

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

**THE CRITERIA FOR CORRECTION OF
QUADRATIC FIELD-DEPENDENT
ABERRATIONS**

by

Chunyu Zhao

Copyright © Chunyu Zhao 2002

A Dissertation Submitted to the Faculty of the
COMMITTEE ON OPTICAL SCIENCES (GRADUATE)

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2002

UMI Number: 3040142

Copyright 2002 by
Zhao, Chunyu

All rights reserved.

UMI[®]

UMI Microform 3040142

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

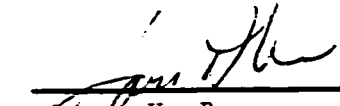
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ©
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Chunyu Zhao

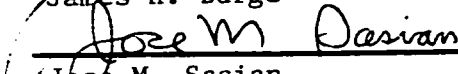
entitled The Criteria for Correction of Quadratic Field - Dependent Aberrations

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy



James H. Burge

Dec 6, 2001
Date



Jose M. Sasian

12-06-01
Date



Roland V. Shack

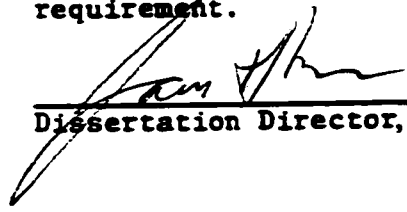
12-06-01
Date

Date

Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.



Dissertation Director, James H. Burge

Dec 6 2001
Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotation from this dissertation is allowable without special permission, provided that accurate acknowledgment of source is made. Request for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED:

A handwritten signature in black ink, appearing to read "Chungyi Zhao", written over a horizontal line.

ACKNOWLEDGMENTS

I would like to thank my adviser Dr. Jim Burge. Dr. Burge has a keen insight in optical design. It is he who suggested I go one step beyond the Sine Condition. His suggestion led to this research. He insisted that I extend my work into the plane symmetric system while I was not confident in getting any meaningful result. Without his insistence, this research would not be complete. For four years and a half since I came to the Optical Sciences Center, Dr. Burge has taught me a lot about optics. What I appreciate most is he taught me not only optics, but everything on how to pursue a successful career in an environment that is so different than what I was familiar with. Besides, I cannot imagine how I could have survived the severe accident without his support.

I thank Professor Jose Sasian and Professor Roland Shack. I learned the basics of optical design from them. When I came across difficulties in my research, they offered great help.

It was fun to work with Erin, Dave, Brian and Feenix. Their enthusiasm in optics and technology in general often motivated me while my spirit was low. I also appreciated their help during the toughest time of my life.

Most importantly, I thank my wife, Fang, for all the joy and care she gives me. Her calm at a turbulent moment helped me win the 2001 Michael Kidger Memorial Scholarship, and her patience at another crucial moment helped me avoid major surgery.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	10
LIST OF TABLES	16
ABSTRACT	17
INTRODUCTION	18
CHAPTER 1	
ABERRATIONS AND HAMILTON'S CHARACTERISTIC FUNCTIONS	21
1.1 INTRODUCTION.....	21
1.2 INTRODUCTION TO ABERRATIONS	21
1.2A RAY ABERRATION AND WAVE ABERRATION.....	23
1.2B RELATIONS BETWEEN WAVE AND RAY ABERRATIONS	25
1.2C THE PUPIL ABERRATIONS	25
1.2D ANOTHER WAY TO GROUP WAVE ABERRATIONS.....	26
1.3 INTRODUCTION TO HAMILTON'S CHARACTERISTIC FUNCTIONS.....	28
1.3A DEFINITION OF CHARACTERISTIC FUNCTIONS	29
1.3B TAYLOR SERIES EXPANSION OF THE CHARACTERISTIC FUNCTIONS	31
1.4 CONCLUSION	33
CHAPTER 2	
THE PUPIL ASTIGMATISM CONDITIONS FOR CORRECTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS	34

TABLE OF CONTENTS – *continued*

	Page
2.1 INTRODUCTION.....	34
2.2 DEPENDENCE OF QUADRATIC FIELD-DEPENDENT ABERRATIONS ON ENTRANCE PUPIL ASTIGMATISM.....	36
2.2.1A SUMMARY FOR SYSTEMS WITH FINITE CONJUGATES	36
2.2.1B DERIVATION OF PUPIL ASTIGMATISM CONDITIONS FOR SYSTEMS WITH FINITE CONJUGATES.....	38
2.2.2A INFINITE CONJUGATE SYSTEMS.....	43
2.2.2B DERIVATION OF THE PUPIL ASTIGMATISM CONDITIONS FOR THE INFINITE CONJUGATE SYSTEM	45
2.3 FORWARD RAY-TRACING VERSION OF THE PUPIL ASTIGMATISM CRITERIA	49
2.3A THE FINITE CONJUGATE SYSTEM.....	50
2.3B THE INFINITE CONJUGATE SYSTEM.....	55
2.4 ANOTHER METHOD TO DERIVE THE FORWARD RAY TRACING VERSION OF THE CRITERIA	60
2.4A FINITE CONJUGATES SYSTEM	60
2.4B INFINITE CONJUGATES SYSTEM	65
2.5 COMPARING THE PUPIL ASTIGMATISM CRITERIA WITH THE CODDINGTON EQUATIONS	66
2.6 CONCLUSION	67
CHAPTER 3	
VALIDATION OF THE PUPIL ASTIGMATISM CRITERIA TO 3RD ORDER	68
3.1 INTRODUCTION.....	68
3.2 IMAGE AND PUPIL SEIDEL ABERRATIONS AND THEIR RELATIONS ...	69

TABLE OF CONTENTS -- *continued*

	Page
3.3 VALIDATION OF THE PUPIL ASTIGMATISM CRITERIA TO 3 RD ORDER	74
3.4 CONCLUSION	77
 CHAPTER 4	
USING THE PUPIL ASTIGMATISM CRITERIA TO ANALYZE THE KNOWN DESIGN EXAMPLES	78
4.1 INTRODUCTION.....	78
4.2 THE APLANATIC CONJUGATE PAIR OF A REFRACTIVE SPHERE	79
4.3 THE LUNEBURG LENS	82
4.4 THE RETROREFLECTIVE CONCENTRIC SYSTEMS	85
4.5 CONCLUSION	89
 CHAPTER 5	
PREDICTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS.....	90
5.1 INTRODUCTION.....	90
5.2 FORMULAS FOR CALCULATING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS.....	90
5.3 SPHERICAL MIRROR	93
5.4 A 4-ELEMENT SYMMETRIC SYSTEM	97
5.5 A SYSTEM WITH FLAT TANGENTIAL IMAGE.....	101
5.6 THE APLANAIC CONJUGATE PAIR OF A REFRACTIVE SPHERE.....	104
5.7 CONCLUSION	108
 CHAPTER 6	
IMPLEMENTING THE PUPIL ASTIGMATISM CRITERIA FOR DESIGN OF OPTICAL SYSTEM	109

TABLE OF CONTENTS -- *continued*

	Page
6.1 INTRODUCTION.....	109
6.2 THE CONSTANT OPL CONDITION AND THE SINE CONDITION	109
6.3 THE PUPIL ASIGMATISM CRITERIA FOR CORRECTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS.....	114
6.4 THE CODDINGTON EQUATIONS	115
6.5 DETERMINE R_t AND R_r FOR GENERAL ASPHERIC SURFACES	116
6.6 DETERMINE THE SURFACE SLOPE WHEN THE RAY DIRECTIONS BEFORE AND AFTER A SURFACE ARE KNOWN.....	117
6.7 NUMBER OF DEGREES OF FREEDOM	118
6.8 DESCRIPTION OF THE ALGORITHM.....	119
6.8A 3-SURFACE SYSTEMS	119
6.8B 4-SURFACE SYSTEMS	121
6.9 EXAMPLE DESIGNS	122
6.10 CONCLUSION	130
 CHAPTER 7	
THE PUPIL ASTIGMATISM CONDITIONS FOR PLANE-SYMMETRIC SYSTEMS	132
7.1 INTRODUCTION.....	132
7.2 THE PLANE SYMMETRIC SYSTEM.....	132
7.3 THE HAMILTON'S CHARACTERISTIC FUNCTIONS FOR PLANE SYMMETRIC SYSTEMS	134
7.4 THE PUPIL ASTIGMATISM CONDITION FOR PLANE SYMMETRIC SYSTEMS.....	136

TABLE OF CONTENTS – *continued*

	Page
7.5 CONCLUSION	143
CHAPTER 8	
CONCLUSION	144
APPENDIX A	
SOURCE FILES FOR DESIGNING OPTICAL SYSTEMS NUMERICALLY	146
APPENDIX B	
SOURCE FILES FOR THE ZEMAX USER DEFINED SURFACE.....	173
REFERENCES	190

LIST OF FIGURES

	Page
Figure 1.1. Illustration of wave aberration and ray aberration.	22
Figure 1.2. Illustration of the pupils. The Entrance Pupil is the image of stop in object space, and the Exit Pupil is the image of stop in image space.	26
Figure 1.3. Interpretation of Hamilton's characteristic functions when the initial and final media are homogeneous.....	29
Figure 2.1. Trace a thin bundle of parallel rays backwards through an aplanatic optical system (finite conjugate case). T is the tangential focus and S is the sagittal focus.	36
Figure 2.2. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ for a finite conjugate system in (a) the tangential plane, and (b) the sagittal plane.....	41
Figure 2.3. Trace a thin bundle of parallel rays backwards through an aplanatic optical system (infinite conjugate case). T is the tangential focus and S is the sagittal focus.....	44
Figure 2.4. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ for an infinite conjugate system in (a) the tangential plane, and (b) the sagittal plane.....	47
Figure 2.5. Trace a thin bundle of parallel rays forwards through an optical system (finite conjugate case). T is the tangential focus and S is the sagittal focus.	51
Figure 2.6. The side view of an optical system with the object a finite distance away. T is the tangential focus of a thin bundle of parallel rays traced backwards through the optical system. T' is its counterpart when a thin bundle of parallel rays are traced forward through the optical system.....	51
Figure 2.7. An anastigmatic optical system with magnification -2.25	54

LIST OF FIGURES -- *continued*

	Page
Figure 2.8. Illustration of a system with object at infinity. A small cone of rays are traced from a point A in a plane which is perpendicular to the optical axis. The tangential image of the cone is T and the sagittal image of the cone is S	56
Figure 2.9. The side view of an optical system with object at infinity. T is the tangential image of the cone of rays originating from a point A in the object space. .	57
Figure 2.10. The Luneburg Lens.	59
Figure 2.11. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ in (a) the tangential plane, and (b) the sagittal plane by tracing rays forward through an optical system with object a finite distance away.	61
Figure 2.12. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ in the tangential plane by tracing rays forward through the optical system with object at infinity....	65
Figure 3.1. Illustration of the ray angles and heights at a surface.	70
Figure 3.2. Illustrates a system with entrance pupil at infinity, s' and t' denote the sagittal and tangential pupil curves respectively.	73
Figure 4.1. The aplanatic conjugate pair of a refractive sphere.	79
Figure 4.2. Illustration of calculating s and t for the aplanatic conjugate pair of a refractive sphere.	80
Figure 4.3 Plots of s , t and $t/\cos^2\theta$ as functions of θ for the aplanatic system shown in Figure 4.1.....	81
Figure 4.4. (a) The Luneburg lens. Object at infinity is perfectly imaged onto a spherical surface. (b) The index of refraction profile of the Luneburg Lens	82
Figure 4.5 Trace a thin bundle of rays from the image point I backward through the Luneburg Lens. The tangential and sagittal foci coincide on the front surface of the lens.	83

LIST OF FIGURES -- *continued*

	Page
Figure 4.6. Trace a small cone of rays from any point in the plane in the object space through the Luneberg lens. s and t are then calculated from the ray tracing information.	84
Figure 4.7. (a) Dyson system. (b) Shows Dyson system is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4 th power of the field height.	86
Figure 4.8. (a) The Offner relay. (b) Shows Offner relay is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4 th power of the field height.	87
Figure 4.9. (a) The modified Bouwers system. (b) Shows Bouwers system is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4 th power of the field height for a moderate field of view.	88
Figure 5.1. (a) A spherical mirror. The object is at the plane that contains the center of curvature of the spherical mirror. (b) Shows the sagittal image of the system is flat. The tangential image is on the circle of radius of curvature $R/4$ and the medial image is on the circle of radius of curvature $R/2$ where R is radius of curvature of the mirror.	94
Figure 5.2. Illustrates the ray tracing method to get t and s for the system shown in Figure 5.1 (a). $t(\theta) = s(\theta) = R/2$, where R is the radius of curvature of the spherical mirror.	95
Figure 5.3. (a) The layout of a symmetric anastigmatic system. (b) The ray fans of the system.	97
Figure 5.4. The plots of $-1/(2s)$ and $-\cos^2\theta/(2t)$ as functions of $\sin\theta$ for the symmetric systems shown in Figure 5.3(a). The continuous curves are the fitted 4 th power curves.	100
Figure 5.5. Plot of the tangential and sagittal ray aberrations vs. $\sin(\theta)$. The discrete values are the ZEMAX ray tracing result. The solid lines are the theoretical predictions.	100
Figure 5.6. An aplanatic system with flattened tangential field.	101

LIST OF FIGURES -- *continued*

	Page
Figure 5.7. (a) The field curve, and (b) the spot diagram of the system shown in Figure 5.6. The maximum field is 10 mm.	102
Figure 5.8. Plots of $W_3(\rho)$ and $p_1^2 W_2(\rho) + W_3(\rho)$ vs. ρ for the system shown in Figure 5.6. The dots are the data calculated from the values of s and t , and the solid lines are the polynomial fits (5.18).	103
Figure 5.9. (a) The aplanatic conjugate pair of a sphere. The lens material has an index of refraction 2.0, the radius of curvature of the spherical surface is 100 mm. (b) The field curve of the system. The maximum field is 1 mm.	105
Figure 5.10. The ray fans of the system shown in Figure 5.9(a).	105
Figure 5.11. Plot of $W_3(\rho)$ vs. $\sin\theta$ where θ is the ray angle in image space. The dots are the real ray trace data listed in Table 5.5 and the solid line is the polynomial fit (5.19a).	107
Figure 6.1. An optical system with object a finite distance away.	110
Figure 6.2. An optical system with object at infinity.	111
Figure 6.3. (a) A normal incidence 3-mirror aplanatic telescope. The primary mirror is spherical, and the secondary and tertiary are general aspheres. (b) The spot diagrams of the aplanatic telescope. Notice the residual aberration is primarily quadratic field-dependent.	112
Figure 6.4. (a) A grazing incidence 3-mirror aplanatic telescope. The primary mirror is spherical, and the secondary and tertiary are general aspheres. (b) The spot diagrams of the aplanatic telescope. Notice the residual aberration is primarily quadratic field-dependent.	113
Figure 6.5. Illustration of the tangential rays of an oblique bundle of rays refracted by a surface.	115
Figure 6.6. Illustrating how to obtain the tangential and sagittal radii of curvature of a general aspheric surface with rotational symmetry. C_1 is the sagittal center of curvature of the segment of the surface at P and C_2 is the tangential center of curvature of the same segment.	117

LIST OF FIGURES – *continued*

	Page
Figure 6.7. Illustration of the Snell's law.	118
Figure 6.8. Illustration of the design procedure of the 3-surface system that is corrected for all spherical aberration, coma and some specific quadratic field-dependent aberrations.	120
Figure 6.9. Illustration of the design procedure of the 4-surface system that is corrected for all spherical aberration, coma and quadratic field-dependent aberrations.	122
Figure 6.10. (a) A 3-surface system with zero quadratic field-dependent aberrations in tangential plane. All three surfaces are general aspheres. (b) The spot diagrams of the system.	124
Figure 6.11. (a) A 3-surface system with zero quadratic field-dependent astigmatism. All three surfaces are general aspheres. (b) The spot diagrams of the system.	125
Figure 6.12. (a) A 3-surface system with zero sagittal quadratic field-dependent aberrations. All three surfaces are general aspheres. (b) The spot diagrams of the system.	126
Figure 6.13. (a) A 4-surface system with no quadratic field-dependent aberrations. All four surfaces are general aspheres. (b) The spot diagrams of the system.	127
Figure 6.14. (a) A 4-surface system with no quadratic field-dependent aberrations. The object is at the infinity. All four surfaces are general aspheres. (b) The spot diagrams of the system.	128
Figure 6.15. (a) A 4-mirror system with no quadratic field-dependent aberrations. The object is at infinity. All four surfaces are general aspheres. (b) The spot diagrams of the system.	129
Figure 6.16. The ray fans of the 4-mirror system shown in Figure 6.15.	130
Figure 7.1. Unobstructed telescope: an example of plane-symmetric system.	133

LIST OF FIGURES -- *continued*

	Page
Figure 7.2. Illustration of the definition of coordinate systems for a plane-symmetric system. The ray that originates from the center of the field and goes through the center of pupil is called Optical Axis Ray (OAR). The object plane is x_0 - y_0 , and the image plane is x_1 - y_1	134
Figure 7.3. Trace a thin parallel bundle of rays centered at a ray from the field center with ray vector $\vec{\rho}_0$ in object space. (a) 3-D illustration of the bundle of rays, and (b) looking down the z_0 axis.	138
Figure 7.4. 3-D illustration of the sagittal ray of the parallel bundle in image space. The ray intersects the image plane at $(h_{x_1}, 0, 0)$. The angle between the sagittal ray and the center ray of the bundle is $\Delta\theta$	140
Figure 7.5. Blow out of the triangle T_sIH_1 in Figure 7.4.	140
Figure 7.6. 3-D illustration of the tangential ray of the parallel bundle in image space. The ray intersects the image plane at $(0, h_{y_1}, 0)$. The angle between the tangential ray and the center ray of the bundle is $\Delta\theta$	141

LIST OF TABLES

	Page
Table 1.1. Grouping aberrations according to their field-dependence.	27
Table 2.1. Surface data for the anastigmatic system shown in Figure 2.7.	54
Table 2.2. The result of forward and backward ray tracing for the Luneburg lens.	59
Table 4.1. Forward ray tracing result of the Luneburg Lens. It is shown that $t/\cos^2\theta = s$	84
Table 5.1. Surface data of the optical system shown in Figure 5.3 (a).	98
Table 5.2. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.3 (a).	99
Table 5.3. Surface data of the optical system shown in Figure 5.6.	101
Table 5.4. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.6.	102
Table 5.5. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.9.	106
Table 6.1. The prescription data of the 3-surface systems.	123
Table 6.2. The prescription data of the 4-surface systems.	123

ABSTRACT

In designing imaging optical systems, the primary task is to correct aberrations. Aberrations are deviations from perfect imagery. They depend on both the field size and pupil position. When the Constant Optical Path Length (OPL) condition is satisfied, an optical system is free of all orders of spherical aberrations, which have zero field dependence. When the Abbe Sine condition is satisfied, all the aberrations with linear field dependence are corrected. The Abbe Sine condition does not involve any off-axis ray properties, but it predicts the correction of off-axis aberrations.

We go one step beyond the Constant OPL condition and the Abbe Sine condition. By using Hamilton's characteristic functions, we developed a set of criteria for correcting the aberrations with quadratic field dependence and all orders of pupil dependence. These criteria involve only properties of the rays originating from the on-axis object point as the Abbe Sine condition does. Using these criteria, we analyzed some known designs and obtained new information about these designs. We also developed an algorithm to implement the criteria in designing well-corrected novel optical systems. Even when the criteria are not exactly satisfied, we now have a way to predict the residual quadratic field-dependent aberrations without tracing rays from any off-axis object point. We extended the Hamiltonian treatment to bilateral systems and developed similar criteria for correcting the quadratic field-dependent aberrations for this type of systems.

INTRODUCTION

The imaging optical systems, except for some very simple ones such as flat mirrors, have aberrations. The aberrations of an optical system are usually classified in terms of the combined order of their dependence on field and pupil. So there are 3rd order, 5th order, 7th order aberrations and so on. For most of the optical systems, 3rd order aberrations are the primary aberrations that need to be corrected, and higher order aberrations are negligible. So this classification is reasonable. But in some special cases, such as the microscope objective, resolution is more important than the field size. These systems have a small field of view and a large numerical aperture. Then the aberrations with higher order field dependence and lower order pupil dependence are not so critical to the system performance as those with lower order field dependence and higher order pupil dependence. Therefore just correcting the general lower order aberrations can not improve the system performance significantly. This suggests the necessity of a different method to group aberrations.

For optical systems that has a small field of view and a large numerical aperture, it is a better way to group aberrations in terms of their field dependence only. Then we have:

1. Spherical aberrations with no field dependence.
2. Linear field-dependent aberrations.
3. Quadratic field-dependent aberrations.

Criteria already exist for the correction of field-independent and linear field-dependent aberrations. They are the Constant OPL condition and the Abbe Sine condition. This dissertation carries out research on the correction of the quadratic field-dependent aberrations. By using the Hamilton's characteristic functions, we developed the criteria for correcting the aberrations with quadratic field dependence and all orders of pupil dependence. Since the criteria involve the astigmatism of the pupil aberration, we name them the *Pupil Astigmatism Conditions*.

Similar to the Constant OPL condition and the Abbe Sine condition, the Pupil Astigmatism conditions involve only the properties of the rays originating from the on-axis object point, and they predict the correction of aberrations of all orders of pupil dependence. Such conditions are completely new to people. As shown in this dissertation, these conditions give new information of the optical systems that people are already familiar with. They can also be used to design new systems with the quadratic field-dependent aberrations fully or partially corrected. They are believed to be very useful in designing high-resolution optical systems.

In chapter 1, a brief introduction to aberrations and Hamilton's characteristic functions is given. The Pupil Astigmatism conditions take two different forms. Both of them are presented and derived in Chapter 2. By using the relations between the image and pupil aberrations, the criteria are validated to 3rd order in Chapter 3. In Chapter 4 some known designs are analyzed with the conditions and new information is extracted. Chapter 5 gives a method to predict the quadratic field-dependent aberrations when the criteria are not exactly satisfied, and some examples are given to verify the method. In

Chapter 6, an algorithm is developed to apply the criteria in the numerical design of optical systems, and some systems designed using the algorithm are shown. Similar criteria for the plane symmetric system are derived in Chapter 7. In Appendix A, the computer programs that apply the algorithm to design new optical systems are listed. Since the algorithm generates an optical surface point by point, we have to use the *user defined surface* feature in ZEMAX to simulate the performance of the generated system. The computer files for implementing the surfaces made up of discrete points in ZEMAX are listed in Appendix B.

CHAPTER 1

ABERRATIONS AND HAMILTON'S CHARACTERISTIC FUNCTIONS

1.1 INTRODUCTION

The work of this dissertation is on the correction of the quadratic field-dependent aberrations of all orders. Criteria are developed to correct certain or all types of aberrations with quadratic field dependence. The main tool used in developing these criteria is the Hamilton's characteristic functions. In Section 1.2, I give an introduction to the aberrations of optical systems which includes the definition and classification of aberrations, the definition of pupil aberrations and the principles to correct certain types of aberrations. Then the Hamilton's characteristic functions are introduced in Section 1.3, which includes the types of the characteristic functions and the use of them in characterizing the general optical system.

1.2 INTRODUCTION TO ABERRATIONS

If an optical system can not image an object perfectly, then aberrations are present. Perfect imagery is achieved in the case that any rays from an object point goes through the same image point and the image as whole is a scaled replica of the object. From the point of view of wave optics, aberrations are present if the spherical wave front from an

object point can not be transformed into a perfect spherical wave front centered at the ideal image point by the optical system, as a result all the rays originating from the object point do not go through the ideal image point. Therefore aberrations can be described in two ways. One is in terms of wave front deviation from the perfect spherical reference; the other is in terms of the lateral ray deviation from a perfect reference point. In the former case, the perfect spherical wave front is usually chosen to be the sphere at the exit pupil of the optical system with the center at the ideal image location. In the latter case, the reference point is chosen to be the ideal image point. See Figure 1.1 for the illustration of wave and ray aberrations.

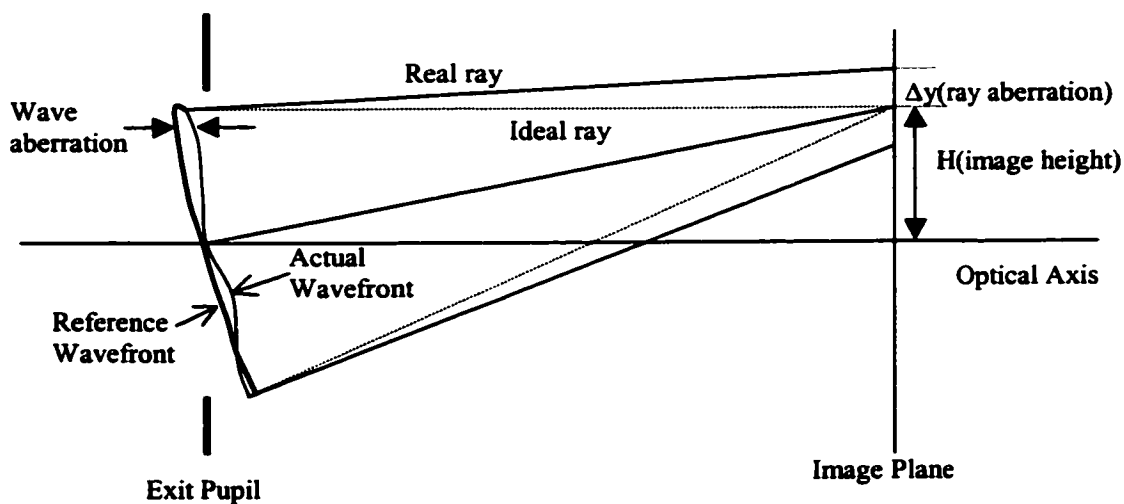


Figure 1.1. Illustration of wave aberration and ray aberration.

The index of refraction of an optical media is a function of the wavelength of light. The aberrations caused by this dependence on wave length are called chromatic

aberrations. The work presented in this dissertation does not deal with chromatic aberrations, so we restrict our discussions to monochromatic aberrations in this chapter and the entire dissertation.

We also restrict our discussion about aberrations to axially symmetric systems until Chapter 6. We will discuss the aberration corrections for bilateral systems in Chapter 7.

1.2A RAY ABERRATION AND WAVE ABERRATION

Figure 1.1 illustrates the definition of the wave and ray aberration. The ray aberration is the lateral deviation of the ray from its ideal position in the image plane. The wave aberration is defined as the difference between the aberrated wave front and the reference wave front, namely,

$$W = W_{ab} - W_{re}. \quad (1.1)$$

The aberration of an optical system depends on both the field and aperture. Assume the maximum field height is H_m , and the exit pupil radius is r_p . Considering a ray originating from a field point \vec{H} and going through point \vec{r} at the exit pupil, the wave aberration of this ray is a function of both \vec{H} and \vec{r} :

$$W = W(\vec{H}, \vec{r}). \quad (1.2)$$

Often times it is convenient to use the normalized field and pupil coordinates \vec{h} and $\vec{\rho}$ instead, where

$$\vec{h} = \frac{\vec{H}}{H_m} \quad \text{and} \quad \vec{\rho} = \frac{\vec{r}}{r_p}. \quad (1.3)$$

Then

$$W = W(\bar{h}, \bar{\rho}). \quad (1.4)$$

Since the wave aberration is a scalar and it does not depend on the rotation of the axially symmetric systems about its optical axis, it can only be a function of three quantities, namely, $(\bar{h} \cdot \bar{h})$, $(\bar{\rho} \cdot \bar{\rho})$ and $(\bar{h} \cdot \bar{\rho})$. Then the aberration can be written in the following form:

$$W(\bar{h}, \bar{\rho}) = \sum_{m,n,k} W_{2m+k, 2n+k, k} (h^2)^m \cdot (\rho^2)^n \cdot (\bar{h} \cdot \bar{\rho})^k. \quad (1.5)$$

Depending on the value of $2(m+n+k)$, the aberration terms $W_{2m+k, 2n+k, k}$ can be classified into 3rd order aberration, 5th order aberration, etc. For example, when $2(m+n+k)=4$, the aberration is 3rd order. The 3rd order aberration includes:

W_{040} (spherical aberration),
 W_{131} (coma),
 W_{222} (astigmatism),
 W_{220} (field curvature),
 W_{311} (distortion).

When $2(m+n+k)=6$, the aberration is 5th order (Shack, Shannon 1997). There are totally 10 types of 5th order aberrations:

W_{060} (spherical aberration),
 W_{151} (5th order coma),
 W_{240} (Oblique spherical aberration),
 W_{242} (tangential oblique spherical aberrations),
 W_{333} (elliptical coma),
 W_{331} (elliptical coma),
 W_{420} (5th order field curvature),
 W_{422} (5th order astigmatism),
 W_{511} (5th order distortion),
 W_{600} (piston).

1.2B RELATIONS BETWEEN WAVE AND RAY ABERRATIONS

Wave and ray aberrations are closely correlated. If one is known, the other can be calculated (Shack). Assume the ray aberration is $\bar{\epsilon}$, and the distance from the image plane to the exit pupil is R, then ray and wave aberrations are related via the following equation:

$$\bar{\epsilon} = -\frac{R}{r_p} \nabla_{\bar{\rho}} W(\bar{h}, \bar{\rho}). \quad (1.6)$$

1.2C THE PUPIL ABERRATIONS

Every optical system has an aperture that determines the diameter of the cone of light that the system will accept from an axial point on the object. This aperture is the stop of the system. The image of the stop in object space is the *entrance pupil* and the image of the stop in image space is the *exit pupil* (see Figure 1.2). The ray that starts from the on-axis object point and goes through the edge of entrance pupil is named as marginal ray, and the ray that starts from the edge of the object and goes through the center of entrance pupil is named as chief ray.

The aberration definitions in previous sections give relationships that model departure from ideal mapping from object points to image points. This modeling is used to give aberrations corresponding to image blur or wavefront error. It is also useful to investigate non-ideal mapping of the pupils, or pupil aberrations. Similar to the image aberrations, 3rd order pupil aberrations include spherical aberration, coma, astigmatism,

field curvature and distortion. The pupil aberrations can be used for understanding the source of image aberrations, and as we show later in this dissertation, they can be used for deriving some general imaging relationships.

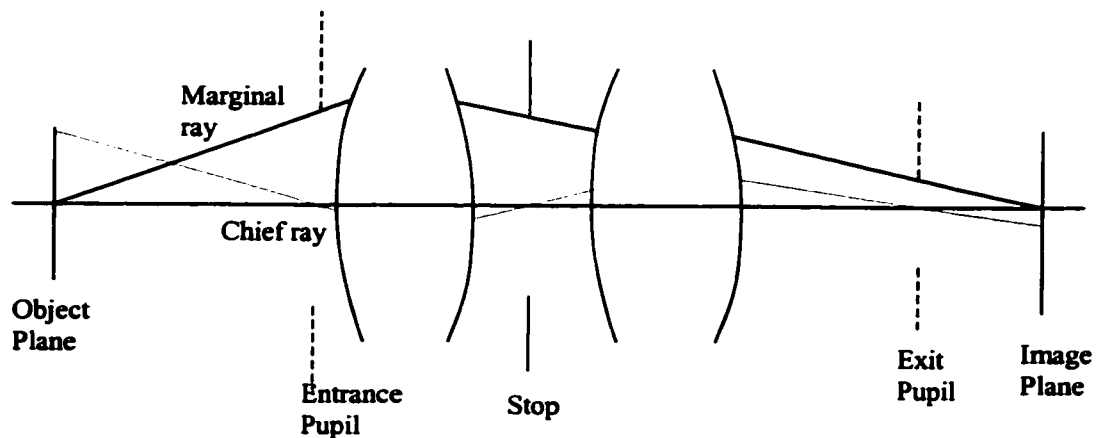


Figure 1.2. Illustration of the pupils. The Entrance Pupil is the image of stop in object space, and the Exit Pupil is the image of stop in image space.

1.2D ANOTHER WAY TO GROUP WAVE ABERRATIONS

A typical optical design procedure starts from first order layout of a system, then steps are taken to correct or balance 3rd order aberrations. In many cases, correction or balance of 3rd order aberrations is sufficient. But in some special cases, such as grazing incidence optics, 3rd order treatment fails. In other cases, such as the high NA microscope objectives where resolution is more concerned than the size of field of view, 3rd order treatment is not enough. These cases suggest the necessity of a different way to group aberrations. In the microscope case, aberrations of higher order pupil dependence

are more vital to the performance than aberrations of higher order field dependence. So we may classify aberrations in terms of their field dependence, then take a different approach to correct them. Then we have:

1. field-independent aberrations including W_{040} , W_{060} , W_{080} ,
 2. linear field-dependent aberrations including W_{131} , W_{151} , W_{171} ,
 3. quadratic field-dependent aberrations including: W_{220} , W_{240} , W_{260} ,
- and W_{222} , W_{242} , W_{262} ,

Table 1.1. Grouping aberrations according to their field-dependence.

Aberrations			
Field Dependence	Aberration Type	Conditions to correct all	Example Systems
Independent	Spherical aberrations of all orders	the Optical Path Length(OPL) from an on-axis object point to its image point along any ray is constant	Cartesian oval, reflecting ellipsoid
Linear	Coma of all orders	The Abbe Sine Condition	Aplanatic Telescopes
Quadratic	Astigmatism, field curvature, etc.	<i>The Pupil Astigmatism Condition</i>	<i>Aplanatic anastigmats</i>

Usually it does not make any sense to know the exact values of W_{080} , W_{171} , W_{260} , W_{262} , etc, but they still need to be corrected. If an on-axis object point is imaged perfectly by an axial symmetric system, all field-independent aberrations are absent; this is what I call the Constant Optical Path Length (OPL) Condition for correcting all orders of spherical aberration. Similarly, the Abbe Sine Condition provides a tool for eliminating all the aberrations with linear field dependence. Here we derive a condition I

call Pupil Astigmatism Condition that is used to correct all the quadratic field-dependent aberrations. Table 1.1 summarizes the new grouping methods. My dissertation work is shown in the bold and italic fonts. The tool I used to develop the condition is Hamilton's characteristic functions, which will be introduced in Chapter 1.3.

1.3 INTRODUCTION TO HAMILTON'S CHARACTERISTIC FUNCTIONS

Characteristic functions were first introduced by Sir W. R. Hamilton (1828), later expanded by Bruns as eikonal (1895). Hamilton's characteristic functions are a set of functions that represent the optical path length along a ray. Even though it is impossible to get the analytical form of the Hamilton's characteristic functions for optical systems except very simple ones, they are nevertheless very powerful tools to study the general properties of an optical system. The Hamilton's characteristic functions have been very useful in optical design. Luneburg (1970) used them to derive the Abbe Sine condition and design the Luneburg lens. Buchdahl (1970) developed the formulas for calculating aberration coefficients. They were used to show generally what imaging properties can or cannot be realized (Walther 1989). Their application in the design of non-axially symmetric systems was explored extensively by Stone and Forbes (1992-1994). They were used to specify lens modules for modular lens design (Chang 1999). As personal computer becomes more and more powerful, application of Hamilton's characteristics in real lens design becomes more practical.

1.3A DEFINITION OF CHARACTERISTIC FUNCTIONS

For my dissertation work, I used Hamilton's characteristic functions to derive the criteria for correcting all the quadratic field-dependence aberrations. In this section, the definition and use of the Hamilton's characteristic functions are described.

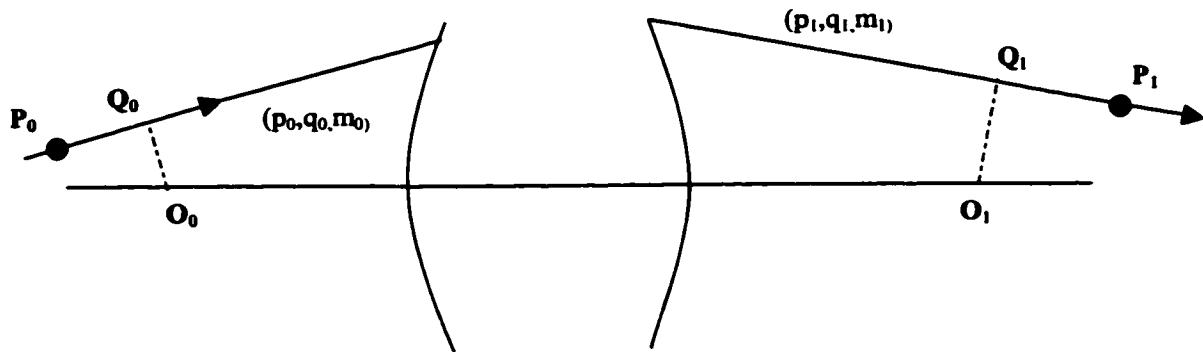


Figure 1.3. Interpretation of Hamilton's characteristic functions when the initial and final media are homogeneous.

In Figure 1.3, a ray originates from a point $P_0(x_0, y_0, z_0)$ in object space, and passes through $P_1(x_1, y_1, z_1)$ in image space. O_0 and O_1 are origins of local coordinate systems. (p_0, q_0, m_0) and (p_1, q_1, m_1) are ray vectors in object space and image space respectively. A ray vector is the vector along the ray with length equal to the index of refraction of the local medium. By definition,

$$p_0^2 + q_0^2 + m_0^2 = n_0^2, \quad (1.7a)$$

and

$$p_1^2 + q_1^2 + m_1^2 = n_1^2, \quad (1.7b)$$

where n_0 and n_1 are the refractive indices in object space and image space respectively. Q_0 and Q_1 are the intersections of perpendiculars drawn from O_0 and O_1 to the ray in object space and image space respectively.

If we make the following definition:

$$[AB] = \text{optical path length along the ray from Point } A \text{ to Point } B, \quad (1.8)$$

then the Hamilton's characteristic functions are defined as follows:

Point characteristic: $V(x_0, y_0, z_0; x_1, y_1, z_1) = [P_0 P_1],$

Mixed characteristic: $W(x_0, y_0, z_0; p_1, q_1) = [P_0 Q_1],$

Angle characteristic: $T(p_0, q_0; p_1, q_1) = [Q_0 Q_1].$

Hamilton's characteristic functions are very powerful tools for investigation of the general properties of optical systems. If one of the Hamilton characteristics is known, we can obtain all the information about any ray. For example,

If V is known, then

$$p_0 = -\frac{\partial V}{\partial x_0}, \quad q_0 = -\frac{\partial V}{\partial y_0}, \quad (1.9a)$$

and

$$p_1 = \frac{\partial V}{\partial x_1}, \quad q_1 = \frac{\partial V}{\partial y_1}. \quad (1.9b)$$

If W is known, then

$$p_0 = -\frac{\partial W}{\partial x_0}, \quad p_1 = -\frac{\partial W}{\partial y_0}, \quad (1.10a)$$

and

$$X_1 = -\frac{\partial W}{\partial p_1}, \quad Y_1 = -\frac{\partial W}{\partial q_1}. \quad (1.10b)$$

Here (X_1, Y_1) is the coordinate of the ray intersection with the plane that passes point O_1 and is perpendicular to the optical axis.

If T is known, then

$$X_0 = \frac{\partial T}{\partial p_0}, \quad Y_0 = \frac{\partial T}{\partial q_0}, \quad (1.11a)$$

and

$$X_1 = -\frac{\partial T}{\partial p_1}, \quad Y_1 = -\frac{\partial T}{\partial q_1}. \quad (1.11b)$$

Here (X_0, Y_0) is the coordinate of the ray intersection with the plane that passes point O and is perpendicular to the optical axis, and (X_1, Y_1) is the coordinate of the ray intersection with the plane that passes point O_1 and is perpendicular to the optical axis.

Since the Hamilton's mixed and angle characteristic functions can be used to calculate the ray intercept at a plane, we can then use them to calculate ray aberrations.

1.3B TAYLOR SERIES EXPANSION OF THE CHARACTERISTIC FUNCTIONS

Similar to the wave aberration described in Chapter 1.2A, the characteristic functions are scalars and they are rotationally invariant for an axially symmetric system. For an axially symmetric system, the mixed characteristic function W only depends on the following 3 quantities:

$$\begin{aligned}
h^2 &= x_0^2 + y_0^2, \\
\rho^2 &= p_1^2 + q_1^2, \\
\vec{h} \cdot \vec{\rho} &= x_0 p_1 + y_0 q_1.
\end{aligned} \tag{1.12}$$

Expand W in the power series of the field, and neglect the 3rd and higher powers:

$$\begin{aligned}
W(x_0, y_0; p_1, q_1; z_0, z_1) &= W_0(\rho) + (x_0 p_1 + y_0 q_1) W_1(\rho) \\
&+ (x_0 p_1 + y_0 q_1)^2 W_2(\rho) + (x_0^2 + y_0^2) W_3(\rho) + \dots \tag{1.13}
\end{aligned}$$

For a field point in x-z plane, $y_0 = 0$. Then

$$\begin{aligned}
W(x_0, y_0; p_1, q_1; z_0, z_1) &= W_0(\rho) + x_0 p_1 W_1(\rho) \\
&+ x_0^2 (p_1^2 W_2(\rho) + W_3(\rho)) + \dots \tag{1.14}
\end{aligned}$$

In analogy, for an axially symmetric system, the angle characteristic function T only depends on the following 3 quantities:

$$\begin{aligned}
h^2 &= p_0^2 + q_0^2, \\
\rho^2 &= p_1^2 + q_1^2, \\
\vec{h} \cdot \vec{\rho} &= p_0 p_1 + q_0 q_1.
\end{aligned} \tag{1.15}$$

Expand T in the power series of the field, and neglect the 3rd and higher powers, we obtain

$$\begin{aligned}
T(p_0, q_0; p_1, q_1; z_0, z_1) &= W_0(\rho) + (p_0 p_1 + q_0 q_1) W_1(\rho) \\
&+ (p_0 p_1 + q_0 q_1)^2 W_2(\rho) + (p_0^2 + q_0^2) W_3(\rho) + \dots \tag{1.16}
\end{aligned}$$

For a field point in x-z plane, $q_0 = 0$. Then

$$\begin{aligned}
T(p_0, q_0; p_1, q_1; z_0, z_1) &= W_0(\rho) + p_0 p_1 W_1(\rho) \\
&+ p_0^2 (p_1^2 W_2(\rho) + W_3(\rho)) + \dots \tag{1.17}
\end{aligned}$$

In Chapter 2, Eq. (1.14) is used to derive the Pupil Astigmatism conditions for correcting the quadratic field-dependent aberrations for optical systems with the object a finite distance away, and Eq. (1.17) is used to derive the similar conditions for optical systems with the object at infinity.

1.4 CONCLUSION

In this chapter, I go over the basics about the aberrations of an optical system, which include the definition of aberration, the relation between the ray aberration and wave aberration, the definition of pupil aberration, the types of monochromatic aberrations and the different ways in grouping aberrations. The conditions for correcting certain types of aberrations are listed in Table 1.1. Hamilton's characteristic functions are introduced in Section 1.3. Their definition, physical meaning and use are described.

CHAPTER 2

THE PUPIL ASTIGMATISM CONDITIONS FOR CORRECTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS

2.1 INTRODUCTION

In this chapter, we use Hamilton's characteristic functions to derive conditions that predict and allow complete correction of all aberrations that have quadratic field dependence. For this analysis, we consider an arbitrary optical imaging system that has a small field of view. We then use Hamilton's characteristic functions to define some simple relationships between the pupil's astigmatism and the image aberrations that have quadratic field dependence.

For this analysis, we first define an exit pupil at infinity and look at the aberrations in entrance pupil, which occurs at an image of the exit pupil. It is important to understand that the pupils defined here are mathematical, and do not correspond to the pupil in the real system which is defined as an image of a real aperture. The analysis proves that a particular relationship must be held between the sagittal and tangential pupil images in order to correct the quadratic field-dependent image aberrations. If these pupil astigmatism relations are maintained for all points in the pupil, then the quadratic field-dependent aberrations will be zero for all points in the field.

The general derivation of the pupil astigmatism relations requires analysis of the exit pupil – entrance pupil transformation, which requires the optical system to be modeled by tracing rays backwards, starting in image space and propagating into object space. We know that light always travels the same forwards as backwards, so mathematically the two are equivalent. Nonetheless, it feels wrong to trace the system backwards, even if it does provide a simple solution. We are justified in violating this directional bias because the exit pupil – entrance pupil relation allows a derivation of the pupil astigmatism relations that are simple and general. By making a second assumption, that the first order field relations are also corrected, we can show a less general, but more useful form of the pupil astigmatism relations that uses the entrance pupil to exit pupil transformation. This lets us use our convenient computer and mental models where light travels from left to right.

Just like the Abbe Sine condition, the pupil astigmatism criteria take different forms for systems with object a finite distance away and systems with object at infinity, so in the presentation and derivation of the criteria, we deal with the two types of systems separately. The tools used in the derivation of the criteria are the Hamilton's characteristic functions. Specifically, the mixed characteristic function is used to derive the criteria for the finite conjugate system, and the angle characteristic function is used for the infinite conjugate systems.

2.2 DEPENDENCE OF QUADRATIC FIELD-DEPENDENT ABERRATIONS ON ENTRANCE PUPIL ASTIGMATISM

2.2.1A SUMMARY FOR SYSTEMS WITH FINITE CONJUGATES

Consider the case of an optical system imaging a point which is a finite distance away. This system forms an image which is nominally a point, but may have some aberrations (which we assume to be small compared to the dimensions of the imaging system). To evaluate an exit pupil at infinity, we treat small bundles of parallel rays that go through the object point. Their angle defines the pupil coordinate. The entrance pupil is defined as an image of this exit pupil, so it can be found by tracing these bundles of rays backward through the system.

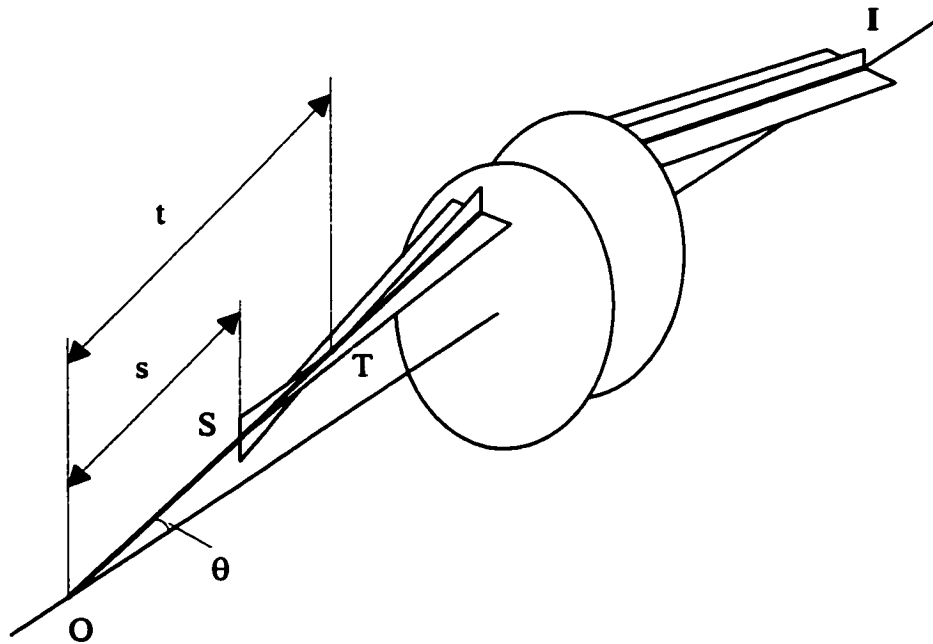


Figure 2.1. Trace a thin bundle of parallel rays backwards through an optical system (finite conjugate case). T is the tangential focus and S is the sagittal focus.

In general, each bundle of rays will come to focus with some amount of astigmatism. The bundles are made small enough that only aberrations up to second order are considered. This geometry is shown in Figure 2.1 where I is the image point. It is useful to map two different versions of this entrance pupil -- one is defined by the tangential focus, and the other by the sagittal focus. This is similar to the field curves that are commonly used for image evaluation, except here we are investigating the image of an infinite exit pupil so these curves are not physically significant.

As shown in Fig. 2.1, the pupil coordinate is defined using the angle from the object to the particular point in the entrance pupil. Denoting the tangential focus as T and the sagittal focus as S , define the distance from object to T as t , and object to S as s . Clearly, t and s are functions of θ . Using these definitions and relationships from Hamilton's characteristic functions, we will prove that the imaging system will be free from all quadratic field dependent aberrations if

$$s(\theta) = \frac{t(\theta)}{\cos^2(\theta)} = \text{constant.} \quad (2.1)$$

The derivation of this condition is given in Section 2.2.1B. From the derivation we also obtain the following criteria for partially correcting the quadratic field-dependent aberrations:

- The criterion for correcting all the quadratic field-dependent aberrations in tangential plane is

$$\frac{t(\theta)}{\cos^2(\theta)} = \text{constant.} \quad (2.2)$$

- The criterion for correcting all the quadratic field-dependent aberrations of the form of W_{2n0} where n is even is

$$s(\theta) = \text{constant.} \quad (2.3)$$

- The criterion for correcting all the quadratic field-dependent astigmatism of the form W_{2n2} where n is even is

$$s(\theta) = \frac{t(\theta)}{\cos^2(\theta)}. \quad (2.4)$$

As these criteria are written in terms of the astigmatic image of the exit pupil, we call them the *pupil astigmatism criteria* or *pupil astigmatism conditions*.

2.2.1B DERIVATION OF PUPIL ASTIGMATISM CONDITIONS FOR SYSTEMS WITH FINITE CONJUGATES

To prove the pupil astigmatism conditions, given above, we start with the Taylor series expansion of the mixed characteristic function W , as given in Chapter 1:

$$\begin{aligned} W(x_0, y_0; p_0, q_0; z_0, z_1) = & W_0(\rho) + (x_0 p_1 + y_0 q_1) W_1(\rho) + (x_0 p_1 + y_0 q_1)^2 W_2(\rho) \\ & + (x_0^2 + y_0^2) W_3(\rho) + \dots \end{aligned} \quad (2.5)$$

Consider an object point in x-z plane, then $y_0 = 0$. According to Eq. (1.10b),

$$\begin{aligned}
x_1 &= -\frac{\partial W}{\partial p_1} \\
&= -\frac{\partial W_0(\rho)}{\partial p_1} - x_0 \frac{\partial}{\partial p_1}(p_1 W_1(\rho)) - x_0^2 \frac{\partial}{\partial p_1}(p_1^2 W_2(\rho) + W_3(\rho)), \quad (2.6a)
\end{aligned}$$

$$\begin{aligned}
y_1 &= -\frac{\partial W}{\partial q_1} \\
&= -\frac{\partial W_0(\rho)}{\partial q_1} - x_0 \frac{\partial}{\partial q_1}(p_1 W_1(\rho)) - x_0^2 \frac{\partial}{\partial q_1}(p_1^2 W_2(\rho) + W_3(\rho)). \quad (2.6b)
\end{aligned}$$

According to Eq. (1.10a),

$$p_0 = -p_1 W_1(\rho), \quad (2.7)$$

for a small field in the vicinity of Point O. And we know $\frac{p_0}{p_1}$ is the magnification, in

general it is a function of ρ . Define the magnification as $M(\rho)$:

$$M(\rho) = \frac{p_0}{p_1}. \quad (2.8)$$

The ideal image of a point at $(x_0, 0, z_0)$ will be at $(x_1', 0, z_1)$ where

$$x_1' = M(0)x_0. \quad (2.9)$$

Let Δx and Δy be the lateral aberrations and substitute Eq. (2.9) into Eqs. (2.6a) and (2.6b), we then obtain

$$\begin{aligned}
\Delta x &= x_1 - x_1' \\
&= -\frac{\partial W_0(\rho)}{\partial p_1} + x_0 \frac{\partial}{\partial p_1}(p_1(M(\rho) - M(0))) - x_0^2 \frac{\partial}{\partial p_1}(p_1^2 W_2(\rho) + W_3(\rho)) \quad (2.10a)
\end{aligned}$$

$$\Delta y = y_1$$

$$= -\frac{\partial W_0(\rho)}{\partial q_1} + x_0 \frac{\partial}{\partial q_1}(p_1 M(\rho)) - x_0^2 \frac{\partial}{\partial q_1}(p_1^2 W_2(\rho) + W_3(\rho)) \quad (2.10b)$$

The first term of the right hand side of equations in (2.10) gives rise to spherical aberration. The second term is coma. For linear field-dependent aberration to be corrected, $M(\rho)$ must be constant,

$$M(\rho) = M(0) = \text{constant}, \quad (2.11)$$

which is the famous Abbe Sine condition. To correct the quadratic field-dependent aberration, the following equation must be true:

$$p_1^2 W_2(\rho) + W_3(\rho) = \text{constant}. \quad (2.12)$$

Apparently, $W_2(\rho)$ determines astigmatism and $W_3(\rho)$ determines field curvature and oblique spherical aberrations.

We would like to know what $p_1^2 W_2(\rho) + W_3(\rho)$ represents physically. After differentiating Eq. (2.5) with respect to x_0 twice and using the relations in Eq. (1.10b), we obtain

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{1}{2} \frac{\partial p_0}{\partial x_0} \quad (2.13)$$

for a field in the close vicinity of point O .

The partial differentiation in Eq. (2.13) means everything else doesn't change, i.e. keep y_0 , p_1 , q_1 unchanged, see how p_0 changes with infinitesimal change of x_0 .

Tracing the thin bundle of parallel rays backward through an optical system maintains the values of y_0, p_1 and q_1 .

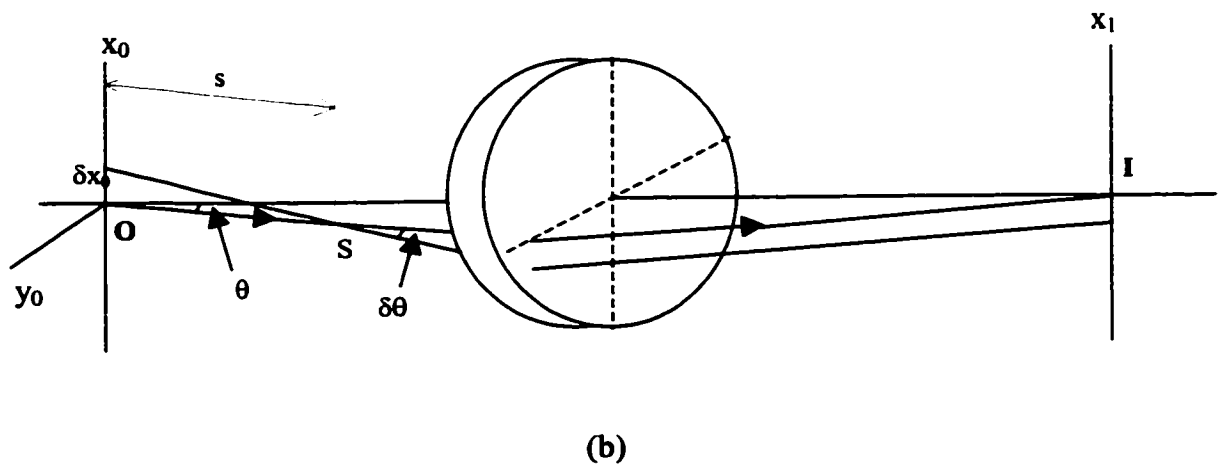
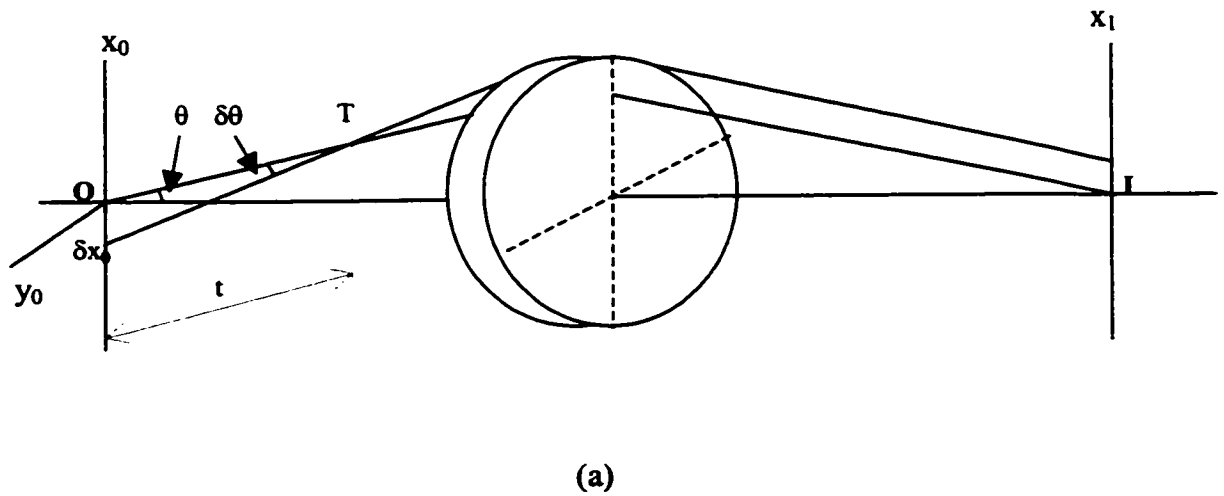


Figure 2.2. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ for a finite conjugate system in (a) the tangential plane, and (b) the sagittal plane.

In Figure 2.2(a), in tangential plane, at $x_0 = 0$, $p_0 = n_0 \sin(\theta)$, let $OT = t$ (< 0 in this case), then

$$\delta p_0 = n_0 \cos(\theta) \delta \theta, \quad (2.14)$$

and

$$\delta \theta = \frac{\delta x_0 \cos(\theta)}{t}. \quad (2.15)$$

So, in the tangential plane,

$$\frac{\delta p_0}{\delta x_0} = \frac{n_0 \cos^2(\theta)}{t}, \quad (2.16)$$

therefore

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{1}{2} \frac{n_0 \cos^2(\theta)}{t}. \quad (2.17)$$

In Figure 2.2(b), in sagittal plane, at $x_0 = 0$, $p_0 = 0$, $p_l = 0$, trace a thin bundle of parallel rays backwards again, S is the focus, denote $OS = s$ (< 0 in this case), then

$$\delta p_0 = n_0 \delta \theta \quad (2.18)$$

and

$$\delta x_0 = s \delta \theta. \quad (2.19)$$

So, in the sagittal plane,

$$\frac{\delta p_0}{\delta x_0} = \frac{n_0}{s}, \quad (2.20)$$

therefore

$$p_1^2 W_2(\rho) + W_3(\rho) = W_3(\rho) = -\frac{1}{2} \frac{n_0}{s}. \quad (2.21)$$

So, combining Equations (2.17), (2.21) and (2.10), we draw the following conclusions:

- When $s(\theta) = t(\theta)/\cos^2(\theta) = \text{constant}$, then $W_2(\rho) = 0$ and $W_3(\rho) = \text{constant}$, and the optical system is free of all the quadratic field-dependent aberrations.
- When $t(\theta)/\cos^2(\theta) = \text{constant}$, then $p_1^2 W_2(\rho) + W_3(\rho)$ is constant in the tangential plane, so all the quadratic field-dependent aberrations in the tangential plane is corrected.
- When $s(\theta) = \text{constant}$, then $W_3(\rho) = \text{constant}$, so all the quadratic field-dependent aberrations in the sagittal plane are corrected.
- When $s = t/\cos^2(\theta)$, then $W_2(\rho) = 0$, so astigmatism of the form W_{2n2} (n is even) is corrected.

2.2.2A INFINITE CONJUGATE SYSTEMS

If the object is at infinity, the definitions of t and s must be modified. For the optical system in Figure 2.3, still trace a thin bundle of parallel rays backwards through the system, the tangential focus is T and the sagittal focus is S . Denote the distance from T to a plane perpendicular to the optical axis as t and the distance from S to the same plane as s , t and s are functions of the ray height h in the object space.

With t and s defined and obtained this way, the criteria for correcting the quadratic field-dependent aberrations are:

To correct all the quadratic field-dependent aberrations:

$$s(h) = t(h) = \text{constant.} \quad (2.22a)$$

To correct all the quadratic field-dependent aberrations in tangential plane:

$$t(h) = \text{constant.} \quad (2.22b)$$

To correct all the quadratic field-dependent aberrations of the form of W_{2n0} where n is even:

$$s(h) = \text{constant.} \quad (2.22c)$$

To correct all the quadratic field-dependent aberrations of the form W_{2n2} where n is even:

$$s(h) = t(h). \quad (2.22d)$$

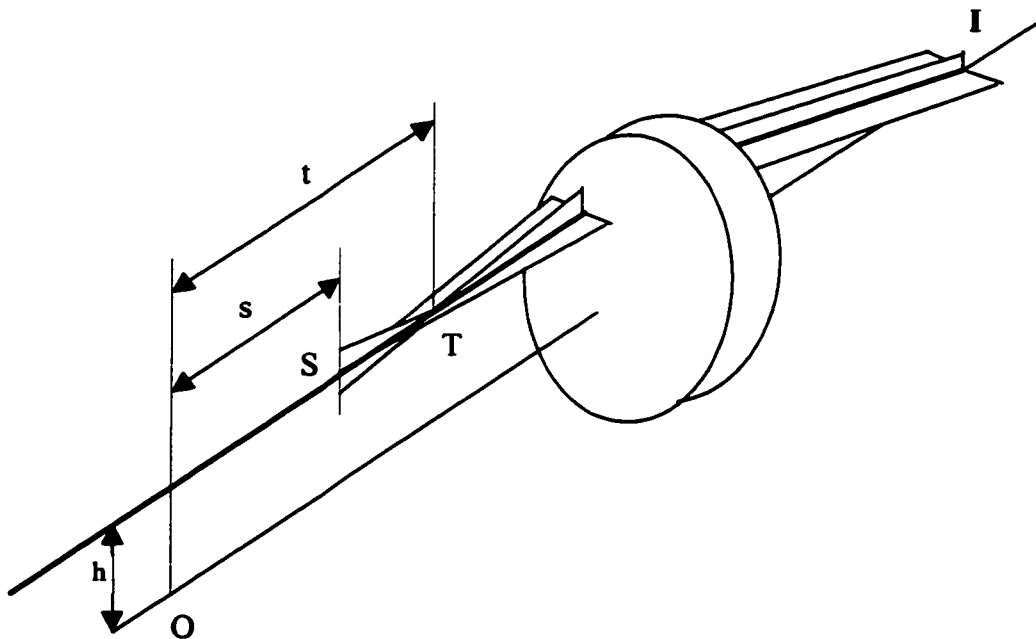


Figure 2.3. Trace a thin bundle of parallel rays backward through an aplanatic optical system (infinite conjugate case). T is the tangential focus and S is the sagittal focus.

Criterion (2.22a) indicates that if a system with object at infinity is to be corrected for all the quadratic field-dependent aberrations, then the 3rd order entrance pupil astigmatism and field curvature are perfectly corrected.

2.2.2B DERIVATION OF THE PUPIL ASTIGMATISM CONDITIONS FOR THE INFINITE CONJUGATE SYSTEM

To characterize the optical system with the object at infinity, the angle characteristic function is more convenient to use. The Taylor series expansion of the angle characteristic function T given in Chapter 1 is

$$T(p_0, q_0; p_1, q_1; z_0, z_1) = W_0(\rho) + (p_0 p_1 + q_0 q_1) W_1(\rho) + (p_0 p_1 + q_0 q_1)^2 W_2(\rho) + (p_0^2 + q_0^2) W_3(\rho) + \dots \quad (2.23)$$

Compared to the mixed characteristic function, the field size is angle (p_0, q_0) now instead of the height (x_0, y_0) . The ray intersection at the image plane is calculated in a similar way.

Consider a ray in the x - z plane, then $p_0 = 0$. According to Eq. (1.11b),

$$\begin{aligned} x_1 &= -\frac{\partial T}{\partial p_1} \\ &= -\frac{\partial W_0(\rho)}{\partial p_1} - p_0 \frac{\partial}{\partial p_1} (p_1 W_1(\rho)) - p_0^2 \frac{\partial}{\partial p_1} (p_1^2 W_2(\rho) + W_3(\rho)), \end{aligned} \quad (2.24a)$$

$$\begin{aligned} y_1 &= -\frac{\partial T}{\partial q_1} \\ &= -\frac{\partial W_0(\rho)}{\partial q_1} - p_0 \frac{\partial}{\partial q_1} (p_1 W_1(\rho)) - p_0^2 \frac{\partial}{\partial q_1} (p_1^2 W_2(\rho) + W_3(\rho)). \end{aligned} \quad (2.24b)$$

According to Eq. (1.11a), the ray height h in object space is

$$h = \frac{\partial T}{\partial p_0} = p_1 W_1(\rho) \quad (2.25)$$

for a small field angle. And we know $-\frac{h}{p_1}$ is the effective focal length, in general it is a

function of ρ . Define the effective focal length as $f(\rho)$, then

$$f(\rho) = -W_1(\rho). \quad (2.26)$$

The ideal image of a field point $(p_0, 0, m_0)$ will be at $(x_1', 0, 0)$ where

$$x_1' = f(0)p_0. \quad (2.27)$$

Let Δx and Δy be the lateral aberrations and substitute Eq. (2.26) into Eqs. (2.24a) and (2.24b), we then obtain

$$\begin{aligned} \Delta x &= x_1 - x_1' \\ &= -\frac{\partial W_0(\rho)}{\partial p_1} + p_0 \frac{\partial}{\partial p_1} (p_1 (f(\rho) - f(0))) - p_0^2 \frac{\partial}{\partial p_1} (p_1^2 W_2(\rho) + W_3(\rho)), \end{aligned} \quad (2.28a)$$

$$\begin{aligned} \Delta y &= y_1 \\ &= -\frac{\partial W_0(\rho)}{\partial q_1} + p_0 \frac{\partial}{\partial q_1} (p_1 f(\rho)) - p_0^2 \frac{\partial}{\partial q_1} (p_1^2 W_2(\rho) + W_3(\rho)). \end{aligned} \quad (2.28b)$$

For the linear field-dependent aberration to be corrected, $f(\rho)$ must be constant,

$$f(\rho) = f(0) = \text{constant}, \quad (2.29)$$

which is the Abbe Sine condition for systems with object at infinity. To correct the quadratic field dependent aberration, again the following equation must be true:

$$p_1^2 W_2(\rho) + W_3(\rho) = \text{constant}. \quad (2.30)$$

We then proceed to find out what $p_1^2 W_2(\rho) + W_3(\rho)$ represents physically. Similar to Eq. (2.13), we now have

$$p_1^2 W_2(\rho) + W_3(\rho) = \frac{1}{2} \frac{\partial x_0}{\partial p_0} \quad (2.31)$$

for a small field angle.

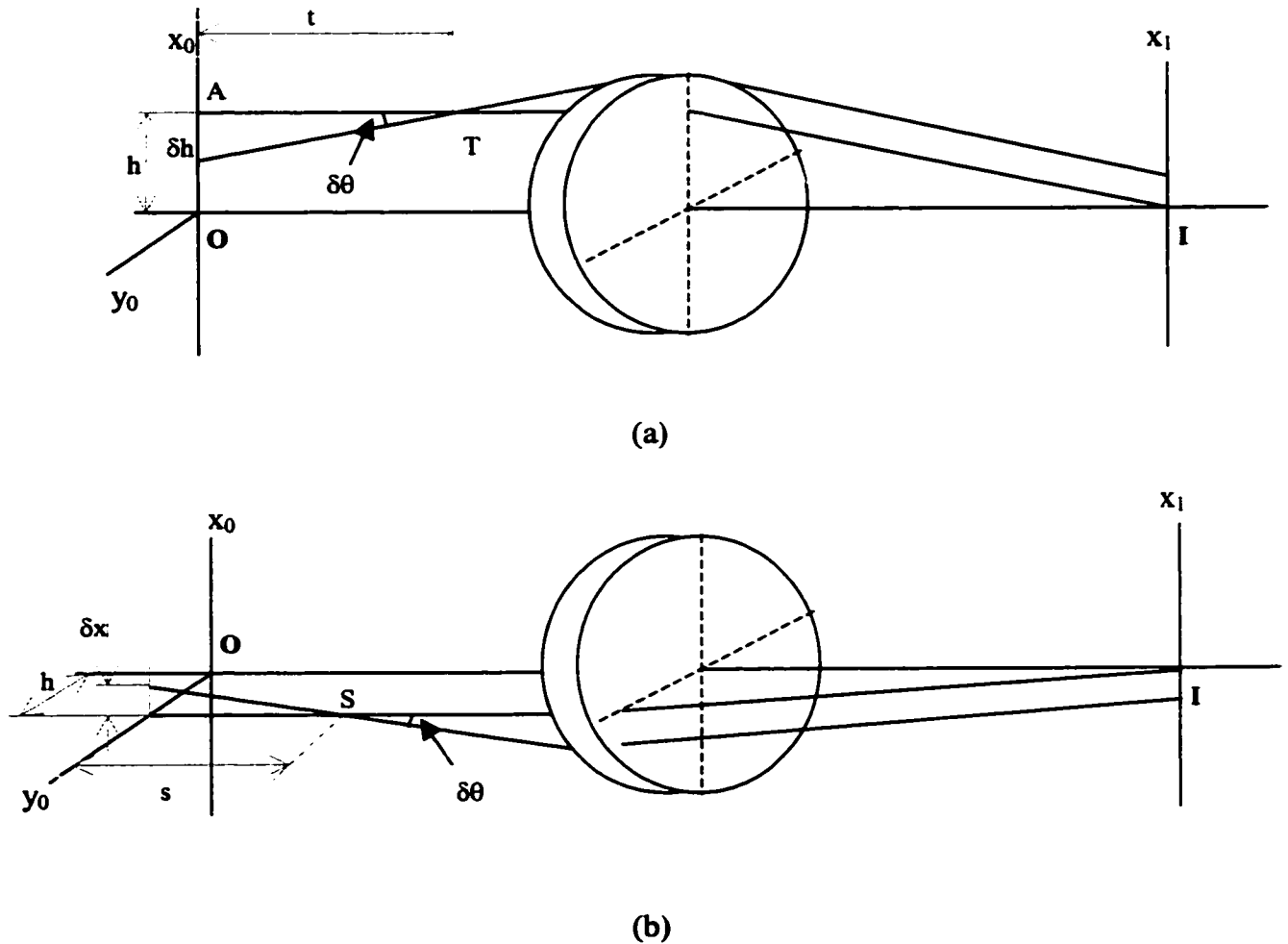


Figure 2.4. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ for an infinite conjugate system in (a) the tangential plane, and (b) the sagittal plane.

We again trace a thin bundle of parallel rays backward through the optical system. In Figure 4(a), in tangential plane, at $p_0 = 0, x_0 = h$, let $OT = t$ (< 0 in this case), then

$$\delta p_0 = \frac{n_0 \delta h}{t}. \quad (2.32)$$

So, in the tangential plane,

$$\frac{\delta h}{\delta p_0} = \frac{t}{n_0} \quad (2.33)$$

and

$$p_1^2 W_2(\rho) + W_3(\rho) = \frac{1}{2} \frac{t}{n_0}. \quad (2.34)$$

In Figure 4(b), in sagittal plane, at $p_0 = 0, x_0 = h, p_l = 0$, trace a thin bundle of parallel rays backwards again, S is the focus, denote $OS = s$ (< 0 in this case), then

$$\delta p_0 = n_0 \delta \theta \quad (2.35)$$

and

$$\delta x = s \delta \theta. \quad (2.36)$$

So, in sagittal plane,

$$\frac{\delta x}{\delta p_0} = \frac{s}{n_0} \quad (2.37)$$

and

$$p_1^2 W_2(\rho) + W_3(\rho) = W_3(\rho) = \frac{1}{2} \frac{s}{n_0}. \quad (2.38)$$

So, combining Equations (2.38), (2.34) and (2.28), we draw the following conclusions similar to those in Section 2.2.1B:

- When $s(h) = t(h) = \text{constant}$, then $W_2(\rho) = 0$ and $W_3(\rho) = \text{constant}$, so all the quadratic field-dependent aberrations are corrected.
- When $t(h) = \text{constant}$, then $p_1^2 W_2(\rho) + W_3(\rho)$ is constant in the tangential plane, so all the quadratic field-dependent aberrations in the tangential plane are corrected.
- When $s(h) = \text{constant}$, then $W_3(\rho) = \text{constant}$, so all the quadratic field-dependent aberrations in the sagittal plane are corrected.
- When $s(h) = t(h)$, then $W_2(\rho) = 0$, astigmatism of the form W_{2n2} (n is even) is corrected.

It is important to notice that we used the ideal image position as the reference and the ray vector as the pupil coordinate when deriving these conditions. Usually in an optical design program, the chief ray position is used as the reference instead. Therefore, in our approach, the change of the stop location does not affect the types of aberrations present in the system. That is why we can choose the exit pupil at infinity.

2.3 FORWARD RAY-TRACING VERSION OF THE PUPIL ASTIGMATISM CRITERIA

The backward ray-tracing version of the Pupil Astigmatism Criteria for correcting the quadratic field-dependent aberrations is presented in Section 2.2.

Although it is quite straightforward to trace rays backward to derive the criteria, it is not

so convenient to use them this way. First of all, it is difficult to perform backward ray tracing in a computer ray trace program. Second, we need ways to validate the criteria, but there is no known way to connect these backward ray tracing properties with the image aberrations. Facing these difficulties, we decided to develop the forward ray tracing version of the criteria. To do this, we define an entrance pupil at infinity and look at the aberrations in exit pupil.

2.3A THE FINITE CONJUGATE SYSTEM

Figure 2.5 shows an optical system with object a finite distance away. A thin bundle of parallel rays are traced forward through the system. The center ray of the bundle is a marginal ray from Point O to Point I . The Tangential focus of the bundle is T' and the sagittal focus is S' ; let $IT' = t'$ and $IS' = s'$. The marginal ray has an angle θ with the optical axis in the object space and an angle θ' in the image space.

Figure 2.6 shows the side view of the optical system shown in Figure 2.5. Both the forward and backward ray tracing are shown in this figure. t and t' can then be calculated by the following formulas:

$$\begin{aligned} t &= dx_o \cos(\theta) / d\theta, \\ t' &= dx_i \cos(\theta') / d\theta'. \end{aligned} \tag{2.39}$$

Then,

$$t' = \frac{dx_i \cos(\theta')}{dx_o \cos(\theta)} \frac{d\theta}{d\theta'} t = m \frac{\cos(\theta')}{\cos(\theta)} \frac{d\theta}{d\theta'} t, \tag{2.40}$$

where m is the magnification.

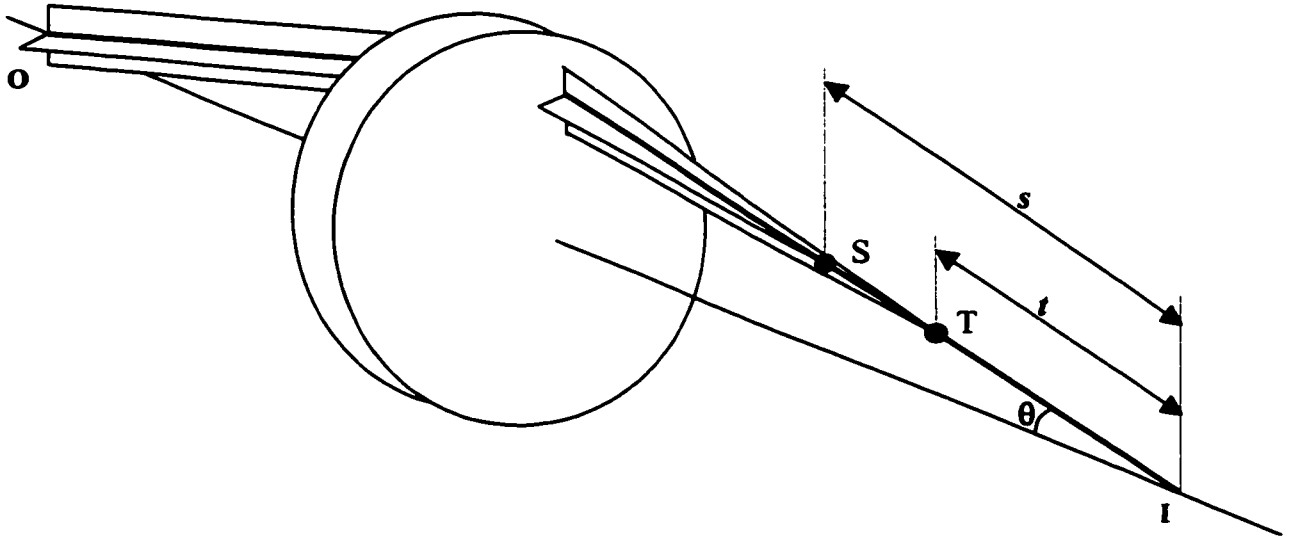


Figure 2.5. Trace a thin bundle of parallel rays forwards through an optical system (finite conjugate case). T is the tangential focus and S is the sagittal focus.

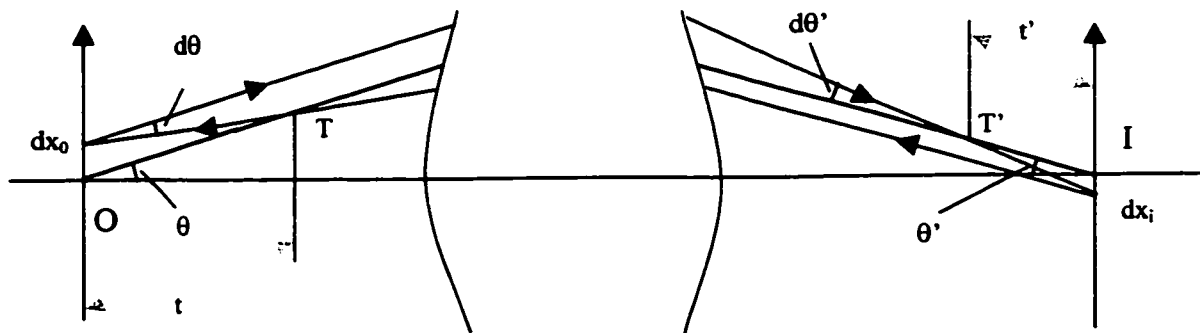


Figure 2.6. The side view of an optical system with the object a finite distance away. T is the tangential focus of a thin bundle of parallel rays traced backwards through the optical system. T' is its counterpart when a thin bundle of parallel rays are traced forward through the optical system.

Now consider a patch of object (dx_o high and dy_o wide) around point O whose image is dx_i high and dy_i wide around point I , consider a beam around the ray with direction cosines $(\sin(\theta), 0, \cos(\theta))$ which spreads $d\theta$ in x-z plane and $d\beta$ in y-z plane in the object space. In the image space, the beam is along the ray with direction cosines $(-\sin(\theta'), 0, \cos(\theta'))$ and spreads $d\theta'$ in x-z plane and $d\beta'$ in y-z plane. Since the energy must be conserved, then the basic throughput of the system is constant, i.e.

$$n^2 d\vec{A} \bullet d\vec{\Omega} = \text{const} , \quad (2.41)$$

where n is the index of refraction, $d\vec{A}$ is the area element, $d\vec{\Omega}$ is solid angle element.

Then we have

$$n_o^2 dx_o dy_o \cos(\theta) d\theta d\beta = n_i^2 dx_i dy_i \cos(\theta') d\theta' d\beta' . \quad (2.42)$$

It follows that

$$\frac{d\theta}{d\theta'} = \frac{dx_i dy_i}{dx_o dy_o} \frac{n_i d\beta'}{n_o d\beta} \frac{n_i \cos(\theta')}{n_o \cos(\theta)} . \quad (2.43)$$

Since

$$\frac{dx_i dy_i}{dx_o dy_o} = m^2 \quad (2.44)$$

and

$$\frac{n_i d\beta'}{n_o d\beta} = \frac{1}{m} , \quad (2.45)$$

then combining Eqs. (2.43), (2.44) and (2.45), we obtain,

$$\frac{d\theta}{d\theta'} = m \frac{n_i \cos(\theta')}{n_o \cos(\theta)} . \quad (2.46)$$

Substitute Eq. (2.46) into Eq. (2.40), we obtain

$$t' = m^2 \frac{n_i \cos^2(\theta')}{n_o \cos^2(\theta)} t, \quad (2.47)$$

which is equivalent to

$$\frac{t'}{n_i \cos^2(\theta')} = m^2 \frac{t}{n_o \cos^2(\theta)}. \quad (2.48)$$

By the same token, we can also prove

$$\frac{s'}{n_i} = m^2 \frac{s}{n_o}. \quad (2.49)$$

It should be noticed that when spherical aberrations are perfectly corrected and the Abbe Sine condition is satisfied, Eqs. (2.40)-(2.49) are rigorously true.

An example is given below to show the relationships in (2.48) and (2.49) are correct. Figure 2.7 shows an anastigmatic optical system whose 3rd order spherical aberration, coma, astigmatism and field curvature are all corrected. Its surface data are listed in Table 2.1. This optical system has a magnification $m = -2.25$. For forward ray tracing, $t' / \cos^2(\theta') = s'(\theta') = 149.334$ when $\theta' = 0$. And for backward ray tracing, $t / \cos^2(\theta) = s(\theta) = 29.498$ when $\theta = 0$. The ratio between the two constants is 5.0625, and the square of the magnification is $m^2 = 5.0625$, so the relations in Eqs. (2.48) and (2.49) are confirmed here.

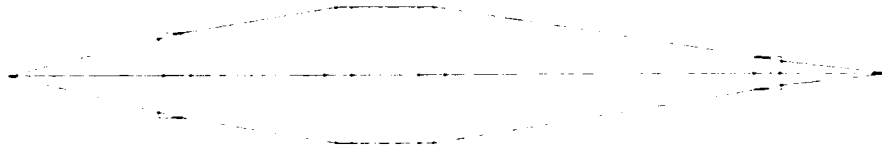


Figure 2.7. An anastigmatic optical system with magnification -2.25 .

Table 2.1. Surface data for the anastigmatic system shown in Figure 2.7.

Surf	Type	Radius	Thickness	Glass	Diameter	Conic
OBJ	STANDARD	Infinity	113.3333		0	0
STO	STANDARD	-113.3333	20	1.5	58.0	0
2	STANDARD	-80	100		66.7	0
3	STANDARD	150	20	1.5	107.1	-2.25
4	STANDARD	Infinity	50.68886		107.1	0
5	STANDARD	Infinity	20	1.5	107.1	0
6	STANDARD	-150	225.1072		107.1	-2.25
7	STANDARD	74.89276	20	1.5	25.6	0
8	STANDARD	32.93566	82.33915		18.5	0
IMA	STANDARD	Infinity			0	0

Then Eqs. (2.48) and (2.49) show that there is only a constant factor difference between the forward and backward ray tracing properties involved in the Pupil Astigmatism Criteria in (2.1)-(2.4), so the form of the criteria based on forward ray tracing properties stays the same as that of the backward ray tracing version. We write them below:

- The criterion for correcting all the quadratic field-dependent aberrations then is:

$$s'(\theta') = \frac{t'(\theta')}{\cos^2(\theta')} = \text{constant.} \quad (2.50a)$$

- The criterion for correcting all the quadratic field-dependent aberrations in the tangential plane is:

$$\frac{t'(\theta')}{\cos^2(\theta')} = \text{constant.} \quad (2.50b)$$

- The criterion for correcting all the quadratic field-dependent aberrations of the form of W_{2n0} where n is even is:

$$s'(\theta') = \text{constant.} \quad (2.50c)$$

- The criterion for correcting all the quadratic field-dependent aberrations of the form W_{2n2} where n is even is:

$$s'(\theta') = \frac{t'(\theta')}{\cos^2(\theta')}. \quad (2.50d)$$

2.3B THE INFINITE CONJUGATE SYSTEM

The forward ray tracing version of the Pupil Astigmatism Criteria for the infinite conjugate system is identical to that of the finite conjugate system (Equations (2.50)), but s and t are obtained differently.

For an infinite conjugate system shown in Figure 2.8, select an arbitrary plane which is perpendicular to the optical axis in object space and serves as the entrance pupil of the system. Then trace a small cone of rays which originates from any point in the plane and are centered on the ray which is parallel to the optical axis, again we have the tangential image at T' , and the sagittal image at S' . Define $t' = IT'$ and $s' = IS'$. With

these definitions, the conditions for correcting the quadratic field-dependent aberrations for infinite conjugate systems are the same as Eqs. (2.50).

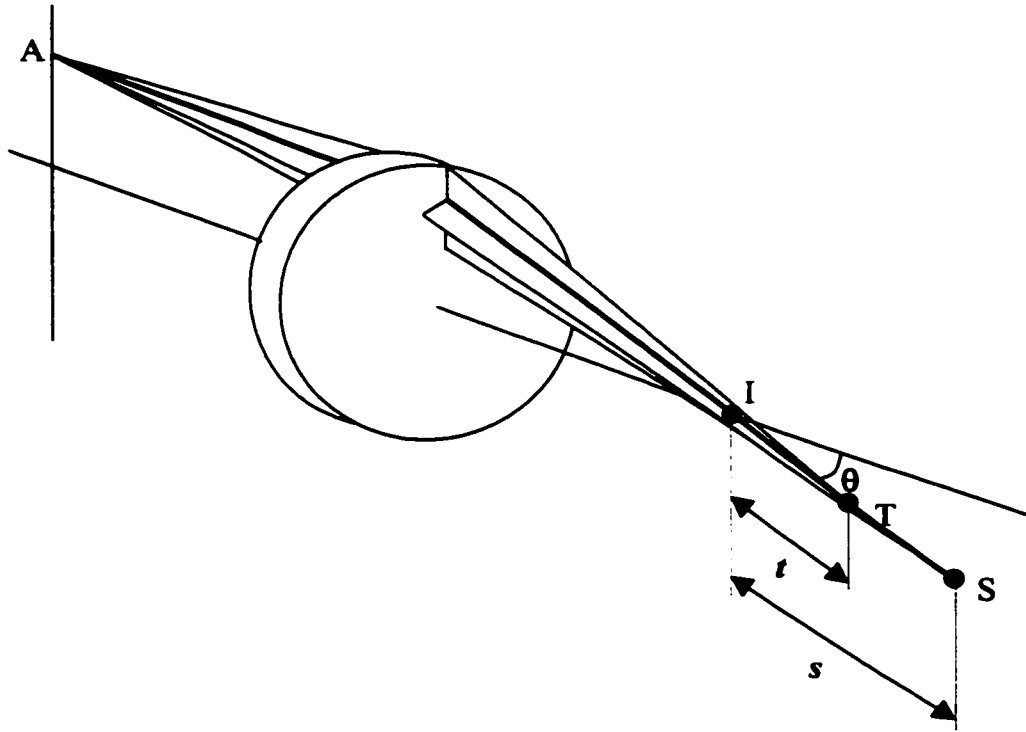


Figure 2.8. Illustration of a system with object at infinity. A small cone of rays are traced from a point A in a plane which is perpendicular to the optical axis. The tangential image of the cone is T and the sagittal image of the cone is S .

In Figure 2.9, we have the following relation:

$$t = dh / d\theta , \quad (2.51a)$$

$$t' = dx_i \cos(\theta') / d\theta' . \quad (2.51b)$$

It follows that

$$t' = t \frac{dx_i}{dh} \frac{d\theta}{d\theta'} \cos\theta' . \quad (2.52)$$

Again using Eq. (2.41), we have

$$n_o^2 dh dy_o \cos(\theta) d\theta d\beta = n_i^2 dx_i dy_i \cos(\theta') d\theta' d\beta'. \quad (2.53)$$

where $d\beta$ and $d\beta'$ are the angles that the cone of the rays spreads in the plane perpendicular to the paper in the object and image space, and dy_o and dy_i are the y-axis dimensions of the areas considered in the x_o - y_o and x_i - y_i planes respectively.

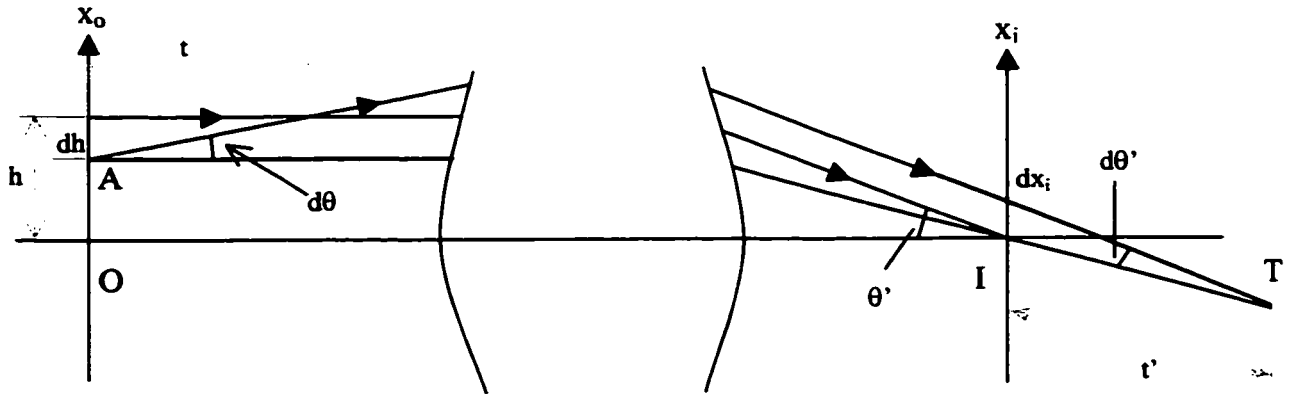


Figure 2.9. The side view of an optical system with object at infinity. T is the tangential image of the cone of rays originating from a point A in the object space.

Then the following equation is obtained:

$$\frac{d\theta}{d\theta'} = \frac{dx_i}{dh} \frac{n_i dy_i d\beta'}{n_o dy_o d\beta} \cos(\theta'). \quad (2.54)$$

The magnification of the system gives

$$dy_i = f n_o d\beta \quad \text{and} \quad dx_i = f n_o d\theta. \quad (2.55)$$

And the Abbe Sine Condition gives

$$dy_o = f n_i d\beta' \quad \text{and} \quad dh = f n_i \cos\theta' d\theta'. \quad (2.56)$$

It then follows that

$$\frac{n_i dy_i d\beta'}{n_o dy_o d\beta} = 1 \quad (2.57)$$

and

$$\frac{dx_i}{dh} = \frac{n_o}{n_i \cos \theta'} \frac{d\theta}{d\theta'} \quad (2.58)$$

Combining Eqs. (2.51b) and (2.56), we obtain

$$\frac{dx_i}{dh} = \frac{t'}{fn_i \cos^2 \theta'} \quad (2.59)$$

Substitute Eqs. (2.58) and (2.59) into Eq. (2.52), we get

$$t = f^2 \frac{n_o n_i \cos^2(\theta')}{t'} \quad (2.60)$$

which is equivalent to

$$\frac{t}{n_o} = f^2 \frac{n_i \cos^2(\theta')}{t'} \quad (2.61)$$

By the same token, we can also prove

$$\frac{s}{n_o} = f^2 \frac{n_i}{s'} \quad (2.62)$$

One example to show Eqs. (2.61) and (2.62) are correct is the Luneburg lens. The

Luneburg lens is a sphere made of a gradient index material with the index profile

$$n(r) = \sqrt{2 - \left(\frac{r}{R}\right)^2},$$

where R is the radius of the sphere. The Luneburg lens is aplanatic. Figure 2.10 shows the Luneburg lens with radius $R = 100$ mm. The effective focal length of this lens is $f = 100$ mm, and the index of refraction in both the object and image space is unity. Table

2.2 shows the forward and backward ray-tracing results. Comparing Column 4 and Column 5, we can see Eqs. (2.61) and (2.62) are verified.

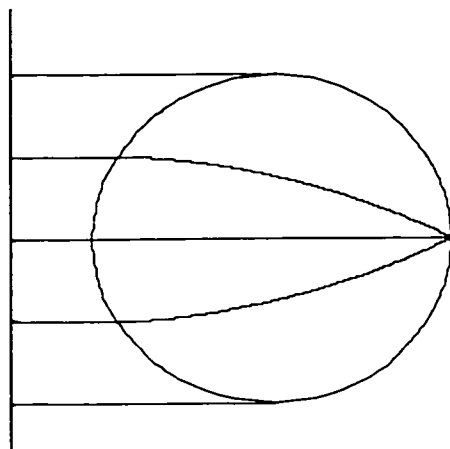


Figure 2.10 The Luneburg Lens.

Table 2.2. The result of forward and backward ray tracing for the Luneburg lens.

h (in mm): Ray height in the object space	θ (in $^\circ$): Ray angle in the image space	$s(\theta) = t(\theta)/\cos^2(\theta)$	$t(h) = s(h)$	$f^2/s(\theta) = f^2 \cos^2(\theta)/t(\theta)$
0	0	100	100	100
8.000756	4.589	99.681	100.3206	100.32
16.00018	9.207	98.728	101.2883	101.2884
24.00078	13.887	97.16	102.9229	102.923
32.00012	18.663	95.004	105.2583	105.2587
39.99972	23.578	92.295	108.3484	108.3482
47.99939	28.685	89.069	112.2728	112.2725
56.00029	34.056	85.36	117.1509	117.1509
64.00024	39.792	81.194	123.1627	123.1618
71.99942	46.054	76.568	130.602	130.6029
79.99989	53.13	71.429	139.9999	139.9992

With the backward ray-tracing version of the Pupil Astigmatism criteria in Section 2.2.2A and the relations in Eqs. (2.61) and (2.62), we conclude that the forward ray-tracing version of the criteria for the infinite conjugate system is exactly the same as that for the finite conjugate system listed in (2.50), although the definitions of s and t are different for the two types of systems.

2.4 ANOTHER METHOD TO DERIVE THE FORWARD RAY TRACING VERSION OF THE CRITERIA

The forward ray tracing version of the Pupil Astigmatism Criteria can also be derived directly without knowing the backward ray tracing version first. From Sections 2.2 and 2.3, we know the criteria are based on the value of $p_1^2 W_2(\rho) + W_3(\rho)$, which is the coefficient of the field square term in the Taylor's series expansion of the Hamilton's characteristic functions. We will derive the expression of the coefficient by forward ray tracing method directly in this section. This derivation also gives us insight in how to deal with plane symmetric optical systems.

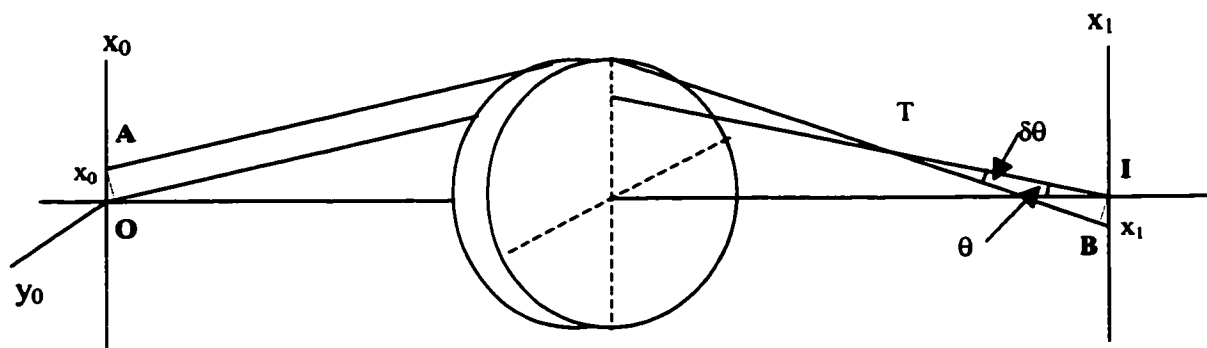
2.4A FINITE CONJUGATES SYSTEM

Consider a field point A in the optical system shown in Figure 2.11(a). A is on the x -axis and has an infinitesimal field height x_0 . Now trace a ray from Point O to Point I , which has an angle θ with the optical axis in the image space. Then trace a parallel ray from A in the tangential plane. The two rays intersect with each other at T in the image

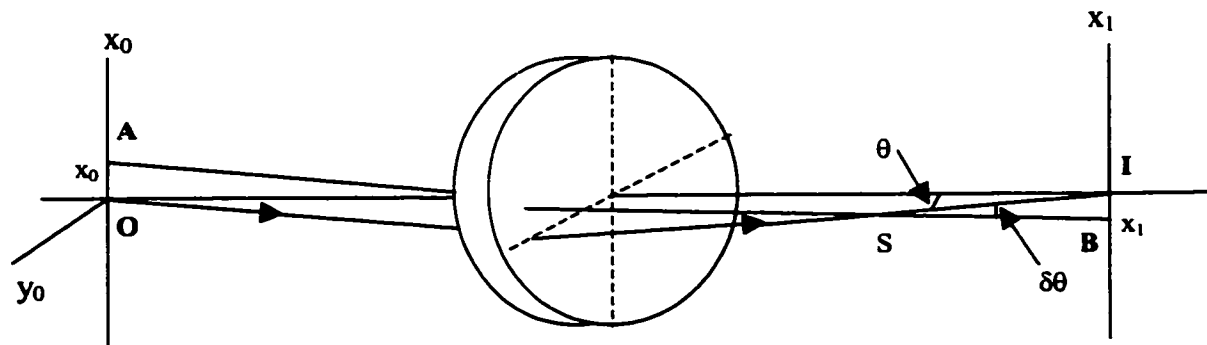
space. Draw a perpendicular from I to the ray originating from A , the foot is denoted as B . Let $TI = t (>0)$, then the optical path length from T to I is

$$[TI] = n_i t, \quad (2.63)$$

where n_i is the index of refraction of the media in image space.



(a)



(b)

Figure 2.11. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ in (a) the tangential plane, and (b) the sagittal plane by tracing rays forward through an optical system with object a finite distance away.

The mixed Hamilton's characteristic function for O is then

$$W(O) = [OT] + [TI], \quad (2.64)$$

and the mixed Hamilton's characteristic function for A is

$$W(A) = [AT] + [TB], \quad (2.65)$$

where

$$[AT] = [OT] - x_0 p_0, \quad (2.66)$$

and

$$\begin{aligned} [TB] &= [TI] \cos(\delta\theta) \\ &\cong [TI] - [TI](\delta\theta)^2 / 2. \end{aligned} \quad (2.67)$$

So combining Eqs. (2.64) - (2.67), we get

$$W(A) = W(O) - x_0 p_0 - [TI](\delta\theta)^2 / 2, \quad (2.68)$$

where

$$\begin{aligned} \delta\theta &= x_1 \cos(\theta) / t \\ &= x_0 M(\rho) \cos(\theta) / t. \end{aligned} \quad (2.69)$$

Then

$$W(A) = W(O) - x_0 p_0 - x_0^2 (n_i M^2(\rho) \cos^2(\theta) / (2t)). \quad (2.70)$$

Compare this equation to Eqs. (2.6), we get

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{n_i M^2(\rho) \cos^2(\theta)}{2t} \quad (2.71)$$

in the tangential plane.

To find out what $W_3(\rho)$ is, we trace a ray from Point O to Point I in sagittal plane, which has an angle θ with the optical axis in the image space (see Figure 2.11(b)). Then trace a parallel ray from A to the image space. The two rays intersect with each other at S in the image space. Draw a perpendicular from I to the ray originating from A , the foot is denoted as B . Let $SI = s$ ($s > 0$), then the optical path length from S to I is

$$[SI] = n_i s. \quad (2.72)$$

The mixed Hamilton's characteristic function for O is

$$W(O) = [OS] + [SI], \quad (2.78)$$

and the mixed Hamilton's characteristic function for A is

$$W(A) = [AS] + [SB], \quad (2.79)$$

where

$$[AS] = [OS], \quad (2.80)$$

and

$$\begin{aligned} [SB] &= [SI] \cos(\delta\theta) \\ &\cong [SI] - [SI] (\delta\theta)^2 / 2. \end{aligned} \quad (2.81)$$

So combining Eqs. (2.78), (2.79), (2.80) and (2.81), we get

$$W(A) = W(O) - [SI] (\delta\theta)^2 / 2, \quad (2.82)$$

where

$$\delta\theta = \frac{x_0 M(\rho)}{s}. \quad (2.83)$$

Then

$$W(A) = W(O) - x_0^2 \left(\frac{n_i M^2(\rho)}{2s} \right). \quad (2.84)$$

Compare this equation to Eqs. (2.6) (note $p_l = 0$ in the sagittal plane), we get

$$W_3(\rho) = -\frac{n_i M^2(\rho)}{2s} \quad (2.85)$$

in the sagittal plane.

Assume the Abbe Sine Condition is satisfied for an optical system, then $M(\rho) =$ constant. Also assume the image space is homogeneous, then combining equations (2.71) and (2.85), we draw the following conclusions:

- When $s = t/\cos^2(\theta) = \text{constant}$, then $W_2(\rho) = 0$, $W_3(\rho) = \text{constant}$, which means all the quadratic field-dependent aberrations of the system are corrected.
- When $t/\cos^2(\theta) = \text{constant}$, then $p_1^2 W_2(\rho) + W_3(\rho)$ is constant in tangential plane, which means all the quadratic field-dependent aberrations in the tangential plane are corrected.
- When $s = \text{constant}$, then $W_3(\rho)$ is constant, which means all the quadratic field-dependent aberrations of the form W_{2n0} (n is even), which include field curvature and oblique spherical aberrations, are corrected.
- When $s = t/\cos^2(\theta)$, then $W_2(\rho) = 0$, all the quadratic field-dependent aberrations of the form W_{2n2} (n is even) are corrected.

2.4B INFINITE CONJUGATES SYSTEM

For an optical system with the object at infinity, following the same procedure in Section 2.4A, we can obtain the mathematical expression for $p_1^2 W_2(\rho) + W_3(\rho)$.

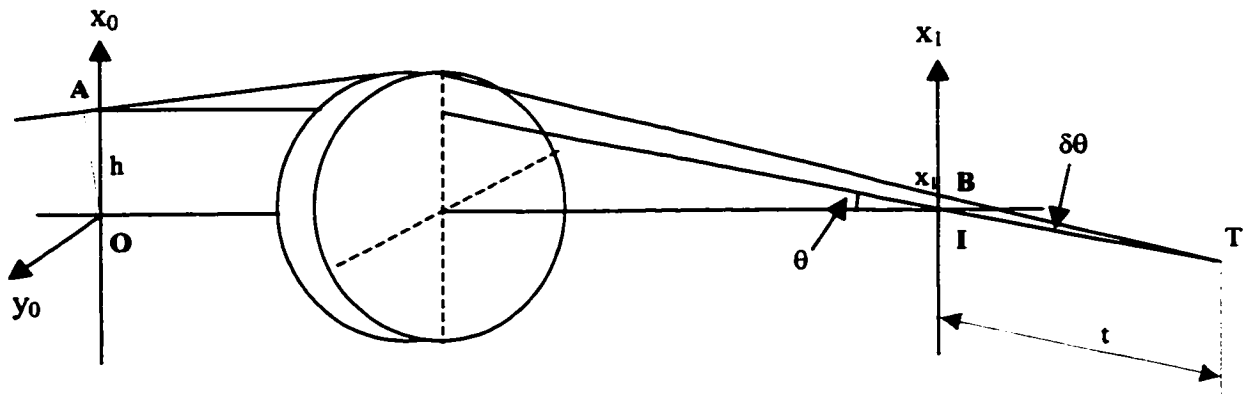


Figure 2.12. Illustration of how to get $p_1^2 W_2(\rho) + W_3(\rho)$ in the tangential plane by tracing rays forward through the optical system with object at infinity.

Figure 2.12 shows an optical system with the object at infinity, the angle characteristic of an off-axis field is approximately

$$T(A) = T(O) + hp_0 + [TI](\delta\theta)^2 / 2, \quad (2.86)$$

where

$$\begin{aligned} \delta\theta &= x_1 \cos(\theta) / t \\ &= p_0 f(\rho) \cos(\theta) / t. \end{aligned} \quad (2.87)$$

Then

$$T(A) = T(O) + x_0 p_0 - p_0^2 (f^2(\rho) \cos^2(\theta) / (2t)), \quad (2.88)$$

where $t < 0$.

Compare (2.88) to (2.23), we obtain

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{n_i f^2(\rho) \cos^2(\theta)}{2t} \quad (2.89)$$

in the tangential plane.

In the sagittal plane, it is easily to show that

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{n_i f^2(\rho)}{2s}. \quad (2.90)$$

When the Abbe Sine Condition is satisfied, $f(\rho) = \text{constant}$. With Eq. (2.89) and (2.90), the Pupil Astigmatism Criteria (2.50) for the infinite conjugates system are again derived.

2.5 COMPARING THE PUPIL ASTIGMATISM CRITERIA WITH THE CODDINGTON EQUATIONS

It is important to understand the distinction between the pupil astigmatism relations derived here, and the Coddington equations, that are used for determining field curves. The conventional application of the Coddington equations in optical design is to find where the tangential and sagittal rays intersect for small bundles of rays that all go through the center of the stop, and are different for each field point. These give information of the aberrations for any field point, but only those with quadratic pupil dependence. The pupil astigmatism analysis looks at bundles of rays that all go through the object point. Each point in the pupil has its own ray bundle. This information is

used to determine aberrations that have quadratic field dependence, for all points in the pupil. So the conventional treatment is complete for wide field of view systems with long focal ratios. The pupil astigmatism is complete for systems with high NA and small fields.

2.6 CONCLUSION

In this chapter, we present the Pupil Astigmatism Conditions for both systems with object a finite distance away and systems with object at infinity. Three different ways are taken to derive the conditions. The backward ray-tracing version of the criteria shows the entrance pupil astigmatism determines the image aberrations that have quadratic field dependence. This method is mathematically straightforward even though it is not so physically. Then using the principle of energy conservation, the relation between the forward and backward ray tracing properties are obtained, the forward ray-tracing version of the criteria then follow. This version of the criteria shows the exit pupil astigmatism determines the quadratic field-dependent aberrations. Finally, the criteria are derived again by a direct forward ray-tracing approach. This approach is taken again to derive the equivalent criteria for plane symmetric systems in Chapter 7.

CHAPTER 3

VALIDATION OF THE PUPIL ASTIGMATISM CRITERIA TO 3RD ORDER

3.1 INTRODUCTION

It is well known that the first order properties of an axisymmetric imaging system can be used to predict the 3rd order, or Seidel aberrations in the image. In addition to this, it is possible to predict similar aberrations in the pupil (Shack, Welford 1986). The relationship between image and pupil aberrations has also been established to third order. (Shack, Welford 1986). Here we show that these relationships to be consistent with a third order approximation to the more general Pupil Astigmatism Conditions.

The Pupil Astigmatism Conditions were derived by investigating the mapping relationships between pupils. For normal applications, these pupils are defined as images of the aperture stop. In Chapter 2, we use the ray vector in image space as the pupil coordinate (see Eq. (2.5)) to define aberrations. Such definition of aberrations and the small field of view we are concerned determines that the actual stop location is not important. It is convenient to set the entrance pupil at infinity for systems with object a finite distance away. For systems with object at infinity, the stop location can be arbitrarily chosen.

3.2 IMAGE AND PUPIL SEIDEL ABERRATIONS AND THEIR RELATIONS

Seidel aberrations are 3rd order aberrations. There are formulas for calculating Seidel aberrations of an optical system using only its first order properties. To better understand the formulas, we first define 2 quantities. One is *Petzval Sum*:

$$P = \sum c \Delta \left(\frac{1}{n} \right), \quad (3.1)$$

where c is the curvature of an optical surface, n is the index of refraction, $\Delta(\cdot)$ is an operator that calculates the difference between the quantity in the parenthesis after and before the surface, sum is over all the surfaces in the system. The other quantity is

Lagrangian Invariant:

$$H = n\bar{u}y - nu\bar{y}, \quad (3.2)$$

where u and y are the marginal ray angle and height, and \bar{u} and \bar{y} are the chief ray angle and height at a surface (see Figure 3.1). Let i and \bar{i} be the incidence angles of marginal ray and chief ray at a surface, then define A and \bar{A} as follows:

$$A = ni \text{ and } \bar{A} = n\bar{i}. \quad (3.3)$$

With the quantities defined in (3.1) – (3.3), we can calculate the coefficients of Seidel aberrations using the following equations:

$$S_I = -\sum A^2 y \Delta \left(\frac{u}{n} \right), \quad (3.4a)$$

$$S_{II} = -\sum A\bar{A}y \Delta \left(\frac{u}{n} \right), \quad (3.4b)$$

$$S_{III} = -\sum \bar{A}^2 y \Delta\left(\frac{u}{n}\right), \quad (3.4c)$$

$$S_{IV} = -\sum H^2 c \Delta\left(\frac{1}{n}\right), \quad (3.4d)$$

$$S_V = -\sum \left\{ \frac{\bar{A}^3}{A} y \Delta\left(\frac{u}{n}\right) + \frac{\bar{A}}{A} H^2 c \Delta\left(\frac{1}{n}\right) \right\}. \quad (3.4e)$$

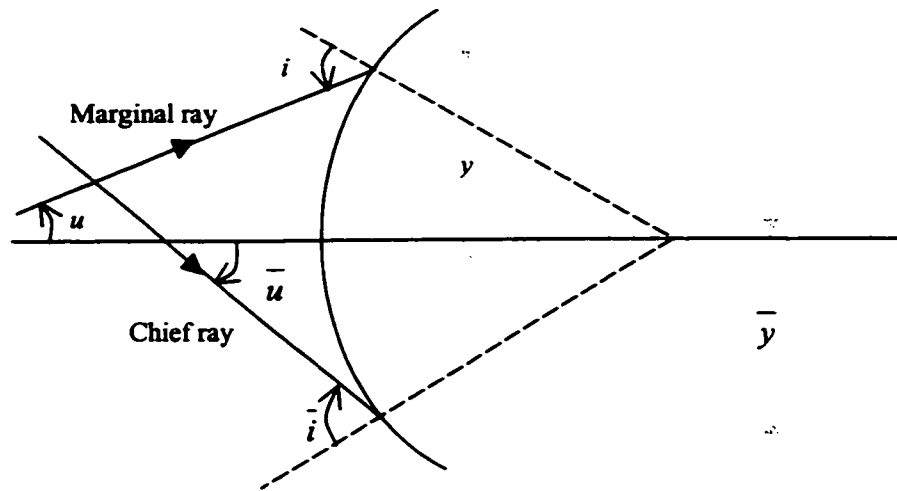


Figure 3.1. Illustration of the ray angles and heights at a surface.

There are Seidel formulas for pupil aberrations as well. Since the chief ray of the image is the marginal ray of the pupil and the marginal ray of the image is the chief ray of the pupil, the image serves as the pupil when the pupil aberration is concerned. So the Seidel formulas for pupil aberrations are similar to Eqs. (3.4) but with the exchanged role of barred and unbarred quantities, i.e. A and \bar{A} are interchanged, so are u and \bar{u} . Then the Seidel formulas for pupil aberrations are:

$$\bar{S}_I = -\sum \bar{A}^2 \bar{y} \Delta \left(\frac{\bar{u}}{n} \right), \quad (3.5a)$$

$$\bar{S}_{II} = -\sum A \bar{A} y \Delta \left(\frac{\bar{u}}{n} \right), \quad (3.5b)$$

$$\bar{S}_{III} = -\sum A^2 \bar{y} \Delta \left(\frac{\bar{u}}{n} \right), \quad (3.5c)$$

$$\bar{S}_{IV} = -\sum H^2 c \Delta \left(\frac{1}{n} \right), \quad (3.5d)$$

$$\bar{S}_V = -\sum \left\{ \frac{A^3}{A} \bar{y} \Delta \left(\frac{\bar{u}}{n} \right) + \frac{A}{A} H^2 c \Delta \left(\frac{1}{n} \right) \right\}. \quad (3.5e)$$

The wave aberration coefficients defined in (1.3) is related to the Seidel coefficients in the following way (Shack):

$$W_{040} = \frac{1}{8} S_I,$$

$$W_{131} = \frac{1}{2} S_{II},$$

$$W_{222} = \frac{1}{2} S_{III},$$

$$W_{220P} = \frac{1}{4} S_{IV},$$

$$W_{311} = \frac{1}{2} S_V.$$

Here we only care about the aberrations with quadratic field dependence, i.e.

astigmatism and field curvature, so we only calculate 3rd and 4th Seidel coefficients:

$$S_{III}, S_{IV} \text{ and } \bar{S}_{III}, \bar{S}_{IV}.$$

The relationships between the image Seidel aberrations and pupil Seidel aberrations are (Shack, Welford 1986):

$$\begin{aligned}\bar{S}_{III} &= S_{III} + H\Delta(u\bar{u}), \\ \bar{S}_{IV} &= S_{IV},\end{aligned}\tag{3.6}$$

where $\Delta(\cdot)$ is the difference between the quantity in the parenthesis in image space and object space. It follows that

$$\bar{S}_{IV} + \bar{S}_{III} = (S_{IV} + S_{III}) + H\Delta(u\bar{u}),\tag{3.7a}$$

$$\bar{S}_{IV} + 2\bar{S}_{III} = (S_{IV} + 2S_{III}) + 2H\Delta(u\bar{u}),\tag{3.7b}$$

$$\bar{S}_{IV} + 3\bar{S}_{III} = (S_{IV} + 3S_{III}) + 3H\Delta(u\bar{u}).\tag{3.7c}$$

The longitudinal pupil aberrations at maximum aperture are (Welford 1986):

$$\bar{\delta z}_{220s} = -\frac{1}{2n\bar{u}^2}(\bar{S}_{IV} + \bar{S}_{III}),\tag{3.8a}$$

$$\bar{\delta z}_{220m} = -\frac{1}{2n\bar{u}^2}(\bar{S}_{IV} + 2\bar{S}_{III}),\tag{3.8b}$$

$$\bar{\delta z}_{220r} = -\frac{1}{2n\bar{u}^2}(\bar{S}_{IV} + 3\bar{S}_{III})\tag{3.8c}$$

and

$$\bar{\delta z}_{220r} - \bar{\delta z}_{220s} = -\frac{\bar{S}_{III}}{n\bar{u}^2},\tag{3.8d}$$

$$\bar{\delta z}_{220r} - 3\bar{\delta z}_{220s} = \frac{\bar{S}_{IV}}{n\bar{u}^2},\tag{3.8e}$$

where n is index of refraction and \bar{u} is the chief ray angle in image space, see Figure 3.2. Here we follow Roland V. Shack's notations for 3rd order longitudinal aberrations such as $\bar{\delta z}_{220s}$, etc.

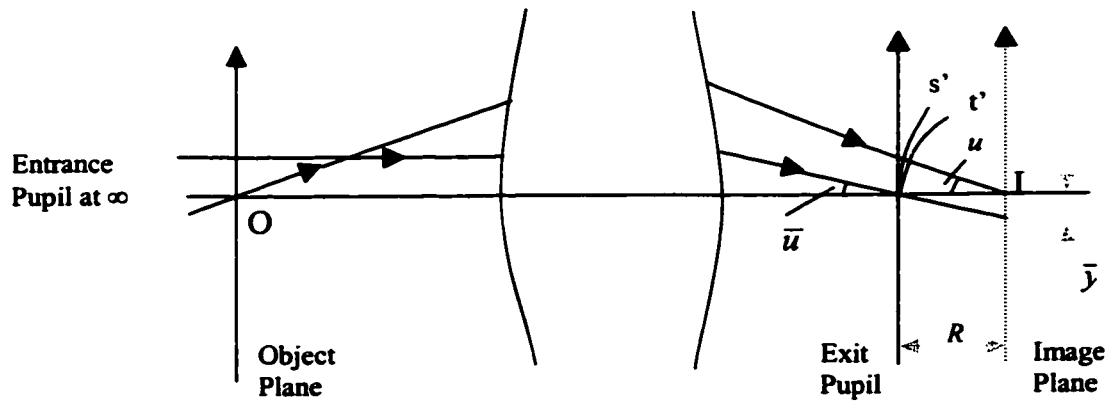


Figure 3.2. Illustrates a system with entrance pupil at infinity, s' and t' denote the sagittal and tangential pupil curves respectively.

In Figure 3.2, according to the sign convention, u , \bar{u} and y are all negative, Let R be positive, then

$$\bar{u} = \frac{\bar{y}}{R}, \quad (3.9)$$

$$H = -nu\bar{y}.$$

So,

$$H\Delta(u\bar{u}) = -nRu^2\bar{u}^2. \quad (3.10)$$

3.3 VALIDATION OF THE PUPIL ASTIGMATISM CRITERIA TO 3RD ORDER

The Pupil Astigmatism Conditions derived in Chapter 2 relate the field curves for the pupil aberrations to the aberrations in the image that have quadratic field dependence. In general these pupil field curves are not simply quadratic and can have any dependence on the pupil parameter. Since we can use the Seidel equations to directly calculate the third order pupil aberrations and the third order image aberrations, we can show the existing pupil – image relationships to be the special cases of the more general Pupil Astigmatism Conditions. Refer to Figure 3.2 for all the parameters used in the calculations in this section.

(a) If condition (2.50b) holds, then

$$\frac{t'(u)}{\cos^2(u)} = R,$$

and the longitudinal pupil aberration in tangential plane is:

$$\delta\bar{z}_l = R - t' \cos(u) = R(1 - \cos^3(u)). \quad (3.11)$$

To the 3rd order,

$$\delta\bar{z}_{220l} = \frac{3Ru^2}{2}. \quad (3.12)$$

Substitute Eq. (3.12) into Eq. (3.8c), we obtain

$$\begin{aligned} \bar{S}_{IV} + 3\bar{S}_{III} &= -3Ru^2\bar{u}^2 \\ &= 3H\Delta(\bar{u}\bar{u}). \end{aligned} \quad (3.13)$$

Then according to Eq. (3.7c),

$$S_{IV} + 3S_{III} = 0, \quad (3.14)$$

which indicates the image's tangential field is flat.

(b) If condition (2.50c) holds, then

$$s'(u) = R,$$

and the longitudinal pupil aberration in sagittal plane is:

$$\bar{\delta z}_s = R(1 - \cos(u)). \quad (3.15)$$

To the 3rd order,

$$\bar{\delta z}_{220s} = \frac{Ru^2}{2}. \quad (3.16)$$

Then, substitute Eq. (3.16) into Eq. (3.8a), we obtain

$$\begin{aligned} \bar{S}_{IV} + \bar{S}_{III} &= -nRu^2\bar{u}^2 \\ &= H\Delta(u\bar{u}). \end{aligned} \quad (3.17)$$

Then according to (3.7a),

$$S_{IV} + S_{III} = 0, \quad (3.18)$$

which indicates the image's sagittal field is flat.

(c) If condition (2.50a) holds, then

$$s'(u) = \frac{t'(u)}{\cos^2(u)} = R,$$

therefore both Eq. (3.14) and Eq. (3.18) are true, so we can solve the equations and get

$$S_{III} = 0 \text{ and } S_{IV} = 0, \quad (3.19)$$

which indicates both the image's astigmatism and field curvature are 0.

(d) If condition (2.50d) holds, then

$$s'(u) = \frac{r'(u)}{\cos^2(u)} = R(u),$$

and the longitudinal pupil aberrations in the tangential plane is

$$\delta\bar{z}_t = R - R(u)\cos^3(u)$$

(3.20a)

and the longitudinal pupil aberrations in the sagittal plane is

$$\delta\bar{z}_s = R - R(u)\cos(u).$$

(3.20b)

Then,

$$\delta\bar{z}_t - \delta\bar{z}_s = R(u)\cos(u)(1 - \cos^2(u)). \quad (3.21)$$

The 3rd order approximation gives

$$\delta\bar{z}_{220t} - \delta\bar{z}_{220s} = Ru^2. \quad (3.22)$$

Combine Eq. (3.22) with Eq. (3.8d), we obtain

$$\bar{S}_{III} = -nRu^2\bar{u}^2. \quad (3.23)$$

It then follows that

$$S_{III} = 0, \quad (3.24)$$

which indicates the image's astigmatism is 0.

3.4 CONCLUSION

The forward ray tracing version of the Pupil Astigmatism Criteria for correcting the quadratic field-dependent aberrations implies the connection of the criteria with the relationships between image and pupil aberrations. The formulas that give the relations of 3rd order image and pupil aberrations were revisited in this chapter. By using these relationships we were able to validate the criteria to 3rd order, i.e., when one of the criteria is satisfied, the corresponding 3rd order quadratic field-dependent aberrations -- astigmatism and/or field curvature -- are indeed corrected.

CHAPTER 4

USING THE PUPIL ASTIGMATISM CRITERIA TO ANALYZE THE KNOWN DESIGN EXAMPLES

4.1 INTRODUCTION

Aberration theories can help people design well-corrected systems. The Seidel aberration coefficient formulas, the Petzval theorem, the Constant Optical Path Length Condition and the Abbe Sine Condition have been known to people for more than a hundred years, and many elegant optical systems were designed under the guide of the above mentioned aberration theories. Since the Pupil Astigmatism Criteria presented and derived in Chapter 2 are new to optical designers, we can use them to analyze the known designs. The reason for doing this is two-fold: one is to verify the criteria to some extent, the other is to extract new information from the old designs.

In this chapter, several systems are analyzed, which include the aplanatic conjugate pairs of a refractive sphere, the well-known Offner relay and Dyson system. An idea inspired by the Offner and Dyson system leads to the design of a system with superior performance: the modified Bouwers system. Lunberg lens is also presented to show the effectiveness of the Pupil Astigmatism Criterion (2.50d) for systems with object at infinity.

4.2 THE APLANATIC CONJUGATE PAIR OF A REFRACTIVE SPHERE

A sphere with radius of curvature R and refractive index n has a pair of conjugate points where perfect imagery is formed, therefore all orders of spherical aberration are corrected. In the Figure 4.1, the index of refraction in the image space is 1, the distance from the object to the vertex of the sphere surface is $(1 + \frac{1}{n})R$, and the distance from the image to the vertex of the sphere is $(1 + n)R$. For this pair of conjugate points, the Sine condition is strictly satisfied, i.e. $\frac{\sin U}{n} = \sin U'$, then we know all orders of coma are corrected as well. In paraxial approximation $\frac{u}{n} = u'$, from Seidel formulas (3.4a) – (3.4c), we know all 3rd order spherical aberration, coma and astigmatism are absent.

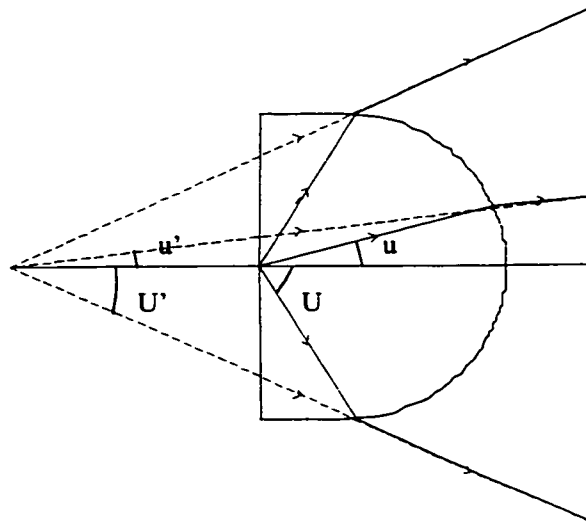


Figure 4.1. The aplanatic conjugate pair of a refractive sphere.

As far as spherical aberration and coma are concerned, the Constant OPL Condition and the Abbe Sine Condition tells us more information than the Seidel formulas because the Seidel formulas are approximated to 3rd order and the Constant OPL and Sine Conditions are exact. Do the Pupil Astigmatism Criteria in (2.50) provide more information than Seidel formula too? The answer is yes. The proof is given below. Seidel formulas say that the 3rd order astigmatism is absent while the Petzval curvature is present, this fact indicates criterion (2.50d) may be true in this case. We will check criterion (2.50d) to see whether it is strictly satisfied. To do that we need to calculate s and t and then compare $\frac{\cos^2 \theta}{t}$ and $\frac{1}{s}$. The Coddington equations are used to calculate s and t . See Chapter 6 for the application of Coddington equations.

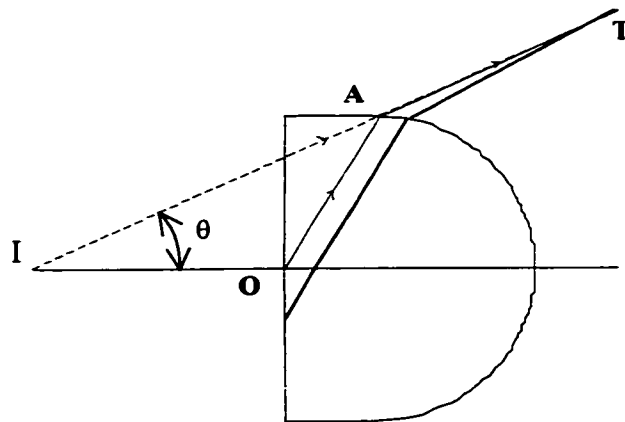


Figure 4.2. Illustration of calculating s and t for the aplanatic conjugate pair of a refractive sphere.

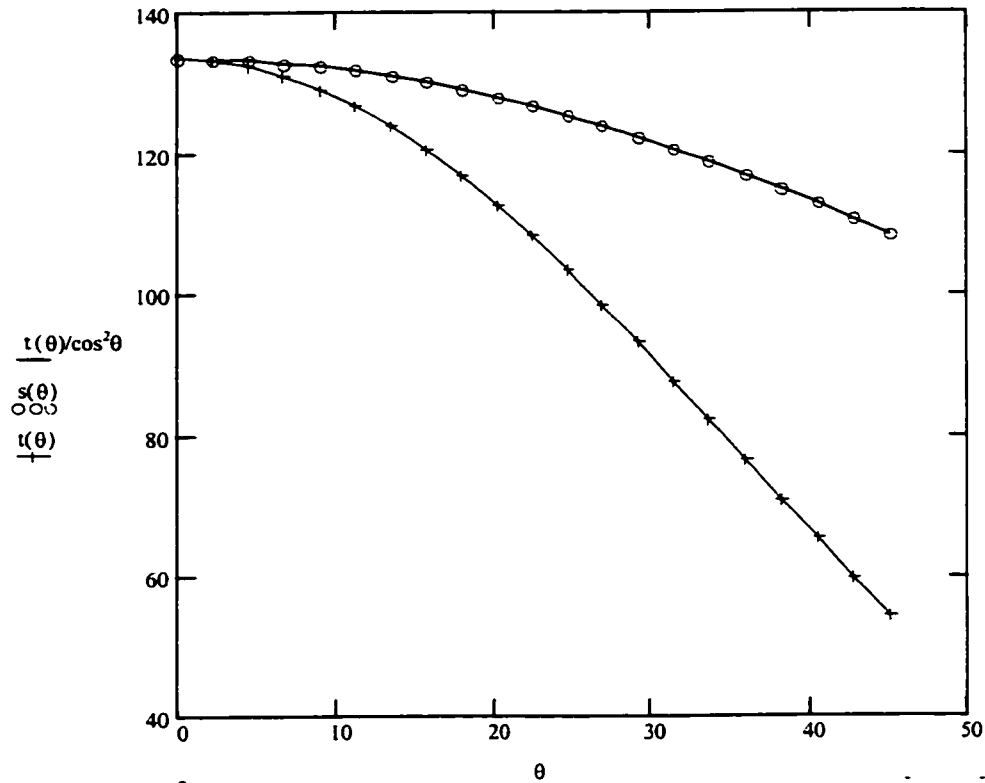


Figure 4.3 Plots of s , t and $t/\cos^2\theta$ as functions of θ for the aplanatic system shown in Figure 4.1.

Figure 4.2. Illustrates how to trace the tangential rays to get t . In this figure, $t = TI$ and is a function of the ray angle θ in image space. Similarly, sagittal rays can be traced to get s . Assume the thin bundle of the sagittal rays focus on point S, then $s = SI$, also a function of θ . Figure 4.3 plots s , t and $t/\cos^2\theta$ as functions of θ . The s curve and $t/\cos^2\theta$ curve coincide exactly, which means criterion (2.50d) is perfectly satisfied, therefore all orders of astigmatism of quadratic field dependence are corrected. The 3rd order prediction of the Seidel formula is consistent with this result.

4.3 THE LUNEBURG LENS

The Luneburg lens, an infinite conjugate system, is presented below to show the application of the Pupil Astigmatism Criterion in (2.50d). The Luneburg lens is a unit sphere with the index of refraction

$$n(r) = \sqrt{2 - r^2}.$$

This lens images the object at infinity perfectly on its surface. Figure 4.4 (a) shows the layout of the Luneburg Lens. Figure 4.4 (b) shows the index of refraction profile of the lens.

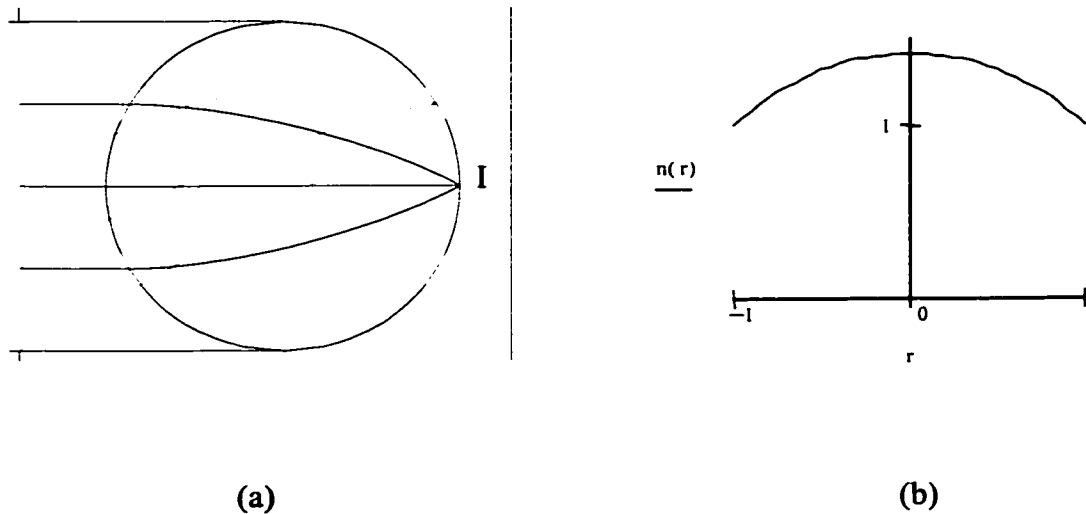


Figure 4.4. (a) The Luneburg lens. Object at infinity is perfectly imaged onto a spherical surface. (b) The index of refraction profile of the Luneburg Lens.

If we trace a thin bundle of parallel rays backward from the image point I, then the tangential and sagittal foci coincide on the front surface of the lens. According to the Pupil Astigmatism Criterion (2.22d), all the quadratic field dependent aberrations of the form W_{2n2} are predicted to be absent. Direct analysis of the system shows that no

astigmatism of any order exists in this imaging system, which conforms to the prediction of the criterion. Because $t(h)=s(h)\neq\text{constant}$, the Pupil Astigmatism Condition predicts the presence of aberrations of the form W_{2n0} , indeed there is significant field curvature W_{220} in this system, so the validity of the criterion is again verified here.

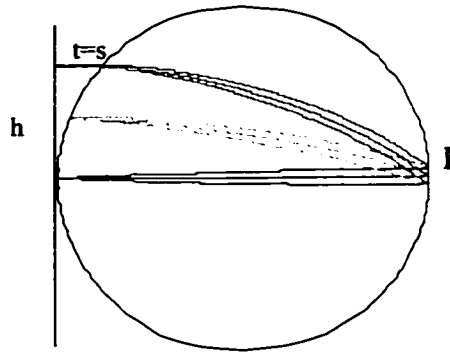


Figure 4.5 Trace a thin bundle of rays from the image point I backward through the Luneburg Lens. The tangential and sagittal foci coincide on the front surface of the lens.

We used the backward ray tracing version of the Pupil Astigmatism Criterion above to show all orders of astigmatism are absent in the Luneburg lens. Below we also show that the forward ray tracing version of the criterion predicts the correct result as well. We use the optical analysis code ZEMAX to trace rays from any point in a plane that is perpendicular to the optical axis in object space. ZEMAX determines the locations of tangential and sagittal images, and s and t are then calculated from this information. Table 4.1 shows the calculated s and $t/\cos^2\theta$. The equality of the two quantities indicates that astigmatism of any order is absolutely corrected according to the Pupil Astigmatism Criterion (2.50d). Again this agrees with the well-known behavior of this imaging system.

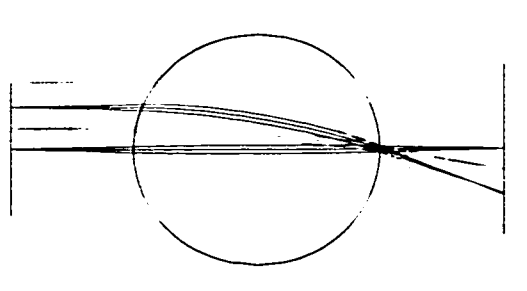


Figure 4.6. Trace a small cone of rays from any point in the plane in the object space through the Luneburg lens. s and t are then calculated from the ray tracing information.

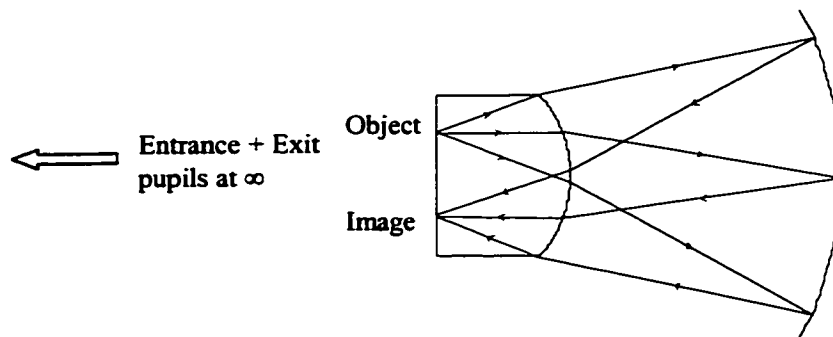
Table 4.1. Forward ray tracing result of the Luneburg Lens. It is shown that $t/\cos^2\theta = s$.

Ray Angle θ (in degree)	$t/\cos^2\theta$	s
0.000	-100.000	-100.000
4.589	-99.681	-99.681
9.207	-98.728	-98.728
13.887	-97.160	-97.160
18.663	-95.004	-95.004
23.578	-92.295	-92.295
28.685	-89.069	-89.068
34.056	-85.360	-85.360
39.792	-81.194	-81.194
46.054	-76.568	-76.568
53.130	-71.429	-71.429

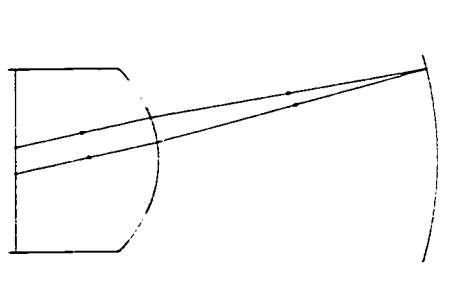
4.4 THE RETROREFLECTIVE CONCENTRIC SYSTEMS

A special case is when both t and s are equal to infinity and independent of the ray angles, then the Pupil Astigmatism Condition (2.50a) is satisfied which predicts all the quadratic field-dependent aberrations are corrected. For a concentric and retroreflective system, if the object is put in the plane that contains the common center of curvature, then t and s are both equal to infinity and independent of the ray angle θ . The system should be free of all the quadratic field-dependent aberrations including the 3rd order astigmatism and field curvature. The Dyson and Offner systems fall into this category, see Figure 4.7 and 4.8. We analyzed the two systems in ZEMAX, where all the Seidel aberrations are shown to be zero. Also the residual aberrations are proportional to the 4th power of the field, which means the aberrations of quadratic field dependence are corrected. So the Pupil Astigmatism Criterion does predict the correct result in this case.

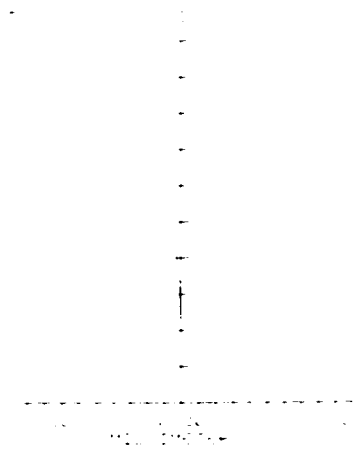
Because of the symmetry, we traced only one bundle of rays, and we know that 3rd order astigmatism and field curvature are absent according to the criterion. If we use the Coddington equations, we have to trace rays for a good number of field points. Next we generate the field curve plot like those shown in Figure 4.7(c) and 4.8(c); we could then see that no 3rd order astigmatism and field curvature are present. The Pupil Astigmatism Criteria are certainly advantageous over the Coddington equations in this case.



(a)

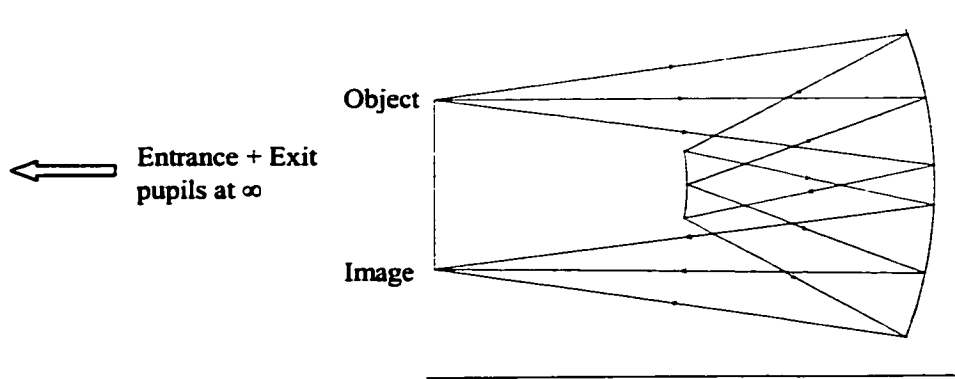


(b)

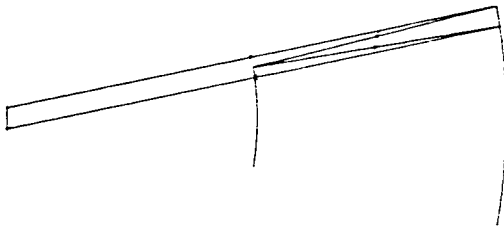


(c)

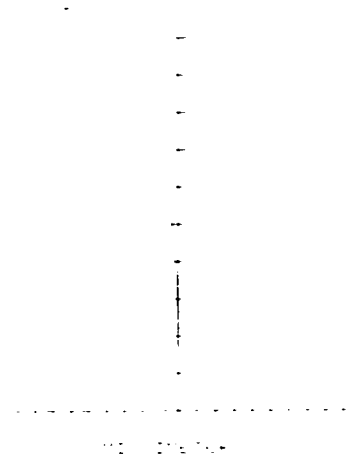
Figure 4.7. (a) Dyson system. (b) Shows Dyson system is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4th power of the field height.



(a)

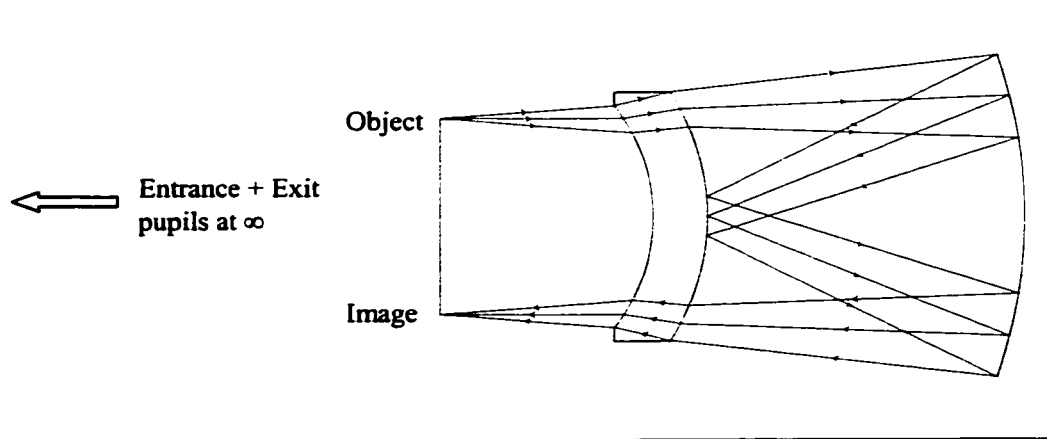


(b)

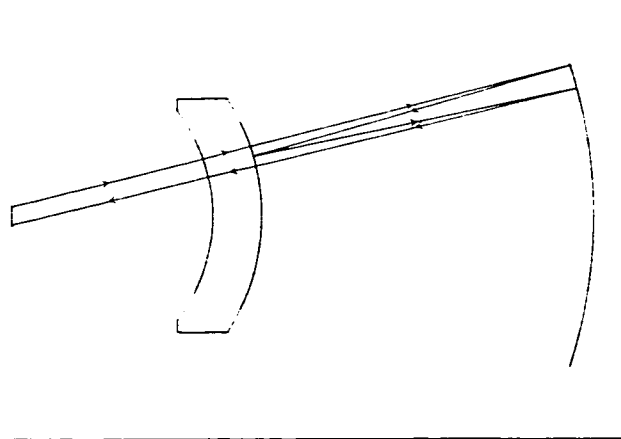


(c)

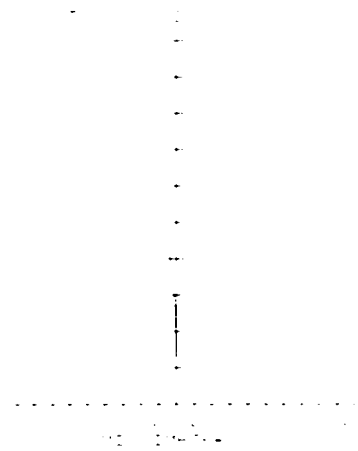
Figure 4.8. (a) The Offner relay. (b) Shows Offner relay is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4th power of the field height.



(a)



(b)



(c)

Figure 4.9. (a) The modified Bouwers system. (b) Shows Bouwers system is retroreflective, which means $t = s = \infty$. (c) Astigmatism is shown to be proportional to the 4th power of the field height for a moderate field of view.

Another example is modified from a Bouwers system, see Figure 4.9. The system is designed so that the plane wave is focused on the convex surface of the meniscus lens. This surface has a reflective coating that makes the system retroreflective. If an object is put in the plane that contains the common center of curvature, then the image will be in the same plane. The system should be free of all quadratic field-dependent aberrations including 3rd order astigmatism and field curvature just like the Offner relay and the Dyson system are, and we verified this with a ZEMAX simulation. Compared to the Offner relay, the residual higher-order astigmatism is greatly reduced for a moderate field.

4.5 CONCLUSION

Several well-known optical systems are examined with the Pupil Astigmatism Criteria in this chapter. These optical systems were designed under the guidance of the known aberration theories, so certain aberrations are corrected as intended. What is not known before is that some other aberrations are also corrected. After analyzing these designs with the Pupil Astigmatism Criteria, we gain new information. The usefulness of the criteria is demonstrated here. In the following chapters, more applications of the criteria will be shown.

CHAPTER 5

PREDICTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS

5.1 INTRODUCTION

The Pupil Astigmatism Conditions tell us when certain or all types of quadratic field-dependent aberrations are absent. When these conditions are not satisfied exactly, we now have a method to predict what types of aberrations with quadratic field dependence are present and what the amount of each type is. In this chapter, we will give the formulas to calculate the quadratic field-dependent aberrations. Several examples are given to support our argument.

5.2 FORMULAS FOR CALCULATING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS

From the equations in (2.6), we know that $p_1^2 W_2(\rho) + W_3(\rho)$ is the coefficient of the x^2 (field square) term in the Taylor's series expansion of the mixed characteristic function $W(\rho)$ and its derivatives with regard to p_l and q_l are the coefficients of the x^2 term of the transverse ray aberration. When an optical system is strictly aplanatic, i.e. no spherical aberration and coma of any order is present, then the primary aberrations are quadratic field-dependent, namely,

$$\Delta x = -x_0^2 \frac{\partial}{\partial p_1} (p_1^2 W_2(\rho) + W_3(\rho)), \quad (5.1a)$$

$$\Delta y = -x_0^2 \frac{\partial}{\partial q_1} (p_1^2 W_2(\rho) + W_3(\rho)). \quad (5.1b)$$

In the derivation of the Pupil Astigmatism Conditions in Chapter 2, we obtained the expressions for calculating $p_1^2 W_2(\rho) + W_3(\rho)$ in both tangential and sagittal plane. For systems with object a finite distance away, we have

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{n_i M^2(\rho) \cos^2(\theta)}{2t}, \quad (5.2a)$$

$$W_3(\rho) = -\frac{n_i M^2(\rho)}{2s}, \quad (5.2b)$$

where n_i is the index of refraction of the medium in image space and $M(\rho)$ is the magnification for a specific pupil point ρ . For systems with object at infinity, we have

$$p_1^2 W_2(\rho) + W_3(\rho) = -\frac{n_i f^2(\rho) \cos^2(\theta)}{2t}, \quad (5.3a)$$

$$W_3(\rho) = -\frac{n_i f^2(\rho)}{2s}, \quad (5.3b)$$

where again n_i is the index of refraction of the medium in image space and $f(\rho)$ is the effective focal length for a specific pupil point ρ .

When a system is strictly aplanatic, $M(\rho) = \text{constant}$ for finite conjugate systems and $f(\rho) = \text{constant}$ for infinite conjugate systems. From Eqs. (5.2) and (5.3), the Pupil Astigmatism Criteria (2.50) follow. Even when the criteria are not satisfied, we can still get the information of the quadratic field-dependent aberrations from t and s .

Assume the Taylor's series expansion of W_2 and W_3 are:

$$W_2(\rho) = W_{222} + W_{242}\rho^2 + W_{262}\rho^4 + \dots, \quad (5.4)$$

$$W_3(\rho) = W_{200} + W_{220}\rho^2 + W_{240}\rho^4 + W_{260}\rho^6 + \dots, \quad (5.5)$$

where $\rho^2 = p_1^2 + q_1^2 = n_i^2 \sin^2 \theta$, then substitute $W_2(\rho)$ and $W_3(\rho)$ into Eqs. (5.1a) and (5.1b), we obtain

$$\Delta x = -x_0^2 \sum_n n(W_{2n2} + W_{2n0})\rho^{n-1}, \quad (5.6a)$$

$$\Delta y = -x_0^2 \sum_n nW_{2n0}\rho^{n-1}. \quad (5.6b)$$

So we expand $\frac{\cos^2(\theta)}{t}$ and $\frac{1}{s}$ in the power series of $\sin^2 \theta$, which is proportional to ρ^2 ,

we can get the coefficients $W_{222}, W_{242}, W_{262}, W_{282}$, etc and $W_{220}, W_{240}, W_{260}, W_{280}$, etc, then detailed information of the quadratic field-dependent aberrations is known.

For a system with object a finite distance away, assume the Taylor series expansion of $1/s$ turns out to be

$$\frac{1}{s(\theta)} = a_1 + b_1 \sin^2 \theta + c_1 \sin^4 \theta + d_1 \sin^6 \theta + \dots, \quad (5.7)$$

and the Taylor series expansion of $\frac{\cos^2 \theta}{t(\theta)} - \frac{1}{s(\theta)}$ turns out to be

$$\frac{\cos^2 \theta}{t(\theta)} - \frac{1}{s(\theta)} = b_2 \sin^2 \theta + c_2 \sin^4 \theta + d_2 \sin^6 \theta + \dots \quad (5.8)$$

Then

$$W_{200} = -\frac{n_i M^2}{2} a_1, \quad (5.9a)$$

$$W_{220} = -\frac{n_i M^2 b_1}{2 n_i^2}, \quad (5.9b)$$

$$W_{240} = -\frac{n_i M^2 c_1}{2 n_i^4}, \quad (5.9c)$$

$$W_{260} = -\frac{n_i M^2 d_1}{2 n_i^6}. \quad (5.9d)$$

and

$$W_{222} = -\frac{n_i M^2 b_2}{2 n_i^2}, \quad (5.10a)$$

$$W_{242} = -\frac{n_i M^2 c_2}{2 n_i^4}, \quad (5.10b)$$

$$W_{262} = -\frac{n_i M^2 d_2}{2 n_i^6}. \quad (5.10c)$$

Below we examine some examples where the criteria are not exactly satisfied.

We expand $1/s$ and $\cos^2\theta/t-1/s$ and calculate the coefficient of W_{2n0} and W_{2n2} (n is even).

We compare the aberrations calculated this way with the ray tracing result in ZEMAX.

All the systems examined are strictly aplanatic, i.e. no spherical aberration of any order and no coma of any order exists.

5.3 SPHERICAL MIRROR

The simplest aplanatic system is a spherical mirror with the object at the plane that contains the center of curvature, see Figure 5.1 (a). Since the marginal ray hits the

mirror normally, then the quantity A in the Seidel formulas (3.4a) and (3.4b) is equal to 0. Therefore the 3rd order spherical aberration and coma are all absent.

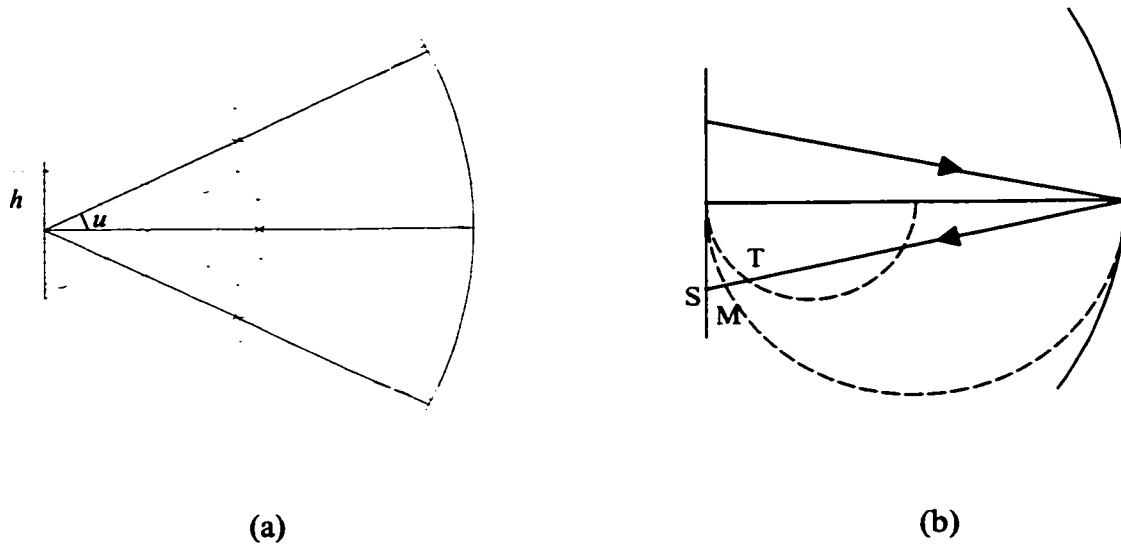


Figure 5.1. (a) A spherical mirror. The object is at the plane that contains the center of curvature of the spherical mirror. (b) Shows the sagittal image of the system is flat. The tangential image is on the circle of radius of curvature $R/4$ and the medial image is on the circle of radius of curvature $R/2$ where R is radius of curvature of the mirror.

As for the astigmatism and field curvature, the Seidel formulas (3.4c) and (3.4d) give the following result:

$$S_{III} = \frac{2}{R} h^2 u^2, \quad (5.11)$$

and

$$S_{IV} = -\frac{2}{R} h^2 u^2. \quad (5.12)$$

It follows that (Welford 1986)

$$\begin{aligned}
 W_{220} &= \frac{1}{h^2 u^2} (S_{III} + S_{IV}) \\
 &= 0,
 \end{aligned} \tag{5.13}$$

and

$$\begin{aligned}
 W_{222} &= \frac{1}{h^2 u^2} (2S_{III} + S_{IV}) \\
 &= \frac{2}{R}.
 \end{aligned} \tag{5.14}$$

Eq. (5.13) indicates the sagittal image is flat. Further analysis shows the tangential image is on the circle with radius of curvature $R/4$ and the medial image is on the circle with radius of curvature $R/2$. Figure 5.1 (b) shows this result.

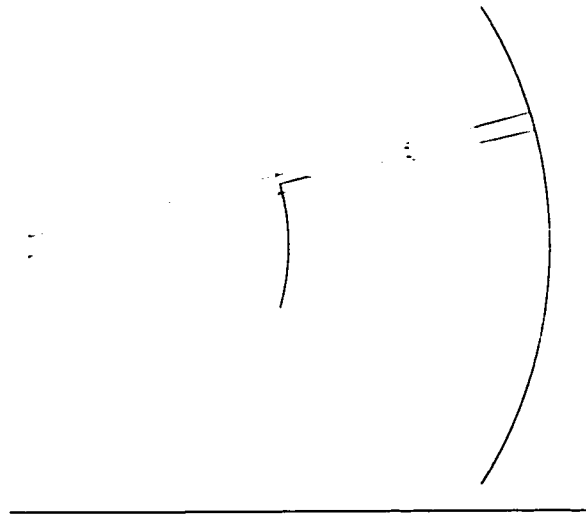


Figure 5.2. Illustrates the ray tracing method to get t and s for the system shown in Figure 5.1 (a). $t(\theta) = s(\theta) = R/2$, where R is the radius of curvature of the spherical mirror.

Besides the 3rd order spherical aberration and coma, the constant OPL condition and the Abbe Sine condition tells us that all orders of spherical aberration and linear coma are zero for this system. These two conditions give us more information about the system than the Seidel formulas do. We would like to know whether the Pupil Astigmatism Criteria (2.50) give us more information than Seidel formulas too.

Figure 5.2 shows the ray tracing method to get t and s . If the radius of curvature of the spherical mirror is R (>0), then it is easily seen that $t = s = -R/2$. And t and s are independent of the ray angle θ . Since $s = -R/2 = \text{constant}$, the Pupil Astigmatism Criterion (2.50c) is strictly satisfied, so the sagittal image is free of any quadratic field-dependent aberrations. According to Eqs. (5.7) and (5.9)s, all the aberrations of the form W_{2n0} , where n is even, are zero. And the fact that $t = -R/2 = \text{constant}$ indicates that the quadratic field-dependent aberrations are present in the tangential plane. Further analysis shows that

$$\frac{\cos^2 \theta}{t} - \frac{1}{s} = \frac{2 \sin^2 \theta}{R}. \quad (5.15)$$

Combine Eqs. (5.15), (5.8) and (5.10)s, we then conclude

$$W_{222} = \frac{2}{R} \quad (5.16)$$

and all higher order quadratic field-dependent astigmatism are zero.

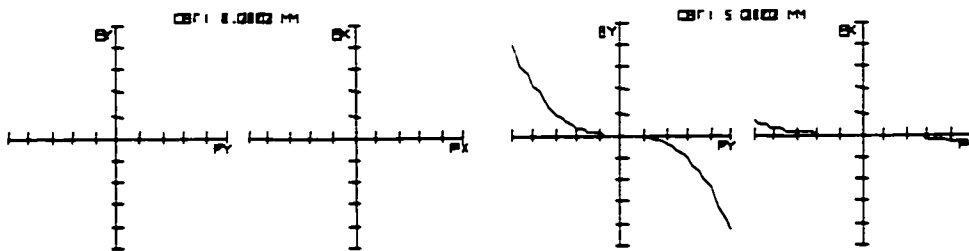
So the Pupil Astigmatism Criteria predict the same result as the Seidel formulas do to 3rd order, but the criteria provide the information on higher order aberrations too. This simple example demonstrates the power of the Pupil Astigmatism Conditions and the result presented in the previous sections that comes with the criteria.

5.4 A 4-ELEMENT SYMMETRIC SYSTEM

Figure 5.3(a) shows a symmetric system with unit magnification -1. Half of the system is made up of a plano-hyperbolic lens and a concentric meniscus lens. The hyperbolic surface has a conic constant $e = -n^2$, where n is the index of refraction of the glass. Thus, the lens focuses the plane wave sharply, i.e. no spherical aberration of any order is present. The concentric meniscus lens makes the Petzval sum go to zero without disturbing the converging wavefront after the plano-hyperbolic lens. The symmetry of the system ensures that the Sine Condition is satisfied. The surface data of the system are displayed in Table 5.1.



(a)



(b)

Figure 5.3. (a) The layout of a symmetric anastigmatic system. (b) The ray fans of the system.

Table 5.1. Surface data of the optical system shown in Figure 5.3 (a).

Surf	Radius	Thickness	Glass	Conic
OBJ	Infinity	30		0
1	-30	7.5	1.50	0
2	-37.5	262.5		0
3	150	10	1.50	-2.25
4	Infinity	76.67		0
STO	Infinity	76.67		0
6	Infinity	10	1.50	0
7	-150	262.5		-2.25
8	37.5	7.5	1.50	0
9	30	30		0
IMA	Infinity			0

Analyzing the system in ZEMAX shows that all 3rd order aberrations are absent.

We also calculate t and s accurately by using the Coddington equations. Table 5.2

displays the result of the calculations. The calculations show that Pupil Astigmatism

Criterion (2.50a) is not exactly satisfied which indicates some quadratic field-dependent

aberrations are present. The ray fans in Figure 5.3 (b) suggest that the oblique spherical

aberration and the 5th order astigmatism are present. The analysis below shows this

matches the prediction of the theory.

Expanding $p_1^2 W_2(\rho)$ and $W_3(\rho)$ in Taylor series, we obtain

$$p_1^2 W_2(\rho) = 35.99 \times 10^{-3} \sin^4 \theta, \quad (5.17a)$$

$$W_3(\rho) = -4.63 + 5.51 \sin^4 \theta. \quad (5.17b)$$

Then according to Eqs. (5.4) and (5.5), we have

$$W_{240} = 5.51 \text{ m}^{-1} \text{ and } W_{242} = 35.99 \times 10^{-3} \text{ m}^{-1},$$

and other aberrations coefficients are zero. Figure 5.4 plots the calculated discrete values of $W_3(\rho)$ and $p_1^2 W_2(\rho) + W_3(\rho)$ versus $\rho (= \sin(\theta))$ and also the continuous 4th order fitted curves.

Using Eqs. (5.17)s and (5.6)s, we plot the theoretical ray aberrations in both tangential and sagittal plane for a field ($x = 0.5\text{mm}$, $y = 0$) as solid lines in Figure 5.5. In the same figure, the discrete dots are the ray tracing values obtained with ZEMAX. The theoretical predictions and the ray tracing results agree to each other quite well. This example verifies our argument in Section 5.2.

Table 5.2. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.3 (a).

θ (degree)	s (mm)	$t/\cos^2(\theta)$ (mm)	$W_3(\rho) = -1/(2s)$ (m^{-1})	$P_1^2 W_2(\rho) = 1/(2s) - \cos^2(\theta)/(2t)$ (m^{-1})
0	108	108	-4.63	0
0.9	108	108	-4.63	0
1.8	108	108.001	-4.63	0.042867
2.7	108.001	108.005	-4.63	0.171458
3.6	108.002	108.016	-4.63	0.600037
4.5	108.005	108.039	-4.629	1.456882
5.4	108.01	108.079	-4.629	2.955383
6.3	108.018	108.145	-4.629	5.435896
7.2	108.031	108.245	-4.628	9.150136
8.1	108.05	108.388	-4.627	14.43048
9.0	108.077	108.583	-4.626	21.55884

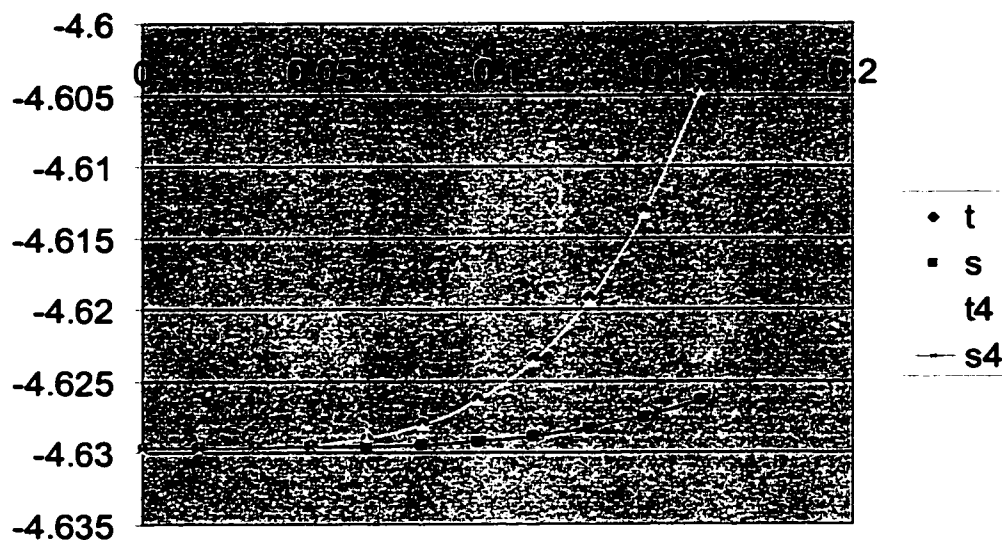


Figure 5.4. The plots of $-1/(2s)$ and $-\cos^2\theta/(2t)$ as functions of $\sin\theta$ for the symmetric systems shown in Figure 5.3(a). The continuous curves are the fitted 4th power curves.

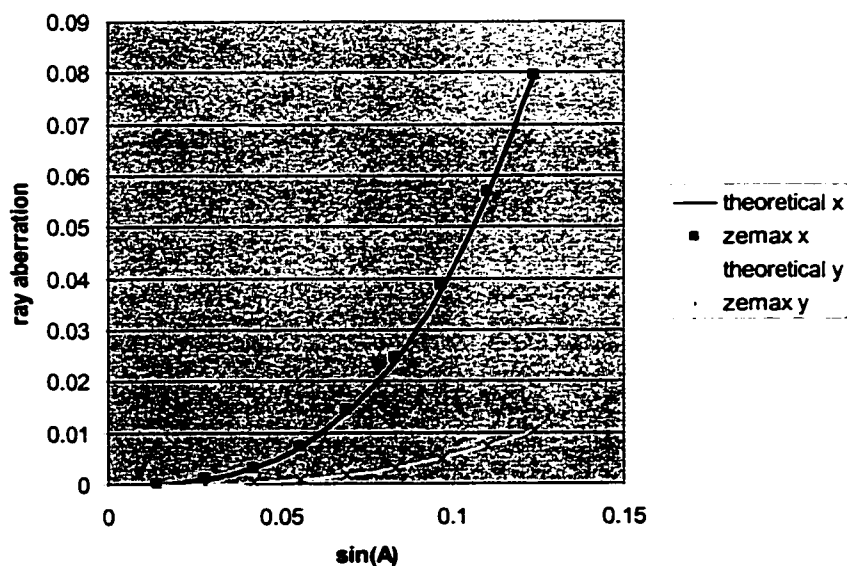


Figure 5.5. Plot of the tangential and sagittal ray aberrations vs. $\sin(\theta)$. The discrete values are the ZEMAX ray tracing result. The solid lines are the theoretical predictions.

5.5 A SYSTEM WITH FLAT TANGENTIAL IMAGE

We will examine another aplanatic system which has a flat tangential field. Figure 5.6 shows the system. The surface data of the system is shown in Table 5.3. The magnification of the system is $m = 5.333418$. The system is strictly aplanatic. We can obtain the values of t and s by tracing a pair of closely spaced parallel rays in ZEMAX. The calculated $t/\cos^2(\theta)$ and $s(\theta)$ for the system are listed in Table 5.4. It is shown that $t/\cos^2(\theta)$ is quite close to constant here. The spot diagram in Figure 5.7 (b) shows that the tangential field is nearly perfect as is consistent with the theory.

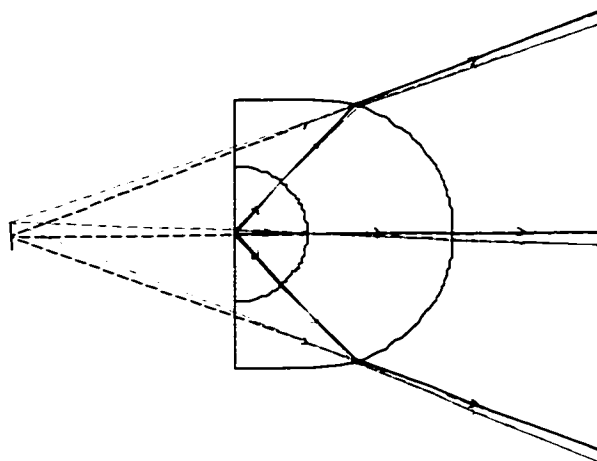


Figure 5.6. An aplanatic system with flattened tangential field.

Table 5.3. Surface data of the optical system shown in Figure 5.6.

Surf	Radius	Thickness	Glass	Conic
OBJ	Infinity	0		0
1	Infinity	50	2.67	0
2	-50	100	2.00	0
STO	-100	100		0
4	Infinity	-400		0
IMA	Infinity			0

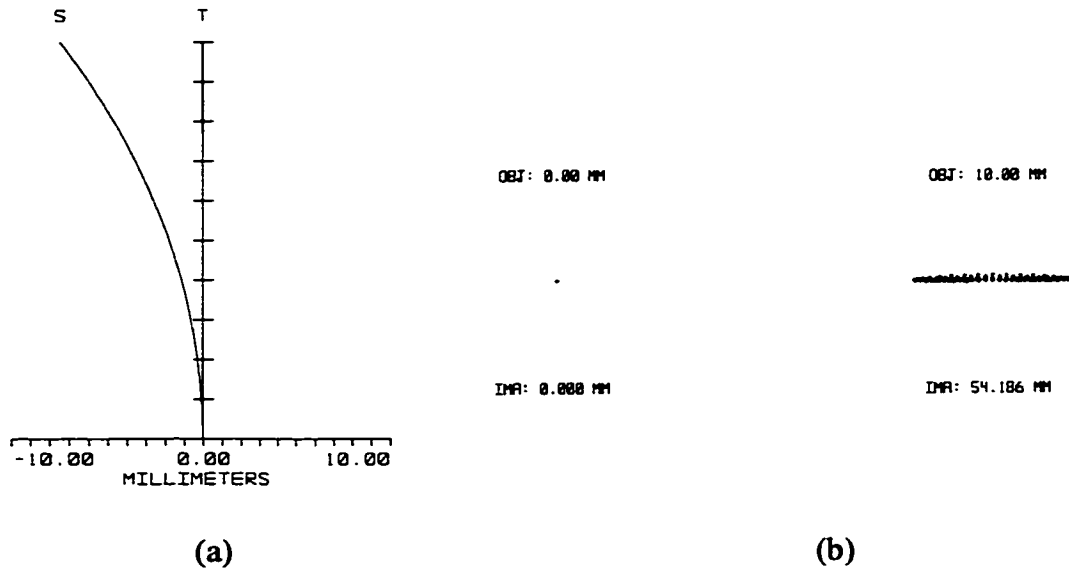


Figure 5.7. (a) The field curve, and (b) the spot diagram of the system shown in Figure 5.6. The maximum field is 10 mm.

Table 5.4. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.6.

θ (degree)	$t/\cos^2(\theta)$ (mm)	s (mm)	$W_3(\rho) = -m^2/(2s)$ (m^{-1})	$-m^2 \cos^2(\theta)/(2t)$ (m^{-1})
0	-319.997	-319.997	44.44627	44.44627
1.264	-319.997	-319.872	44.46364	44.44627
2.478	-319.996	-319.517	44.51304	44.44641
3.601	-319.99	-318.983	44.58756	44.44724
4.606	-319.979	-318.336	44.67818	44.44877
5.485	-319.96	-317.638	44.77636	44.45141
6.238	-319.934	-316.94	44.87497	44.45502
6.875	-319.903	-316.277	44.96904	44.45933
7.412	-319.869	-315.668	45.0558	44.46405
7.862	-319.835	-315.12	45.13415	44.46878
8.239	-319.8	314.634	45.20387	44.47365

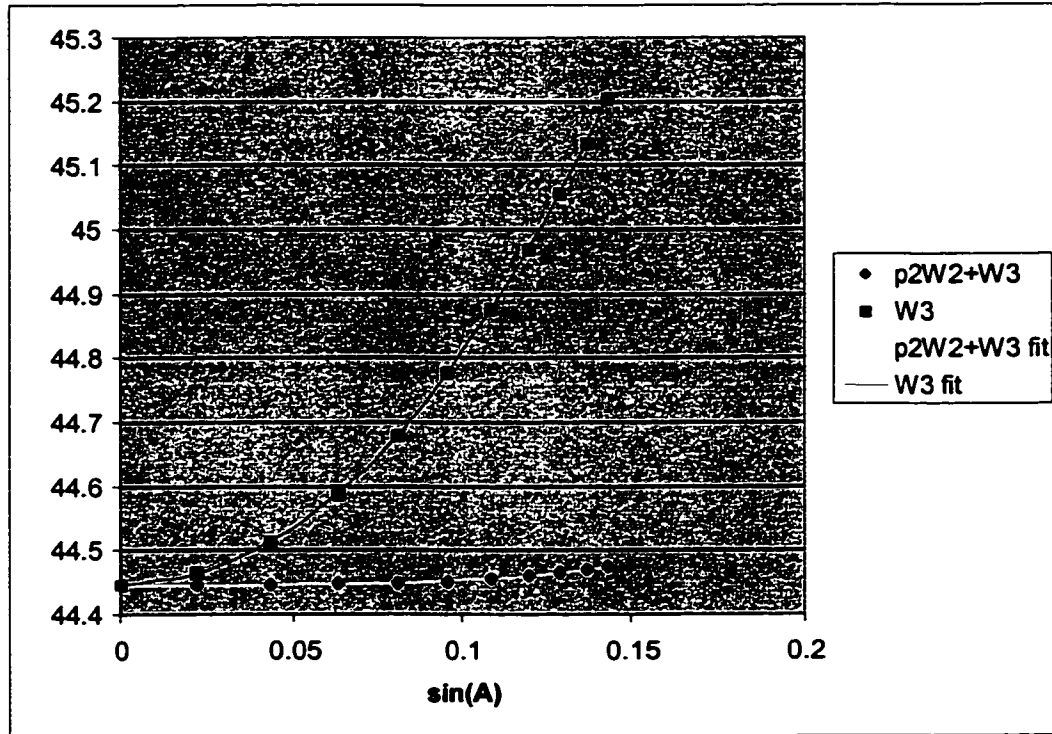


Figure 5.8. Plots of $W_3(\rho)$ and $p_1^2 W_2(\rho) + W_3(\rho)$ vs. ρ for the system shown in Figure 5.6. The dots are the data calculated from the values of s and t , and the solid lines are the polynomial fits (5.18).

With the values of t and s obtained by tracing rays in ZEMAX, we can calculate

$W_3(\rho)$ and $p_1^2 W_2(\rho) + W_3(\rho)$ using Eqs. (5.2). Then fit the calculated data, we get

$$W_3(\rho) = 44.44627 + 35.52\rho^2 + 66.423\rho^4, \quad (5.18a)$$

$$p_1^2 W_2(\rho) + W_3(\rho) = 44.44627 + 64.77\rho^4. \quad (5.18b)$$

Applying Eqs. (5.4) and (5.5), we then obtain

$$W_{200} = 44.44627 \text{ m}^{-1},$$

$$W_{220} = 35.52 \text{ m}^{-1},$$

$$W_{240} = 66.423 \text{ m}^{-1},$$

$$W_{222} = -35.52 \text{ m}^{-1},$$

$$W_{242} = -1.653 \text{ m}^{-1}.$$

For a field point at ($x = 0.5\text{mm}$, $y = 0$), pupil coordinate $\sin\theta = 0.1866577$, using Eqs. (5.1) and the above coefficients, we calculate its ray aberration: ($\Delta x = 3.747 \mu\text{m}$, $\Delta y = 0.421 \mu\text{m}$). Compared to the ZEMAX ray tracing result: ($\Delta x = 3.7771 \mu\text{m}$, $\Delta y = 0.441 \mu\text{m}$), this is a very good prediction.

5.6 THE APLANAIC CONJUGATE PAIR OF A REFRACTIVE SPHERE

Finally we look at the aplanatic conjugate pair of a refractive sphere. Such a system is free of quadratic field dependent astigmatism of all orders (see Chapter 4.2), but significant field curvature is present. The system layout is shown in Figure 5.9(a) and the ray fans of this system are plotted in Figure 5.10. Since the index of refraction of the sphere is $n = 2$, the magnification of this system is $m = 4$. For a specific field point, the identical tangential and sagittal ray fans indicate the absence of astigmatism. The curved ray fans for non-zero fields show the presence of the oblique spherical aberrations besides the 3rd order field curvature. Below we will determine the amount of each type of aberrations.

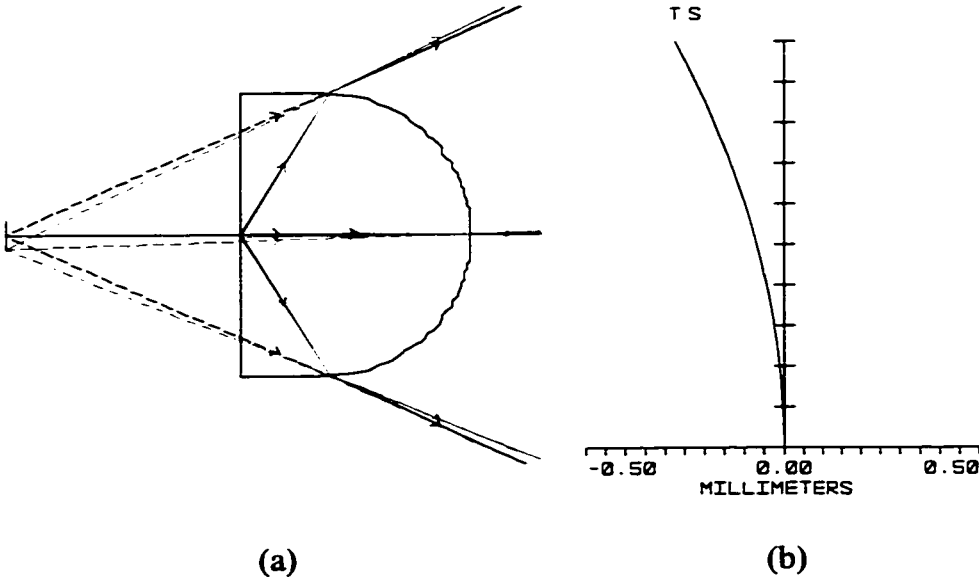


Figure 5.9. (a) The aplanatic conjugate pair of a sphere. The lens material has an index of refraction 2.0, the radius of curvature of the spherical surface is 100 mm. (b) The field curve of the system. The maximum field is 1 mm.

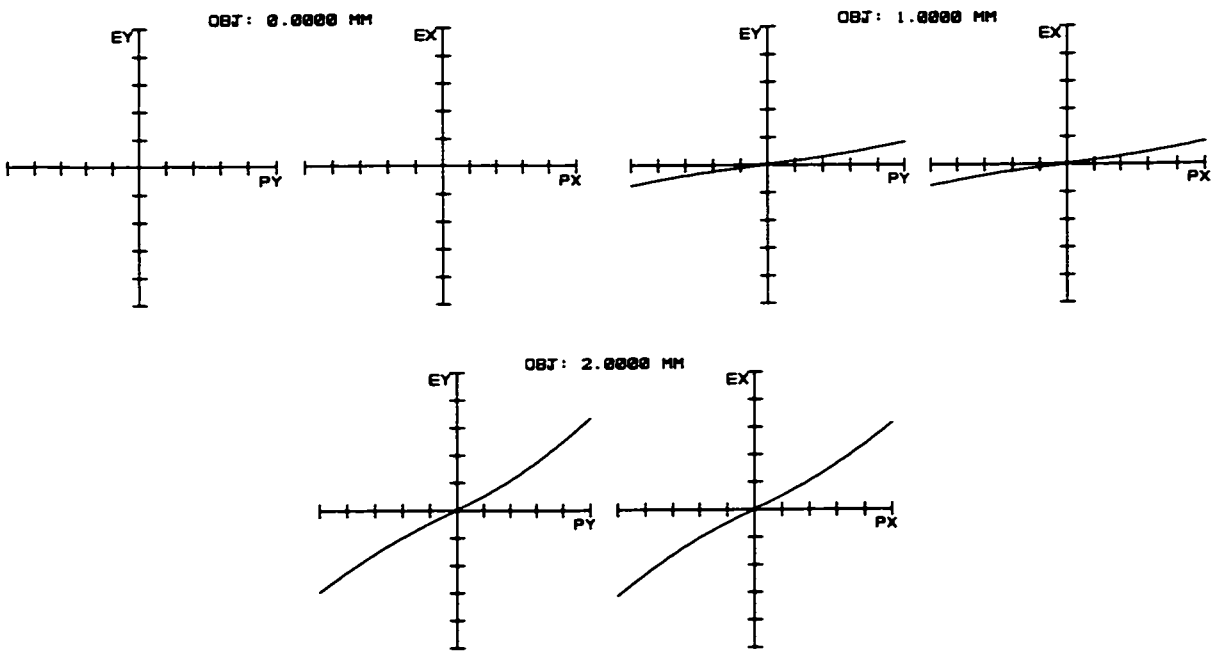


Figure 5.10. The ray fans of the system shown in Figure 5.9(a).

Table 5.5. The calculated s , $t/\cos^2(\theta)$ and the aberration coefficients for the system shown in Figure 5.9.

θ (degree)	s (mm)	$t/\cos^2(\theta)$ (mm)	$W_3(\rho) = -m^2/(2s)$ (m^{-1})	$P_1^2 W_2(\rho)$ $= m^2/(2s) - m^2 \cos^2(\theta)/(2t)$
0	-400	-400	20	$W_{200} = 20$ $W_{220} = 18.4$ $W_{240} = 63.3$ $W_{260} = 0.00$
3.415	-398.577	-398.577	20.07	0
6.701	-394.497	-394.497	20.28	0
9.751	-388.266	-388.266	20.6	0
12.496	-380.553	-380.553	21.02	0
14.907	-372.025	-372.025	21.5	0
16.987	-363.237	-363.237	22.02	0
18.759	-354.595	-354.595	22.56	0
20.259	-346.355	-346.355	23.1	0
21.523	-338.660	-338.659	23.62	0
22.589	-331.568	-331.568	24.13	0

Again t and s can be obtained by tracing a pair of closely spaced parallel rays in ZEMAX, we can then calculate $W_3(\rho)$ and $p_1^2 W_2(\rho) + W_3(\rho)$ using Eqs. (5.2). The results of ray tracing and calculation are listed in Table 5.5. With the values of t and s obtained by ray tracing in ZEMAX, we can then fit the calculated data and get

$$W_3(\rho) = 20 + 18.4\rho^2 + 63.3\rho^4, \quad (5.19a)$$

$$p_1^2 W_2(\rho) = 0. \quad (5.19b)$$

Applying Eqs. (5.4) and (5.5), we then obtain

$$W_{200} = 20 \text{ m}^{-1},$$

$$W_{220} = 18.4 \text{ m}^{-1},$$

$$W_{240} = 63.3 \text{ m}^{-1},$$

$$W_{2n2} = 0 \text{ (n is even).}$$

For a field point at $(x = 0.5\text{mm}, y = 0)$, and pupil coordinate $\sin\theta = 0.3535233$, using Eqs. (5.1a) and (5.1b) we calculate its ray aberration: $\Delta x = \Delta y = 6.05 \text{ }\mu\text{m}$.

Compared to the ZEMAX ray tracing result: $\Delta x = \Delta y = 6.22 \text{ }\mu\text{m}$, this is a very good agreement.

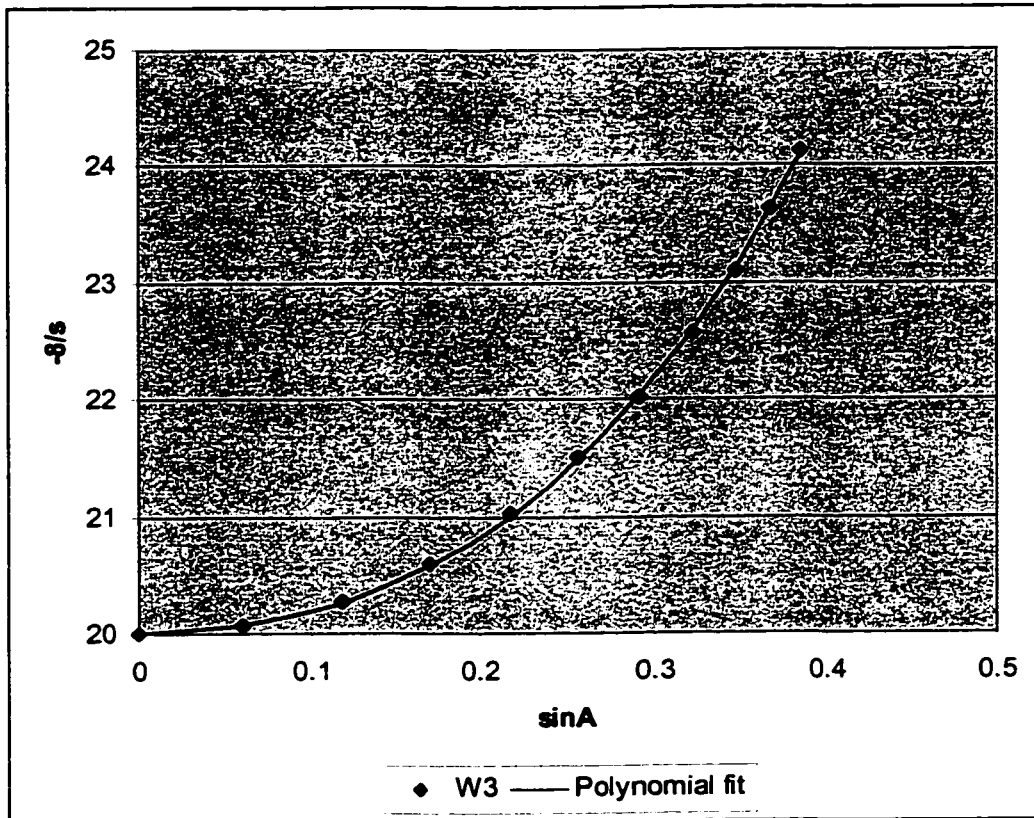


Figure 5.11. Plot of $W_3(\rho)$ vs. $\sin\theta$ where θ is the ray angle in image space. The dots are the real ray trace data listed in Table 5.5 and the solid line is the polynomial fit (5.19a).

5.7 CONCLUSION

For general cases where the Pupil Astigmatism Criteria in (2.50) are not satisfied, the quadratic field-dependent aberrations can be calculated from the values of t and s obtained by Coddington equations or ray tracing. The formulas for doing this are given first in this chapter. They can be used for determining the true quadratic field-dependent aberrations in the same way that Offense against Sine Condition (OSC) is used for predicting coma. This analysis was confirmed using several examples.

CHAPTER 6

IMPLEMENTING THE PUPIL ASTIGMATISM CRITERIA FOR DESIGN OF OPTICAL SYSTEM

6.1 INTRODUCTION

The pupil astigmatism criteria were derived in Chapter 2. In Chapter 3, we showed them to be consistent with and more general than the relationships between image and pupil Seidel aberrations. Chapter 4 demonstrated that fully corrected systems obey the pupil astigmatism conditions and Chapter 5 shows how offense against the pupil astigmatism conditions can be used to calculate field aberrations. All of this is interesting and adds insight to the behavior of optical systems, but is not directly applicable for making new systems. We now apply the Pupil Astigmatism Conditions to define a class of imaging systems that is fully corrected for quadratic field aberrations, for all points in the pupil. For correcting the quadratic field-dependent aberrations to make sense, the spherical aberration and linear coma must be corrected first, then the Constant OPL Condition and the Abbe Sine Condition must be satisfied.

6.2 THE CONSTANT OPL CONDITION AND THE SINE CONDITION

The Pupil Astigmatism Criteria in (2.50) can be used by an optical designer to correct the quadratic field-dependent aberrations. For such correction to make sense, the

field-independent aberrations (spherical) and linear field-dependent aberrations (coma) must be corrected first. Described in Table 1.1, the condition for correcting all the spherical aberrations are the Constant Optical Path Length (OPL) Condition and the condition for correcting all coma is the Abbe Sine condition. Figure 6.1 shows a finite conjugate system. The Constant OPL condition and the Sine condition for such systems are:

$$\text{OPL}[\text{OI}] = \text{constant (independent of } \theta \text{)}. \quad (6.1)$$

$$\frac{\sin(\alpha)}{\sin(\theta)} = \text{constant (independent of } \theta \text{)}. \quad (6.2)$$

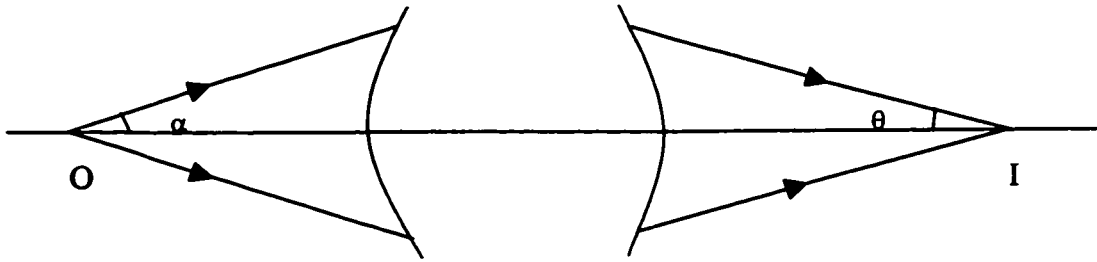


Figure 6.1. An optical system with object a finite distance away.

Figure 6.2 shows an infinite conjugate system, the Constant OPL condition and the Sine condition take different forms for such systems:

$$\text{OPL}[\text{AI}] = \text{constant (independent of } \theta \text{)}. \quad (6.3)$$

$$\frac{h}{\sin(\theta)} = \text{constant (independent of } \theta \text{)}. \quad (6.4)$$

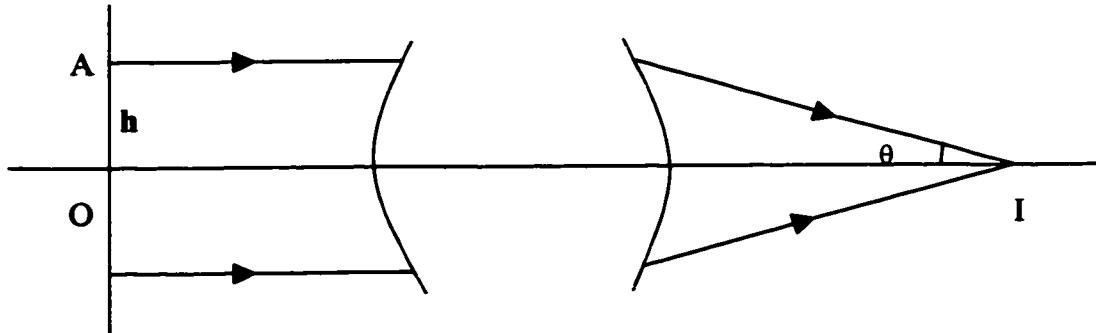
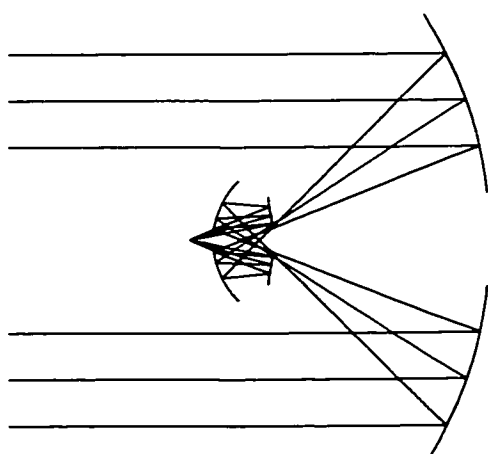
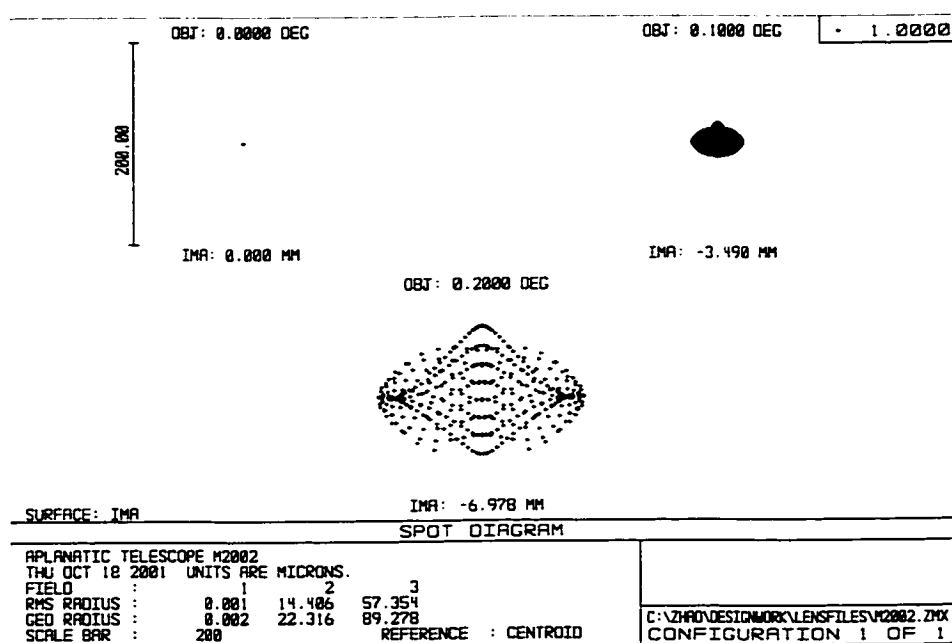


Figure 6.2. An optical system with object at infinity.

It has long been known that we can satisfy the Constant OPL Condition and the Abbe Sine Condition simultaneously to design the optical systems free of all orders of spherical aberrations and coma (Luneburg 1966). Such systems are said to be strictly aplanatic. Only two degrees of freedom are needed to design such systems. Each degree of freedom will result in a general aspheric surface. These systems can be normal incidence or grazing incidence systems since the Constant OPL Condition and the Sine Condition is universal, unlike the Seidel formulas which are not. Mertz (1996) proposed a novel method for designing such systems. Figure 6.3 and 6.4 show two examples designed using Mertz's method. Notice that the RMS spot size is proportional to the square of the field angle. Cubic splines are used to simulate each general aspheric surface generated by numerical method. For the grazing incidence system shown in Figure 6.4, the surface function is expressed as $r = f(z)$ instead of $z = f(r)$ to make it single-valued.

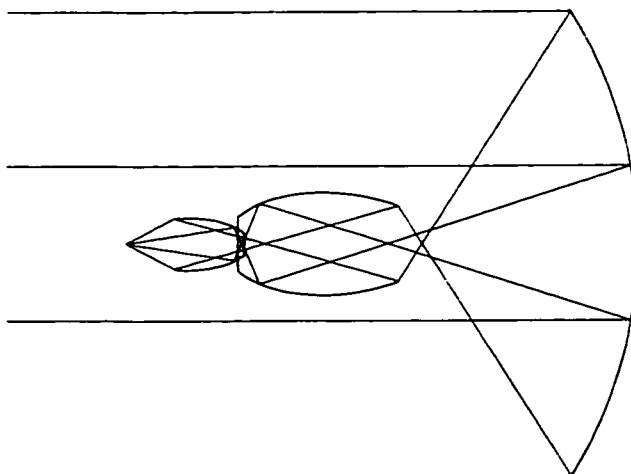


(a)

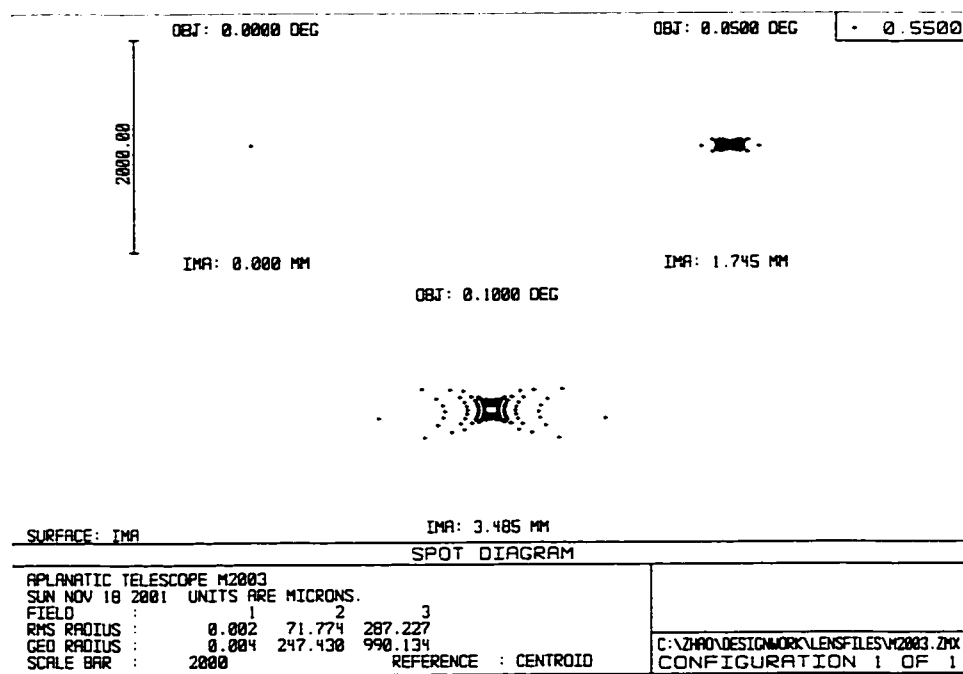


(b)

Figure 6.3. (a) A normal incidence 3-mirror aplanatic telescope. The primary mirror is spherical, and the secondary and tertiary are general aspheres. (b) The spot diagrams of the aplanatic telescope. Notice the residual aberration is primarily quadratic field-dependent.



(a)



(b)

Figure 6.4. (a) A grazing incidence 3-mirror aplanatic telescope. The primary mirror is spherical, and the secondary and tertiary are general aspheres. (b) The spot diagrams of the aplanatic telescope. Note the residual aberration is primarily quadratic field-dependent.

6.3 THE PUPIL ASIGMATISM CRITERIA FOR CORRECTING THE QUADRATIC FIELD-DEPENDENT ABERRATIONS

Since the Pupil Astigmatism Criteria (2.50) involve only the on-axis ray properties, in this regard it is just like the Constant OPL Condition and the Abbe Sine Condition. So we can implement them the same way as we implement Constant OPL Condition and Sine condition in the design of the perfectly aplanatic systems and design system free of certain or all types of quadratic field-dependent aberrations. Below we rewrite the criteria (2.50):

$$s(\theta') = \frac{t(\theta)}{\cos^2(\theta)} = \text{constant}, \quad (6.5a)$$

$$\frac{t(\theta)}{\cos^2(\theta)} = \text{constant}, \quad (6.5b)$$

$$s(\theta) = \text{constant}, \quad (6.5c)$$

$$s(\theta) = \frac{t(\theta)}{\cos^2(\theta)}, \quad (6.5d)$$

where t , s and θ are quantities in image space.

When conditions (6.1) and (6.2) (or (6.3) and (6.4) for infinite conjugate system) are combined with condition (6.5a), then all the aberrations of quadratic field dependence are corrected besides all spherical aberrations and coma. When one of the conditions of (6.5b)-(6.5c) are satisfied instead of (6.5a), then specific quadratic field-dependent aberrations are corrected.

Since thin bundle of rays are traced to obtain t and s , the Coddington equations are the proper tools to use to implement the Pupil Astigmatism Conditions (6.5a)-(6.5d).

6.4 THE CODDINGTON EQUATIONS

When a narrow beam of light is obliquely incident on a refracting surface, astigmatism is introduced. That is to say, the tangential rays and sagittal rays focus on different locations. The Coddington Equations can calculate the locations of the tangential and sagittal images (Kingslake 1978).

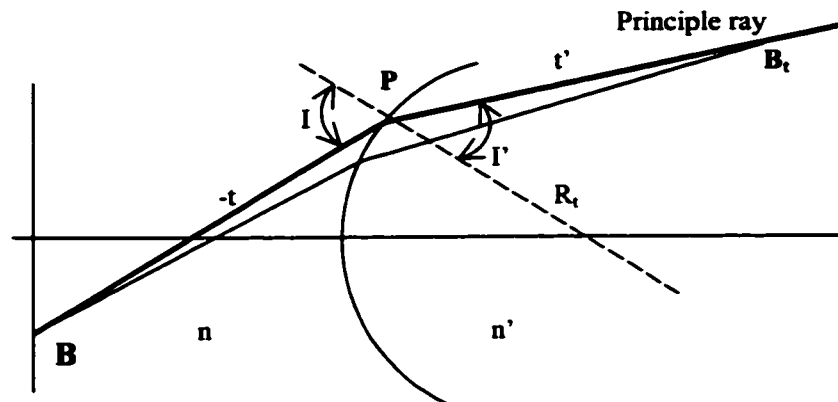


Figure 6.5. Illustration of the tangential rays of an oblique bundle of rays refracted by a surface.

In Figure 6.5, a thin bundle of rays originating from a point B are incident on a refractive surface. The principle ray hit the surface at P . The incidence angle of this ray is I and the refraction angle is I' . At Point P , the tangential radius of curvature of the surface is R_t and the sagittal radius of curvature of the surface is R_s . The index of refraction of the medium is n before the surface and n' after the surface. The physical

distance from point B to P is $-t$ ($t < 0$), and the distance from the tangential image point B_t to Point P is t' (> 0). Then the Coddington equation to determine t' from the given parameters is:

$$\frac{n' \cos^2 I'}{t'} - \frac{n \cos^2 I}{t} = \frac{n' \cos I' - n \cos I}{R_t}. \quad (6.6a)$$

Similarly, the Coddington equation below determines the location of the sagittal image:

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n' \cos I' - n \cos I}{R_s}, \quad (6.6b)$$

where s is the distance from Point P to the sagittal image (denoted as B_s), and $-s$ is the distance from B to P (therefore $s = t$).

6.5 DETERMINE R_t AND R_s FOR GENERAL ASPHERIC SURFACES

For a general aspheric surface, R_t and R_s are not equal. In our discussion, we are concerned with the rotationally symmetric surface only. Figure 6.6 shows a typical rotational symmetric surface. P is a point on the surface with coordinate (r, z) , and the first and second derivatives of the curve at P is z' and z'' . Line PA is the tangent and Line PC_2 is the normal at P . Because of the rotational symmetry, the sagittal radius of curvature at P is then

$$R_s = |PC_1| = r \frac{\sqrt{1 + z'^2}}{z'}, \quad (6.7)$$

where C_1 is the intersection of the surface normal at P and the optical axis.

And the tangential radius of curvature is

$$R_t = |PC_2| = \frac{\sqrt{1+z'^2}}{z''}, \quad (6.8)$$

where C_2 is the center of curvature of the surface segment around P .

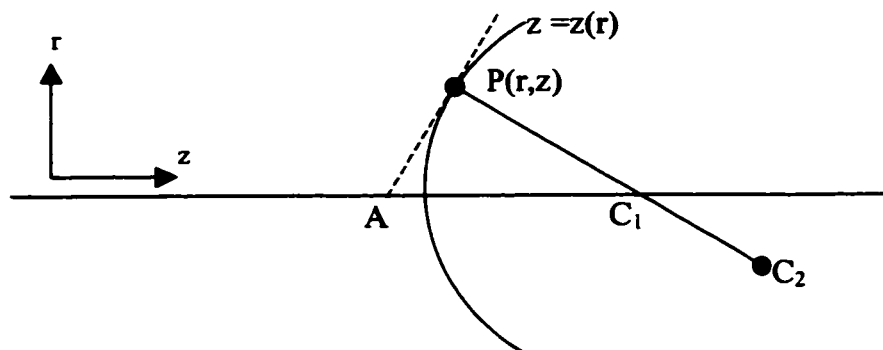


Figure 6.6. Illustrating how to obtain the tangential and sagittal radii of curvature of a general aspheric surface with rotational symmetry. C_1 is the sagittal center of curvature of the segment of the surface at P and C_2 is the tangential center of curvature of the same segment.

6.6 DETERMINE THE SURFACE SLOPE WHEN THE RAY DIRECTIONS BEFORE AND AFTER A SURFACE ARE KNOWN

When the indices of refraction and the ray directions before and after an optical surface are known, the surface normal (therefore the slope) can be calculated by the Snell's Law. In Figure 6.7, a ray with ray vector $n_i \hat{A}$ is incident on an optical surface, and the refracted ray has a ray vector $n_r \hat{B}$. The Snell's Law says

$$n_i \hat{A} \times \hat{n} = n_r \hat{B} \times \hat{n}, \quad (6.9)$$

where \hat{n} is the unit vector along the surface normal.

That is equivalent to

$$(n_i \hat{A} - n_r \hat{B}) \times \hat{n} = 0. \quad (6.10)$$

This equation indicates that the surface normal \hat{n} at where the ray hits is parallel to the vector $n_i \hat{A} - n_r \hat{B}$. Once the normal is known, it is trivial to obtain the surface slope. In the numerical method to design systems free of certain or all types of quadratic field-dependent aberrations, the slope of each surface is calculated this way when the ray directions are known first.

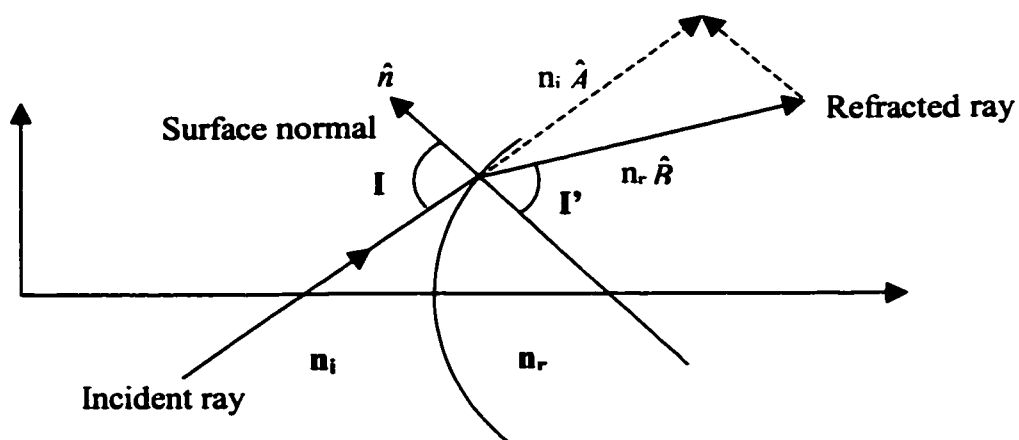


Figure 6.7. Illustration of the Snell's law.

6.7 NUMBER OF DEGREES OF FREEDOM

The Constant OPL Condition and the Sine Condition need two degrees of freedom to implement, and the Condition (6.5a) needs two additional degrees of freedom to be implemented, so a total of four degrees of freedom are needed to design a system with no spherical aberrations, no linear and quadratic field-dependent aberrations. Then a 4-surface system can be designed to correct all three types of aberrations. To maintain only one of the Pupil Astigmatism Conditions (6.5b)-(6.5d), a

three-surface system has enough degrees to correct all spherical aberrations, all linear field-dependent aberrations and some specific quadratic field-dependent aberrations.

6.8 DESCRIPTION OF THE ALGORITHM

The three- and four-surface solutions are made numerically by defining each surface directly, point by point. Since the relations used, the Constant OPL Condition, the Sine Condition, and the Pupil Astigmatism Conditions are defined using on-axis ray tracing, the surfaces are unique and easily defined. To simulate the performance of the designed systems in ZEMAX, cubic splines are used to represent each surface.

The Constant OPL Condition involves only physical distance, so it determines the coordinates of a point on each surface. While the Sine Condition involves the ray directions in object and image space, it is appropriate to say it determines the slope of a point on each surface. And the Pupil Astigmatism Conditions in (6.5) involve the paraxial image locations of a point on a marginal ray, so they determine the local power, therefore the second derivatives of a point on each surface. Each surface will be generated following the above analysis. Below I will describe the algorithm used to generate the surfaces in three- and four-surface systems.

6.8A 3-SURFACE SYSTEMS

If we first assume all the spacings, all the indices of refraction and the magnification of a 3-surface optical system, then we can generate each surface unambiguously. Figure 6.8 shows such a 3-surface system. O is the on-axis object point and I is its image. At the vertices of each surface, A_0 , B_0 , C_0 , the slope of the surface is 0.

The vertices are the starting points of the surface generation. Assume the optical path length from O to I is L , then we have

$$L = [OA_0] + [A_0B_0] + [B_0C_0] + [C_0I]. \quad (6.11)$$

Any the optical path length along any other ray from O to I must also be equal to L .

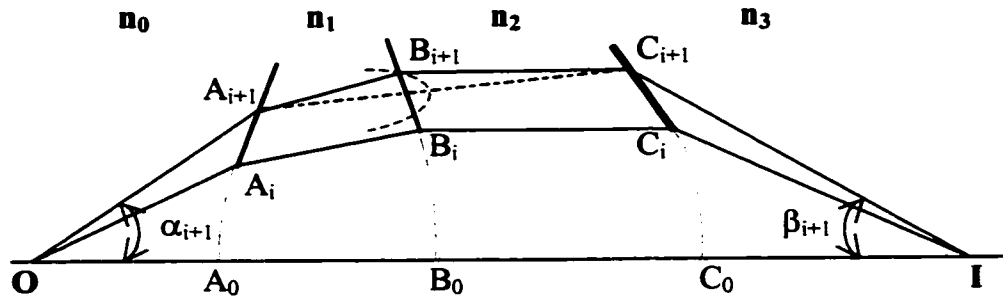


Figure 6.8. Illustration of the design procedure of the 3-surface system that is corrected for all spherical aberration, coma and some specific quadratic field-dependent aberrations.

Assume the i^{th} point of each surface has been generated, i.e. its coordinate and slope are known. Then the $(i+1)^{\text{th}}$ point of each surface should be on the tangent at the i^{th} point. We start with a ray from Point O . The ray has an angle α_{i+1} with the optical axis in object space. This ray will go through Point I in image space and has an angle β_{i+1} with the optical axis. Since the Sine condition needs to be satisfied as well, then β_{i+1} is known,

$$n_3 \sin \beta_{i+1} = \frac{n_0 \sin \alpha_{i+1}}{|m|}, \quad (6.12)$$

where m is the magnification of the system. Then the coordinates of A_{i+1} and C_{i+1} can be calculated because they are the ray intersections with the first and third surface. The last

step is to figure out the location of B_{i+1} and the surface slope at A_{i+1} , B_{i+1} and C_{i+1} . To ensure that the Constant OPL Condition is satisfied, optical path length $[A_{i+1}C_{i+1}]$ must be a fixed number, i.e.

$$[A_{i+1}C_{i+1}] = L - [OA_{i+1}] - [C_{i+1}I]. \quad (6.13)$$

Then we know B_{i+1} must be on the Cartesian oval that images Point A_{i+1} on Point C_{i+1} perfectly. If the second surface is reflective, then B_{i+1} must be on an elliptical surface. The dashed curve around B_{i+1} in Figure 6.8 is the Cartesian oval. When we move B_{i+1} along the Cartesian oval, the slopes at A_{i+1} , B_{i+1} and C_{i+1} change as the ray direction change and can be calculated using the method described in Section 6.6. Then the second derivative at A_{i+1} , B_{i+1} , C_{i+1} can be calculated since the coordinates of and the slopes at A_i , B_i , C_i are already known. Then we can use the Coddington Equations (6.6a) or (6.6b) to determine the image location of the point at infinity on the ray (see Figure 6.5). We can make the tangential or sagittal image or both at the desired location so that one of the Pupil Astigmatism Criteria (6.5b)-(6.5d) is satisfied. Then the coordinate of B_{i+1} is uniquely determined and the surface slope at each point can be calculated. The same procedure is then followed to determine A_{i+2} , B_{i+2} and C_{i+2} . In this way the entire system is generated.

6.8B 4-SURFACE SYSTEMS

For a 4-surface system, there are 4 conditions to be satisfied. Figure 6.9 shows such a 4-surface system. If the coordinates of A_i , B_i , C_i and D_i and slopes of each surface at these points are known already, we can generate the next points A_{i+1} , B_{i+1} , C_{i+1} and D_{i+1} . Just like in the design of 3-surface system, A_{i+1} and D_{i+1} are easily determined

using the Abbe Sine condition. Once C_{i+1} is chosen, following the same procedure described in 6.8A, we can determine B_{i+1} while satisfying the Constant OPL condition plus one condition in (6.5a). For the other condition in (6.5a) to be satisfied, we can move C_{i+1} along the tangent at C_i . The same procedure is then followed to determine A_{i+2} , B_{i+2} , C_{i+2} and D_{i+2} . In this way the entire system is generated.

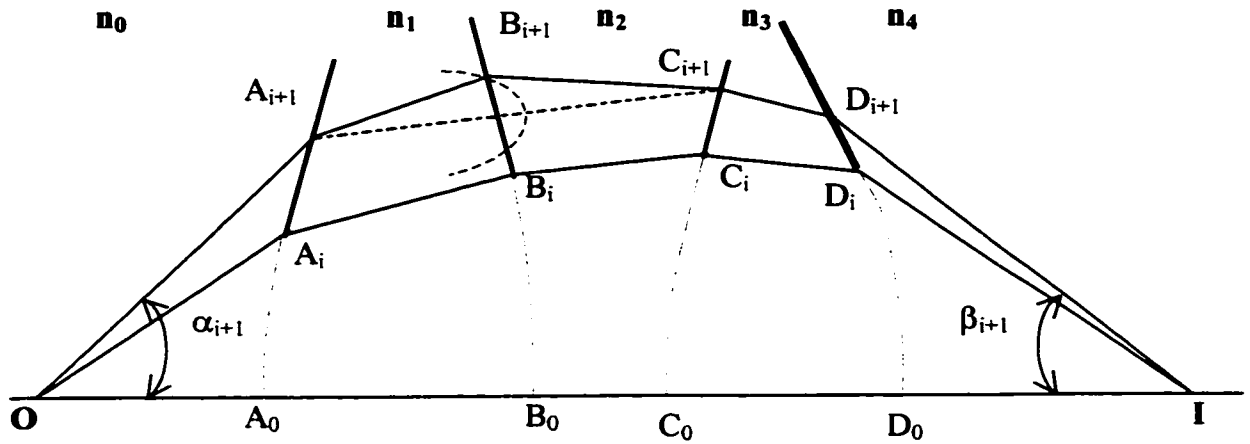


Figure 6.9. Illustration of the design procedure of the 4-surface system that is corrected for all spherical aberration, coma and quadratic field-dependent aberrations.

6.9 EXAMPLE DESIGNS

By using the algorithm described in Section 6.8, three 3-surface systems and three 4-surface systems are designed. The prescription data of these systems are summarized in Table 6.1 and Table 6.2. Figure (6.10)-(6.12) show the 3-surface systems. Figure (6.13) shows a 4-surface refractive system with object a finite distance away. The system in Figure (6.14) is a refractive 4-surface system with object at infinity. The system in Figure (6.15) is a 4-mirror system with object at infinity. The ray fans of this system is shown in Figure 6.16. The spot diagrams of the 3-surface systems show

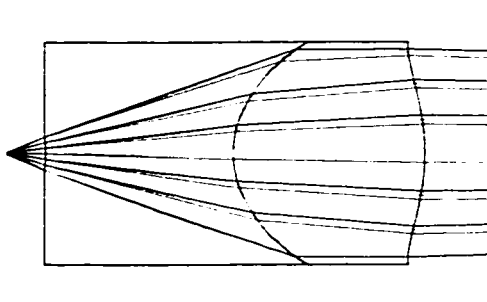
that the aberrations we intended to correct are indeed corrected, and the spot diagrams of the 4-surface systems show that the residual aberrations are third order in field.

Table 6.1. The prescription data of the 3-surface systems.

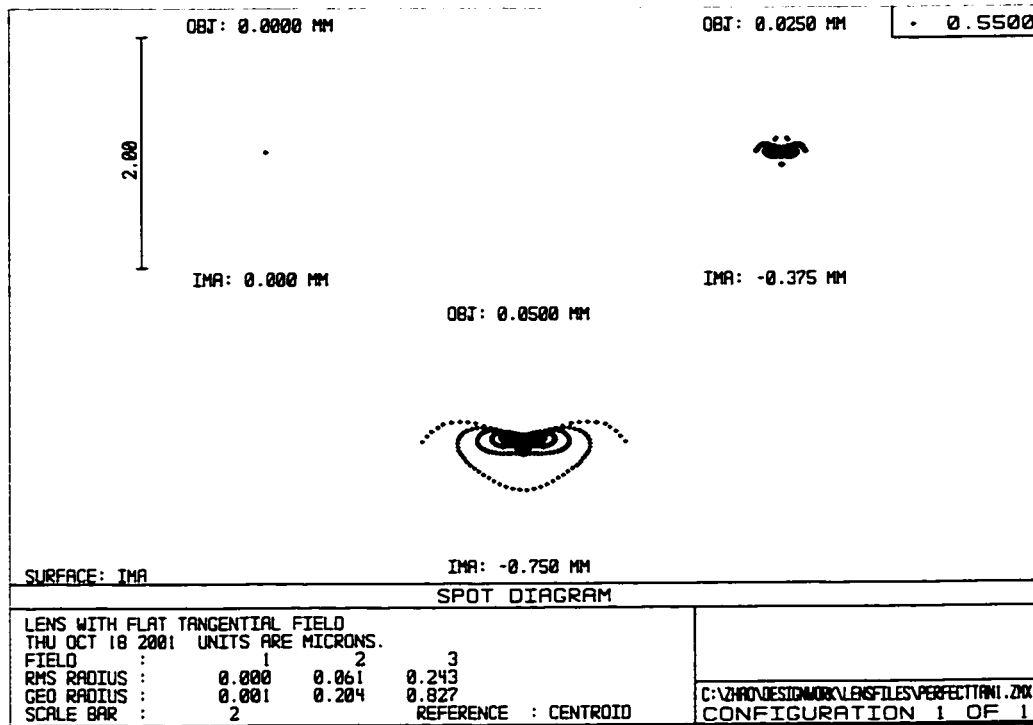
Optical System	3-surface systems (shown in Figure 6.10 – 6.12.)	
Magnification m	-15	
Surface	Thickness (mm)	Index of refraction
Object Plane – 1 st surface	1	1
1 st – 2 nd surface	5	1.5
2 nd – 3 rd surface	5	1.8
3 rd surface – image plane	100	1

Table 6.2. The prescription data of the 4-surface systems.

Optical System	4-surface system (shown in Figure 6.13)		4-surface system (shown in Figure 6.14)		4-mirror system (shown in Figure 6.15)	
Magnification m or Focal Length f	$m = -15$		$f = 18.16539$ mm		$f = 177.9$ mm	
Surface	Thickness (mm)	Index of refraction	Thickness (mm)	Index of refraction	Thickness (mm)	Index of refraction
Object Plane – 1 st surface	1	1	infinity	1	infinity	1
1 st – 2 nd surface	5	1.5	5	1.5	-50	-1
2 nd – 3 rd surface	5	1	5	1	60	1
3 rd – 4 th surface	5	1.5	5	1.5	-20	-1
4 th surface – image plane	100	1	23.233582	1	32.296034	1

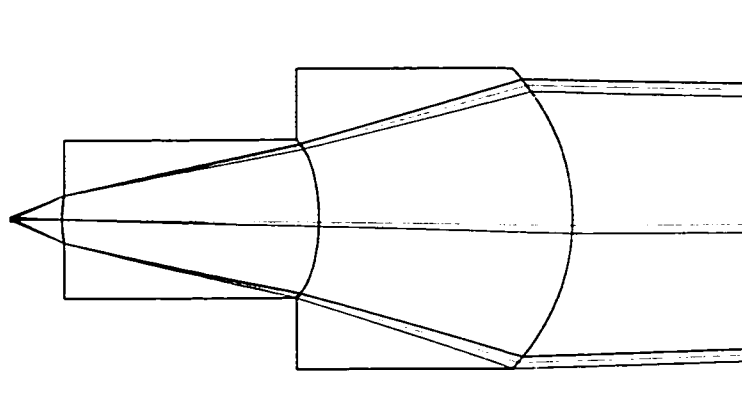


(a)

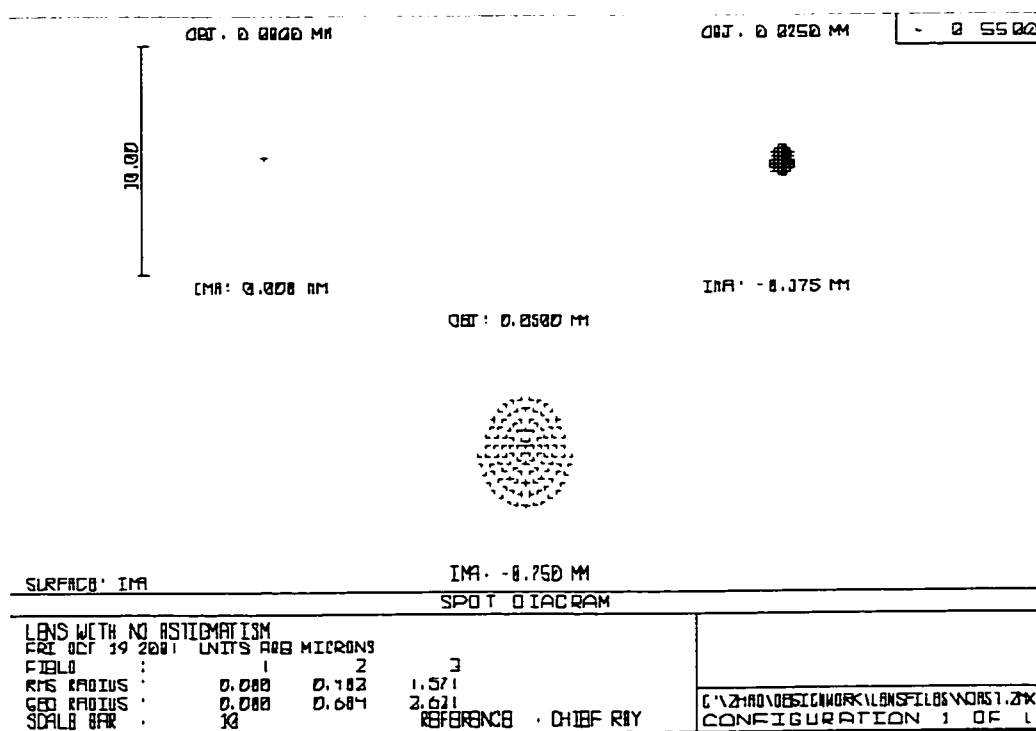


(b)

Figure 6.10. (a) A 3-surface system with zero quadratic field-dependent aberrations in tangential plane. All three surfaces are general aspheres. (b) The spot diagrams of the system.

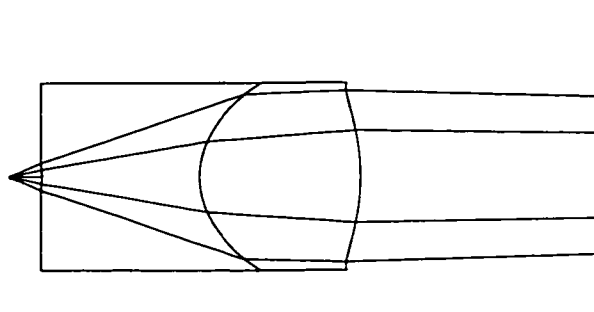


(a)

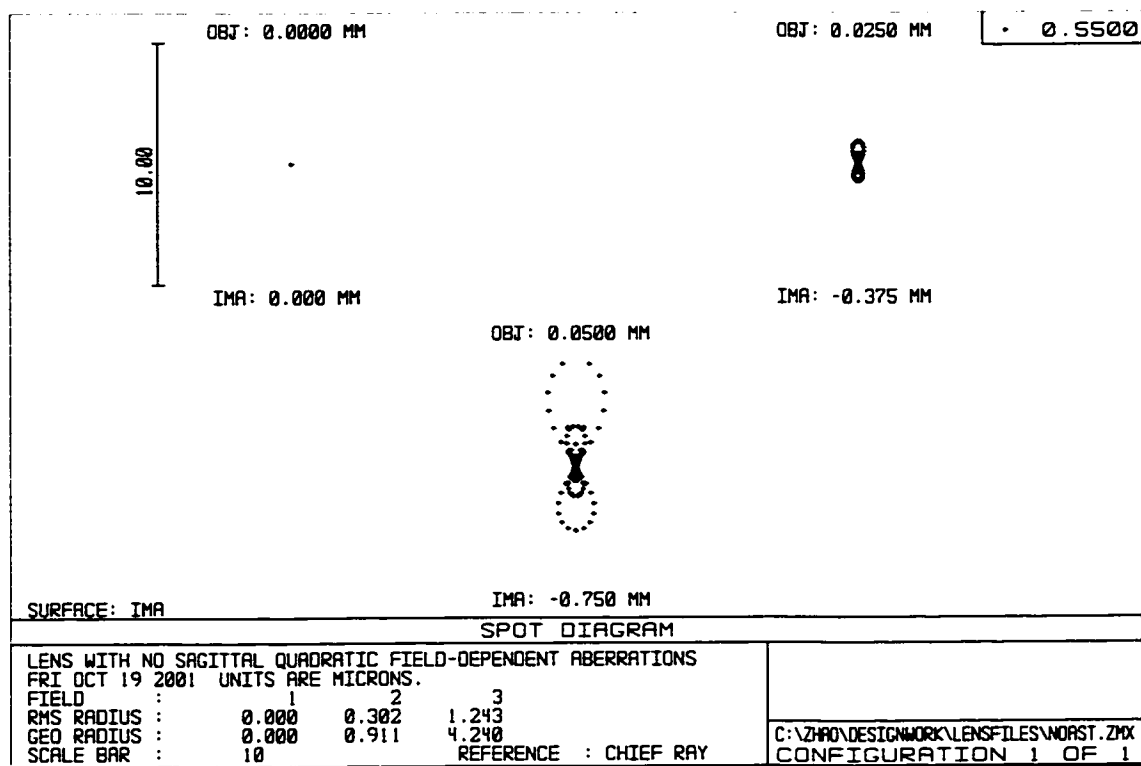


(b)

Figure 6.11. (a) A 3-surface system with zero quadratic field-dependent astigmatism. All three surfaces are general aspheres. (b) The spot diagrams of the system.

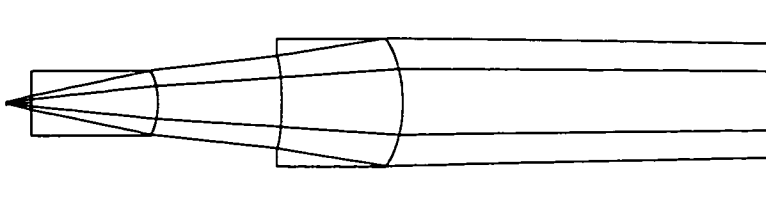


(a)

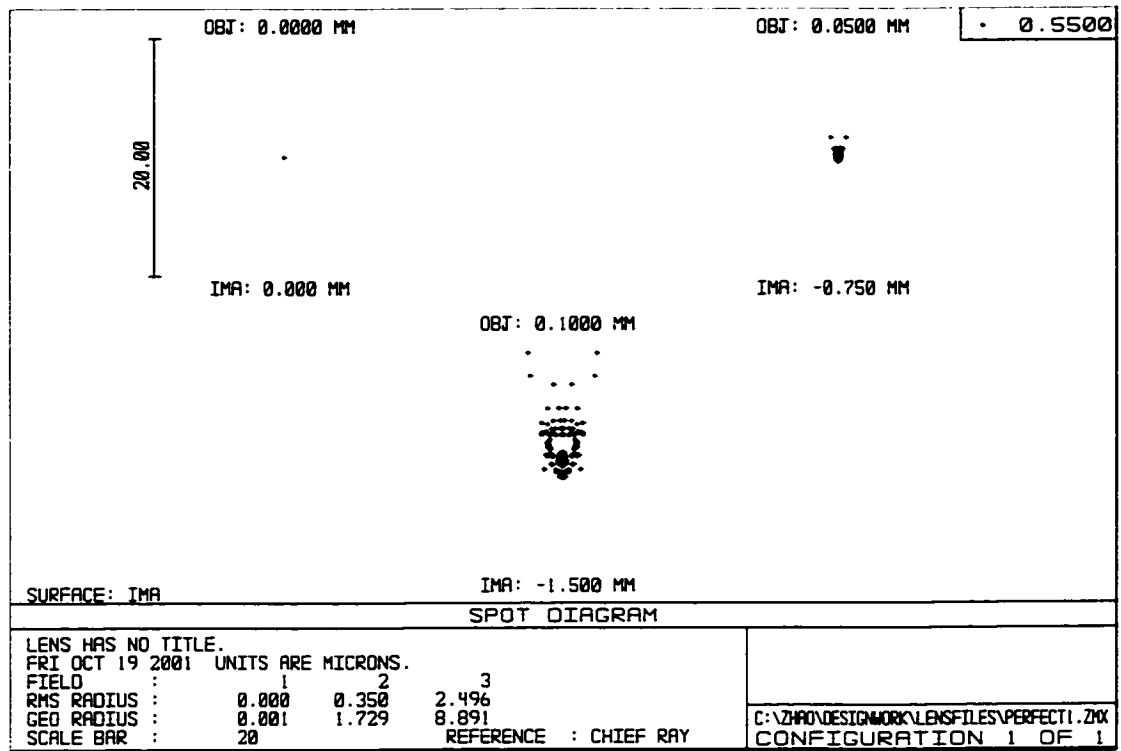


(b)

Figure 6.12. (a) A 3-surface system with zero sagittal quadratic field-dependent aberrations. All three surfaces are general aspheres. (b) The spot diagrams of the system.

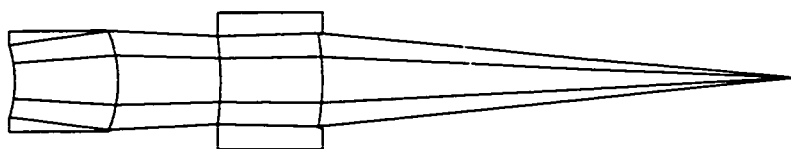


(a)

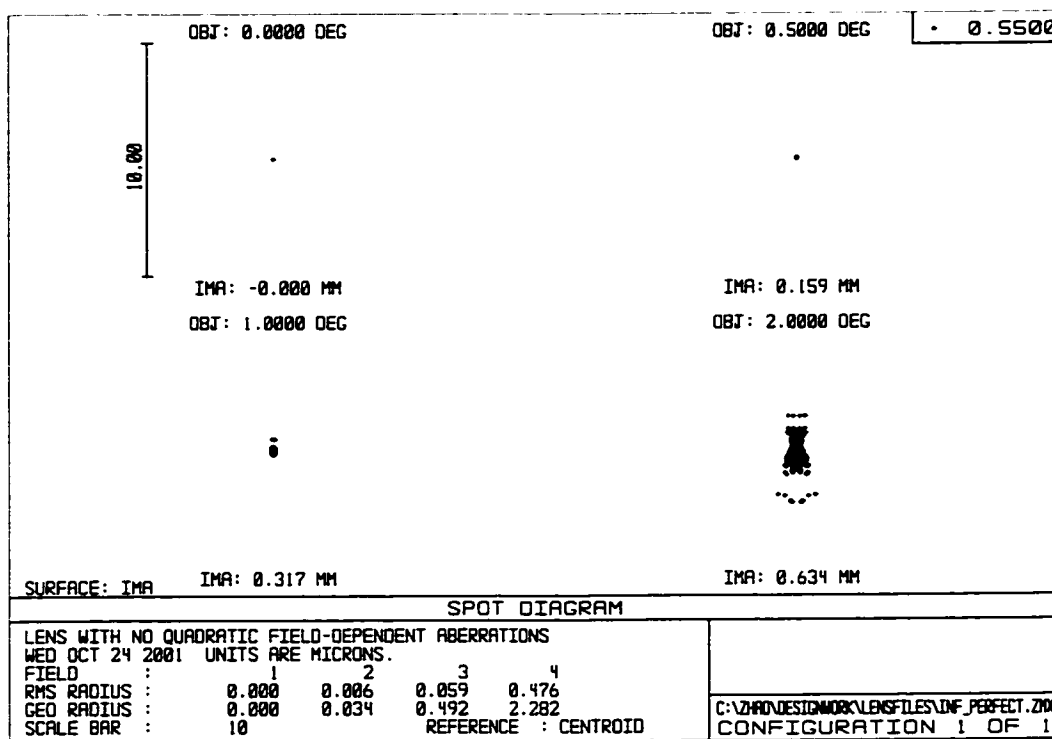


(b)

Figure 6.13. (a) A 4-surface system with no quadratic field-dependent aberrations. All four surfaces are general aspheres. (b) The spot diagrams of the system.

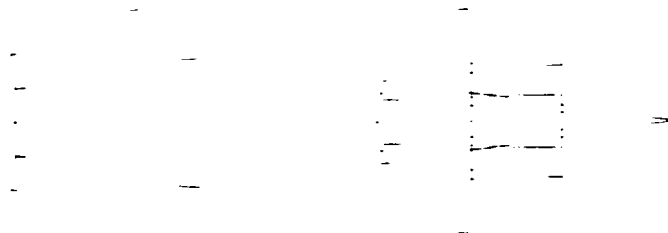


(a)

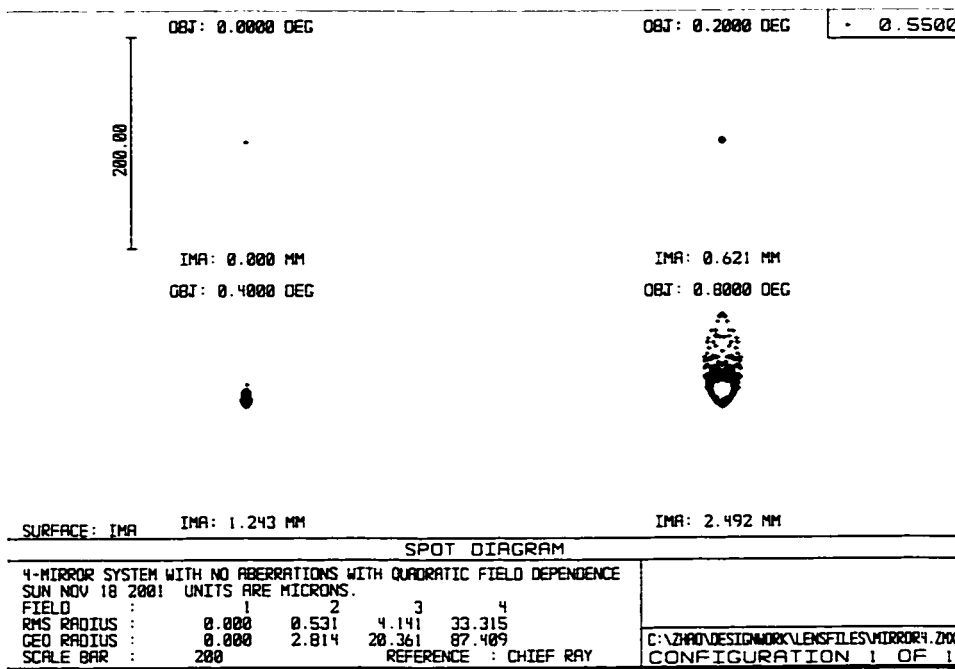


(b)

Figure 6.14. (a) A 4-surface system with no quadratic field-dependent aberrations. The object is at the infinity. All four surfaces are general aspheres. (b) The spot diagrams of the system.



(a)



(b)

Figure 6.15. (a) A 4-mirror system with no quadratic field-dependent aberrations. The object is at infinity. All four surfaces are general aspheres. (b) The spot diagrams of the system.

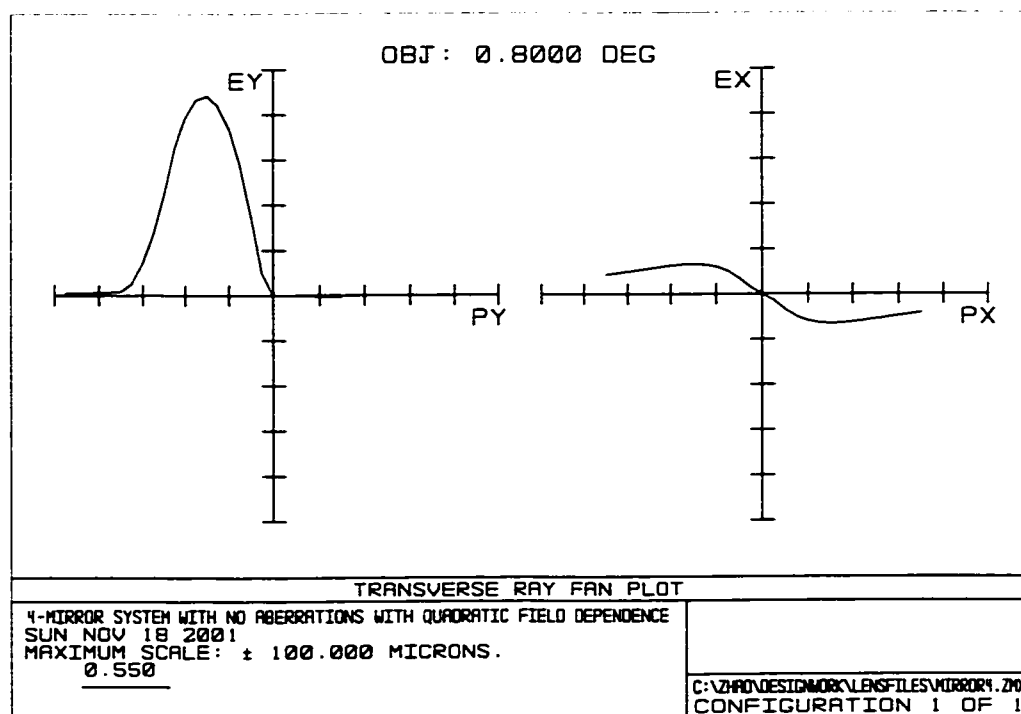


Figure 6.16. The ray fans of the 4-mirror system shown in Figure 6.15.

In designing the above systems, one problem often encountered was the numerical generation could not go on because the solution to one of the pupil astigmatism conditions could not be found. We believe the algorithm described in Section 6.8 is correct, the problem may lie in its numerical realization in the computer program. We were unable to debug the program. Nevertheless, the designs we present here do show the usefulness of the Pupil Astigmatism Conditions in optical design. Finding out what causes the failure of the program will continue to be my task.

6.10 CONCLUSION

The fact that the Pupil Astigmatism Conditions presented in Chapter 2 involve only the properties of the rays originating from the on-axis object point makes it very

convenient to implement them precisely in designing an optical system just like the Constant OPL Condition and the Abbe Sine Condition. In this chapter, the necessary tools in implementing the conditions are introduced, the algorithm used to implement the conditions is described and some examples designed are presented. The actual computer programs for designing these systems are listed in Appendix A. The programs for simulating the surface made up of the generated discrete points in ZEMAX are listed in Appendix B.

CHAPTER 7

THE PUPIL ASTIGMATISM CONDITIONS FOR PLANE-SYMMETRIC SYSTEMS

7.1 INTRODUCTION

The derivation of the pupil astigmatism conditions is general, and does not rely on symmetry of the optical systems. The implementation of the conditions has so far been limited to axisymmetric systems, as they are most common and the symmetry can be exploited to make the analysis easy. We now develop the equivalent criteria for plane symmetric systems. It is the application to these systems that the Pupil Astigmatism Condition may have the most value for guiding system design.

7.2 THE PLANE SYMMETRIC SYSTEM

There is a plane of symmetry in a plane symmetric system, which means one half of the system is the mirror image of the other. Axially symmetric system is a special type of the more general plane symmetric systems. Both unobstructed reflective telescopes and Extreme Ultra-Violet Lithography systems are plane symmetric. As more and more plane symmetric systems need to be designed while no well-accepted method exists to assist such design, knowing basic theories on how to correct aberrations for

such systems are more and more important in modern optical design. Figure 7.1 is one example of plane symmetric system.

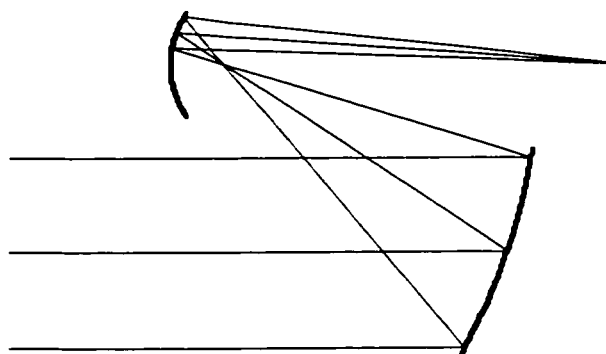


Figure 7.1. Unobstructed telescope: an example of plane-symmetric system.

In an axially symmetric system, the optical axis is always chosen as an axis of the coordinate system. The object plane and image plane is perpendicular to the optical axis. In a plane symmetric system, there is no obvious choice of axes. In literature, a special ray, which originates from the field center and goes through the center of the stop, is chosen as the optical axis ray (OAR). The OAR plays the same role as the optical axis in the rotationally symmetric system: it becomes an axis of the coordinate system referencing the variables describing the system. In this approach to extend the Hamiltonian treatment to the plane-symmetric system, I did not choose the OAR as an axis of the coordinate system. Instead, the intersection line of the plane of symmetry and the object plane is chosen to be y_0 -axis of the coordinate system $x_0y_0z_0$ in the object plane, the x_0 -axis is perpendicular to the plane of symmetry and the z_0 -axis is in the

plane of symmetry. The coordinate system $x_1y_1z_1$ in the image space is defined similarly: y_1 -axis is along the intersection line of the plane of symmetry and the object plane, the x_1 -axis is perpendicular to the plane of symmetry and the z_1 -axis is in the plane of symmetry. The direction cosines in the object space and image space are relative to the corresponding coordinate systems. Figure 7.2 shows the definition of the coordinate systems for a plane symmetric optical system.

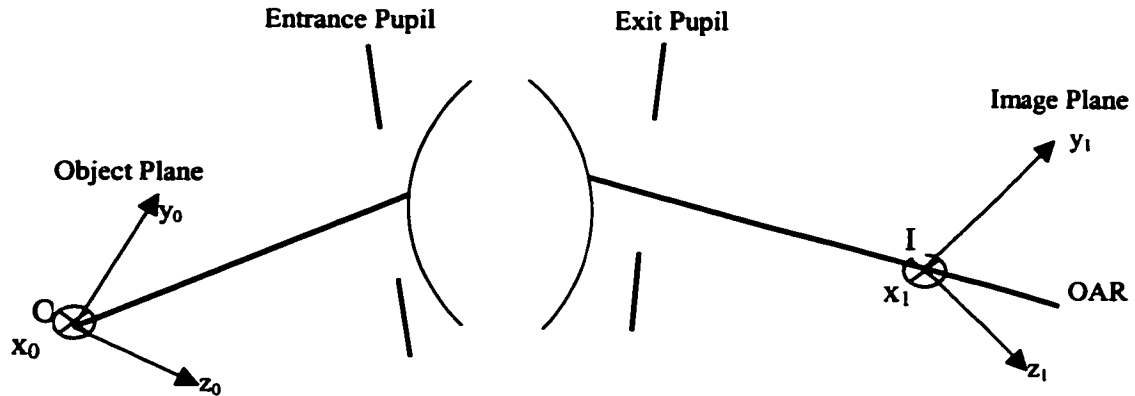


Figure 7.2. Illustration of the definition of coordinate systems for a plane-symmetric system. The ray that originates from the center of the field and goes through the center of pupil is called Optical Axis Ray (OAR). The object plane is x_0 - y_0 , and the image plane is x_1 - y_1 .

7.3 THE HAMILTON'S CHARACTERISTIC FUNCTIONS FOR PLANE SYMMETRIC SYSTEMS

In dealing with the plane symmetric systems, I still use the Hamilton's Characteristic functions: mixed characteristic for finite conjugate systems and angle characteristic for infinite conjugate systems. The terms in the Taylor series expansion of

the characteristic functions are now different than those for the axially symmetric systems.

For a system with object a finite distance away, we denote a field point in the object plane as \vec{h} , and the ray vector in the image space as $\vec{\rho}$. Now we define a unit vector \vec{i} , which is in the plane of symmetry. Then we make following definitions:

$$\vec{h} \cdot \vec{h} = h_{x_0}^2 + h_{y_0}^2,$$

$$\vec{\rho} \cdot \vec{\rho} = \rho_{x_1}^2 + \rho_{y_1}^2,$$

$$\vec{h} \cdot \vec{\rho} = h_{x_0} \rho_{x_1} + h_{y_0} \rho_{y_1},$$

$$\vec{h} \cdot \vec{i} = h_{y_0},$$

$$\vec{\rho} \cdot \vec{i} = \rho_{y_1},$$

where h_{x_0} and h_{y_0} are the x_0 and y_0 components of \vec{h} , and ρ_{x_1} and ρ_{y_1} are the x_1 and y_1 components of $\vec{\rho}$. Since the mixed characteristic function is a scalar, and it must be invariant when the system is reversed about the plane of symmetry, it must depend solely on the dot products of the three vectors: \vec{h} , $\vec{\rho}$ and \vec{i} . Now a term in the Taylor series expansion of the mixed characteristic function takes the following form:

$$W_{2k+n+p, 2m+n+q, n, p, q} = (\vec{h} \cdot \vec{h})^k (\vec{\rho} \cdot \vec{\rho})^m (\vec{h} \cdot \vec{\rho})^n (\vec{i} \cdot \vec{h})^p (\vec{i} \cdot \vec{\rho})^q. \quad (7.1)$$

Now we expand the mixed characteristic in Taylor series:

$$\begin{aligned}
W(\vec{h}, \vec{\rho}) = & W_0(\vec{\rho} \cdot \vec{\rho}, \vec{i} \cdot \vec{\rho}) \\
& + (\vec{h} \cdot \vec{\rho})W_{1a}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) + (\vec{h} \cdot \vec{i})W_{1b}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) \\
& + (\vec{h} \cdot \vec{h})W_{2a}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) \\
& + (\vec{h} \cdot \vec{\rho})^2 W_{2b}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) \\
& + (\vec{h} \cdot \vec{i})^2 W_{2c}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) \\
& + (\vec{h} \cdot \vec{i})(\vec{h} \cdot \vec{\rho})W_{2d}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) \\
& + \dots
\end{aligned} \tag{7.2}$$

The first term in Eq. (7.2), W_0 , does not depend on \vec{h} . If it is a constant, the center field point is imaged stigmatically. If it is not a constant, there must be aberrations that are independent of field size. The second and third terms determine the magnification and linear field-dependent aberrations. The 4th-7th terms determines the aberrations that are quadratic in field.

For a system with object at infinity, we use the Hamilton's angle characteristic function instead. Replace the field vector \vec{h} in Eq. (7.2) with the ray vector $\vec{\rho}_0$ in object space, we then obtain the angle characteristic.

7.4 THE PUPIL ASTIGMATISM CONDITION FOR PLANE SYMMETRIC SYSTEMS

By using Eq. (1.10a), we can calculate the ray vector in object space for $\vec{h} = 0$ from Eq. (7.2):

$$\begin{aligned}
\vec{\rho}_0 = & -\nabla_{\vec{h}} W(\vec{h}, \vec{\rho}) \\
= & -\vec{\rho}W_{1a}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}) - \vec{i}W_{1b}(\vec{\rho} \cdot \vec{\rho}, \vec{\rho} \cdot \vec{i}).
\end{aligned} \tag{7.3}$$

If the system is corrected for field-independent aberrations, then

$$W_0(\bar{\rho} \cdot \bar{\rho}, \bar{i} \cdot \bar{\rho}) = \text{constant},$$

and the ray intercept at the image plane is

$$\begin{aligned} \bar{h}_i &= -\nabla_{\bar{\rho}} W(\bar{h}, \bar{\rho}) \\ &= -\bar{h} W_{1a}(\bar{\rho} \cdot \bar{\rho}, \bar{\rho} \cdot \bar{i}) - (\bar{h} \cdot \bar{\rho}) \nabla_{\bar{\rho}} W_{1a}(\bar{\rho} \cdot \bar{\rho}, \bar{\rho} \cdot \bar{i}) - (\bar{h} \cdot \bar{i}) \nabla_{\bar{\rho}} W_{1b}(\bar{\rho} \cdot \bar{\rho}, \bar{\rho} \cdot \bar{i}). \end{aligned} \quad (7.4)$$

For the system to form a perfect image of the infinitesimal area around the center of field, W_{1a} and W_{1b} must be constant, i.e.

$$W_{1a} = -m = \text{constant}, \text{ and } W_{1b} = -a = \text{constant}, \quad (7.5)$$

where m is the system magnification.

Then Eq. (7.3) becomes

$$\bar{\rho}_0 = m\bar{\rho} + a\bar{i}. \quad (7.6)$$

And it is equivalent to

$$\rho_{x_0} = m\rho_{x_1} \text{ and } \rho_{y_0} = m\rho_{y_1} + a, \quad (7.7)$$

where ρ_{x_0} and ρ_{y_0} are the x_0 and y_0 components of $\bar{\rho}_0$. This is the general Sine Condition for plane symmetric systems. As the axially symmetric system is a special type of plane symmetric system, the Sine condition of the axially symmetric system (2.11) is a special form of Eq. (7.7) when $a=0$.

We derived the general Sine condition for the plane symmetric system above.

Once this condition is fulfilled, then there is no linear field-dependent aberrations

present in the system which include anamorphism: $(\bar{i} \cdot \bar{h})(\bar{i} \cdot \bar{\rho})$, linear

astigmatism: $(\bar{i} \cdot \bar{\rho})(\bar{h} \cdot \bar{\rho})$, etc.

We now proceed to correct the quadratic field-dependent aberrations using the powerful Hamiltonian characteristic functions. It is pointed out before that the last four terms of Eq. (7.2) represent the aberrations that are quadratic in field. For the aberrations to be corrected, each term must be independent of ρ , then the following statement must be true:

$$W_{2a} = \text{constant } c_1,$$

$$W_{2b} = W_{2d} = 0,$$

$$W_{2c} = \text{constant } c_2.$$

To find out what the physical meaning of the coefficients W_{2a} , W_{2b} , W_{2c} and W_{2d} , We will follow the same procedure as the derivation of the Pupil Astigmatism Criteria (2.50) in Chapter 2.

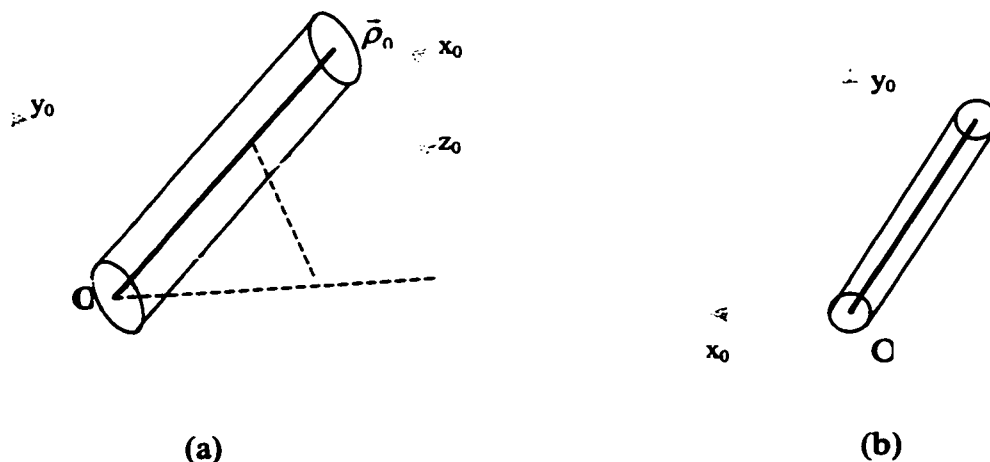


Figure 7.3. Trace a thin parallel bundle of rays centered at a ray from the field center with ray vector $\vec{\rho}_0$ in object space. (a) 3-D illustration of the bundle of rays, and (b) looking down the z_0 axis.

Assume the center of field is imaged perfectly, then consider a ray from the field center with the ray vector $\vec{\rho}$ in image space and $\vec{\rho}_0$ in object space, and trace a thin bundle of parallel rays centered on this ray through the system. This is equivalent to looking at the image of an entrance pupil at infinity. Figure 7.3 illustrates the bundle of rays being traced. The ray that originates from a field point $\vec{h} = (h_{x_0}, 0, 0)$ with ray vector $\vec{\rho}_0$ intersects the image plane at $\vec{h}_1 = (h_{x_1}, 0, 0)$. We now call this ray the sagittal ray of the parallel bundle. Similarly, the ray that originates from $\vec{h} = (0, h_{y_0}, 0)$ with the ray vector $\vec{\rho}_0$ is called the tangential ray of the bundle. If the generalized Sine Condition (7.7) is satisfied, then $h_{x_1} = mh_{x_0}$ for any $\vec{\rho}_0$ when h_{x_0} is infinitesimal. If this ray intersects the center ray of the bundle with ray vector $\vec{\rho}$ in image space at the point T_s , then let $IT_s = t_s$ and the angle between the two rays be $\Delta\theta$. Figure 7.4. shows the center and sagittal rays in image space. In analogy to the derivation of criteria (2.50) in Chapter 2, the mixed characteristic W_s for the sagittal ray is:

$$\begin{aligned} W_s(h_{x_0}, \vec{\rho}) &= W(0, \vec{\rho}) - h_{x_0} \rho_{x_0} - n_1 t_s + n_1 t_s \cos(\Delta\theta) \\ &\cong W(0, \vec{\rho}) - h_{x_0} \rho_{x_0} - n_1 t_s \frac{(\Delta\theta)^2}{2}, \end{aligned} \quad (7.8)$$

where n_1 is the refractive index of the medium in image space, and ρ_{x_0} is the x_0 component of ray vector $\vec{\rho}_0$.

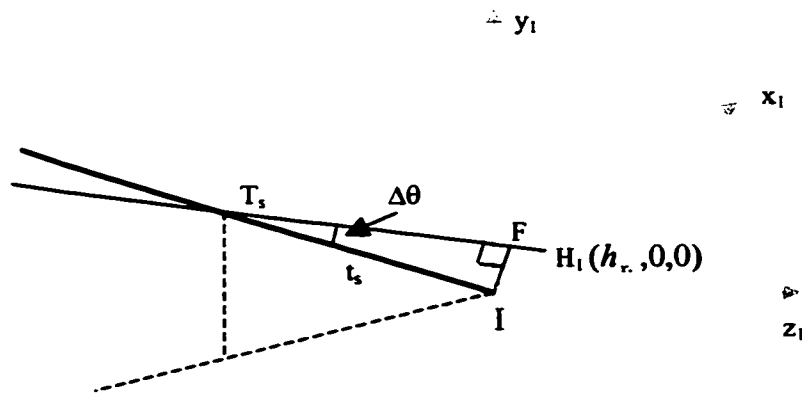


Figure 7.4. 3-D illustration of the sagittal ray of the parallel bundle in image space. The ray intersects the image plane at $(h_x, 0, 0)$. The angle between the sagittal ray and the center ray of the bundle is $\Delta\theta$.

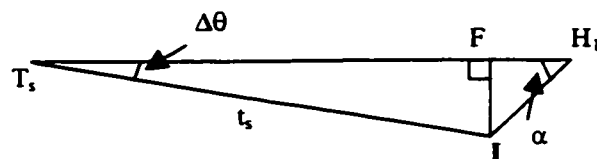


Figure 7.5. Blow out of the triangle T_sIH_1 in Figure 7.4.

$\Delta\theta$ in Eq. (7.8) can be determined by the following equation (refer to Figure 7.5):

$$\Delta\theta = \frac{|IF|}{|t_s|} = \frac{mh_{x_0}}{t_s} \frac{|\vec{\rho} \times \hat{x}_1|}{n_1}. \quad (7.9)$$

Substitute (7.9) in (7.8), we obtain

$$W(h_{x_0}, \bar{\rho}) \cong W(0, \bar{\rho}) - h_{x_0} \rho_{x_0} - \frac{m^2 |\bar{\rho} \times \hat{x}_1|^2}{2n_1 t_s} h_{x_0}^2. \quad (7.10)$$

For the system to be free of the quadratic field-dependent aberrations, the coefficient of the $h_{x_0}^2$ term in the Taylor's series expansion of the mixed characteristic must be a constant, i.e.

$$\frac{m^2 |\bar{\rho} \times \hat{x}_1|^2}{2n_1 t_s} = \text{constant } A. \quad (7.11)$$

Similarly, for the tangential ray, we have the following condition for an optical system to be free of quadratic field-dependent aberration (see Figure 7.6 for the definition of t_t):

$$\frac{m^2 |\bar{\rho} \times \hat{y}_1|^2}{2n_1 t_t} = \text{constant } B. \quad (7.12)$$

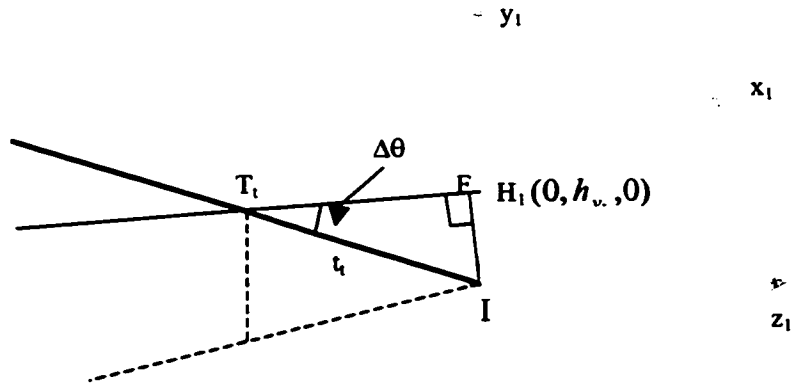


Figure 7.6. 3-D illustration of the tangential ray of the parallel bundle in image space. The ray intersects the image plane at $(0, h_{y_t}, 0)$. The angle between the tangential ray and the center ray of the bundle is $\Delta\theta$.

In Eqs. (7.11) and (7.12), \hat{x}_1 and \hat{y}_1 are the unit vectors along x_1 and y_1 axis respectively. For lack of the rotational symmetry, A and B are not necessarily equal. When the general Sine condition is satisfied, m is a constant. Also for the most desirable system, the image space is homogeneous and isotropic, so n_l is also a constant. Then the simplified form of the condition becomes

$$\frac{|\bar{\rho} \times \hat{x}_1|^2}{t_s} = \text{constant}, \quad (7.13a)$$

$$\frac{|\bar{\rho} \times y_1|^2}{t_t} = \text{constant}. \quad (7.13b)$$

Eqs. (7.13) are the Pupil Astigmatism Conditions for plane symmetric systems. Again the axially symmetric system is a special type of plane symmetric system, so the Pupil Astigmatism Criteria (2.50) should be a special example of (7.13). For axisymmetric systems, we need to consider only the rays in y - z plane because of the symmetry. It is easily shown that criteria (2.50) are the special form of (7.13).

Knowing the general Sine Condition (7.7) and the general Pupil Astigmatism Conditions (7.13) for the plane symmetric system enables us to design new types of systems that are corrected for aberrations up to the second order of field dependence and all orders of pupil dependence. Like in the design of axisymmetric systems, Coddington equations will play an important role in designing these new systems. Tools for implementing the criteria to design such systems and tools for evaluating the performance of the designed systems in an optical design code are yet to be developed. This will be an interesting area of future work.

7.5 CONCLUSION

Plane symmetric systems are studied in this chapter. The Hamilton's characteristic functions for the plane symmetric system are constructed and used to derive the general Sine Condition and Pupil Astigmatism Criteria. The general Sine Condition and Pupil Astigmatism Criteria contain their counterparts for the axially symmetric systems as special cases. Again these criteria can be used in the numerical design of the plane symmetric systems that are free of the field-independent, linear and quadratic field-dependent aberrations. This work definitely adds a useful tool in the design of plane symmetric systems.

CHAPTER 8

CONCLUSION

We derived a set of conditions, which we call Pupil Astigmatism Conditions, for correcting quadratic field-dependent aberrations. Hamilton's characteristic functions were used in the derivation. Just like the Constant OPL Condition is used for correcting all orders of spherical aberrations and the Abbe Sine Condition is used for correcting all orders of linear coma, the Pupil Astigmatism Conditions can be used to correct all orders of aberrations that have quadratic field dependence. In this dissertation, we show that these conditions are valuable new tools for analyzing and designing imaging systems.

The Pupil Astigmatism Conditions give the relations between the quadratic field-dependent image aberrations and the astigmatism and field curvature of the pupil. It is shown that the relations between image and pupil Seidel aberrations are the 3rd order approximation of these more general relations.

Examples of numerous systems were analyzed to demonstrate the power of the Pupil Astigmatism Conditions. Compared to Seidel formulas, these conditions provide more information about the aberrations that have quadratic field dependence.

The pupil astigmatism relations can be used not only to assess perfect systems, but to quantify the aberrations with quadratic field dependence without going off axis.

The Pupil Astigmatism Conditions were used to explicitly design 4-surface systems that are fully corrected for all orders of aberrations with quadratic field dependence. Some 3-surface systems were also designed to correct certain types of quadratic field-dependent aberrations. In this dissertation, the methods for making these designs is established, but not fully exercised. One area for future work may be in the application of this powerful design tool to develop a new class of systems.

Finally, a more general model of the pupil astigmatism conditions was developed for plane symmetric systems. The work presented here only shows that such conditions exist. There has been no attempt to use these conditions to design plane symmetric optical systems. It remains an important area of future work to follow this up and develop software and analysis tools that can exploit the pupil astigmatism conditions to design new types of plane symmetric imaging system efficiently.

APPENDIX A

SOURCE FILES FOR DESIGNING OPTICAL SYSTEMS NUMERICALLY

Numeric method is used to design optical systems free of all orders of spherical aberrations, all the linear and quadratic field-dependent aberrations. "Ana.cpp" is the source file for generating the surfaces of an infinite conjugate optical system. "Def.h" is the header file which defines all the data structures and general variables used in "Ana.cpp".

Ana.cpp

```

/*****
This program designs 4-surface systems that correct all spherical aberrations and all
linear and quadratic field-dependent aberrations. The object is at infinity.
by Chunyu Zhao 10-21-2001
*****/

```

```

#include <iostream.h>
#include <math.h>
#include <fstream.h>
#include <iomanip.h>
#include "def.h"

```

```

double Codd1(double,double,double,double,double,double,double, double);
double Codd2(double,double,double,double,double,double,double,double
*,double *,double *,double *);
double Codd_st(double, codd *);
double Cal_t(pnts *, slope *, double *, double *, double*, double *);
double Cal_s(pnts *, double *, double *, double*, double *);

```

```

void main()
{
//system parameter
    n0 = 1;
    n1 = 1.5;

```

```

n2 = 1;
n3 = 1.5;
n4 = 1;
t0 = 1;           //for infinite conjugate case,the distance from a plane to 1st surf
t1 = 5;
t2 = 5;
t3 = 5;
t4 = 23.233582;
m = -15;         //for infinite conjugate case, should be focal length
f = 18.16539;    //focal length
t0 = 10;

to0 = t0/n0;
to1 = t1/n1;
to2 = t2/n2;
to3 = t3/n3;
to4 = t4/n4;

double p1, p2, p3, p4, r1, r2, r3, r4;
cout<<"Press any key to continue:\n";
char yn;

r1 = -4.006932;
r2 = -4.233832;
r3 = -9.259410;
r4 = -8.239058;

p1 = (n1 - n0)/r1;
p2 = (n2 - n1)/r2;
p3 = (n3 - n2)/r3;
p4 = (n4 - n3)/r4;

// dfi = Codd1(p1, p2, p3, p4, t1, t2, t3, t4);
dfi = Codd1(p1, p2, p3, p4, t0, t1, t2, t3, t4);
cout<<"\ndfi = "<<dfi<<"\n";
cout<<"Press any key to continue:\n";
cin>>yn;

//define the output data streams
char* Filename="C:\\Zemax\\Data Files\\InfSurfaceData0.dat";
char* FileSurf1 = "C:\\Zemax\\Data Files\\InfSurf1.dat";
char* FileSurf2 = "C:\\Zemax\\Data Files\\InfSurf2.dat";
char* FileSurf3 = "C:\\Zemax\\Data Files\\InfSurf3.dat";

```

```
char* FileSurf4 = "C:\\Zemax\\Data Files\\InfSurf4.dat";
char* FileBoundary = "C:\\Zemax\\Data Files\\InfBdata.dat";
```

```
fstream SurfaceData(Filename, ios::out);
fstream Surf1(FileSurf1, ios::out|ios::binary);
fstream Surf2(FileSurf2, ios::out|ios::binary);
fstream Surf3(FileSurf3, ios::out|ios::binary);
fstream Surf4(FileSurf4, ios::out|ios::binary);
fstream Bdata(FileBoundary, ios::out|ios::binary);
```

```
//Set the vertex values of the 4 surfaces
```

```
xSurf1[0]=0;
zSurf1[0]=0;
slopeSurf1[0]=0;
xSurf2[0]=0;
zSurf2[0]=0;
slopeSurf2[0]=0;
xSurf3[0]=0;
zSurf3[0]=0;
slopeSurf3[0]=0;
xSurf4[0]=0;
zSurf4[0]=0;
slopeSurf4[0]=0;
opd[0]=n0*t0 + n1*t1 + n2*t2 + n3*t3 + n4*t4; //infinte conjugate case
```

```
prevd0 = t0;//new
prevcd.p1 = p1;
prevcd.p2 = p2;
prevcd.p3 = p3;
prevcd.p4 = p4;
prevcd.d1 = t1;
prevcd.d2 = t2;
prevcd.d3 = t3;
prevcd.d4 = t4;
prevcd.cs1 = 1;
prevcd.cs1p = 1;
prevcd.cs2 = 1;
prevcd.cs2p = 1;
prevcd.cs3 = 1;
prevcd.cs3p = 1;
prevcd.cs4 = 1;
prevcd.cs4p = 1;
dfti = dfti;
```

```

//out put the coordinates, slope of points at each surface to the file
SurfaceData<<setw(12)<<setprecision(8)<<xSurf1[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf1[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf1[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf2[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf2[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf2[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf3[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf3[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf3[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf4[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf4[0]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf4[0]<<" ";
SurfaceData<<setw(12)<<setprecision(12)<<opd[0];
SurfaceData<<"\n";

//define variables
int i, j, k, k1;
int k2;
int NMax = nps;
double maxSin = 0.30; //NA; //Max height for inf case
double H = 2.5; //max height
double dh, h0; //height increment, specific height
double dSin;
double x1, z1, x4, z4;
double x2, z2, x20, z20, x2p, z2p;
double x3, z3, x30, z30, x3p, z3p, x3pp, z3pp;
double slp1, slp2, slp3, slp4;
double maxCt, dCt, ct0, Sct0, Tct0, Sct4, C2ct4, Tct4;
double ds1, ds2, ds3, df;
double dt1, dt2, dt3;
double theta, alpha, alpha0, betap, slp;
double curv1, curv2, curv3, curv4;
double opl;
double b1, c1, b2, c2;
double g0, dg;
double rt1, rt2, dz3, ddz3, l3, Cphi3, Sphi3;
double l2, Cphi2, Sphi2;
double root;
double dr, dx, dz, dfx, dfz, dt3x, dt3z, ang;
pnts pnts_4;
slope slope1;

curv1 = 1/r1;

```



```

    curv2 = 1/r2;
    curv3 = 1/r3;
    curv4 = 1/r4;
//    maxCt = asin(maxSin);
//    dCt = maxCt/nps/nis;
//    dSin = maxSin/nps/nis; //delta sine between adjacent iteration
    dh = H/nps/nis;

    slope1.x1 = 0;
    slope1.x2 = 0;
    slope1.x3 = 0;
    slope1.x4 = 0;
    slope1.slp1 = 0;
    slope1.slp2 = 0;
    slope1.slp3 = 0;
    slope1.slp4 = 0;

//Calculate surface data
//total number of data points at each surface is nps+1.
for(j=0; j<nps; j++){
    //set the starting values of new iteration
    xSurf1 1[0]=xSurf1[j];
    zSurf1 1[0]=zSurf1[j];
    slopeSurf1 1[0]=slopeSurf1[j];
    xSurf2 1[0]=xSurf2[j];
    zSurf2 1[0]=zSurf2[j];
    slopeSurf2 1[0]=slopeSurf2[j];
    xSurf3 1[0]=xSurf3[j];
    zSurf3 1[0]=zSurf3[j];
    slopeSurf3 1[0]=slopeSurf3[j];
    xSurf4 1[0]=xSurf4[j];
    zSurf4 1[0]=zSurf4[j];
    slopeSurf4 1[0]=slopeSurf4[j];

    //iterates "nis" number of times to get values at the (j+1)th point
    for(i=0; i<nis;i++)
    {
/*
        ct0 = (double)(i + 1 +j*nis)*dCt;
        Sct0 = sin(ct0);
        Sct4 = n0*Sct0/n4/m; //the sign conventions for ct0 and ct4 are
same
        Tct0 = Sct0/sqrt(1-Sct0*Sct0);

```

```

C2ct4 = 1 - Sct4*Sct4;
Tct4 = Sct4/sqrt(C2ct4);

*/

h0 = (double)(i + 1 + j*nis)*dh;
Sct0 = 0;
Sct4 = -h0/f;
Tct0 = Sct0/sqrt(1-Sct0*Sct0);
C2ct4 = 1 - Sct4*Sct4;
Tct4 = Sct4/sqrt(C2ct4);

dftip = dfti;
dfti = dfti*C2ct4;

//find out x4, z4: trace rays from image point back to the 4th
surface

x4 = xSurf41[i] + (Tct4*(zSurf41[i]-t4) - xSurf41[i])/(1-
Tct4*slopeSurf41[i]);
z4 = zSurf41[i] + slopeSurf41[i]*(x4-xSurf41[i]);

theta = asin(Sct4);
alpha = atan(slopeSurf41[i] + curv4*(x4-xSurf41[i]));
betap = asin(n4*sin(theta + alpha)/n3);
slp = tan(betap - alpha);

//find out (x3, z3): trace rays from 4th surface to 3rd surface

x3p = xSurf31[i] + (slp*(zSurf31[i]-t3-z4) - (xSurf31[i]-x4))/(1-
slp*slopeSurf31[i]);
z3p = zSurf31[i] + slopeSurf31[i]*(x3-xSurf31[i]);

x3 = x3p;
z3 = z3p;

//find out x1, z1

x1 = h0; //infinite conjugate case
z1 = zSurf11[i] + slopeSurf11[i]*(x1-xSurf11[i]);

theta = asin(Sct0); //incidence ray angle with z axis
alpha = atan(slopeSurf11[i] + curv1*(x1-xSurf11[i])); //surface
slope angle with x axis

```

```

z-axis          betap = asin(n0*sin(theta + alpha)/n1); //refractive ray angle with
                slp = tan(betap - alpha);

                //assign values to data structure pnts_4
                pnts_4.x1 = x1;
                pnts_4.z1 = z1;
                pnts_4.x4 = x4;
                pnts_4.z4 = z4;

                //find out (x2, z2): trace rays from 1st surface to 2nd surface

                x2 = xSurf21[i] + (slp*(zSurf21[i]+t1-z1) - (xSurf21[i]-x1))/(1-
slope*slopeSurf21[i]);
                z2 = zSurf21[i] + slopeSurf21[i]*(x2-xSurf21[i]);

                x20 = x2;
                z20 = z2;

                theta = betap - alpha; //incidence ray angle with z axis
                alpha = atan(slopeSurf21[i] + curv2*(x2-xSurf21[i])); //surface
slope angle with x axis
                betap = asin(n1 *sin(theta + alpha)/n2); //refractive ray angle with
z-axis
                slp = tan(betap - alpha); //refracted ray direction?
                alpha0 = alpha;

                dx = 0.01*(x2 - xSurf21[i]);
                dz = 0.01*(z2 - zSurf21[i]);

                k1 = 1; //an indicator
                k2 = 0;

                //-----

                do{

                if(k1 == 2){
                l2 = 0.05*sqrt(pow(x2 - xSurf21[i], 2) + pow(z2 - zSurf21[i], 2));
                Cphi2 = cos(alpha0);
                Sphi2 = sin(alpha0);

```

```

x2 = l2*Cphi2 + x2;
z2 = l2*Sphi2 + z2;
theta = atan((x2-x1)/(t1+z2-z1));
alpha = atan(slopeSurf21[i] + curv2*(x2-xSurf21[i]));
betap = asin(n1*sin(theta + alpha)/n2);
slp = tan(betap - alpha);
}

if(k1>2){
    x2 = x2 - dt1*df;
    z2 = z20 + tan(alpha0)*(x2-x20) + 0.5*curv2*(x2-
x20)*(x2-x20);

    //z2 accurate to quadratic term
    theta = atan((x2-x1)/(t1+z2-z1));
    alpha = atan(slopeSurf21[i] + curv2*(x2-xSurf21[i]));
    betap = asin(n1*sin(theta + alpha)/n2);
    slp = tan(betap - alpha);
}
//assign values to data structure pnts_4
pnts_4.x2 = x2;
pnts_4.z2 = z2;

//make sure total opd is constant
opl = opd[0] - fabs(n0)*(z1+t0) - fabs(n1)*sqrt((x2-x1)*(x2-x1) +
(z2+t1-z1)*(z2+t1-z1))
- fabs(n4)*sqrt(x4*x4 + (t4-z4)*(t4-z4));
//Cartesian oval: function of (x3,z3), opd from (x2, z2) to (x4, z4)
is equal to opl.
//In (x3, z3) local coordinates: (x2,z2)->(x2,z2-t2), (x4,z4)-
>(x4,z4+t3).
b1 = -2*(z2 - t2);
c1 = (z2 - t2)*(z2 - t2) + (x3 - x2)*(x3 - x2);
b2 = -2*(z4 + t3);
c2 = (z4 + t3)*(z4 + t3) + (x3 - x4)*(x3 - x4);

//given x3, find z3. (x3, z3) is on the cartesian oval from (x2,z2)
to (x4,z4)
//Newton's method is used, previously calculated (x3,z3) is the
starting point:
do{
    rt1 = sqrt(z3*z3 + b1*z3 + c1);
    rt2 = sqrt(z3*z3 + b2*z3 + c2);
    g0 = n2*rt1 + n3*rt2 - opl;
    dg = n2*(z3 + b1/2)/rt1 + n3*(z3 + b2/2)/rt2;

```

```

        z3 = z3 - g0/dg;
    }while(fabs(g0)>1e-8);

    //assign values to data structure pnts_4
    pnts_4.x3 = x3;
    pnts_4.z3 = z3;

    ds1 = Cal_s(&pnts_4, &slp1, &slp2, &slp3, &slp4) - dfi;

    rt1 = sqrt(z3*z3 + b1*z3 + c1);
    rt2 = sqrt(z3*z3 + b2*z3 + c2);
    dz3 = -(n2*(x3-x2)*rt2 + n3*(x3-x4)*rt1)/(n2*(z3-z2+t2)*rt2 +
n3*(z3-z4-t3)*rt1);
    Cphi3 = cos(atan(dz3));
    Sphi3 = sin(atan(dz3));

    //calculate the 2nd derivative at (x3, z3) along the cartesian oval
    l3 = 0.05*sqrt(pow(x3 - xSurf31[i], 2) + pow(z3 - zSurf31[i], 2));
    x3pp = l3*Cphi3 + x3;
    z3pp = l3*Sphi3 + z3;
    x30 = x3; //the parabola segment of the 3rd surface is centered
    z30 = z3; //at (x30, z30) which is a fixed point
    ddz3 = 1/(n2*(z3-z2+t2)/rt1 + n3*(z3-z4-t3)/rt2)*(-
(n2/rt1+n3/rt2)*(1+dz3*dz3)
        +n2/pow(rt1,3)*pow((x3-x2)+(z3-z2+t2)*dz3,2) +
n3/pow(rt2,3)*pow((x3-x4)+(z3-z4-t3)*dz3,2));

    //assign values to data structure pnts_4
    pnts_4.x3 = x3pp;
    pnts_4.z3 = z3pp;

    ds3 = Cal_s(&pnts_4, &slp1, &slp2, &slp3, &slp4) - dfi;

    df = (ds3 - ds1)/(x3pp - x3);
    //Newton's method to find out the (x3, z3) that satisfies the perfect
sagittal

    //field condition
    k = 0;
    do{
        x3 = x3 - ds1/df;
        z3 = z30 + dz3*(x3-x30) + 0.5*ddz3*(x3-x30)*(x3-x30);
//accurate to quadratic term

        c1 = (z2 - t2)*(z2 - t2) + (x3 - x2)*(x3 - x2);

```

```

c2 = (z4 + t3)*(z4 + t3) + (x3 - x4)*(x3 - x4);

/*
do {
rt1 = sqrt(z3*z3 + b1*z3 + c1);
rt2 = sqrt(z3*z3 + b2*z3 + c2);
g0 = n2*rt1 + n3*rt2 - opl;
dg = n2*(z3 + b1/2)/rt1 + n3*(z3 + b2/2)/rt2;
z3 = z3 - g0/dg;
}while(fabs(g0)>1e-8);*/

//assign values to data structure pnts_4
pnts_4.x3 = x3;
pnts_4.z3 = z3;

ds2 = Cal_s(&pnts_4, &slp1, &slp2, &slp3, &slp4) - dfi;

df = (ds2 - ds3)/(x3 - x3pp);
ds3 = ds2;
ds1 = ds2;
x3pp = x3;
k++;
}while((fabs(ds2)>1e-4)&&(k<10));

if(k1 == 1){
x2p = x2;
z2p = z2;

prevcd.p1 = (slp1 - (slope1.slp1))/(x1 -
(slope1.x1))/pow(sqrt(1 + slope1.slp1*slope1.slp1), 3)*(n1*prevcd.cs1p -
n0*prevcd.cs1);
prevcd.p2 = (slp2 - (slope1.slp2))/(x2 -
(slope1.x2))/pow(sqrt(1 + slope1.slp2*slope1.slp2), 3)*(n2*prevcd.cs2p -
n1*prevcd.cs2);
prevcd.p3 = (slp3 - (slope1.slp3))/(x3 -
(slope1.x3))/pow(sqrt(1 + slope1.slp3*slope1.slp3), 3)*(n3*prevcd.cs3p -
n2*prevcd.cs3);
prevcd.p4 = (slp4 - (slope1.slp4))/(x4 -
(slope1.x4))/pow(sqrt(1 + slope1.slp4*slope1.slp4), 3)*(n4*prevcd.cs4p -
n3*prevcd.cs4);

dt3 = Codd_st(prevd0, &prevcd) - dftip;
}

```

```

else if(k1 == 2){
    prevcd.p1 = (slp1 - (slope1.slp1))/(x1 -
(slope1.x1))/pow(sqrt(1 + slope1.slp1*slope1.slp1), 3)*(n1*prevcd.cs1p -
n0*prevcd.cs1);
    prevcd.p2 = (slp2 - (slope1.slp2))/(x2 -
(slope1.x2))/pow(sqrt(1 + slope1.slp2*slope1.slp2), 3)*(n2*prevcd.cs2p -
n1*prevcd.cs2);
    prevcd.p3 = (slp3 - (slope1.slp3))/(x3 -
(slope1.x3))/pow(sqrt(1 + slope1.slp3*slope1.slp3), 3)*(n3*prevcd.cs3p -
n2*prevcd.cs3);
    prevcd.p4 = (slp4 - (slope1.slp4))/(x4 -
(slope1.x4))/pow(sqrt(1 + slope1.slp4*slope1.slp4), 3)*(n4*prevcd.cs4p -
n3*prevcd.cs4);

    dt1 = Codd_st(prevd0, &prevcd) - dftip;
    df = (x2p - x2)/(dt3 - dt1);
}

else{
    prevcd.p1 = (slp1 - (slope1.slp1))/(x1 -
(slope1.x1))/pow(sqrt(1 + slope1.slp1*slope1.slp1), 3)*(n1*prevcd.cs1p -
n0*prevcd.cs1);
    prevcd.p2 = (slp2 - (slope1.slp2))/(x2 -
(slope1.x2))/pow(sqrt(1 + slope1.slp2*slope1.slp2), 3)*(n2*prevcd.cs2p -
n1*prevcd.cs2);
    prevcd.p3 = (slp3 - (slope1.slp3))/(x3 -
(slope1.x3))/pow(sqrt(1 + slope1.slp3*slope1.slp3), 3)*(n3*prevcd.cs3p -
n2*prevcd.cs3);
    prevcd.p4 = (slp4 - (slope1.slp4))/(x4 -
(slope1.x4))/pow(sqrt(1 + slope1.slp4*slope1.slp4), 3)*(n4*prevcd.cs4p -
n3*prevcd.cs4);

    dt2 = Codd_st(prevd0, &prevcd) - dftip;
    df = (x2 - x2p)/(dt2 - dt3);
    dt3 = dt2;
    dt1 = dt2;
    x2p = x2;
}

k1++;

}while((fabs(dt3)>1e-6)&&(k1<20));

```

```

//-----

prevcd.d1 = prevcd1.d1;
prevcd.d2 = prevcd1.d2;
prevcd.d3 = prevcd1.d3;
prevcd.d4 = prevcd1.d4;
prevcd.cs1 = prevcd1.cs1;
prevcd.cs1p = prevcd1.cs1p;
prevcd.cs2 = prevcd1.cs2;
prevcd.cs2p = prevcd1.cs2p;
prevcd.cs3 = prevcd1.cs3;
prevcd.cs3p = prevcd1.cs3p;
prevcd.cs4 = prevcd1.cs4;
prevcd.cs4p = prevcd1.cs4p;
prevd0 = t0 + z1;//new

if(k1 == 20) {
    cout<<"\nNecessary precision could not be obtained!!!\n";
    NMax = j+1;

    //debugging
    int i1;
    for(i1=0; i1<=i;i1++){
        Bdata<<setw(12)<<setprecision(8)<<xSurf21[i1]<<" ";
        Bdata<<setw(12)<<setprecision(8)<<zSurf21[i1]<<" ";
        Bdata<<"\n";
    }

    break;
}

//assign values to data structure slope
slope1.x1 = x1;
slope1.x2 = x2;
slope1.x3 = x3;
slope1.x4 = x4;
slope1.slp1 = slp1;
slope1.slp2 = slp2;
slope1.slp3 = slp3;
slope1.slp4 = slp4;

xSurf11[i+1] = x1;
zSurf11[i+1] = z1;
slopeSurf11[i+1] = slp1;

```



```

    curv1 = (slp1-slopeSurf1 1[i])/(x1-xSurf1 1[i]);
    xSurf21[i+1] = x2;
    zSurf21[i+1] = z2;
    slopeSurf21[i+1] = slp2;
    curv2 = (slp2-slopeSurf21[i])/(x2-xSurf21[i]);
    xSurf31[i+1] = x3;
    zSurf31[i+1] = z3;
    slopeSurf31[i+1] = slp3;
    curv3 = (slp3-slopeSurf31[i])/(x3-xSurf31[i]);
    xSurf41[i+1] = x4;
    zSurf41[i+1] = z4;
    slopeSurf41[i+1] = slp4;
    curv4 = (slp4-slopeSurf41[i])/(x4-xSurf41[i]);

    }//i

    cout<<j+1<<" "<<i<<" "<<k1<<" "<<k<<" "<<ds3<<"
"<<dt3<<" "<<"\n";

//      cin>>yn;

    if(NMax == j+1) break;

    //set the values of a new iteration the end values of a previous iteration
    xSurf1[j+1]=xSurf1 1[nis];
    zSurf1[j+1]=zSurf1 1[nis];
    slopeSurf1[j+1]=slopeSurf1 1[nis];
    xSurf2[j+1]=xSurf21[nis];
    zSurf2[j+1]=zSurf21[nis];
    slopeSurf2[j+1]=slopeSurf21[nis];
    xSurf3[j+1]=xSurf31[nis];
    zSurf3[j+1]=zSurf31[nis];
    slopeSurf3[j+1]=slopeSurf31[nis];
    xSurf4[j+1]=xSurf41[nis];
    zSurf4[j+1]=zSurf41[nis];
    slopeSurf4[j+1]=slopeSurf41[nis];

    opd[j+1]=fabs(n0)*(t0 + zSurf1[j+1])
                + fabs(n1)*sqrt(pow(xSurf2[j+1] - xSurf1[j+1], 2) +
pow(t1 + zSurf2[j+1] - zSurf1[j+1], 2))
                + fabs(n2)*sqrt(pow(xSurf3[j+1] - xSurf2[j+1], 2) +
pow(t2 + zSurf3[j+1] - zSurf2[j+1], 2))

```

```

+ fabs(n3)*sqrt(pow(xSurf4[j+1] - xSurf3[j+1], 2) +
pow(t3 + zSurf4[j+1] - zSurf3[j+1], 2))
+ fabs(n4)*sqrt(pow(xSurf4[j+1], 2) + pow(t4 -
zSurf4[j+1], 2));

```

```

//write the data to the file
SurfaceData<<setw(12)<<setprecision(8)<<xSurf1[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf1[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf1[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf2[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf2[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf2[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf3[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf3[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf3[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<xSurf4[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<zSurf4[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(8)<<slopeSurf4[j+1]<<" ";
SurfaceData<<setw(12)<<setprecision(12)<<opd[j+1]<<" ";
SurfaceData<<"\n";
} //j

```

```

//Out put the data of each surface to a file. The data structure is as follows
// Number of data points
// An unused number
// Slope at the starting point
// Slope at the end point
// Pairs of radial and sag corrdinates

```

```

//output the data of 1st surface to a file
Surf1<<setw(25)<<setprecision(15)<<NMax<<" ";
Surf1<<setw(25)<<setprecision(15)<<0<<" ";
Surf1<<setw(25)<<setprecision(15)<<0<<" ";
Surf1<<setw(25)<<setprecision(15)<<slopeSurf1[NMax-1]<<" ";
Surf1<<"\n";
for(i = 0; i<NMax; i++){
    Surf1<<setw(25)<<setprecision(15)<<xSurf1[i]<<" ";
    Surf1<<setw(25)<<setprecision(15)<<zSurf1[i]<<" ";
    Surf1<<"\n";
}

```

```

//output the data of 2nd surface to a file
Surf2<<setw(25)<<setprecision(15)<<NMax<<" ";

```

```

Surf2<<setw(25)<<setprecision(15)<<0<<" ";
Surf2<<setw(25)<<setprecision(15)<<0<<" ";
Surf2<<setw(25)<<setprecision(15)<<slopeSurf2[NMax-1]<<" ";
Surf2<<"\n";
for(i = 0; i<NMax; i++){
    Surf2<<setw(25)<<setprecision(15)<<xSurf2[i]<<" ";
    Surf2<<setw(25)<<setprecision(15)<<zSurf2[i]<<" ";
    Surf2<<"\n";
}

//output the data of 3rd surface to a file
Surf3<<setw(25)<<setprecision(15)<<NMax<<" ";
Surf3<<setw(25)<<setprecision(15)<<0<<" ";
Surf3<<setw(25)<<setprecision(15)<<0<<" ";
Surf3<<setw(25)<<setprecision(15)<<slopeSurf3[NMax-1]<<" ";
Surf3<<"\n";
for(i = 0; i<NMax; i++){
    Surf3<<setw(25)<<setprecision(15)<<xSurf3[i]<<" ";
    Surf3<<setw(25)<<setprecision(15)<<zSurf3[i]<<" ";
    Surf3<<"\n";
}

//output the data of 4th surface to a file
Surf4<<setw(25)<<setprecision(15)<<NMax<<" ";
Surf4<<setw(25)<<setprecision(15)<<0<<" ";
Surf4<<setw(25)<<setprecision(15)<<0<<" ";
Surf4<<setw(25)<<setprecision(15)<<slopeSurf4[NMax-1]<<" ";
Surf4<<"\n";
for(i = 0; i<NMax; i++){
    Surf4<<setw(25)<<setprecision(15)<<xSurf4[i]<<" ";
    Surf4<<setw(25)<<setprecision(15)<<zSurf4[i]<<" ";
    Surf4<<"\n";
}
}

//calculating the distance from focus to image plane, negative
//value is expected, the first 4 arguments are powers of each element
//the 5th is the distance from a plane to the 1st surface
//the last 4 arguments are spacings
//also used are the system parameters such as indicis of refraction.
// The formula is Coddington equation. s1 is finite.
double Codd1(double p1,double p2,double p3, double p4, double d0, double d1, double
d2, double d3, double d4)

```

```

{
    double s1, s1p, s2, s2p, s3, s3p, s4, s4p;
    double s;

    s1 = -d0; //distance from the selected plane to 1st surface

    if(s1==0) s1p = 0;
    else if(fabs(p1+n0/s1)<(1e-20)) s1p = n1*1e30;
    else s1p = n1/(p1+n0/s1);

    s2 = s1p - d1;
    if(s2==0) s2p = 0;
    else if(fabs(p2+n1/s2)<(1e-20)) s2p = n2*1e30;
    else s2p = n2/(p2+n1/s2);

    s3 = s2p - d2;
    if(s3 == 0) s3p = 0;
    else if(fabs(p3+n2/s3)<(1e-20)) s3p = n3*1e30;
    else s3p = n3/(p3+n2/s3);

    s4 = s3p - d3;
    if(s4 == 0) s4p = 0;
    else if(fabs(p4+n3/s4)<(1e-20)) s4p = n4*1e30;
    else s4p = n4/(p4+n3/s4);

    s = s4p - d4;
    return s;
}

//calculating the distance from focus to image plane, negative
//value is expected, the first 8 arguments are coordinates of ray intersection
//at each element. The last 4 arguments hold the slopes. Also used are the
//system parameters such as indices of refraction and spacings. The formula
//is Coddington equation.
double Codd2(double x1, double z1, double x2, double z2, double x3, double z3,
             double x4, double z4, double * slp1, double * slp2, double * slp3,
             double * slp4)
{
    double cs1, cs1p, cs2, cs2p, cs3, cs3p, cs4, cs4p; //cosines of incident and
refractive angle
    double p1, p2, p3, p4; //oblic power at each surface
    double d0, d1, d2, d3, d4; //equivalent spacings between elements
    double r1, r2, r3, r4; //sagittal radius of curvature
    double Cct0, Sct0, Cct1, Sct1, Cct2, Sct2, Cct3, Sct3, Cct4, Sct4; //ray angles

```

```

double N1z, N1x, N2z, N2x, N3z, N3x, N4z, N4x; //normal at each surface
double N1l, N2l, N3l, N4l; //length of the normal at each surface
// double slope3;

// d0 = sqrt(x1*x1 + (t0+z1)*(t0+z1)); //finite case
d0 = t0+z1; //infinite case
d1 = sqrt((x2-x1)*(x2-x1) + (t1+z2-z1)*(t1+z2-z1));
d2 = sqrt((x3-x2)*(x3-x2) + (t2+z3-z2)*(t2+z3-z2));
d3 = sqrt((x4-x3)*(x4-x3) + (t3+z4-z3)*(t3+z4-z3));
d4 = sqrt(x4*x4 + (t4-z4)*(t4-z4));

// Cct0 = (t0+z1)/d0; //finite case
// Sct0 = x1/d0;
Cct0 = 1; //infinite case
Sct0 = 0;
Cct1 = (t1+z2-z1)/d1;
Sct1 = (x2-x1)/d1;
Cct2 = (t2+z3-z2)/d2;
Sct2 = (x3-x2)/d2;
Cct3 = (t3+z4-z3)/d3;
Sct3 = (x4-x3)/d3;
Cct4 = (t4-z4)/d4;
Sct4 = -x4/d4;

N1z = n1*Cct1 - n0*Cct0;
N1x = n1*Sct1 - n0*Sct0;
N1l = sqrt(N1z*N1z + N1x*N1x);

N2z = n2*Cct2 - n1*Cct1;
N2x = n2*Sct2 - n1*Sct1;
N2l = sqrt(N2z*N2z + N2x*N2x);

N3z = n3*Cct3 - n2*Cct2;
N3x = n3*Sct3 - n2*Sct2;
N3l = sqrt(N3z*N3z + N3x*N3x);

N4z = n4*Cct4 - n3*Cct3;
N4x = n4*Sct4 - n3*Sct3;
N4l = sqrt(N4z*N4z + N4x*N4x);

/*
(*slp1) = -atan2(N1x, N1z);
(*slp2) = -atan2(N2x, N2z);
(*slp3) = -atan(N3x/N3z);
slope3 = -atan(N3x/N3z);

```

```

(*slp1) = -tan(atan2(N1x, N1z));
(*slp2) = -tan(atan2(N2x, N2z));
(*slp3) = -tan(atan2(N3x, N3z));
*/
(*slp1) = -N1x/N1z;
(*slp2) = -N2x/N2z;
(*slp3) = -N3x/N3z;
(*slp4) = -N4x/N4z;

/*
r1 = ((*slp1)==0)? 0 : x1/sin(atan((*slp1)));
r2 = ((*slp2)==0)? 0 : x2/sin(atan((*slp2)));
r3 = ((*slp3)==0)? 0 : x3/sin(atan((*slp3)));
*/
r1 = (fabs((*slp1))<= fabs((1e-10)*x1))? 0 : x1*sqrt(1+pow((*slp1),2))/(*slp1);
r2 = (fabs((*slp2))<= fabs((1e-10)*x2))? 0 : x2*sqrt(1+pow((*slp2),2))/(*slp2);
r3 = (fabs((*slp3))<= fabs((1e-10)*x3))? 0 : x3*sqrt(1+pow((*slp3),2))/(*slp3);
r4 = (fabs((*slp4))<= fabs((1e-10)*x4))? 0 : x4*sqrt(1+pow((*slp4),2))/(*slp4);

cs1 = fabs((N1z*Cct0 + N1x*Sct0)/N1l);
cs1p = fabs((N1z*Cct1 + N1x*Sct1)/N1l);

cs2 = fabs((N2z*Cct1 + N2x*Sct1)/N2l);
cs2p = fabs((N2z*Cct2 + N2x*Sct2)/N2l);

cs3 = fabs((N3z*Cct2 + N3x*Sct2)/N3l);
cs3p = fabs((N3z*Cct3 + N3x*Sct3)/N3l);

cs4 = fabs((N4z*Cct3 + N4x*Sct3)/N4l);
cs4p = fabs((N4z*Cct4 + N4x*Sct4)/N4l);

p1 = (r1 == 0)? 0 : (n1*cs1p - n0*cs1)/r1;
p2 = (r2 == 0)? 0 : (n2*cs2p - n1*cs2)/r2;
p3 = (r3 == 0)? 0 : (n3*cs3p - n2*cs3)/r3;
p4 = (r4 == 0)? 0 : (n4*cs4p - n3*cs4)/r4;

return Codd1(p1, p2, p3, p4, d0, d1, d2, d3, d4);

}

```

//General Coddington equation: both tangential and sagittal versions are embedded.
//return: the distance from finite image location to the focus.

```

double Codd_st(double d0, codd * cd)
{
    double s1, s1p, s2, s2p, s3, s3p, s4, s4p;
    double s;

    s1 = -d0; //distance from the selected plane to 1st surface

//object at finite distance away, ray trace along a certain ray

    //at the 1st surface
//    s1p = (cd->p1==0)? n1*1e30 : n1*pow(cd->cs1,2)/(cd->p1);
//    s2 = s1p - (cd->d1);
    if(s1==0) s1p = 0; //right at the 1st surface
    else if(fabs((cd->p1)+n0*pow(cd->cs1,2)/s1)<(1e-20)) s1p = n1*1e30; //right at
the local front focus
    else s1p = n1*pow(cd->cs1,2)/((cd->p1)+n0*pow(cd->cs1,2)/s1); //general case
    s2 = s1p - (cd->d1);

    //at the 2nd surface
    if(s2==0) s2p = 0; //right at the 2nd surface
    else if(fabs((cd->p2)+n1*pow(cd->cs2,2)/s2)<(1e-20)) s2p = n2*1e30; //right at
the local front focus
    else s2p = n2*pow(cd->cs2,2)/((cd->p2)+n1*pow(cd->cs2,2)/s2); //general case
    s3 = s2p - (cd->d2);

    //at the 3rd surface
    if(s3 == 0) s3p = 0;
    else if(fabs((cd->p3)+n2*pow(cd->cs3,2)/s3)<(1e-20)) s3p = n3*1e30;
    else s3p = n3*pow(cd->cs3,2)/((cd->p3)+n2*pow(cd->cs3,2)/s3);
    s4 = s3p - (cd->d3);

    //at the 4th surface
    if(s4 == 0) s4p = 0;
    else if(fabs((cd->p4)+n3*pow(cd->cs4,2)/s4)<(1e-20)) s4p = n4*1e30;
    else s4p = n4*pow(cd->cs4,2)/((cd->p4)+n3*pow(cd->cs4,2)/s4);
    s = s4p - (cd->d4);

    //return the distance between the local focus to the system on-axis image point
    return s;
}

/*****
Cal_t: apply Coddington equations to tangential rays.

```

input: the ray intersection at each surface and the previous sections' slopes
 return: the distance between tangential focus and the image point and
 the slopes at each surface

```

*****/
double Cal_t(pnts * pnt_4, slope * slp, double *slp01, double *slp02, double *slp03,
double *slp04)
{
    codd cd1;

    double cs1, cs1p, cs2, cs2p, cs3, cs3p, cs4, cs4p; //cosines of incident and
refractive angle
    double d0, d1, d2, d3, d4; //equivalent spacings between elements
    double r1, r2, r3, r4; //sagittal radius of curvature
    double Cct0, Sct0, Cct1, Sct1, Cct2, Sct2, Cct3, Sct3, Cct4, Sct4; //ray angles
    double N1z, N1x, N2z, N2x, N3z, N3x, N4z, N4x; //normal at each surface
    double N1l, N2l, N3l, N4l; //length of the normal at each surface
    double slp1, slp2, slp3, slp4; //slopes at each surface

//    d0 = sqrt(pnt_4->x1*pnt_4->x1 + (t0+pnt_4->z1)*(t0+pnt_4->z1));
    d0 = pnt_4->z1 + t0; // the ray is parallel to the optical axis
    d1 = sqrt((pnt_4->x2-pnt_4->x1)*(pnt_4->x2-pnt_4->x1) + (t1+pnt_4->z2-
pnt_4->z1)*(t1+pnt_4->z2-pnt_4->z1));
    d2 = sqrt((pnt_4->x3-pnt_4->x2)*(pnt_4->x3-pnt_4->x2) + (t2+pnt_4->z3-
pnt_4->z2)*(t2+pnt_4->z3-pnt_4->z2));
    d3 = sqrt((pnt_4->x4-pnt_4->x3)*(pnt_4->x4-pnt_4->x3) + (t3+pnt_4->z4-
pnt_4->z3)*(t3+pnt_4->z4-pnt_4->z3));
    d4 = sqrt(pnt_4->x4*pnt_4->x4 + (t4-pnt_4->z4)*(t4-pnt_4->z4));

//    Cct0 = (t0+pnt_4->z1)/d0;
//    Sct0 = pnt_4->x1/d0;
    Cct0 = 1;
    Sct0 = 0;
    Cct1 = (t1+pnt_4->z2-pnt_4->z1)/d1;
    Sct1 = (pnt_4->x2-pnt_4->x1)/d1;
    Cct2 = (t2+pnt_4->z3-pnt_4->z2)/d2;
    Sct2 = (pnt_4->x3-pnt_4->x2)/d2;
    Cct3 = (t3+pnt_4->z4-pnt_4->z3)/d3;
    Sct3 = (pnt_4->x4-pnt_4->x3)/d3;
    Cct4 = (t4-pnt_4->z4)/d4;
    Sct4 = -pnt_4->x4/d4;

    N1z = n1*Cct1 - n0*Cct0;
    N1x = n1*Sct1 - n0*Sct0;

```



```

N1l = sqrt(N1z*N1z + N1x*N1x);

N2z = n2*Cct2 - n1*Cct1;
N2x = n2*Sct2 - n1*Sct1;
N2l = sqrt(N2z*N2z + N2x*N2x);

N3z = n3*Cct3 - n2*Cct2;
N3x = n3*Sct3 - n2*Sct2;
N3l = sqrt(N3z*N3z + N3x*N3x);

N4z = n4*Cct4 - n3*Cct3;
N4x = n4*Sct4 - n3*Sct3;
N4l = sqrt(N4z*N4z + N4x*N4x);

slp1 = -tan(atan2(N1x, N1z));
slp2 = -tan(atan2(N2x, N2z));
slp3 = -tan(atan2(N3x, N3z));
slp4 = -tan(atan2(N4x, N4z));

r1 = (fabs(slp1)<= fabs((1e-10)*pnt_4->x1))? 0 : pnt_4-
>x1*sqrt(1+pow(slp1,2))/slp1;
r2 = (fabs(slp2)<= fabs((1e-10)*pnt_4->x2))? 0 : pnt_4-
>x2*sqrt(1+pow(slp2,2))/slp2;
r3 = (fabs(slp3)<= fabs((1e-10)*pnt_4->x3))? 0 : pnt_4-
>x3*sqrt(1+pow(slp3,2))/slp3;
r4 = (fabs(slp4)<= fabs((1e-10)*pnt_4->x4))? 0 : pnt_4-
>x4*sqrt(1+pow(slp4,2))/slp4;

cs1 = fabs((N1z*Cct0 + N1x*Sct0)/N1l);
cs1p = fabs((N1z*Cct1 + N1x*Sct1)/N1l);

cs2 = fabs((N2z*Cct1 + N2x*Sct1)/N2l);
cs2p = fabs((N2z*Cct2 + N2x*Sct2)/N2l);

cs3 = fabs((N3z*Cct2 + N3x*Sct2)/N3l);
cs3p = fabs((N3z*Cct3 + N3x*Sct3)/N3l);

cs4 = fabs((N4z*Cct3 + N4x*Sct3)/N4l);
cs4p = fabs((N4z*Cct4 + N4x*Sct4)/N4l);

//trace tangential rays

cd1.p1 = (n1*cs1p - n0*cs1)*(slp1 - (slp->slp1))/(pnt_4->x1 - (slp-
>x1))/pow(sqrt(1 + slp1*slp1), 3);

```

```

        cd1.p2 = (n2*cs2p - n1*cs2)*(slp2 - (slp->slp2))/(pnt_4->x2 - (slp-
>x2))/pow(sqrt(1 + slp2*slp2), 3);
        cd1.p3 = (n3*cs3p - n2*cs3)*(slp3 - (slp->slp3))/(pnt_4->x3 - (slp-
>x3))/pow(sqrt(1 + slp3*slp3), 3);
        cd1.p4 = (n4*cs4p - n3*cs4)*(slp4 - (slp->slp4))/(pnt_4->x4 - (slp-
>x4))/pow(sqrt(1 + slp4*slp4), 3);

        cd1.cs1 = cs1;
        cd1.cs1p = cs1p;
        cd1.cs2 = cs2;
        cd1.cs2p = cs2p;
        cd1.cs3 = cs3;
        cd1.cs3p = cs3p;
        cd1.cs4 = cs4;
        cd1.cs4p = cs4p;

        cd1.d1 = d1;
        cd1.d2 = d2;
        cd1.d3 = d3;
        cd1.d4 = d4;

        (*slp01) = slp1;
        (*slp02) = slp2;
        (*slp03) = slp3;
        (*slp04) = slp4;

        return Codd_st(d0,&cd1);
    }

/*****
Cal_s: apply Coddington equations to sagittal rays.
input: the ray intersection at each surface
return: the distance between sagittal focus and the image point and
the slopes at each surface.
*****/

double Cal_s(pnts * pnt_4, double *slp1, double *slp2, double *slp3, double *slp4)
{
    codd cd1;

    double cs1, cs1p, cs2, cs2p, cs3, cs3p, cs4, cs4p; //cosines of incident and
refractive angle
//    double p1, p2, p3, p4; //oblic power at each surface
    double d0, d1, d2, d3, d4; //equivalent spacings between elements

```

```

double r1, r2, r3, r4; //sagittal radius of curvature
double Cct0, Sct0, Cct1, Sct1, Cct2, Sct2, Cct3, Sct3, Cct4, Sct4; //ray angles
double N1z, N1x, N2z, N2x, N3z, N3x, N4z, N4x; //normal at each surface
double N1l, N2l, N3l, N4l; //length of the normal at each surface

// d0 = sqrt(pnt_4->x1*pnt_4->x1 + (t0+pnt_4->z1)*(t0+pnt_4->z1));
d0 = pnt_4->z1 + t0; //infinte case
d1 = sqrt((pnt_4->x2-pnt_4->x1)*(pnt_4->x2-pnt_4->x1) + (t1+pnt_4->z2-
pnt_4->z1)*(t1+pnt_4->z2-pnt_4->z1));
d2 = sqrt((pnt_4->x3-pnt_4->x2)*(pnt_4->x3-pnt_4->x2) + (t2+pnt_4->z3-
pnt_4->z2)*(t2+pnt_4->z3-pnt_4->z2));
d3 = sqrt((pnt_4->x4-pnt_4->x3)*(pnt_4->x4-pnt_4->x3) + (t3+pnt_4->z4-
pnt_4->z3)*(t3+pnt_4->z4-pnt_4->z3));
d4 = sqrt(pnt_4->x4*pnt_4->x4 + (t4-pnt_4->z4)*(t4-pnt_4->z4));

//each segment's ray direction
// Cct0 = (t0+pnt_4->z1)/d0;
// Sct0 = pnt_4->x1/d0;
Cct0 = 1;
Sct0 = 0;
Cct1 = (t1+pnt_4->z2-pnt_4->z1)/d1;
Sct1 = (pnt_4->x2-pnt_4->x1)/d1;
Cct2 = (t2+pnt_4->z3-pnt_4->z2)/d2;
Sct2 = (pnt_4->x3-pnt_4->x2)/d2;
Cct3 = (t3+pnt_4->z4-pnt_4->z3)/d3;
Sct3 = (pnt_4->x4-pnt_4->x3)/d3;
Cct4 = (t4-pnt_4->z4)/d4;
Sct4 = -pnt_4->x4/d4;

//surface normal at each surface
N1z = n1*Cct1 - n0*Cct0;
N1x = n1*Sct1 - n0*Sct0;
N1l = sqrt(N1z*N1z + N1x*N1x);

N2z = n2*Cct2 - n1*Cct1;
N2x = n2*Sct2 - n1*Sct1;
N2l = sqrt(N2z*N2z + N2x*N2x);

N3z = n3*Cct3 - n2*Cct2;
N3x = n3*Sct3 - n2*Sct2;
N3l = sqrt(N3z*N3z + N3x*N3x);

N4z = n4*Cct4 - n3*Cct3;
N4x = n4*Sct4 - n3*Sct3;

```

```

N4l = sqrt(N4z*N4z + N4x*N4x);

//slope at each surface
(*slp1) = -tan(atan2(N1x, N1z));
(*slp2) = -tan(atan2(N2x, N2z));
(*slp3) = -tan(atan2(N3x, N3z));
(*slp4) = -tan(atan2(N4x, N4z));

r1 = (fabs((*slp1))<= fabs((1e-10)*pnt_4->x1))? 0 : pnt_4-
>x1*sqrt(1+pow((*slp1),2))/(*slp1);
r2 = (fabs((*slp2))<= fabs((1e-10)*pnt_4->x2))? 0 : pnt_4-
>x2*sqrt(1+pow((*slp2),2))/(*slp2);
r3 = (fabs((*slp3))<= fabs((1e-10)*pnt_4->x3))? 0 : pnt_4-
>x3*sqrt(1+pow((*slp3),2))/(*slp3);
r4 = (fabs((*slp4))<= fabs((1e-10)*pnt_4->x4))? 0 : pnt_4-
>x4*sqrt(1+pow((*slp4),2))/(*slp4);

//cosines of incident angle and refractive angle
cs1 = fabs((N1z*Cct0 + N1x*Sct0)/N1l);
cs1p = fabs((N1z*Cct1 + N1x*Sct1)/N1l);

cs2 = fabs((N2z*Cct1 + N2x*Sct1)/N2l);
cs2p = fabs((N2z*Cct2 + N2x*Sct2)/N2l);

cs3 = fabs((N3z*Cct2 + N3x*Sct2)/N3l);
cs3p = fabs((N3z*Cct3 + N3x*Sct3)/N3l);

cs4 = fabs((N4z*Cct3 + N4x*Sct3)/N4l);
cs4p = fabs((N4z*Cct4 + N4x*Sct4)/N4l);

//trace sagittal rays

cd1.p1 = (r1 == 0)? 0 : (n1*cs1p - n0*cs1)/r1;
cd1.p2 = (r2 == 0)? 0 : (n2*cs2p - n1*cs2)/r2;
cd1.p3 = (r3 == 0)? 0 : (n3*cs3p - n2*cs3)/r3;
cd1.p4 = (r4 == 0)? 0 : (n4*cs4p - n3*cs4)/r4;

cd1.cs1 = 1;
cd1.cs1p = 1;
cd1.cs2 = 1;
cd1.cs2p = 1;
cd1.cs3 = 1;
cd1.cs3p = 1;

```

```

    cd1.cs4 = 1;
    cd1.cs4p = 1;

    cd1.d1 = d1;
    cd1.d2 = d2;
    cd1.d3 = d3;
    cd1.d4 = d4;

    prevcd1.d1 = d1;
    prevcd1.d2 = d2;
    prevcd1.d3 = d3;
    prevcd1.d4 = d4;
    prevcd1.cs1 = cs1;
    prevcd1.cs1p = cs1p;
    prevcd1.cs2 = cs2;
    prevcd1.cs2p = cs2p;
    prevcd1.cs3 = cs3;
    prevcd1.cs3p = cs3p;
    prevcd1.cs4 = cs4;
    prevcd1.cs4p = cs4p;

    return Codd_st(d0, &cd1);
}

```

Def.h

```

/*****
This file defines data structures and general variables.
*****/

#define nps 500 // output values at 1000 off axis points
#define nis 1000 // iterate 10000 times to get values at (j+1)th point
                // from those at jth point

//data structure that hold the coordinates of 4 points at 4 surfaces
typedef struct {
    double x1;
    double z1;
    double x2;
    double z2;
    double x3;
    double z3;
    double x4;
    double z4;
}

```

```

}pnts;

//data structure that hold the quantities used in Coddington equations
typedef struct{
    //power
    double p1;
    double p2;
    double p3;
    double p4;

    //cosines of incident and refractive angles at each surface
    double cs1;
    double cs1p;
    double cs2;
    double cs2p;
    double cs3;
    double cs3p;
    double cs4;
    double cs4p;

    //spacings between two adjacent surfaces along a ray
    double d1;
    double d2;
    double d3;
    double d4;
}codd;

//data structure that holds the slopes at certain points of ach surface
//it will be used to calculate the second derivative at a point
typedef struct{
    //x-coordinate at the previous iteration
    double x1;
    double x2;
    double x3;
    double x4;
    //slopes at the previous iteration
    double slp1;
    double slp2;
    double slp3;
    double slp4;
}slope;

//Define the arrays to hold the coordinates and slopes of the
//1st, 2nd, 3rd and 4th surfaces at output points and intermediate iterate points

```

```

double xSurf1[nps+1], xSurf11[nis+1];
double zSurf1[nps+1], zSurf11[nis+1];
double slopeSurf1[nps+1], slopeSurf11[nis+1];
double xSurf2[nps+1], xSurf21[nis+1];
double zSurf2[nps+1], zSurf21[nis+1];
double slopeSurf2[nps+1], slopeSurf21[nis+1];
double xSurf3[nps+1], xSurf31[nis+1];
double zSurf3[nps+1], zSurf31[nis+1];
double slopeSurf3[nps+1], slopeSurf31[nis+1];
double xSurf4[nps+1], xSurf41[nis+1];
double zSurf4[nps+1], zSurf41[nis+1];
double slopeSurf4[nps+1], slopeSurf41[nis+1];
double opd[nps+1];

//system parameters
double n0, n1, n2, n3, n4; //index of refraction
double t0, t1, t2, t3, t4; //spacing
double to0, to1, to2, to3, to4;
double m, f; //focal length for infinite conjugate case
double dfi; //on-axis distance from focus to image plane

double dfti; //distance between tangential focus and image plane
double dftip; //distance between sagittal focus and image plane
codd prevcd;
codd prevcd1;
double prevd0;//new

```

APPENDIX B

SOURCE FILES FOR THE ZEMAX USER DEFINED SURFACE

Since all the designs presented in Chapter 6 are generated point by point, i.e. each surface is made up of discrete points, I have to use the user defined surface feature provided by ZEMAX to simulate the system performance. I wrote a program which take the coordinates of the data points and fit them with the cubic spines. “CubSpln.cpp” is the c++ source file for the user defined surface. “CubSpln.rc” is the resource file. “Dialogdef.h” and “resource.h” are the header files. These four files combined with the “Usersurf.h” and “usersurf.def” provided by ZEMAX make the project that realizes this type of user defined surface in ZEMAX.

CubSpln.cpp

```
//This is the ZEMAX user defined surface source file. The program takes the coordinates
// of surface data points (up to 1000) generated by other programs and approximate the
//surface using the cublic spilines.
```

```
#include <math.h>
#include <string.h>
#include "usersurf.h"
#include "resource.h"
#include <windows.h>
#include <fstream.h>
#include <iomanip.h>

char filepath[100];
HINSTANCE hglobalInst;
BOOL bData;
int NumPoints;
double CubCoeff[1000][4];
double r[1000], z[1000], p[9];
double normal, derv0, derv97, dervN_1;
double * pCubCoeff[1000];
```



```

int __declspec(dllexport) APIENTRY UserDefinedSurface(USER_DATA *UD,
FIXED_DATA *FD);
LRESULT CALLBACK DlgProc (HWND hDlg, UINT message, WPARAM wParam,
LPARAM lParam);

```

```

/* a generic Snells law refraction routine */
int Refract(double thisn, double nextn, double *l, double *m, double *n, double ln,
double mn, double nn);
void CalCoeff(double **, double *, double *, int, double, double, double);
int Locate(double *, int, double);
double Sag(double, int);
double dSag(double, int);
double PolyTerms(double);
double dPolyTerms(double);

```

```

BOOL WINAPI DllMain (HANDLE hInst, ULONG ul_reason_for_call, LPVOID
lpReserved)
{
    hglobalInst = hInst;
    return TRUE;
}

```

/* this DLL models a standard ZEMAX surface type, either plane, sphere, or conic */

```

int __declspec(dllexport) APIENTRY UserDefinedSurface(USER_DATA *UD,
FIXED_DATA *FD)
{
    int i;
    int beginIndex, middleIndex, endIndex, bIndex, mIndex, eIndex;
    int N = 1024;
    char coef[4];
    double radial, a, b, r1, r2, x, slp;
    double t0, t1, t, f, df, alpha, phi;
    double rt, zt, dr;
    double l1, m1, n1, ln1, mn1, nn1, power;
    switch(FD->type)
    {
    case 0:
        /* ZEMAX is requesting general information about the surface */
        switch(FD->numb)
        {
        case 0:
            /* ZEMAX wants to know the name of the surface */

```

```

        /* do not exceed 12 characters */
        strcpy(UD->string,"CubSpln");
    break;
case 1:
    /* ZEMAX wants to know if this surface is rotationally symmetric */
    /* it is, so return any character in the string; otherwise, return a null string */
    strcpy(UD->string, "1");
    break;
case 2:
    /* ZEMAX wants to know if this surface is a gradient index media */
    /* it is not, so return a null string */
    UD->string[0] = '\0';
    break;
}
break;
case 1:
    /* ZEMAX is requesting the names of the parameter columns */
    /* the value FD->numb will indicate which value ZEMAX wants. */
    /* they are all "Unused" for this surface type */
    /* returning a null string indicates that the parameter is unused. */
    switch(FD->numb)
    {
    default:
        strcpy(UD->string, "a");
        strcat(UD->string, _itoa(2*(FD->numb-1), coef, 10));
        break;
    }
    break;
case 2:
    /* ZEMAX is requesting the names of the extra data columns */
    /* the value FD->numb will indicate which value ZEMAX wants. */
    /* they are all "Unused" for this surface type */
    /* returning a null string indicates that the extradata value is unused. */
    switch(FD->numb)
    {
        case 1:
            strcpy(UD->string, "N Points");
            break;
        case 2:
            strcpy(UD->string, "Normal?");
            break;
        case 3:
            strcpy(UD->string, "Derivative@r0");
            break;
    }
}

```

```

case 4:
    strcpy(UD->string, "Derivative@rN_1");
    break;

default:
    if(FD->numb%2){
        strcpy(UD->string, "r");
        strcat(UD->string, _itoa(FD->numb/2-2, coef,
10));
    }
    else{
        strcpy(UD->string, "z");
        strcat(UD->string, _itoa(FD->numb/2-3, coef,
10));
    }

    break;
}
break;
case 3:
    /* ZEMAX wants to know the sag of the surface */
    /* if there is an alternate sag, return it as well */
    /* otherwise, set the alternate sag identical to the sag */
    /* The sag is sag1, alternate is sag2. */

    for(i = 1; i<9; i++) p[i] = FD->param[i];
    radial = sqrt(UD->x*UD->x + UD->y*UD->y);

    i = Locate(r, NumPoints, radial);
    if(i == -1) {
        //if radial is just greater than the maximum r, still draw it
        if((radial>r[NumPoints-1]&&(radial<1.1*r[NumPoints-1] -
0.1*r[NumPoints-2]))
        {
            i = NumPoints - 2;
            UD->sag1 = CubCoeff[i][0]*pow(r[i+1] - radial, 3)
                + CubCoeff[i][1]*pow(radial - r[i], 3)
                + CubCoeff[i][2]*(radial - r[i])
                + CubCoeff[i][3]*(r[i+1] - radial)
                + PolyTerms(radial);
        }
        else return (-1);
    }
    else{

```

```

        UD->sag1 = CubCoeff[i][0]*pow(r[i+1] - radial, 3)
                + CubCoeff[i][1]*pow(radial - r[i], 3)
                + CubCoeff[i][2]*(radial - r[i])
                + CubCoeff[i][3]*(r[i+1] - radial)
                + PolyTerms(radial);
    }
    UD->sag2 = 0.0;

    break;
case 4:
    /* ZEMAX wants a paraxial ray trace to this surface */
    /* x, y, z, and the optical path are unaffected, at least for this surface type */
    /* for paraxial ray tracing, the return z coordinate should always be zero. */
    /* paraxial surfaces are always planes with the following normals */

    l1 = 0;
    m1 = 0;
    n1 = 1;

    alpha=atan(dSag(r[0], 0));
    ln1 = 0;
    mn1 = -sin(alpha);
    nn1 = cos(alpha);

    Refract(1, 2, &l1, &m1, &n1, ln1, mn1, nn1);
    power = (FD->n2 - FD->n1)*dz2/(-r[0]*n1/m1);

/*  UD->ln = 0.0;
    UD->mn = 0.0;
    UD->nn = -1.0;
        dz = dSag(r[0], 0);
        dz2 = 6*CubCoeff[0][0]*(r[1] - r[0]);

    power = (FD->n2 - FD->n1)*dz2/pow(1 + dz*dz, 1.5);
*/
    if ((UD->n) != 0.0)
    {
        (UD->l) = (UD->l)/(UD->n);
        (UD->m) = (UD->m)/(UD->n);

        (UD->l) = (FD->n1*(UD->l) - (UD->x)*power)/(FD->n2);
        (UD->m) = (FD->n1*(UD->m) - (UD->y)*power)/(FD->n2);

    /* normalize */

```

```

(UD->n) = sqrt(1/(1 + (UD->l)*(UD->l) + (UD->m)*(UD->m) ));
/* de-paraxialize */
(UD->l) = (UD->l)*(UD->n);
(UD->m) = (UD->m)*(UD->n);
}
break;
case 5:
/* ZEMAX wants a real ray trace to this surface */
for(i = 1; i<9; i++) p[i] = FD->param[i];
beginIndex = 0;
endIndex = NumPoints - 1;
if((UD->n) == 0) return(FD->surf);
r1 = (UD->l)/(UD->n);
r2 = (UD->m)/(UD->n);

//Find out between which neighboring pair of sample points
//the ray hits the mirror

a = pow((UD->x) + (z[endIndex]+PolyTerms(r[endIndex]))*r1, 2)
      + pow((UD->y) +
(z[endIndex]+PolyTerms(r[endIndex]))*r2, 2);
b = r[endIndex] * r[endIndex];
if(a>b) return(FD->surf); //ray misses the surface
a = pow(UD->x + PolyTerms(r[beginIndex])*r1, 2)
      + pow(UD->y + PolyTerms(r[beginIndex])*r2, 2);
b = r[beginIndex] * r[beginIndex];
if(a<b) return(FD->surf); //ray misses the surface

else do {
middleIndex = (beginIndex + endIndex)/2;
a = pow((UD->x) +
(z[middleIndex]+PolyTerms(r[middleIndex]))*r1, 2)
      + pow((UD->y) +
(z[middleIndex]+PolyTerms(r[middleIndex]))*r2, 2);
b = r[middleIndex] * r[middleIndex];
if(a < b) endIndex = middleIndex;
else beginIndex = middleIndex;
}while(endIndex > beginIndex + 1);

//Bisection method to find out optical path t approximately
bIndex = 0;
eIndex = N;
dr = (r[beginIndex+1] - r[beginIndex])/(double)N;
do{

```

```

    mIndex = (bIndex + eIndex)/2;
    rt = (double)mIndex * dr + r[beginIndex];
    zt = Sag(rt, beginIndex) + PolyTerms(rt);
    a = pow((UD->x) + zt*r1, 2)
        + pow((UD->y) + zt*r2, 2);
    b = rt * rt;
    if(a < b) eIndex = mIndex;
    else bIndex = mIndex;
} while(eIndex > bIndex + 1);

//Newtonian method to find out optical path t more accurately
t0 = zt/UD->n;

do{
    x = sqrt(pow(UD->x + UD->l*t0, 2) + pow(UD->y + UD->m*t0,
2));
    f = (Sag(x, beginIndex)+PolyTerms(x)) - UD->z - UD->n*t0;
    df = (dSag(x, beginIndex)+dPolyTerms(x))/x*((UD->x + UD-
>l*t0)*UD->l +
                                (UD->y + UD->m*t0)*UD->m) - UD->n;
    t1 = t0 - f/df;
    t0 = t1;
} while(f>1e-10);

//calculate surface slope
t = t0;
(UD->x) += t*(UD->l);
(UD->y) += t*(UD->m);
(UD->z) += t*(UD->n);
radial = sqrt(pow(UD->x, 2) + pow(UD->y, 2));
slp = dSag(radial, beginIndex) + dPolyTerms(radial);
alpha=atan(slp);
phi=atan2((UD->y), (UD->x));
(UD->ln) = sin(alpha)*cos(phi);
(UD->mn) = sin(alpha)*sin(phi);
(UD->nn) = -cos(alpha);

    UD->path = t;

    if (Refract(FD->n1, FD->n2, &UD->l, &UD->m, &UD->n, UD->ln, UD->mn,
UD->nn)) return(-FD->surf);
    break;
case 6:

```

```

    /* ZEMAX wants the index, dn/dx, dn/dy, and dn/dz at the given x, y, z. */

    /* This is only required for gradient index surfaces, so return dummy values */
    UD->index = FD->n2;
    UD->dndx = 0.0;
    UD->dndy = 0.0;
    UD->dndz = 0.0;
    break;
case 7:
    /* ZEMAX wants the "safe" data. */
    /* this is used by ZEMAX to set the initial values for all parameters and extra data
    */
    /* when the user first changes to this surface type. */
    /* this is the only time the DLL should modify the data in the FIXED_DATA FD
    structure */
    DialogBox(hglobalInst, MAKEINTRESOURCE(IDD_DIALOG1), NULL,
    (DLGPROC)DlgProc);
    if(bData){
        fstream infile(filepath, ios::in|ios::binary);
        infile>>NumPoints;
        infile>>normal;
        infile>>derv0;
        infile>>dervN_1;
        for(i = 0; i<NumPoints; i++){
            infile>>r[i];
            infile>>z[i];
        }

        if(r[0]<0){
            derv0 *= -1;
            dervN_1 *= -1;
        }

        for(i = 0; i<NumPoints; i++){
            r[i] = (r[i]<0)? -r[i] : r[i];
        }

        FD->xdata[1] = (double)NumPoints;
        FD->xdata[2] = normal;
        FD->xdata[3] = derv0;
        FD->xdata[4] = dervN_1;

        for (i = 5; i <= 200; i++) {
            if(i%2) FD->xdata[i] = r[(i/2-2)*(NumPoints-1)/97];
        }
    }

```

```

        else FD->xdata[i] = z[(i/2-3)*(NumPoints-1)/97];
    }

    for(i = 0; i<NumPoints; i++) pCubCoeff[i] = &CubCoeff[i][0];
    CalCoeff(pCubCoeff, r, z, NumPoints, normal, derv0, dervN_1);

    fstream outfile("E:\\Zemax\\Data Files\\cub.dat", ios::out);
    for(i = 0; i<NumPoints-1; i++){
        outfile<<i<<" ";
        outfile<<setw(11)<<setprecision(8)<<CubCoeff[i][0]<<" ";
        outfile<<setw(11)<<setprecision(8)<<CubCoeff[i][1]<<" ";
        outfile<<setw(11)<<setprecision(8)<<CubCoeff[i][2]<<" ";
        outfile<<setw(11)<<setprecision(8)<<CubCoeff[i][3]<<" ";
        outfile<<"\n";
    }
}

else for (i = 1; i <= 200; i++) FD->xdata[i] = 0.0;
        for (i = 1; i <= 8; i++) FD->param[i] = 0.0;

        break;
    }
return 0;
}

int Refract(double thisn, double nextn, double *l, double *m, double *n, double ln,
double mn, double nn)
{
double nr, cosi, cosi2, rad, cosr, gamma;
if (thisn != nextn)
    {
nr = thisn / nextn;
cosi = fabs((*l) * ln + (*m) * mn + (*n) * nn);
cosi2 = cosi * cosi;
if (cosi2 > 1) cosi2 = 1;
rad = 1 - ((1 - cosi2) * (nr * nr));
if (rad < 0) return(-1);
cosr = sqrt(rad);
gamma = nr * cosi - cosr;
(*l) = (nr * (*l)) + (gamma * ln);
(*m) = (nr * (*m)) + (gamma * mn);
(*n) = (nr * (*n)) + (gamma * nn);
    }
return 0;
}

```



```

}

//Dialogbox invoked when cubic splines surface type is selected
//Ask whether data points are provided for the purpose of sag optimization
//Data points should be stored in a file in the directory "E:\zemax\data files\..."
//Up to 1000 data points can be specified.
LRESULT CALLBACK DlgProc (HWND hDlg, UINT message, WPARAM wParam,
LPARAM lParam)
{
    WORD wID;

    wID = LOWORD(wParam);
    switch(message)
    {
        case WM_INITDIALOG:

            SetWindowText(hDlg, "Specify data file path");
            strcpy(filepath, "E:\\Zemax\\Data Files\\Surfl.dat");
            SetDlgItemText(hDlg, IDC_EDIT1, filepath);
            return TRUE;

        case WM_COMMAND:
            switch(wID)
            {
                case IDOK:
                    GetDlgItemText(hDlg, IDC_EDIT1, filepath,
sizeof(filepath)-1);

                    EndDialog(hDlg, wParam);
                    bData = TRUE;
                    return TRUE;
                case IDCANCEL:
                    bData = FALSE;
                    EndDialog(hDlg, wParam);
                    return TRUE;
            }
    }

    return FALSE;
}

```

```

void CalCoeff(double ** Coeff, double * r1, double * z1, int Num, double norm, double
d0, double dN_1)
{
    ofstream outfile("E:\\Zemax\\Data Files\\cub1.dat", ios::out);
    int i;
    double h[1000], df[1000];
    double a[1000], b[1000], c[1000], u[1000], rr[1000], gam[1000];
    double bet;

    for(i = 0; i < Num-1; i++) df[i] = z1[i+1] - z1[i];
    for(i = 0; i < Num-1; i++) h[i] = r1[i+1] - r1[i];
    //set matrix lower diagonal matrix elements;
    a[Num-1] = h[Num-2];
    for(i = 1; i < Num-1; i++) a[i] = h[i-1];

    //set matrix diagonal elements;
    b[0] = 2*h[0];
    b[Num-1] = 2*h[Num-2];
    for(i = 1; i < Num-1; i++) b[i] = 2*(h[i] + h[i-1]);

    //set matrix upper diagonal matrix elements;
    c[0] = h[0];
    for(i = 1; i < Num-1; i++) c[i] = h[i];

    //set the right hand side elements of the equations
    rr[0] = 6*(df[0]/h[0] - d0);
    rr[Num-1] = 6*(dN_1 - df[Num-2]/h[Num-2]);
    for(i = 1; i < Num-1; i++) rr[i] = 6*(df[i]/h[i] - df[i-1]/h[i-1]);

    for(i = 0; i < Num-1; i++){
        outfile << i << " ";
            outfile << setw(11) << setprecision(8) << h[i] << " ";
            outfile << setw(11) << setprecision(8) << df[i] << " ";
            outfile << setw(11) << setprecision(8) << a[i] << " ";
            outfile << setw(11) << setprecision(8) << b[i] << " ";
            outfile << setw(11) << setprecision(8) << c[i] << " ";
            outfile << setw(11) << setprecision(8) << rr[i] << " ";
            outfile << "\n";
    }
    //solve the tridiagonal equations: algorithm copied from Numerical recipes
    bet = b[0];
    u[0] = rr[0]/bet;
    for(i = 1; i < Num; i++){

```

```

        gam[i] = c[i-1]/bet;
        bet = b[i] - a[i]*gam[i];
        u[i] = (rr[i] - a[i]*u[i-1])/bet;
    }
    for(i = Num-2; i >= 0; i--) u[i] = u[i] - gam[i+1]*u[i+1];

    //calculating cubic spline fitting coefficients
    for(i = 0; i<Num-1; i++){
        Coeff[i][0] = u[i]/h[i]/6;
        Coeff[i][1] = u[i+1]/h[i]/6;
        Coeff[i][2] = z1[i+1]/h[i] - u[i+1]*h[i]/6;
        Coeff[i][3] = z1[i]/h[i] - u[i]*h[i]/6;
    }
}

// A double array xx[N], given a value x, find j where x is between xx[j]
// and xx[j+1], return j; if x is out of range, return -1.
int Locate(double *xx, int N, double x)
{
    int begin, middle, end;
    begin = 0;
    end = N - 1;

    //out of range, return -1;
    if((xx[N-1]>xx[0])&&((x>xx[N-1])||(x<xx[0]))) return -1;
    if((xx[N-1]<xx[0])&&((x<xx[N-1])||(x>xx[0]))) return -1;

    do {
        middle = (begin + end)/2;
        if((xx[N-1]>xx[0]) == (x>xx[middle])) begin = middle;
        else end = middle;
    } while((end - begin)>1);

    return begin;
}

//calculate sag given a coordinate x and the range index i: x[i]<=x<=x[i+1]
double Sag(double x, int i)
{
    double f;
    f = CubCoeff[i][0]*pow(r[i+1] - x, 3)
        + CubCoeff[i][1]*pow(x - r[i], 3)

```

```

+ CubCoeff[i][2]*(x - r[i])
+ CubCoeff[i][3]*(r[i+1] - x);

    return f;
}

//calculate derivative given coordinate x and the range index i
double dSag(double x, int i)
{
    double g;
    g = -3*CubCoeff[i][0]*pow(r[i+1] - x, 2)
        + 3*CubCoeff[i][1]*pow(x - r[i], 2)
        + CubCoeff[i][2]
        - CubCoeff[i][3];

    return g;
}

//calculate the added sag
double PolyTerms(double radial)
{
    int i;
    double asag;
    asag = p[1];
    for(i = 2; i<=8; i++){
        asag += p[i]*pow(radial/r[NumPoints-1], (double)(i-1)*2);
    }
    return asag;
}

//calculate the added slope
double dPolyTerms(double radial)
{
    int i;
    double dasag;
    dasag = 0;
    for(i = 2; i<=8; i++){
        dasag += 2*(double)(i-1)*p[i]*pow(radial/r[NumPoints-1], (double)(i-
1)*2 - 1)/r[NumPoints-1];
    }
    return dasag;
}

```

CubSpln.rc

```

//Microsoft Developer Studio generated resource script.
//
#include "resource.h"

#define APSTUDIO_READONLY_SYMBOLS
////////////////////////////////////
//
// Generated from the TEXTINCLUDE 2 resource.
//
#include "afxres.h"

////////////////////////////////////
#undef APSTUDIO_READONLY_SYMBOLS

////////////////////////////////////
// English (U.S.) resources

#if !defined(AFX_RESOURCE_DLL) || defined(AFX_TARG_ENU)
#ifdef _WIN32
LANGUAGE LANG_ENGLISH, SUBLANG_ENGLISH_US
#pragma code_page(1252)
#endif // _WIN32

////////////////////////////////////
//
// Dialog
//

IDD_DIALOG1 DIALOG DISCARDABLE 0, 0, 187, 78
STYLE DS_MODALFRAME | WS_POPUP | WS_CAPTION | WS_SYSMENU
CAPTION "Dialog"
FONT 8, "MS Sans Serif"
BEGIN
    DEFPUSHBUTTON "OK",IDOK,130,7,50,14
    PUSHBUTTON "Cancel",IDCANCEL,130,24,50,14
    LTEXT "Specify the data file path:",IDC_STATIC,14,20,109,8
    EDITTEXT IDC_EDIT1,15,46,156,14,ES_AUTOHSCROLL
END

////////////////////////////////////
//

```



```

#ifndef APSTUDIO_INVOKED
//
// Generated from the TEXTINCLUDE 3 resource.
//

```

```

//
//
#endif // not APSTUDIO_INVOKED

```

Dialogdef.h

```
#include <afxwin.h>
```

```

class CDatafileDialog : public CDialog
{
public:
    CDatafileDialog(UINT id, CWnd *pWnd) : CDialog::CDialog(id, pWnd){};
    void DoDataExchange(CDataExchange * pDX);
    DECLARE_MESSAGE_MAP()
};

```

Resource.h

```

//{{NO_DEPENDENCIES}}
// Microsoft Developer Studio generated include file.
// Used by CunSpln.rc
//
#define IDD_DIALOG1          101
#define IDC_EDIT1            1000
#define IDC_STATIC           -1

// Next default values for new objects
//
#ifdef APSTUDIO_INVOKED
#ifndef APSTUDIO_READONLY_SYMBOLS
#define _APS_NEXT_RESOURCE_VALUE        102
#define _APS_NEXT_COMMAND_VALUE        40001
#define _APS_NEXT_CONTROL_VALUE        1001
#define _APS_NEXT_SYMED_VALUE          101

```

#endif
#endif

REFERENCES

1. Alonso, M. A. and G. W. Forbes, "generalization of Hamilton's formalism for geometrical optics", *J. Opt. Soc. Am. A*, 12, 2744-2752 (1995).
2. Born, M. and E. Wolf, *Principles of Optics*, MacMillan, New York (1964).
3. Buchdahl, H. A., *An Introduction To Hamiltonian Optics*, Cambridge University Press (1970).
4. Chang, M., "Pupil aberration in modular zoom lens design", PhD dissertation, University of Arizona.
5. Dyson, J., "Unit Magnification Optical System without Seidel Aberrations," *J. Opt. Soc. Am.* **49**, 713-716, (1959).
6. Forbes, G. W. and B. D. Stone, "Hamilton's angle characteristic in closed form for generally configured conic and toric interfaces", *J. Opt. Soc. Am. A*, 10, 1270-1278 (1993).
7. Forbes, G. W. and B. D. Stone, "Restricted characteristic functions for general optical configurations", *J. Opt. Soc. Am. A*, 10, 1263-1269 (1993).
8. Forbes, G. W. and J. K. Wallace, "Can the bounds to system performance in geometrical optics be attained?" *J. Opt. Soc. Am. A*, 12, 2064-2071 (1995).
9. Johnson, L. W. and R. D. Riess, *Numerical Analysis*, Chap. 5, pp. 200-209, Addison-Wesley (1977).
10. Kingslake, R., *Lens Design Fundamentals*, Academic, New York (1978).
11. Korsch, D., *Reflective optics*. Academic Press, Boston (1991).
12. Luneburg, R. K., *Mathematical Theory of Optics*, University of California Press, Berkeley and Los Angeles (1966).
13. Malacara, D. and Z. Malacara, *Handbook of Lens design*, Marcel Dekker, Inc. (1994).
14. Mansuripur, M., "Abbe's Sine Condition", *OPN* Feb, 1998, 56-60.

REFERENCES -- *continued*

15. Mertz, L., *Excursions in Astronomical Optics*, Springer, New York (1996).
16. Offner, A., "New Concepts in Projection Mask Aligners," *Opt. Eng.* **14**, 130-132, (1975).
17. Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes*, Chap. 2, pp. 40-41, Cambridge University Press (1987).
18. Sasian, J. M., "How to approach the design of a bilateral symmetric optical system", *Opt. Eng.* **33**(6), 2045-2061 (1994).
19. Shack, R. V., "Advanced aberration theory" class notes, Optical Sciences Center, the University of Arizona.
20. Shack, R. V., "Introduction to aberration theory" class notes, Optical Sciences Center, the University of Arizona.
21. Shannon, R. R., *The Art and Science of Optical Design*, Cambridge University Press (1997).
22. Smith, W. J., *Modern Lens Design: A resource Manual*, McGraw-Hill, Inc. (1992).
23. Smith, W. J., *Modern Optical Engineering*, Chap 3, McGraw-Hill Inc. (1990).
24. Stone, B. D., "Analog of third-order methods in the design of asymmetric systems", *OSA Proceedings of the International Lens Design*, G. W. Forbes Eds (1994).
25. Stone, B.D. and G. W. Forbes, "Characterizations of first-order optical properties for asymmetric systems", *J. Opt. Soc. Am. A*, **9**, 478-489 (1992).
26. Stone, B.D. and G. W. Forbes, "First-order layout of asymmetric systems composed of three spherical mirrors", *J. Opt. Soc. Am. A*, **9**, 110-120 (1992).
27. Stone, B.D. and G. W. Forbes, "Foundations of first-order layout for asymmetric systems: an application of Hamilton's methods", *J. Opt. Soc. Am. A*, **9**, 96-109 (1992).
28. Sygne, J. L., "Hamilton's Method in Geometrical Optics", *J. Opt. Soc. Am.* **27** 75-82 (1937).

REFERENCES -- *continued*

29. Walther, A., "Eikonal theory and computer algebra II", J. Opt. Soc. Am. A, 13, 1763-1765(1989).
30. Walther, A., "Eikonal theory and computer algebra", J. Opt. Soc. Am. A, 13, 523-531(1986).
31. Walther, A., "Irreducible aberrations of a lens used for a range of magnifications", J. Opt. Soc. Am. A, 6, 415-422 (1989).
32. Walther, A., "Numerical techniques in eikonal function theory", J. Opt. Soc. Am. A, 5, 511-515 (1988).
33. Welford, W. T., *Aberrations of Optical Systems*, Adam Hilgler Ltd, Bristol & Boston (1986).
34. ZEMAX optical design software. Focus Software Inc., Tucson, Arizona.
35. Zhao, C., and J. H. Burge, "The criteria for correction of the quadratic field-dependent aberrations", Submitted to JOSA A.
36. Zhao, C. and J. H. Burge, "Comparison of exact pupil astigmatism conditions with Seidel approximation", to be submitted for publication.
37. Zhao, C. and J. H. Burge, "Application of the Pupil Astigmatism Criteria in optical design", to be submitted for publication.
38. Zhao, C. and J. H. Burge, "The Pupil Astigmatism Criteria for plane symmetric systems", to be submitted for publication.