods such as Newton’s method combined with bracketing [37] to determine the roots of equation (4.3.2) where \( F \) of equation (4.3.7) is used.

An important case is e.g. the description of rotationally symmetric aspheric surfaces with their axis of rotation along \( z \) by using the function [33]:

\[
z = f(x,y) = f(h) = \frac{ch^2}{1 + \sqrt{1 - (K + 1)c^2h^2}} + \sum_{i=1}^{i_{\text{max}}} a_i h^i
\]

with \( h = \sqrt{x^2 + y^2} \). \( c = 1/R \) is the curvature of the conical part of the surface with the conic constant \( K \) (\( K < -1 \) for a hyperboloid, \( K = -1 \) for a paraboloid and \( K > -1 \) for ellipsoids with the special case \( K = 0 \) for a sphere). \( a_i \) are aspheric coefficients describing a polynomial of \( h \). In most cases, only coefficients with even integers \( i \geq 4 \) are used and \( i_{\text{max}} \) is in most cases less or equal to ten. But in modern aspheric surfaces there may also be odd terms of \( i \) and \( i_{\text{max}} > 10 \).

4.3.4 Coordinate transformation

In many cases there is a quite simple description of a surface in a local coordinate system (e.g. the description of a rotationally symmetric aspheric surface by using equation (4.3.8)) and it would not be useful to find the implicit function \( F \) in the global coordinate system if the surface is e.g. tilted. In these cases it is more useful to transform the ray parameters \( p \) and \( a \) from the global coordinate system to the local system. Then, the finding of the point of intersection with the surface and the refraction or reflection (or "diffraction" if the element is a diffractive optical element) are done in the local coordinate system. Afterwards, the new ray is transformed back into the global coordinate system.
Assume that the origin of the local coordinate system has the position vector $C$ in the global coordinate system and the three unit vectors along the coordinate axes of the local system are $n_x$, $n_y$ and $n_z$ in the global system (see figure 4.5). For the transformation between the position vector $p = (p_x, p_y, p_z)$ in the global system and $p' = (p'_x, p'_y, p'_z)$ in the local system there are the equations:

$$p = C + p'_x n_x + p'_y n_y + p'_z n_z \quad (4.3.9)$$

and

$$p'_x = (p - C) \cdot n_x$$
$$p'_y = (p - C) \cdot n_y$$
$$p'_z = (p - C) \cdot n_z \quad (4.3.10)$$

For the ray direction vector with coordinates $a' = (a'_x, a'_y, a'_z)$ in the local system and $a = (a_x, a_y, a_z)$ in the global system the analogous equations are valid (but with $C = 0$ because direction vectors are measured from the origin of the respective coordinate system and can be shifted arbitrarily):

$$a = a'_x n_x + a'_y n_y + a'_z n_z \quad (4.3.11)$$

and

$$a'_x = a \cdot n_x$$
$$a'_y = a \cdot n_y$$
$$a'_z = a \cdot n_z \quad (4.3.12)$$

Of course it would also be possible to write the coordinate transformation using 3x3 matrices with the vectors $n_x$, $n_y$ and $n_z$ as column or row vectors. But, we have preferred here the vector notation.
4.4 Calculation of the optical path length

The optical path length $L$ along a ray at the point of intersection with the next surface is calculated by adding to the original optical path length $L_0$ at the starting point $p$ of the ray the distance $s_0$ between the starting point of the ray and the point of intersection with the next surface multiplied with the refractive index $n$ of the material in which the ray propagates. Therefore, the optical path length is:

$$L = L_0 + ns_0 \quad (4.4.1)$$

If the optical path length on another point $r = p + sa$ on the ray has to be calculated this is done by just replacing $s_0$ in equation (4.4.1) by $s$.

4.5 Determination of the surface normal

If the function $F$ of the implicit representation of the surface is known the surface normal $N$ at the point of intersection is defined as the normalized gradient of $F$ at the point of intersection $r_0$

$$N = \frac{\nabla F}{|\nabla F|}, \quad (4.5.1)$$

Some examples of surface normals are given in the following.

4.5.1 Plane surface

It is according to equation (4.3.3):

$$F(r) = (r - C) \cdot n_z = 0$$

$$\Rightarrow N = n_z \quad (4.5.2)$$

4.5.2 Spherical surface

A spherical surface is described by equation (4.3.5):

$$F(r) = |r - C|^2 - R^2 = 0$$

$$\Rightarrow N = \frac{r_0 - C}{|r_0 - C|} \quad (4.5.3)$$

4.5.3 General surface $z=f(x,y)$

$$F(r) = z - f(x,y) = 0$$

$$\Rightarrow N = \left(-\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}, -\frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}, 1\right) \quad (4.5.4)$$

where $f_x := \partial f/\partial x$ and $f_y := \partial f/\partial y$ are the partial derivatives of $f$ at the point of intersection $r_0$ with the surface.
4.6 Law of refraction

For the ray tracing a vectorial formulation of the law of refraction is necessary. In equation (1.6.3) an implicit formulation of the law of refraction (and also of the law of reflection) has been given

$$\mathbf{N} \times (n_2 \mathbf{a}_2 - n_1 \mathbf{a}_1) = 0$$

where $n_1$ and $n_2$ are the refractive indices of the two materials and $\mathbf{a}_1$ and $\mathbf{a}_2$ are the unit direction vectors of the incident and refracted ray, respectively (see figure 4.6). $\mathbf{N}$ is the local surface normal at the point of intersection of the incident ray with the surface.

A solution of this equation can be found by the following steps:

$$\left( \mathbf{a}_2 - \frac{n_1}{n_2} \mathbf{a}_1 \right) \times \mathbf{N} = 0$$

This means that the term in round brackets has to be parallel to $\mathbf{N}$ or itself zero. The later case is only possible for $n_1 = n_2$ so that for $n_1 \neq n_2$ we have:

$$\mathbf{a}_2 = \frac{n_1}{n_2} \mathbf{a}_1 + \gamma \mathbf{N}$$

with a real value $\gamma$. By taking the square of both sides it is ($\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{N}$ are all unit vectors, i.e. $|\mathbf{a}_1| = |\mathbf{a}_2| = |\mathbf{N}| = 1$)

$$1 = \left( \frac{n_1}{n_2} \right)^2 + \gamma^2 + 2\gamma \frac{n_1}{n_2} \mathbf{a}_1 \cdot \mathbf{N}$$

and therefore

$$\gamma_{1,2} = -\frac{n_1}{n_2} \mathbf{a}_1 \cdot \mathbf{N} \pm \sqrt{1 - \left( \frac{n_1}{n_2} \right)^2 \left[ 1 - (\mathbf{a}_1 \cdot \mathbf{N})^2 \right]}$$

In total the result is:

$$\mathbf{a}_2 = \frac{n_1}{n_2} \mathbf{a}_1 - \frac{n_1}{n_2} (\mathbf{a}_1 \cdot \mathbf{N}) \mathbf{N} \pm \sqrt{1 - \left( \frac{n_1}{n_2} \right)^2 \left[ 1 - (\mathbf{a}_1 \cdot \mathbf{N})^2 \right]} \mathbf{N}$$

(4.6.1)
4.7 Law of Reflection

Figure 4.7: Parameters for the reflection of a ray at a surface.

The vector term in front of the square root is parallel to the surface (scalar product with $N$ is zero). This means that the sign in front of the square root decides whether the component of $a_2$ along $N$ is parallel or antiparallel to $N$. Since the ray is refracted the sign of the component of $a_1$ along $N$ has to be equal to the sign of the component of $a_2$ along $N$:

$$\text{signum}(a_1 \cdot N) = \text{signum}(a_2 \cdot N)$$

(4.6.2)

where signum is the sign function which is +1 for a positive argument and -1 for a negative argument.

Therefore, equation (4.6.1) can be written independent of the relative direction of $N$ with respect to $a_1$ as:

$$a_2 = \frac{n_1}{n_2} a_1 - \frac{n_1}{n_2} (a_1 \cdot N) N + \text{signum}(a_1 \cdot N) \sqrt{1 - \left( \frac{n_1}{n_2} \right)^2 \left[ 1 - (a_1 \cdot N)^2 \right]} N$$

(4.6.3)

So, this equation allows the calculation of the direction vector $a_2$ of the refracted ray if the incident ray (direction vector $a_1$), the local surface normal $N$ and the two refractive indices $n_1$ and $n_2$ are known.

4.7 Law of reflection

Also the law of reflection is formally described by equation (1.6.3) and therefore also by equation (4.6.1). But in the case of reflection first of all the refractive indices are identical for the incident and the reflected ray, i.e. $n_1 = n_2$, and second the component of $a_2$ along $N$ has the opposite sign as the component of $a_1$ along $N$ (see figure 4.7). This means that the other sign in front of the square root has to be taken and equation (4.6.1) results in:

$$a_2 = a_1 - (a_1 \cdot N) N - \text{signum}(a_1 \cdot N) \sqrt{(a_1 \cdot N)^2} N = a_1 - 2 (a_1 \cdot N) N$$

(4.7.1)

It is easy to prove that this equation describes correctly the reflection of a ray at a surface because all three vectors are lying in a common plane (linearly dependent vectors) and the angle
of incidence is equal to the angle of the reflected ray. The later can be seen by calculating the modulus of the cross product of equation (4.7.1) with \( N \). Third, \( a_2 \) describes really a reflected ray since double of the component of \( a_1 \) along \( N \) is subtracted from \( a_1 \) to obtain \( a_2 \).

Besides refraction and reflection there is also a third quite important law for deflecting a ray at a surface, the vectorial local grating equation which is used for the ray tracing on holographic and more general diffractive optical elements. But for this equation and its solution we refer to the literature [25],[27],[44],[49].

4.8 Non–sequential ray tracing and other types of ray tracing

The normal mode in most ray tracing computer programs is the so called sequential ray tracing, i.e. the user defines the order in which the different surfaces of the optical system are passed by a ray. But this method is e.g. not useful for the analysis of illumination systems where the path of a ray and the order of surfaces can be different for each ray. The stability analysis of laser resonators [22],[42] is also quite exhausting with the sequential mode because the user knows the order of the surfaces but not how many times they will be hit by a ray. Of course, a stable resonator will be crossed by a light ray with an infinite number of cycles. But for unstable resonators there is a finite number of cycles before the ray leaves the resonator.

Therefore, the non–sequential ray tracing is used in these cases. There, the computer calculates automatically the physically hit next surface for each ray. This is e.g. done by calculating the points of intersection of the ray with all surfaces and taking that surface with the smallest positive distance \( s_0 \). If there is no point of intersection with \( s_0 > 0 \) the ray does not hit any surface of the system. Of course, non–sequential ray tracing is quite extensive concerning the computing time and therefore it is normally only used if it is really necessary.

Another speciality of non–sequential ray tracing is that a ray can be split at a surface into a refracted and a reflected ray (and in the case of diffractive optical elements also in more than two rays representing the different diffraction orders). Each ray is then recursively traced through the optical system.

Some interesting modern optical systems such as Shack–Hartmann wavefront sensors [12] or beam homogenizers [6] use microlens arrays in combination with macroscopic optics. These array systems can also be analyzed with sequential or non–sequential ray tracing to obtain a first insight [28],[29]. Of course, one has to be careful in these cases because diffraction and interference effects (for coherent or partially coherent illumination) may not be neglected in several cases [4].

Sophisticated modern computer programs for sequential or non–sequential ray tracing implement in addition polarization ray tracing [7],[47]. There, the local polarization state of each ray is taken into account and for example the split–up of the local power transported by each ray to the refracted and reflected ray for refraction/reflection at a surface is done according to the Fresnel equations [1].

A third type of ray tracing is the so called differential ray tracing or generalized ray tracing [21],[27],[46]. In this case, each ray is assumed to represent a local wave front with two principal curvatures and two principal directions. These parameters are then traced additionally to the normal ray parameters for each ray during the propagation through the optical system. This allows for example the calculation of the local astigmatism of the wave front belonging to the ray by just tracing one ray. It allows also to calculate the change of the local intensity of
the wave during the propagation.
Chapter 5

Aberrations

Whereas in the paraxial case the imaging quality of an optical system is ideal there are in practice aberrations (dt.: Aberrationen oder Abbildungsfehler) of an optical system which deteriorate its imaging quality [1],[9],[32],[35],[48]. To explain the nature of aberrations look at figure 5.1. At the exit pupil of an optical system there is a real wave front (solid line), i.e. the surface of equal optical path length, which intersects the exit pupil on the optical axis and which has its paraxial focus at the point P which lies in the image plane of the system (in figure 5.1 we assume that the object plane is at infinity so that the image plane is identical to the focal plane). But there are deviations between an ideal spherical wave front (dashed line) with the center of curvature at P and the real wave front in the non–paraxial region. So, a ray starting at the point \((x', y')\) of the exit pupil has between the real wave front and the ideal spherical wave front an optical path length difference called the \textbf{wave aberration} \(W(x', y')\). Additionally, a ray with aberrations does not intersect the image plane in the focus P but at a point with the lateral distances \(\Delta x\) and \(\Delta y\) in x and y–direction. These lateral deviations from the paraxial focus P are called the \textbf{ray aberrations}. Of course, the wave aberrations and the ray aberrations are not independent of each other (see [38]) and with good approximation the ray aberrations are proportional to the partial derivatives of the wave aberrations with respect to \(x'\) and \(y'\):

\[
\Delta x \approx R \frac{\partial W}{\partial x'} \\
\Delta y \approx R \frac{\partial W}{\partial y'}
\]  \(\text{(5.0.1)}\)

Here, \(R\) is the distance from the exit pupil to the image plane.

5.1 Calculation of the wave aberrations

The wave aberrations (dt.: Wellenaberrationen) can be calculated by ray tracing. To do this a sphere which intersects the exit pupil on the optical axis and which has the (paraxial) focus P as center of curvature is defined. Then, the optical path lengths \(L(x', y')\) of the points of intersection of this reference sphere with rays starting at the exit pupil at the points \((x', y')\) are calculated by using equations (4.3.6) and (4.4.1). The optical path length \(L(0, 0)\) of the chief ray is subtracted from the optical path length values of all other rays resulting in the wave aberrations \(W(x', y')\):

\[
W(x', y') = L(x', y') - L(0, 0)
\]  \(\text{(5.1.1)}\)
5.1. **CALCULATION OF THE WAVE ABERRATIONS**

Figure 5.1: Explanation of the wave aberrations $W$ and the ray aberrations $\Delta x, \Delta y$. The solid curve is the real wave front and the dashed curve the ideal spherical wave front. The solid rays are rays starting from the real wave front whereas the dashed rays are rays starting from the ideal spherical wave front. $P$ is the (paraxial) focus of the wave front.
Figure 5.2: Wave aberrations for the on–axis point of the microscopic objective of figure 4.1 (NA=0.4, magnification 20x, focal length \(f'=11.5\) mm). The reference sphere is around the best focus of the wave aberrations. The focused spot is diffraction–limited since the peak–to–valley value is just 0.1 wavelengths (\(\lambda=587.6\) nm).

So, the wave aberrations are known for a grid of rays, i.e. points \((x', y')\) in the exit pupil. In some cases it is useful not to take the paraxial focus for \(P\) but the so called ”best focus”. This is that point where either the wave aberrations or the ray aberrations have the smallest mean value (see figure 5.2). So, in fact there are two different definitions of the ”best focus”. If there is e.g. field curvature the best focus will not be in the focal plane but on a sphere which intersects the optical axis in the focal plane.

### 5.2 The ray aberrations and the spot diagram

The ray aberrations (dt.: Strahlaberrationen oder Queraberrationen) can also be calculated by ray tracing. They are just the lateral deviations \(\Delta x\) and \(\Delta y\) between the focus \(P\) itself (which can be the paraxial focus or the best focus) and the points of intersection of the rays and a plane through the focus \(P\). The surface normal of this plane is assumed to be \(n_z\) (in most cases \(n_z\) will be parallel to the optical axis) and the focus \(P\) has the position vector \(P\). Additionally, the two unit vectors \(n_x\) and \(n_y\) lying in this plane and defining the local x– and y–axis are known (\(n_x, n_y\) and \(n_z\) form an orthogonal triad of unit vectors). Then, a ray number \(i\) with the starting point \(p_i\) and the direction vector \(a_i\) has its point of intersection \(r_i\) with the plane according to equations (4.2.1) and (4.3.4) at:

\[
r_i = p_i + \frac{(P - p_i) \cdot n_z}{a_i \cdot n_z} a_i
\]

The ray aberrations are then defined as:

\[
\Delta x = (r_i - P) \cdot n_x
\]

\[
\Delta y = (r_i - P) \cdot n_y
\]

A quite demonstrative representation of the ray aberrations is a **spot diagram**. There, the points of intersection of the rays with a plane are graphically displayed by just drawing them.
5.3. THE SEIDEL TERMS AND THE ZERNIKE POLYNOMIALS

Figure 5.3: Spot diagram for the on–axis point of the microscopic objective of fig. 4.1 with NA=0.4, magnification 20x, focal length \( f' = 11.5 \) mm. Since the numerical aperture of the lens used to focus the light is NA=0.4 on the image side and the wavelength is \( \lambda = 587.6 \) nm the diffraction–limited airy disc would have a diameter of \( 1.22\lambda/NA = 1.8 \) µm, i.e. larger than the ray aberrations. So, like in figure 5.2 it can also be seen from the ray aberrations that the on–axis spot of this lens is diffraction–limited.

as points (see figure 5.3). This means that the spot diagram is a graphical representation of the ray aberrations \((\Delta x, \Delta y)\). Sometimes it is useful to determine the spot diagram not only in a plane through the focus but also in other planes to track the focussing of the rays.

5.3 The Seidel terms and the Zernike polynomials

In classical aberration theory [1],[13] the primary aberration terms of Seidel (fourth order wave aberration terms or third order ray aberration terms) play an important role. The different terms are: spherical aberration, coma, astigmatism, curvature of field and distortion. Whereas, the first three terms are point aberrations, i.e. aberrations which generate a blurred image point, the last two terms just cause a shift of the image point relative to the ideal paraxial image point but the image point itself would be sharp. It is no time in this chapter to go into details and to give a mathematical derivation so that only some facts will be stated to the different aberration terms. The distance of the object point from the optical axis will be called in the following the object height \( r_O \) whereas the distance of a ray from the optical axis in the exit pupil will be called \( r_A \) (from aperture height). For lenses with a small numerical aperture the maximum value \( r_A \) is proportional to the numerical aperture NA. Therefore, in the following the numerical aperture NA and the object height \( r_O \) will be used to describe the functionality
of the different Seidel terms.

5.3.1 Spherical aberration

The spherical aberration (dt.: sphärische Aberration oder Öffnungsfehler) is the only classical aberration which occurs also for object points on the optical axis of a rotationally symmetric optical system, i.e. for \( r_O = 0 \). The spherical aberration of a normal single lens causes that rays with a large height \( r_A \) in the exit pupil of the lens are refracted stronger so that they intersect the optical axis in front of the paraxial focus. In general optical systems the off-axis rays can also intersect the optical axis behind the paraxial focus. A typical property of spherical aberration is that it increases with the fourth power of the numerical aperture \( NA \) of the ray pencil forming the image point:

\[
\text{spherical aberration} \propto NA^4 \quad (5.3.1)
\]

As already mentioned above the spherical aberration is independent of the object height \( r_O \).

5.3.2 Coma

Coma (dt.: die Koma) is an aberration which occurs only for off-axis points (of a rotationally symmetric optical system), i.e. \( r_O \neq 0 \). The name coma is caused by the deformation of the image point which looks like the coma of a comet. The coma depends on the third power of the numerical aperture and linearly on the image height:

\[
\text{coma} \propto r_O NA^3 \quad (5.3.2)
\]

This is the reason why coma occurs especially for large numerical apertures whereas astigmatism dominates for small numerical apertures and large object heights (see next paragraph).

Coma can for example be generated in the microscopic objective of figure 4.1 by a lateral shift of the first lens. A shift of 0.1 mm results in the aberrations of figure 5.4 which are dominated by coma although the spherical aberration of the original lens is still present.
5.3. THE SEIDEL TERMS AND THE ZERNIKE POLYNOMIALS

Figure 5.5: Wave aberrations for an off-axis object point of the microscopic objective of figure 4.1 (NA=0.4, magnification 20x, focal length $f'=11.5$ mm). The image point shows in this case mainly astigmatism but of course mixed with the spherical aberration of figure 5.2. There is nearly no coma because the well-adjusted microscopic objective fulfills the sine condition.

5.3.3 Astigmatism

Astigmatism (dt.: Astigmatismus) means that rays of the meridional plane and of the sagittal plane focus in different planes perpendicular to the optical axis. So, the geometrical shape of the image point is in general an ellipse. In two special planes, called meridional and sagittal focal plane, the ellipses degenerate into two focal lines. The focal lines are perpendicular to each other. Between the meridional and the sagittal focal plane there is another plane where the shape of the image point is a circle. But of course, this circle is extended whereas an ideal image point in geometrical optics would be a mathematical point. The astigmatism of an optical system is proportional to the square of the numerical aperture and the square of the object height:

$$\text{astigmatism} \propto r_O^2 \text{NA}^2$$  \hspace{1cm} (5.3.3)

As mentioned above, this functionality is the reason that astigmatism occurs also for quite narrow pencils of rays. If there are cylindrical or toric surfaces in an optical system astigmatism occurs also on the optical axis whereas in the usual case of rotationally symmetric optical systems astigmatism occurs only for off-axis points.

If we take again the microscopic objective of figure 4.1 but now with an off-axis object point (object height 15 mm, resulting image height 0.74 mm because of curvature of field) the resulting aberrations are mainly astigmatism showing the typical saddle shape. Of course, the spherical aberration which is present on-axis remains so that in fact the resulting aberrations represented in figure 5.5 are a mixture of about one wavelength peak-to-valley value astigmatism and 0.1 wavelength spherical aberration. Nearly no coma appears for off-axis points because a microscopic objective fulfills the sine condition (see equation (3.1.5) on page 54) which guarantees that object points in the neighborhood of the optical axis are imaged without coma.
5.3.4 Curvature of field

As mentioned above the curvature of field (dt.: Bildfeldkrümmung) is not a point aberration but a field aberration, i.e. the image point can be sharp but the position of the image point is shifted relative to the ideal paraxial value. In the case of the curvature of field the image points are situated on a spherical surface and in connection with astigmatism there are even two different spheres for rays in the meridional plane and in the sagittal plane.

Figure 5.6 shows the curvature of field in the image plane of the microscopic objective of figure 4.1. For the off-axis points the best focus of the image points is behind the focal plane (light is coming as usual from left). Of course, the off-axis points show also astigmatism so that the image points are blurred.

5.3.5 Distortion

The last Seidel term is distortion (dt.: Verzeichnung) which is also a field aberration and not a point aberration. It means that the lateral magnification for imaging is not a constant for all off-axis points but depends to some extent on the object height \( r_O \). The result is that each straight line in the object plane which does not pass the optical axis is curved in the image plane. A regular grid like in figure 5.7 b) is either pincushion distorted (see a)) or barrel distorted (see...
5.4. CHROMATIC ABERRATIONS

Figure 5.7: Effect of distortion. The regular grid of b) in the object plane is either distorted in the image plane to a pincushion shape a) or a barrel shape c). In a) the lateral magnification increases with increasing object height whereas it decreases in c).

5.3.6 The Zernike polynomials

A quite important method to calculate the different terms of the wave aberrations of an optical system is to fit the so called Zernike polynomials [1],[17],[24] to it. The wave aberration data for this procedure can either be theoretically determined, e.g. by ray tracing, or experimentally determined, e.g. by interferometry.

The condition for using Zernike polynomials is that the aperture of the optical system is circular because the Zernike polynomials are only orthogonal on the unit circle. There, they build a complete set of orthogonal polynomials and some of the terms correspond to the classical Seidel terms for the point aberrations, i.e. spherical aberration, coma and astigmatism. Besides this there are other terms such as trifoil or tetrafoil which result for example if the optical elements are stressed by fixing them at three or four points. It has to be emphasized that there are no terms corresponding to the Seidel terms curvature of field and distortion since the Zernike polynomials can only represent point aberrations and no field aberrations.

5.4 Chromatic aberrations

Up to now it was implicitly assumed that only light of one wavelength is considered and the presented aberrations were all monochromatic aberrations. Besides this, there are so called chromatic aberrations (dt.: Chromatische Aberrationen oder Farbfehler) which are a result of the dispersion of a material, i.e. the dependence of the refractive index of a material on the wavelength (or if there are diffractive optical elements in the system the dispersion results from the strong dependence of the grating equation on the wavelength). The dispersion changes the paraxial parameters like the focal length of a lens. For a thin lens with refractive index $n$ in air we have for example (see equation (2.5.3)):

$$\frac{1}{f'} = (n - 1) \left( \frac{1}{R_1} - \frac{1}{R_2} \right)$$

(5.4.1)
CHAPTER 5. ABERRATIONS

If now \( n \) depends on the wavelength \( \lambda \) we have:

\[
\frac{d}{d\lambda} \left( \frac{1}{f'} \right) = -\frac{df'/d\lambda}{f'^2} = \frac{dn}{d\lambda} \left( \frac{1}{R_1} - \frac{1}{R_2} \right) = \frac{dn/d\lambda}{n-1} \frac{1}{f'}
\]

(5.4.2)

By replacing the differentials by finite differences we can write with a good approximation:

\[
\frac{\Delta f'}{f'} = -\frac{\Delta n}{n-1}
\]

(5.4.3)

To characterize the dispersion of a material the so called **Abbe number** \( V_d \) is used which is defined as:

\[
V_d = \frac{n_d - 1}{n_F - n_C} = \frac{n(\lambda_d = 587.6 \text{ nm}) - 1}{n(\lambda_F = 486.1 \text{ nm}) - n(\lambda_C = 656.3 \text{ nm})}
\]

(5.4.4)

So, we have with a good approximation:

\[
\frac{f'(\lambda_C = 656.3 \text{ nm}) - f'(\lambda_F = 486.1 \text{ nm})}{f'(\lambda_d = 587.6 \text{ nm})} = \frac{1}{V_d}
\]

(5.4.5)

For glasses with normal dispersion the Abbe number is a positive constant which has a small value for materials with high dispersion (e.g. materials like SF10) and a large value for materials with small dispersion (like e.g. BK7). The positive sign indicates that the focal length of a lens increases with increasing wavelength.

It has to be mentioned that the definition of the Abbe number by equation (5.4.5) is only an approximation which results for refractive lenses in an error of about 1-2%. The exact definition using the focal lengths at different wavelengths is according to equation (5.4.4) using (5.4.1):

\[
V_d = \frac{n_d - 1}{n_F - n_C} = \frac{n_d - 1}{(n_F - 1) - (n_C - 1)} = \frac{1/f_d'}{1/f_F' - 1/f_C'}
\]

(5.4.6)

Here, \( f_d' \), \( f_F' \) and \( f_C' \) are the focal lengths at the respective wavelengths with the same indices. To calculate for example the Abbe number of a diffractive lens (see next chapter) the exact definition has to be taken. Using the approximation of equation (5.4.5) would result in an error of about 10% for a diffractive lens!
Chapter 6

Some important optical instruments

In this section some important optical elements and instruments such as the achromatic lens, the camera, the human eye, the telescope and the microscope will be discussed. However, this will be done in some cases quite shortly because there are many textbooks of geometrical optics which treat these subjects quite ample [13],[31],[35]. We will start with a diffractive lens which is not so widely known like a refractive lens.

6.1 The diffractive lens

A diffractive lens which is also known as Fresnel zone lens (FZL) is an optical element which is based on the diffraction of light at (nearly) periodic structures. A more general collective term for elements based on diffraction is diffractive optical element (DOE). DOEs comprise for example gratings, diffractive lenses, diffractive beam shaping elements, and so on.

A Fresnel zone lens must not be mistaken for a Fresnel lens, although they are related to each other. A Fresnel lens is based on refraction of light at locally prism–like segments of a lens. However, the height of such segments is for a Fresnel lens many wavelengths. Fresnel lenses are known in daily life as thin lenses (thin in the meaning of less than 1 mm which is still more than 1000 wavelengths) which are for example stuck onto the rear window of cars to enhance the clear view. They are also known from light houses where very large lenses are needed to collect the light. Then, a Fresnel lens needs much less material as a normal full lens.

6.1.1 Formation of a Fresnel zone lens from a full lens

If now the height $\Delta h$ of the single segments of a Fresnel lens is such that it causes a relative difference in the optical path lengths of one wavelength (or a small integer number of wavelengths) between neighbored segments, a so called blazed Fresnel zone lens (i.e. element with a continuous surface structure) is formed. Figure 6.1 shows how the formation of a Fresnel lens from a full lens can be imagined and if, as mentioned above, the height $\Delta h$ is $\lambda/(n-1)$ (refractive index $n$ of the element which is assumed to be in air) it is called a blazed FZL. By further simplifying the blazed element to a binary element with height $\Delta h = \lambda/(2(n-1))$ and equal size of bar and rill (strictly this equality is only valid at the rim of the FZL where the grating frequency is high) a Ronchi–type binary FZL (see last part of figure 6.1) is obtained. Whereas the blazed element has for the design wavelength $\lambda$ theoretically 100% diffraction efficiency in the first order (which is then identical to the light which would also result by the law of refraction), the
binary element with the height $\Delta h = \lambda/(2(n-1))$ has only about 40% diffraction efficiency in the desired first order (exactly it is $4/\pi^2 = 40.5\%$). However, due to symmetry reasons there is also a minus first diffraction order with 40.5% efficiency. This means in the case of a FZL that there are a convergent and a divergent spherical wave (for an incident plane wave) with the same modulus of radius of curvature but opposite signs and both have 40.5% efficiency. The zeroth diffraction order as well as all even diffraction orders have zero diffraction efficiency in this special case of a binary Ronchi–type FZL with adapted height. If the height is different (or equivalently if the actual wavelength deviates from the design wavelength) there is also light in the zeroth diffraction order passing just through the element without being diffracted. If the duty cycle, i.e. the ratio of the widths of bar and rill of the binary element, is different from 1:1 there will also be even diffraction orders. So, the diffraction efficiency of 40.5% in the first orders is the maximum value which can be obtained with a binary FZL. If there are deviations from the design wavelength or fabrication errors the diffraction efficiency in the first orders will always be smaller.

### 6.1.2 Grating model of a Fresnel zone lens

In the last subsection the formation of a Fresnel zone lens from a full lens has been explained. For the design wavelength the deflection of the light at a blazed FZL is in the same direction as for a ray which is refracted at the respective segment. However, this model does not explain the behavior of the FZL for an arbitrary wavelength or for the case of a binary element. So, another model based on diffraction has to be found. In the grating model the FZL is just assumed to behave locally like an infinitely extended diffraction grating with the local grating period and grating vector orientation of the FZL. Then, the well–known grating equation (2.8.1) explains how the light is locally deflected in the respective diffraction order $m$:

$$\sin \varphi' = \sin \varphi + m \frac{\lambda}{\Lambda} \quad (6.1.1)$$

Here, $\varphi$ is the angle of the incident light ray (local plane wave) with the grating normal and $\varphi'$ the angle of the diffracted ray with the grating normal (see figure 6.2). $\Lambda$ is the local grating
6.1. THE DIFFRACTIVE LENS

Figure 6.2: Explanation of the grating equation: Only if the optical path length difference between two neighbored periods is an integer multiple of the wavelength there is positive interference.

period and \( \lambda \) the wavelength of the used light.
This equation follows from simple wave–optical considerations by calculating the optical path length differences of plane waves propagating along the direction of the rays. There will only be positive interference if the optical path length difference between two neighbored periods at the same position in each period is an integer multiple \( m \) of the wavelength \( \lambda \) (see figure 6.2):

\[
\Lambda \sin \varphi' - \Lambda \sin \varphi = m \lambda \quad \Rightarrow \quad \sin \varphi' = \sin \varphi + m \frac{\lambda}{\Lambda}
\]  

Of course, for a finite grating with only a few periods the diffraction orders will not be quite sharp. But, assuming an infinite grating there is only light in the direction given by equation (6.1.2).

Similar, for a Fresnel zone lens there is a local grating period which changes slightly from period to period. For a FZL with a small numerical aperture the grating frequency (reciprocal value of the grating period) \( \nu = \frac{1}{\Lambda} \) increases linearly with the distance \( r \) from the optical axis to achieve focusing of an incident plane wave (\( c \) is a constant with the physical dimension mm\(^{-2} \)):

\[
\nu (r) = \frac{1}{\Lambda (r)} = cr \quad \Rightarrow \quad \varphi' = \varphi + m \lambda cr
\]  

Here, the paraxial approximation of the grating equation has been used. By comparing this equation with the paraxial equation (2.5.5) of a (thin) lens and replacing \( r \) by \( x \) it can be seen that the focal length \( f' \) of a FZL is just:

\[
-\frac{1}{f'} = m \lambda c \quad \Rightarrow \quad f' = -\frac{1}{m \lambda c}
\]  

This means in particular that the product \( f' \lambda \) of a FZL is constant. So, the dispersion relation for the focal length of a FZL can be written as:

\[
f' (\lambda) = \frac{\lambda_0}{\lambda} f' (\lambda_0)
\]  

(6.1.5)
The wavelength $\lambda_0$ is the design wavelength.

So, the Abbe number $V_d$ of a diffractive lens can be calculated using the definition (5.4.6):

$$V_d = \frac{1/f_d'}{1/f'_{F} - 1/f'_{C}} = \frac{\lambda_d}{\lambda_0 f'(\lambda_0)} - \frac{\lambda_d}{\lambda_0 f'(\lambda_0)} = \frac{\lambda_d}{\lambda_F - \lambda_C} = \frac{587.6 \text{ nm}}{486.1 \text{ nm} - 656.3 \text{ nm}} = -3.452 \ (6.1.6)$$

So, the Abbe number of a diffractive lens is a negative constant which is independent of the material of the lens or other parameters of the lens. Keep in mind that the Abbe number of a refractive material is always positive so that a DOE has a negative dispersion compared to a refractive lens.

### 6.1.3 Short glossary of DOEs

There are different types of DOEs. Here, just a short glossary for some of these different types will be given. More information about DOEs in general and their applications is available for example in [23],[30],[43].

Diffractive optical elements which are also called computer generated holograms (CGH) can be divided into amplitude and phase elements.

In an **amplitude hologram** just the local absorption is varied. Mostly, the local transmittance of an amplitude hologram is either (nearly) zero or (nearly) one, i.e. light is completely transmitted or completely absorbed. Such holograms are called binary amplitude holograms (see figure 6.3a). Without being able to derive here the equations for the diffraction efficiency it should be mentioned that the maximum diffraction efficiency of a binary amplitude hologram is reached if the duty cycle is 1:1 (i.e. the absorbing area and the light transmitting area in each period has equal size) and results in an efficiency of 10.1% (1/$\pi^2$) in each of the first orders. 50% of the light is then just absorbed as is clear from simple geometrical considerations, and 25% of the light is in the zeroth diffraction order.

A more efficient implementation of DOEs are the **phase holograms** where the local phase is varied by changing the local height of a dielectric material like glass. For a material with refractive index $n$ which is working in air (refractive index of air is here set to exactly one) the
relation between the local phase change $\Phi$ and the local height $h$ of the element is simply:

$$\Phi = \frac{2\pi}{\lambda} (n - 1) h$$  \hspace{1cm} (6.1.7)

Again, $\lambda$ is the wavelength of the illuminating light. If the height profile forms a continuous surface with only sharp edges at the rim of each period the element is called blazed phase hologram or kinoform hologram (see figure 6.3d). In such a blazed hologram the light can theoretically be diffracted in only one diffraction order with 100% efficiency if the phase change from each period to the next is exactly $2\pi$, i.e. if the maximum height of the structures is $\Delta h = \lambda/(n - 1)$. We have also seen this in subsection 6.1.1 for the special case of a FZL. If the height is different from the ideal case or equivalently if the element is used for light with a different wavelength as in the design, the diffraction efficiency decreases. But, for a normal glass element with $n \approx 1.5$ the diffraction efficiency can be more than 80% over the whole visible range if the design wavelength is in the middle of the visible range (wavelength of about 500 nm).

The next best approximation of a blazed phase hologram is a multiple step phase hologram which can be fabricated by a consecutive application of the method of fabricating binary elements (see figure 6.3c). There, first the finest structures are written and at the end the coarsest structures. By making $M$ binary processes an element with $N = 2^M$ steps can be fabricated in this way. The optimum overall height $\Delta h$ (for perpendicular incidence) and the maximum diffraction efficiency in the first order $\eta$ are then for $N$ equidistant steps [23]:

$$\Delta h = \frac{N - 1}{N} \frac{\lambda}{n - 1} \Rightarrow \eta = \frac{N^2}{\pi^2} \sin^2 \left( \frac{\pi}{N} \right)$$  \hspace{1cm} (6.1.8)

Of course, for the limiting case of an infinite number of steps $N$ the values of the blazed hologram result: $\Delta h = \lambda/(n - 1)$ and $\eta = 1$. On the other, for the binary phase hologram with only $N = 2$ steps (see figure 6.3b) the optimum values are $\Delta h = \lambda/(2(n - 1))$ and $\eta = 4/\pi^2 = 0.405$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>40.5%</td>
</tr>
<tr>
<td>4</td>
<td>81.1%</td>
</tr>
<tr>
<td>8</td>
<td>95.0%</td>
</tr>
<tr>
<td>16</td>
<td>98.7%</td>
</tr>
<tr>
<td>32</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

Table 6.1: Maximum diffraction efficiency $\eta$ of multiple step phase holograms in the first diffraction order for $N$ equidistant steps.

More general, the maximum diffraction efficiency (provided the height is correct) of a multiple step phase hologram is listed in table 6.1. So, for more than $N = 16$ steps the hologram will work in practice like a blazed hologram if no fabrication errors are present.

As mentioned above all these diffraction efficiency values are optimum values. In the case of fabrication errors or deviations of the wavelength of the used light from the design wavelength there will be a deterioration of the diffraction efficiency. For an amplitude hologram the diffraction efficiency is independent of the wavelength, but of course it is quite small at all.
6.1.4 Phase function of a diffractive optical element

In subsection 6.1.2 the Fresnel zone lens was defined by the local grating frequency \( \nu(r) \). By integrating this function the so called phase function \( \Phi \) of the FZL is obtained. This method is not limited to FZLs, but it can also be used for quite general DOEs. Exactly, there is the following relation between the phase function and the local grating frequency:

\[
\nu(x, y) = \frac{1}{2\pi} |\nabla \Phi(x, y)|
\]  

(6.1.9)

The phase function gives directly the position and shape of the structures of the DOE. If the phase function increases from one point to another point by \( 2\pi \) it means that the next period of the DOE is reached. Mostly, the phase function is defined as continuous function. The local height \( h \) of the structures of a blazed phase hologram is then just given by using equation (6.1.7):

\[
h(x, y) = \frac{\lambda}{2\pi(n - 1)} \left[ \Phi(x, y) \mod 2\pi \right]
\]  

(6.1.10)

Here, the operation \( \mod \) designates the floating point rest of a modulo operation. For example it is: \( 8.5 \mod 1.5 = 1.0 \) (since it is \( 8.5=5\times1.5+1.0 \)) and \( 9.0 \mod 1.5 = 0 \) (since it is \( 9.0=6\times1.5+0.0 \)).

For a multiple step phase hologram with \( N \) steps the local height \( h \) can also easily be calculated from the continuous phase function \( \Phi \):

\[
h(x, y) = \frac{\lambda}{N(n - 1)} \left[ \text{floor} \left( \frac{N\Phi(x, y)}{2\pi} \right) \mod N \right]
\]  

(6.1.11)

Here, the operation floor designates like in the programming language C that integer value which is smaller or equal to the argument of the floor operation. In total, the floor operation combined with the modulo operation gives for example: \( \text{floor}(5.7) \mod 8 = 5 \), \( \text{floor}(8.0) \mod 8 = 0 \) and \( \text{floor}(12.5) \mod 8 = 4 \).

Simple examples of phase functions are:

- \( \Phi(x, y) = ax + by \): Linear grating with constant grating frequency \( \nu = \sqrt{a^2 + b^2}/(2\pi) \).
  The grating lines are oriented relative to the \( x \)-axis by an angle \( \phi = \arctan(a/b) \). If for example \( b = 0 \) (and \( a \neq 0 \)) it is \( \phi = \pi/2 \) and the grating lines are along the \( y \)-axis.

- \( \Phi(x, y) = a(x^2 + y^2) = ar^2 \): Phase function of a lens in the paraxial approximation.
  The structures of the DOE are concentric rings where the grating frequency \( \nu(r) = ar/\pi \) increases linearly with the radius \( r \). Then, the focal length \( f' \) in the diffraction order \( m \) is according to equation (6.1.4) \( f' = -\pi/(m\lambda a) \).

The phase function of a wave front shaping DOE can be easily calculated if the incident wave and the desired output wave are known. Assume that the incident wave has a phase distribution \( \Phi_{\text{in}} \) in the plane of the DOE and that the output wave has a phase \( \Phi_{\text{out}} \) in the DOE plane. Then, using the principle of holography the phase function \( \Phi \) of the DOE using the diffraction order \( m \) is:

\[
\Phi_{\text{out}} = \Phi_{\text{in}} + m\Phi \quad \Rightarrow \quad \Phi = \frac{\Phi_{\text{out}} - \Phi_{\text{in}}}{m}
\]  

(6.1.12)

Of course, in most cases \( m \) is set to one, i.e. the first diffraction order is used.
6.1. THE DIFFRACTIVE LENS

6.1.5 Example of designing the phase function

As an example light from a point source should be imaged to another point (see figure 6.4). The hologram plane is perpendicular to the axis which is defined by the two points. The object point is in front of the hologram plane by a distance $g$. The image point is behind the hologram plane by a distance $b$. Then, the optical path length and the corresponding phase function $\Phi_{\text{in}}$ from the object point to a point on the hologram plane in the distance $r$ from the optical axis is according to the Pythagorean theorem:

$$\Phi_{\text{in}} = \frac{2\pi}{\lambda} \sqrt{g^2 + r^2} \quad (6.1.13)$$

In the same way the phase function $\Phi_{\text{out}}$ of the output wave to the image point can be calculated:

$$\Phi_{\text{out}} = -\frac{2\pi}{\lambda} \sqrt{b^2 + r^2} \quad (6.1.14)$$

The negative sign takes into account that the wave is now converging to the image point. So, the phase function $\Phi$ of the DOE using the first diffraction order is according to equation (6.1.12):

$$\Phi = -\frac{2\pi}{\lambda} \left( \sqrt{b^2 + r^2} + \sqrt{g^2 + r^2} \right) \quad (6.1.15)$$

The negative sign indicates that the DOE structure which is on the backside of the hologram plane is oriented like the structure which is symbolized in figure 6.4.

In the paraxial case it is of course $r \ll g$ and $r \ll b$. Then, the square roots can be replaced by the first two terms of their Taylor expansions and by neglecting a constant phase term $-2\pi(g + b)/\lambda$, which is irrelevant, a parabolic phase function results:

$$\Phi \approx -\frac{\pi}{\lambda} \left( \frac{1}{b} + \frac{1}{g} \right) r^2 \quad (6.1.16)$$

So, in the paraxial case it is a simple FZL with parabolic phase function and focal length $f' = (1/b + 1/g)^{-1}$. However, in the non–paraxial case a FZL with a parabolic phase function
would result in large spherical aberration for the image point and the phase function of equation (6.1.15) has to be used in order to obtain a diffraction limited imaging. It has to be mentioned that such a diffractive lens is of course only aberration corrected for the special pair of object and image point from the design. For off–axis object points or an on–axis object point with a distance different from $g$ there will be aberrations.

### 6.2 The aplanatic meniscus

A very simple but useful lens to increase the numerical aperture of an imaging system without introducing a remarkable amount of aberrations is an aplanatic meniscus lens. In order to understand its mode of operation the aplanatic points of imaging with a sphere shall be discussed.

#### 6.2.1 The aplanatic points of a sphere

It is possible to image all points on a sphere aberration–free to points on another concentric sphere by using a refracting sphere. The radius of curvature of the refracting sphere (solid circle in figure 6.5) shall be $R$ and its material has a refractive index $n'$. The refractive index of the surrounding material is $n$. In figure 6.5 only the case $n' > n$ is shown. Then, all points on a sphere with radius of curvature $n'R/n$ (outer dashed sphere) are imaged to points on the internal sphere (dotted sphere) with radius of curvature $nR/n'$. To understand this, figure 6.5 is used. An arbitrary ray which would travel to the point $P$ on the outer sphere hits the refracting sphere in the point $Q$. Then, the ray is refracted and hits the internal sphere in the point $P'$ which lies on the line $OP$. The point $O$ is the center of all concentric spheres. The angles $i$ and $i'$ are the angles of the incident and refracted ray with the surface normal.
6.2. THE APLANATIC MENISCUS

Figure 6.6: Scheme of an aplanatic meniscus lens. The first lens surface makes an aplanatic imaging of points on the dashed large circle around O to points on the small dotted circle around O. The lens back surface is concentric around the on–axis image point P’.

The proof of this can be done using the following relations:

\[ [OQ] = R \]
\[ [OP] = \frac{n'}{n} R \]
\[ [OP'] = \frac{n}{n'} R \]  \hspace{1cm} (6.2.1)

So, it is:

\[ \frac{[OQ]}{[OP]} = \frac{n}{n'} = \frac{[OP']}{[OQ]} \]  \hspace{1cm} (6.2.2)

Hence, the two triangles QOP’ and POQ are similar because they have one common angle \( \hat{OQP} \) and the ratio of the two sides of both triangles adjacent to this angle are identical, respectively. Then, the angle \( \hat{OPQ} \) is identical to \( i' \) and by using the law of sines in a triangle it is:

\[ \frac{\sin \hat{PQO}}{\sin \hat{OPQ}} = \frac{\sin i}{\sin i'} = \frac{[OP]}{[OQ]} = \frac{n'}{n} \]  \hspace{1cm} (6.2.3)

But, this is the law of refraction and hence it is clear that all rays which travel to the point P are perfectly refracted by the sphere with radius of curvature \( R \) to the image point P’ without any aberrations. So, because of the rotational symmetry of our system all points on the outer sphere are imaged to points on the internal sphere without any aberrations.

6.2.2 Properties of an aplanatic meniscus

Now, we assume \([OP]\) to be the optical axis in figure 6.5. Then, the sine of the aperture angle \( i \) of the refracted convergent ray bundle propagating to P’ is increased compared to the sine of the aperture angle \( i' \) of the incident ray bundle by a factor \( n'/n \) according to equation (6.2.3). In figure 6.6 this situation is displayed again, but now the new names for the aperture angle \( \varphi = i' \) of the incident ray bundle and \( \varphi' = i \) of the refracted ray bundle are introduced. In the
6.3 The achromatic lens

In paragraph 5.4 the chromatic aberrations of a single lens, i.e. the dependence of the focal length on the wavelength, were treated. An achromatic lens should have in the ideal case no chromatic aberrations. However, in practice the most important achromatic lens is an achromatic doublet consisting of two cemented lenses and then the focal length can be identical for only two different wavelengths. So, in technical optics the term achromatic lens (dt.: Achromat) normally means a lens where the focal length is identical for two different wavelengths. For applications in the visible range these two wavelengths are commonly $\lambda_F = 486.1$ nm (blue line of atomic hydrogen) and $\lambda_C = 656.3$ nm (red line of atomic hydrogen) which are near the border area of the visible range. A lens where the focal length is identical for three wavelengths is called an apochromatic lens (dt.: Apochromat).

To understand the principle of the achromatic correction of a lens doublet the paraxial matrix $M$ of a combination of two thin lenses (paraxial matrices $M_1$ and $M_2$) with a zero distance situated in air is calculated. Of course, this is a simplification because in practice no thin lens really exists and if the principal points of thick lenses are taken as reference elements the distance between two lenses will normally be different from zero. But nevertheless, the calculation with two zero–distant thin lenses explains the principle and according to equation (2.5.2) the result is

$$M = M_2 M_1 = \begin{pmatrix} 1 & 0 \\ -f'_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -f'_1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -f'_1 - f'_2 & 1 \end{pmatrix}$$

(6.3.1)

where the focal lengths of the two thin lenses are $f'_1$ and $f'_2$. Therefore, the focal length $f'$ of the combination of these two lenses is

$$\frac{1}{f'} = \frac{1}{f'_1} + \frac{1}{f'_2},$$

(6.3.2)

i.e. the optical powers of the single lenses are just added.

The refractive index of the first lens is $n_1$ and that of the second lens $n_2$, both situated in air. The optical powers $1/f'_i$ ($i \in \{1, 2\}$) of refractive thin lenses are then according to equation (2.5.3)

$$\frac{1}{f'_i(\lambda)} = [n_i(\lambda) - 1] \left( \frac{1}{R_{i,1}} - \frac{1}{R_{i,2}} \right) =: [n_i(\lambda) - 1] C_i$$

(6.3.3)

where the term $C_i$ depends only on the two radii of curvature $R_{i,1}$ and $R_{i,2}$ of the thin lenses and is independent of the wavelength $\lambda$ whereas the refractive index $n_i$ depends on the wavelength. For an achromatic lens the optical powers at the two wavelengths $\lambda_F$ and $\lambda_C$ (or for two other wavelengths depending on the application) have to be identical. By using equations (6.3.2) and
6.3. THE ACHROMATIC LENS

(6.3.3) this means:

\[
\frac{1}{f'_1(\lambda_F)} + \frac{1}{f'_2(\lambda_F)} = \frac{1}{f'_1(\lambda_C)} + \frac{1}{f'_2(\lambda_C)}
\]

\[\Rightarrow [n_1(\lambda_F) - 1]C_1 + [n_2(\lambda_F) - 1]C_2 = [n_1(\lambda_C) - 1]C_1 + [n_2(\lambda_C) - 1]C_2 \]

\[\Rightarrow [n_1(\lambda_F) - n_1(\lambda_C)]C_1 = -[n_2(\lambda_F) - n_2(\lambda_C)]C_2 \]

By using again equation (6.3.3) the terms \(C_1\) and \(C_2\) can be expressed by the refractive indices at a medium wavelength between \(\lambda_F\) and \(\lambda_C\), in our case \(\lambda_d = 587.6\) nm (yellow line of helium), and the focal lengths at this wavelength and the result is:

\[
\frac{n_1(\lambda_F) - n_1(\lambda_C)}{[n_1(\lambda_d) - 1]f'_1(\lambda_d)} = -\frac{n_2(\lambda_F) - n_2(\lambda_C)}{[n_2(\lambda_d) - 1]f'_2(\lambda_d)} \Rightarrow V_{1,d}f'_1(\lambda_d) = -V_{2,d}f'_2(\lambda_d) \quad (6.3.4)
\]

Here, the Abbe numbers \(V_{i,d}\) \((i \in \{1, 2\})\) of the materials with refractive indices \(n_i\) are defined by equation (5.4.4).

Since the Abbe number of a refractive material is always positive one of the refractive thin lenses has to be a negative lens and one a positive lens to fulfill equation (6.3.4). However, if one of the two thin lenses is not a refractive but a diffractive lens it has formally a constant negative Abbe number \(V_d = -3.452\) (see equation (6.1.6) or [15], chapter 10 or [30]). So, in the case of a so called hybrid achromatic lens consisting of a refractive and a diffractive lens both lenses will have the same sign of their optical powers and so a positive hybrid achromatic lens consists of two positive single lenses: a refractive lens with a high optical power and a high Abbe number and a diffractive lens with a small optical power and the negative Abbe number with small modulus (see figure 6.7b).

However, a positive purely refractive achromatic lens consists of a positive lens with high optical power and high Abbe number (made of a crown glass such as BK7) and a negative lens with smaller optical power and smaller Abbe number (made of a high dispersive flint glass such as SF10), so that in total a positive optical power results. Figure 6.7a shows the principal scheme of such an achromatic doublet. The first spherical surface of the crown glass lens has to be highly curved in order to guarantee a positive lens. The last surface of the flint glass lens has

---

Figure 6.7: Principal schemes of a) a refractive achromatic doublet and b) a hybrid achromatic doublet.
only a small curvature so that together with the common surface of both lenses which has a
medium curvature the flint glass lens has a negative power.
In our mathematical description only the case of two thin lenses with zero distance is treated. But
it is no problem to use the paraxial matrix theory to calculate the matrix of a real achromatic
doublet consisting of two cemented lenses, i.e. three refractive spherical surfaces with finite
distances embedding two different materials. However, in this case not only the focal length will
depend on the wavelength but also to some degree the position of the principal planes. So, the
position of the focus itself can vary a little bit although the focal length is identical for the two
selected wavelengths $\lambda_F$ and $\lambda_C$.
In practice, a refractive achromatic doublet which can be bought is a lens which does not only
correct the chromatic errors but which also fulfills the sine condition (see equation (3.1.5) on
page 54). This is possible because there are three surfaces with different radii of curvature
whereas to fulfill the paraxial properties only two of these three parameters are determined.

6.3.1 Examples of designing achromatic doublets
In this paragraph the paraxial properties of different achromatic doublets will be calculated and
compared with those of single refractive lenses. It will be assumed like above that the two lenses
of the achromatic doublet are thin lenses with zero distance between the two lenses. This is of
course a simplification, but nevertheless it is a good approximation for most cases.
Due to equation (6.3.4) the focal lengths $f'_1(\lambda_d)$ and $f'_2(\lambda_d)$ of the two lenses of the achromatic
doublet at the wavelength $\lambda_d = 587.6$ nm have to fulfill the condition:

$$V_{1,d}f'_1(\lambda_d) = -V_{2,d}f'_2(\lambda_d) \Rightarrow f'_1(\lambda_d) = -\frac{V_{2,d}}{V_{1,d}}f'_2(\lambda_d) \quad \text{or} \quad f'_2(\lambda_d) = -\frac{V_{1,d}}{V_{2,d}}f'_1(\lambda_d)$$

Here, $V_{1,d}$ and $V_{2,d}$ are the Abbe numbers of the materials of the two lenses. Additionally, the
focal length $f'$ of the achromatic doublet can be calculated according to equation (6.3.2) by

$$\frac{1}{f'} = \frac{1}{f'_1} + \frac{1}{f'_2}$$

By combining both equations the focal lengths of the two single lenses can be expressed as
function of the focal length of the achromatic doublet:

$$f'_1(\lambda_d) = \frac{V_{1,d} - V_{2,d}}{V_{1,d}} f'(\lambda_d)$$
$$f'_2(\lambda_d) = \frac{V_{2,d} - V_{1,d}}{V_{2,d}} f'(\lambda_d) \quad (6.3.5)$$

A refractive achromatic doublet made of BK7 and SF10 has for example the Abbe numbers
$V_{1,d} = 64.17$ (BK7) and $V_{2,d} = 28.41$ (SF10). Therefore, the focal lengths of the two single
lenses are in this case due to the equations (6.3.5):

Lens made of BK7: \quad $f'_1(\lambda_d) = 0.557 f'(\lambda_d)$
Lens made of SF10: \quad $f'_2(\lambda_d) = -1.259 f'(\lambda_d)$

So, the second lens made of the highly dispersive material SF10 is a negative lens if the achro-
matic doublet itself is a positive lens.
As mentioned previously a diffractive lens (DOE) can be described by a negative and material-independent Abbe number $V_d = -3.452$. Therefore, in the following the focal lengths of the two single lenses of a hybrid achromatic doublet made of one refractive lens and one DOE shall be considered. First, the refractive lens with focal length $f'_1$ is made of BK7 and the second lens with focal length $f'_2$ is a DOE. According to equations (6.3.5) the focal lengths are:

- **Lens made of BK7:** $f'_1(\lambda_d) = 1.054f'(_d)\lambda$
- **DOE:** $f'_2(\lambda_d) = 19.588f'(_d)\lambda$

So, as mentioned previously, both lenses are positive lenses if the achromatic doublet has a positive optical power. Of course, most of the optical power is delivered by the refractive lens. A second hybrid achromatic doublet can be made for example by taking a refractive lens made of SF10 and a DOE:

- **Lens made of SF10:** $f'_1(\lambda_d) = 1.122f'(_d)\lambda$
- **DOE:** $f'_2(\lambda_d) = 9.230f'(_d)\lambda$

The remaining chromatic aberrations of an achromatic doublet, i.e. the variation of the focal length with the wavelength of the illuminating light, can be calculated by using equation (6.3.2). In this equation the optical power of a refractive lens as function of the wavelength is calculated by equation (6.3.3) and the optical power of a diffractive lens as a function of the wavelength is [15]:

$$\frac{1}{f'_{DOE}(\lambda)} = \frac{\lambda}{\lambda_d f'_{DOE}(\lambda_d)} =: C \lambda$$

Here, $C = 1/(\lambda_d f'_{DOE}(\lambda_d))$ is a constant value which depends on the focal length $f'_{DOE}(\lambda_d)$ of the DOE at the wavelength $\lambda_d = 587.6$ nm. So, the optical power of the DOE increases linearly
6.4 The camera

One of the simplest optical instruments is a camera [1],[35]. Of course, modern cameras are highly sophisticated technical instruments with complex wide angle or zoom lenses. But the basic principle of each camera (see figure 6.10) is that a lens forms a real inverted picture of an object on a photosensitive surface which can be a photographic film or an electronic detector like a CCD chip. Additionally, each camera has a diaphragm near the lens.

A standard camera objective of a miniature camera has a focal length of $f' = 50 \text{ mm}$ so that each object with a distance of several meters can be assumed to be at an infinite distance and the object distance $d_O$ in the "lens equation" (2.4.7) can be assumed to be $d_O \rightarrow -\infty$. Then, the image is practically formed in the focal plane of the lens, i.e. $d_I \approx f'$. So, the size $x$ of the image of an object is determined by the angular extension $\varphi$ of the object by

$$x \approx \varphi f'$$  \hspace{1cm} (6.4.1)
Figure 6.10: Principle of a camera. The distance of the object to the lens of the camera compared to the focal length of the lens is in the presented case so large that the object can be assumed to be at infinite distance.

The moon has for example an angular extension of about half a degree if observed from the earth so that its image on a standard camera would just be $x = 0.44 \, \text{mm}$. This is the reason why the moon on a photo made with a miniature camera with a film size of $24 \, \text{mm} \times 36 \, \text{mm}$ is really small and details cannot be detected. But, this can be changed by using a telescope (see section 6.6) in front of the camera, which changes the angular extension $\varphi$ of the object. In astronomical cameras the eyepiece of the telescope is commonly omitted and the detector is positioned directly in the focal plane of the objective lens or mirror which has a large focal length $f'$ and which serves as camera lens. Nevertheless, such a device is still called "astronomical telescope".

6.4.1 The depth of field

In geometrical optics an ideal camera lens (without aberrations) images one object plane really sharp onto the photosensitive image plane. But in reality each image point is first due to the wave nature of light not an ideal mathematical point but an airy disc, and second in the case of a camera the resolution of the detector is in many cases smaller than the maximum possible resolution given by the wave nature of light. Object points in other planes as the ideal object plane are imaged to planes in front of or behind the detector plane (see figure 6.11). Therefore, in the detector plane they will form small "image discs". But, if the diameter of the "image discs" is smaller than the pixel distance $p$ of the detector also these other planes will be imaged without loss of resolution onto the detector which limits the resolution.

The ideal object plane which is really imaged sharply onto the detector has the object distance $d_O$ and the detector plane has the image distance $d_I$ (where $d_O < 0$ and $d_I > 0$ for a real image in a camera). That object plane which is nearer to the camera lens than $|d_O|$ and where the light rays of the object points form small discs in the detector plane with a diameter of exactly $p$ is the nearest object plane which is imaged onto the detector with the maximum resolution given by the pixel distance $p$. Its object distance is called $d_{O,N}$ (index "N" for "near") and its image distance is called $d_{I,N}$ (see figure 6.11). Similarly, that object plane with a larger distance than $|d_O|$ from the lens where the rays coming from the object points also form discs...
in the detector plane with a diameter \( p \) is the farthest object plane which is imaged with the maximum resolution given by the detector. It has the object distance \( d_{O,F} \) and corresponding image distance \( d_{I,F} \) (index "F" for "far").

The depth of field (dt.: Schärftiefe) is now defined as the axial extension of the object space between the "near" object plane and the "far" object plane which are both just imaged with the maximum resolution of the detector. The depth of field depends of course on the diameter \( D \) of the aperture stop and on the resolution of the detector, i.e. the pixel distance \( p \). We assume in the following that we have a thin ideal lens with a focal length \( f' \) on the image side and that the aperture stop is directly in the plane of the lens. Then, the aperture stop is also the entrance pupil and the exit pupil. An important quantity is the so called F number \( f# \) of the lens (dt.: Öffnungszahl oder Blendenzahl) which is defined as the ratio of the focal length \( f' \) of the lens and the diameter \( D \) of the entrance pupil:

\[
f# = \frac{f'}{D} \tag{6.4.2}
\]

If the image is formed in the focal plane like it is nearly the case for a camera imaging a far distant object and if the diameter \( D \) is small compared to the focal length \( f' \) the F number and the numerical aperture on the image side defined by equation (3.1.3) are connected with a good approximation by:

\[
NA_I = n_I \sin \varphi_I \approx n_I \frac{D}{2f'} = n_I \frac{1}{2f#} \tag{6.4.3}
\]

There, \( \varphi_I \) is the half aperture angle of the light cone on the image side and \( n_I \) is the refractive index on the image side. In most cases there will be air on the image side, i.e. \( n_I = 1 \). But for some camera–like systems such as the human eye \( n_I \) will be different from 1 (see section 6.5).

According to the "lens equation" (2.4.7), where \( n_O \) and \( n_I \) are the refractive indices on the object and image side, respectively, we have three equations for the different object and image
6.4. THE CAMERA

distances:

\[
\frac{n_I}{d_I} - \frac{n_O}{d_O} = \frac{n_I}{f'} \Rightarrow \quad d_I = \frac{n_I f' d_O}{n_O f' + n_I d_O} \quad (6.4.4)
\]

\[
\frac{n_I}{d_{I,N}} - \frac{n_O}{d_{O,N}} = \frac{n_I}{f'} \Rightarrow \quad d_{I,N} = \frac{n_I f' d_{O,N}}{n_O f' + n_I d_{O,N}} \quad (6.4.5)
\]

\[
\frac{n_I}{d_{I,F}} - \frac{n_O}{d_{O,F}} = \frac{n_I}{f'} \Rightarrow \quad d_{I,F} = \frac{n_I f' d_{O,F}}{n_O f' + n_I d_{O,F}} \quad (6.4.6)
\]

Additionally, according to the theorem on intersecting lines we have two additional equations (see fig. 6.11):

\[
\frac{D}{d_{I,N}} = \frac{p}{d_{I,N} - d_I} \Rightarrow \quad d_{I,N} - d_I = \frac{p D}{d_{I,N}} \quad (6.4.7)
\]

\[
\frac{D}{d_{I,F}} = \frac{p}{d_I - d_{I,F}} \Rightarrow \quad d_I - d_{I,F} = \frac{p D}{d_{I,F}} \quad (6.4.8)
\]

By putting equations (6.4.4) and (6.4.5) in equation (6.4.7) and solving for \(d_{O,N}\) the result is:

\[
d_{O,N} = \frac{n_O f' d_O}{n_O f' + p D (n_O f' + n_I d_O)} = \frac{d_O}{1 - \frac{p D}{n_I d_O}} \quad (6.4.9)
\]

In the same way by combining equations (6.4.4), (6.4.6) and (6.4.8) the result for \(d_{O,F}\) is:

\[
d_{O,F} = \frac{n_O f' d_O}{n_O f' + p D (n_O f' + n_I d_O)} = \frac{d_O}{1 + \frac{p D}{n_I d_O}} \quad (6.4.10)
\]

It is common practice in photography to use the lateral magnification \(\beta\) which was defined by equation (2.3.1) as the ratio of the image height \(x_I\) and the object height \(x_O\). For a lens which fulfills the sine condition (3.1.5) the principal planes are in reality "principal spheres" which are centered around the object and the image point, respectively. The same is valid for the entrance and exit pupil [13]. Then, using equation (3.1.5) the lateral magnification can be expressed as:

\[
\beta = \frac{x_I}{x_O} = \frac{n_O \sin \varphi_O}{n_I \sin \varphi_I} = \frac{n_O D/(2d_O)}{n_I D/(2d_I)} = \frac{n_O}{n_I} \frac{d_I}{d_O} \quad (6.4.11)
\]

By multiplying equation (6.4.4) with \(d_O/n_O\) it holds:

\[
\frac{n_I d_O}{n_O d_I} - 1 = \frac{n_I d_O}{n_O f'} \Rightarrow \quad 1 = \frac{1}{\beta} = 1 + \frac{n_I d_O}{n_O f'} \quad (6.4.12)
\]

So, equations (6.4.9) and (6.4.10) can be expressed by:

\[
d_{O,N} = \frac{d_O}{1 - \frac{p D}{\beta}} = \frac{d_O}{1 - \frac{p f'}{f'}} \quad (6.4.13)
\]

\[
d_{O,F} = \frac{d_O}{1 + \frac{p D}{\beta}} = \frac{d_O}{1 + \frac{p f'}{f'}} \quad (6.4.14)
\]
In the last step, the F number \( f\# \) defined by equation (6.4.2) is used. In the case of a camera the focal length \( f' \) is positive and the lateral magnification \( \beta \) is always negative since a real image is formed, i.e. \( \beta < 0 \). So, there is the interesting special case that the denominator of equation (6.4.14) can be zero:

\[
1 + \frac{p}{D\beta} = 0 \Rightarrow \beta = -\frac{p}{D} \Rightarrow d_{O,C} = -\frac{n_O f'}{n_I} \left( 1 + \frac{D}{p} \right) = -\frac{n_O f'}{n_I} \left( 1 + \frac{f'}{p f\#} \right) \tag{6.4.15}
\]

At the right side, equation (6.4.12) was used and solved for \( d_O \). So, if the camera is focused to the critical object distance \( d_{O,C} \) given by equation (6.4.15) it holds \( |d_{O,F}| \to \infty \) and all objects which are farther from the camera lens than \( |d_{O,N}| = |d_{O,C}|/2 \) (this follows from equation (6.4.13)) will be imaged onto the detector with the maximum resolution, i.e. the image will look sharp. Of course, if the modulus \( |d_O| \) of the actual object distance is larger than the modulus of the critical value \( |d_{O,C}| \) given by equation (6.4.15), \( d_{O,F} \) will formally be positive. This means that also a virtual object behind the lens with distance \( d_{O,F} \), which can be produced by some auxiliary optics, can be imaged sharp onto the detector. In fact, this means that still all real objects with larger distance from the lens than \( |d_{O,N}| \) will be imaged sharp onto the detector.

If we have for example a camera with \( f' = 50 \text{ mm} \), a minimum F number \( f\# = 2.8 \), \( n_O = n_I = 1 \), and a pixel distance \( p = 11 \text{ \mu m} \) (typical CCD chip) the critical object distance \( d_{O,C} \) of equation (6.4.15) is \( d_{O,C} = -81.2 \text{ m} \). Therefore, all objects with a distance of more than \( |d_{O,N}| = |d_{O,C}|/2 = 40.6 \text{ m} \) to the camera will be imaged sharp if the camera is focused to \( d_{O,C} \).

If the F number is \( f\# = 16 \) all objects with a distance of more than \( 7.1 \text{ m} \) will be imaged sharp for a focusing distance of \( |d_{O,C}| = 14.3 \text{ m} \). However, for larger F numbers the wave nature of light begins to limit the resolution because the radius \( r_{\text{diff}} \) of a diffraction limited spot will be \( r_{\text{diff}} = 0.61 \lambda/\text{NA} \approx 1.22 \lambda f\# = 10.7 \text{ \mu m} \approx p \) for a wavelength \( \lambda = 550 \text{ nm} \) and \( f\# = 16 \).

Of course, a larger F number means that the light intensity on the detector decreases because the light intensity on the detector is proportional to the effective area \( \pi D^2/4 \) of the light gathering lens and therefore proportional to \( 1/f\#^2 = D^2/f^2 \). So, a larger F number means that the exposure time (dt.: Belichtungszeit) has to be increased proportional to \( f\#^2 \). All these facts are well-known from photography.

If \( d_{O,F} \) has a finite value, e.g. if the camera is focused to a near object, it is useful to calculate the axial extension \( \Delta d = d_{O,N} - d_{O,F} \) of the sharply imaged object space. By using equations (6.4.13) and (6.4.14) the result is:

\[
\Delta d = d_{O,N} - d_{O,F} = 2d_O \frac{pf\#}{f^2} \frac{\left( \frac{1}{\beta} - 1 \right)^{\frac{1}{\beta}}}{1 - \left( \frac{pf\#}{f^2} \right)^2} = 2n_O pf\# \frac{\left( \frac{1}{\beta} - 1 \right)^{\frac{1}{\beta}}}{1 - \left( \frac{pf\#}{f^2} \right)^2} \tag{6.4.16}
\]

In the last step equation (6.4.12) has been used to express the object distance \( d_O \) by the lateral magnification \( \beta \) because these two quantities are of course coupled with each other.

Again, we see at equation (6.4.16) the limiting case that the denominator can approach zero (if equation (6.4.15) is fulfilled) and that therefore the depth of field has an infinite range. But, for near objects (for example \( |d_O| \leq 1 \text{ m} \)) we normally have the case that \( f'/\beta \gg pf\# \). Then, \( \Delta d \) has first of all a finite positive value and second there is a quite good approximation which is often used for the photography of near objects [35]:

\[
\Delta d = 2n_O pf\# \frac{\left( \frac{1}{\beta} - 1 \right)^{\frac{1}{\beta}}}{1 - \left( \frac{pf\#}{f^2} \right)^2} \approx 2n_O pf\# \frac{1}{\beta} \tag{6.4.17}
\]
As an example we take again a common electronic camera with \( f' = 50 \text{ mm} \), \( p = 11 \mu \text{m} \) and \( n_O = n_I = 1 \). The F number is assumed to be \( f\# = 10 \) and the object is at \( d_O = -1 \text{ m} \). Then, the lateral magnification is according to equation (6.4.12) \( \beta = -0.05263 \). The extension of the depth of field \( \Delta d \) is according to the exact equation (6.4.16) \( \Delta d = 83.75 \text{ mm} \) and according to the approximate equation (6.4.17) \( \Delta d = 83.60 \text{ mm} \). So, the error of the approximate equation is just about 0.2\% and the depth of field has an extension of about 8.4 cm, i.e. objects with an axial extension in this range (for a medium object distance \( |d_O| = 1 \text{ m} \)) will be imaged without loss of resolution onto the detector.

### 6.5 The human eye

The human eye is from its principle a camera which builds an inverted real image of the surrounding area on the retina [1],[35]. However, the actual structure and performance of the human eye is quite complex [5],[10],[16] so that only the most important features of the normal emmetropic eye (dt.: normalsichtiges Auge) can be discussed in this section.

The optical power of the eye is delivered by the cornea (dt.: Hornhaut) and the deformable crystalline lens (dt.: Kristalllinse oder kurz Augenlinse) (see figure 6.12). The main part of the optical power is delivered by the cornea with about 43 diopters (1 dioptr=1 dpt=1 m\(^{-1}\), dt.: 1 Dioptrie) because at the first surface the difference between the refractive indices of air and
The cornea (1.376) is quite high. The crystalline lens with a refractive index between 1.386 in the outer parts and 1.406 in the core is immersed on the one side in the aqueous humour (dt.: Kammerwasser) and on the other side in the vitreous body (dt.: Glaskörper) which both have a refractive index of 1.336. Therefore, the lens has just about 19 diopters in the case of distant vision. The resulting total optical power of the eye is due to the finite distance between cornea and crystalline lens about 59 diopters for distant vision. The accommodation of the eye for near objects which is performed by the crystalline lens can vary between about 14 diopters in young age and nearly 0 diopters above 50 years of age because the crystalline lens loses its flexibility with increasing age. Since the normal distance for reading is about 25–30 cm an accommodation of less than 3-4 diopters has to be corrected by wearing eyeglasses for reading.

The photosensitive surface of the eye is the curved retina (dt.: Netzhaut) and the diaphragm of the eye is the iris (dt.: Iris oder Regenbogenhaut), which can change its diameter between about 2 mm and 8 mm to control the irradiance on the retina depending on the intensity of the illuminating light. The effective focal length of the eye, which is as mentioned above an immersion system, is \( f'/n' \approx 1/(59 \text{ diopeters}) \approx 17 \text{ mm} \) (\( n' = 1.336 \) is the refractive index of the vitreous body between the eye lens and the retina). The so-called least distance of distinct vision (dt.: deutliche Sehweite oder Normsehweite) of a normal adult eye is about 25 cm requiring an accommodation of 4 diopters.

The angular resolution \( \Delta \varphi \) of a normal eye is about 1’ (one arc minute) and can achieve for optimal conditions 30”. The later corresponds to a distance \( \Delta x = \Delta \varphi f'/n' = 2.5 \mu \text{m} \) on the retina. So, the light sensitive cells (cones) in the fovea (dt.: Sehgrube oder wissenschaftlich fovea centralis) (about 200 \( \mu \text{m} \) diameter), which is the central part of the retina, have to be about 2.5 \( \mu \text{m} \) or less in diameter and distance. In the fovea there are mainly the color-sensitive cones (dt.: Zapfen) whereas in the outer parts the rods (dt.: Stäbchen) dominate which are more sensitive to light but which cannot distinguish between different colors.

It is interesting to note that a human eye with normal vision is a nearly diffraction-limited optical system for a diameter of the pupil of up to 3 mm (diameter for sharpest vision). This can be seen because in this case the radius \( r \) of the airy disc which limits the resolution according to the Rayleigh criterion (see the lecture about wave optics) is about \( r = 0.61 \lambda/\text{NA}=3.8 \mu \text{m} \) for a wavelength of \( \lambda = 0.55 \mu \text{m} \) and a numerical aperture \( \text{NA} \approx 1.5 \text{ mm}/17 \text{ mm} \approx 0.088 \). So, the above given value \( \Delta x = 2.5 \mu \text{m} \) for the smallest resolvable distance on the retina is even a little bit smaller than the distance \( r \) given by the Rayleigh criterion due to diffraction which assumes that a drop of 26 percent in irradiance can be detected. The reason is that the Rayleigh criterion is a little bit arbitrary and for optimal conditions the eye can also detect smaller drops in irradiance between two adjacent points. For larger diameters of the pupil than 3–4 mm the spherical aberration and chromatic aberrations of the eye reduce the resolution. Therefore, at night or in badly illuminated rooms the resolution of the eye is reduced and all tasks which need a high resolution, for example reading, are more difficult or impossible if the irradiance on the retina is also for the largest pupil diameter too small.

### 6.6 The Telescope

One of the most important optical instruments is the telescope (dt.: Fernrohr) [1],[13],[35]. It has well-known applications in terrestrial and astronomical observations [39]. But there are at least as important applications in optics to expand or compress a collimated (laser) beam,
6.6. THE TELESCOPE

A telescope consists in principle of two lenses or two other focusing optical elements like spherical or aspheric mirrors. Here, to demonstrate the principle we assume that it consists of two lenses with focal lengths \( f'_1 \) and \( f'_2 \) and a distance \( d \) between the two lenses. In order to have a telescope the image–sided focus \( F'_1 \) of the first lens and the object–sided focus \( F'_2 \) of the second lens have to coincide (see fig. 6.13). Additionally, we assume that the two lenses are situated in air, so that we have \( f'_2 = -f_2 \) for the image– and object–sided focal length. So, by taking into account the sign conventions for the focal lengths the condition for the distance between the two lenses of a telescope is:

\[
d = f'_1 - f_2 = f'_1 + f'_2
\]  

The paraxial matrix \( M \) of the telescope from the object–sided principal plane \( U'_1 \) of the first lens to the image–sided principal plane \( U'_2 \) of the second lens is:

\[
M = \begin{pmatrix}
1 & 0 & -1 & 0 \\
-1 & 1 & \frac{1}{f_2} & 1 \\
0 & 1 & \frac{1}{f_1} & 1 \\
\end{pmatrix}
\begin{pmatrix}
1 & d \\
0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
-1 & 1 \\
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{f_1} & f'_1 + f'_2 \\
\frac{1}{f_2} & 0 \\
\end{pmatrix}
\begin{pmatrix}
\frac{1}{f_1} & 0 \\
\frac{1}{f_2} & 1 \\
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{f_1} & f'_1 + f'_2 \\
\frac{1}{f_2} & 0 \\
\end{pmatrix}
\]

So, the coefficient \( C \) of the ABCD–matrix \( M \), which is according to equation (2.3.17) the negative value of the optical power, is zero and therefore the focal length of the telescope is infinity. Such
a system with zero optical power is called an **afocal system**. So, a telescope can also be defined to be an afocal optical system, where the trivial case that all lenses have themselves zero optical power, i.e. \(1/f'_1 = 1/f'_2 = 0\), is excluded.

### 6.6.1 Telescope as beam expander and imaging system for far–distant objects

An important property of an afocal system is that it transforms a collimated bundle of rays into another collimated bundle of rays. The application as a beam expander (dt.: Strahlauflaufung) for a collimated beam or as an imaging system for far–distant objects can easily be seen from equation (6.6.2) by taking two parallel rays with paraxial ray parameters \((x_1, \varphi_1)\) and \((x_2, \varphi_2)\) \((\varphi_2 = \varphi_1)\) in front of the telescope. The paraxial ray parameters \((x'_1, \varphi'_1)\) and \((x'_2, \varphi'_2)\) of the rays behind the telescope are then:

\[
\begin{pmatrix}
  x'_1 \\
  \varphi'_1
\end{pmatrix} = \begin{pmatrix}
  -f'_2/f'_1 & f'_1 + f'_2 \\
  0 & -f'_2/f'_2
\end{pmatrix} \begin{pmatrix}
  x_1 \\
  \varphi_1
\end{pmatrix} = \begin{pmatrix}
  -f'_2 x_1 + (f'_1 + f'_2) \varphi_1 \\
  -f'_2 \varphi_1
\end{pmatrix}
\tag{6.6.3}
\]

with \(i \in \{1, 2\}\).

The angular magnification \(\gamma\) defined by equation (2.3.2), i.e. the ratio of the angle \(\varphi' := \varphi'_1 = \varphi'_2\) between the bundle of rays and the optical axis behind the afocal system to the angle \(\varphi_i := \varphi_1 = \varphi_2\) in front of the system, is in the paraxial case:

\[
\gamma = \frac{\varphi'}{\varphi} = -\frac{f'_1}{f'_2}
\tag{6.6.4}
\]

So, the angular magnification, which determines the size of the image of a far–distant object, only depends on the ratio of the focal lengths of the two lenses.

The beam expanding property can be seen by calculating the distance \(\Delta x\) between two parallel rays \((\varphi_2 = \varphi_1)\) in front of the telescope and the distance \(\Delta x'\) behind the telescope:

\[
\Delta x' = x'_2 - x'_1 = -\frac{f'_2}{f'_1} (x_2 - x_1) = -\frac{f'_2}{f'_1} \Delta x
\tag{6.6.5}
\]

So, the beam expanding ratio \(\Delta x'/\Delta x\) is the reciprocal value of the angular magnification.

### 6.6.2 Imaging property of a telescope for finite distant objects

Although a telescope has zero optical power it images an object from one plane to another plane. This can be seen by calculating the paraxial matrix \(M'\) from an object plane with a distance \(d_1\) to the principal plane \(U_1\) of the first lens (keep in mind that, different from the normal sign conventions of paraxial optics, in the paraxial matrix theory \(d_1\) is positive if the object plane is in front of \(U_1\) and negative if it is behind \(U_1\)) to an image plane with the distance \(d_2\) behind the principal plane \(U'_2\) of the second lens (\(d_2\) is positive if the image plane is real and behind \(U'_2\) and negative if it is a virtual image plane in front of \(U'_2\)). Fig. 6.14 shows the parameters to calculate \(M'\):

\[
M' = \begin{pmatrix}
  1 & d_2 \\
  0 & 1
\end{pmatrix} \begin{pmatrix}
  -f'_2/f'_1 & f'_1 + f'_2 \\
  0 & -f'_2/f'_2
\end{pmatrix} \begin{pmatrix}
  1 & d_1 \\
  0 & 1
\end{pmatrix} =
\begin{pmatrix}
  -f'_2/f'_1 & f'_1 + f'_2 - d_1 f'_1/f'_2 - d_2 f'_2/f'_2 \\
  0 & -f'_2/f'_2
\end{pmatrix}
\tag{6.6.6}
\]
6.6. THE TELESCOPE

Figure 6.14: Parameters to calculate the paraxial matrix $M'$ for imaging of an object point to an image point with the help of a telescope.

In the case of imaging the parameter $B$ of the matrix has to be zero. So, the condition for the distances $d_1$ and $d_2$ is:

$$f'_1 + f'_2 - d_1 \frac{f'_2}{f'_1} - d_2 \frac{f'_1}{f'_2} = 0 \quad \Rightarrow \quad d_2 = f'_2 + \frac{(f'_2)^2}{f'_1} - d_1 \frac{(f'_2)^2}{(f'_1)^2} \quad (6.6.7)$$

As mentioned above, the image is real if $d_2$ is positive and it is virtual for a negative value $d_2$. So, a real image of a real object point (i.e. $d_1 \geq 0$) means:

$$d_2 \geq 0 \quad \Rightarrow \quad f'_1 + \frac{f'_2}{f'_2} \geq d_1 \geq 0 \quad \Rightarrow \quad \frac{1}{f'_1} + \frac{1}{f'_2} \geq 0 \quad (6.6.8)$$

It can easily be seen that the Galilean telescope (see next paragraph) cannot deliver a real image of a real object point whereas the astronomical telescope delivers real images as long as $0 \leq d_1 \leq f'_1 + f'_2/f'_2$.

A quite interesting property of a telescopic imaging system is the lateral magnification $\beta$ (see equation (2.3.1)). It is according to equation (6.6.6) for the imaging case, i.e. matrix element $B = 0$, equal to the matrix element $A$:

$$\beta = \frac{x'}{x} = -\frac{f'_2}{f'_1} \quad (6.6.9)$$

So, the lateral magnification of the telescopic system depends only on the focal lengths of the two lenses and is independent of the axial position of the object point. If we additionally place the aperture stop into the focal plane of the first lens (only for astronomical telescope) the telescopic imaging system is telecentric (see also page 54).

A very important system is the so called 4f-system with $f := f'_1 = f'_2 > 0$. Then, we have with the help of equation (6.6.7):

$$d_2 = f + f - d_1 \quad \Rightarrow \quad d_1 + d_2 = 2f \quad (6.6.10)$$

This means that the sum of the two distances $d_1$ and $d_2$ is always $2f$ and in total the distance from the object plane to the image plane is $4f$ (in the case of thin lenses where the thickness


Figure 6.15: The aperture stop and the field stop for the imaging of (a) infinite distant objects or (b) finite distant objects (here shown for the case that the object plane is in the front focal plane of the first lens) with the help of an astronomical telescope.

of the lenses can be neglected compared to $4f$) because the length of the telescope has to be added. It also means that for a $4f$–system the shift of the image plane is equal to the shift of the object plane and therefore, the telescope itself can for example be shifted relative to the object and image plane without changing the imaging situation. Of course, in the non–paraxial realm aberrations will change the imaging quality if the telescope of a $4f$–system is moved because the aberrations depend on the actual position of the telescope relative to the object and image plane.

6.6.3 The astronomical and the Galilean telescope

There are two different types of telescopes (see fig. 6.13). The astronomical telescope (also called Kepler telescope, dt.: Astronomisches Fernrohr oder Kepler–Fernrohr) and the Galilean telescope (also called Dutch telescope, dt.: Galilei–Fernrohr oder Holländisches Fernrohr).

6.6.3.1 Astronomical telescope

The astronomical telescope (see fig. 6.13 (a) or fig. 6.15) consists of two positive lenses so that the first lens (called objective, dt.: Objektiv oder Objektiv–Linse) forms a real image of a far
distant object near the focal plane (or exactly in the focal plane for an object with infinite distance). Then, the second lens (called eyepiece, *dt.*: Okular) forms an also infinite distant image but with an increased angular magnification. Since the focal lengths $f'_1$ and $f'_2$ are both positive the angular magnification $\gamma$ is according to equation (6.6.4) $\gamma = -f'_1/f'_2 < 0$. Therefore, the image is upside down so that an astronomical telescope without additional optics to reverse the image is not practical for terrestrial inspections. But, for astronomical purposes or for image transfers in optical systems this is no disadvantage. Additionally, the advantage of the astronomical telescope is that the entrance pupil coincides with the objective in the case of the imaging of infinite distant objects. This means that the exit pupil which is the image of the objective formed by the eyepiece is typically near the focal plane of the eyepiece if $f'_1 \gg f'_2$ which is necessary to obtain an image with angular magnification $|\gamma| \gg 1$. Therefore, the pupil of the eye can be positioned at the exit pupil of the telescope and all light rays with the same off–axis angle (i.e. from the same infinite distant object point) entering the telescope contribute to the image on the retina of the eye. Another advantage of the astronomical telescope is, as mentioned previously, that it can deliver a telecentric real image of objects with a finite distance.

It is quite interesting to think a little bit more about the position of the aperture stop and the field stop for the two cases of imaging of infinite distant objects or finite distant objects (see figure 6.15). As mentioned, the aperture stop for the imaging of infinite distant objects (see fig. 6.15 (a)) is the aperture of the first lens. The field stop lies in this case in the back focal plane of the first lens. For the case of the imaging of finite distant objects (see fig. 6.15 (b)) the situation is different and it is useful to put the aperture stop in the back focal plane of the first lens to have a well–defined numerical aperture for all object points which are not too far away from the optical axis. Then, the aperture of the first lens can act as field stop. Of course, in this case the field stop has no sharp rim because parts of the light cone of points with a similar distance from the optical axis as the radius of the first lens can pass the system if the aperture stop is large enough. In this case a kind of vignetting occurs. So, an additional stop directly in the object or image plane serving as a field stop would be desirable.

For astronomical observations most modern telescopes use mirrors as focussing elements instead of lenses [39]. There are in the meantime telescopes with a primary mirror of $D = 8$ m diameter. From wave optics it is well–known that two (infinite distant) object points can just be resolved by an aberration–free telescope with a diameter $D$ of the primary mirror and a wavelength $\lambda$ of the observed light if their angular separation is larger or equal to $\Delta \varphi$ with

$$\Delta \varphi = k \frac{\lambda}{D}, \quad (6.6.11)$$

where $k$ is a constant of approximately $k = 1$ (for a full circular aperture it is $k = 1.22$). The exact value of $k$ depends on the actual design of the instrument because a reflective telescope has in many cases an annular or more complicate aperture because the secondary mirror and the mounting shadow the central and other parts of the primary mirror. So, a telescope with a large diameter of the primary mirror has of course an increased light–gathering power and an increased angular resolution.

Telescopes on earth have of course the disadvantage that the turbulence of the atmosphere disturbs the resolution. Therefore, modern telescope mirrors are often adaptive mirrors which can be deformed locally by actuators. For the measurement of the necessary deformations a so called Shack–Hartmann wave front sensor is used which observes the light of a distant star (i.e. point object) using the telescope mirror and an eyepiece. The deviations of the resulting wave
front from a plane wave are measured with the help of the Shack–Hartmann sensor and the mirror is deformed until the resulting wave front is plane. The adaptive optics is not only used for correcting the turbulence of the atmosphere, but it is also necessary to correct deformations of a very large mirror due to its large weight if it is moved to be directed to an astronomical object.

6.6.3.2 Galilean telescope

The Galilean telescope (fig. 6.13 (b)) consists of a positive lens (the objective) with the focal length $f'_1 > 0$ and a negative lens (the eyepiece) with focal length $f'_2 < 0$ or $f_2 = -f'_2 > 0$ and $|f'_2| < |f'_1|$. Of course, the telescope can also be rotated by 180 degree so that it reduces the angular magnification. But, in the following we assume that $f'_1 > 0$ and $f'_2 < 0$. The total length of the Galilean telescope is only $|f'_1| - |f'_2|$ (for thin lenses) compared to $|f'_1| + |f'_2|$ for the astronomical telescope (we use here the absolute values of the focal lengths although $f'_1$ is always positive and only $f'_2$ has a different sign for an astronomical and a Galilean telescope).

Another advantage of the Galilean telescope is that the angular magnification $\gamma$ is according to equation (6.6.4) positive: $\gamma = -f'_1/f'_2 > 0$. Therefore, the image is upright and can be directly used for terrestrial inspections. However, a disadvantage of the Galilean telescope is that the image of the objective made by the second lens is between the two lenses. Therefore, the exit pupil of the Galilean telescope is not accessible for the eye and the pupil of the eye works itself as the aperture stop of the complete system whereas the diameter of the objective limits the field. So, Galilean telescopes have a limited field of view and only small magnifications of two to five are useful. Another disadvantage is that the Galilean telescope cannot deliver a real image of a real object. So, a Galilean telescope cannot be used to transport a real intermediate image to another plane in an optical system.

But, the compact overall length and the positive angular magnification provide applications for a Galilean telescope as beam expander or terrestrial telescope like a lorgnette (dt.: Opernglas).

6.7 The Microscope

The last important optical instrument that will be discussed here is a microscope [1],[13],[35]. Whereas, a telescope, especially an astronomical telescope, is used to achieve an angular magnification of distant objects, a microscope is used to obtain a magnified image of a very small near object.

6.7.1 The magnifier

If somebody wants to see details of a small object he brings the object as close to the eye as possible since then the image of the object on the retina of the eye is as large as possible. But, a typical human eye can only form a sharp image of an object for a smallest distance of about $d_S = 25$ cm, which is the standard distance for distinct vision. So, it is obvious that a positive lens, called magnifier (dt.: Lupe oder Vergrößerungsglas), directly in front of the eye can be used to obtain in the distance $|d_I| = d_S$ a magnified virtual image of an object which has itself a smaller distance $|d_O|$ to the eye than the standard distance $d_S$. The image with the distance $d_I$ from the image sided principal plane of the lens with focal length $f'$ has to fulfill the imaging equation (2.4.7) whereby the refractive index on the image side has to be $n' = 1$ because the
6.7. THE MICROSCOPE

Figure 6.16: Principle of a magnifier. A thin lens is used and it is assumed that the refractive indices in the object and image space are both equal so that \( \varphi_I = \varphi_O \).

Human eye is normally used in air and delivers only in this case a sharp image. Then, the imaging equation is:

\[
\frac{1}{d_I} - \frac{n}{d_O} = \frac{1}{f'}
\]  

(6.7.1)

Here, \( n \) is the refractive index on the object side which is often 1 (object in air) but sometimes also larger than 1 if the object is in immersion (for example in water or oil). Due to the sign conventions of geometrical optics \( d_O \) is negative since the object is in front of the lens. The image distance \( d_I \) is also negative for a virtual image. Then, the lateral magnification \( \beta \) of the image is according to equation (2.3.1), equation (6.7.1) and figure 6.16:

\[
\beta = \frac{x_I}{x_O} = \frac{\varphi_I d_I}{\varphi_O d_O} = \frac{n I d_I}{d_O} = 1 - \frac{d_I}{f'} = 1 + \frac{d_S}{f'}
\]  

(6.7.2)

Hereby, it is used that in the paraxial case the angles \( \varphi_I \) and \( \varphi_O \) have to fulfill the condition \( n' \varphi_I = n \varphi_O \), whereby \( n' = 1 \) is valid in our case. Additionally, it is used that the virtual image is at the standard distance for distinct vision so that the image distance \( d_I \), which is negative, is replaced by \(-d_S\), whereby \( d_S \) is the absolute value of the standard distance for distinct vision.

If the lens has for example a focal length \( f' = 5 \text{ cm} \) a lateral magnification of \( \beta = 1 + 25/5 = 6 \) is obtained. In order to have a large field of view without aberrations and especially without chromatic aberrations the magnifier itself is in practice not a single lens but an achromatic combination of different single lenses.

6.7.2 The two-stage microscope

There is of course a limitation for the lateral magnification by using a magnifier because the object has to be very close to the magnifier and therefore also to the eye to achieve large lateral magnifications. Therefore, the **microscope** has been invented which makes a magnification
CHAPTER 6. SOME IMPORTANT OPTICAL INSTRUMENTS

Figure 6.17: Principle of a microscope illustrated by using thin lenses. The objective forms a real magnified intermediate image of the object which is then transformed by the eyepiece in a virtual further magnified image. The distance of this virtual image from the eyepiece and the eye, which is directly behind the eyepiece, has to be the standard distance of distinct vision.

of the object in two stages (see figure 6.17). First, a magnified real image of the object with magnification $\beta_{\text{objective}}$ is formed by using a lens with a small focal length, called objective ($dt.$: Objektiv). This real image is of course inverted. Then, a magnifier, called eyepiece ($dt.$: Okular), with a (mostly) larger focal length is used to form a magnified virtual image of the intermediate real image which is at the standard distance of distinct vision of the eye. The lateral magnification for this second operation is $\beta_{\text{eyepiece}}$. This means, that the lateral magnifications of both operations are multiplied and the total lateral magnification of the microscope $\beta_{\text{microscope}}$ is:

$$\beta_{\text{microscope}} = \beta_{\text{objective}} \beta_{\text{eyepiece}}$$  \hspace{1cm} (6.7.3)

In practice, the objective of a microscope is a quite complex lens consisting of many single lenses to correct the aberrations (especially spherical aberration, coma and chromatic aberrations) of the objective and to guarantee a large field of view[35]. Moreover, modern microscope objectives are corrected for infinity. This means that their aberrations are only corrected if the object is exactly in the object sided focal plane. Therefore, the image would be at infinite distance and an additional lens (called tubus lens) with a fixed focal length (the so called tubus length which is often 160 mm) must be used to get the real image with the magnification imprinted on the objective. For biological investigations, where the object is often covered by a thin coverslip, the spherical aberrations, which result by a high–NA spherical wave passing through a plane–parallel plate, have also to be corrected.

Another very important parameter of the objective is its numerical aperture NA (see equation (3.1.2)). It determines on the one hand the light gathering power of the objective and on the other the resolution which is possible. From wave optics we know that the smallest distance $\Delta x$
between two points which can be resolved by a microscope is

\[ \Delta x = k \frac{\lambda}{NA}, \tag{6.7.4} \]

whereby \( \lambda \) is the wavelength of the used light and \( k \) is a constant (typically about 0.5) which depends on the illumination conditions (coherence) and the exact aperture shape of the objective (mostly circular).

If the image of a microscope has to be on a camera chip (for example CCD chip) a real image has to be on the camera chip. Therefore, the eyepiece which produces a virtual image cannot be used and indeed it is sufficient just to bring the CCD chip at the position of the real image of the objective (plus tubus lens). A typical magnification of a high NA objective in air with for example \( |\beta| = 50 \) is sufficient if a CCD chip with a pixel size of typically 11 \( \mu \text{m} \) is used. This would mean that a structure size of 0.22 \( \mu \text{m} \) on the object is magnified to the size of a pixel of the CCD chip. But, due to equation (6.7.4) 0.22 \( \mu \text{m} \) is approximately the resolution of an objective with \( NA < 1 \) and a wavelength in the visible spectral range. By bringing an immersion oil between the object and the objective which has to be a special immersion objective the NA can be increased up to about 1.4. So, the resolution can be increased accordingly. Another possibility is of course to reduce the wavelength. Modern microscopes for the inspection of integrated circuits use ultraviolet light with a wavelength of 248 nm.
Chapter 7

Radiometry and Photometry

- Up to now the imaging of point–like objects (position and size of the image) have been treated.
- However, from a physical point of view energy (photons) is transmitted in optical imaging.
- **Light sources** have specific properties like:
  - Size of emitting surface/volume
  - Directional characteristics
  - Spectral energy distribution
  - Total emitted light power (spectral light power \( \Phi_{e\lambda} \))

\[
\Phi_e = \int \Phi_{e\lambda}(\lambda)d\lambda
\]

- The **transmission channel** (free–space, optical system) has a transmission factor and other properties which influence the radiation.
- The **detector** (physical detector or human eye) has a certain sensitivity to the incident radiation, for example:

\[
\Phi = K \int \Phi_{e\lambda}V_\lambda d\lambda \tag{7.0.1}
\]

\(V_\lambda\): spectral sensitivity of the detector (for example of the human eye).

7.1 Definition of radiometric and photometric parameters

It has to be distinguished between **physical detectors** (Light power measured in Watt [W]) and the **human eye** (Light flux in Lumen [lm]).

7.1.1 Radiometric parameters

Radiometric parameters (*dt.*: strahlungsphysikalische Größen) are measured with a physical detector which is sensitive to the integrated light power. This can be done with thermal receivers (bolometer). Radiometric parameters are here designated with an index 'e' (coming from 'energy').
7.1. DEFINITION OF RADIOMETRIC AND PHOTOMETRIC PARAMETERS

1. **Radiant flux** (dt.: Strahlungsfluß) \( \Phi_e \) [W]
   total emitted radiation power (≡ energy/time)

2. **Radiant intensity** (dt.: Strahlstärke) \( I_e \) [W/sr] (sr: steradian, full solid angle: \( 4\pi \) sr)
   \[
   I_e = \frac{d\Phi_e}{d\Omega} \quad (7.1.1)
   \]
   \( I_e \) is the part \( d\Phi_e \) of the radiant flux which is emitted into a small solid angle \( d\Omega \).

3. **Irradiance** (dt.: Bestrah lungsstärke) \( E_e \) [W/m\(^2\)]
   \[
   E_e = \frac{d\Phi_e}{dF} \quad (7.1.2)
   \]
   \( E_e \) is the part \( d\Phi_e \) of the radiant flux which illuminates a small detector surface element \( dF \). Sometimes the word intensity is used instead of irradiance although it is not really correct.

4. **Radiance** (dt.: Strahldichte) \( L_e \) [W/(m\(^2\) sr)]
   \[
   L_e = \frac{dI_e}{\cos \vartheta dA} \quad (7.1.3)
   \]
   \( L_e \) is the part \( d\Phi_e \) of the radiant flux which is emitted from the small surface area \( dA \) of the radiation source into the small solid angle \( d\Omega \). \( \vartheta \) is the angle between the surface normal and the direction of the emitted light. The factor \( \cos \vartheta \) is necessary since only the projection of the surface element \( dA \) perpendicular to the direction of emission is relevant.

7.1.2 Photometric parameters (related to human eye)

The photometric parameters (dt.: lichtechnische Größen) are the parameters of an optical system related to the sensitivity of the human eye.

1. **Luminous flux** (dt.: Lichtfluß) \( \Phi \) [lm] (lumen)

2. **Luminous intensity** (dt.: Lichtstärke) \( I \) [cd] (candela, 1 cd = 1 lm/sr)
   \[
   I = \frac{d\Phi}{d\Omega} \quad (7.1.4)
   \]

3. **Illuminance** (dt.: Beleuchtungsstärke) \( E \) [lx] (lux, 1 lx = 1 lm/m\(^2\))
   \[
   E = \frac{d\Phi}{dF} \quad (7.1.5)
   \]

4. **Luminance** (dt.: Leuchtdichte) \( L \) [lm/(m\(^2\) sr)=cd/m\(^2\)]
   \[
   L = \frac{d^2\Phi}{\cos \vartheta d\Omega dA} \quad (7.1.6)
   \]
   Old unit: 1 sb (stilb) = 1 cd/cm\(^2\)
The radiant flux of a thermal detector and the luminous flux of the human eye are connected to each other by taking into account the sensitivity of the eye to different wavelengths:

\[ \Phi = K \int_{380 \text{ nm}}^{780 \text{ nm}} \Phi_e \lambda V(\lambda) d\lambda \]  

(7.1.7)

\( V(\lambda) \): Spectral sensitivity of the standard eye (max. sensitivity at 555 nm) for daylight. \( \Phi_e \lambda \): Spectral light power, i.e. light power \( d\Phi_e \) per small wavelength range \( d\lambda \). For scotopic vision (dt.: Nachtsehen) the maximum sensitivity is shifted to the blue range.

<table>
<thead>
<tr>
<th>( \lambda / \text{nm} )</th>
<th>380</th>
<th>430</th>
<th>510</th>
<th>555</th>
<th>610</th>
<th>633</th>
<th>720</th>
<th>780</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V(\lambda) )</td>
<td>( 10^{-5} )</td>
<td>0.01</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>0.26</td>
<td>0.01</td>
<td>( 10^{-9} )</td>
</tr>
</tbody>
</table>

For photopic vision (dt.: Tagessehen) \( K \) is:

\[ K = 683 \text{ lm/W} \]

### 7.1.3 Some illustrating examples

1. **He–Ne laser**: \( \lambda = 633 \text{ nm} \Rightarrow V(\lambda) = 0.26 \), 1 mW radiant flux
   
   \( \Rightarrow \) Luminous flux \( \Phi = K \Phi_e V(\lambda) = 683 \text{ lm/W} \cdot 1 \text{ mW} \cdot 0.26 = 0.18 \text{ lm} \)
   
   So, the luminous flux is quite small.
   
   **Attention**: Luminous intensity \( I = d\Phi / d\Omega \approx 0.18 \text{ lm}/10^{-6} \text{ sr} = 0.18 \cdot 10^6 \text{ cd} \),

   by assuming a typical solid angle of the laser of \( d\Omega \approx 10^{-6} = (1 \text{ mrad})^2 \).

2. **100 W bulb**: a 100 W bulb produces about 1500 lm luminous flux. But, it has a luminous intensity of only about 1500 lm/4\( \pi \) sr \( \approx 125 \text{ cd} \), since it emits nearly isotropically into the whole solid angle.

3. **Black body radiation**:
   
   - Realization by a hole in a cavity.
   
   \[ S(\lambda) \propto \frac{1}{\lambda^5} \frac{1}{e^{h\omega/(kT)} - 1} \propto \frac{h\omega^3}{e^{h\omega/(kT)} - 1} \]

   - Increasing temperature \( T \Rightarrow \) strong power increase for short wavelengths.
   
   - Total emitted power is proportional to \( T^4 \) (Law of Stefan–Boltzmann)
   
   - The spectral maximum fulfills Wien’s displacement law \( \lambda_m T = \text{const.} \)

Some typical values for the light efficiency (=luminous flux per total expended power like for example electric power)

\[ \frac{\Phi}{\Phi_e} = K \frac{\int_{380 \text{ nm}}^{780 \text{ nm}} \Phi_e \lambda V(\lambda) d\lambda}{\int_0^\infty \Phi_e \lambda d\lambda} \]

| Black body radiation at \( T=6000 \text{ K} \) | 100 lm/W |
| bulbo | 10–20 lm/W |
| fluorescent lamb | 40 lm/W |
7.2. Imaging of light sources

In the following it is not distinguished between radiometric and photometric parameters since the considerations are valid for both types of parameters as long as radiometric and photometric parameters are not mixed with each other. To simplify the notation the photometric parameters without the index ‘e’ are used.

7.2.1 Small (point–like) light source

The solid angle $d\Omega$ is defined as the ratio of the corresponding surface area $dF$ of a sphere and the square of the radius of curvature $a$ (distance from the light source) of this sphere (see figure 7.1):

$$d\Omega = \frac{dF}{a^2}$$

Using this simple relation the illuminance is:

$$E = \frac{d\Phi}{dF} = \frac{d\Phi}{a^2d\Omega} = \frac{I}{a^2} \quad (7.2.1)$$

This is the well–known inverse–square law $E \propto 1/a^2$ (see figure 7.1). However, for a light source with a finite size $D$ this law is only valid if the distance from the light source is larger than a certain threshold value $a_g \gg D$. 

Typical illuminance

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Street lighting</td>
<td>15 lx</td>
</tr>
<tr>
<td>At writing table</td>
<td>300 lx</td>
</tr>
<tr>
<td>Cinema</td>
<td>100 lx</td>
</tr>
<tr>
<td>Lighting for precision work</td>
<td>1000 lx</td>
</tr>
</tbody>
</table>

Figure 7.1: Definition of the solid angle $d\Omega$ and illustration of the inverse–square law of photometry.
7.2.2 Behavior of the luminous intensity in the case of imaging

An object of the diameter $x$ is imaged by an optical system to an image with the diameter $x'$ (see figure 7.2). For small aperture angles the Helmholtz–Lagrange invariant (or Smith–Helmholtz invariant as it is named in the English literature) (2.4.14) is valid:

$$nux = n'u'x' \Rightarrow \beta = \frac{x'}{x} = \frac{u}{u'} \quad \text{for} \quad n = n'$$

Here, $u$ and $u'$ are the aperture angles in the object and image space.

Now, a circular surface element with radius $r$ and surface area $dF$ is considered in the entrance pupil. Then, by taking the definition of the solid angle $d\Omega = a^2d\Omega$ and of the circular area $dF = \pi r^2 = \pi(ua)^2$ the solid angle is:

$$d\Omega = \pi u^2 \quad (7.2.2)$$

On the image side an analogous equation is valid for the solid angle and finally it is:

$$\frac{d\Omega}{d\Omega'} = \frac{\pi u^2}{\pi u'^2} = \beta^2 \Rightarrow d\Omega = \beta^2d\Omega' \quad (7.2.3)$$

By using equation (7.2.2) it is assumed that the angle $u$ and therefore also the solid angle $d\Omega$ are so small that the surface element on a sphere can be replaced with good approximation by a plane circle.

If there is no absorption in the optical system the luminous flux which enters the system in the object space has also to leave the optical system in the image space (energy conservation). This means:

$$I d\Omega = d\Phi = I' d\Omega'$$

Therefore, the behavior of the luminous intensity is:

$$I' = \beta^2 I \quad (7.2.4)$$

So, the luminous intensity increases with the square of the scaling factor. The explanation for this is of course that an increased image results in a decreased aperture angle. Therefore, the light power is emitted into a smaller solid angle and so the luminous intensity is increased.
7.2.3 Practical example: Slide–projector

Figure 7.3 shows the principle of a slide–projector. A condenser lens, which consists in practice of an aspheric collimator lens with a high numerical aperture and a second spherical lens, images the light source into the entrance pupil of the projector objective. The slide behind the condenser lens is imaged by the projector lens into the plane of projection which has a distance $a$ from the projector lens.

A typical condenser lens has a scaling factor of about $\beta = -4 \Rightarrow \beta^2 = 16$ or in practice only $\beta^2 = 11.8$. Since the condenser lens changes the luminous intensity of the light source the illuminance in the projection plane is also changed because the distance $a$ between projector lens and projection plane is given by the desired size of the projected slide. By assuming that the illuminance without condenser lens would be $E = 10$ lx the illuminance with condenser lens is according to equations (7.2.1) and (7.2.4):

$$E' = \frac{I'}{a^2} = \frac{\beta^2 I}{a^2} = \beta^2 E$$

In our example: $E' = 118$ lx. For $a = 4.7$ m this means $I' = 2600$ cd.

7.3 Transition from a point source to an extended light source

In the following, rotational invariance of the optical system is assumed. This means that the luminous intensity $I$ is only a function of the angle $\vartheta$ which is the angle between the regarded direction and the optical axis. So, it is $I = I(\vartheta)$.

7.3.1 Radiator types according to Straubel

Some important radiator types are listed in the following. They are characterized according to their luminous intensity $I$ (see figure 7.4):

$$I(\vartheta) = I_0 \cos^m \vartheta$$  \hspace{1cm} (7.3.1)
Figure 7.4: Polar diagram of the luminous intensity of different radiator types according to Straubel. The length of the arrows gives the value of the luminous intensity in the direction of the angle \( \vartheta \). 

- \( m=0 \): Spherical radiator (dt.: Kugelstrahler)
- \( m=1 \): Lambertian radiator (dt.: Lambertstrahler)
- \( m=3 \): Radiator with a preferential direction (dt.: Keulenstrahler)

The corresponding luminance shall be independent of the position on the emitting surface. Therefore, the differential quotient can be replaced by a normal quotient and the luminance for the different radiators is:

\( m=0 \):

\[
L = \frac{I}{A \cos \vartheta} = \frac{I_0}{A \cos \vartheta}
\]

This means that the luminance increases with \( \vartheta \).

\( m=1 \):

\[
L = \frac{I}{A \cos \vartheta} = \frac{I_0 \cos \vartheta}{A \cos \vartheta} = \frac{I_0}{A} = \text{const.}
\]

So, the luminance is constant for a Lambertian radiator and this means that independent of the direction under which you look onto the surface it will seem to have the same brightness. Of course, the luminous intensity decreases with increasing angle \( \vartheta \) by \( \cos \vartheta \). But, the effective area perpendicular to the direction of vision from which the light comes also decreases by \( \cos \vartheta \) so that the visual impression of brightness remains constant for a Lambertian radiator which has a constant luminance. Therefore, a sphere and a circular plane surface which both emit like a Lambertian radiator cannot be distinguished! The sun is a little bit darker at their rim \( \Rightarrow \) the sun is not exactly a Lambertian radiator. However, many light sources in daily live (for example bulb, fluorescent lamb or LED without additional optics) can be approximated by a Lambertian radiator.
7.3  TRANSITION FROM A POINT SOURCE TO AN EXTENDED LIGHT SOURCE

\( m=3: \)

\[ L = \frac{I}{A \cos \vartheta} = \frac{I_0 \cos^2 \vartheta}{A} \]

So, for \( m=3 \) there is a preferential direction and most of the light is emitted to small angles \( \vartheta \). So, such a radiator is useful if a small solid angle shall be illuminated efficiently.

### 7.3.2  Luminous flux in a \( 2\pi \) solid angle

The luminous flux in the half space, i.e. \( 2\pi \) solid angle, is obtained by an integration of the luminous intensity over this solid angle:

\[ \Phi_{\text{half}} = \int_0^{2\pi} d\varphi \int_0^{\pi/2} I(\vartheta) \sin \vartheta d\vartheta = 2\pi \int_0^{\pi/2} I(\vartheta) \sin \vartheta d\vartheta \]

Using the equation of Straubel (7.3.1) for the different radiator types gives:

\[ \Phi_{\text{half}} = 2\pi I_0 \int_0^{\pi/2} \cos^m \vartheta \sin \vartheta d\vartheta = -2\pi I_0 \int_1^{\vartheta} \cos^m \vartheta d(\cos \vartheta) \]

\[ \Rightarrow \Phi_{\text{half}} = \frac{2\pi I_0}{m+1} \Rightarrow \frac{I_0}{\Phi_{\text{half}}} = \frac{m+1}{2\pi} \quad (7.3.2) \]

Comparison of the different radiator types

A light converting efficiency of 13 lm/W is assumed. Then, the luminous intensity \( I_0 \) along the symmetry axis \( \vartheta=0 \) is for the different radiator types (assuming that light is emitted in both half spaces, i.e. \( 4\pi \) solid angle):

- \( m=0 \)  1 cd/W
- \( m=1 \)  2 cd/W
- \( m=3 \)  4 cd/W

So, for \( m=3 \) only \( 1/4 \) of the total luminous flux or electric power is needed compared to a spherical radiator to achieve the same luminous intensity on–axis!

### 7.3.3  Illuminance on a plane screen for free space propagation

The light from a small extended light source propagates in free space to a plane screen which is perpendicular to the optical axis, where the optical axis is defined as the surface normal of the extended light source. The distance from the light source to the screen along the optical axis is \( a_0 \) and the distance from the light source to a certain point of the screen is \( a \) (see figure 7.5).

Here, the term ”small extended light source” means that the distance \( a_0 \) has to be much larger than the size of the light source. Then, for a small detector element \( dF \) it is (see figure 7.5):

\[ d\Omega = \frac{dF \cos \vartheta}{a^2} = \frac{dF}{a_0^2} \cos^3 \vartheta \quad (7.3.3) \]

So, the luminous flux which is emitted into the small solid angle \( d\Omega \) is:

\[ d\Phi = I(\vartheta)d\Omega = I(\vartheta)\frac{dF}{a_0^2} \cos^3 \vartheta \quad (7.3.4) \]
Therefore, the illuminance $E(\vartheta)$ on the screen is:

$$E(\vartheta) = \frac{d\Phi}{dF} = \frac{I(\vartheta)}{a_0^2} \cos^3 \vartheta$$

With the help of equation (7.3.1) of Straubel and using the abbreviation $E(0) = E_0 = I_0/a_0^2$ (illuminance on–axis) the final result is:

$$E(\vartheta) = E_0 \cos^{m+3} \vartheta$$

(7.3.5)

For the important case of a Lambertian radiator it is:

$$E(\vartheta) = E_0 \cos^4 \vartheta$$

(7.3.6)

Example: $\vartheta = 45^\circ \Rightarrow E(45^\circ) = E_0 \sqrt{1/2^4} = E_0/4$, which means that there is a strong decrease of the illuminance at the rim of the screen.

This is a well–known effect for example in photography. If a normal flashlight is used the rim of a scene is often not illuminated very well because of this behavior of a Lambertian radiator during free space propagation. Therefore, modern flash lamps try to illuminate a scene as homogeneous as possible by shaping the light with optical elements like for example special free–form surfaces which are fabricated in practice like Fresnel lenses with several local segments. Each segment is in some cases just a local prism which deflects the light in a certain direction.

### 7.3.4 Behavior of the radiator types in the case of imaging

Especially, the **Lambertian radiator**, i.e. $I(\vartheta) = I_0 \cos \vartheta$, will be treated. It is assumed that the imaging is **without spherical aberration** and using a small extended light source.

Figure 7.5: Illustration of the behavior of the illuminance on a plane screen for light which propagates from a small extended light source to the screen in free space.
Regard an annular small solid angle $d\Omega$ ($d\Omega'$) in the object (image) space which has the angle $\vartheta$ ($\vartheta'$) with the optical axis (see figure 7.6):

$$d\Omega = 2\pi \sin \vartheta \, d\vartheta \quad \text{and} \quad d\Omega' = 2\pi \sin \vartheta' \, d\vartheta' \quad (7.3.7)$$

Using energy conservation (assuming no absorption) it is:

$$d\Phi = d\Phi' \quad \Rightarrow \quad I \, d\Omega = I' \, d\Omega'$$

$$\Rightarrow \quad 2\pi I(\vartheta) \sin \vartheta \, d\vartheta = 2\pi I'(\vartheta') \sin \vartheta' \, d\vartheta'$$

Now, it is requested that a Lambertian radiator shall be imaged to another Lambertian radiator. This requires:

$$I_0 \cos \vartheta \sin \vartheta \, d\vartheta = I'_0 \cos \vartheta' \sin \vartheta' \, d\vartheta' \quad \Rightarrow \quad I_0 d(\sin^2 \vartheta) = I'_0 d(\sin^2 \vartheta')$$

From equation (7.2.4) which has been calculated using paraxial optics (which is sufficient in this case since it is on-axis) it is known $I'_0 = \beta^2 I_0$. So, it is required:

$$d(\sin^2 \vartheta) = \beta^2 d(\sin^2 \vartheta') \quad \Rightarrow \quad \sin^2 \vartheta = \beta^2 \sin^2 \vartheta' \quad (7.3.8)$$

This is the case if the optical imaging system fulfills the sine condition (3.1.5) (here, for $n = n'$)

$$x \sin \vartheta = x' \sin \vartheta'$$

Result: If an optical imaging system fulfills the sine condition a Lambertian radiator is imaged to another Lambertian radiator.
**7.3.5 Illuminance in the image**

A small solid angle around the optical axis is taken. Then, there is energy conservation in the case of no absorption and the luminous flux $d\Phi$ in the object space is identical to the luminous flux $d\Phi'$ in the image space:

$$d\Phi = I d\Omega = d\Phi'$$

The solid angles in the object and image space are related to each other via equation (7.2.3)

$$d\Omega = \beta^2 d\Omega'$$

and due to the definition of the scaling factor $\beta$ the surface areas $A$ in object space and $A'$ in image space are connected by:

$$A' = x'^2 = \beta^2 x^2 = \beta^2 A.$$  

So, the illuminance in the image space is:

$$E' = \frac{d\Phi'}{A'} = \frac{d\Phi}{\beta^2 A} = \frac{I d\Omega}{\beta^2 A} = \frac{I d\Omega'}{A}$$

Per definition the luminance $L$ of the light source in a small area around the optical axis ($\cos \vartheta \approx 1$) is

$$L = \frac{I}{A}$$

and finally the result is:

$$E' = L d\Omega'$$

**Result:** The illuminance $E'$ in the image of a small extended light source with surface area $A$ is equal to the product of the luminance $L$ of the light source and the illuminating solid angle $d\Omega'$. 

Figure 7.7: Illustration of the behavior of the illuminance in the image of an extended light source.
7.3. TRANSITION FROM A POINT SOURCE TO AN EXTENDED LIGHT SOURCE

So, only the solid angle with which the lens is seen from the image appears in the equation. The aperture angle in the object space is not contained in the equation!

**Interpretation:** The lens radiates with the luminance $L$ into the solid angle $d\Omega'$ if the observer is in the image of the extended light source!

**Example 1:** Burning glass
Lambert stated long ago that the illuminance in the image of the sun is as large as if the burning glass would radiate with the luminance of the sun. **But attention:** This is not valid for the half space ($2\pi$ sr) but only for a small image of the sun. Therefore, the statement is not completely exact.

The angular extension of the sun is about $2u' = 0.01$. So, the illuminance without lens is (using equation (7.2.2)):

$$E_{\text{no lens}}' = Ld\Omega' = L\pi u'^2$$

If $F_{EP}$ is the surface area of the exit pupil of the lens and $f$ the focal length of the lens, then the result with lens is:

$$E_{\text{with lens}}' = L\frac{F_{EP}}{f^2} = L\pi \left( \frac{D_{EP}}{2f} \right)^2 = \frac{\pi}{4} \frac{L}{f^#} \quad (7.3.10)$$

$f#$ is the F number $f/D_{EP}$ ($D_{EP}$: diameter of the exit pupil). Here, it is assumed that the exit pupil coincides with the principal plane $H'$ of the lens what is the case for a burning glass.

The ratio of the illuminance is:

$$\frac{E_{\text{with lens}}'}{E_{\text{no lens}}'} = \frac{1}{(2u'f^#)^2} = \left( \frac{D_{EP}}{2fu'} \right)^2$$

In the case of the sun:

$$\frac{E_{\text{with lens}}'}{E_{\text{no lens}}'} = \left( \frac{100D_{EP}}{f} \right)^2$$

For $f# = f/D_{EP} = 1 \Rightarrow E_{\text{with lens}}'/E_{\text{no lens}}' = 10^4$.

So, the concentrating effect of the lens is proportional to the inverse square of the F number.

**Example 2:** Taking a picture of a landscape or the moon
The moon has nearly the same luminance like a sunny landscape on earth since both are secondary radiators which are illuminated by the sun. So, for a camera both objects are nearly at infinity and therefore, also the illuminance on the camera detector is nearly identical for both objects.

**Example 3:** The sun
The irradiance of the sun near the earth (but above the atmosphere of the earth) is

$$E_e = 1.3 \text{ kW/m}^2$$

The angular extension of the sun on earth is about 0.5 degree or more exactly $2u' = 0.0092$ (measured in radians). Therefore, the solid angle of the sun seen from earth is $d\Omega = \pi u'^2$. Therefore, the radianc $L_S$ of the sun is:

$$L_S = \frac{E_e}{d\Omega} = 19.3 \text{ MW/m}^2\text{sr}$$
From this the light power $M_h$ per surface element which is emitted by the sun into the half space (away from the sun) can be calculated:

$$M_h = \frac{d\Phi_h}{dA} = 2\pi \int_0^{\pi/2} L_S \cos \vartheta \sin \vartheta d\vartheta = -2\pi L_S \int_1^0 x dx = \pi L_S = 60.7 \text{ MW/m}^2$$

Here, the factor $\cos \vartheta$ during the integration is due to the fact that $L_S = \frac{d^2\Phi_e}{(d\Omega dA \cos \vartheta)}$. So, $L_S \cos \vartheta = \frac{d^2\Phi_e}{(d\Omega dA)}$ has to be integrated. Additionally, it is assumed that the sun is a Lambertian radiator with constant $L_S$. So, the sun emits per m$^2$ of its surface about 60 MW light power.

Of course, this value can also be calculated in another way if the radius of the sun ($r_S = 6.94 \cdot 10^8$ m) and the distance of the earth from the sun ($r_E = 1.50 \cdot 10^{11}$ m) are known. The total radiant flux of the sun is obtained by multiplying the solar constant (=irradiance of the sun near the earth) with the surface area of a sphere around the sun which has the radius of curvature $r_E$:

$$\Phi_S = 4\pi r_E^2 E_e = 3.68 \cdot 10^{20} \text{ MW}$$

The total radiant flux of the sun divided by its surface area delivers the radiant flux per surface element of the sun:

$$M_h = \frac{\Phi_S}{4\pi r_S^2} = 60.7 \text{ MW/m}^2$$

As expected, the values of both algorithms are identical.

### 7.3.6 Difference between the imaging of an extended (sun) and a point–like (star) light source

- **Point source** ⇒ Airy disc
- **Extended light source** ⇒ Geometrical optical laws of imaging

Airy disc for an astronomical telescope

For an astronomical telescope (or more exactly an astronomical camera since only the primary
mirror is used, see figure 7.8) the aperture angle \( \varphi' \) in the image is for an infinitely distant object (for example star):
\[
\varphi' = \frac{D}{2f}
\]  
(7.3.11)

Here, \( D \) is the diameter of the telescope mirror/lens and \( f \) its focal length. Here, the approximation for small aperture angles \( \varphi' \) is used (\( \sin \varphi' \approx \tan \varphi' \approx \varphi' \)), which is valid for astronomical telescopes which have a small numerical aperture. The Airy disc (see the lecture about wave optics) which is then formed in the focal plane of the telescope has a surface area (area limited by the ring of the first minimum of the Airy disc) of:
\[
F_{\text{Airy}} = \pi \rho'^2 \quad \text{with} \quad \rho' = 0.61 \frac{\lambda}{\sin \varphi'} \approx 0.61 \frac{\lambda}{\varphi'}
\]  
(7.3.12)

\( \lambda \) is the wavelength of the used light.

### 7.3.6.1 Contrast improvement in a telescope

Let \( E_0 \) be the irradiance of the star in the entrance pupil of the telescope mirror and \( E_* \) the irradiance in the image plane (focal plane). Of course, the irradiance in the image plane is not constant since an Airy disc is formed. But, \( E_* \) is defined as the irradiance averaged over the area of the Airy disc. Then, the radiant flux is:
\[
\Phi = E_0 \frac{\pi}{4} D^2
\]
\[
E_* = \frac{\Phi}{F_{\text{Airy}}} = \frac{E_0 D^2 \pi/4}{\pi (0.61)^2 \lambda^2 4 f^2 / D^2} = \frac{E_0 D^4}{(2.44)^2 \lambda^2 f^2}
\]  
(7.3.13)

Therefore:
\[
E_* \propto E_0 D^4
\]  
(7.3.14)

The irradiance \( E_B \) of the background radiation with radiance \( L_B \) is:
\[
E_B = L_B d\Omega' = L_B \pi \varphi'^2 = L_B \frac{\pi D^2}{4 f^2} \propto D^2
\]

So, the ratio of the irradiance coming from the star and the irradiance coming from the background radiation is:
\[
\frac{E_*}{E_B} = \frac{E_0 D^2}{(1.22)^2 \pi \lambda^2 L_B} \propto D^2
\]  
(7.3.15)

So, the contrast between the image of the star and the background radiation increases with the square of the aperture diameter \( D \) (i.e. with the surface area of the mirror) of the telescope.

### 7.3.7 Imaging with a telescope

In this subsection only extended objects are treated and no point objects like stars. The object shall be circular, but this is of course no restriction.

The objective of the telescope has a diameter \( 2p \) and the two lenses of the telescope have focal lengths of \( f_1 \) and \( f_2 \) (see figure 7.9). The angular magnification is \( \Gamma = \tan w_t / \tan w_e = f_1 / f_2 \)
(\(w_t\): half angular extension of the object with telescope, \(w_e\): half angular extension of the object with the pure eye). The observer looks with relaxed eye through the telescope, i.e. the telescope images an infinitely distant object. The exit pupil of the telescope has a diameter \(2p' = 2p/\Gamma\).

Retina image without telescope:
On the retina of the eye with focal length \(f_e\) the object has the radius \(x_e\) and covers the surface area \(A_e\):

\[
x_e = w_e f_e \Rightarrow A_e = \pi x_e^2 = \pi w_e^2 f_e^2
\]

Retina image with telescope:
Now, the object has the radius \(x_t\) on the retina and covers the surface area \(A_t\). Here, again an approximation for small angles is used:

\[
w_t = \Gamma w_e \Rightarrow x_t = w_t f_e = \Gamma w_e f_e \Rightarrow A_t = \pi x_t^2 = \pi \Gamma^2 w_e^2 f_e^2
\]

So, the ratio of the surface areas on the retina is:

\[
\frac{A_t}{A_e} = \frac{\Gamma^2 w_e^2 f_e^2}{\pi w_e^2 f_e^2} = \Gamma^2 \tag{7.3.16}
\]

The radiant flux coming from the far distant object which enters the eye without instrument (\(\Phi_e\)) and which enters the telescope (\(\Phi_t\)) are:

\[
\text{Eye: } \Phi_e = \pi \rho^2 E_0 \\
\text{Telescope: } \Phi_t = \pi p^2 E_0
\]

Here, \(E_0\) is the irradiance of the object and \(\rho\) is the radius of the eye pupil.

Irradiance on the retina
To get the brightness of the image of the object (i.e. irradiance on the retina) the following cases have to be distinguished:
7.4. PHOTOMETRIC UNITS

1. $p' \leq \rho$: All light which enters the telescope also enters the eye. Then, it is:

$$\frac{E_t}{E_e} = \frac{\Phi_t}{\Phi_e} \frac{A_t}{A_e} = \frac{p^2}{\rho^2} = \frac{p'^2}{\rho^2} \leq 1 \quad (7.3.17)$$

2. $p' > \rho$: The eye pupil is now the limiting exit pupil and the effective radius of the objective of the telescope from which light enters the eye is only $p_{\text{eff}} = \Gamma \rho$. Then, the following is valid:

$$\frac{E_t}{E_e} = \frac{\Phi_t}{\Phi_e} \frac{A_t}{A_e} = \frac{p_{\text{eff}}^2}{\rho^2 \Gamma^2} = \frac{\Gamma^2 \rho^2}{\Gamma^2 \rho^2} = 1 \quad (7.3.18)$$

This means that the brightness of the image of an extended far distant object (for example the sun) on the retina cannot be increased by using a telescope. However, the image is larger and shows more details.

Attention by observing the sun: Although, the irradiance of an extended object cannot be increased by a telescope it is absolutely forbidden to observe the sun through a telescope without using filters. In fact, also the observation of the sun with the pure eye is dangerous and can cause severe damages of the eye. But, the image of the sun on the retina (focal length of the eye is about 20 mm) has without a telescope about 0.2 mm diameter. The fovea centralis (the region of sharp vision of the eye) has a diameter of about 1.5 mm. So, the image of the sun on the retina is much smaller than the fovea and the small automatic movements of the eye distribute the light power additionally. However, if a telescope is used the image of the sun on the retina covers more than the whole area of the fovea and the total amount of light power is so high that there are thermal damages since the eye cannot dissipate this high amount of light power.

7.4 Photometric units

7.4.1 The candela

The basic unit of photometry is the candela, the unit of luminous intensity $I$. The modern definition of the candela is:

The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian.

It has to be mentioned that the frequency of $540 \cdot 10^{12}$ hertz corresponds to a wavelength of about 555 nm, where the eye has its maximum sensitivity. So, the spectral sensitivity of the eye is $V(\lambda = 555 \text{ nm}) = 1$.

In former days, the reference standard of the photometric units was the black body radiation at the melting temperature of platinum at 2045 K. A black body radiator is a Lambertian radiator and its luminance is independent of the direction. So, the old definition of the candela is:

1 cd is the luminous intensity of a black body radiator perpendicular to its surface at 2045 K and having a surface area of $1/60 \text{ cm}^2$. This means that 1 cm$^2$ of this black body radiator emits with 60 cd perpendicular to its surface.
CHAPTER 7. RADIOMETRY AND PHOTOMETRY

Figure 7.10: Illustration of the general radiation formula showing a two-dimensional section. In principle the surface elements can also be tilted with respect to the connecting line by arbitrary angles out of the plane. The radiator element with surface area \( dA \) emits light which is received by the detector element with surface area \( dF \).

### 7.4.2 Luminance \( \bar{L} \) of the black body radiation at 2045 K

\[
\bar{L} = 60 \, \text{sb} = 60 \frac{\text{cd}}{\text{cm}^2}
\]

So, the basic unit of luminance 1 cd/m\(^2\) is:

\[
1 \frac{\text{cd}}{\text{m}^2} = \frac{1}{60 \cdot 10^4} \bar{L}
\]

Luminance of some light sources (\( L \) in sb)

<table>
<thead>
<tr>
<th>Light Source</th>
<th>Luminance (L in sb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun</td>
<td>120000</td>
</tr>
<tr>
<td>Xe high pressure lamp</td>
<td>250000</td>
</tr>
<tr>
<td>Scenery illuminated by the sun</td>
<td>( 1/3 - 5 )</td>
</tr>
<tr>
<td>Light bulb</td>
<td>200 – 2000</td>
</tr>
<tr>
<td>Opal light bulb</td>
<td>5</td>
</tr>
<tr>
<td>Fluorescent lamp</td>
<td>( \approx 1 )</td>
</tr>
</tbody>
</table>

### 7.4.3 Luminous flux \( \Phi \)

It is:

\[
1 \text{lm} = 1 \text{cd} \cdot 1 \text{sr}
\]

This means that 1 lumen corresponds to the luminous flux which is emitted in a solid angle of 1 sr at a constant, i.e. isotropic, luminous intensity of 1 cd.

### 7.5 Generalization of the radiation formulas

A radiator with a small element of surface area \( dA \) emits to a detector element with surface area \( dF \) (see figure 7.10). The distance between both elements is \( r \) and the angles of the surface normals of the elements to the connecting line of both elements are \( \epsilon_r \) (for radiator) and \( \epsilon_d \) (for detector).

Then, for the total luminous flux there is an equation which is symmetric with respect to radiator and detector:

\[
\Phi = \int_A \int_F \frac{L}{r^2} dA \cos \epsilon_r \, dF \cos \epsilon_d
\]  \hspace{1cm} (7.5.1)
Here, \( L, r, \epsilon_r \) and \( \epsilon_d \) depend in the general case on the surface elements. For small surface areas \( A \) and \( F \) with distance \( r = a \) this results in:

\[
\Phi_{rd} = L \frac{A \cos \epsilon_r F \cos \epsilon_d}{a^2}
\] (7.5.2)

This equation is sufficient for many estimations.

### 7.5.1 Applications

#### 7.5.1.1 Surface elements within a sphere

Regard two (small) surface elements with areas \( F_a \) and \( F_b \) at the interior surface of a sphere (see figure 7.11), as it is for example the case in a so called Ulbricht’s sphere. Then, due to symmetry reasons it is \( \epsilon_r = \epsilon_d =: \epsilon \) and:

\[
\Phi_{ab} = L \frac{F_a F_b \cos^2 \epsilon}{a^2}
\]

By looking at figure 7.11 it is:

\[
a = 2 R \cos \epsilon
\]

\[
\Rightarrow \quad \Phi_{ab} = L \frac{F_a F_b \cos^2 \epsilon}{(2R)^2 \cos^2 \epsilon} = L \frac{F_a F_b}{4R^2}
\] (7.5.3)

This means that independent of the position of the two surface elements on the sphere the same luminous flux is exchanged between the two elements if it is a Lambertian radiator, i.e. if the luminance \( L \) is independent of the direction \( \epsilon \). Also the illuminance on the detector is independent of the position of the surface elements since:

\[
E_b = \frac{\Phi_{ab}}{F_b} = L \frac{F_a}{4R^2}
\]
This method is used for the (integral) measurement of light since each surface element radiates in the same way. Also the luminance \( L \) can be detected with this method, where the calibration is made with a reference light source.

### 7.5.1.2 Decrease of the illuminance by imaging a screen

Regard the imaging of a plane homogeneously illuminated screen by a lens which is free from distortion (see figure 7.12). The aperture stop of the lens shall be nearly in the principle plane of the lens on the object side. By regarding the chief ray it follows

\[
\epsilon_r = \epsilon_d = w
\]

and the distance \( \rho \) of the surface element with area \( A \) to the center of the aperture stop is

\[
\rho = \frac{s}{\cos w},
\]

where \( s \) is the distance of the screen to the aperture stop on-axis. Therefore, \( s \) is also the distance of the object to the principal plane.

The total luminous flux emitted by the surface element with area \( A \) (Lambertian radiator) which hits the aperture stop with surface area \( F \) is:

\[
\Phi_{AF} = L \frac{AF}{\rho^2} \cos^2 w = L \frac{AF}{s^2} \cos^4 w = \Phi_{AF}(w = 0) \cos^4 w
\]

Since the lens shall be free from distortion it is guaranteed that the image area \( A' \) is connected to the area \( A \) of the radiating surface element by the scaling factor \( \beta \) and \( \beta \) is independent of \( w \):

\[
A' = \beta^2 A
\]
7.6. INVARiance OF THE LUMINANCE BY IMAGING

Figure 7.13: Scheme to illustrate the invariance of the luminance by imaging with a lens. $s$ and $s'$ are the object and image distance, $F$ is the surface area of the lens aperture which is assumed here to be small.

Now, the total luminous flux which transmits the aperture stop has also to be in the image. Therefore, the illuminance in the image is:

$$E(w) = \frac{\Phi_A F}{A'} = \frac{F}{\beta^2 s^2} L \cos^4 w$$

And finally:

$$E(w) = E_0 \cos^4 w$$  \hspace{1cm} (7.5.4)

Here, $E_0 = LF/(\beta^2 s^2)$ is the illuminance on-axis for $w = 0$. Further on it is:

$$\text{Image distance: } s' = \beta s \quad \Rightarrow \quad E_0 = \frac{L F}{s'^2} = L \Omega',$$

where $\Omega'$ is the solid angle under which the aperture stop is seen from the on-axis image point.

7.6 Invariance of the luminance by imaging

**Statement:** The luminance $L$ cannot be increased by imaging.

In the derivation a small lens aperture with the surface area $F$ is taken (see figure 7.13 for the special case $n = n'$). The surface area $A$ of the radiating light source is also assumed to be small. So, the derivative in the definition of the luminance can be replaced by normal quotients. Additionally, the ray bundle is in a small region around the optical axis so that the cosine factor is one. So, the luminance $L$ and $L'$ in the object and image space are:

$$L = \frac{\Phi}{\Omega A} = \frac{\Phi s^2}{FA} \quad \Rightarrow \quad \Phi = LF \frac{A}{s^2}$$

$$L' = \frac{\Phi'}{\Omega' A'} = \frac{\Phi' s'^2}{FA'} \quad \Rightarrow \quad \Phi' = L' F \frac{A'}{s'^2}$$
The definition of the scaling factor $\beta$ delivers:

$$s' = \beta s \quad \text{and} \quad A' = \beta^2 A \quad \Rightarrow \quad \frac{A'}{s'^2} = \frac{A}{s^2}$$

Using energy conservation $\Phi = \Phi'$ results in:

$$L' = L$$  \hspace{1cm} (7.6.1)

### 7.6.1 More general treatment with different refractive indices

Now, a more general treatment with different refractive indices $n$ and $n'$ in object and image space is made. Again, figure 7.13 illustrates the different parameters. First, some invariants are used:

1. Helmholtz–Lagrange–invariant (equation (3.1.6))
   $$n xu = n' x'u'$$

2. $$us = h = u's' \quad \Rightarrow \quad \frac{u}{u'} = \frac{s'}{s}$$

This means:

$$\beta = \frac{x'}{x} = \frac{n u}{n'u'} = \frac{n}{n'} \frac{s'}{s} \quad \Rightarrow \quad s' = \beta s \frac{n'}{n}$$

The relation between the surface areas in object and image space is again (assuming circular surfaces):

$$A' = \pi (x'/2)^2 = \pi \beta^2 (x/2)^2 = \beta^2 A$$

Using energy conservation $\Phi' = \Phi$ the luminance is:

$$L' = \frac{\Phi' s'^2}{A' F} = \frac{\Phi s^2 n'^2}{A F n^2} = \frac{n'^2}{n^2} L$$

Therefore, the effective luminance is invariant:

$$L_{\text{eff}} = \frac{L}{n^2} = \frac{L'}{n'^2}$$  \hspace{1cm} (7.6.3)

Summarizing, the following statements are valid by imaging of an extended object:

- $\Phi$ and $L_{\text{eff}}$ are invariant
- $E$ changes with $1/\beta^2$
  $$E' = \frac{\Phi'}{A'} = \frac{\Phi}{\beta^2 A} = \frac{E}{\beta^2}$$
- $I$ changes with $\beta^2 n'^2/n^2$
  $$I' = \frac{\Phi'}{\Omega'} = \frac{\Phi' s'^2}{F} = \frac{\Phi \beta^2 s^2 n'^2}{F n^2} = I \beta^2 \frac{n'^2}{n^2}$$
7.7 Etendue

The etendue (dt.: Lichtleitwert oder geometrischer Fluß) is another invariant in optics. It is a purely geometrical parameter. It is assumed that there is no absorption, scattering or reflection during the light propagation.

It has been shown that the following parameters are invariants:

- Radiant flux $\Phi$
- Effective radiance $L_{\text{eff}}$

But then, there is also another invariant:

$$\frac{\Phi}{L_{\text{eff}}} \quad \text{Invariant of energy propagation}$$

In the following, only small surface areas for the radiating surface and the detector are assumed. Additionally, the refractive index is everywhere $n = 1$. So, $L_{\text{eff}}$ can be replaced by $L$.

Equation (7.5.2) is used which calculates the radiant flux $d\Phi$ on the detector surface $F$ with area $dF$ which is emitted by the radiating surface $A$ with area $dA$. Both surfaces have the distance $a$ and the surface normals enclose with the connecting line an angle $\epsilon_r$ for surface $A$ and an angle $\epsilon_d$ for surface $F$. Then it is:

$$\frac{d\Phi}{L} = \frac{dA \cos \epsilon_r \, dF \cos \epsilon_d}{a^2} = \Lambda$$

(7.7.1)

So, the so called etendue $\Lambda$ is totally symmetric with respect to the surfaces $A$ and $F$ and it is a purely geometrical parameter (see figure 7.14). An exchange of $A$ and $F$ will not change anything. By using the concept of the etendue equation (7.7.1) is formally equal to Ohm’s law of electricity $I = U/R$ (intensity of electric current $I$, voltage $U$ and conductance $1/R$ or electric resistance $R$), whereby the radiant flux $\Phi$ corresponds to $I$, the radiance $L$ to $U$ and the etendue $\Lambda$ to $1/R$. 

Figure 7.14: Scheme to illustrate the etendue. The angles $\epsilon_r$ or $\epsilon_d$ between the surface normals of surface $A$ or $F$ and the connecting line of both surfaces are not shown.
7.8 Scatterer plate in the ray path

An ideal scatterer plate (possible realization: small stochastically distributed etched negative lenses) is a Lambertian radiator with (see equation (7.3.1)):

\[ I'(\vartheta') = I'_0 \cos \vartheta' \]

The radiant flux which is emitted in the whole half space is according to equation (7.3.2) for a Lambertian radiator \((m=1)\):

\[ \Phi_{\text{half}} = \pi I'_0 \]

An ideal scatterer plate does not absorb any radiation so that the incident radiant flux \(d\Phi\) (which has a preferred direction) is also completely emitted (but without a preferred direction):

\[ d\Phi = d\Phi'_{\text{half}} \]

In front of the scatterer plate with the surface element \(dF\) the irradiance \(E\) is (see figure 7.15):

\[ E = \frac{d\Phi}{dF} \]

Behind the scatterer plate it is:

\[ L' = \frac{dI'}{\cos \vartheta' dF} \implies dI'_0 = L' dF \]

This means:

\[ E \, dF = d\Phi = d\Phi'_{\text{half}} = \pi dI'_0 = \pi L' \, dF \]

\[ \implies L' = \frac{1}{\pi} E \] (7.8.1)

An old unit (the apostilb) for the luminance, which is no longer used, was defined as: 1 asb = \(\pi^{-1} \cdot 10^{-4}\) sb = \(\pi^{-1} \cdot 10^{-4}\) cd/cm\(^2\) = \(\pi^{-1}\) cd/m\(^2\).

So, for the case of a scatterer plate the luminance in asb is equal to the illuminance in lx. This is for example important to calculate the luminance on a cinema screen.
Bibliography


Index

Abbe number, 76
Aberrations, 68
   Astigmatism, 73
   Chromatic, 75
   Coma, 72
   Curvature of field, 74
   Distortion, 74
   Point aberrations, 71
   Seidel terms, 71
   Spherical aberration, 72
   Zernike polynomials, 75
Achromatic lens, 86
Afocal system, 98
Angular magnification, 22
Aperture stop, 51
   for telescope, 101
Aplanatic meniscus, 84, 86
Aplanatic points of a sphere, 84
Ball lens, 39
Camera, 90, 95
   Astronomical, 91
Cardinal points of an optical system, 21
Chief ray, 54
Chromatic aberrations, 75
Computer generated hologram, 80

Depth of field, 91
Diffractive lens, 77
DOE, 77
   Amplitude hologram, 80
   Glossary, 80
   Phase function, 82
   Phase hologram, 80
Eikonal equation, 1, 3
   Limitations, 5
Entrance pupil, 51
   Entrance window, 55
   Etendue, 127
   Exit pupil, 51
   Exit window, 55
   Extended light source, 111
   Eye, 95
   F number, 92
   Field stop, 55
      for telescope, 101
   Focal length, 22, 36
      Relation between object and image space, 27
      Thick lens, 38
   Focal planes, 22
   Focal points, 22
   Fresnel lens, 77
   Fresnel zone lens, 77
      Grating model, 78
Geometrical optics, III
   Energy conservation, 8
   Limitations, 1
   Grating equation, 44
Helmholtz–Lagrange invariant, 34
Hoegh’s meniscus, 40

Image construction, graphically, 32
Imaging condition, 24
Imaging equation, 31
Lambertian radiator, 112
Lateral magnification, 21, 32
Law of reflection, 10
Law of refraction, 9, 10
   paraxial, 17
   Lens equation, 31
   Light tube, 8
<table>
<thead>
<tr>
<th>Index Entry</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing method</td>
<td>34</td>
</tr>
<tr>
<td>Longitudinal magnification</td>
<td>32</td>
</tr>
<tr>
<td>Magnifier, 102</td>
<td></td>
</tr>
<tr>
<td>Maxwell equations</td>
<td></td>
</tr>
<tr>
<td>in isotropic and linear materials</td>
<td>1</td>
</tr>
<tr>
<td>Material equations</td>
<td>2</td>
</tr>
<tr>
<td>Time–independent form</td>
<td>2</td>
</tr>
<tr>
<td>Transition to geometrical optics</td>
<td>1</td>
</tr>
<tr>
<td>Meridional plane</td>
<td>12</td>
</tr>
<tr>
<td>Microscope, 102, 103</td>
<td></td>
</tr>
<tr>
<td>Eyepiece, 104</td>
<td></td>
</tr>
<tr>
<td>Numerical aperture</td>
<td>104</td>
</tr>
<tr>
<td>Objective, 104</td>
<td></td>
</tr>
<tr>
<td>Newton equation</td>
<td>31</td>
</tr>
<tr>
<td>Nodal points, 22, 25, 28</td>
<td></td>
</tr>
<tr>
<td>Numerical aperture</td>
<td>54</td>
</tr>
<tr>
<td>Optical axis</td>
<td>12</td>
</tr>
<tr>
<td>Optical imaging</td>
<td>13</td>
</tr>
<tr>
<td>Optical ray, 3</td>
<td></td>
</tr>
<tr>
<td>Orthogonality condition</td>
<td>4</td>
</tr>
<tr>
<td>Paraxial geometrical optics</td>
<td>12</td>
</tr>
<tr>
<td>Paraxial matrix theory</td>
<td>15</td>
</tr>
<tr>
<td>3x3 matrices</td>
<td>45</td>
</tr>
<tr>
<td>Plane parallel plate</td>
<td>18</td>
</tr>
<tr>
<td>Refraction at plane surface</td>
<td>17</td>
</tr>
<tr>
<td>Sign conventions</td>
<td>19</td>
</tr>
<tr>
<td>Thin prism</td>
<td>47</td>
</tr>
<tr>
<td>Transfer matrix</td>
<td>16</td>
</tr>
<tr>
<td>Paraxial ray</td>
<td>13</td>
</tr>
<tr>
<td>Definition</td>
<td>15</td>
</tr>
<tr>
<td>Paraxial ray tracing</td>
<td>13</td>
</tr>
<tr>
<td>at diffraction grating</td>
<td>44</td>
</tr>
<tr>
<td>Phase function</td>
<td>82</td>
</tr>
<tr>
<td>Photography</td>
<td>94</td>
</tr>
<tr>
<td>Photometry, 106, 107</td>
<td></td>
</tr>
<tr>
<td>Illuminance, 107</td>
<td></td>
</tr>
<tr>
<td>Luminance, 107</td>
<td></td>
</tr>
<tr>
<td>Luminous flux, 107</td>
<td></td>
</tr>
<tr>
<td>Luminous intensity, 107</td>
<td></td>
</tr>
<tr>
<td>Plane wave</td>
<td>7</td>
</tr>
<tr>
<td>Plane–concave lens</td>
<td>41</td>
</tr>
<tr>
<td>Plane–convex lens</td>
<td>41</td>
</tr>
<tr>
<td>Polarization ray tracing</td>
<td>4</td>
</tr>
<tr>
<td>Principal planes</td>
<td>22</td>
</tr>
<tr>
<td>Principal points</td>
<td>22, 24, 28</td>
</tr>
<tr>
<td>Pupils, 51</td>
<td></td>
</tr>
<tr>
<td>Entrance pupil</td>
<td>51</td>
</tr>
<tr>
<td>Exit pupil</td>
<td>51</td>
</tr>
<tr>
<td>Radiometry, 106</td>
<td></td>
</tr>
<tr>
<td>Irradiance, 107</td>
<td></td>
</tr>
<tr>
<td>Radiance, 107</td>
<td></td>
</tr>
<tr>
<td>Radiant flux, 107</td>
<td></td>
</tr>
<tr>
<td>Radiant intensity, 107</td>
<td></td>
</tr>
<tr>
<td>Ray aberration</td>
<td>13</td>
</tr>
<tr>
<td>Ray aberrations</td>
<td>68</td>
</tr>
<tr>
<td>Ray equation</td>
<td>4</td>
</tr>
<tr>
<td>GRIN materials</td>
<td>5</td>
</tr>
<tr>
<td>Homogeneous materials</td>
<td>5</td>
</tr>
<tr>
<td>Ray tracing</td>
<td></td>
</tr>
<tr>
<td>Coordinate transformation</td>
<td>61</td>
</tr>
<tr>
<td>Description of a ray</td>
<td>58</td>
</tr>
<tr>
<td>Differential ray tracing</td>
<td>66</td>
</tr>
<tr>
<td>Law of reflection</td>
<td>65</td>
</tr>
<tr>
<td>Law of refraction</td>
<td>64</td>
</tr>
<tr>
<td>Non–sequential</td>
<td>66</td>
</tr>
<tr>
<td>Optical path length</td>
<td>63</td>
</tr>
<tr>
<td>Point of intersection with a surface</td>
<td>59</td>
</tr>
<tr>
<td>Polarization ray tracing</td>
<td>4</td>
</tr>
<tr>
<td>Principle</td>
<td>56</td>
</tr>
<tr>
<td>Ray aberrations</td>
<td>70</td>
</tr>
<tr>
<td>Surface normal</td>
<td>63</td>
</tr>
<tr>
<td>Wave aberrations</td>
<td>68</td>
</tr>
<tr>
<td>Rays, 3</td>
<td></td>
</tr>
<tr>
<td>Refraction</td>
<td></td>
</tr>
<tr>
<td>at tilted plane surface</td>
<td>46</td>
</tr>
<tr>
<td>Refractive index</td>
<td>3</td>
</tr>
<tr>
<td>Sagittal plane</td>
<td>13</td>
</tr>
<tr>
<td>Sagittal rays</td>
<td>13</td>
</tr>
<tr>
<td>Seidel aberrations</td>
<td>71</td>
</tr>
<tr>
<td>Smith–Helmholtz invariant</td>
<td>34</td>
</tr>
<tr>
<td>Snell’s law, 9, 10</td>
<td></td>
</tr>
<tr>
<td>Spherical wave</td>
<td>7</td>
</tr>
<tr>
<td>Spot diagram, 70</td>
<td></td>
</tr>
<tr>
<td>Stops, 51</td>
<td></td>
</tr>
<tr>
<td>Aperture stop</td>
<td>51</td>
</tr>
<tr>
<td>Field stop</td>
<td>55</td>
</tr>
<tr>
<td>Telescope</td>
<td>96</td>
</tr>
</tbody>
</table>
astronomical, 100
Galilean, 102
Thick lens, 37
  Ball lens, 39
  Hoegh’s meniscus, 40
  in air, 39
  Plane-concave, 41
  Plane-convex, 41
Thin lens, 35
  Focal length, 36
Thin prism, 47
Total internal reflection, 10
Unit plane, 22
Vignetting, 55
Wave aberration, 15
Wave aberrations, 68
Zernike polynomials, 75