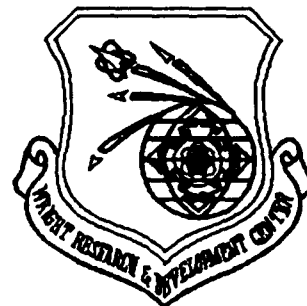


WRDC-TR-90-3058
Volume I

AD-A235 931



**OPTICAL ANALYSIS OF AIRCRAFT TRANSPARENCIES
(OPTRAN)
VOLUME I: THEORETICAL MANUAL**

J. Loomis
J. Fielman
University of Dayton Research Institute
300 College Park Avenue
Dayton, Ohio 45469-0110

October 1990

Final Report for Period December 1988 - May 1990

Approved for public release; distribution unlimited

DTIC
ELECTE
MAY 24 1991
S B D

FLIGHT DYNAMICS LABORATORY
WRIGHT RESEARCH AND DEVELOPMENT CENTER
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6553

91-00246




91 5 22 088

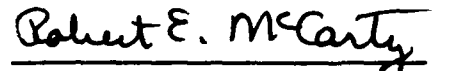
NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publications.


RICHARD A. SMITH
Aerospace Engineer


ROBERT E. MCCARTY, Supervisor
Aircrew Protection Branch

FOR THE COMMANDER


RICHARD E. COLCLOUGH, JR.
Chief
Vehicle Subsystems Division

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify WRDC/FIVR, WPAFB, OH 45433-6553 to help us maintain a current mailing list.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188		
1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b RESTRICTIVE MARKINGS			
2a SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.			
2b DECLASSIFICATION / DOWNGRADING SCHEDULE						
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UDR-TR-90-63			5 MONITORING ORGANIZATION REPORT NUMBER(S) WRDC-TR-90-3058, Vol I			
6a. NAME OF PERFORMING ORGANIZATION University of Dayton Research Institute		5b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION Flight Dynamics Directorate (WL/FIVR) Wright Laboratories			
6c. ADDRESS (City, State, and ZIP Code) 300 College Park Dayton, Ohio 45469-0110			7b. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6553			
8a. NAME OF FUNDING / SPONSORING ORGANIZATION		8b OFFICE SYMBOL (if applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-86-C-3414			
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS			
			PROGRAM ELEMENT NO 62291F	PROJECT NO 2402	TASK NO 03	WORK UNIT ACCESSION NO 60
11 TITLE (Include Security Classification) Optical Analysis of Aircraft Transparencies (OPTRAN) Volume I: Theoretical Manual						
12 PERSONAL AUTHOR(S) J. Loomis and J. Fielman						
13a. TYPE OF REPORT Final		13b TIME COVERED FROM 12-88 TO 5-90		14. DATE OF REPORT (Year, Month, Day) 31 October 1990		
15. PAGE COUNT 62						
16 SUPPLEMENTARY NOTATION						
17. COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)			
FIELD	GROUP	SUB-GROUP				
19. ABSTRACT (Continue on reverse if necessary and identify by block number)						
<p>This document describes the theory of the Optical Transmission (OPTRAN) code developed to predict the optical performance of aircraft aircrew enclosure transparency designs subjected to operational loads. The theory on which the code is based is described. The OPTRAN ray trace code accounts for thermal and stress optical effects. Orthotropic indices of refraction are computed throughout the transparency volume. Geometry is defined by parametric cubic solids. The refraction of incoming light is computed as a function of parametric solid surface outward normals. Angular deviation, polarization effects, and lensing are computed for each ray.</p>						
20 DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED			
22a NAME OF RESPONSIBLE INDIVIDUAL RICHARD A. SMITH			22b TELEPHONE (Include Area Code) (513) 255-2516		22c OFFICE SYMBOL WL/FIVR	

Foreword

This report was prepared by the University of Dayton Research Institute under Rockwell International PO. #L9FM-60048-W-439, Project Title, "Transparency Optical Analysis Capability Development" under United State Air Force Contract F33615-86-C-3414, The project was administered by the Wright Research and Development Center, Wright-Patterson Air Force Base, Ohio. Mr. Richard A. Smith, AFWAL/FIVR, was Laboratory Project Engineer.

This is the final report submitted under the University of Dayton's efforts and documents the software developed in the period from December 1988 to May 1990. Project supervision and technical assistance was provided through the Aerospace Mechanics Division of UDRI, Mr. Dale Whiford, Supervisor. Project manager for this effort was Mr. Blaine E. West, and the Principal Investigator was Mr. John W. Fielman, who was also responsible code development and code interfaces. Dr. John S. Loomis was the primary contributor of the optics theory and ray trace analysis code. Dr. Robert A. Brockman made strategic modifications to MAGNA and the MAGNA interface codes without which the effort could not have been completed.

This report is published in two volumes. Volume I, OPTRAN Theoretical Manual, describes the mathematical theory and working equations on which the OPTRAN code is based. OPTRAN was developed to predict the optical performance of current and future canopy designs. Orthotropic optical effects are computed as a function of temperatures and stresses predicted by finite element codes. Volume II, OPTRAN User's Manual, describes the operation of OPTRAN and the finite element codes with which OPTRAN is interfaced. The operation of pre- and post-processor software is also described.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Contents

1	Introduction	1
1.1	Optical Raytrace	3
1.2	Optical Analysis	4
1.3	Polarization	6
2	Raytracing Parametric Surfaces	11
2.1	Overview	11
2.2	Hyperpatches	12
2.3	Parametric Surfaces	13
2.4	Surface Intersection	14
2.5	Ray Refraction	15
2.6	Refraction and Reflection Coefficients	16
2.7	Mueller Matrices	18
2.8	Output Data	19
3	Nonlinear Least-Squares Optimization	20
3.1	Mathematic Preliminaries	20
3.2	Method of Least-Squares	22
3.2.1	Homogeneous Matrix Formulation	23
3.2.2	Nonlinear Effects	23
3.2.3	Damped Least-Squares	24
3.3	Method of QU Factorization	24

4	Differential Rays	30
4.1	Surface Intersection	30
4.2	Ray Refraction	31
4.3	Differential Trace of Parametric Surface	31
5	Anisotropic Refractive Index Ellipsoid	33
5.1	Orthotropic Indices of Refraction	33
5.2	Geometric Transformation	34
5.3	Stress Birefringence	35
5.4	Tensor Rotation Transformation	37
5.5	Principal Axes of Projected Ellipse	38
6	Optical Waves in Anisotropic Materials	40
6.1	Light Propagation in Anisotropic Materials	40
6.2	Double Refraction	47
6.3	Reflection and Refraction Coefficients	49
6.3.1	Reflection and Refraction in Isotropic Media	51

List of Figures

1.1	Angular deviation of a point in space by aircraft transparency	5
1.2	Diagram illustrating the definition of angular coordinates	5
1.3	Using a grid to show distortion and angular deviation	7
1.4	Examples of polarization states	9
1.5	Angles used to characterize the polarization ellipse	10
6.1	Construction of the \mathbf{D} vectors belonging to a wave normal \mathbf{k} [1].	41
6.2	Intersection of the normal surface with xz plane for (a) biaxial crystals, (b) positive uniaxial crystals, and (c) negative uniaxial crystals [3].	44
6.3	Orientation of rays and waves in a uniaxial crystal [2].	44
6.4	Optical path for rays and wavefronts.	46
6.5	A light beam with two orthogonal field components traversing a calcite principal section [2].	46
6.6	Double refraction at the boundary of an anisotropic material.	47
6.7	Ordinary reflection at the boundary of an anisotropic material.	48
6.8	Extraordinary reflection at the boundary of an anisotropic material. . . .	49
6.9	Refraction and reflection at boundary between two anisotropic materials.	50

List of Tables

6.1	Typical refractive indices of some crystals [3].	43
-----	--	----

Section 1

Introduction

OPTRAN is a raytrace code which evaluates the optical quality of aircraft transparencies subjected to operation load conditions. This volume describes the theoretical background on which the code is based.

The raytrace optical code is interfaced to finite element thermal and stress codes to permit the effects of operational loads to be modeled. Thermal, displacement, and stress field definition data computed by the finite element codes are input to the optics code. This information is required to compute the orthotropic indices of refraction throughout the material volume of the aircraft transparency. This computation is performed at each step along the propagation path of each ray.

The optics code tracks rays of various wavelengths through the transparency. The deformed geometry generated by the stress analysis is used to determine angles of reflection and refraction at transparency layer boundaries. Birefringent indices of refraction are computed as a function of material, temperature, and stress state at the refracting surfaces and within the transparency material.

Post-processing graphics codes display the angular deviation, transmittance, and polarization effects over specified regions of the transparency. Plots of displacement vectors and deformed grids can be also generated.

The PATRAN [1] finite element pre- and post-processing software provides the common interface between the thermal and stress finite element codes and the optical analysis code. PATRAN is a software product of the PATRAN Division of PDA Engineering, Costa Mesa, CA. PATRAN software is available for many computer systems and offers device drivers for many interactive graphics terminals. PATRAN provides excellent tools for defining the model geometries and generating graphic displays of the

models. It is widely used, well supported, and interfaced to many of the more popular finite element codes, thus offering the opportunity of using other analysis tools in conjunction with OPTRAN.

Isoparametric cubic hyperpatches defined by the same mathematical formulation as those used by PATRAN [1] (Chapter 37) are used in OPTRAN. They are also used to model the deformed transparency layer solid geometry and map the temperature and stress parameters within the transparency material layers. The temperature (T) and the six orthogonal stress parameters (σ_{xx} σ_{yy} σ_{zz} τ_{xy} τ_{yz} τ_{zx}) are required at each incremental step along the optical ray trace paths to compute continuously varying orthotropic indices of refraction. Displacements are critical in determining reflected and refracted optical ray path directions at layer boundaries.

Temperature, pressure, and density fields in the surrounding air stream can also be mapped using a series of isoparametric hyperpatches. These parameters determine index of refraction in the atmosphere.

The mathematical formulation of hyperpatches is presented in detail in Chapter 37 of the PATRAN Plus User's Manual [1]. The hyperpatch formulation is that of a cubic solid (64 node) isoparametric finite element. Coordinates and other data parameters are mapped within the hyperpatch with the same parametric equations.

For raytracing, the entrance and exit surfaces of each part must be identified. Surfaces must be numbered sequentially from the outside of the aircraft to the eye. The eye position and a set of pilot reference axes must also be defined. Rays are specified by direction angles with respect to the pilot coordinate system. Rays can also be specified indirectly by defining a mesh of nodes over the first entrance surface. The 3D coordinates of these nodes are used to generate ray directions.

The OPTRAN code accounts for three-dimensional orthotropic optical effects. An orthotropic index of refraction ellipsoid is computed as a function of the stress and temperature values. Orthotropic effects can result from either orthotropic material properties or from form birefringence caused by stress optic effects.

Optical material properties include orthotropic indices of refraction, orthotropic temperature coefficients of the indices of refraction, and a six by six matrix of stress optic coefficients. Orthotropic optical material properties input to OPTRAN are defined with respect to material axes.

The material axes are defined with respect to reference axes. The reference axes can be either the global coordinate axes or axes defined by the derivatives of the global

coordinates with respect to the geometric hyperpatch parametric variables. This latter reference axis option allows the optic axis orientation to vary with the curvature of the transparency. By default the material axes are aligned with the reference axes and the global coordinate axes are the reference axes.

The following method is used to find the ray from a known target location that intersects the eye point. First we aim a ray from the target to the eye and trace an actual ray until the ray intersects the eye plane. Then we trace differential rays (close to the original ray), differing first in azimuth and then in elevation. The intersection of these rays at the eye plane (XY plane in pilot coordinates) are used to generate a first-order matrix that can be solved to give a correction to the original ray, that is a new ray that now intersects the eye point.

An extensive list of variables are generated at each intersection point. These include the hyperpatch ID and face number, material ID, hyperpatch parameters (u , v , w), corresponding 3D coordinates (x , y , z), the direction of the surface normal, the direction of the refracted ray, a reference polarization direction, polarization and transmittance arrays, and auxiliary variables needed to generate differential rays. A detailed surface-by-surface list of this information can be generated on the output listing.

There are three output files generated by OPTRAN. The first is the output listing, which contains a copy of the input parameters, detailed ray trace information, error messages, and summary tables. The second is a PATRAN nodal results file, which contains 14 columns of summary data. The PATRAN nodal results file can be used in PATRAN to produce a variety of 3D plots. The third file is an optical results file, which contains outline segments that help define the field of view and raytrace information on a uniform grid of azimuth/elevation variables. The optical results file is used to generate two-dimensional grid distortion plots, angular deviation fields, and polarization ellipses.

1.1 Optical Raytrace

Rays are represented as straight lines in space. Rays are refracted at the point they intersect an optical surface. The two operations involved in raytracing, therefore, are finding the intersection of a ray with the surface and refracting the ray at the surface. For parametric surfaces, ray intersection is an iterative procedure, requiring a two-dimensional nonlinear optimization. Calculating the direction of propagation of the refracted ray is also an iterative process, since the index of refractive varies with direction.

Finding the first intersection on an entrance surface requires a search over available patches. For an intersection point to be valid, the parametric variables corresponding

to the intersection point must lie within the bounds of $0 \leq (u, v, w) \leq 1$. Once the entrance hyperpatch face has been identified, the ray may be traced through the part without additional searching because the PATRAN geometry file identifies adjacent hyperpatches. After a ray exits a part, however, another search of a list of patches may be required to find the entrance into the next part.

Stress and temperature affect light rays in two ways. First, the surfaces of the transparency deform as a result of stress and temperature changes, so that the geometry of the transparency changes. Second, the index of refraction of the transparency layers depends on both temperature and stress. OPTRAN uses the isoparametric surface geometry definitions from PATRAN to locate boundaries for ray refraction and reflection. The results of finite-element heat and stress programs define the volumetric temperature and stress states of the transparency, from which the index ellipsoid is calculated.

At an interface between two dielectric materials, a plane of incidence is defined by the normal vector to the surface and the direction vector of the incident ray. The light ray is split into reflected and refracted rays, propagating in the plane of incidence. Snell's law determines the direction of propagation. The polarization of the incident ray, defined by the electric field vector, is decomposed into components parallel and perpendicular to the plane of incidence. The Fresnel equations determine the reflectance and transmittance of each polarization component.

Within a birefringent material, the electric field vector must be decomposed into components parallel to the principal axes of the dielectric material, as determined by the dielectric tensor. The components propagate with slightly different phase shifts causing one polarization state to be retarded in phase with respect to the other.

1.2 Optical Analysis

Rays are traced from the outer world to the eye point, as shown in Figure 1.1. Azimuth and elevation angles are used to specify the actual direction \mathbf{k} of a target point in the outer world. This is the dotted line drawn from the object to the eye in Figure 1.1. The direction of the exit ray \mathbf{k}' shows the apparent direction of the target, as shown by the dotted line from the image to the eye. The difference between the apparent direction and the actual direction is the angular deviation. The angular deviation $\delta\mathbf{k}$ is calculated from $\delta\mathbf{k} = \mathbf{k}' - \mathbf{k}$. Deviation causes objects to be seen at other than their true direction from the observer.

In OPTRAN, 2D direction coordinates of a vector are defined with respect to the pilot reference system. Note that direction vectors point toward the eye, and that zero

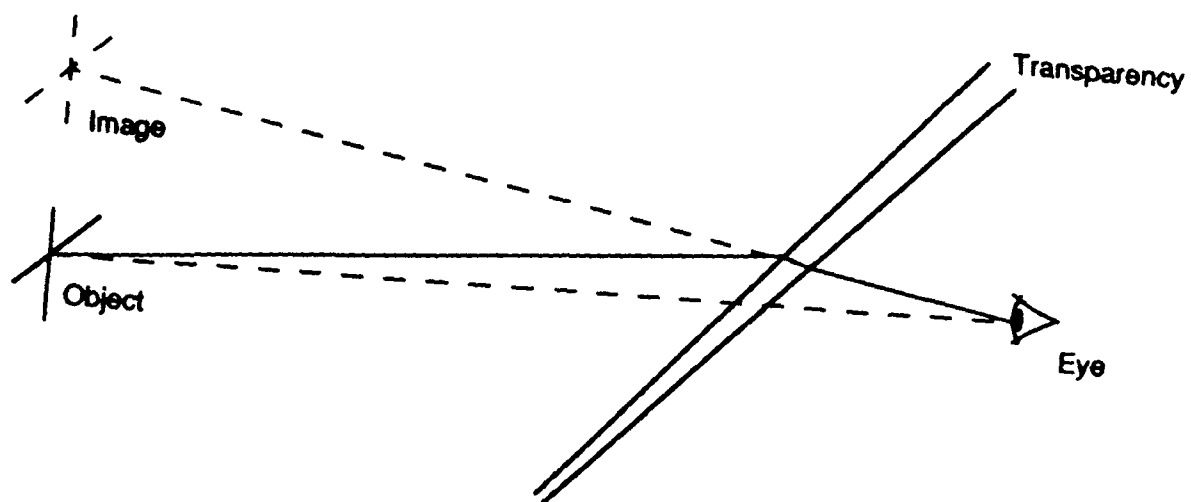


Figure 1.1: Angular deviation of a point in space by aircraft transparency

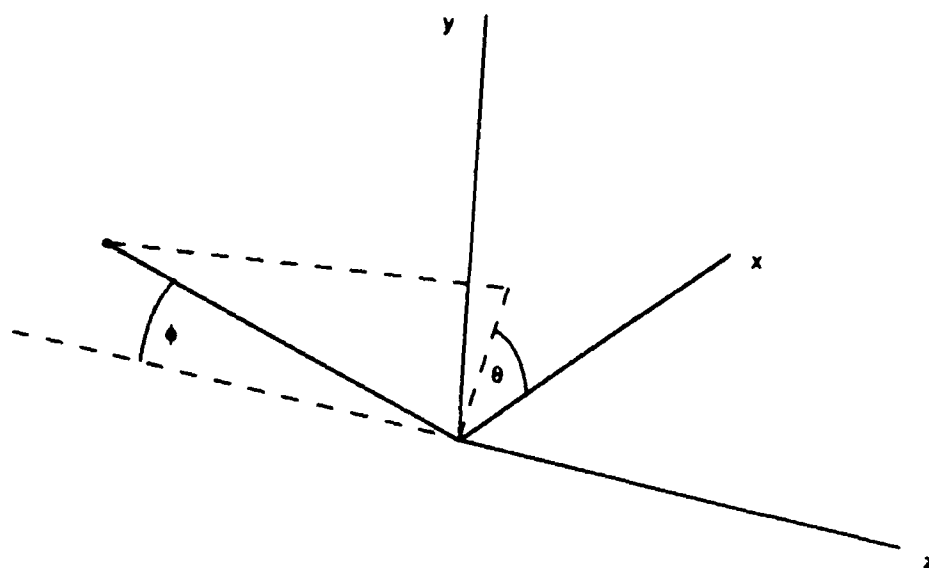


Figure 1.2: Diagram illustrating the definition of angular coordinates

angle corresponds to a direction parallel to the pilot z-axis. For this case the point of regard (object or image point) lies along the negative z-axis, directly in front of the pilot.

The magnitude of direction coordinates is the angle ϕ in degrees between the vector and the pilot z-axis, as shown in Figure 1.2. The orientation of the angular coordinates is obtained by projecting the direction vector onto the xy-plane and finding the angle θ with respect to the x-axis, as shown in Figure 1.2. Then azimuth and elevation direction angles A_z and A_y are found from

$$A_z = \phi \cos \theta$$

$$A_y = \phi \sin \theta$$

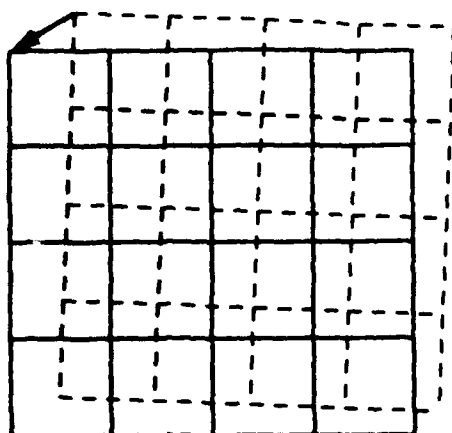
The magnitude of direction angles extends from 0 to 180°. A direction angle of 180° refers to a point directly behind the pilot. For the 180° direction the orientation θ is indeterminate. The horizontal or x-component of direction is called azimuth, and will be positive for points of regard to the right of the pilot and negative to the left. The vertical or y-component of direction is called elevation, and will be positive for points above the horizon (xz-plane) and negative below.

Angular deviation is the most direct measure of optical quality of a transparency. One method of demonstrating angular deviation is to show how a rectilinear grid in angle coordinates is distorted by the transparency. The data from a square array of rays, equally spaced in azimuth and elevation, can be displayed in single cross-sections of elevation or azimuth error or as deformed grids, emulating the typical grid board photograph.

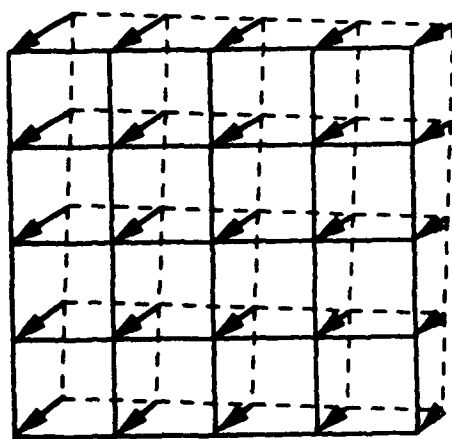
Grid distortion is shown in Figure 1.3 (a). The undistorted grid is shown in dashed lines and the distorted grid by solid lines. The arrow is the angular deviation for a single point on the grid. If arrows are drawn at every grid point, then the representation shown in Figure 1.3 (b) is obtained. Finally, by removing the grids we can produce a plot of angular deviation vectors as shown in Figure 1.3 (c). Both grid distortion and angular deviation plots can be produced from OPTRAN optical results files.

1.3 Polarization

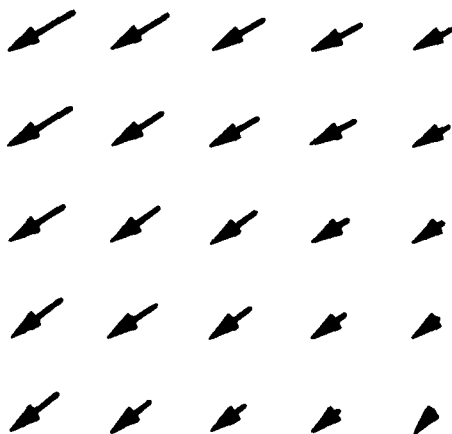
Characterizing the state of polarization is discussed in Born and Wolf [3] pp. 30-32, Hecht [2] pp. 321-326, and in Azzam and Bashara [9]. The different possible states of



(a)



(b)



(c)

Figure 1.3: Using a grid to show distortion and angular deviation

polarization of monochromatic light can be represented by a set of four real quantities, called the Stokes parameters, denoted by the vector

$$\mathbf{s}^\dagger = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \end{pmatrix},$$

each component of which has the dimensions of irradiance. The first term, s_0 , gives the total irradiance of the light wave and therefore is always positive. The next three terms give the difference between the irradiance of three different sets of orthogonal polarization states, and can therefore be either positive, negative, or zero.

Suppose we had our choice of ideal linear and circular polarizers. Let I_0 be the total irradiance of the wave and $I_x, I_y, I_{45}, I_{-45}, I_r, I_l$ be the intensities transmitted by the corresponding ideal polarizer. Then

$$\begin{aligned} s_0 &= I_0 = I_x + I_y = I_{45} + I_{-45} = I_r + I_l \\ s_1 &= I_x - I_y \\ s_2 &= I_{45} - I_{-45} \\ s_3 &= I_r - I_l \end{aligned}$$

where s_1 gives the difference between the irradiance of x and y linear polarization states, s_2 represents the preference for polarization between $+45^\circ$ and -45° , and s_3 is the difference between right- and left-circular polarizations.

For unpolarized light, there is no preference for any particular polarization so that $s_1 = s_2 = s_3 = 0$ and the Stokes vector has the simple form

$$\mathbf{s}^\dagger = \begin{pmatrix} s_0 & 0 & 0 & 0 \end{pmatrix}.$$

The Stokes parameters of a totally polarized wave satisfy the condition

$$s_0^2 = s_1^2 + s_2^2 + s_3^2$$

The general case can be treated by splitting the wave into two components, one totally polarized and one unpolarized. Let

$$s_p = \sqrt{s_1^2 + s_2^2 + s_3^2}.$$

Then

$$\begin{aligned} \mathbf{s}_{un.}^\dagger &= \begin{pmatrix} (s_0 - s_p) & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{s}_{tp.}^\dagger &= \begin{pmatrix} s_p & s_1 & s_2 & s_3 \end{pmatrix} \end{aligned}$$

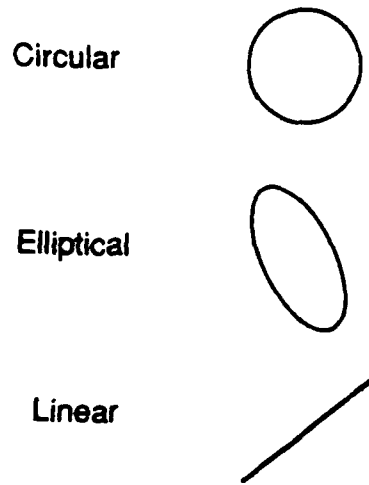


Figure 1.4: Examples of polarization states

We can also define the degree of polarization D_p as the ratio of the irradiance of the totally polarized component to the total irradiance

$$D_p = \frac{s_p}{s_0}.$$

The degree of polarization varies from zero for unpolarized light to unity for totally polarized light, with intermediate values for partially polarized light.

Polarization has a simple geometric interpretation as an ellipse. The shape and orientation of the ellipse relate to the polarization state, as shown in Figure 1.4. The size of the ellipse relates to the relative brightness of the polarized component of the transmitted ray. The shape of the ellipse may be characterized by two angles, the ellipticity angle ϵ and the orientation angle θ , as shown in Figure 1.5. The relationship between the Stokes parameters and the polarization ellipse is

$$s_1 = s_p \cos 2\epsilon \cos 2\theta$$

$$s_2 = s_p \cos 2\epsilon \sin 2\theta$$

$$s_3 = s_p \sin 2\epsilon$$

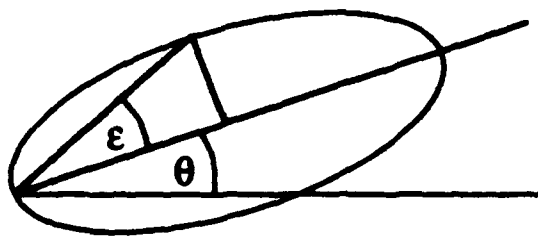


Figure 1.5: Angles used to characterize the polarization ellipse

Section 2

Raytracing Parametric Surfaces

Raytracing is the process of following a ray of light through an optical system. Rays are represented lines in space. The intersection point is calculated for a ray with a surface. For parametric surfaces, this is an iterative procedure, requiring a two-dimensional non-linear optimization. The ray is next refracted or reflected at the surface, and a new ray direction vector is calculated. Then the raytrace is repeated for the succeeding surfaces until the image surface is reached.

This chapter presents the working equations necessary to find the intersection between rays and parametric surfaces.

2.1 Overview

Three ray transfer mechanisms have been devised for tracing rays through optical transparencies. The first is an insertion process. Eligible hyperpatches for ray entry are those associated with the entrance surface. Ray intersections are tested for each eligible hyperpatch until a legal intersection is found. If no intersection is found, the action taken depends on the transmission code for that component. If the transmission code is zero, the ray is assumed to be blocked. The ray trace is continued to the next component if the transmission code is set. The second mechanism is that of internal propagation in which a ray is traced from one side of the hyperpatch to another. The third is ray extraction at the exit face of a hyperpatch. The adjoining face is either a new hyperpatch, the exit surface, or an edge face. If the adjoining face is a new hyperpatch, the local parametric ray coordinates must be calculated for the new hyperpatch. If the exit surface is encountered, the ray is refracted into air and traced to the eye plane or to the entrance surface of the next optical component. Finally, if an edge face is encountered, the ray is either reflected from the interface or totally absorbed at the point of intersection.

2.2 Hyperpatches

The geometric model used in OPTRAN is the hyperpatch defined by PATRAN (chapter 37) [1]. A hyperpatch is a 3-dimensional mapping of the parametric unit cube in (u, v, w) to world coordinates (x, y, z) .

The following method evaluates a tricubic hyperpatch at (u,v,w) to obtain (X,Y,Z) coordinates.

Formal Parameters:

- C (Input) = Hyperpatch coefficients. For each coordinate (X,Y,Z) , these are arranged in a $4 \times 4 \times 4$ matrix of the form:

C	(0,0,0)	C	(0,1,0)	C_v	(0,0,0)	C_v	(0,1,0)
C	(1,0,0)	C	(1,1,0)	C_v	(1,0,0)	C_v	(1,1,0)
C_u	(0,0,0)	C_u	(0,1,0)	C_{uv}	(0,0,0)	C_{uv}	(0,1,0)
C_u	(1,0,0)	C_u	(1,1,0)	C_{uv}	(1,0,0)	C_{uv}	(1,1,0)
C	(0,0,1)	C	(0,1,1)	C_v	(0,0,1)	C_v	(0,1,1)
C	(1,0,1)	C	(1,1,1)	C_v	(1,0,1)	C_v	(1,1,1)
C_u	(0,0,1)	C_u	(0,1,1)	C_{uv}	(0,0,1)	C_{uv}	(0,1,1)
C_u	(1,0,1)	C_u	(1,1,1)	C_{uv}	(1,0,1)	C_{uv}	(1,1,1)
C_w	(0,0,0)	C_w	(0,1,0)	C_{vw}	(0,0,0)	C_{vw}	(0,1,0)
C_w	(1,0,0)	C_w	(1,1,0)	C_{vw}	(1,0,0)	C_{vw}	(1,1,0)
C_{uw}	(0,0,0)	C_{uw}	(0,1,0)	C_{uvw}	(0,0,0)	C_{uvw}	(0,1,0)
C_{uw}	(1,0,0)	C_{uw}	(1,1,0)	C_{uvw}	(1,0,0)	C_{uvw}	(1,1,0)
C_w	(0,0,1)	C_w	(0,1,1)	C_{vw}	(0,0,1)	C_{vw}	(0,1,1)
C_w	(1,0,1)	C_w	(1,1,1)	C_{vw}	(1,0,1)	C_{vw}	(1,1,1)
C_{uw}	(0,0,1)	C_{uw}	(0,1,1)	C_{uvw}	(0,0,1)	C_{uvw}	(0,1,1)
C_{uw}	(1,0,1)	C_{uw}	(1,1,1)	C_{uvw}	(1,0,1)	C_{uvw}	(1,1,1)

- UVW (Input) = Parametric coordinates on the interval $[0,1]$
- XYZ (Output) = Cartesian coordinates at point (u,v,w) of the patch

Method:

For each coordinate, evaluate:

$$\begin{aligned} \mathbf{H}_u &= \mathbf{M}\mathbf{U} \\ \mathbf{H}_v &= \mathbf{M}\mathbf{V} \\ \mathbf{H}_w &= \mathbf{M}\mathbf{W} \end{aligned}$$

in which:

$$\begin{aligned} \mathbf{U}^\dagger &= \begin{pmatrix} u^3 & u^2 & u & 1 \end{pmatrix} \\ \mathbf{V}^\dagger &= \begin{pmatrix} v^3 & v^2 & v & 1 \end{pmatrix} \\ \mathbf{W}^\dagger &= \begin{pmatrix} w^3 & w^2 & w & 1 \end{pmatrix} \end{aligned}$$

\mathbf{C} is the 4x4x4 matrix of patch coefficients for coordinate X , and \mathbf{M} is a 4x4 matrix defining the coefficients of the first order (cubic) Hermite polynomials:

$$\mathbf{M} = \begin{bmatrix} 2 & -3 & 0 & 1 \\ -2 & 3 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}$$

$$H_l = (H_u)_i * (H_v)_j * (H_w)_k \left. \vphantom{H_l} \right\} \text{for } i=1-4, j=1-4, k=1-4 \\ l = i + 4(j - 1) + 16(k - 1)$$

Note: The products $\mathbf{M}\mathbf{U}$, $\mathbf{M}\mathbf{V}$, and $\mathbf{M}\mathbf{W}$ are the same for all three coordinates. The matrix \mathbf{M} above is the transpose of the matrix \mathbf{M} defined in the PATRAN theory chapter.

$$X = \sum_{l=1}^{64} H_l C_l, \quad Y = \sum_{l=1}^{64} H_l C_{l+64}, \quad Z = \sum_{l=1}^{64} H_l C_{l+128}$$

2.3 Parametric Surfaces

Parametric surfaces are found by setting one of the parametric variables to one of its limiting values (0 or 1). This yields one of the six faces of the hyperpatch. Although any one of the six faces may be used in OPTRAN, we show the case for w held constant. This parametric surface is defined by

$$\mathbf{S}(u, v) = \begin{pmatrix} X(u, v) & Y(u, v) & Z(u, v) \end{pmatrix}$$

Surface Normal

At any point (u, v) on the parametric surface, we can construct a vector \mathbf{N} perpendicular to the patch by computing the cross product of the tangent vectors \mathbf{S}^u and \mathbf{S}^v .

$$\mathbf{N} = \mathbf{S}^u \times \mathbf{S}^v$$

where

$$\begin{aligned}\mathbf{S}^u &= \frac{\partial \mathbf{S}(u, v)}{\partial u} = \left(\frac{\partial X(u, v)}{\partial u} \quad \frac{\partial Y(u, v)}{\partial u} \quad \frac{\partial Z(u, v)}{\partial u} \right) \\ \mathbf{S}^v &= \frac{\partial \mathbf{S}(u, v)}{\partial v} = \left(\frac{\partial X(u, v)}{\partial v} \quad \frac{\partial Y(u, v)}{\partial v} \quad \frac{\partial Z(u, v)}{\partial v} \right).\end{aligned}$$

2.4 Surface Intersection

A general discussion of geometric modeling based on parametric cubic surfaces is found in Mortenson [4]. Raytracing of parametric surfaces has been applied to studies of computer graphics [5,6,7].

Given the parametric surface

$$\mathbf{S}(u, v) = \begin{pmatrix} X(u, v) & Y(u, v) & Z(u, v) \end{pmatrix}$$

and a ray defined by the point \mathbf{p} and the unit direction \mathbf{k} , the square of the distance from the surface to the ray is

$$\phi^2 = |\mathbf{F}|^2 = |\mathbf{p}' - \mathbf{p} - q\mathbf{k}|^2$$

where \mathbf{F} is the defect vector and \mathbf{p}' the intersection point,

$$\mathbf{p}' = \mathbf{S}(u, v)$$

and the distance q along the ray to the point of intersection is

$$q = (\mathbf{p}' - \mathbf{p}) \cdot \mathbf{k}.$$

A local minimum of $\phi^2 = 0$ corresponds to a point (u, v) where the ray intersects the surface. All components of the defect vector \mathbf{F} must vanish for an intersection point to exist. A solution can be obtained by adjusting the variables u , v , and q . Those points where ϕ^2 is a minimum, but $\phi^2 > 0$, indicate that the ray missed the surface by a finite distance. Thus a minimization algorithm will still converge near a "silhouette edge" of the surface, but will give a non-zero minimum.

2.5 Ray Refraction

Refraction at a boundary between two dielectric media is determined by Snell's law, that

$$n(\mathbf{k} \times \mathbf{N}) = n'(\mathbf{k}' \times \mathbf{N})$$

where \mathbf{N} is the surface normal vector and \mathbf{k} the unit vector in the direction of propagation.

We can obtain another representation of Snell's Law by examining the following vector triple product:

$$n'(\mathbf{k}' \times \mathbf{N}) \times \mathbf{N} = n(\mathbf{k} \times \mathbf{N}) \times \mathbf{N}$$

Using a standard vector identity, we can write this as

$$(\mathbf{N} \cdot \mathbf{N})n'\mathbf{k}' - (\mathbf{N} \cdot n'\mathbf{k}')\mathbf{N} = (\mathbf{N} \cdot \mathbf{N})n\mathbf{k} - (\mathbf{N} \cdot n\mathbf{k})\mathbf{N}$$

Let

$$\begin{aligned} g^2 &= \mathbf{N} \cdot \mathbf{N} \\ \Gamma &= n\mathbf{k} \cdot \mathbf{N} = ng \cos \theta \\ \Gamma' &= n'\mathbf{k}' \cdot \mathbf{N} = n'g \cos \theta' \end{aligned}$$

These definitions allow us to write

$$n'\mathbf{k}' = n\mathbf{k} + \gamma\mathbf{N}$$

where

$$\gamma = \frac{\Gamma' - \Gamma}{g^2}.$$

Now let us address the problem of finding Γ' . We seek a solution that avoids trigonometric functions.

$$\begin{aligned} n' \sin \theta' &= n \sin \theta \\ n'^2(1 - \cos^2 \theta') &= n^2(1 - \cos^2 \theta) \\ g^2 n'^2 - \Gamma'^2 &= g^2 n^2 - \Gamma^2 \\ \Gamma'^2 &= \Gamma^2 + g^2(n'^2 - n^2) \\ \Gamma' &= \pm \sqrt{\Gamma^2 + g^2(n'^2 - n^2)} \end{aligned}$$

If the argument of the square root is negative, we have total internal reflection.

The choice of sign can be made as follows:

Refraction: Γ' and Γ have the same sign
Reflection: Γ' and Γ change signs

For reflection, $\Gamma' = -\Gamma$.

2.6 Refraction and Reflection Coefficients

The refraction and reflection coefficients determine the energy which is transmitted or reflected at a dielectric interface.

The calculation of these coefficients is best performed in a local coordinate system defined as follows. Let \mathbf{z} be the normal to the surface, with orthogonal vectors \mathbf{x} and \mathbf{y} lying in the plane of the surface, and let \mathbf{x} be perpendicular to the plane of incidence. then

$$\begin{aligned}\mathbf{x} &= \mathbf{k} \times \mathbf{z} \\ \mathbf{y} &= \mathbf{z} \times \mathbf{x}\end{aligned}$$

If the propagation vector \mathbf{k}_o is parallel to the surface normal (normal incidence), no plane of incidence is defined. The choice of \mathbf{x} is then arbitrary, and we let \mathbf{x} be in the direction of the incident polarization reference axis \mathbf{d}_o .

The Maxwell equations for reflection and refraction at the boundary between two isotropic materials can be solved for amplitude reflection/refraction coefficients,

The following amplitude reflection/transmission coefficients may be obtained

$$t_i = \frac{E_m}{E_o}$$

where i is one of four possible combinations, (reflecting or refracting) and (p-polarization or s-polarization), o represents the entrance medium and m represents the exit medium.

Energy reflection/transmission coefficients are obtained from

$$T_i = \left| \frac{n_m \mathbf{k}_m \cdot \mathbf{z}}{n_o \mathbf{k}_o \cdot \mathbf{z}} \right| |t_m|^2$$

For reflection, the index of refraction of exit and entrance media are the same.

S-Polarization

For an incident ray polarized perpendicular to the plane of incidence (s-polarization), the electric field vector $\mathbf{e}_o = \mathbf{x}$ and the reflection and transmission amplitude coefficients

are

$$\begin{aligned} t_s &= \frac{2n_o \cos \theta}{n_o \cos \theta + n_m \cos \theta'} \\ r_s &= \frac{n_o \cos \theta - n_m \cos \theta'}{n_o \cos \theta + n_m \cos \theta'} \end{aligned}$$

where θ is the angle of incidence and θ' the angle of refraction. For this case, the amplitude coefficients simplify to

$$\begin{aligned} t_s &= \frac{2\Gamma}{\Gamma + \Gamma'} \\ r_s &= \frac{\Gamma - \Gamma'}{\Gamma + \Gamma'} \end{aligned}$$

P-Polarization

For an incident ray polarized parallel to the plane of incidence (p-polarization), $\mathbf{e}_o = \mathbf{y}$ and

$$\begin{aligned} t_p &= \frac{2n_o \cos \theta}{n_m \cos \theta + n_o \cos \theta'} \\ r_p &= \frac{n_m \cos \theta - n_o \cos \theta'}{n_m \cos \theta + n_o \cos \theta'} \end{aligned}$$

Normal incidence

For normal incidence, there is no plane of incidence defined. We can then choose $\mathbf{x} = \mathbf{e}_o$. Then the equations simplify to

$$\begin{aligned} t &= \frac{E_b}{E_o} = \frac{2n_o}{n_o + n_m} \\ r &= \frac{E_a}{E_o} = \frac{n_m - n_o}{n_m + n_o} \end{aligned}$$

2.7 Mueller Matrices

In a general polarizing system, the output polarization state is related to the input polarization state by

$$\begin{pmatrix} s'_0 \\ s'_1 \\ s'_2 \\ s'_3 \end{pmatrix} = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix}$$

where the matrix \mathbf{M} is called a Mueller matrix. [9]

Following are Mueller matrices for common optical components. The Mueller matrix for a polarizer oriented at 0° :

$$\mathbf{P}(0) = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Linear retarder, oriented along x-axis:

$$\mathbf{Q}(\delta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \cos \delta & \sin \delta \\ 0 & 0 & \sin \delta & \cos \delta \end{bmatrix}$$

The Mueller matrix for reflection/transmission is equivalent to a partial polarizer with transmission factors t_x t_y

$$\mathbf{P} = \frac{1}{2} \begin{bmatrix} t_x + t_y & t_x - t_y & 0 & 0 \\ t_x - t_y & t_x + t_y & 0 & 0 \\ 0 & 0 & 2\sqrt{t_x t_y} & 0 \\ 0 & 0 & 0 & 2\sqrt{t_x t_y} \end{bmatrix}$$

Rotation matrix:

$$\mathbf{R}(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\alpha & \sin 2\alpha & 0 \\ 0 & -\sin 2\alpha & \cos 2\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

To find the Mueller matrix of a device oriented at an azimuth of θ :

$$\mathbf{M} = \mathbf{R}(-\theta)\mathbf{M}\mathbf{R}(\theta)$$

2.8 Output Data

The result of a ray trace is a set of ray data, consisting of the following information at intersection points:

NPAT	geometric hyperpatch ID
(u, v, w)	parametric coordinates
\mathbf{p}	ray position vector (x, y, z)
\mathbf{n}	surface normal unit vector
\mathbf{k}	direction of wave propagation
\mathbf{d}	polarization reference vector
N	mean index of refraction
δN	difference in refractive index
d	distance to next surface

These data are calculated for the object surface, each intermediate surface intersected by the ray, and finally for the image surface. The optical path to the next surface is Nd . The optical retardance is δNd .

Section 3

Nonlinear Least-Squares Optimization

The surface intersection problem defined in the previous section is solved by the nonlinear least-squares optimization technique presented in this section.

The optimization method seeks to minimize the sum of squares of defect terms, and the ideal solution is obtained if each individual term vanishes. The method expands the merit function in a Taylor series about a starting point, keeping the linear and quadratic terms in the variables. This approximate function is then minimized, and a solution vector is found. This vector may extend outside the region of validity for the Taylor series, so the solution vector is reduced in magnitude or damped to keep changes within the region of validity.

3.1 Mathematic Preliminaries

Merit Function

The defect functions f_i are functions of a set of N variables (x_1, x_2, \dots, x_N) :

$$\begin{aligned} f_1 &= f_1(x_1, x_2, \dots, x_N) \\ f_2 &= f_2(x_1, x_2, \dots, x_N) \\ &\vdots \\ f_M &= f_M(x_1, x_2, \dots, x_N) \end{aligned}$$

The merit function is of the type

$$\phi^2 = \sum_{i=1}^M f_i^2$$

or

$$\phi^2 = \mathbf{f}^t \mathbf{f} = \mathbf{f} \cdot \mathbf{f} = \|\mathbf{f}\|^2$$

where \mathbf{f} is a $(M \times 1)$ vector and \mathbf{f}^t is the $(1 \times M)$ transpose of \mathbf{f} . The first form of the expression uses the notation of matrix multiplication. The second form shows a vector dot (or inner) product, and the last form is a vector norm over defect space.

Linear Defect Model

Over a small region about the starting point, the defects may be approximated by a Taylor series,

$$\mathbf{f} = \mathbf{a}_0 + \mathbf{A}\mathbf{x}.$$

where \mathbf{A} is a $(M \times N)$ matrix of first derivatives:

$$A_{ij} = \frac{\partial f_i}{\partial x_j}$$

and \mathbf{x} are changes in the variables from the starting point.

Gradient

The gradient \mathbf{g} is a $(N \times 1)$ vector given by

$$\mathbf{g} = \nabla \phi^2$$

Its components are

$$g_i = \frac{\partial \phi^2}{\partial x_i} = 2 \left(f_1 \frac{\partial f_1}{\partial x_i} + f_2 \frac{\partial f_2}{\partial x_i} + \cdots + f_M \frac{\partial f_M}{\partial x_i} \right)$$

then

$$\mathbf{g} = 2\mathbf{A}^t \mathbf{f}$$

A very inefficient method for finding a solution point is to search for a minimum along the vector $-\mathbf{g}$. This is known as the gradient search method, and for a highly nonlinear function may be a good strategy. Experience has shown, however, that for most cases the local gradient is not a good predictor of the final minimum.

3.2 Method of Least-Squares

Using the linear model for the defects allows us to express the merit function as

$$\phi^2 = (\mathbf{a}_0 + \mathbf{A}\mathbf{x}) \cdot (\mathbf{a}_0 + \mathbf{A}\mathbf{x}) = \mathbf{a}_0 \cdot \mathbf{a}_0 + 2\mathbf{c}_0 \cdot \mathbf{x} + \mathbf{x}^\dagger \mathbf{C} \mathbf{x}$$

where

$$\begin{aligned} \mathbf{c}_0 &= \mathbf{A}^\dagger \mathbf{a}_0 \\ \mathbf{C} &= \mathbf{A}^\dagger \mathbf{A} \end{aligned}$$

Let \mathbf{a}_j represent column j of matrix \mathbf{A} . The matrix \mathbf{C} is a symmetric ($N \times N$) matrix, whose elements can be written as a sum over the defects,

$$\begin{aligned} c_j &= \sum_{i=1}^M A_{ij}(a_i)_0 = \mathbf{a}_j \cdot \mathbf{a}_0 \\ C_{jk} &= \sum_{i=1}^M A_{ij}A_{ik} = \mathbf{a}_j \cdot \mathbf{a}_k \end{aligned}$$

The matrix \mathbf{C} is called the covariance array. The gradient is

$$\mathbf{g} = 2\mathbf{A}^\dagger \mathbf{f} = 2(\mathbf{c}_0 + \mathbf{C}\mathbf{x})$$

The minimum of ϕ^2 is obtained by setting $\mathbf{g} = 0$ and solving for \mathbf{x} . The resulting matrix equation

$$\mathbf{c}_0 + \mathbf{C}\mathbf{x} = 0$$

is a set of simultaneous linear equations known as the normal equations of least-squares. Providing that the matrix \mathbf{C} is not singular, these equations can always be solved, and the formal solution \mathbf{x}_m may be written

$$\mathbf{x}_m = -\mathbf{C}^{-1}\mathbf{c}_0$$

At the minimum, the merit function becomes

$$\phi^2 = \mathbf{a}_0 \cdot \mathbf{a}_0 + \mathbf{c}_0 \cdot \mathbf{x}_m$$

In fact, the matrix \mathbf{C} is frequently nearly singular, so that direct matrix inversion is an inappropriate numeric method for solving the normal equations. Before exploring a more appropriate method, we present an alternative formulation of the matrix equations, using homogeneous matrices, and discuss some problems associated with nonlinear effects.

3.2.1 Homogeneous Matrix Formulation

Homogeneous, or augmented matrices A' and x' , can be constructed such that

$$f = a_0 + Ax = A'x' = \left(A \mid a_0 \right) \begin{pmatrix} x \\ 1 \end{pmatrix}$$

where A' has $N' = N + 1$ columns, and x' has N' rows. The augmented form of the covariance array C is given by:

$$C' = \left(\begin{array}{c|c} C & c_0 \\ \hline c_0^\dagger & a_0 \cdot a_0 \end{array} \right)$$

Homogeneous matrices provide a compact way of representing and storing the components of the normal equations,

$$C'x' = 0$$

Since C' is a symmetric matrix, a triangular partition can be used to represent the full matrix. The homogeneous C' matrix may be multiplied by a nonzero scalar value without changing the solution of the system of equations. For example, we could choose to normalize the matrix so that the lower right corner element is unity. This element is zero only if the current design represents an ideal solution, in which case, no further processing is necessary.

3.2.2 Nonlinear Effects

A Taylor series expansion of a scalar function of several variables may be written as

$$\phi^2 = \phi_0^2 + g_0 \cdot x + \frac{1}{2} x^\dagger H x$$

where H is the Hessian matrix of second derivatives, defined as

$$H_{jk} = \frac{\partial^2 \phi^2}{\partial x_j \partial x_k}$$

The Hessian matrix may be expressed as

$$\begin{aligned} H_{jk} &= \sum_{i=1}^M \frac{\partial^2 f_i^2}{\partial x_j \partial x_k} \\ &= 2 \sum_{i=1}^M \left(\frac{\partial f_i}{\partial x_j} \frac{\partial f_i}{\partial x_k} + f_i \frac{\partial^2 f_i}{\partial x_i \partial x_j} \right) \end{aligned}$$

A linear defect model predicts that

$$\phi^2 = \mathbf{a}_0 \cdot \mathbf{a}_0 + 2\mathbf{c}_0 \cdot \mathbf{x} + \mathbf{x}^\dagger \mathbf{C} \mathbf{x}$$

and matches the first and second terms of the Taylor series but it leaves out the second derivatives in the Hessian matrix.

$$H_{jk} = 2 \left(C_{jk} + \sum_{i=1}^M f_i \frac{\partial^2 f_i}{\partial x_i \partial x_j} \right)$$

3.2.3 Damped Least-Squares

As an alternative to performing lengthy calculations of second derivatives, a damping term of the form $p\mathbf{I}$, where \mathbf{I} is the identity matrix, may be incorporated to eliminate singularities and restrict the least-squares solution to

$$\mathbf{x} = -(\mathbf{C} + p\mathbf{I})^{-1} \mathbf{c}_0$$

In the limit as $p \rightarrow 0$, the solution approaches the least-squares prediction. An exact solution is obtained if the defects are linear functions of the variables. Sufficiently large values of p guarantee a nonsingular matrix inversion by making the eigenvalues non-negative. As p becomes large, the solution approaches a small step opposed to the direction of the local gradient, which usually results in a smaller merit function and thus a design improvement. The method of damped least-squares has been used extensively and successfully in optical design. Damping plays an important role in any nonlinear optimization program, but we choose to introduce it at a later stage in the process. Let us turn our attention instead to the technique of orthogonal variables.

3.3 Method of QU Factorization

QU factorization is a method of transforming the normal equations into a set of orthonormal equations. We seek to transform the matrix \mathbf{A}' into another ($M \times N'$) matrix \mathbf{Q} whose columns \mathbf{q}_j are orthonormal over the defects,

$$\mathbf{q}_j \cdot \mathbf{q}_k = \sum_{i=1}^M Q_{ij} Q_{ik} = \lambda_j^2 \delta_{jk}$$

The constant λ_j represents the magnitude of q_j . It is zero only if the corresponding q_j vector is zero. The transformation will be an augmented upper triangular matrix U such that

$$QU = A'$$

The matrix U has the form

$$U = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1N} & \alpha_{10} \\ 0 & 1 & \alpha_{23} & \cdots & \alpha_{2N} & \alpha_{20} \\ 0 & 0 & 1 & \cdots & \alpha_{3N} & \alpha_{30} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \alpha_{N0} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

where

$$\alpha_{jk} = \frac{q_j \cdot a_k}{\lambda_j^2}$$

We define a new set of variables w such that

$$w = Ux'$$

The defect vectors can now be written as

$$f = A'x' = Qw = a_0 + \sum_{j=1}^N q_j w_j$$

The merit function itself can be expressed as a quadratic sum,

$$\phi^2 = \lambda_0^2 + \sum_{j=1}^N \lambda_j^2 (w_j + \alpha_{j0})^2$$

where the minimum value of ϕ^2 is obtained by inspection as $w_j = -\alpha_{j0}$. The solution can be expressed compactly as

$$w = -u_0$$

The following steps are repeated as often as necessary to obtain a final solution:

1. Calculate the required derivatives.
2. Carry out the QU factorization.
3. Determine the predicted location of the minimum by setting $w = -u_0$.
4. Solve the triangular matrix equation $w = Ux'$ for x .
5. Search for improvement in the direction predicted by the model.

Each step will now be described in detail.

Calculation of Derivatives

Derivatives of the defect functions may be calculated either analytically or numerically. In raytracing parametric surfaces, the partial derivatives are calculated analytically.

These derivatives are given by

$$\begin{aligned} \mathbf{F}^u &= \mathbf{S}^u - (\mathbf{S}^u \cdot \mathbf{k})\mathbf{k} \\ \mathbf{F}^v &= \mathbf{S}^v - (\mathbf{S}^v \cdot \mathbf{k})\mathbf{k} \end{aligned}$$

Cholesky Decomposition

The Cholesky decomposition algorithm is one way of carrying out the factorization of the C matrix indicated in the previous section.

The general column \mathbf{q}_k is given by

$$\mathbf{q}_k = \mathbf{a}_k - \sum_{i=1}^{k-1} \mathbf{q}_i \alpha_{ik}$$

where for $j < k$

$$\begin{aligned} \mathbf{q}_j \cdot \mathbf{q}_k &= \mathbf{q}_j \cdot \mathbf{a}_k - \lambda_j^2 \alpha_{jk} = 0 \\ \alpha_{jk} &= \frac{\mathbf{q}_j \cdot \mathbf{a}_k}{\lambda_j^2} \end{aligned}$$

and for $j = k$

$$\lambda_k^2 = \mathbf{q}_k \cdot \mathbf{q}_k = \mathbf{q}_k \cdot \mathbf{a}_k$$

The Cholesky algorithm involves the following a recursive process:

$$\begin{aligned} \alpha_{jk} &= \frac{1}{\lambda_j^2} \left(\mathbf{a}_j \cdot \mathbf{a}_k - \sum_{i=1}^{j-1} [\mathbf{q}_i \cdot \mathbf{a}_j] \alpha_{ik} \right) \\ &= \frac{1}{\lambda_j^2} \left(C_{jk} - \sum_{i=1}^{j-1} \alpha_{ij} \alpha_{ik} \lambda_i^2 \right) \end{aligned}$$

and

$$\lambda_k^2 = C_{kk} - \sum_{j=1}^{k-1} \alpha_{jk}^2 \lambda_j^2$$

The algorithm is a simple iterative procedure. For the first column, $q_1 = a_1$ with $\lambda_1^2 = C_{11}$. The next column is constructed by requiring that

$$q_2 = a_2 - q_1 \alpha_{12}$$

where

$$\begin{aligned}\alpha_{12} &= \frac{C_{12}}{\lambda_1^2} \\ \lambda_2^2 &= C_{22} - \alpha_{12}^2 \lambda_1^2\end{aligned}$$

Each new column is constructed to be orthogonal to the previous columns. This is the general principle of Gram-Schmidt orthogonalization except that the columns are calculated from the covariance matrix rather than the original A' matrix. If the sequence of calculations causes a negative value of λ^2 , we set that coefficient to zero and then continue the calculations. If the merit function is independent of a particular variable x_j then the matrix element C_{jj} will be zero. The coefficients α_{ij} and λ_j^2 must also be zero. If the variable is linearly dependent on earlier variables, then the algorithm used for λ^2 should sum to zero (or within a specified tolerance of zero).

Obtaining a Solution

The merit function can be reduced to a quadratic form through the following steps:

$$\begin{aligned}\phi^2 &= \mathbf{f} \cdot \mathbf{f} \\ &= \mathbf{a}_0 \cdot \mathbf{a}_0 + 2 \sum_{j=1}^N (\mathbf{a}_0 \cdot \mathbf{q}_j) w_j + \sum_{j=1}^N \sum_{k=1}^N (\mathbf{q}_j \cdot \mathbf{q}_k) w_j w_k \\ &= \lambda_0^2 + \sum_{j=1}^N \lambda_j^2 (\alpha_{j0}^2 + 2\alpha_{j0} w_j + w_j^2) \\ &= \lambda_0^2 + \sum_{j=1}^N \lambda_j^2 (w_j + \alpha_{j0})^2\end{aligned}$$

where

$$\lambda_0^2 = \mathbf{a}_0 \cdot \mathbf{a}_0 - \sum_{j=1}^N \lambda_j^2 \alpha_{j0}^2$$

The minimum value of ϕ^2 can be obtained by inspection as $w_j = -\alpha_{j0}$. The solution can be expressed compactly as

$$\mathbf{w} = -\mathbf{u}_0$$

The eigenvalues $\{\lambda_j^2\}$ determine the relative influence of the transformed variables on the magnitude of the merit function. If $\lambda_j^2 = 0$, then w_j has no effect on the solution. Zero, or extremely small values of λ^2 , are common in least-squares equations. Such values results from variables which are either linearly related to other variables or which have no effect on the merit function. This situation leads to an ill-conditioned matrix inverse (indeterminate values). Using QU factorization, we can recognize variables for which $\lambda^2 = 0$, and exclude them from further processing. Small values of λ^2 are associated with vectors oriented along valleys of the merit functions. Larger values of λ^2 are associated with directions normal to the walls of the valleys. If λ is zero, we have an unused degree of freedom in the design. There is a continuum of designs that satisfy the merit function.

Round-off errors can lead to imaginary values for some λ_j rather than small positive values. If we set λ_j to zero at this stage of the calculation, we can avoid some of the problems of ill-conditioning.

Back Substitution

A triangular matrix equation, such as

$$\mathbf{w} = \mathbf{U}\mathbf{x}'$$

is easy to solve for \mathbf{x}' by using the method of back substitution. First, we use the definition of an upper triangular matrix to write

$$w_i = x_i + \sum_{j=i+1}^N \alpha_{ij}x_j$$

Then beginning at the bottom of the matrix, we can generate the following sequence:

$$\begin{aligned} x_N &= w_N \\ x_{N-1} &= w_{N-1} - \alpha_{N-1,N}x_N \\ &\vdots \\ x_i &= w_i - \sum_{j=i+1}^N \alpha_{ij}x_j \end{aligned}$$

Search in One Dimension

Searches in one dimension may be made along any single variable x_j , transformed variable w_j , or solution vector \mathbf{u}_0 . The general case is to define a search vector \mathbf{v} and parameter

p such that

$$\mathbf{x} = p\mathbf{v}$$

where $p = 0$ is the current point and $p = 1$ is the predicted solution to

$$\phi(p) = 0$$

Our technique is to first calculate $\phi(1)$. If the merit function is smaller than the original value $\phi(0)$, we calculate $\phi(1.5)$. Otherwise, we find $\phi(0.5)$. At this stage, we have enough information to construct an interpolating parabola $\phi(p) = b_0 + b_1p + b_2p^2$. We locate either the minimum value or the closest root p_m and then find $\phi(p_m)$ directly. The value of p associated with the smallest value of the merit function is used to generate the starting position for the next iteration of optimization.

Section 4

Differential Rays

Differential raytracing is a simplified method of finding rays that are close to some reference ray. If the reference ray is the central ray traveling along the optic axis, then the close rays are called paraxial rays. A differential raytrace is the generalization of the idea of paraxial rays. Differential rays are faster to calculate than exact rays and can be used to determine the focussing properties of a narrow bundle of rays, like those entering the eye from an external object point.

Differential rays are used to calculate the astigmatism and defocus of a small bundle of rays. Given a central ray from \mathbf{p} traveling along the unit direction \mathbf{k} , the differential ray is given by $\delta\mathbf{p}$ and $\delta\mathbf{k}$.

4.1 Surface Intersection

The differential ray intersects a surface at a distance $q + \delta q$ along the ray, yielding $\delta\mathbf{p}'$ subject to the condition that the differential ray lies in the surface tangent plane,

$$\delta\mathbf{p}' \cdot \mathbf{N} = 0.$$

Taking the derivative of the ray translation equation gives

$$\delta\mathbf{p}' = \delta\mathbf{p} + q \delta\mathbf{k} + \delta q \mathbf{k}$$

Let

$$\tilde{\mathbf{p}} = \delta\mathbf{p} + q \delta\mathbf{k}.$$

Then

$$\delta q = -\frac{\tilde{\mathbf{p}} \cdot \mathbf{N}}{\mathbf{k} \cdot \mathbf{N}}$$

and

$$\delta\mathbf{p}' = \tilde{\mathbf{p}} + \delta q \mathbf{k}.$$

4.2 Ray Refraction

The refracted ray is given by

$$n' \mathbf{k}' = n \mathbf{k} + \gamma \mathbf{N}$$

and the differential ray is

$$n' \delta \mathbf{k}' = n \delta \mathbf{k} + \gamma \delta \mathbf{N} + \delta \gamma \mathbf{N}$$

subject to the condition that \mathbf{k}' is a unit vector, so that

$$\mathbf{k}' \cdot \delta \mathbf{k}' = 0$$

Let

$$\tilde{\mathbf{k}} = n \delta \mathbf{k} + \gamma \delta \mathbf{N}$$

Then

$$\delta \gamma = -\frac{\mathbf{k}' \cdot \tilde{\mathbf{k}}}{\mathbf{k}' \cdot \mathbf{N}}$$

and

$$n' \delta \mathbf{k}' = \tilde{\mathbf{k}} + \delta \gamma \mathbf{N}$$

4.3 Differential Trace of Parametric Surface

For a parametric surface, the ray intersection differential $\delta \mathbf{S}$ is given by

$$\delta \mathbf{S} = \mathbf{S}^u \delta u + \mathbf{S}^v \delta v$$

Then

$$\mathbf{S}^u \delta u + \mathbf{S}^v \delta v = \tilde{\mathbf{p}} + \mathbf{k} \delta q$$

represents a set of three linear equations in three unknowns $(\delta u, \delta v, \delta q)$ which may be solved using Cramer's rule:

$$\begin{aligned} \delta u &= \frac{\mathbf{k} \cdot \tilde{\mathbf{p}} \times \mathbf{S}^v}{\mathbf{k} \cdot \mathbf{N}} \\ \delta v &= \frac{\mathbf{k} \cdot \mathbf{S}^u \times \tilde{\mathbf{p}}}{\mathbf{k} \cdot \mathbf{N}} \\ \delta q &= -\frac{\tilde{\mathbf{p}} \cdot \mathbf{N}}{\mathbf{k} \cdot \mathbf{N}} \end{aligned}$$

The differential surface normal $\delta\mathbf{N}$ is also required. The surface normal for parametric surfaces is given by

$$\mathbf{N} = \mathbf{S}^u \times \mathbf{S}^v$$

from which the differential normal can be obtained as

$$\delta\mathbf{N} = (\delta\mathbf{S}^u) \times \mathbf{S}^v + \mathbf{S}^u \times (\delta\mathbf{S}^v)$$

where

$$\begin{aligned}\delta\mathbf{S}^u &= \mathbf{S}^{uu} \delta u + \mathbf{S}^{uv} \delta v \\ \delta\mathbf{S}^v &= \mathbf{S}^{uv} \delta u + \mathbf{S}^{vv} \delta v.\end{aligned}$$

Section 5

Anisotropic Refractive Index Ellipsoid

This section explains how OPTRAN computes the coefficients of an ellipsoid whose principal axes define the orthotropic indices of refraction at a particular location as a function of temperature and stress.

5.1 Orthotropic Indices of Refraction

The equation for the index ellipsoid of an unstressed material with respect to the principal axes of the material (x, y, z) is given by

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1.$$

The orthotropic indices of refraction (n_x, n_y, n_z) are given by

$$\begin{aligned} n_x &= \hat{n}_x + \frac{dn_x}{dT} \Delta T \\ n_y &= \hat{n}_y + \frac{dn_y}{dT} \Delta T \\ n_z &= \hat{n}_z + \frac{dn_z}{dT} \Delta T \end{aligned}$$

where ($\hat{n}_x, \hat{n}_y, \hat{n}_z$) is a set of orthotropic indices of refraction, ($dn_x/dT, dn_y/dT, dn_z/dT$) is a set of orthotropic temperature coefficients along the same axes, and ΔT is the temperature change.

A geometric transformation is required if the axes of the material are not aligned with the global axes.

5.2 Geometric Transformation

The direction cosines of the material axes are calculated as direction cosines with respect to a set of global coordinates and stored as column vectors in the matrix \mathbf{X} . If the axes of the material are aligned with the global axes, this matrix is the unit matrix. A user provided transformation matrix may be defined for materials whose axes have a fixed orientation with respect to a set of reference axes, which may be the global axes or an orthogonal set of axes aligned with respect to the parametric parameters. This option was intended to accommodate materials that are deformed to follow an irregular shape.

The choices are distinguished by setting the following flags for each material:

IUVWAX(IMATL) = 0 use the global axes as reference axes
 = 1 use the derivatives of the parametric parameters
 with respect to the global axes to define a set
 of reference axes

IOPMTR(IMATL) = 0 do not use the OPMATR matrix to define the material
 axes with respect to the reference axes
 = 1 use the OPMATR matrix

Let \mathbf{D}_o contain the direction cosines of the material axes with respect to the reference axes stored as column vectors,

$$\mathbf{D}_o = \begin{bmatrix} \mathbf{l} & \mathbf{m} & \mathbf{n} \end{bmatrix}$$

where \mathbf{l} , \mathbf{m} , \mathbf{n} , are unit direction cosine vectors which define the direction of the three material axes with respect to the reference axes.

If the global axes are used as the reference axes \mathbf{X} is simply set equal to input material axis direction cosine matrix OPMATR for material IMATL.

$$\mathbf{X} = \mathbf{D}_o$$

Otherwise a set of orthogonal reference axes is computed as a function of the derivatives of the x , y , z coordinates with respect to the u , v , w geometric hyperpatch parametric parameters. Then

$$\mathbf{X} = \mathbf{D}_o \mathbf{D}_{uvw}$$

where

$$\mathbf{D}_{uvw} = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}$$

The vectors defined by partial derivative triples $V_u = (X_u, Y_u, Z_u)$, $V_v = (X_v, Y_v, Z_v)$, $V_w = (X_w, Y_w, Z_w)$ are not orthogonal except in special cases where the geometry has curvilinear orthogonality. An orthogonal set of direction cosines is computed by taking cross products of partial derivative triples and normalizing the results to unit vectors. In general only the direction of one of the vectors remains unaltered and can be used as reference. By an arbitrary decision, the following convention has been adopted for defining the a reference axis system (V_1, V_2, V_3) based on the parametric partial derivatives is computed as follows:

If IVOPT = 1 (Parameter u held constant)

$$\begin{aligned} V_2 &= V_v / |V_v| \\ V_1 &= (V_v \times V_w) / |V_v \times V_w| \\ V_3 &= (V_1 \times V_2) / |V_1 \times V_2| \end{aligned}$$

If IVOPT = 2 (Parameter v held constant)

$$\begin{aligned} V_1 &= V_u / |V_u| \\ V_2 &= (V_w \times V_u) / |V_w \times V_u| \\ V_3 &= (V_1 \times V_2) / |V_1 \times V_2| \end{aligned}$$

If IVOPT = 3 (Parameter w held constant)

$$\begin{aligned} V_1 &= V_u / |V_u| \\ V_3 &= (V_u \times V_v) / |V_u \times V_v| \\ V_2 &= (V_3 \times V_1) / |V_3 \times V_1| \end{aligned}$$

where the program variable IVOPT is set as a function of the incident face.

5.3 Stress Birefringence

When a stress is applied to a material, the index ellipsoid is modified, and the changes in the components of the dielectric tensor are linearly related to the six stress components. [3][p. 703-704]

On applying stress, the index ellipsoid is changed into another whose equation is

$$a_{xx}x^2 + a_{yy}y^2 + a_{zz}z^2 + 2a_{xy}xy + 2a_{yz}yz + 2a_{zx}xz = 1.$$

The coefficients of the ellipse are defined by

$$\{A\} = \{A_o\} + [q]\{\sigma\}$$

where

$$\{A\} = \begin{pmatrix} a_{xx} \\ a_{yy} \\ a_{zz} \\ a_{xy} \\ a_{yz} \\ a_{zx} \end{pmatrix}, \quad \{A_o\} = \begin{pmatrix} 1/n_x^2 \\ 1/n_y^2 \\ 1/n_z^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \{\sigma\} = \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{pmatrix}.$$

The first term contains the orthotropic indices and associated thermal coefficients. The second term contains stress-related terms where $\{\sigma\}$ is the stress tensor and $[q]$ is a set of 36 stress-optical coefficients.

Although all the stress-optic coefficients may have unique nonzero values, symmetry considerations for crystalline or isotropic materials prescribe relationships among the coefficients and reduce the number of independent values that must be specified. For cubic crystals, the three principal axes (x, y, z) are equivalent, and consequently the following relations hold among the stress-optics coefficients:

$$\begin{aligned} q_{11} &= q_{22} = q_{33} \\ q_{12} &= q_{21} = q_{23} = q_{32} = q_{13} = q_{31} \\ q_{44} &= q_{55} = q_{66} \end{aligned}$$

with all the remaining coefficients being zero.

For isotropic materials, the above relations must remain unaltered for any change of axes. This is only possible if the stress-optical coefficients satisfy the additional relation

$$2q_{44} = q_{11} - q_{12}.$$

If the global coordinate system is different from that of the dielectric material, the stresses must first be converted to the material coordinates. Then the resulting dielectric tensor must be transformed back into the global coordinate system. If $\{\sigma\}'$ denotes material coordinates and $\{\sigma\}$ global coordinates, then

$$\{\sigma\}' = [s]\{\sigma\}$$

where $[s]$ is the coordinate transformation matrix. In material coordinates,

$$\{A\}' = \{A_o\}' + [q]'\{\sigma\}'$$

which may be written in reference coordinates as

$$\{A\} = [s]^{-1}\{A_o\}' + [s]^{-1}[q]''[s]\{\sigma\}.$$

The matrix $[s]^{-1}[q]''[s]$ is not a function of stress and in many cases is not a function of location, and therefore can be precomputed.

5.4 Tensor Rotation Transformation

Given a point expressed in (x, y, z) global coordinates, we find the corresponding representation in (ℓ, m, n) coordinates. Let ℓ, m, n be the unit direction of the material axes in (x, y, z) global coordinates. Then

$$\begin{pmatrix} \ell & m & n \end{pmatrix} = \begin{pmatrix} x & y & z \end{pmatrix} \begin{bmatrix} \ell_x & m_x & n_x \\ \ell_y & m_y & n_y \\ \ell_z & m_z & n_z \end{bmatrix}.$$

The reverse transformation is

$$\begin{pmatrix} x & y & z \end{pmatrix} = \begin{pmatrix} \ell & m & n \end{pmatrix} \begin{bmatrix} \ell_x & \ell_y & \ell_z \\ m_x & m_y & m_z \\ n_x & n_y & n_z \end{bmatrix}.$$

The coordinate transformation matrix $[s]$ is given by

$$[s] = \begin{bmatrix} \ell_x^2 & \ell_y^2 & \ell_z^2 & 2\ell_x\ell_y & 2\ell_y\ell_z & 2\ell_x\ell_z \\ m_x^2 & m_y^2 & m_z^2 & 2m_xm_y & 2m_ym_z & 2m_xm_z \\ n_x^2 & n_y^2 & n_z^2 & 2n_xn_y & 2n_y n_z & 2n_xn_z \\ \ell_xm_x & \ell_y m_y & \ell_z m_z & (\ell_xm_y + m_x\ell_y) & (\ell_y m_z + m_y\ell_z) & (\ell_xm_z + m_x\ell_z) \\ m_xn_x & m_y n_y & m_z n_z & (m_xn_y + n_xm_y) & (m_y n_z + n_y m_z) & (m_xn_z + m_zn_x) \\ \ell_xn_x & \ell_y n_y & \ell_z n_z & (\ell_xn_y + n_x\ell_y) & (\ell_y n_z + n_y\ell_z) & (\ell_xn_z + \ell_zn_x) \end{bmatrix}$$

5.5 Principal Axes of Projected Ellipse

Given an ellipsoid A in the (x, y, z) coordinate system and an axis of projection given by the unit vector (u, v, w) . Then

$$ux + vy + wz = 0$$

is the equation of a plane through the origin and perpendicular to the projection axis. The intersection of this plane with the ellipsoid is an ellipse. We want to find the major and minor axes of this ellipse.

We find the lengths of the axes by finding the extrema of $r^2 = x^2 + y^2 + z^2$ subject to the constraints that the extrema lies on the plane and the ellipsoid.

By using Lagrange's method of undetermined multipliers (λ_1 and λ_2), we define the following:

$$h(x, y, z) = x^2 + y^2 + z^2 + \lambda_1 h_1(x, y, z) + \lambda_2 h_2(x, y, z)$$

where

$$\begin{aligned} h_1(x, y, z) &= a_{xx}x^2 + a_{yy}y^2 + a_{zz}z^2 + 2a_{xy}xy + 2a_{yz}yz + 2a_{zx}xz - 1 \\ h_2(x, y, z) &= ux + vy + wz. \end{aligned}$$

Then setting the partial derivatives of $h(x, y, z) = 0$ gives

$$\begin{aligned} \frac{\partial h}{\partial x} &= 2x + 2\lambda_1 [a_{xx}x + a_{xy}y + a_{xz}z] + \lambda_2 u = 0 \\ \frac{\partial h}{\partial y} &= 2y + 2\lambda_1 [a_{xy}x + a_{yy}y + a_{yz}z] + \lambda_2 v = 0 \\ \frac{\partial h}{\partial z} &= 2z + 2\lambda_1 [a_{xz}x + a_{yz}y + a_{zz}z] + \lambda_2 w = 0. \end{aligned}$$

Then we may obtain

$$x \frac{\partial h}{\partial x} + y \frac{\partial h}{\partial y} + z \frac{\partial h}{\partial z} = 2r^2 + 2\lambda_1 = 0$$

and

$$u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} + w \frac{\partial h}{\partial z} = 2\lambda_1 [ua_x + va_y + wa_z] + \lambda_2 = 0$$

where

$$\begin{aligned} a_x &= a_{xx}x + a_{xy}y + a_{xz}z \\ a_y &= a_{xy}x + a_{yy}y + a_{yz}z \\ a_z &= a_{xz}x + a_{yz}y + a_{zz}z. \end{aligned}$$

Then

$$\begin{aligned}\frac{\partial h}{\partial x} &= 2x - 2r^2 a_x + 2r^2 u [u a_x + v a_y + w a_z] = 0 \\ \frac{\partial h}{\partial y} &= 2y - 2r^2 a_y + 2r^2 v [u a_x + v a_y + w a_z] = 0 \\ \frac{\partial h}{\partial z} &= 2z - 2r^2 a_z + 2r^2 w [u a_x + v a_y + w a_z] = 0.\end{aligned}$$

Factoring gives the following system of homogeneous equations.

$$\begin{aligned}x[r^{-2} - a_{xx} + u a_u] + y[-a_{xy} + u a_v] + z[-a_{xz} + u a_w] &= 0 \\ x[-a_{xy} + v a_u] + y[r^{-2} - a_{yy} + v a_v] + z[-a_{yz} + v a_w] &= 0 \\ x[-a_{xz} + w a_u] + y[-a_{yz} + w a_v] + z[r^{-2} - a_{zz} + w a_w] &= 0\end{aligned}$$

where

$$\begin{aligned}a_u &= a_{xx}u + a_{xy}v + a_{xz}w \\ a_v &= a_{xy}u + a_{yy}v + a_{yz}w \\ a_w &= a_{xz}u + a_{yz}v + a_{zz}w.\end{aligned}$$

The determinant of coefficients of this set of three homogeneous equations must vanish for there to be a nontrivial solution for x , y , and z . Using this requirement gives

$$\begin{vmatrix} [r^{-2} - a_{xx} + u a_u] & [-a_{xy} + u a_v] & [-a_{xz} + u a_w] \\ [-a_{xy} + v a_u] & [r^{-2} - a_{yy} + v a_v] & [-a_{yz} + v a_w] \\ [-a_{xz} + w a_u] & [-a_{yz} + w a_v] & [r^{-2} - a_{zz} + w a_w] \end{vmatrix} = 0$$

Expanding the determinant gives

$$ar^4 + br^2 + 1 = 0$$

where

$$a = (a_1 - a_2) - (a_3 - a_4)$$

$$\begin{aligned}a_1 &= a_{xx}a_{yy} + a_{yy}a_{xz} + a_{xz}a_{zx} \\ a_2 &= a_{xy}^2 + a_{yz}^2 + a_{zx}^2 \\ a_3 &= a_{xx}(v a_v + w a_w) + a_{yy}(u a_u + w a_w) + a_{zz}(u a_u + v a_v) \\ a_4 &= a_{yz}(v a_w + w a_v) + a_{xy}(v a_u + u a_v) + a_{xz}(u a_w + w a_u),\end{aligned}$$

and

$$b = u a_u + v a_v + w a_w - a_{xx} - a_{yy} - a_{zz}.$$

This equation is a simple quadratic in r^2 . The solutions are the major and minor axes. The system of homogeneous equations can then be solved for the direction of the respective principal axes.

Section 6

Optical Waves in Anisotropic Materials

Transparent materials may be optically anisotropic due to intrinsic crystal properties or as a result of mechanical stresses. In either case, the index of refraction becomes a function of direction within the material and can be described by the equation of an ellipsoid. An anisotropic material permits two monochromatic plane waves with two different polarizations and two different velocities to propagate in any given direction. The directions of the two displacement field vectors \mathbf{D} corresponding to a given direction of propagation \mathbf{k} are perpendicular to each other. The phenomenon of wave propagation in an anisotropic material is called birefringence or double refraction.

6.1 Light Propagation in Anisotropic Materials

As discussed in the previous section, the equation for the index ellipsoid with respect to the principal axes of the material (x, y, z) is given by

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1.$$

After rotating the coordinate system, this ellipsoid is changed into another whose equation is

$$a_{xx}x^2 + a_{yy}y^2 + a_{zz}z^2 + 2a_{xy}xy + 2a_{yz}yz + 2a_{zx}xz = 1.$$

In a birefringent medium two orthogonal polarized waves may be propagated in a given direction. Let a set of principal axes (x, y, z) be defined, with the direction of propagation in the z -direction, the allowed polarization directions along the x and y axes,

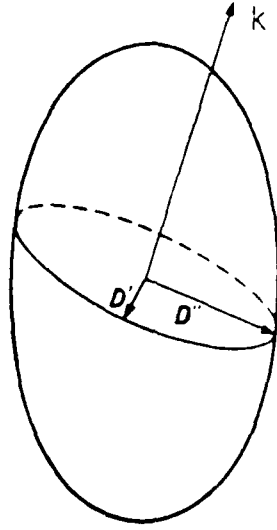


Figure 6.1: Construction of the \mathbf{D} vectors belonging to a wave normal \mathbf{k} [1].

and the associated indices of refraction be given by n_x and n_y . The Jones matrix for the propagating wave is then

$$\mathbf{J}' = \begin{pmatrix} e^{-2\pi j n_x d/\lambda} & 0 \\ 0 & e^{-2\pi j n_y d/\lambda} \end{pmatrix} \mathbf{J}$$

where d is the distance propagated. This relationship may be factored by defining

$$n = \frac{n_x + n_y}{2}$$

$$\delta_n = \frac{n_x - n_y}{2}.$$

The optical path δ along the ray is given by

$$\delta = nd$$

and the Jones matrix is now given by

$$\mathbf{J}' = \begin{pmatrix} e^{-2\pi j \delta_n d/\lambda} & 0 \\ 0 & e^{2\pi j \delta_n d/\lambda} \end{pmatrix} \mathbf{J}.$$

which may be expressed as the Mueller matrix of a linear retarder.

As shown in Figure 6.1, the intersection of the dielectric ellipsoid with a plane through the origin and perpendicular to the direction of wave propagation \mathbf{k} defines an ellipse whose major and minor axes determine the direction of the displacement field vectors \mathbf{d}

and the associated indices of refraction for the two orthogonal polarized waves propagating in the direction \mathbf{k} . The vector \mathbf{d} is a unit vector in the direction of the displacement \mathbf{D} .

Wavefront propagation is governed by the triplet of unit vectors $(\mathbf{d}, \mathbf{h}, \mathbf{k})$ such that

$$\mathbf{k} = \mathbf{d} \times \mathbf{h}$$

Rays in a birrefringent medium do not propagate in the same direction as the wavefront. Ray propagation is determined by the Poynting vector \mathbf{S} given by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$

The direction of ray propagation and polarization is represented by the triplet of unit vectors $(\mathbf{e}, \mathbf{h}, \mathbf{s})$ such that

$$\mathbf{s} = \mathbf{e} \times \mathbf{h}$$

The electric field \mathbf{E} is related to the displacement field \mathbf{D} by the dielectric tensor.

$$\mathbf{D} = \epsilon \mathbf{E}$$

The index ellipsoid is an equivalent representation of the dielectric tensor that relates the electric and displacement fields as follows:

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \begin{bmatrix} a_{xx} & a_{xy} & a_{xz} \\ a_{xy} & a_{yy} & a_{yz} \\ a_{xz} & a_{yz} & a_{zz} \end{bmatrix} \begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} \quad (6.1)$$

The indices of refraction may be graphed on a polar plot as a function of the direction of wave propagation. This defines a two-sheeted surface called the normal surface. Where the sheets intersect, the two orthogonal polarized waves have the same index of refraction. These directions are called the optic axes of the medium.

A material is classified as isotropic if the three principal indices of refraction are equal, as uniaxial if two of the three principal indices are equal, and as biaxial if none of the principal indices are the same. Table 6.1 lists typical refractive indices of some crystals.

The wave normal surface for an isotropic material is a sphere, since the index of refraction does not vary with direction. For a uniaxial material, the normal surface consists of a sphere and an ellipsoid of revolution. If $n_x = n_y$, then these two sheets touch at two points on the z -axis because both wavefronts propagate along the z -axis with the

Table 6.1: Typical refractive indices of some crystals [3].

Crystal		Refractive Indices		
Isotropic	Fluorite	1.392		
	Sodium chloride, NaCl	1.544		
	Diamond, C	2.417		
	CdTe	2.69		
	GaAs	3.40		
Uniaxial (Positive)		n_o	n_e	
	Ice, H ₂ O	1.309	1.310	
	Quartz, SiO ₂	1.544	1.553	
	Beryllium oxide, BeO	1.717	1.732	
	Zircon, ZrSiO ₄	1.923	1.968	
	ZnS	2.354	2.358	
Uniaxial (Negative)		n_o	n_e	
	ADP, (NH ₄)H ₂ PO ₄	1.522	1.478	
	Beryl, Be ₃ Al ₂ (SiO ₃) ₆	1.598	1.590	
	KDP, KH ₂ PO ₄	1.507	1.467	
	Sodium nitrate, NaNO ₃	1.587	1.366	
	Calcite, CaCO ₃	1.658	1.486	
	Tourmaline	1.638	1.618	
	Sapphire, Al ₂ O ₃	1.768	1.760	
	Lithium niobate LiNbO ₃	2.300	2.208	
	Barium titanate, BaTiO ₃	2.416	2.364	
Biaxial		n_x	n_y	n_z
	Gypsum	1.520	1.523	1.530
	Feldspar	1.522	1.526	1.530
	Mica	1.552	1.582	1.588
	Topaz	1.619	1.620	1.627
	Sodium nitrite	1.344	1.411	1.651
	YAlO ₃	1.923	1.938	1.947
	SbSI	2.7	3.2	3.8

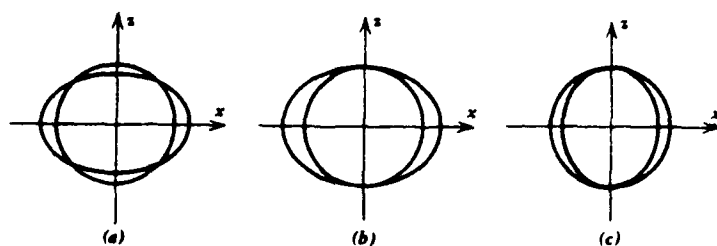


Figure 6.2: Intersection of the normal surface with xz plane for (a) biaxial crystals, (b) positive uniaxial crystals, and (c) negative uniaxial crystals [3].

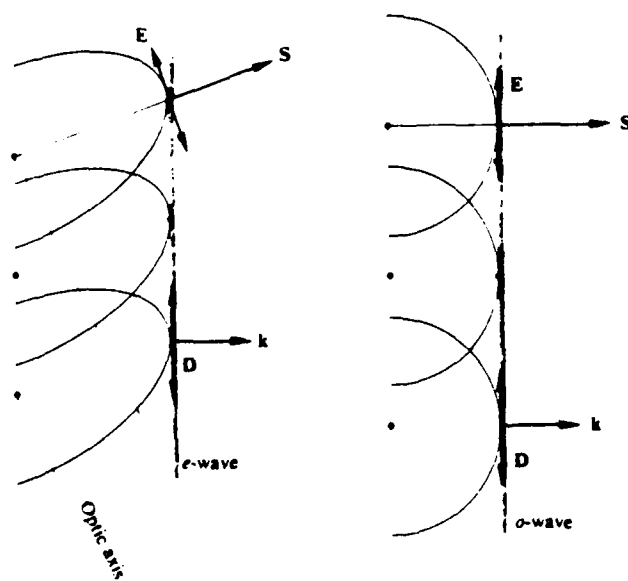


Figure 6.3: Orientation of rays and waves in a uniaxial crystal [2].

same index of refraction. This direction is called the optic axis of the material. A biaxial material has two separate optic axes coplanar with longest and shortest principal axes. Figure 6.2 shows cross-sections of the normal surface for biaxial and uniaxial crystals.

Figure 6.3 shows Huygens wavelets for two modes of propagation in an anisotropic crystal. Huygens wavelets for which the index of refraction is independent of direction are spherical wavefronts, and the associated rays are called ordinary waves. The direction of wavefront propagation and the direction of energy transfer (ray direction) are the same. Wavelets for which the index of refraction varies with direction are ellipsoidal wavefronts, and the associated waves are called extraordinary waves. The direction of ray propagation is different from that of wavefront propagation. The envelope of all wavelets in both cases is a plane wave propagating horizontally to the right.

Given the direction of wave propagation \mathbf{k} and the refractive index ellipsoid \mathbf{A} , we can calculate \mathbf{d}_1 , n_1 , \mathbf{d}_2 , and n_2 for the principal polarization states, as shown in Section 5. If $n_1 = n_2$ there is no unique choice of principal directions, and we may arbitrarily select the principal polarization states.

For each polarization mode, we calculate $\mathbf{h} = \mathbf{k} \times \mathbf{d}$ and the electric field. Then using the directions of the electric and magnetic fields, we can calculate the ray direction \mathbf{s} from

$$\mathbf{s} = \mathbf{e} \times \mathbf{h}$$

The ray index of refraction n_r is obtained from the wave index of refraction by

$$n_r = n \mathbf{k} \cdot \mathbf{s}$$

For a wavefront propagating normal to a plane-parallel slab of material of thickness d , as shown in Figure 6.4, the optical path δ is given by

$$\delta = nd$$

If the ray direction makes an angle α with the wave direction, the distance traveled by the ray in the material is $d/\cos \alpha$ and the index is $n \cos \alpha$, so that the calculated optical path is the same whether calculated for the wavefront or the ray.

Figure 6.5 is an illustration of the propagation of light through a calcite crystal. If we send a narrow beam of normal (unpolarized) light into a calcite crystal, it will split into two beams. Rotating the crystal causes one of the rays to remain stationary and the other to move in a circle about it, following the motion of the crystal. The fixed

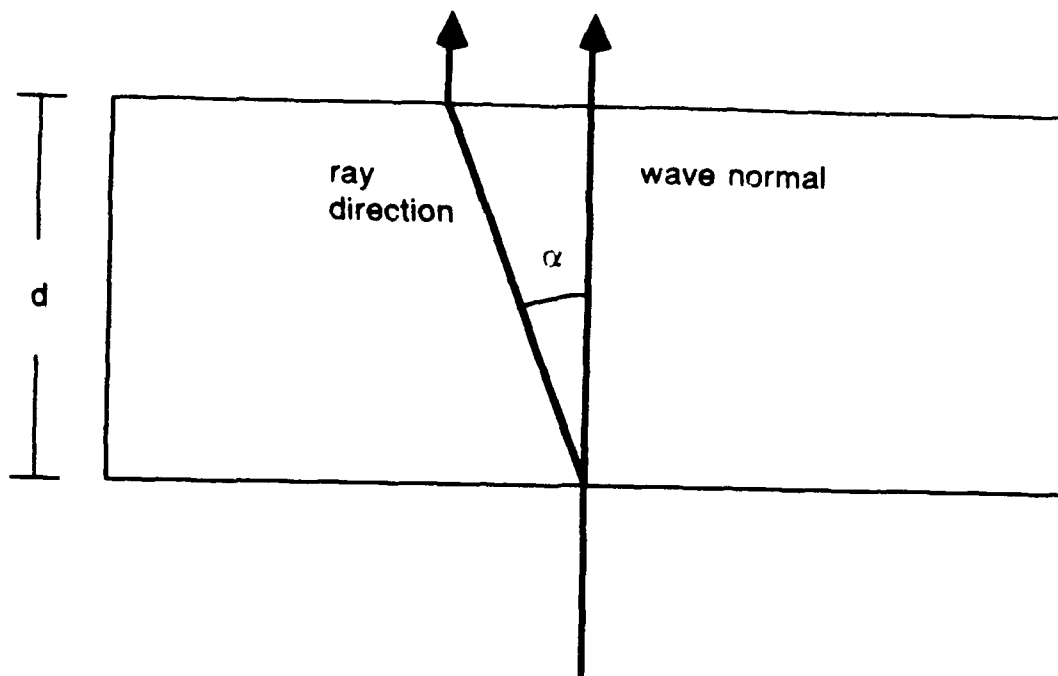


Figure 6.4: Optical path for rays and wavefronts.

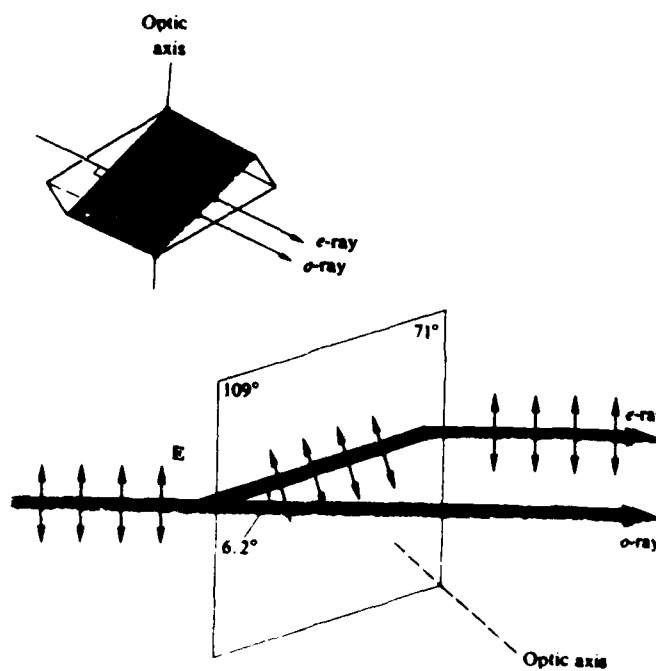


Figure 6.5: A light beam with two orthogonal field components traversing a calcite principal section [2].

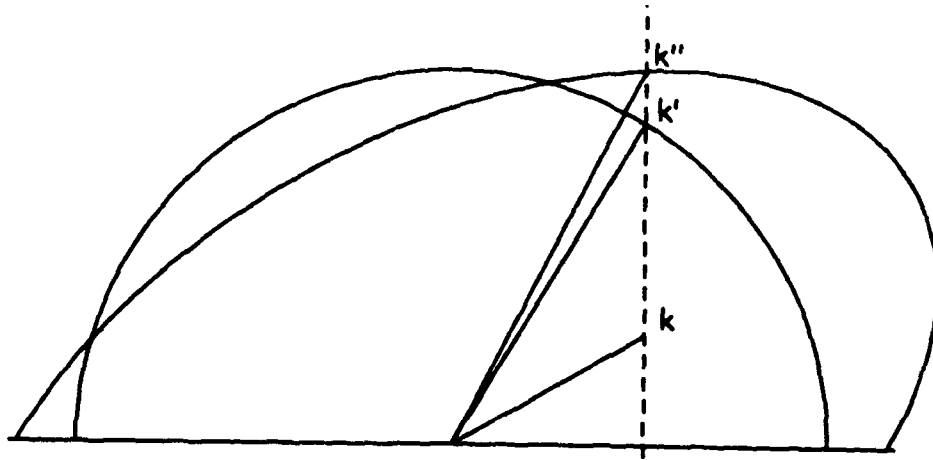


Figure 6.6: Double refraction at the boundary of an anisotropic material.

ray, which follows the usual rules of refraction, is the ordinary ray. The moving ray, not following the usual rules, is the extraordinary ray.

In the discussion so far, the incident wavefront has been normal to the crystal surface, so that the direction of refraction was also normal to the crystal surface. In general, however, rays will be incident at an oblique angle to a dielectric boundary surface.

6.2 Double Refraction

Consider an unpolarized plane wave incident on the surface of an anisotropic material. The refracted wave, in general, is a mixture of two propagation modes. The conditions for refraction, established by the application of Fermat's Principle, required that the refracted wavefront lie in the plane of incidence established by the incident propagation vector and the surface normal and that Snell's Law is satisfied,

$$n \sin \theta = n_1 \sin \theta_1 = n_2 \sin \theta_2$$

For an extraordinary ray, the index of refraction is a function of the angle of incidence, leading to the following transcendental equation

$$n \sin \theta = n_r(\theta_r) \sin \theta_r$$

which must be solved for θ_r . The solution can be obtained graphically as shown in Figure 6.6. The wavefront optical cosines are vectors whose magnitude is the index of

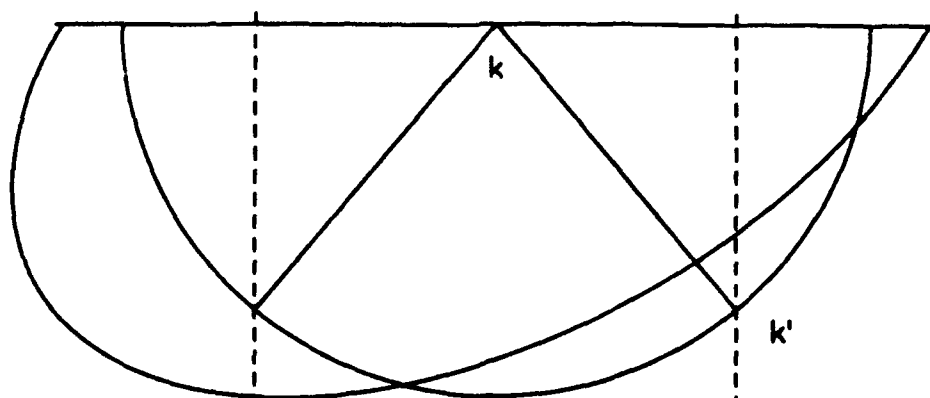


Figure 6.7: Ordinary reflection at the boundary of an anisotropic material.

refraction. The wavefront normal surface traces the index of refraction as a function of angle of incidence for the two possible propagation modes. The projection of the wavefront propagation vector onto the boundary is equal to $n \sin \theta$. This projection must be the same for the incident and refracted rays to satisfy Snell's Law. The refracted waves may be found by extending a vertical line from the incident wavefront propagation vector through the normal surface. The intersection points with the normal surface represent the two refracted waves.

If a polarized wave propagating within an anisotropic material is incident at an exit surface, part of the wave will be reflected at the surface. For an ordinary wave, shown in Figure 6.7, the angle of reflection will be equal to the angle of incidence. For an extraordinary wave, shown in Figure 6.8, the angle of reflection is not equal to the angle of incidence because the index of refraction is different for the reflected wave.

In examining Figures 6.7 and 6.8, we observe that there are two possible solutions for the reflected wave. Only one solution is reported for each situation. The polarization modes for this geometry turn out to be uncoupled. The ordinary wave reflects only as an ordinary wave, and the extraordinary incident wave reflects only as an extraordinary wave. This situation is shown in Figure 6.9. The short rays denote the Poynting vector and the longer waves the wave propagation vector. The p-polarization case corresponds to the extraordinary wave. The associated wave normal surfaces are shown for each polarization mode.

In general, at the boundary between two different anisotropic materials, there are two possible reflected waves and two refracted waves for a polarized incident wave. The energy carried by the incident wave is divided among the four possible output waves.

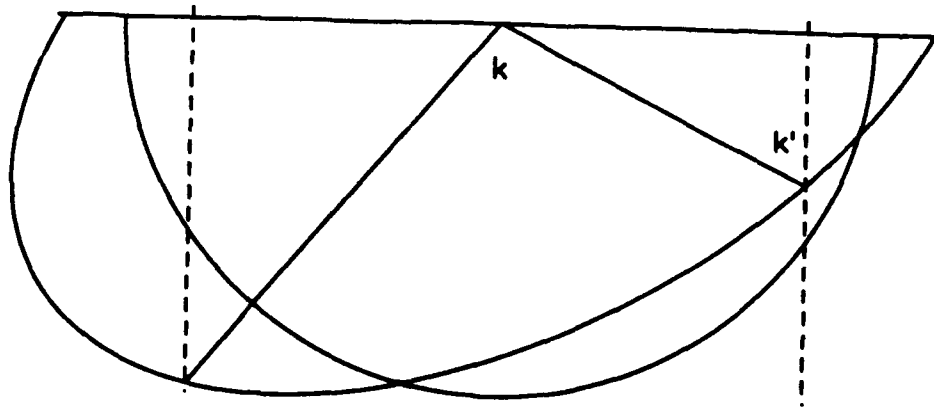


Figure 6.8: Extraordinary reflection at the boundary of an anisotropic material.

The distribution of energy is determined by using the boundary conditions imposed by Maxwell's equations on the transverse components of the electric and magnetic fields. This distribution will be discussed in the next section.

6.3 Reflection and Refraction Coefficients

Let \mathbf{s}_o be the unit Poynting vector of the incident ray, \mathbf{e}_o the electric field vector, and \mathbf{h}_o the magnetic field vector, such that

$$\mathbf{s}_o = \mathbf{e}_o \times \mathbf{h}_o$$

Let \mathbf{z} be the normal to the surface, with orthogonal vectors \mathbf{x} and \mathbf{y} lying in the plane of the surface. There will be two reflected beams, \mathbf{s}_a and \mathbf{s}_c and two refracted beams \mathbf{s}_b and \mathbf{s}_d . At the boundary, the transverse components of electric and magnetic fields are continuous.

These boundary conditions are:

$$\begin{aligned} E_o \mathbf{e}_o \cdot \mathbf{x} + E_a \mathbf{e}_a \cdot \mathbf{x} + E_c \mathbf{e}_c \cdot \mathbf{x} &= E_b \mathbf{e}_b \cdot \mathbf{x} + E_d \mathbf{e}_d \cdot \mathbf{x} \\ E_o \mathbf{e}_o \cdot \mathbf{y} + E_a \mathbf{e}_a \cdot \mathbf{y} + E_c \mathbf{e}_c \cdot \mathbf{y} &= E_b \mathbf{e}_b \cdot \mathbf{y} + E_d \mathbf{e}_d \cdot \mathbf{y} \\ n_o E_o \mathbf{h}_o \cdot \mathbf{x} + n_a E_a \mathbf{h}_a \cdot \mathbf{x} + n_c E_c \mathbf{h}_c \cdot \mathbf{x} &= n_b E_b \mathbf{h}_b \cdot \mathbf{x} + n_d E_d \mathbf{h}_d \cdot \mathbf{x} \\ n_o E_o \mathbf{h}_o \cdot \mathbf{y} + n_a E_a \mathbf{h}_a \cdot \mathbf{y} + n_c E_c \mathbf{h}_c \cdot \mathbf{y} &= n_b E_b \mathbf{h}_b \cdot \mathbf{y} + n_d E_d \mathbf{h}_d \cdot \mathbf{y} \end{aligned}$$

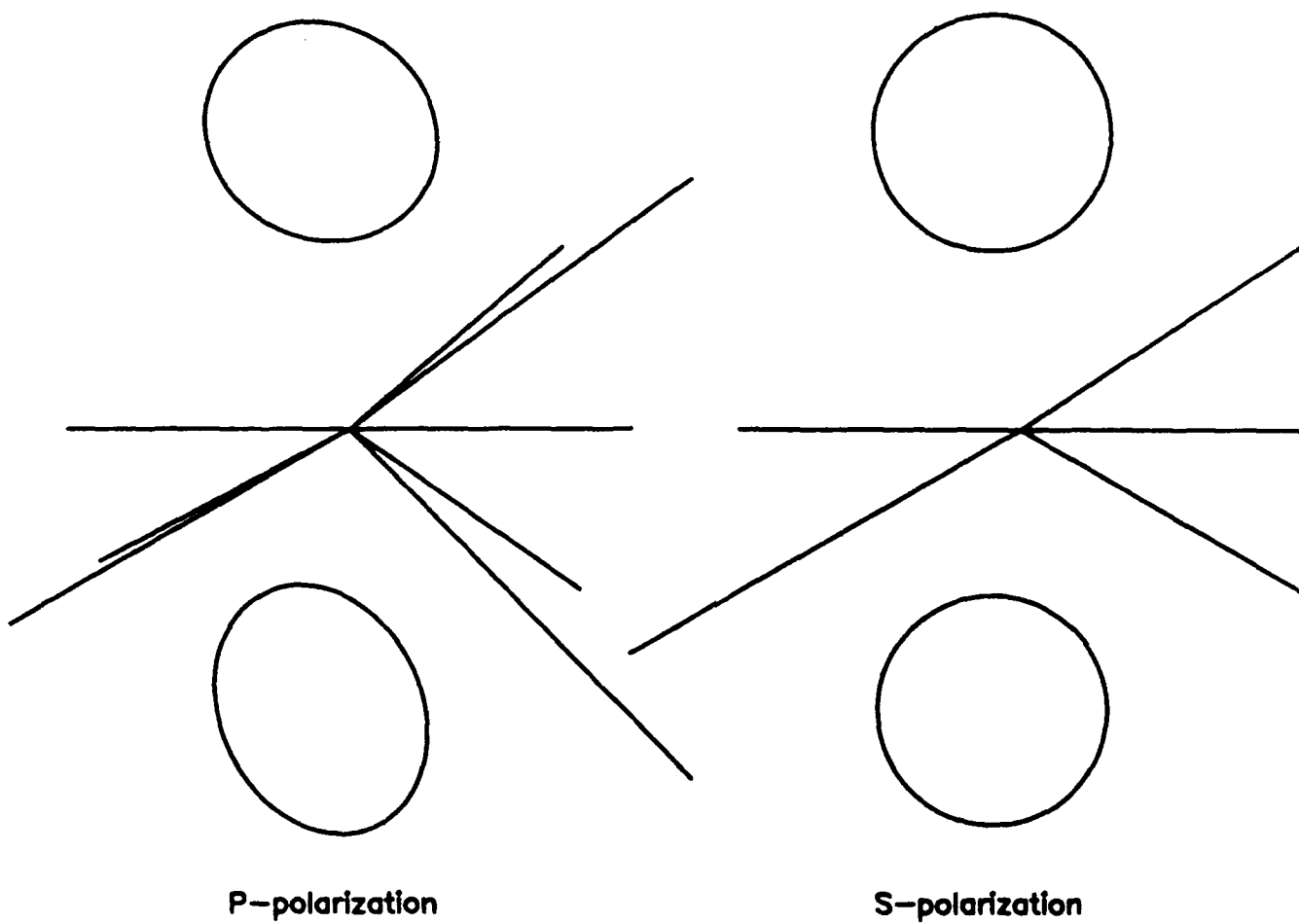


Figure 6.9: Refraction and reflection at boundary between two anisotropic materials.

$$\begin{bmatrix} -\mathbf{e}_a \cdot \mathbf{x} & \mathbf{e}_b \cdot \mathbf{x} & -\mathbf{e}_c \cdot \mathbf{x} & \mathbf{e}_d \cdot \mathbf{x} \\ -\mathbf{e}_a \cdot \mathbf{y} & \mathbf{e}_b \cdot \mathbf{y} & -\mathbf{e}_c \cdot \mathbf{y} & \mathbf{e}_d \cdot \mathbf{y} \\ -n_a \mathbf{h}_a \cdot \mathbf{x} & n_b \mathbf{h}_b \cdot \mathbf{x} & -n_c \mathbf{h}_c \cdot \mathbf{x} & n_d \mathbf{h}_d \cdot \mathbf{x} \\ -n_a \mathbf{h}_a \cdot \mathbf{y} & n_b \mathbf{h}_b \cdot \mathbf{y} & -n_c \mathbf{h}_c \cdot \mathbf{y} & n_d \mathbf{h}_d \cdot \mathbf{y} \end{bmatrix} \begin{bmatrix} E_a \\ E_b \\ E_c \\ E_d \end{bmatrix} = \begin{bmatrix} \mathbf{e}_o \cdot \mathbf{x} \\ \mathbf{e}_o \cdot \mathbf{y} \\ n_o \mathbf{h}_o \cdot \mathbf{x} \\ n_o \mathbf{h}_o \cdot \mathbf{y} \end{bmatrix} E_o$$

Let \mathbf{x} be perpendicular to the plane of incidence, then

$$\mathbf{x} = \mathbf{s}_o \times \mathbf{z}$$

$$\mathbf{y} = \mathbf{z} \times \mathbf{x}$$

If the Poynting vector is parallel to the surface normal (normal incidence), no plane of incidence is defined. The choice of \mathbf{x} is then arbitrary. We typically let $\mathbf{x} = \mathbf{e}_o$.

The following amplitude reflection/transmission coefficients may be obtained

$$t_m = \frac{E_m}{E_o}$$

where m is one of the four reflecting or refracting rays (a, b, c, d). Energy reflection/transmission coefficients are obtained from

$$T_m = \left| \frac{n_m \mathbf{s}_m \cdot \mathbf{z}}{n_o \mathbf{s}_o \cdot \mathbf{z}} \right| |t_m|^2$$

6.3.1 Reflection and Refraction in Isotropic Media

The equations for reflection and refraction at the boundary between two isotropic materials define the Fresnel coefficients. In this section we show how they may be derived from the general expressions given above.

S-Polarization

For an incident ray polarized perpendicular to the plane of incidence (s-polarization), $\mathbf{e}_o = \mathbf{x}$ and

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & \cos \theta' \\ 0 & 0 & n_o & -n_s \\ n_o \cos \theta & n_s \cos \theta' & 0 & 0 \end{bmatrix} \begin{bmatrix} E_a \\ E_b \\ E_c \\ E_d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ n_o \cos \theta \end{bmatrix} E_o$$

from which we obtain $E_c = E_d = 0$ and

$$\begin{bmatrix} -1 & 1 \\ n_o \cos \theta & n_s \cos \theta' \end{bmatrix} \begin{bmatrix} E_a \\ E_b \end{bmatrix} = \begin{bmatrix} 1 \\ n_o \cos \theta \end{bmatrix} E_o$$

These equations can be solved for the reflection and transmission amplitude coefficients

$$\begin{aligned} t_s &= \frac{2n_o \cos \theta}{n_o \cos \theta + n_s \cos \theta'} \\ r_s &= \frac{n_o \cos \theta - n_s \cos \theta'}{n_o \cos \theta + n_s \cos \theta'} \end{aligned}$$

P-Polarization

For an incident ray polarized parallel to the plane of incidence (p-polarization), $\mathbf{e}_o = \mathbf{s}_o \times \mathbf{x}$ and

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & \cos \theta' \\ 0 & 0 & n_o & -n_s \\ n_o \cos \theta & n_s \cos \theta' & 0 & 0 \end{bmatrix} \begin{bmatrix} E_a \\ E_b \\ E_c \\ E_d \end{bmatrix} = \begin{bmatrix} 0 \\ \cos \theta \\ -n_o \\ 0 \end{bmatrix} E_o$$

from which we obtain $E_a = E_b = 0$ and

$$\begin{bmatrix} \cos \theta & \cos \theta' \\ n_o & -n_s \end{bmatrix} \begin{bmatrix} E_c \\ E_d \end{bmatrix} = \begin{bmatrix} \cos \theta \\ -n_o \end{bmatrix} E_o$$

These equations can be solved for the reflection and transmission amplitude coefficients

$$\begin{aligned} t_p = \frac{E_d}{E_o} &= \frac{2n_o \cos \theta}{n_s \cos \theta + n_o \cos \theta'} \\ r_p = \frac{E_c}{E_o} &= \frac{n_s \cos \theta - n_o \cos \theta'}{n_s \cos \theta + n_o \cos \theta'} \end{aligned}$$

Normal incidence

For isotropic media at normal incidence, there is no plane of incidence defined. We can then choose $\mathbf{x} = \mathbf{e}_o$. Then the equations simplify to

$$\begin{aligned} E_o + E_a &= E_b \\ n_o E_o - n_o E_a &= n_s E_b \end{aligned}$$

from which we can find the transmission and reflection amplitude coefficients

$$\begin{aligned} t &= \frac{E_b}{E_o} = \frac{2n_o}{n_o + n_s} \\ r &= \frac{E_a}{E_o} = \frac{n_s - n_o}{n_s + n_o} \end{aligned}$$

References

- [1] "PATRAN Plus User Manual," PDA Engineering, PATRAN Division, Costa Mesa, California, 1987.
- [2] Eugene Hecht, *Optics*, Second Edition, Addison-Wesley, 1987.
- [3] Max Born and Emil Wolf, *Principles of Optics*, Fifth Edition, Pergamon Press, 1975.
- [4] Michael Mortenson, *Geometric Modeling*, John Wiley & Sons, 1985.
- [5] Kenneth I. Joy and Murthy N. Bhetanabhotla, "Ray Tracing Parametric Surface Patches Utilizing Numerical Techniques and Ray Coherence," *Computer Graphics*, Proceedings of SIGGRAPH 1986, volume 20, number 4, pp.279-285, (July, 1986).
- [6] J. T. Kajiya, "Ray Tracing Parametric Patches," *Computer Graphics*, Proceedings of SIGGRAPH 1982, volume 16, number 3, pp. 245-254, (July, 1982).
- [7] Mikio Shinya, Tokiichiro Takahashi, and Seichiro Naito, "Principles and Applications of Pencil Tracing," *Computer Graphics*, Proceedings of SIGGRAPH 1987, volume 21, number 4, pp. 45-54, (July, 1987).
- [8] Pochi Yeh, *Optical Waves in Layered Media*, John Wiley & Sons, 1988.
- [9] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam, 1977.
- [10] John W. Fielman and John S. Loomis, "Transparency Optical Performance Prediction," in proceedings of Conference on Aerospace Transparent Materials and Enclosures, Monterey, CA, (January, 1989).