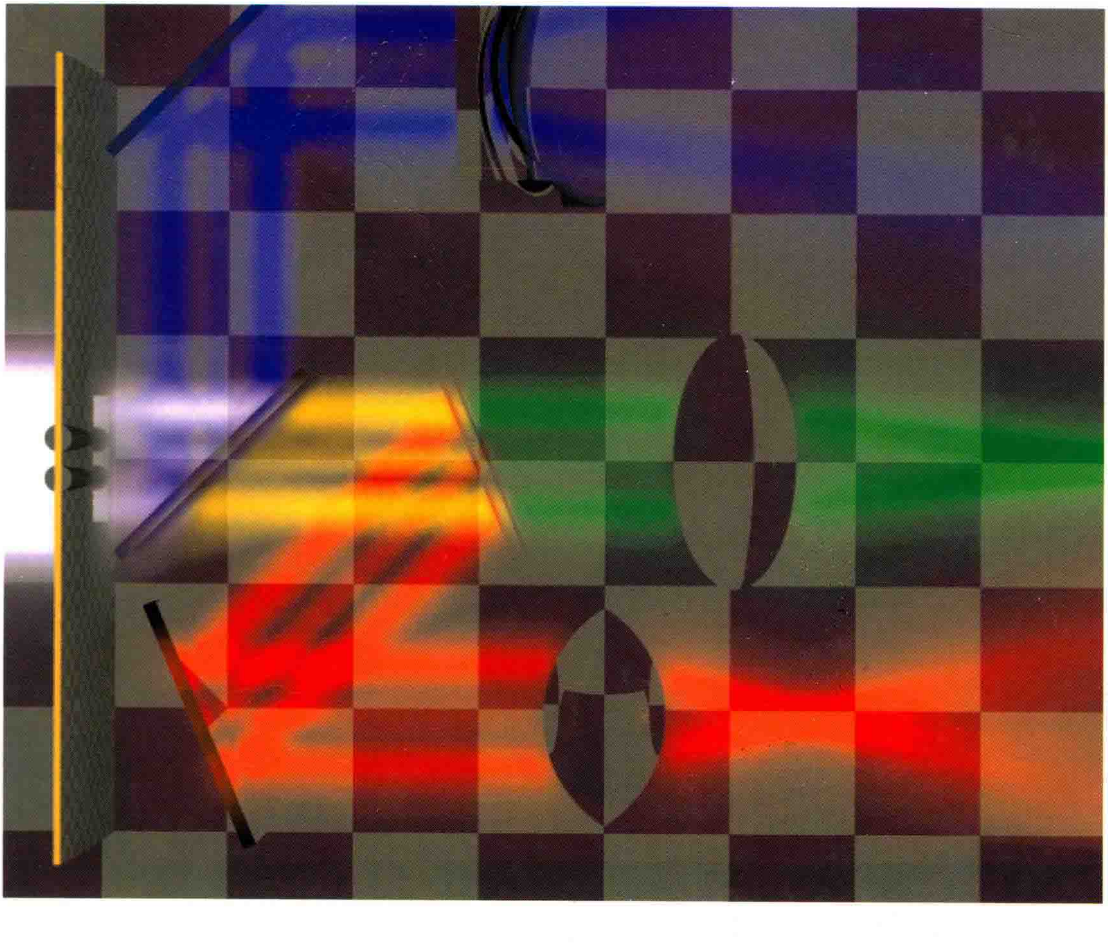
Orestes N. Stavroudis

WILEY-VCH

The Mathematics of Geometrical and Physical Optics

The k-function and its Ramifications



Orestes N. Stavroudis

The Mathematics of Geometrical and Physical Optics

The k-function and its Ramifications

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

The Author
Prof. Dr. Orestes N. Stavroudis
Centro de Investigaciones en Optica,
León, Guanajuato, Mexico
ostavro@cio.mx

Cover picture
Persistence of Vision Raytracer Version 3.5
Sample File
Author: Christopher J. Huff

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Printed in the Federal Republic of Germany

Typesetting Steingraeber Satztechnik GmbH, Ladenburg

Printing Strauss GmbH, Mörlenbach

Binding Schäffer GmbH, Grünstadt

ISBN-13: 978-3-527-40448-3
ISBN-10: 3-527-40448-1

Orestes N. Stavroudis

The Mathematics of Geometrical and Physical Optics

Acknowledgements

Many hands make the work light. It is my pleasure to acknowledge and thank those whose efforts made this work possible. These are (in alphabetic order): Maximiliano Avendaño-Alejo, Isidro Cornejo, Lark London, Christopher Stavroudis, Dorle Stavroudis, and many former students whose helpful comments and snide remarks were of immensurable value. Thanks are also due to my project editor, Ulrike Werner, who skillfully and tactfully, led me in the proper direction.

So I pass from a task, which has filled the greater part of many years of my life, which has broadened in my view as they passed, and which has suffered interruptions that threatened to end it before its completion. Many of its defects are known to me; after it has gone from me, others will become apparent. Nevertheless, my hope is that my work will ease the labour of those who, coming after me, may desire to possess a systematic account of this branch of pure mathematics.

A. R. Forsyth
Trinity College, Cambridge
October, 1906.

Introduction

This work is about geometrical optics though it shall extend into some fundamental areas of physical optics as well. It makes heavy use of several branches of mathematics which, perhaps, the reader will find disturbingly unfamiliar. These I will describe with some care but with only lip service to mathematical rigor and vigor.

Keep in mind that geometrical optics is a peculiar science. Its fundamental artifacts are rays, which do not exist, and wavefronts, which indeed do exist but are not directly observable. A third item is the caustic, a surface in image space which is certainly observable, defined variously as the envelope of an array of rays associated with some point object, the locus of the principal centers of wavefront curvatures, or as the locus of points where the differential element of area of a wavefront vanishes. Of course, these wavefronts must be in a wavefront train generated by a lens and associated with some fixed object point.

The peculiarities of geometrical optics go even further. Rays, which do not exist, are trajectories of corpuscles, which also do not exist. These trajectories, according to the principle of Fermat, are those paths over which the time of transit of a corpuscle, passing from one point to another, is either a maximum or a minimum.

Yet it works. Geometrical optics, anachronistic as it is, remains the basis for modern optical design, the highly successful engineering application built on the sandiest foundation imaginable. There is hardly one area of modern science in which instruments are used whose design depends ultimately on Fermat's postulate on the intrinsic laziness of mother nature.

In what follows I shall use a method best described as axiomatic, the axiom being Fermat's principle. This we must modify, however. Since point-to-point transit times can be maxima as well as minima we must use, in the language of the Calculus of Variations, *extrema* (singular: *extremum*) as our criterion in applying Fermat's principle.

Indeed the interpretation of the principle of Fermat in terms of the language of the variational calculus will lead us to ray paths in *inhomogeneous* media; media in which the refractive index

The Mathematics of Geometrical and Physical Optics: The k-function and its Ramifications. O.N. Stavroudis

ISBN: 3-527-40448-1

is a continuous function of position. These ray paths will be expressed in the form of a system of ordinary differential equations that can be applied to any specified media.

These ray paths are then subject to analysis using the techniques of the differential geometry of space curves. Using these differential equations for a ray path we can deduce its shape and its relationship to the refracting medium itself. From these results we can determine, quickly and easily, the nature of rays in, say, Maxwell's fish eye.

From here we pass on to the Hilbert integral, developed originally for dealing with the problem of the variable end point in the Calculus of Variations. This very rich theory leads us to a number of very important deductions in geometrical optics; conditions for the existence of wavefronts, Snell's law, The Hamilton-Jacobi equations (though both Hamilton and Jacobi preceded Hilbert by as much as a half century), the eikonal equation, among others. In this context the theorem of Malus becomes trivial. From this context Herzberger recognized the importance of the *normal congruence* or the *orthotomic system* of rays.

With the concept of the wavefront in hand we proceed to the differential geometry of surfaces and to partial differential equations of the first order. One such is the eikonal equation, mentioned above, obtained from the Hamilton-Jacobi equation, for which we find a general solution descriptive of any wavefront train in a *homogeneous optical medium*; one with a constant refractive index.

In terms of the differential geometry of surfaces we can find, for the general wavefront train, wavefront principal directions and curvatures. This leads to the important concept of the caustic, that surface that is the locus of the principal centers of wavefront curvature. In the caustic resides all of the monochromatic aberrations associated with a wavefront train and, ultimately, with the lens and object point that give rise to it. The structure of this caustic describes completely the *image errors*: *spherical aberration, coma* and *astigmatism*. Its location in space indicates the *field errors*; *distortion* and *field curvature*.

Along the way we look at *generalized ray tracing*, more properly, a generalization of the *Coddington equations*, that determines the principal directions and principal curvatures at any point on a wavefront through which a traced ray passes.

This we apply to the prolate spheroid, a rotationally symmetric ellipsoid generated by rotating an ellipse about its major axis. This leads to a reflecting optical system, consisting of two confocal spheroids, that I have called the *modern schiefspiegler*.

We also look at Herzberger's fundamental optical invariant and his diapoint theory and apply it to the representation of wavefronts obtained form the solution of the eikonal equation. This leads to a hierarchial system of aberrations.

The canon that I have described here, based on Fermat's principle, omits many important items. Outstanding among these is *paraxial theory* and *paraxial ray tracing*. Although is it is of tremendous practical importance, it is based on an approximation that, in my opinion, does not belong here.

A far more fundamental omission is Gaussian optics, in particularly, its model as developed by Maxwell. He began with certain assumptions about perfect lenses from which he represented perfect image formation by a fractional-linear transformation. Upon assuming that his perfect lens is rotationally symmetric, he was able to derive its cardinal points; the foci, the nodal points and the principal points.

Other omissions are the *Seidel aberrations* and their higher order extensions. These are from a solution of the eikonal equation in the form of a power series that has never been shown to converge.

Huygens' principle is omitted. It is clearly independent of any corpuscular concepts and is based on wavefront propagation as the envelope of spherical wavelets, which also do not exist, centered on a previous position of the wavefront. It also leads to Snell's law. It was for many centuries the main competitor to Fermat's corpuscules.

But nowadays the photon incorporates the best of both the corpuscule and the wavelet, a compromise that has resulted in a far more useful theory with applications far beyond the dreams of Fermat and Huygens.

Contents

The Mathematics of Geometrical and Physical Optics: The k-function and its Ramifications. O.N. Stavroudis

ISBN: 3-527-40448-1

Part I

Preliminaries

The Mathematics of Geometrical and Physical Optics: The k-function and its Ramifications. O.N. Stavroudis

ISBN: 3-527-40448-1

1 Fermat's Principle and the Variational Calculus

In the seventeenth century light was believed to be a flow of *corpuscules*, 'little bodies'; their trajectories were called *rays*. Pierre de Fermat asserted that Nature was intrinsically lazy and that those corpuscules 'chose' a trajectory that made their time of transit from point to point a minimum. We refer to this anthropomorphism as *Fermat's Principle*. It was a successful hypothesis. With it, Fermat was able to derive the law of refraction, *Snell's law*, in an economical and precise way.[1]

The connection between Optics and the variational calculus came some years after Fermat when the Swiss mathematician Jacob Bernoulli proposed a problem, the *brachistochrone*, and offered a prize for its solution. Consider a rigid wire connecting a pair of points, fixed in space, on which a bead slides under the force of gravity but without friction. The problem was to find that shape of the wire for which the time of transit of the bead, from one point to the other, was a minimum.[2]

The connection between geometrical optics and Fermat's principle is clear. Jacob's solution was to calculate the vertical force on the bead, taking into account the constraint imposed by the rigid wire. He related this force to an index of refraction function that depended on the height of the bead on the wire. He partitioned the space between the initial and terminal points into horizontal lamina each having a constant refractive index that was determined by its height. Then he could use Snell's law to trace a ray down from the initial point, resulting in a polygonal ray path that approximated the desired solution. As the number of lamina increased and as each thickness approached zero, the polygonal figure approached a continuous curve which was the desired shape of the rigid wire. This curve turned out to be an arc of a cycloid.[3]

Jacob Bernoulli was very pleased with his solution, so much so that he awarded to himself the prize that he had offered, and disregarded the efforts of his brother Jean, who also solved the brachistochrone problem, from an entirely different point of view.

Jean made use of the newly discovered differential calculus and the fact that the first derivative of a function vanishes at its maximum or minimum value. He expressed the time of transit of the bead from the initial point to its terminal point as a an integral of the reciprocal of its velocity. The first derivative of this integral must vanish at a minimum and he obtained conditions that the solution curve must satisfy. Subsequently Leonard Euler extended Jean's

[1] Sabra 1967, Chapter V. An account of the history and background of Fermat's principle.

[2] Bliss 1925, pp. 65–72. Caratheodory 1989, pp. 235–236 uses the Hamiltonian which we will encounter in Chapter 3. Woodhouse 1964, Chapters I and II provides a more detailed historical account. Courant & Robbins 1996, pp. 381–384. In Smith 1959, pp. 644–655 there is an English translation of Bernoulli's original paper and announcement.

[3] Bliss 1946, Chapter VI. Jean's use of the calculus in generating the variational calculus.

The Mathematics of Geometrical and Physical Optics: The k-function and its Ramifications. O.N. Stavroudis

ISBN: 3-527-40448-1

method to more general problems and obtained differential equations for their solution. Jean's method can rightfully be called the beginning of modern Calculus of Variations.[4]

It is natural to refer to a solution of a variational problem as an *extremal arc* or more simply as an *extremal*. We will interpret the principle of Fermat in terms of the language of the variational calculus and apply modern mathematics to that basic axiom of geometrical optics and develop it as far as we can.

1.1 Rays in Inhomogeneous Media

We have seen that the basic assumption of geometrical optics is Fermat's principle: A ray path that connects two points in any medium is that path for which the time of transit is an extremum. To be more explicit, out of the totality of all possible paths connecting the two points, A and B, a ray is that unique path for which the time of transit is either a maximum or a minimum. Of course if A and B are conjugates, if B is a perfect image of A, then the ray path is not unique; every ray passing through A must also pass through B.

The time of transit between two points, A and B, is given by the equation

$$T = \int_A^B dt = \int_A^B \frac{ds}{v} = \int_A^B \frac{nds}{c}, \tag{1.1}$$

where c is the velocity of light *in vacuo*, v its velocity in the medium through which it propagates and n the refractive index of that medium. The arc length along the ray or trajectory is s. The optical medium is said to be *homogeneous* if n is constant; it is *inhomogeneous* but *isotropic* if n is a function of position. It is *anisotropic* if the refractive index of the medium depends on the ray's direction.

The convention most used is to drop c from the equations and to use the *optical path length* I, instead of the *time of transit* T, as the variational integral. Thus

$$I = \int_A^B nds. \tag{1.2}$$

In what follows we take the medium to be inhomogeneous so that the refractive index is a function of position $n = n(x,\ y,\ z)$. A possible path connecting A and B is given parametrically by the three coordinate functions $x(t),\ y(t),\ z(t)$ where the choice of the parameter t is entirely arbitrary. If A has the coordinates $(a_1,\ a_2,\ a_3)$ and B, $(b_1,\ b_2, b_3)$ then it must be that

$$\begin{aligned} x(t_0) &= a_1, \quad y(t_0) = a_2, \quad z(t_0) = a_3, \\ x(t_1) &= b_1, \quad y(t_1) = b_2, \quad z(t_1) = b_3, \end{aligned} \tag{1.3}$$

[4]Bliss 1946, Chapter I. Bolza 1961, Chapter 1. Clegg 1968, Chapter 3.

so that

$$I(A,\ B) = \int_{t_0}^{t_1} n(x,\ y,\ z)ds = \int_{t_0}^{t_1} n(x,\ y,\ z)\frac{ds}{dt}dt, \tag{1.4}$$

where the Pythagorean theorem gives us

$$\frac{ds}{dt} = s_t = \sqrt{x_t^2 + y_t^2 + z_t^2}. \tag{1.5}$$

Here, the subscript $(_t)$ denotes differentiation with respect to the parameter t. This subscript notation for both ordinary and partial differentiation will be used extensively in what follows.

In these terms then the problem is to find that curve, given by $x(t),\ y(t),\ z(t)$, for which $I(A,\ B)$ is an extremum.

1.2 The Calculus of Variations

This problem is a special case of a more general problem that belongs to that body of mathematics known as the Calculus of Variations. That more general problem is to find the curve in space, given by $y(x),\ z(x)$ for which the integral

$$I = \int_a^b f\big(x,\ y(x),\ z(x),\ y_x(x),\ z_x(x)\big)dx, \tag{1.6}$$

is an extremum. The function f is always known since it is determined by the nature of the problem; for example, in Eq. 1.4, f is equal to $n(x,\ y,\ z)ds/dt$.

Here we need to find expressions for $y(x)$ and $z(x)$ that make Eq. 1.6 an extremum. First assume that $\overline{y}(x)$ and $\overline{z}(x)$ represent a solution, a curve for which Eq. 1.6 is an extremum. In addition let $\eta(x),\ \zeta(x)$ be any two functions, sufficiently differentiable, such that

$$\begin{aligned} \eta(a) &= \eta(b) = 0, \\ \zeta(a) &= \zeta(b) = 0. \end{aligned} \tag{1.7}$$

Now form a one-parameter family of curves given by

$$y(x) = \overline{y}(x) + h\,\eta(x), \qquad z(x) = \overline{z}(x) + h\,\zeta(x), \tag{1.8}$$

where h is the parameter. By virtue of Eq. 1.7 these curves all pass through the end points of the integral; when the parameter h is *zero* we have, by definition, the solution curve. We replace $y(x)$ and $z(x)$ in the variational integral, Eq. 1.6, by using Eq. 1.8 to get

$$\begin{aligned} I(h) \ &= \int_a^b f\Big(x,\ \overline{y}(x) + h\ \eta(x),\ \overline{z}(x) + h\,\zeta(x), \\ &\qquad \overline{y}_x(x) + h\ \eta_x(x),\ \overline{z}_x(x) + h\ \zeta_x(x)\Big)dx. \end{aligned} \tag{1.9}$$