METROLOGY AND DESIGN OF GRADIENT-INDEX OPTICAL MATERIALS

A Thesis

Submitted to the Faculty of the Graduate School

of the University of Minnesota

by

Di Lin

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

James R. Leger, Adviser

August, 2015

Acknowledgements

I wish to express my sincerest thanks to Prof. James Leger for his continuous encouragement and guidance throughout the course of this research and learning process, for demonstrating the virtues and qualities that I aspire to in all other aspects of my life. He is truly an outstanding role model, and I could not have asked for a better advisor during my graduate studies at the University of Minnesota. Thanks are also extended Profs. Joey Talghader, Mo Li and Prisca Cushman for careful reading of this thesis and serving on the examination committee.

Furthermore, I would like to acknowledge the contributions and valuable insight of Jeremy Teichman from the Institute for Defense Analyses (IDA), whose suggestions in our many fruitful discussions formed the basis for this work. I would also like to acknowledge the Defense Advanced Research Projects Agency (DARPA) for taking an interest in my research and providing the funding support for the bulk of this work. The opportunity to collaborate with Prof. Duncan Moore's group at the Institute of Optics at the University of Rochester, who provided manufactured samples and measurements for the metrology portion of this research, is also greatly appreciated.

Finally, I would like to extend my thanks to my family members, friends and colleagues. Without their support, I would never have come this far in my academic endeavor. I am truly blessed to have so many positive influences in my life.

Abstract

Gradient-index (GRIN) materials provide interesting ways to direct light propagation inside a bulk medium. Their application in optical systems as compact optical elements offer many advantages such as convenient form factor, unique dispersion characteristics, aberration correction capabilities, *etc*. With the recent technological advances in the fabrication techniques for these materials, it is reasonable to speculate that arbitrary refract index distributions in GRIN media can be realized in the near future.

The integration of GRIN components into optical systems requires accurate knowledge of their refractive index distribution. Numerical methods for recovering the refractive index of the material using boundary value measurements of position and slope for interrogating rays that transit the medium are described. For one-dimensional index profiles, we develop a bootstrap algorithm for recovering the refractive index in successive regions of the overall profile from the boundary value data. We then compare the reconstructed index profile obtained in this method with that of a different method based on ray displacement and show good agreement in computer simulation as well as in experimental measurement. In the case of two-dimensional refractive index distributions, we show that the path integrals describing beam deflection inside the material can be cast in the form of linear algebraic equations using a simplifying assumption that decouples unknown ray trajectories inside the medium from the refractive index. The resulting linear system is inverted numerically to recover the refractive index distribution, and the ray trajectories are subsequently ascertained through an iterative ray trace procedure. Using boundary values of ray position and slope generated from a numerical ray trace, we show that this method can achieve RMS index errors less than 0.5% of the refractive index range.

In addition, we explore the application of GRIN components in designing optical resonators. Using a Green's function approach, we show that wave propagation inside GRIN media follows the Huygens-Fresnel principle and can be calculated from the superposition of secondary wavelets. A design procedure for achieving coherent mode conversion in GRIN media is described, and a tool for analyzing optical resonators employing an intracavity GRIN component is developed. We use this tool to calculate the spatial eigenmodes of a flat-mirror resonator employing a Gaussian-to-flat-top GRIN mode converter and determine its modal properties.

Table of Contents

List of Figures

List of Abbreviations

1-D – One-dimensional

2-D – Two-dimensional

3-D – Three-dimensional

CVD – Chemical vapor deposition

DoF – Degree of freedom

EM – Electromagnetic

GRIN – Gradient-index

LSQR – Least squares QR factorization

MLS – Moving least squares

OPL – Optical path length

RIU – Refractive index units

RMS – Root-mean-squared

1. Introduction

1.1 Introduction to GRIN Media

We introduce the notion of gradient-index (GRIN) optics through a familiar and naturally occurring optical phenomenon known as the *mirage effect*, in which the refraction of light rays in transit between a distant object and an observer produces a displaced image of the object. The cause of the phenomenon may be explained as follows. On a hot day, the sun may heat the ground to such high temperatures that a significant gradient develops in the temperature profile of air just above the ground. According to the ideal gas law, the temperature of air molecules is related to their local density, and density is related to the air's refractive index. Thus, the varying density of air molecules induced by the temperature gradient produces a gradient in the air's refractive index, which in turn causes light rays to bend away from the ground. From the observer's perspective, it looks as if the rays originating from a distant object are coming directly from the ground, creating a mirage of the object. We illustrate the mirage effect in Figure 1.1.1.

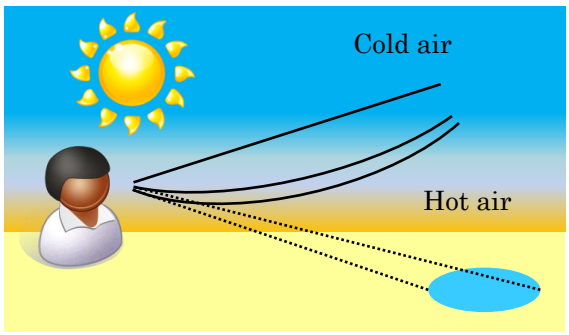


Figure 1.1.1. The mirage effect caused by refraction of light rays in air.

1.2 Fabrication of GRIN Materials

More generally, GRIN materials refer to a class of inhomogeneous optical media whose refractive index can vary arbitrarily with position. The subject of GRIN optics dates back to the 1850s. For many years, the manufacturing capabilities of GRIN materials have been the limiting factor in successfully implementing GRIN elements in optical systems. However, there are now many different material systems (mainly glasses and plastics) in which gradients in the refractive index can be made. Myriad techniques have been studied for fabricating increasingly complex index variations [Moore, 1980; Teichman, *et al.*, 2013]. Current methods for fabricating GRIN materials include neutron irradiation [Sinai, 1971], chemical vapor deposition (CVD) [Pickering, *et al.*, 1986], ion exchange [Ohmi, *et al.*, 1988; Messerschmidt, *et al.*,

1995], ion stuffing [Mohr, *et al.*, 1979], and variations of polymer-based processes [Wu, *et al.*, 1996; Liu, *et al.*, 2006; Liu, *et al.*, 2009]. In neutron irradiation, a boron-rich glass is bombarded with neutrons to induce a change in the boron concentration, which dictates the local refractive index. CVD involves the deposition of different types of glass materials with varying refractive indices onto a surface (*e.g.* the inside or outside of a tube); this method can be employed to achieve a continuous overall index gradient provided that the dimensions of the distinct layers in the composite structure are smaller than the wavelength of interest. Ion exchange is perhaps the most prevalent method of GRIN manufacture due to its simplicity in instrumentation and control; in this system, a glass material is immersed in a salt bath containing lithium ions, which are exchanged with sodium ions in the glass by diffusion to produce an index gradient. Ion stuffing uses a similar technique and refers to the process of filling pores in specific glass materials with a variety of salts through diffusion to produce varying refractive indices; the pores are formed when the glass material is heated and phase separates. In polymerization, the monomer of an organic material is partially polymerized by exposing it to ultraviolet light at varying intensities in order to achieve varying refractive indices in the plastic material.

The underlying physical principles behind these fabrication methods have always limited their ability to produce arbitrary index profiles. Specifically, these limitations include: (1) the physical depth into which the material's refractive index can be changed; (2) the maximum change that can be achieved in the refractive index; and (3) the types of refractive index variations that can be manufactured. However, several new methods have been developed that permit unprecedented control over the index profile in two and three dimensions [*e.g.*, Urness, *et al.*, 2015]. Given the various technological advances in GRIN fabrication techniques described above, it is reasonable to speculate that arbitrary index distributions in GRIN optical components may be realizable in the near future.

1.3 Applications of GRIN Optics

In response to the advances in GRIN fabrication over the last three decades, there has been increasing interest in exploring novel GRIN materials to develop compact, lightweight and robust optics. GRIN optics offer appealing form factors as well as additional degrees of freedom in controlling the propagation of light and have found application in telecommunications and compact imaging. In the former application, a refractive index profile that varies radially from the center of an optical fiber can be chosen properly so that the rays that transit the fiber are guided by refraction, rather than by total internal reflection in step-index optical fibers, for instance. In addition, GRIN fibers can be designed so that all modes propagate with the same velocity, reducing modal dispersion and thereby increasing the bandwidth and the repeater distance of optical communication systems [Moore, 1980]. Furthermore, the cylindrical form factor of GRIN optics simplifies coupling between optical fibers and sources. The latter application of GRIN optics is primarily based on exploiting the optical power provided by GRIN materials. It follows that the optical power of a lens is not only determined by its surface geometry but also by its refractive index

distribution. By combining the two effects, new approaches to chromatic as well as spherical aberration correction become possible [Beadie, *et al.*, 2008; Ji, *et al.*, 2013]. Furthermore, GRIN optics can be designed to redistribute irradiance in an optical beam and perform coherent mode conversion in beam shaping applications [Wang & Shealy, 1993; Zahreddine, *et al.*, 2008; Lin & Leger, 2012].

1.4 Overview of Thesis

The work presented in this thesis is divided into 2 distinct sections. The first section covers the measurement of GRIN elements by deflectometry, which encompasses both Chapters 2 & 3. Our methods involve solving the inverse problem of calculating an unknown refractive index distribution from measured boundary values of position and slope for probe laser beams that transit the medium. Chapter 2 focuses on measuring one-dimensional (1-D) index profiles in thick GRIN elements using a fundamental relation between beam slopes and refractive indices. In Chapter 3, we develop a more complicated mathematical model based on path integrals to address the deflection of probe laser beams in two-dimensional (2-D) and three-dimensional (3-D) index fields. Using a simplified approach, we show that the inverse problem in higher dimensions can be formulated as a linear system. The second section of this thesis covers the analysis and design of GRIN optics for beam shaping and their application in laser cavities. In Chapter 4, we develop a numerical method for computing the propagation of optical beams inside GRIN media based on the Fresnel-Kirchoff formulation of scalar wave diffraction theory. We then apply the method to compute spatial eigenmodes in optical resonators that employ GRIN elements. In addition, we provide a design procedure for achieving arbitrary coherent mode conversions using GRIN materials. Finally, a summary and outlook for future work is provided in Chapter 5.

2. Deflectometry for One-dimensional Index Profiles in Thick GRIN Media

2.1 Background Material

The advent of advanced fabrication techniques developed over the last three decades has enabled us to design, manufacture, and measure a wide variety of useful GRIN optical components. Since the refractive index of the GRIN element dictates the propagation of light inside the optical material, accurate knowledge of its index distribution is essential for integration into optical systems. Many methods for measuring 1-D GRIN profiles have been described in the literature over the years and have found application in the study of diffusion properties of particles in liquids and gases outside the realm of GRIN optics. Perhaps the most straightforward of these methods are based on the deflectometry of probe laser beams used for interrogation. Such methods generally use either beam displacement [Wiener, 1893; Barnard, 1975] or angular deflection [Bodnar & Ratajczyk, 1965] to provide a direct measurement of the index gradient. However, due to the simplifying geometrical assumptions, the measurements are only reliable when the index gradient remains relatively constant across the entire trajectory of the probe beam inside the medium. Interferometric measurement techniques that use fringe patterns to recover the phase profile of an interrogating beam have also been described [Saunders, 1969; Iwata, 1970; Junginger & Haeringer, 1972; Moore & Ryan, 1978]. However, resolving fringes becomes impractical for moderately steep gradients or if the interrogating beam incurs a substantial propagation distance inside the medium. Furthermore, ambiguity in the direction of the index gradient can result from such interferometric measurements without prior knowledge of the index variation present in the GRIN element. This restricts the application of interferometric methods to specific types of index profiles, such as those produced by diffusion-based processes where the profiles are restricted to regions of monotonic index variations. Older (and more complicated) methods that are typically used to study 1-D index gradients in fluids include prism methods [Kovalskii & Polanskii, 1966; Miller & Malisek, 1968] and moiré patterns [Nishijima, *et al.*, 1964; Oster, *et al.*, 1964]. These methods are generally difficult to use and lack practical methods of automatic data acquisition. Optical coherence tomography (OCT) is capable of providing high-resolution imaging of layered GRIN profiles [Meemon, *et al.*, 2013], but the principles behind this approach require scattering elements, such as discontinuities in the index, inside the medium.

In this chapter, we will develop a numerical approach to calculating 1-D index profiles in thick GRIN media based on deflectometry principles; the presented material expands on our previous work [Lin, *et al.*, 2013]. Our nondestructive measurement technique uses boundary value measurements of position and slope for a series of probe beams used to interrogate the GRIN element. We begin in Section 2.2 by first developing the fundamental relation on which our method is based; we then describe a bootstrap algorithm for recovering the entire 1-D index profile from the boundary measurements starting at a location

of known index. For the purposes of verifying our new measurement technique, we provide an alternative approach based on an older beam displacement method in Section 2.3, where we improve upon its mathematical model of beam displacement and apply a corrective procedure in computing the index profile. The ramifications of beam shaping effects for finite-width optical beams used to interrogate GRIN media are discussed in Section 2.4. In addition, a general procedure for converting the field description of an optical beam to an equivalent ray distribution (and vice versa) is established for later use. In Section 2.5, we demonstrate the reconstruction of a hypothetical 1-D index profile from simulated deflectometry data using the methods described in Sections 2 & 3, and a comparison of the results from both methods is made. An error analysis of the bootstrap algorithm is provided in Section 2.6. Finally, the experimental measurement of a real GRIN sample is detailed in Section 2.7.

To preface the mathematical tools we will develop later in this chapter, we provide a brief introduction to the description of waves in inhomogeneous media within the framework of geometrical optics [Elmore & Heald, 1985] and the fundamental equations that lay out the foundations for our work. Let us consider a wave quantity ψ that satisfies the scalar wave equation,

$$\nabla^2\psi = \frac{\partial^2}{\partial x^2}\psi + \frac{\partial^2}{\partial y^2}\psi + \frac{\partial^2}{\partial z^2}\psi = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi \,, \tag{2.1.1}$$

where $c = c(x, y, z)$ is the wave's propagation speed inside the optical medium. Physically, ψ can describe the amplitude of the electric field associated with an electromagnetic (EM) wave, for instance. We assume a trial solution of the form

$$\psi(x, y, z, t) = A(x, y, z)e^{i\left[\frac{2\pi}{\lambda_0}S(x,y,z)-\omega t\right]} \,, \tag{2.1.2}$$

where it is understood that the real part of the complex expression represents the physical value of the field. In the expression, ω is the frequency of oscillation for the EM wave and $\lambda_0 = \frac{2\pi c}{\omega}$ is its wavelength. For the time being, let us examine the trial solution at a snapshot in time so that the time dependence in Equation 2.1.2 can be omitted,

$$\psi(x, y, z) = A(x, y, z)e^{i\frac{2\pi}{\lambda_0}S(x,y,z)} \,. \tag{2.1.3}$$

$S(x, y, z)$ in the expression is termed the *Eikonal*; its significance can be understood by observing the electric field along constant contours of $S(x, y, z)$. These contours represent *geometrical wavefronts* along which the phase of the wave remains constant. As we vary the spatial coordinates x, y and z, the phase of the EM wave changes along with the amplitude of the electric field. It is typically assumed that $A(x, y, z)$ is slow-varying relative to spatial variations produced by $\Re e\left\{e^{i\frac{2\pi}{\lambda_0}S(x,y,z)}\right\}$ such that $A(x, y, z)$ can be treated as an envelope function.

Using a generalized definition of the refractive index,

$$n(x,y,z) \equiv \frac{c_0}{c(x,y,z)}\,, \tag{2.1.4}$$

where c_0 is the speed of light in vacuum, Equation 2.1.1 can be rewritten as

$$\nabla^2\psi = \frac{\partial^2}{\partial x^2}\psi + \frac{\partial^2}{\partial y^2}\psi + \frac{\partial^2}{\partial z^2}\psi = \frac{n}{{c_0}^2}\frac{\partial^2}{\partial t^2}\psi\,. \tag{2.1.5}$$

We note that the refractive index is a dimensionless quantity and is measured in refractive index units (RIU). Substituting the trial solution from Equation 2.1.2 into Equation 2.1.5 yields the simultaneous equations

$$\nabla^2 A + [n^2 - (\nabla S)\cdot(\nabla S)]\left(\frac{2\pi}{\lambda_0}\right)^2 A = 0\,, \tag{2.1.6a}$$

$$2(\nabla S)\cdot(\nabla A) + (\nabla^2 S)A = 0\,. \tag{2.1.6b}$$

Under the slow-varying assumption for $A(x,y,z)$, the $\nabla^2 A$ term in Equation 2.1.6a can be neglected. This simplifies Equation 2.1.6a to

$$|\nabla S|^2 = n^2\,, \tag{2.1.7}$$

which is known as the *Eikonal equation*. From this equation, we infer that

$$\nabla S = n(x,y,z)\hat{s}(x,y,z)\,, \tag{2.1.8}$$

where $\hat{s}(x,y,z)$ is a unit vector pointing in the direction of ∇S. It is well-known from elementary vector calculus that ∇S lies orthogonal to constant contours of $S(x,y,z)$. Hence, $\hat{s}(x,y,z)$ is always perpendicular to the geometrical wavefronts of the EM wave.

Continuous curves can be constructed from $\hat{s}(x,y,z)$, called *rays*, that intersect all geometrical wavefronts at a perpendicular angle. To make further progress, we need to eliminate the Eikonal S from Equation 2.1.8. We achieve this by examining the rate of change of $n(x,y,z)\hat{s}(x,y,z)$ along the ray,

$$\frac{d}{ds}(n\hat{s}) = \frac{d}{ds}(\nabla S) = \hat{s}\cdot\nabla(\nabla S) = \frac{\nabla S}{n}\cdot\nabla(\nabla S)$$

$$= \frac{1}{2n}\cdot\nabla(\nabla S)^2 = \frac{1}{2n}\cdot\nabla n^2 = \nabla n\,, \tag{2.1.9}$$

where we have made repeated use of Equation 2.1.8. Equation 2.1.9 is sometimes called *ray equation*; it enables us to find the trajectory of a ray given the refractive index distribution $n(x,y,z)$ and the initial ray vector $\hat{s}_0$.

2.2 Ray Deflection Method

In this section, we will continue developing the mathematics from the ray equation to establish a fundamental relation between the refractive index values along a ray trajectory in a thick 1-D GRIN element and the ray slopes at their respective locations. We will then provide two interrogation schemes for recovering the index profile from boundary value measurements of ray position and slope using knowledge of the refractive index at some location. We start our analysis by following a derivation that closely parallels the mathematics used for measuring axial, radial, and spherical refractive index profiles [Teichman, 2013] and using it to simplify the ray equation for 1-D index profiles $n = n(y)$. Starting with the ray equation from Equation 2.1.9, we have

$$\frac{d}{ds}\left(n\frac{d\vec{r}}{ds}\right) = \nabla n\,, \tag{2.2.1}$$

where ds is the arc length along the ray trajectory and $\vec{r} = x\hat{\imath} + y\hat{\jmath}$ is the position vector of a point along the trajectory in 2-D space. Assuming ray trajectories inside the medium are of the form $y = y(x)$, Equation 2.2.1 can be expressed in Cartesian coordinates as follows. The first derivative with respect to arc length s in Cartesian coordinates is

$$\frac{d}{ds} = \frac{dx}{ds}\frac{d}{dx} = f\frac{d}{dx}\,, \tag{2.2.2}$$

where $f = \frac{dx}{ds} = \frac{dx}{\sqrt{dx^2+dy^2}} = \frac{1}{\sqrt{1+\left(\frac{dy}{dx}\right)^2}}$ using the fact that $ds^2 = dx^2 + dy^2$. The second derivative is then

$$\frac{d^2}{ds^2} = f\frac{d}{dx}\left(f\frac{d}{dx}\right) = f\frac{df}{dx}\frac{d}{dx} + f^2\frac{d^2}{dx^2}\,. \tag{2.2.3}$$

Rewriting Equation 2.2.1 in scalar form for each Cartesian component, we have

$$\frac{dn}{ds}\frac{dx}{ds} + n\frac{d^2x}{ds^2} = \frac{\partial n}{\partial x}\,, \tag{2.2.4a}$$

$$\frac{dn}{ds}\frac{dy}{ds} + n\frac{d^2y}{ds^2} = \frac{\partial n}{\partial y}\,. \tag{2.2.4b}$$

With the help of Equation 2.2.2 & 2.2.3, Equation 2.2.4a becomes

$$\left(f\frac{dn}{dx}\right)\left(f\frac{dx}{dx}\right) + n\left(f\frac{df}{dx}\frac{dx}{dx} + f^2\frac{d^2x}{dx^2}\right) = \frac{\partial n}{\partial x}\,,$$

$$f^2\frac{dn}{dx} = \frac{\partial n}{\partial x} - nf\frac{df}{dx}\,, \tag{2.2.5a}$$

where we have used $\frac{dx}{dx} = 1$ and $\frac{d^2x}{dx^2} = 0$. Similarly, Equation 2.2.4b becomes

$$\left(f\frac{dn}{dx}\right)\left(f\frac{dy}{dx}\right) + n\left(f\frac{df}{dx}\frac{dy}{dx} + f^2\frac{d^2y}{dx^2}\right) = \frac{\partial n}{\partial y}\,,$$

$$\frac{dy}{dx} f^2 \frac{dn}{dx} + n\left(f\frac{df}{dx}\frac{dy}{dx} + f^2\frac{d^2y}{dx^2}\right) = \frac{\partial n}{\partial y}\,. \tag{2.2.5b}$$

Substituting Equation 2.2.5a into Equation 2.2.5b, canceling terms, and dividing by n yields

$$\frac{dy}{dx}\frac{\partial n}{\partial x} - \frac{dy}{dx} nf\frac{df}{dx} + nf\frac{df}{dx}\frac{dy}{dx} + nf^2\frac{d^2y}{dx^2} = \frac{\partial n}{\partial y}\,,$$

$$\frac{1}{n}\frac{\partial n}{\partial y} - \frac{1}{n}\frac{\partial n}{\partial x} y' = \frac{y''}{1+(y')^2}\,, \tag{2.2.6}$$

where the primes indicate differentiation with respect to x. Since the index does not vary along x, we have $\frac{\partial n}{\partial x} = 0$ and Equation 2.2.6 simplifies to

$$ny''\frac{1}{1+(y')^2} = \frac{\partial n}{\partial y}\,. \tag{2.2.7}$$

Multiplying both sides by $2y'$ and rearranging, we find that

$$\frac{2y'y''}{1+(y')^2} = 2y'\frac{1}{n}\frac{\partial n}{\partial y}\,, \tag{2.2.8a}$$

which can be expressed in differential form as

$$\frac{d}{dx} ln[1 + (y')^2] = \frac{d}{dx} 2ln(n)\,. \tag{2.2.8b}$$

It is straightforward to show that carrying out the chain rule of differentiation on Equation 2.2.8b yields Equation 2.2.8a. Integrating both sides of Equation 2.2.8b and taking their exponents yields

$$1 + (y')^2 = Cn^2\,, \tag{2.2.9}$$

where C is some unknown constant. If the index and ray slope are known at some location along the ray trajectory, we can use this information to apply boundary values to the general solution in Equation 2.2.9 and eliminate C in the expression, *e.g.*,

$$\frac{1+(y')^2}{1+(y_0')^2} = \frac{n^2}{n_0^2}\,, \tag{2.2.10}$$

where the index n_0 is known at position $\vec{r}_0 = x_0\hat{\imath} + y_0\hat{\jmath}$ along the ray trajectory. Under this scenario, Equation 2.2.10 can be used to ascertain index values along the entire ray trajectory if local ray slopes are readily available from measurement, *e.g.*, imaged from scattered light. For our proposed measurement scheme based on nondestructive deflectometry, however, we assume ray slopes inside the GRIN medium to be inaccessible for direct measurement. Moreover, the ray slopes obtained at the boundaries are measured externally in the ambient medium and changes according to Snell's law,

$$n_{amb} sin\theta_{amb} = nsin\theta\,, \tag{2.2.11}$$

upon entering or exiting the GRIN medium, where n and θ correspond to the index and angle (relative to the normal of the local interface) inside the GRIN medium, respectively, and n_{amb} and θ_{amb} refer to their values in the ambient medium. In order to obtain an analogous expression to Equation 2.2.10 for external ray slope values, it is more convenient to rewrite Equation 2.2.11 as

$$n_{amb}\frac{y'_{amb}}{\sqrt{1+(y'_{amb})^2}} = n\frac{y'}{\sqrt{1+(y')^2}}, \tag{2.2.12}$$

where we have used the trigonometric relation $sin\theta = \frac{dy}{\sqrt{dx^2+dy^2}} = \frac{y'}{\sqrt{1+(y')^2}}$. Rearranging the above expression, we find that

$$1 + (y')^2 = \frac{1+(y'_{amb})^2}{1+\left(1-\frac{n_{amb}{}^2}{n^2}\right)(y'_{amb})^2}. \tag{2.2.13}$$

Replacing the ray slopes in Equation 2.2.10 with external ray slopes and utilizing Equation 2.2.13 yields

$$\frac{1+(y')^2}{1+(y'_0)^2}\left[\frac{1+\left(1-\frac{n_{amb}{}^2}{n_0{}^2}\right)(y'_0)^2}{1+\left(1-\frac{n_{amb}{}^2}{n^2}\right)(y')^2}\right] = \frac{n^2}{n_0{}^2}, \tag{2.2.14}$$

where it is understood that ray slopes in the expression refer to external values in the ambient medium. Rearranging and expressing n^2 in terms of the known quantities, we find that

$$n^2 = n_0{}^2 + n_{amb}{}^2\left[\frac{(y')^2}{1+(y')^2} - \frac{(y'_0)^2}{1+(y'_0)^2}\right]. \tag{2.2.15}$$

Equation 2.2.15 is a fundamental relation that can be used to compute the index at one end of the interrogating ray from a known index value at the opposite end, which we illustrate in Figure 2.2.1, provided that the external ray slopes at both ends are known. For instance, if we assume the index is known at y_{in} in the figure, then y'_{in} replaces y'_0 in Equation 2.2.15 and y'_{out} replaces y'. Likewise, n refers to $n(y_{out})$ in the expression and $n_0 = n(y_{in})$.

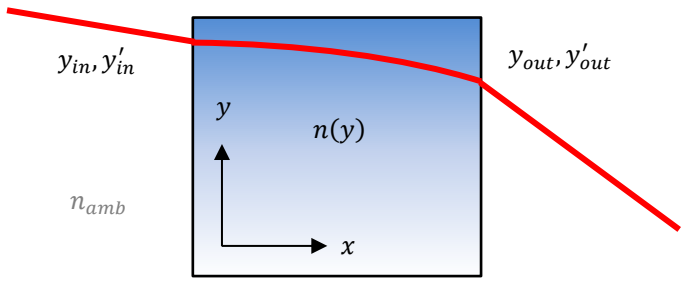


Figure 2.2.1. Geometry of the fundamental relation expressed in Equation 2.2.15.

In theory, it is possible to develop an integral relation along the entire trajectory of the interrogating ray and invert a series of integrals across multiple interrogating rays to solve for the index distribution [Beliakov & Buckley, 2003]. However, the inversion process does not guarantee the existence of a solution for noisy data and numerical techniques used to deal with the noise tend to be cumbersome. On the other hand, applying Equation 2.2.15 over a full sweep of interrogating rays launched from a location of known index can be used to ascertain all index values at their exit locations using only external ray slopes measured along the boundaries of the GRIN element. We illustrate this geometry in Figure 2.2.2. Using this interrogation scheme, a continuum of exit ray positions and associated external ray slopes fully describes $n(y)$.

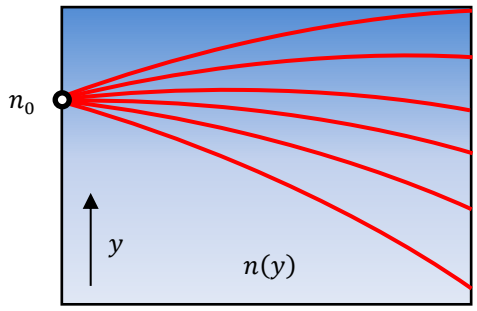


Figure 2.2.2. Full sweep measurement of a 1-D refractive index profile from a location of known index.

In practice, the sweep range of the interrogating rays is restricted by total internal reflection (TIR) conditions specific to the discontinuities in refractive index at the boundary interface. This effectively limits the region of the GRIN element that can be measured using the interrogation scheme depicted in Figure 2.2.2. In addition, the index is typically known at the ends of the GRIN element, especially for diffusion-based fabrication techniques. In order to address these issues, we propose a different interrogation scheme based on bootstrapping multiple interrogating rays, which we illustrate in Figure 2.2.3. In this measurement scheme, we use angled incidence to effectively couple the fundamental relation in Equation 2.2.15 across multiple interrogating rays.

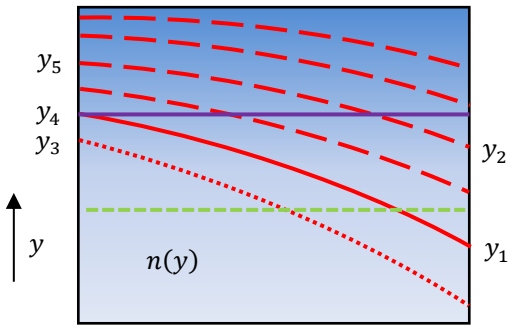


Figure 2.2.3. Angled incidence for bootstrapping refractive index calculations using Equation 2.2.15.

The angle of incidence is chosen to ensure there is at least one interrogating ray traveling from known region(s) of the index profile to the unknown region(s). The expansion of the known region through the augmentation of newly calculated refractive indices is illustrated in Figure 2.2.3, where the refractive index $n(y)$ is initially known below the dashed horizontal line. The interrogating ray whose boundary location is the farthest from the initially known region in the figure has output position y_1 and input position y_4, depicted as a solid line. Using Equation 2.2.15, we can compute $n(y_4)$ from the externally measured ray slopes at y_1 and y_4. $n(y_3)$ is calculated in a similar manner from the interrogating ray depicted as a dotted line. Then, using the calculated index values $n(y_3)$, $n(y_4)$ as well as the index from the initially known region, any index value for $y < y_4$ can be interpolated. This effectively increases our knowledge of $n(y)$ and expands the known region up to y_4.

The process of expanding the known index region through the application of Equation 2.2.15 to all applicable rays defines one bootstrap in the algorithm. In the first bootstrap that we have just described, the known region was expanded up to y_4, allowing the interrogating ray with input position y_5 and output position y_2 to contribute to the next bootstrap. In this manner, each bootstrap in the overall procedure will yield new information about the index profile, provided that a ray traveling from a known region to the unknown region(s) exists. This process is repeated until all interrogating rays have been exhausted and the entire index distribution has been recovered. We hasten to add that the bootstrap process can terminate prematurely in the presence of steep index gradients if the angle of incidence is too shallow and the interrogating rays are bent aggressively toward the unknown region(s) such that both boundary ray points reside in the unknown region(s). However, one can generally guarantee the completion of the algorithm by choosing the appropriate incident angles and using a sufficient number of interrogating rays.

2.3 Ray Displacement Method

In order to assess the performance of the bootstrap algorithm detailed in the previous section, we will develop the mathematics for a measurement method used to measure 1-D index profiles based on ray displacement [Barnard & Ahlborn, 1975]. Barnard's method is based on measurements of overall displacement across multiple interrogating rays and makes use of several geometrical approximations to simplify the mathematical model. Hence, it is expected to yield larger errors in the computed index profile compared to the bootstrap method. Improving upon Barnard's approximate model, we will provide a correction term to his simplified expressions. In addition, we will employ an iterative ray trace algorithm developed at the Institute of Optics at the University of Rochester [McCarthy, 2015] to optimize the calculated index profile and minimize reconstruction error. We illustrate the measurement geometry for Barnard's ray displacement method in Figure 2.3.1. With the appropriate assumptions, the local index gradient $\frac{\partial n}{\partial y}$ can be related to the total displacement Δ by

$$\frac{\partial n}{\partial y} \approx \frac{\Delta}{LD}, \tag{2.3.1}$$

where L is the thickness of the GRIN medium and D is the propagation distance from the exit surface to the measurement plane.

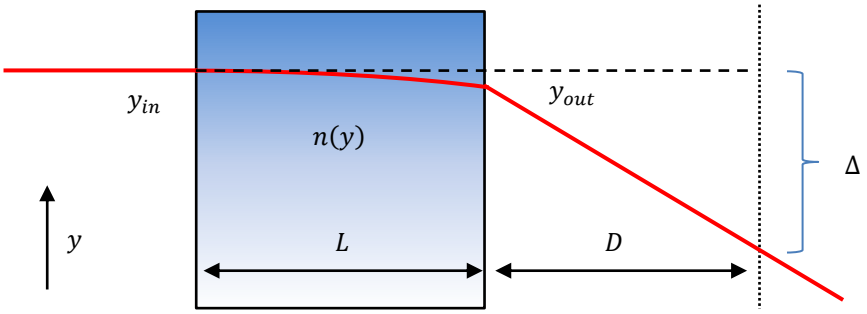


Figure 2.3.1. Measurement geometry for Barnard's ray displacement method.

The measured displacement $\Delta(y)$ obtained from multiple interrogating rays can be integrated to arrive at the index distribution $n(y)$ if the refractive index is known at some location. While Barnard's method allows for data acquisition in a single measurement plane, its accuracy is ultimately limited by the geometrical assumptions associated with its simplified model. For instance, Barnard assumes the displacement inside the GRIN medium is negligible and does not account for it in his calculations. The difference between the actual geometry of the interrogating ray and Barnard's model is illustrated in Figure 2.3.2.

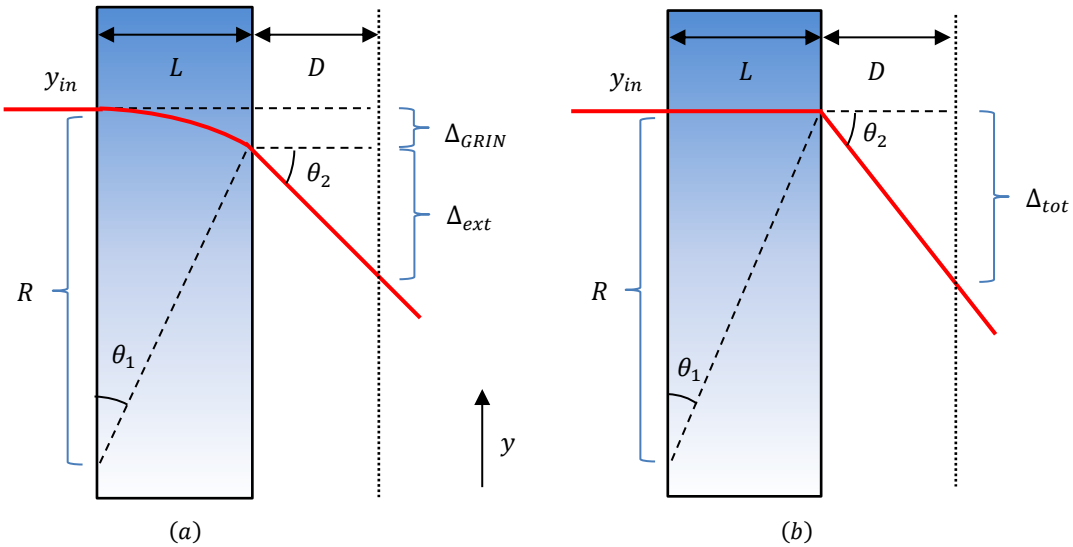


Figure 2.3.2. Difference between (a) actual deflection geometry and (b) Barnard's simplified model.

Improving upon Barnard's simplified model, let us retain the 2nd order term in small angle approximations. This results in a nonzero contribution to the overall displacement occurring strictly inside the GRIN medium,

$$\Delta_{GRIN} = R[1 - cos(\theta_1)] = n\left(\frac{\partial n}{\partial y}\right)^{-1}\frac{1}{2}{\theta_1}^2\,, \tag{2.3.2}$$

where we use a circular arc to approximate the trajectory of a normally incident ray under a constant index gradient along y, whose radius R is given by

$$\frac{1}{R} = \frac{1}{n}\frac{\partial n}{\partial y}\,. \tag{2.3.3}$$

It is straightforward to show that $y'' = \frac{1}{R}$ when $y' = 0$ for a circle of radius R from the geometrical formula $x^2 + y^2 = R^2$. The expression in Equation 2.3.3 can then be deduced directly from Equation 2.2.6, which becomes

$$y'' = \frac{1}{n}\frac{\partial n}{\partial y}\,, \tag{2.3.4}$$

where we have used the fact that $\frac{\partial n}{\partial x} = 0$ and the approximation $y' \approx 0$ for the normally incident ray. The validity of the circular arc approximation will illustrated through ray trace results at the end of this section.

The output angle θ_2 in Figure 2.3.2 is related to the total displacement Δ_{tot} from the input position and the internal displacement Δ_{GRIN} by

$$\theta_2 = \frac{\Delta_{tot} - \Delta_{GRIN}}{D} = \frac{\Delta_{tot}}{D} - \frac{n{\theta_1}^2}{2D}\left(\frac{\partial n}{\partial y}\right)^{-1}\,, \tag{2.3.5}$$

where we have made use of Equation 2.3.2 and the 2nd order small angle approximation $tan(\theta) = \theta$. Application of Snell's law at the exit surface of the GRIN element under the same small angle approximation yields

$$\theta_2 = n\theta_1\,, \tag{2.3.6}$$

where we have assumed the index of the ambient medium to be unity. From the depicted geometry, we can also relate θ_1 to the thickness of the GRIN element and the local index gradient by

$$\sin\left(\frac{\theta_1}{2}\right) = \frac{\theta_1}{2} = \frac{1}{2R}\sqrt{L^2 + {\Delta_{GRIN}}^2}\,. \tag{2.3.7}$$

With the help of Equation 2.3.3, Equation 2.3.7 becomes

$$\theta_1 = \frac{L}{n}\frac{\partial n}{\partial y}\sqrt{1 + \frac{{\Delta_{GRIN}}^2}{L^2}} \approx \frac{L}{n}\frac{\partial n}{\partial y}\left(1 + \frac{{\Delta_{GRIN}}^2}{2L^2}\right)\,, \tag{2.3.8}$$

where we have assumed that $\Delta_{GRIN} \ll L$ under small angle approximations. Substituting Equation 2.3.2 into Equation 2.3.8 for Δ_{GRIN} and neglecting the resulting 4th order term in θ_1, we find that

$$n\theta_1 = L\frac{\partial n}{\partial y}. \tag{2.3.9}$$

According to Equation 2.3.6, we can equate the right-hand sides of Equation 2.3.5 and Equation 2.3.9,

$$L\frac{\partial n}{\partial y} = \frac{\Delta_{tot}}{D} - \frac{n{\theta_1}^2}{2\frac{\partial n}{\partial y}D} = \frac{\Delta_{tot}}{D} - \frac{L^2}{2nD}\frac{\partial n}{\partial y}. \tag{2.3.10}$$

where we have made use of Equation 2.3.9 in the intermediate expression. Solving for the index gradient yields

$$\frac{\partial n}{\partial y} = \frac{\Delta_{tot}}{LD+\frac{L^2}{2n}} = \frac{\Delta_{tot}}{LD}\frac{1}{1+\frac{L}{2nD}}. \tag{2.3.11}$$

Note that the index n appears on the right-hand side in the differential equation. If we assume that the index range is small for the index profile under investigation, then substituting an approximate index value (*e.g.*, a known index value at some location) for n in the above expression will suffice for an initial computation of the index profile. In principle, $n(y)$ can be solved recursively by substituting the calculated $n(y)$ back into Equation 2.3.11 and then recalculating $n(y)$ until successive reconstructions of the index profile are in agreement.

In order to improve the accuracy of the ray displacement method, the University of Rochester has developed an iterative correction procedure based on a numerical ray trace through the reconstructed refractive index profile [McCarthy, 2015]. Using the initial computation of Equation 2.3.11 as a starting point, the ray trace computes the overall displacement of normally incident rays propagating through the calculated index profile over the thickness of the GRIN medium and over the propagation distance from the exit surface to the measurement plane. Cubic splines are used in the implementation of the procedure to parametrize $n(y)$, and the value of spline knots are optimized based on the difference between measured ray displacement data and the calculated displacement obtained through the ray trace. The values of the spline knots are refined following each ray trace until the trace results are consistent with the measured data. We note that employing this corrective procedure to optimize the calculated index profile allows us to forgo the aforementioned iterative process used to solve Equation 2.3.11 recursively.

It is important to note the differences in the geometry of measurement between the ray displacement method and the ray deflection method from Section 2.2. The former requires normal incidence to satisfy geometrical assumptions while angled incidence is needed in the latter method in order to guarantee successful completion of the bootstrap algorithm. More importantly, the fundamental relation (Equation 2.2.15) used in the ray deflection method is exact and does not contain any intrinsic errors. This comes at the expense of requiring slope measurements in our calculations, which can be obtained from ray

positions measured in two parallel planes, for instance. On the other hand, measurements in the ray displacement method can be conducted in a single plane.

Finally, for the purposes of checking the validity of geometrical approximations used in the ray displacement method, we provide a discrepancy plot, shown in Figure 2.3.3, between a circular arc and the actual trajectory of a ray under a constant index gradient of $\left|\frac{\partial n}{\partial y}\right| = 0.2$ RIU/cm over a propagation distance of 5 cm.

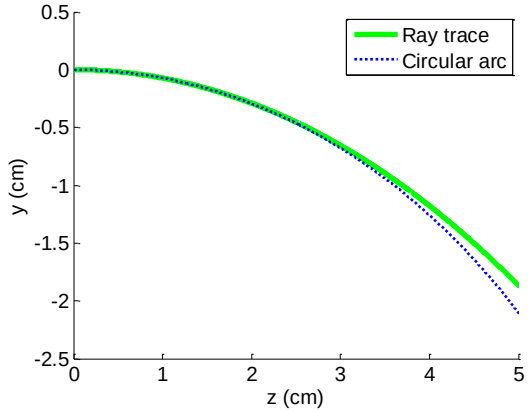


Figure 2.3.3. Comparison between a circular arc and ray trace results under a constant linear index gradient.

The amount of ray displacement inside the GRIN medium under this scenario should raise concern because the accuracy of the circular arc approximation deteriorates as the ray deviates from a predominantly horizontal trajectory. Furthermore, small angle approximations used in the model break down as the ray is directed away from the horizontal axis in an aggressive manner. This restriction can lead to large errors in the internal ray displacement for (1) steep index gradients or (2) long propagation lengths inside the GRIN medium.

2.4 Beam Shaping Analysis

Before we can demonstrate the measurement methods discussed in the preceding sections, it is important to discuss the non-ideal beam shaping effects associated with optical beams of finite width used for interrogation. In this section, we will develop a numerical procedure for converting the field description of any optical beam to an equivalent ray distribution based on energy conservation laws. This technique is particularly useful in evaluating the performance of our deflectometry methods as well as designing GRIN mode converters in Chapter 4. To address the issue at hand, let us represent a finite-width optical beam as a collective bundle of rays that transit the medium, as shown in Figure 2.4.1.

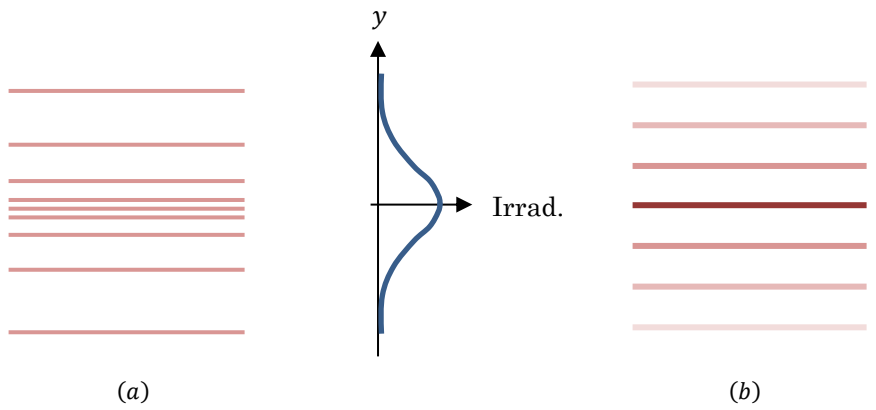


Figure 2.4.1.Ray representation for a finite-width optical beam using (a) the local density of rays with equal amounts of radiant energy and (b) varying amounts of radiant energy associated with uniformly-spaced rays, where lighter shades correspond to less energy.

Each ray contained inside the bundle is associated with a finite amount of radiant energy. The geometrical optics intensity law for energy propagation states that the distribution of ray positions can be interpreted as a direct mapping of the optical beam's irradiance profile; this law has been used to calculate redistribute the irradiance of an optical beam using two aspheric refractive surfaces [Rhodes & Shealy, 1980]. According to this principle, the local ray density represents the irradiance profile of an optical beam, as seen in Figure 2.4.1a, where all rays carry equal amounts of radiant energy. In the figure, the spatial distribution of the constituent rays reflects the irradiance profile of a collimated optical beam with a Gaussian irradiance profile along y. Alternatively, Figure 2.4.1b illustrates a different scheme used to represent the same optical beam using varying amounts of radiant energy for the constituent rays in the collective bundle. Both representations are equally valid if the distribution of radiant energy calculated from the ray positions matches the optical beam's irradiance profile. Of course, a combination of the two schemes can also be used.

Using the ray representation, it is much easier to see how a GRIN medium might alter the irradiance profiles of finite-width optical beams. As is evident from Equation 2.2.1, local index gradients induce bends in the constituent rays' trajectories as they propagate through the GRIN medium, modifying their positions relative to each other by the time they reach the output plane and thereby distorting the overall irradiance profile of the optical beam. If the index gradient is always constant across the transverse profile of a collimated optical beam as it propagates through the GRIN medium and all the constituent rays remain parallel to each other, then the distortion effects are restricted to stretching and squeezing of the entire irradiance profile as a whole. In the general case, however, any change in the index gradient across the optical beam will redistribute the positions of constituent rays relative to each other.

In order to assign values to the measured position of a finite-width optical beam without ambiguity, it is useful to establish a method for representing its position using a single ray. This begs the

question of how the propagation behavior of a finite-width optical beam differs from that of an ideal ray inside the GRIN medium. Specifically, what location along the measured output irradiance profile of the optical beam should be designated as its measured position? Based on energy conservation laws, we argue that the median along the optical beam's irradiance profile provides a good estimate of its measured position. Our reasoning is as follows: in the limit of infinitesimally small beam width, the optical beam is reduced to an ideal ray, and Equation 2.2.1 dictates its trajectory. Next, suppose that as the beam width increases, we maintain the ideal ray trajectory as is and introduce additional constituent rays to the overall distribution on both sides of the ideal ray to describe the beam's irradiance profile, which now occupies a finite extent in the transverse direction. It is then straightforward to show through energy conservation that if the individual ray trajectories in the bundle do not intersect with the ideal ray's trajectory, then the ratio of radiant energies contained on either side of the ideal ray remains constant. Therefore, if we place the ideal ray at the median of the optical beam's irradiance profile under this scenario, the position of the ideal ray will always coincide with the median. We illustrate this principle in Figure 2.4.2, where all constituent rays are assigned equal amounts of radiant energy. It is apparent in the figure that the radiant energies contained on either side of the ideal ray are always equal to each other between the two planes.

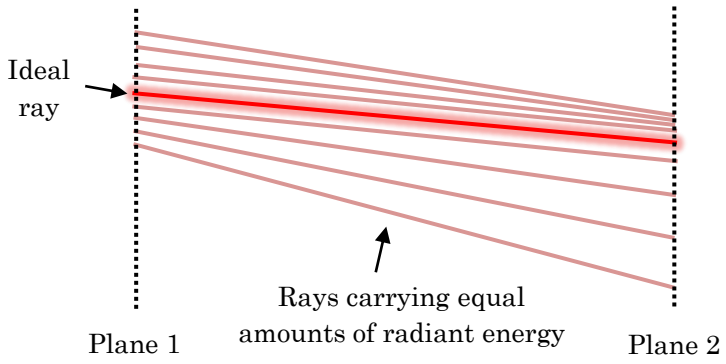


Figure 2.4.2. Ideal ray trajectory following the median of the optical beam's irradiance profile.

Therefore, it is perfectly reasonable to designate the median of an optical beam's measured irradiance profile as its measured position, provided that constituent rays in its ray representation do not cross the ideal ray between any two planes of measurement. If the constituent rays violate this restriction, a discrepancy between the median of the beam's irradiance profile and the ideal ray trajectory will develop, and the beam's measured position becomes ambiguous. For our purposes, we assume that such circumstances do not occur in our measurements; any ambiguity resulting from violating this restriction is treated as a source of error in the measured beam position. We hasten to add that aside from the beam shaping effects associated with the GRIN medium, other sources can also contribute to the distortion of an optical beam's irradiance profile in the measurement system. These include roughness, tilt, and opacity of the GRIN element's boundary surfaces. If the model does not explicitly account for these mechanisms, they are treated as sources of measurement error as well.

It follows that beam shaping effects can also be used, in theory, to interrogate the entire 1-D index profile of the GRIN medium. This is possible because the distortion in the optical beam's irradiance profile can be treated as a distributed version of deflectometry if we were to examine the deflection of individual rays in the collective bundle. However, in order to match the individual rays in the input distribution with those in the output distribution without ambiguity, any ray crossing between the two observation planes is forbidden. In practice, the positions of individual rays in the measurement planes are extremely sensitive to noise in the measured irradiance profile. The combination of these factors generally renders the conversion between the measured irradiance profile of a wide interrogating beam and its equivalent ray distribution unreliable.

In the remainder of this section, we will briefly describe a numerical procedure for converting the irradiance profile of a 1-D optical beam into an equivalent ray distribution. We start by integrating the irradiance $I(y)$ along y,

$$H(y) = \int_{y_1}^{y_2} I(y)\, dy\,, \tag{2.4.1}$$

where $H(y_1) = 0$ and $I(y)$ is assumed to be negligible for $y < y_1$ and $y > y_2$. Note that $H(y)$ is a monotonic function in y since irradiance is a strictly positive quantity. Next, a stepwise function is used to approximate $H(y)$ so that each discontinuity in the approximation corresponds to the inclusion of an additional ray in the collective bundle contributing a discrete amount of radiant energy. This stepwise function can be interpreted as follows: Starting at $y = y_1$ and moving in the $+y$ direction, a new constituent ray is added to the overall ray distribution whenever we encounter a discontinuity in the stepwise approximation; the ray is placed at the location of the discontinuity and its radiant energy is equal to the size of the discontinuity. We depict this conversion mapping for a 1-D optical beam with a Gaussian irradiance profile in Figure 2.4.3. In this particular case, $H(y)$ is an *erf* function in y and is evaluated starting at the center of the beam due to symmetry.

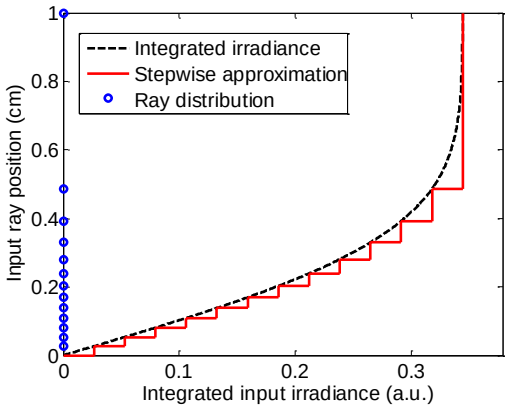


Figure 2.4.3. Conversion mapping between a 1-D irradiance profile and its equivalent ray representation.

Recalling that rays lie perpendicular to geometrical wavefronts and their OPLs are identical along these wavefronts (see Section 2.1), it is also possible to specify the optical beam's phase profile using ray angles in the collective ray distribution. Let us suppose that the field disturbance of the optical beam has the form

$$\psi(y,z) = A(y,z)e^{ik_0 S(y,z)}\ , \tag{2.4.2}$$

where $k_0 = \frac{2\pi}{\lambda_0}$ is the wavenumber in free space, $|A(y,z)|^2$ is proportional to the optical beam's irradiance, and constant contours of the Eikonal $S(y,z)$ describe geometrical wavefronts. Taking the projection of $\psi(y,z)$ in the y-plane at any particular value $z = z_0$, we differentiate $S(y,z_0)$ with respect to y in that plane to calculate the normal direction to the local wavefront and assign directions to the constituent rays accordingly. For instance, in the simple case where $\psi(y,z)$ is a plane wave propagating along z and $S(y,z_0)$ is constant, all the constituent rays are aligned parallel to the z-axis. We hasten to add that the phase information of an optical beam can also be encapsulated in the OPLs of the constituent rays. However, because the directions and the OPLs of the constituent rays are both derived from the Eikonal and must be self-consistent, specifying both quantities in a ray distribution is redundant.

The numerical procedure we have just described will be used in a later chapter to design index distributions for GRIN mode converters. For the purposes of measuring index profiles using deflectometry, we are only concerned with the beam shaping effects of the medium and apply the same principles in assigning an effective beam position for the measured irradiance profile of a finite-width optical beam.

2.5 Computing the Index Profile

We are now ready to demonstrate the reconstruction of a hypothetical test index profile in a thick GRIN element from simulated deflectometry data using the nondestructive methods detailed in Sections 2.2 & 2.3. For simplicity, the interrogation is first performed without taking beam shaping effects into consideration. We will incorporate these effects into the model in a separate simulation toward the end of this section.

We set the thickness of our hypothetical GRIN element to 0.3 cm and gave it the following index profile,

$$n(y) = \begin{cases} n_0 + \Delta n\ , & y \le -0.5 \\ n_0 + \Delta n \cdot sin(\pi y) + \frac{\Delta n}{5} \cdot e^{-100y^2}\ , & -0.5 < y < 0.5\ , \\ n_0 - \Delta n\ , & y \ge 0.5 \end{cases} \tag{2.5.1}$$

where $n_0 = 1.5$ and $\Delta n = 0.1$. Figure 2.5.1 shows a plot of the test index profile. To generate the deflectometry data to be used as measurements for both methods, we traced a set of incident interrogating rays through the GRIN medium. For 1-D index profiles, a number of implementations based on Euler's method will suffice [Southwell, 1982; Sharma, *et al.*, 1982; Richerzhagen, 1996]. During the numerical ray

trace, we employed constrained cubic splines [Kruger, 2002] for interpolating refractive indices between the discrete values of the index field, which was sampled on a 2-D uniform rectangular grid. Further details on the numerical ray trace used to generate the data are omitted in this section for brevity. We provide a more detailed overview of accurate ray tracing for 2-D index fields in Appendix 1, which is applicable here and will be utilized in later chapters.

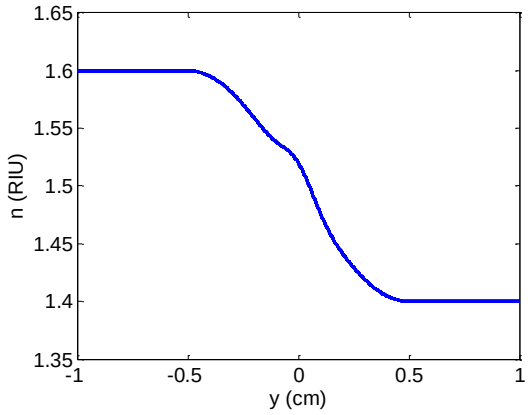


Figure 2.5.1. Index profile for a hypothetical thick GRIN element used for testing the bootstrap method.

In testing the ray deflection method of Section 2.2, we chose a set of interrogating rays that is uniformly spaced in y (separation $\Delta = 0.02$ cm) and incident at a common external angle of $-\frac{\pi}{7}$ radians relative to the horizontal axis (z). The traced trajectories of these interrogating rays inside the GRIN medium are shown in Figure 2.5.2, where the application of Snell's law at the boundary interface (not shown in figure) between the GRIN medium and the ambient medium, which we assumed to be air in this simulation, is implied.

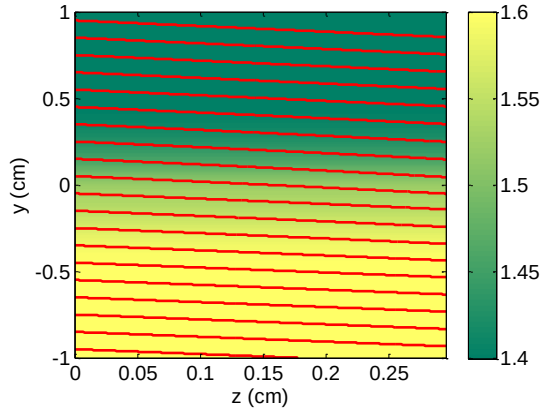
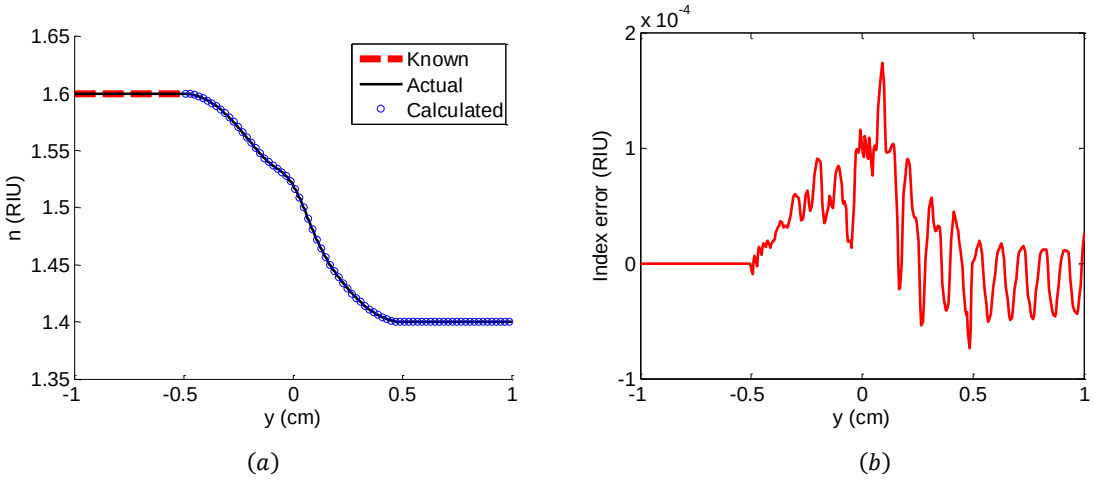


Figure 2.5.2. Traced trajectories for interrogating rays used to test the bootstrap method, where the number of rays has been reduced in the figure for illustration purposes.

The traced trajectories are discarded immediately following the ray trace as our methods of reconstruction do not assume knowledge of this information; only the external values of ray slope outside the GRIN medium and the corresponding boundary positions are retained.

The boundary values obtained from the ray trace were fed into the bootstrap algorithm from Section 2.2 for reconstructing the test index profile. For this simulation, we assumed $n(y)$ is initially known for $y \leq -0.5$ cm and use constrained cubic splines to interpolate index values within the known region of the index profile during each bootstrap. Figure 2.5.3a shows the reconstructed refractive index profile using the bootstrap algorithm, and the index error relative to the test index profile is shown in Figure 2.5.3b. Recalling that the fundamental relation used in the bootstrap algorithm is an exact analytical expression, the index error in the reconstruction can only be a manifestation of the noise sources present in measurement (*e.g.*, quantization effects associated with the numerical ray trace used to generate the measurement data). We will defer the discussion on these noise sources as well as error propagation in the bootstrap algorithm to Section 2.6.



(a) (b)

Figure 2.5.3. (a) Bootstrapped reconstruction of the index profile from external boundary values of ray slope and position and (b) its error relative to the test index profile shown in Figure. 2.5.1, with an absolute maximum value of 1.74×10^{-4} *RIU.*

For the sole purpose of providing a comparison in reconstruction accuracy, our variant of Barnard's ray displacement method developed in Section 2.3 is also used to reconstruct the test index profile. In the following simulation, we used a set of normally incident rays uniformly spaced along y (separation $\Delta = 0.02$ cm) to interrogate the same GRIN element, and the overall ray displacement is measured 0.5 cm behind the GRIN element, as shown in Figure 2.5.4.

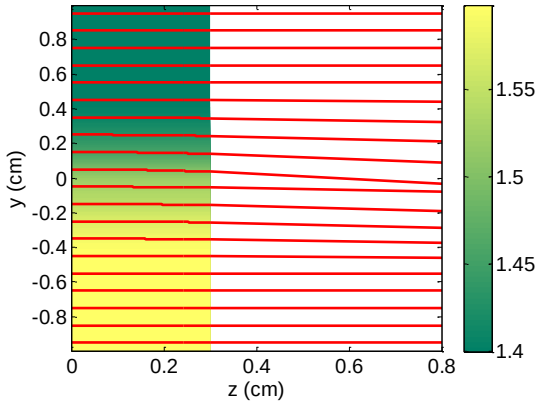
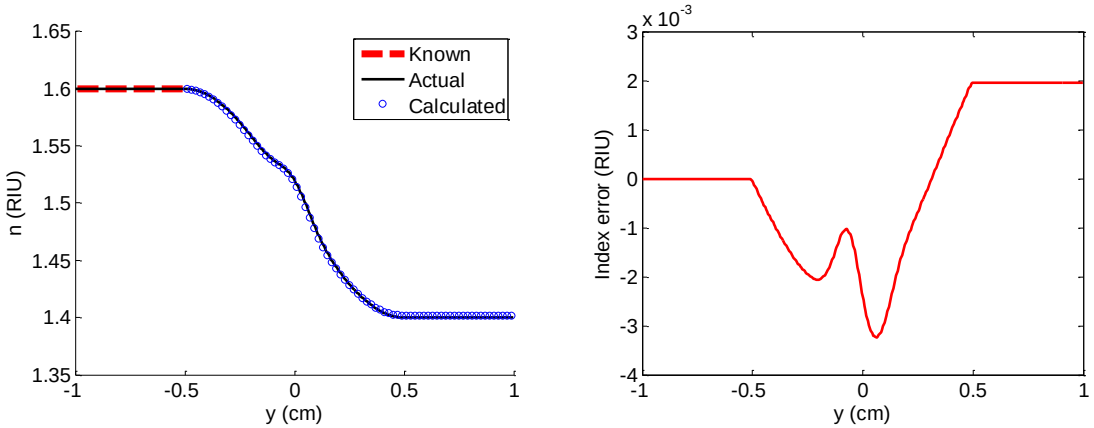


Figure 2.5.4. Traced trajectories for interrogating rays used for testing the ray displacement method, where the number of rays has been reduced in the figure for illustration purposes.

As before, we assumed that $n(y)$ is known for $y \leq -0.5$ cm and performed an initial reconstruction of the refractive index profile using Equation 2.3.11 and substituting $n = 1.6$ RIU into the right-hand side of the expression. The initial reconstruction was subsequently used to refine the reconstruction by substituting the calculated $n(y)$ into Equation 2.3.11 and reevaluating the expression in a recursive manner. We repeated the recursion process until successive reconstructions were in agreement. Figure 2.5.5 shows the resulting index profile along with its error in index relative to the test index profile shown in Figure 2.5.1.



(a) (b)

Figure 2.5.5. (a) Reconstructed index profile computed from the displacement of normally incident rays using Equation 2.3.9 and (b) its error relative to the test index profile shown in Figure 2.5.1, with an absolute maximum error of 3.20×10^{-3} *RIU.*

As is apparent from the comparison between Figures 2.5.3 & 2.5.5, the reconstruction error produced by the ray displacement method is roughly an order of magnitude larger than that of the ray deflection method employing the bootstrap algorithm. This is a direct consequence of the geometrical assumptions made in Barnard's approach, although the sensitivity to noise in the simulated data also plays a role. We then used the University of Rochester's corrective ray trace procedure to improve the reconstruction accuracy of the ray displacement method. In this procedure, the index profile was modified to minimize inconsistencies between numerical traces of the interrogating rays in the calculated solution and the measured (or simulated) data [McCarthy, 2015]. A dramatic improvement can be seen in the optimized refractive index profile in Figure 2.5.6. We hasten to add that the results of this optimization do not provide a fair comparison for assessing the performance of the bootstrap algorithm; a similar corrective procedure can be applied to improve reconstruction accuracy in the latter method as well.

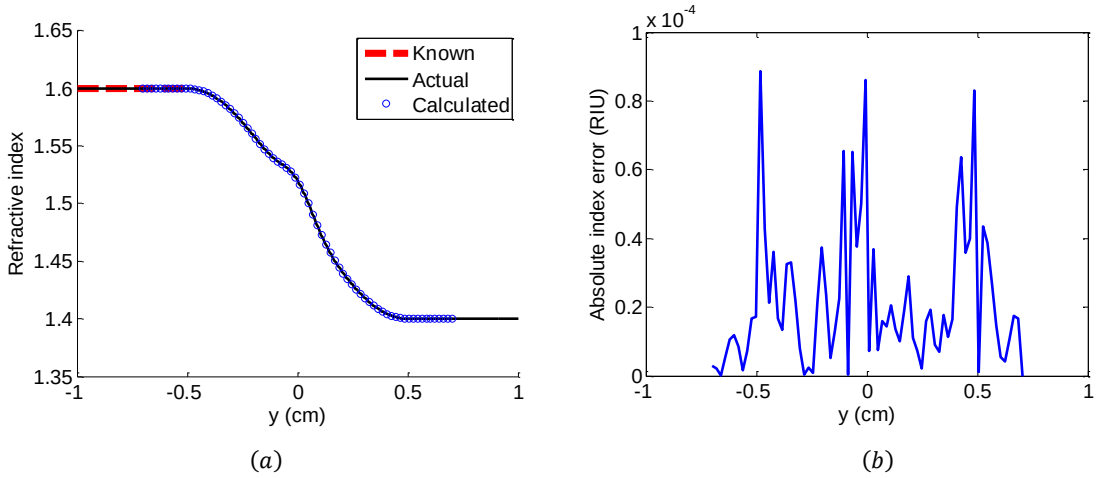


Figure 2.5.6. (a) Optimized index profile computed after applying an iterative ray trace algorithm to the ray displacement data and (b) its absolute error in index relative to the test index profile shown in Figure 2.5.1, with an absolute maximum value of 8.87×10^{-5} *RIU.*

In order to mimic experimental measurements made in a practical realization of the proposed deflectometry system and incorporate beam shaping effects into the simulation model, each interrogating ray incident upon the GRIN element in the simulation was replaced by a bundle of parallel rays representing a collimated Gaussian beam of beam width σ. The constituent rays in the individual bundles were then traced through the GRIN element and their positions taken in two parallel planes separated by distance b, as shown in Figure 2.5.7. The ray distributions in the measurement planes were subsequently converted to 1-D irradiance profiles (as outlined in Section 2.4), whose medians were used to designate the respective beam positions in those planes. From these positions, the external ray slopes were computed as

$$y'_{out} = \frac{y_b - y_a}{b}, \tag{2.5.2}$$

and extrapolated back to the exit surface of the GRIN medium to determine the boundary beam positions,

$$y_{out} = y_a - a y'_{out} \, . \tag{2.5.3}$$

The hypothetical GRIN element used in this simulation was unchanged from the previous simulations.

Upon obtaining the external boundary values, y_{out} and y'_{out}, the interrogating rays were bootstrapped to reconstruct the test index profile, identical to the process used to compute the reconstruction shown in Figure 2.5.3a. As before, we assumed $n(y)$ is initially known for $y \leq -0.5$ cm. Figure 2.5.8a shows the resulting index profile as well as its error relative to the test index profile.

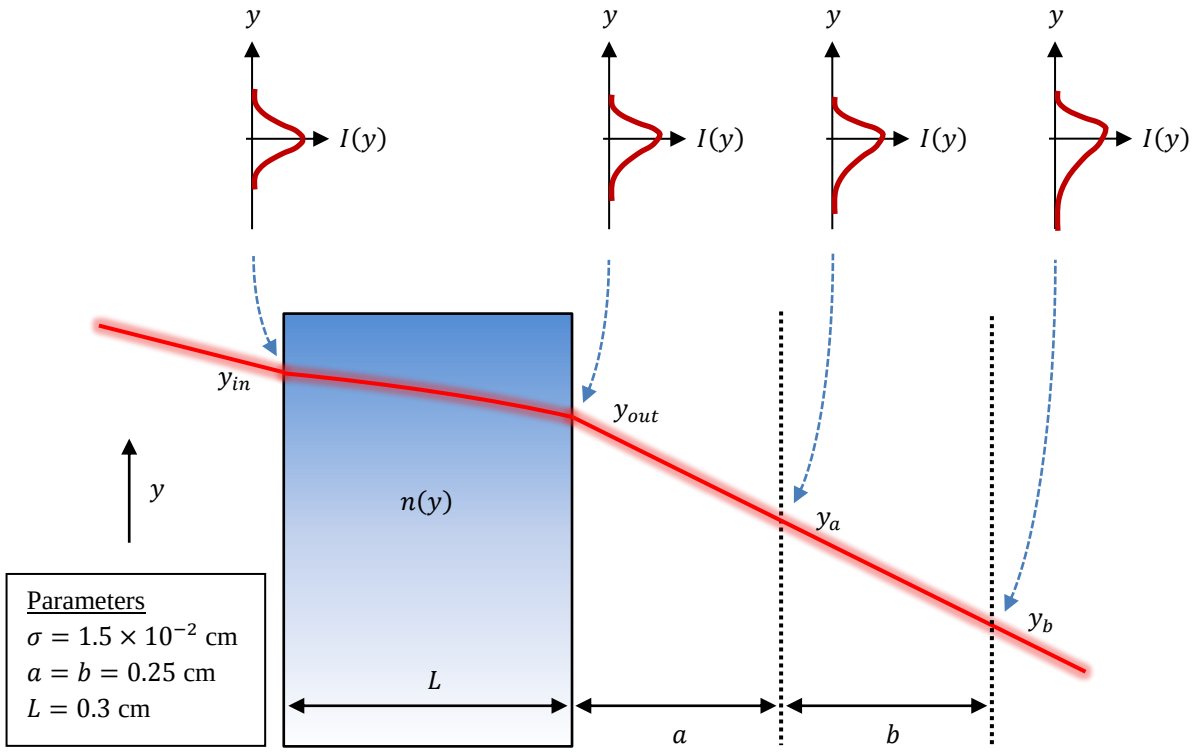


Figure. 2.5.7. Simulation geometry used to model experimental measurement of a hypothetical GRIN element for a single interrogating ray.

It stands to reason that the slight increase in error compared to Figure 2.5.3b is due to the beam shaping effects discussed in Section 2.4. In particular, we expect to encounter errors in the measured positions of finite-width optical beams if any prohibited ray crossings occur between the input and measurement planes. These errors are exacerbated by longer propagation distances a and b, which can be reasoned from the extreme limit $a + b \to \infty$. In this limit, any constituent ray approaching the median after exiting the GRIN medium will result in a prohibited crossing. The cause of the apparent oscillatory behavior in Figures 2.5.3b & 2.5.8b is examined in the ensuing error analysis.

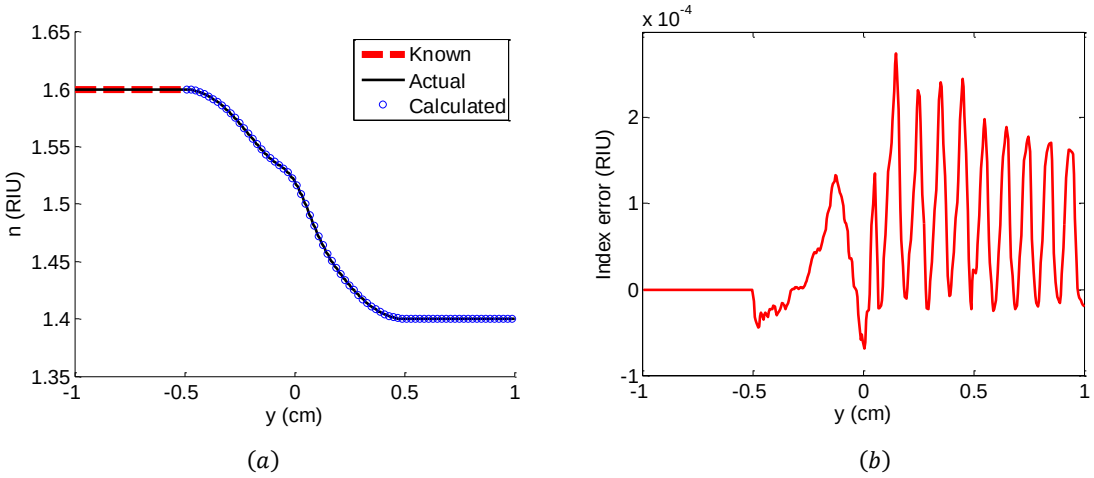


Figure 2.5.8. (a) Bootstrapped reconstruction of the index profile from external boundary values of ray slope and position after subjecting measurements to beam shaping effects and (b) its error relative to the test index profile from Figure 2.5.1, with an absolute maximum value of 2.74×10^{-4} *RIU.*

2.6 Error Analysis

We showed in the previous section that geometrical assumptions in the ray displacement method produced reconstruction errors that were roughly an order of magnitude larger than those produced by the ray deflection method for the same hypothetical GRIN element. In this section, we will analyze in detail the accumulation of error in the bootstrap algorithm for the ray deflection method and test the robustness of University of Rochester's corrective procedure for the ray displacement method for noisy data.

Because the calculated index values are used in subsequent computations in the bootstrap process, propagation of computational error is of particular concern when the deflectometry data is contaminated by noise. The following analysis leads to an expression that can be used to estimate error accumulation based on noise level in the data. The analysis is straightforward because all the refractive index calculations in the ray deflection method are based on an analytical expression. Starting with the fundamental relation from Equation 2.2.15 and using the parameters depicted in Figure 2.6.1, the expression can be rewritten as

$$n_{in} = \left[{n_{out}}^2 - \frac{{n_{amb}}^2 (y'_{out})^2}{1+(y'_{out})^2} + \frac{{n_{amb}}^2 (y'_{in})^2}{1+(y'_{in})^2} \right]^{1/2}, \qquad (2.6.1)$$

where $n_{in} = n(y_{in})$ and $n_{out} = n(y_{out})$. It is assumed that the index is initially known at the bottom region of the GRIN element, as indicated by the figure, and that n_{out} is the known index value in bootstrap calculations since interrogating rays generally propagate toward the known region for the depicted incident angle. The incident ray slope y'_{in} is constant for all interrogating rays while the output ray slope is calculated from ray positions measured in two parallel planes,

$$y'_{out} = \frac{1}{b}(y_b - y_a)\ . \tag{2.6.2}$$

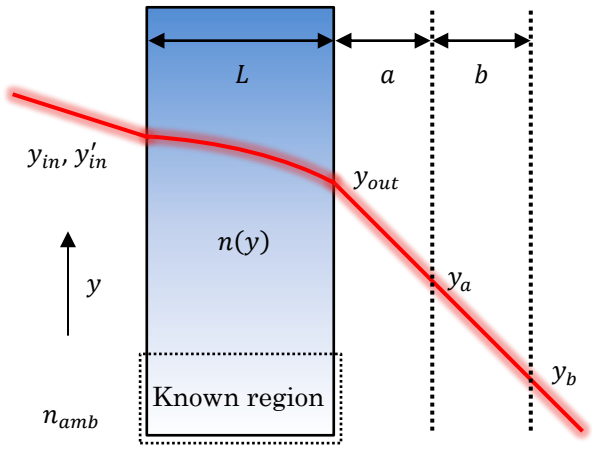


Figure 2.6.1. Measurement geometry used to interrogate a thick GRIN element.

Using Equation 2.6.1 for the depicted geometry in our bootstrap calculations, computational error can result from uncertainty in ray positions in each of the measurement planes as well as uncertainty in the positions of the measurement planes themselves. To make the analysis tractable, let us decompose the overall error in Equation 2.6.1 into three separate contributions,

$$dn_{in} = \frac{\partial n_{in}}{\partial n_{out}} dn_{out} + \frac{\partial n_{in}}{\partial (y'_{out})} d(y'_{out}) + \frac{\partial n_{in}}{\partial (y'_{in})} d(y'_{in})\ . \tag{2.6.3}$$

Differentiating Equation 2.6.1 with respect to n_{out}, y'_{in} and y'_{out} yields the partial derivatives

$$\frac{\partial n_{in}}{\partial n_{out}} = n_{out}\left[{n_{out}}^2 - \frac{{n_{amb}}^2 (y'_{out})^2}{1+(y'_{out})^2} + \frac{{n_{amb}}^2 (y'_{in})^2}{1+(y'_{in})^2}\right]^{-1/2}, \tag{2.6.4a}$$

$$\frac{\partial n_{in}}{\partial (y'_{out})} = \left[-\frac{{n_{amb}}^2 (y'_{out})}{1+2(y'_{out})^2+(y'_{out})^4}\right]\left[{n_{out}}^2 - \frac{{n_{amb}}^2 (y'_{out})^2}{1+(y'_{out})^2} + \frac{{n_{amb}}^2 (y'_{in})^2}{1+(y'_{in})^2}\right]^{-1/2}, \tag{2.6.4b}$$

$$\frac{\partial n_{in}}{\partial (y_{in}{}')} = \left[\frac{{n_{amb}}^2 (y'_{in})}{1+2(y'_{in})^2+(y'_{in})^4}\right]\left[{n_{out}}^2 - \frac{{n_{amb}}^2 (y'_{out})^2}{1+(y'_{out})^2} + \frac{{n_{amb}}^2 (y'_{in})^2}{1+(y'_{in})^2}\right]^{-1/2}, \tag{2.6.4c}$$

which will be utilized later in our accumulated error expression. Since the error in computation is a direct consequence of the uncertainty in the measured quantities a, b, y_a and y_b, it is more convenient to express dn_{out} and $d(y'_{out})$ in Equation 2.6.3 in terms of the listed quantities. Starting with $d(y'_{out})$, we have

$$d(y'_{out}) = \frac{\partial y'_{out}}{\partial y_a} dy_a + \frac{\partial y'_{out}}{\partial y_b} dy_b + \frac{\partial y'_{out}}{\partial b} db$$

$$= \frac{1}{b}(dy_b - dy_a) - \frac{y_b - y_a}{b^2}db = \frac{1}{b}(dy_b - dy_a - y_{out}'db)\ , \tag{2.6.5}$$

where we have used Equation 2.6.2. The error in y'_{out} affects the calculated output ray position

$$y_{out} = y_a - \frac{a}{b}(y_b - y_a) = \left(1 + \frac{a}{b}\right)y_a - \frac{a}{b}y_b\ , \tag{2.6.6}$$

which incurs the error

$$dy_{out} = \frac{\partial y_{out}}{\partial y_a}dy_a + \frac{\partial y_{out}}{\partial y_b}dy_b + \frac{\partial y_{out}}{\partial a}da + \frac{\partial y_{out}}{\partial b}db$$

$$= \left(1 + \frac{a}{b}\right)dy_a - \frac{a}{b}dy_b - y'_{out}da + \frac{a}{b}y'_{out}db\ . \tag{2.6.7}$$

Consequently, the resulting error in the index calculated at $y = y_{out}$ due to the discrepancy in the value of y_{out} is related to the index gradient by

$$dn_{out} = \left.\frac{\partial n}{\partial y}\right|_{y_{out}} dy_{out}\ . \tag{2.6.8}$$

We note that this expression only accounts for the error contribution due to a shift in the calculated output ray position and is hence only appropriate for the first bootstrap where values of n_{out} are taken from the initially known region with presumably no uncertainty. In contrast, refractive index values outside the initially known region used for the computation of Equation 2.6.1 are interpolated from values calculated in previous bootstraps, whose associated computational errors propagate through subsequent bootstraps. Thus, it is more accurate to write the error in n_{out} in the general case as

$$dn_{out} = \left.\frac{\partial n}{\partial y}\right|_{y_{out}} dy_{out} + dn_{interp} + dn^*_{in}\ , \tag{2.6.9}$$

where dn_{interp} is the error associated with interpolating n_{out} from previous computations of Equation 2.6.1 and dn^*_{in} is the calculated index error in these sample values (which constitute spline knots in the interpolation). Substituting Equations 2.6.5, 2.6.7, & 2.6.9 into the error expression in Equation 2.6.2 and rearranging, we find that

$$dn_{in} = \left[-\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}\left(1 + \frac{a}{b}\right) - \frac{\partial n_{in}}{\partial y'_{out}}\frac{1}{b}\right]dy_a + \left[\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}\frac{a}{b} + \frac{\partial n_{in}}{\partial y'_{out}}\frac{1}{b}\right]dy_b +$$

$$\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}y'_{out}da + \left[-\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}\frac{a}{b}y'_{out} - \frac{\partial n_{in}}{\partial (y'_{out})}\frac{1}{b}y'_{out}\right]db +$$

$$+ \frac{\partial n_{in}}{\partial (y'_{in})}dy'_{in} + \frac{\partial n_{in}}{\partial n_{out}}\left(dn_{interp} + dn^*_{in}\right)\ , \tag{2.6.10}$$

where $\frac{\partial n}{\partial y}$ is evaluated at y_{out} and the partial derivatives of n_{in} are given in Equation 2.6.4. Equation 2.6.10 captures all error contributions due to the uncertainty in the measured quantities depicted in Figure 2.6.1. As is evident from the expression, the accumulated error dn_{in} for the current bootstrap is a function of the accumulated index error from the previous bootstrap, dn_{in}^*. Thus, if N bootstraps are needed to compute the entire index profile, we need to apply Equation 2.6.8 N times to determine the accumulated error as a function of y. The accumulated error compounds with each successive bootstrap, starting with $dn_{in}^* = 0$ in the initial calculation. The last term in Equation 2.6.10 implies that the partial derivative $\frac{\partial n_{in}}{\partial n_{out}}$ dictates how fast this error compounds. In cases where $\left|\frac{\partial n_{in}}{\partial n_{out}}\right| = 1$ (*e.g.*, in a homogeneous medium where y'_{in} and y'_{out} are always equal), we expect the accumulated error to increase linearly with each bootstrap. In cases where $|y'_{in}| > |y'_{out}|$, we have $\left|\frac{\partial n_{in}}{\partial n_{out}}\right| < 1$ and the accumulated error will exhibit asymptotic behavior. On the other hand, if $|y'_{in}| < |y'_{out}|$ such that $\left|\frac{\partial n_{in}}{\partial n_{out}}\right| > 1$, the error will grow exponentially with each bootstrap. This behavior enables us to use prior knowledge about the index to our advantage and mitigate error build-up during the bootstrap algorithm in measuring monotonic index profiles, for instance.

In order to verify the error expression in Equation 2.6.10, we introduced a systematic error to the deflection data used to reconstruct the test index profile shown in Figure 2.5.1. The simulation parameters, depicted and labeled in Figure 2.6.1, are $y'_{in} = -\frac{\pi}{7}$ radians (relative to horizontal axis), $L = 0.3$ cm, $a = b = 0.25$ cm. Using unaltered ray trace data, the bootstrap algorithm will yield the reconstruction shown in Figure 2.5.3 since the amount of interrogating rays and their initial conditions have not changed. Once again, we employed constrained cubic splines [Kruger, 2002] in interpolating refractive index values within the known region of the index profile during the bootstrap process. The noise from the interpolator was characterized to be roughly fourth-order in the separation between output ray positions ($\Delta \approx 0.02$ cm) and was assumed to be negligible in the simulation. Following the ray trace, all traced values for y_a were shifted by -1 μm and values for y_b are shifted by $+1$ μm for each interrogating ray in order to simulate a systematic error. The error accumulation due to this systematic shift is plotted in Figure 2.6.2*b* and is in good agreement with the expression in Equation 2.6.10, where we have assumed knowledge of the index gradient $\frac{\partial n}{\partial y}$. For illustration purposes, we plot the predicted error accumulation as a stepwise error bound using the worst case scenario for each bootstrap in the computation of Equation 2.6.10. Naturally, a growing discrepancy between the accumulated error and the error bound develops as we increase in y-value. This is because the error in Equation 2.6.10 depends on deflectometry data specific to each interrogating ray, which manifests as a latency effect in the error plot relative to the worst case scenario. From the figure, it is apparent that the accumulated error only reaches the predicted error bound for the first couple bootstraps in the simulation. We hasten to add that a similar systematic perturbation in the recovered index profile can result in practice from the uncertainty in propagation distances a and b shown in Figure 2.6.1.

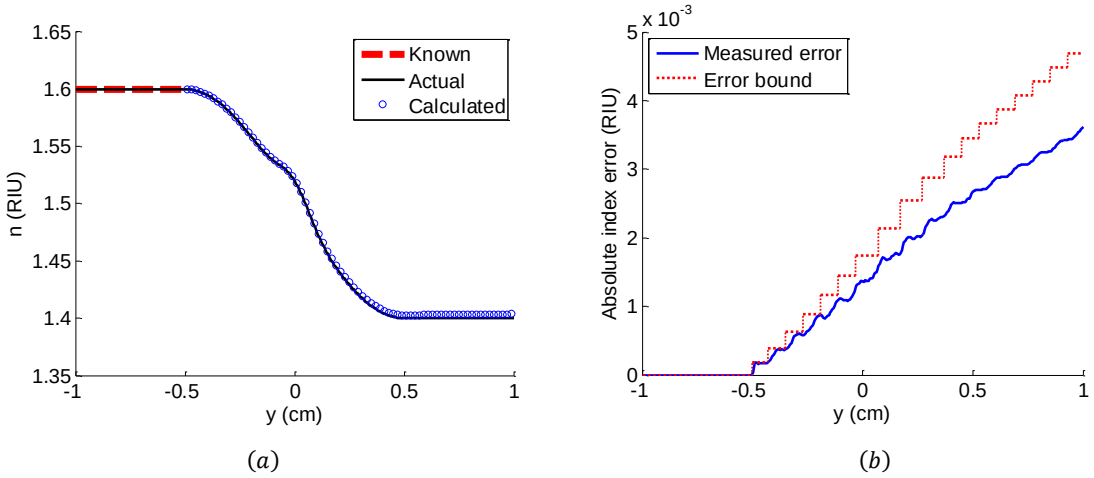


Figure 2.6.2. (a) Calculated index profile after introducing systematic errors to ray position measurements and (b) accumulated index error in the bootstrap process relative to the test index profile.

Aside from systematic errors in the deflection data, we also need to account for the random aspect of uncertainty in position measurements. We incorporate this type of uncertainty in a separate simulation by introducing normally distributed white noise to y_a and y_b, with a standard deviation of 1 μm. In order to compute the expected error accumulation resulting from such a noise source, Equation 2.6.10 is modified so that the independent error contributions are summed in the RMS sense, *i.e.*,

$$\Delta n_{in,rms}{}^2 = \left|-\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}\left(1+\frac{a}{b}\right)-\frac{\partial n_{in}}{\partial y'_{out}}\frac{1}{b}\right|^2 \Delta y_{a,rms}{}^2 +$$

$$\left|\frac{\partial n_{in}}{\partial n_{out}}\frac{\partial n}{\partial y}\frac{a}{b}+\frac{\partial n_{in}}{\partial y'_{out}}\frac{1}{b}\right|^2 \Delta y_{b,rms}{}^2 + \left|\frac{\partial n_{in}}{\partial y'_{in}}\right|^2 \Delta y'_{in,rms}{}^2 +$$

$$\left|\frac{\partial n_{in}}{\partial n_{out}}\right|^2 \left(\Delta n_{interp,rms}{}^2 + \Delta n^*_{in,rms}{}^2\right). \qquad (2.6.11)$$

We note that the error contributions due to the uncertainty in quantities a and b are systematic in nature and have therefore been omitted from the RMS expression. Figure 2.6.3a shows the accumulated error in the reconstructed index profile resulting from one instance of applying normally distributed noise to y_a and y_b. The ensemble average over 10^3 trials of the simulation is shown in Figure 2.6.3b, with good agreement to the expected error in Equation 2.6.11. Note that the variance in the ensemble average also compounds as we increase in y in the latter figure, another direct consequence of the bootstrap process.

It is clear from both Equations 2.6.10 & 2.6.11 that minimizing a and maximizing b makes the bootstrap algorithm more resilient to noise in the deflectometry data. However, one must keep in mind that increasing the overall distance $a + b$ exacerbates the uncertainty of position measurements for finite-width interrogating beams due to beam shaping effects (see Section 2.4). This is because a larger propagation

distance after the optical beam exits the GRIN medium increases the likelihood of prohibited crossings amongst its constituent rays; such an occurrence would cause its measured position to become ambiguous.

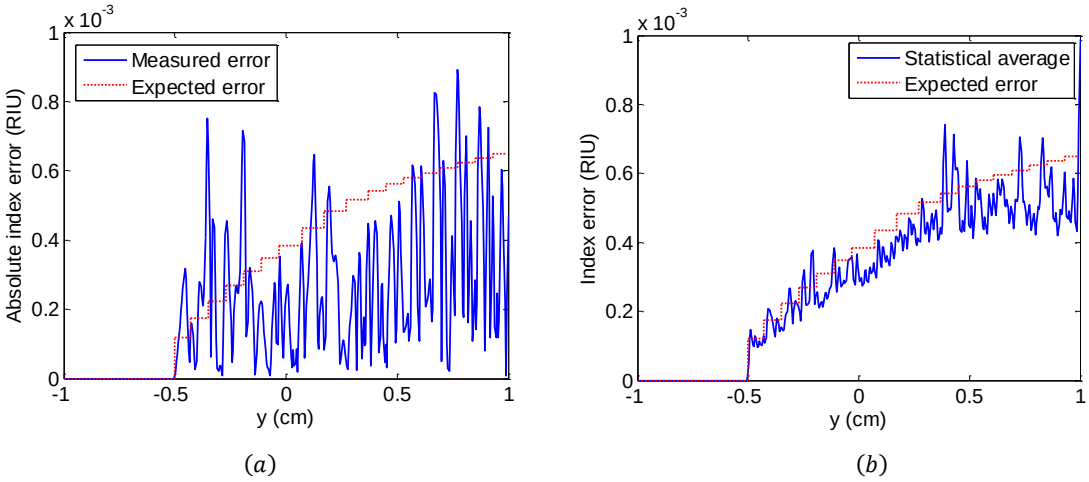


Figure 2.6.3. Accumulated index error due to random errors in ray position measurements taken from (a) a single trial and (b) the ensemble average over a large number of trial simulations.

We performed a similar test for analyzing the error sensitivity of University of Rochester's corrective ray trace procedure [McCarthy, 2015] used to improvement reconstruction accuracy for the ray displacement method detailed in Section 2.3. In one instance, displacement values used for the reconstruction shown in Figure 2.5.6 were uniformly shifted by $+1$ μm in the measurement plane to simulate a systematic error in ray position measurements. In a separate simulation, the displacement values were contaminated by normally distributed white noise with a standard deviation of 1 μm, and the error was averaged over 10^3 trials of the simulation. The resulting errors in the optimized index profile for both cases are plotted in Figure 2.6.4. Compared to the ray deflection method, the reconstruction error is significantly smaller in the ray displacement method employing a corrective procedure for similar noise levels in position measurements. This is partially because computing ray slopes from two random variables amplifies the measurement noise in the former method. In addition, a significant portion of the error reduction can be attributed to the iterative ray trace algorithm used to optimize the index profile (we showed in Section 2.5 that the reconstruction error from the ray displacement method without the corrective procedure is an order of magnitude larger than that of the ray deflection method). We expect the error accumulation in the ray displacement method to be linear in y since the index profile is obtained by integrating the index gradient calculated from individual ray displacement measurements. Furthermore, we expect the variance in index error to remain relatively constant across the index profile in this method, provided that there is no overlap between the y-values traversed by different interrogating rays so that the computed errors in the index gradient at different locations are completely independent of each other. Both of these features are evident in Figure 2.6.4.

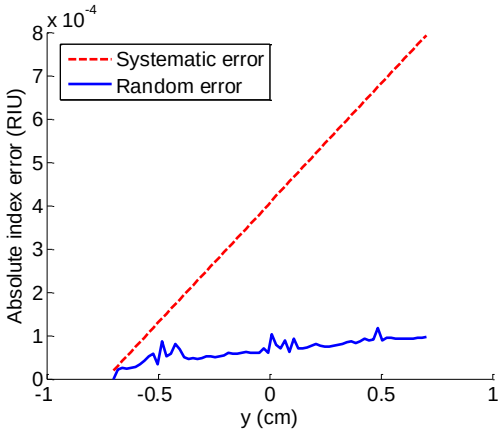


Figure 2.6.4. Reconstruction error from the University of Rochester's iterative ray trace algorithm due to both systematic and random errors introduced in measured ray displacement values.

2.7 Experimental Measurement

We will conclude Chapter 2 with the experimental demonstration of our 1-D deflectometry technique. In this section, we employ both measurement techniques developed in Sections 2.2 & 2.3 in measuring a real GRIN sample provided by the Institute of Optics at the University of Rochester. The GRIN sample is fabricated using a polymer blend of polymethyl methacrylate and polystyrene, with a measured thickness of 1.403 cm. Similar to the interrogation geometry depicted in Figure 2.6.1, the benchtop setup for measuring the GRIN sample using the ray deflection method is depicted in Figure 2.7.1, where beam profilers are used to capture the optical beam's irradiance profiles in observation planes "A" and "B". In accordance with Section 2.4, the medians in the measured irradiance profiles were used to designate the positions y_a and y_b in their respective planes for an ideal ray used to represent the optical beam. In the experiment, the GRIN sample was translated along the y-axis to provide a systematic interrogation of its index profile so that the angle of incidence remains constant for all interrogating rays. Ray positions and slopes at the output surface of the sample were extrapolated from the measured values of y_a and y_b for the corresponding ray. We chose an incident angle of 0.4363 radians (25°) relative to the horizontal axis and interrogated the GRIN sample at 66 discrete positions along y to provide boundary value measurements for 66 individual interrogating rays. Propagation distances to the measurement planes in the setup were set to $a = 0.473$ cm and $b = 1.791$ cm. If we assume the index is initially known at the bottom of the GRIN sample, then the positive angle of incidence used in the interrogation implies that y_{in} resides in the known region of the index profile and that $n(y_{out})$ is the unknown quantity in Equation 2.6.1. Hence, Equation 2.6.1 should be rewritten as

$$n_{out} = \left[{n_{in}}^2 + \frac{{n_{amb}}^2 (y'_{out})^2}{1+(y'_{out})^2} - \frac{{n_{amb}}^2 (y'_{in})^2}{1+(y'_{in})^2} \right]^{1/2} . \qquad (2.7.1)$$

According to the University of Rochester, the bulk index value of the optical material at the bottom end of the GRIN sample is equal to 1.566 RIU.

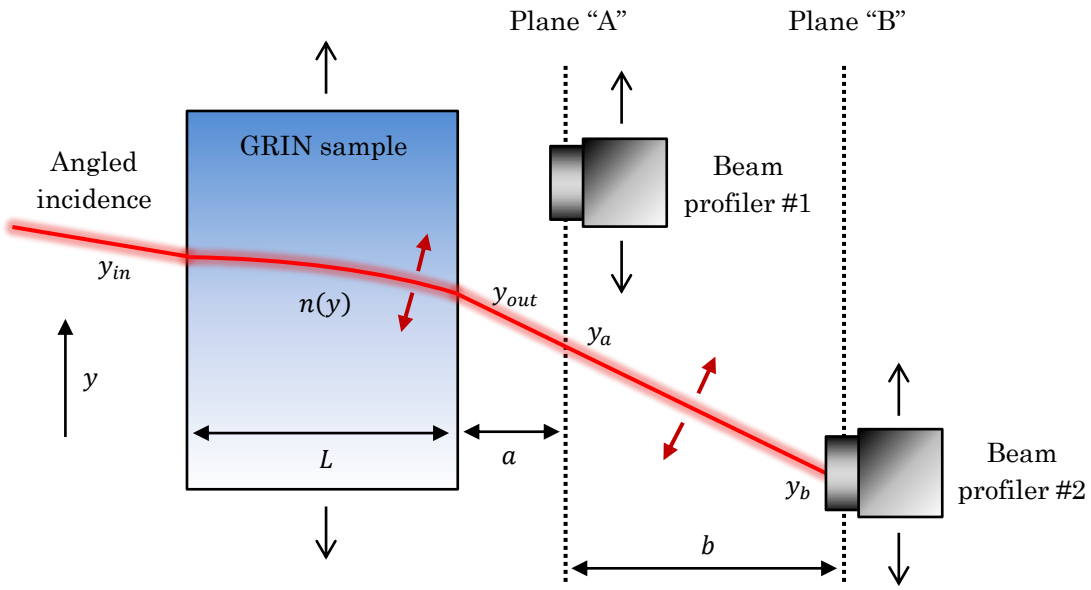


Figure 2.7.1. Experimental setup for measuring a thick GRIN sample using angled incidence.

Upon acquiring the necessary boundary values for ray position and slope for each of the 66 interrogating rays, the boundary data was fed into the bootstrap algorithm to compute the refractive index profile of the GRIN sample. A separate measurement was performed by the University of Rochester using the ray displacement method in conjunction with their corrective procedure based on iterative ray tracing [McCarthy, 2015]. Figure 2.7.2*a* shows the calculated index profiles obtained from both measurement techniques. Although the reconstructed profiles are in good agreement, we cannot say with absolute certainty which of the two methods is more accurate since the actual refractive index profile of the sample is unknown outside of these measurements. However, due to the built-in error correction capabilities in their iterative ray trace algorithm, it is reasonable to assume that the measurement provided by the University of Rochester is closer to the ground truth; the built-in capabilities not only correct the errors resulting from the geometrical assumptions of the model but also dramatically reduce the recovery algorithm's sensitivity to measurement noise. The discrepancy between the calculated index profiles from both methods is plotted in Figure 2.7.2*b*, with an absolute maximum of 1.5×10^{-3} RIU. This is a reasonable value considering the uncertainty in position measurements (~0.5 μm) and the short propagation distance between the measurement planes.

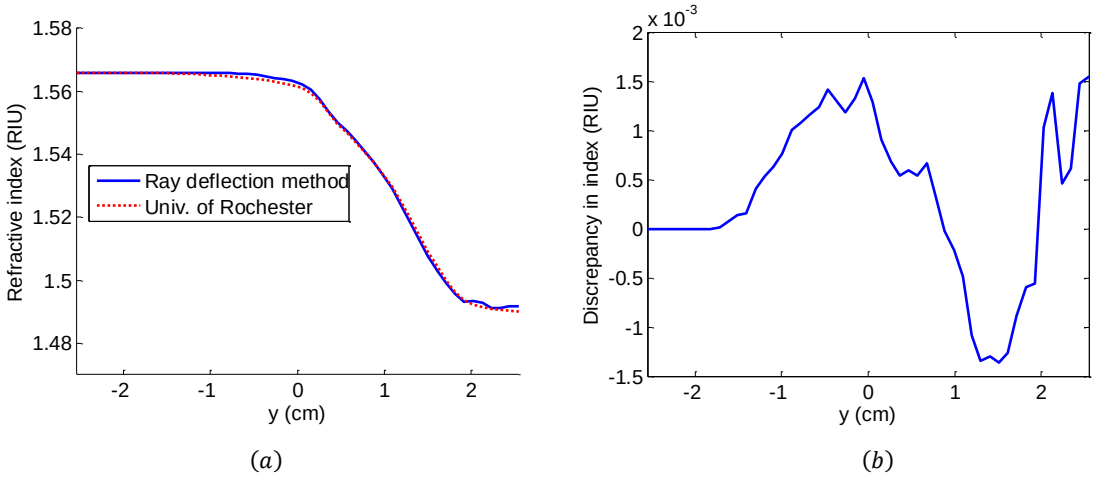


Figure 2.7.2. Comparison between index profiles measured in a real GRIN sample using the ray deflection method and the ray displacement method.

Figure 2.7.3 shows the distortion in the irradiance profile of a sample incident Gaussian beam used for interrogation after the beam passed through the GRIN sample and propagated to plane "B." In addition to the beam shaping effects induced by the index profile of the sample, the measured beam irradiance profile in Figure 2.7.3*b* includes the distortion effects resulting from roughness, tilt, and opacity of the front and rear boundary surfaces of the GRIN element. These effects also contribute to the measurement noise in determining ray positions at planes "A" and "B".

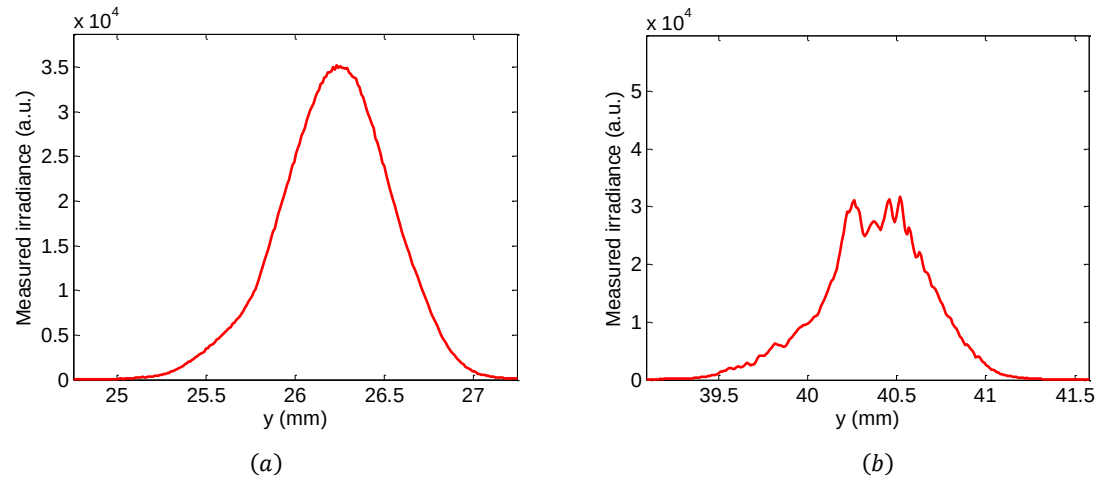


Figure 2.7.3. Irradiance profiles of (a) a sample incident Gaussian beam and (b) the same optical beam after passing through the GRIN sample and propagating to the far measurement plane.

In the next chapter, we will examine the measurement of GRIN elements with 2-D index distributions using nondestructive methods. A summary of the measurement principles developed in this chapter (specific to 1-D index distributions) is provided in Chapter 5, as well as suggestions for possible improvements in our measurement technique.

3. Deflectometry for Two-dimensional Index Fields

3.1 Background Material

At the time of writing this thesis, there has been no reported method for measuring arbitrary two-dimensional (2-D) index distributions in the general case. The methods described in the literature typically assume linear trajectories in order to simplify the mathematical model or are only applicable to specific geometries. For instance, analytical solutions for deflectometry measurements have been shown for radially symmetric index distributions of optical fiber preforms [Chu, 1977]. In this particular measurement system, a set of laser beams are shone perpendicularly to the preform axis; the beam angles after exiting the preform as a function of the distance from the horizontal axis passing through the center of the preform can be integrated to give the radial index profile. For asymmetric weakly refracting index fields in higher dimensions, tomographic measurement techniques have been described, where approximate linear trajectories for the interrogating beams are used to simplify the mathematical model. Of course, we know that rays do not travel in straight lines in an inhomogeneous medium from Equation 2.2.1. However, such approximations allow the projection slice theorem to be used in the Fourier synthesis of the index field from optical path length measurements, which can be obtained using multidirectional holographic interferometry [Sweeney & Vest, 1973]. An attempt has been made to address curved trajectories in the reconstruction of strongly refracting index fields using this method, but the technique involves the use of an imaging system during the formation of the interferogram and is not generalizable to other measurement techniques [Cha & Vest, 1981]. Using the same tomographic principles, measurement of 2-D index fields by deflectometry have also been demonstrated using x-rays [Yuasa, *et al.*, 2006; Gasilov, *et al.*, 2014]. In the latter paper, the deflection of an optical beam over a linear trajectory is expressed as a complex-valued radon transform, and the deflection of multiple beams is inverted using a filtered back-projection algorithm to reconstruct the index gradient. The linear trajectory approximation is justified in this case because the index of materials is close to unity within the x-ray frequency band. Furthermore, Snell's law at the boundary interface between the ambient medium (air) and the GRIN medium can be neglected due to the negligible discontinuity in the refractive index. Due to these assumptions, however, the methods cannot be applied to longer wavelengths because significant refraction occurring at the boundary interface and curved beam trajectories inside the GRIN medium introduce complications into the mathematical model, making the measured deflectometry data difficult to interpret.

In this chapter, we will develop a method that does not restrict the wavelength of measurement; the method is applicable to the general case where probe laser beams can encounter modest index gradients in the optical material under interrogation. In such cases, the medium cannot be assumed to be weakly refracting and beam trajectories are seldom linear. Furthermore, our method incorporates Snell's law of refraction along the boundary interface into the mathematical model. We begin by discussing the

simplifying assumptions in our model that enable us to formulate the inverse problem as a linear system in Section 3.2, where we use optical path length measurements to explain our method in a simpler context. An extension of the model to deflectometry data is detailed in Section 3.3, where we establish the primary deflectometry system equation to be inverted for recovering the index field. A detailed discussion of the numerical techniques for achieving accurate discrete representations of the physical system is provided in Section 3.4. In Section 3.5, we demonstrate a numerical implementation of the proposed method, where we reconstruct a hypothetical index field from internal boundary measurements using simulated deflectometry data and invert the primary system equation using numerical methods. We then use the preliminary results to justify the simplifying assumptions made in Section 3.2 and demonstrate a corrective procedure in Section 3.6 to address the reconstruction error associated with these approximations. In Section 3.7, we identify the primary sources of error in the reconstruction process and address various numerical aspects of system inversion. Following the error analysis, we develop a procedure for reconstructing the index field directly from external boundary value measurements in Section 3.8 by incorporating Snell's law into the overall inversion process. Finally, we make an extension to three-dimensional (3-D) index distributions in Section 3.9 using the same measurement principles.

3.2 Linear System Formulation

We begin by developing the formalism for describing path integrals in a scalar field as linear algebraic equations of the sampled values in the field quantity. The same principles are fully extendable to the Cartesian components of vector field quantities, as we will soon see in Section 3.3. A wide selection of powerful tools for solving linear systems provides the primary incentive for taking a linear system approach to the inverse problem. In the latter half of this section, we will demonstrate a practical application of our method in using OPL measurements to determine an unknown index field. In the next section, we will show how these concepts can be applied to deflectometry measurements for 2-D index distributions.

Let us suppose that the scalar field quantity in question, $f(x,y)$, is represented on a uniform rectangular grid of discrete sample points, as shown in Figure 3.2.1, where the discrete indices l and k correspond to the spatial variables x and y, respectively. In the discrete representation, the sample values $f_{l,k}$ are treated as the system's unknown quantities, which we will use synonymously with its degrees of freedom (DoFs). Of course, a sufficient number of sample points are required in the discrete representation $f(x,y)$ in order to avoid loss of information. Assuming that $f(x,y)$ is band-limited, the minimum number of sample points is associated with its space-bandwidth product [Lohmann, *et al.*, 1996].

For the sake of simplicity, let us assume that $f(x,y)$ is constant within each discrete rectangular element depicted in the figure and that its value is equal to the value of the sample point $f_{l,k}$ located at its center. Next, suppose that we wish to express a path integral of $f(x,y)$ in terms of the sample values $f_{l,k}$ for a given trajectory inside the field quantity. In the analog domain, such a path integral can be expressed as

$$\varphi = \int_a^b ds \cdot f(x,y)\,, \tag{3.2.1}$$

where a and b are the end-points for the path of interest and ds is the differential arc length.

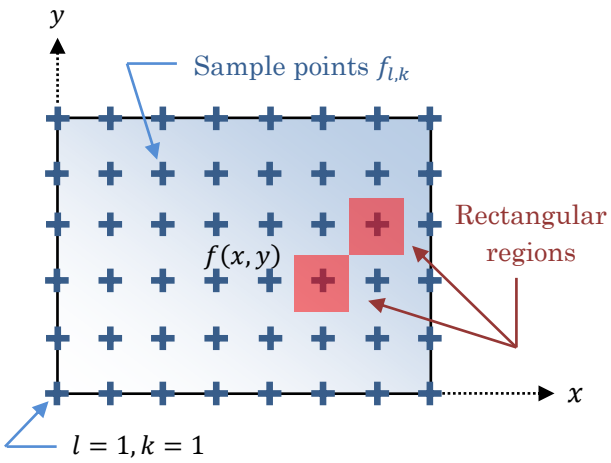


Figure 3.2.1. Discretization representation of a continuous field quantity on a uniform rectangular grid.

Switching to a discrete representation under our crude approximation, the path integral can be expressed as the Riemann sum,

$$\varphi = \sum_l \sum_k ds_{l,k} f_{l,k}\,, \tag{3.2.2}$$

where $ds_{l,k}$ is the arc length of the path segment contained within the rectangular element corresponding to $f_{l,k}$. If the path of the integral is given, then $ds_{l,k}$ is fully specified and Equation 3.2.2 is a linear expression in $f_{l,k}$. A visualization of the discretization process is shown in Figure 3.2.2. For the depicted path integral, Equation 3.2.1 becomes

$$\varphi_1 = \int_A^D ds \cdot f(x,y) \approx \overline{AB} f_{2,1} + \overline{BC} f_{2,2} + \overline{CD} f_{1,2}\,, \tag{3.2.3}$$

where $\overline{AB}$ is the arc length between A and B, and likewise for $\overline{BC}$ and $\overline{CD}$.

Next, suppose that we have multiple path integrals discretized in the form of Equation 3.2.2, whose integral values are known. It is then straightforward to assemble the path integrals into a linear system of algebraic equations whose unknowns correspond to the values of the discrete elements $f_{l,k}$, *i.e.*,

$$[S] \cdot \vec{f} = \vec{\varphi}\,, \tag{3.2.4}$$

where $[S]$ is the system matrix containing all the discrete arc lengths from multiple path integrals, and each row of the matrix contains lexicographically ordered pieces of the arc length $ds_{l,k}$ from a particular path

integral. The column vector $\vec{\varphi}$ contains the integral values associated with each path integral, and $\vec{f}$ contains all sampled values of $f_{l,k}$ lexicographically ordered into a single column vector.

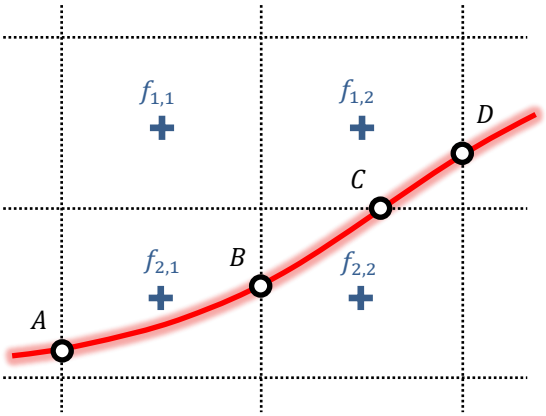


Figure 3.2.2. Visualizing the discrete representation of a path integral in a scalar field quantity.

To demonstrate the construction of such a linear system from N discrete path integrals for a hypothetical scalar field quantity consisting of the sample values shown in Figure 3.2.2, let the first of these path integrals be the path integral depicted in the figure. Equation 3.2.4 then takes the form

$$\begin{bmatrix} 0 & \overline{CD} & \overline{AB} & \overline{BC} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \cdot \begin{bmatrix} f_{1,1} \\ f_{1,2} \\ f_{2,1} \\ f_{2,2} \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \cdot \\ \cdot \\ \varphi_N \end{bmatrix}, \tag{3.2.5}$$

where we have used Equation 3.2.3 as the first equation in the system. Assuming N to be sufficiently large such that $[S]$ is a full rank matrix, we can solve for the system's DoFs, in principle, using the inverse formula $\vec{f} = [S]^{-1} \cdot \vec{\varphi}$. If N is greater than the number of elements in $\vec{f}$, then the system is overdetermined and can be solved in the least squares sense by invoking the pseudoinverse formula $\vec{f} = ([S]^T \cdot [S])^{-1} \cdot [S]^T \cdot \vec{\varphi}$. Thus, the inverse problem of solving for sample values in an unknown scalar field quantity can be reduced to inverting a simple linear system of algebraic equations if the paths of the integrals are known. On the other hand, if the paths of the integrals are not given, a new method for obtaining the coefficients in $[S]$ is needed; such is the case when interrogating unknown index fields using nondestructive methods such as deflectometry, where only quantities along the boundaries of (and external to) the index field are available for direct measurement. Provided that the trajectories of the path integrals are well behaved, a crude approximation based on measured boundary values will generally suffice.

In order to provide a physical context for our model and develop further insight on our linear system formulation, let us suppose that the trajectories of path integrals follow rays of light propagating inside a GRIN medium with an unknown index distribution $n(x, y)$ and that the integrals report the total

accumulated OPL inside the medium for each light ray. Figure 3.2.3 illustrates the geometry of such a measurement.

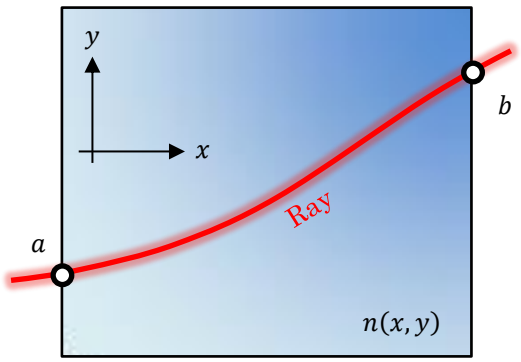


Figure 3.2.3. Ray of light propagating through a GRIN medium with an unknown refractive index distribution.

Under this scenario, the total OPL from point a to point b is given by the integral in Equation 3.2.1, where scalar field $f(x,y)$ is replaced by $n(x,y)$. The OPL measurement we have just described yields no concrete information regarding the ray's trajectory between a and b; we only know that the trajectory obeys the ray equation, *i.e.*, Equation 2.2.1, inside the GRIN medium. Given $n(x,y)$, the ray equation can be used to compute the ray trajectories; this is known as the forward problem of ray tracing, which we discuss in detail in Appendix 1. However, since $n(x,y)$ is the field quantity we seek, we cannot use it to ascertain the trajectories of our path integrals. Under these circumstances, it is clear that the path of the integral is a function of the unknown field quantity and Equation 3.2.1 is not a linear expression in $f(x,y)$. In spite of this fact, we would still prefer to formulate the inverse problem as a system of linear algebraic equations due to the tools and techniques we have at our disposal for solving such systems. This can be achieved by making a rather daring assumption, albeit an unphysical one: Let us assume that the ray trajectory inside the GRIN medium is independent of its refractive index distribution and does not need to obey the ray equation. At this point in time, the reader may consider this to be an over-simplification of the inverse problem, but we will show in Section 3.6 that trajectories consistent with the ray equation can be obtained using this assumption as a starting point. With this in mind, let us assume that the boundary measurements of ray positions y and its slope $\frac{dy}{dx}$ are readily available at the end points a and b in addition to the total OPL traveled by the light ray. Using this information, we can construct our own ray trajectories to use as tentative approximations for the time being. Doing so enables us to treat the path of the integral as a known quantity so that a linear system in the form of Equation 3.2.4 can be constructed from the collection of OPL measurements.

To construct an approximate ray trajectory inside the GRIN medium using measured boundary values of ray position and slope at the end points, we assume that the ray follows a trajectory of the form $y = y(x)$. A cubic polynomial of the form

$$y(x) = Ax^3 + Bx^2 + Cx + D \tag{3.2.7}$$

will generally suffice, where A, B, C and D are the polynomial coefficients. The measured boundary positions and slopes for the ray trajectory at a and b yield four algebraic equations that can be arranged in matrix form as

$$\begin{bmatrix} {x_a}^3 & {x_a}^2 & x_a & 1 \\ 3{x_a}^2 & 2x_a & 1 & 0 \\ {x_b}^3 & {x_b}^2 & x_b & 1 \\ 3{x_b}^2 & 2x_b & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} y_a \\ y_a' \\ y_b \\ y_b' \end{bmatrix}, \tag{3.2.8}$$

where the prime denotes differentiation with respect to x. Upon inverting the matrix and solving for the polynomial coefficients, we obtain approximate trajectories for the path integrals and proceed with the construction of our linear system from OPL measurements. The resulting system can then be inverted, in principle, to solve for the unknown index field $n(x, y)$.

3.3 Extension to Deflectometry

We have just developed the mathematical framework for a linear system formulation of the inverse problem and explained the mathematical model in the context of OPL measurements. While it is possible to interrogate an unknown index field using a series of OPL measurements, the boundary measurements of ray position and slope used to determine ray trajectories are sufficient for recovering the gradient of the index field. Furthermore, measuring the absolute OPL traveled inside the index field requires interferometry and is generally more difficult to implement. In this section, we will extend the measurement principles developed in our linear system approach to deflectometry.

We begin by deriving the equations that describe the deflection of a light ray propagating inside a GRIN medium with a refractive index distribution $n(x, y)$. Starting with Equation 2.2.6, we rewrite the expression as

$$\frac{\partial w}{\partial y} - \frac{\partial w}{\partial x} y' = \frac{y''}{1+(y')^2}, \tag{3.3.1}$$

where $w = \ln(n)$ is the logarithmic refractive index. Equation 3.3.1 can be reduced to a far more fundamental relationship between the ray's deflection and the local index gradient. To illustrate this relationship, let us switch to relative coordinates where the x-axis is defined parallel to the local ray trajectory at a particular point of interest. Consequently, we have $y' = 0$ and Equation 3.3.1 simplifies to

$$\frac{\partial w}{\partial y} = y'' \,, \tag{3.3.2}$$

where y is taken to be the normal direction to the local ray vector in the relative coordinate system. Hence, the directional derivative of the refractive index along the normal of the local ray vector is responsible for the ray's deflection. This is a fundamental result that we will use as the basis for our deflectometry model. Rewriting y'' as $\frac{d}{dx}y'$ in the above expression results in the differential

$$\Delta y' = \frac{\partial w}{\partial y}\Delta x \,. \tag{3.3.3}$$

It is insightful to view the geometry of ray deflection in relative coordinates as a right triangle, as shown in Figure 3.3.1. From this geometry, we deduce that the following trigonometric relation,

$$tan(\Delta\theta) \approx \Delta\theta = \frac{\Delta y}{\Delta x} = \Delta y' \,, \tag{3.3.4}$$

describes the angular deflection $\Delta\theta$ of the light ray from its current direction $y' = 0$, where $\Delta y'$ is given by Equation 3.3.3.

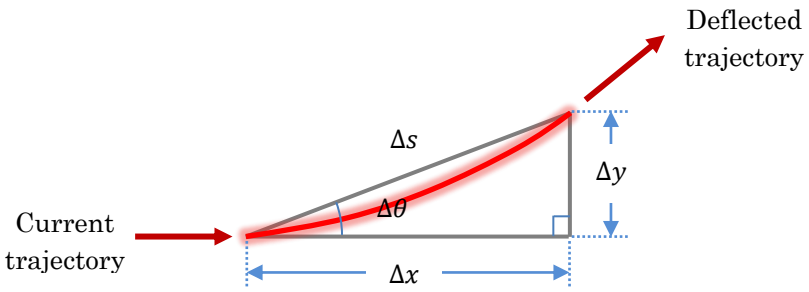


Figure 3.3.1. Angular deflection of a propagating ray from its current trajectory.

Ideally, we desire an expression for $\Delta\theta$ in terms of the arc length Δs because it allows for the straightforward expression of path integrals along the ray's trajectory. This can be achieved by invoking the small angle approximation $cos(\Delta\theta) = \frac{\Delta x}{\Delta s} \approx 1$ such that

$$\Delta\theta \approx \Delta y' = \frac{\partial w}{\partial y}\Delta x \approx \frac{\partial w}{\partial y}\Delta s \,. \tag{3.3.5}$$

The use of small angles in Equations 3.3.4 & 3.3.5 is the only geometrical approximation we will make in establishing the mathematical model for ray propagation in 2-D GRIN media. We note that this assumption does not restrict the types of index fields we can measure accurately using this approach because Δx in Equation 3.3.5 can always be made arbitrarily small to satisfy the approximation. Next, using the fact that y

is the normal direction to the local ray vector in relative coordinates, the differential in Equation 3.3.5 can be rewritten in laboratory coordinates as

$$\Delta\theta = \nabla w \cdot \hat{n} \cdot \Delta s = \left[-\frac{\partial w}{\partial x} sin(\theta) + \frac{\partial w}{\partial y} cos(\theta)\right] \Delta s \,, \tag{3.3.6}$$

where $\hat{n} = -sin(\theta)\hat{\imath} + cos(\theta)\hat{\jmath}$ is a unit vector normal to the local ray trajectory and $\vec{\nabla} w = \frac{\partial w}{\partial x}\hat{\imath} + \frac{\partial w}{\partial y}\hat{\jmath}$ is the gradient of the logarithmic refractive index in the laboratory coordinate system. θ is the angle of the local ray vector relative to the laboratory x-axis, which implies that $tan(\theta) = y'$. Equation 3.3.6 describes the differential angular deflection of a light ray from its current direction due to the presence of a local index gradient. Therefore, the total deflection accumulated by the light ray in transit from point a to point b inside the GRIN medium is given by the integral

$$\Delta\theta_{total} = \int_a^b ds \cdot \left[-\frac{\partial w}{\partial x} sin(\theta) + \frac{\partial w}{\partial y} cos(\theta)\right] . \tag{3.3.7}$$

We now have a deflectometry path integral that will form the basis for constructing a linear system similar to Equation 3.2.4. It is important to note that the integral now depends on two field quantities, $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$. However, because $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$ are the partial derivatives of an underlying potential function, *i.e.*, the logarithmic refractive index $w(x, y)$, they are related to each other and must satisfy the irrotational constraint:

$$\nabla \times \nabla w = \frac{\partial}{\partial y}\frac{\partial w}{\partial x} - \frac{\partial}{\partial x}\frac{\partial w}{\partial y} = 0 \,. \tag{3.3.8}$$

In other words, the mixed partial derivatives for w must be equal in order for w to exist; any gradient field ∇w with non-zero curl is unphysical.

It follows that the logarithmic refractive index $w(x, y)$ is only specified up to an unknown constant by the deflectometry data. This constant can only be identified with additional measurements such as the refractive index at some position inside the GRIN medium. However, both partial derivatives of w are fully specified by the deflectometry data, provided that the resulting system of equations is of full rank or is overdetermined. Thus, it is more convenient to treat $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$ as two individual field quantities and impose the constraint in Equation 3.3.8 as algebraic equations to augment the primary system of equations. Similar to the discretization process presented in the previous section and employing the same crude approximation for the partial derivative fields, the integral in Equation 3.3.7 can be expressed as

$$\Delta\theta_{total} = \sum_l \sum_k ds_{l,k} \left[-\left.\frac{\partial w}{\partial x}\right|_{l,k} sin\left(\theta_{l,k}\right) + \left.\frac{\partial w}{\partial y}\right|_{l,k} cos\left(\theta_{l,k}\right)\right] , \tag{3.3.9}$$

where $ds_{l,k}$ is the arc length of the ray segment, which we approximate as a straight line in this particular expression, contained within the rectangular region corresponding to the sample field quantity at

indices l, k. $\theta_{l,k}$ is ray segment's angle relative to the x-axis. A visualization of the discretization process is shown in Figure 3.3.2.

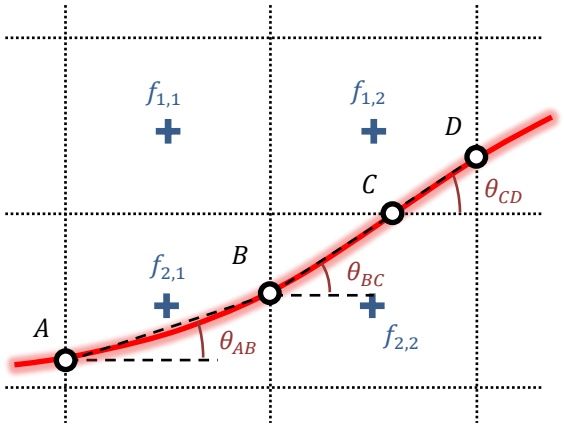


Figure 3.3.2. Visualizing the discrete representation of a deflectometry path integral.

Because two distinct field quantities are involved in evaluating the deflectometry path integral, each grid point $f_{l,k}$ shown in the figure has an additional DoF associated with it. To avoid confusion, let $f_{l,k}$ denote the discrete field quantity $\frac{\partial w}{\partial x}\Big|_{l,k}$ and let $g_{l,k}$ denote $\frac{\partial w}{\partial y}\Big|_{l,k}$. Using this notation and writing out each term in the summation in Equation 3.3.9, the deflectometry path integral depicted in the figure can be written as

$$\Delta\theta_{total} = \overline{AB}\left[-f_{2,1}sin(\theta_{AB}) + g_{2,1}cos(\theta_{AB})\right] + \overline{BC}\left[-f_{2,2}sin(\theta_{BC}) + g_{2,2}cos(\theta_{BC})\right]$$

$$+\overline{CD}\left[-f_{1,2}sin(\theta_{CD}) + g_{1,2}cos(\theta_{CD})\right], \qquad (3.3.10)$$

where $\overline{AB}$ is the arc length between A and B, and likewise for $\overline{BC}$ and $\overline{CD}$. We hasten to add that the trajectory of the light ray inside the index field, or, alternatively, a tentative approximation to its trajectory is needed in order for us to determine the arc lengths and angles in the above expression. If the exact trajectory is available from measurement, it is possible to recover the gradient of the refractive index using ray crossings (see Appendix 2) in lieu of our linear system approach.

Next, suppose that we have N path integrals in the form of Equation 3.3.9 and a system with 8 DoFs representing the partial derivative fields of the logarithmic index, 4 of which correspond to $f_{l,k}$ (as seen in Figure 3.3.2) and another 4 for $g_{l,k}$ at an identical set of sample locations. In addition, let the first of these path integrals correspond to the light ray shown in Figure 3.3.2. Using these integrals to construct the primary system of equations yields the matrix equation

$$\begin{bmatrix} 0 & -\overline{CD}sin\theta_{AB} & -\overline{AB}sin\theta_{BC} & \overline{BC}sin\theta_{CD} & 0 & \overline{CD}cos\theta_{AB} & \overline{AB}cos\theta_{BC} & \overline{BC}cos\theta_{CD} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \cdot \begin{bmatrix} f_{1,1} \\ f_{1,2} \\ f_{2,1} \\ f_{2,2} \\ g_{1,1} \\ g_{1,2} \\ g_{2,1} \\ g_{2,2} \end{bmatrix} = \begin{bmatrix} \Delta\theta_1 \\ \cdot \\ \cdot \\ \Delta\theta_N \end{bmatrix}, \tag{3.3.11}$$

where we have used Equation 3.3.10 as the first equation in the system and $\Delta\theta_1$ in the expression corresponds to the angular deflection value $\Delta\theta_{total}$ in the equation. Note that the column vector being multiplied in the system equation now contains DoFs from both field quantities $f_{l,k}$ and $g_{l,k}$. A quick comparison with Equation 3.2.4 shows that the number of DoFs has doubled in the system equation; this is a direct consequence of treating the partial derivatives as distinct field quantities. This does not, however, imply that twice the number of independent path integrals is needed to fully determine the solution. The irrotational constraint in Equation 3.3.8 provides additional algebraic equations that effectively reduce the DoFs by a factor of 2. These constraints can be discretized in the form

$$\left(f_{l,k+1} - f_{l,k}\right) - \left(g_{l+1,k} - g_{l,k}\right) = 0 \tag{3.3.12}$$

and be used to augment the linear system in Equation 3.3.11. For a $L \times K$ rectangular grid used to sample the partial derivative fields, Equation 3.3.12 provides $(L-1) \times (K-1)$ additional equations. If $N = L \times K$, the slight deficit of equations leads to an underdetermined system. In practice, however, this deficit becomes inconsequential because the number of independent path integrals (N) needed to accurately invert the system is significantly larger than the number of sample points on the rectangular grid.

The resulting system of equations constructed from the measured deflectometry path integrals and the augmented irrotational constraints takes the form

$$[S] \cdot \vec{\delta} = \vec{P}\,, \tag{3.3.13}$$

where $[S]$ is the system matrix and $\vec{\delta}$ contains all the unknown values of $f_{l,k}$ and $g_{l,k}$ lexicographically ordered into a single column vector. The number of rows in $[S]$ and the number of elements in $\vec{P}$ are both equal to $N + (L-1) \times (K-1)$, whereas the number of elements in $\vec{\delta}$ is equal to $2 \times L \times K$. The top N rows of $[S]$ correspond to the deflectometry path integrals and are composed of coefficients that are multiplied with the system's DoFs, as seen in the discretized expression in Equation 3.3.9, for instance. The bottom $(L-1) \times (K-1)$ rows correspond to the irrotational constraints in the form of Equation 3.3.12 and consist of strictly binary values. Accordingly, the first N elements of $\vec{P}$ correspond to the respective angular deflection values, while the remaining spots are filled with zeros. Assuming that there are enough equations to invert the system, we can, in principle, solve for the DoFs of the system using the inverse

formula $\vec{\delta} = [S]^{-1} \cdot \vec{P}$. If the system is overdetermined, it can instead be inverted by invoking the least squares formula $\vec{\delta} = ([S]^T \cdot [S])^{-1} \cdot [S]^T \cdot \vec{P}$.

3.4 Improved System Construction

In this section, we will demonstrate more advanced techniques in improving the discrete representation of the deflectometry path integral in constructing our linear system. In particular, we will employ linear constrained cubic spline interpolation in our model to help reduce the quantization effects in the discretization process. The material presented in this section is highly numerical in nature and is not considered essential to the remaining sections of this chapter. However, any numerical implementation of our deflectometry model for 2-D index distributions will utilize these techniques to improve reconstruction accuracy. With this in mind, readers who wish to view these improvements as a black box are encouraged to forego the material in this section.

In the previous section, we showed that the underlying principles behind our linear system approach to the inverse problem can be applied to the deflection of light rays inside a GRIN medium. In particular, we showed that the overall angular deflection can be expressed as a path integral in the form of Equation 3.3.7. If we assume the partial derivatives of the index field to be constant within each rectangular element associated with each sample point (see, *e.g.*, Figure 3.2.1), then the path integral can be discretized as a Riemann sum in the form of Equation 3.3.9. While this crude approximation is useful in understanding the application of our linear system model for deflectometry, the inherent quantization effects often lead to significant errors in the reconstructed index field. In order to mitigate these effects, especially where relatively few sample points are used to represent the partial derivative fields, we employ a linear variant of constrained cubic splines, which we develop in Appendix 3, to interpolate the quantities in Equation 3.3.7 along the trajectory of the path integral.

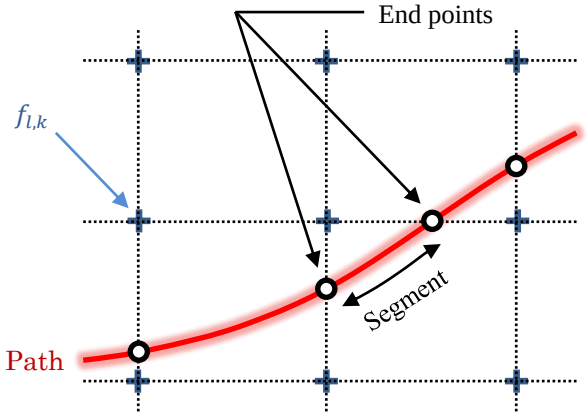


Figure 3.4.1. Partitioning a field quantity into rectangular regions for numerical evaluation of path integrals.

Before we describe the formalism for the interpolation process, let us first simplify the problem by writing an analogous path integral to Equation 3.3.7,

$$\sigma = \int_a^b ds \cdot cos(\theta)\, f(x,y)\,, \tag{3.4.1}$$

where $f(x,y)$ is the unknown field quantity we wish to interrogate. We start by partitioning the field quantity into rectangular regions so that the system's DoFs are located at the corners, as shown in Figure 3.4.1, rather than at the center (as shown previously in Figure 3.3.2). Similar to the discretization process outlined in Sections 3.2 & 3.3, the trajectory of the path integral is divided into segments contained within each rectangular region, as seen in the figure. Using the depicted partitioning scheme, 1-D linear constrained cubic splines are sufficient in interpolating $f(x,y)$ from the sample values $f_{l,k}$. Of course, the interpolated values for $f(x,y)$ at the end points must be expressed as linear combinations of $f_{l,k}$ since the values of $f_{l,k}$ are unknown at this point (see Appendix 3). In order to further reduce quantization error in the numerical evaluation of Equation 3.3.7 for each path segment, intermediate values of $f(x,y)$ are introduced between the end points along each segment, as shown in Figure 3.4.2. These intermediate values can also be obtained through linear constrained cubic spline interpolation using the interpolated values of $f(x,y)$ at the end points of the path segment as the spline knots. Note that the independent parameter for these splines changes to arc length s along the trajectory. Incorporating these intermediate points into the numerical evaluation of Equation 3.4.1 for a particular path segment requires knowledge of θ and ds at the intermediate points as well as the end points of the segment. These quantities are assumed to be independent of $f(x,y)$ in our linear system model and can therefore be interpolated directly from a tentative approximation to the trajectory of the path integral (*e.g.*, a cubic polynomial generated from boundary values of position and slope).

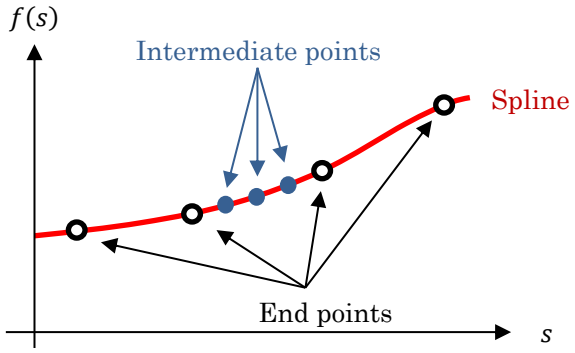


Figure 3.4.2. Splines used for interpolation at intermediate points along each path segment.

We are now ready to improve upon the construction procedure of our linear system by utilizing spline interpolation to mitigate undesirable quantization effects. As usual, matrix algebra will be applied liberally to facilitate the organization of coefficients associated with the system's DoFs in the path

integrals. Suppose that we wish to evaluate Equation 3.4.1 for a particular path segment inside a discrete field quantity $f_{l,k}$ with 16 DoFs, as shown in Figure 3.4.3.

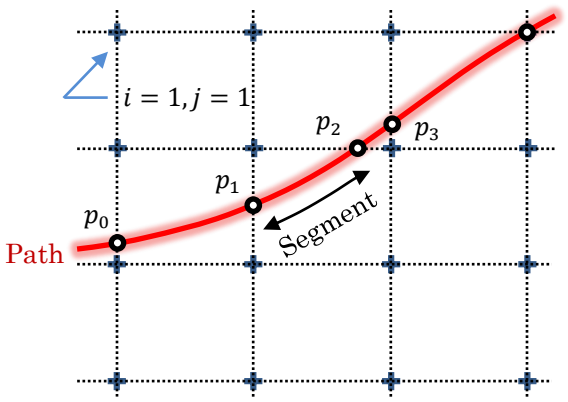


Figure 3.4.3. Partitioning a path integral for numerical evaluation.

We start by lexicographically scanning the discrete field $f_{l,k}$ into a column vector,

$$\begin{bmatrix} f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} \\ f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} \\ f_{3,1} & f_{3,2} & f_{3,3} & f_{3,4} \\ f_{3,1} & f_{3,2} & f_{3,3} & f_{4,4} \end{bmatrix} \rightarrow \begin{bmatrix} f_{1,1} \\ f_{1,2} \\ f_{1,3} \\ f_{1,4} \\ f_{2,1} \\ f_{2,2} \\ \vdots \\ f_{4,4} \end{bmatrix} \equiv \vec{f} \, . \tag{3.4.2}$$

We then determine the end points of the path segment by locating the intersections between the trajectory and the boundaries of the rectangular region containing the segment. For the highlighted path segment, its left end point p_1 occurs on a vertical boundary, whereas its right end point p_2 resides on a horizontal boundary. We emphasize orientation here because the 1-D spline used to interpolate $f(x, y)$ at the end points lies along the boundary line. Starting with p_1 at Cartesian coordinates (x_1, y_1), the second column of vertices (or DoFs) in the figure is used to construct the cubic spline. The four vertices are arranged into a column vector in ascending y-value, and the value of $f(x, y)$ at p_1 is readily obtained by evaluating the spline using the formula (see Appendix 3)

$$f_1 = f(x_1, y_1) = [\boldsymbol{X}][\boldsymbol{M}]^{-1}[\boldsymbol{L}] \begin{bmatrix} f_{2,1} \\ f_{2,2} \\ f_{2,3} \\ f_{2,4} \end{bmatrix}, \tag{3.4.3}$$

where $[\boldsymbol{X}] = [{y_1}^3 \quad {y_1}^2 \quad y_1 \quad 1]$ and $f_{l,k}$ corresponds to the vertex located in the l^{th} column and the k^{th} row in the figure. Note that that y is the independent parameter for this spline, as implied by $[\boldsymbol{X}]$.

In order to include multiple splines into a single equation, it is convenient to designate a *pertinence* matrix $[\boldsymbol{P}]$ containing strictly binary values so that the product of $[\boldsymbol{P}]$ and $\vec{f}$ defined in Equation 3.4.2 selects the pertinent vertices to be used for interpolation. For instance, the interpolation performed in Equation 3.4.3 implies that

$$[\boldsymbol{P}]\vec{f} = \begin{bmatrix} f_{2,1} \\ f_{2,2} \\ f_{2,3} \\ f_{2,4} \end{bmatrix}. \tag{3.4.4}$$

Using this nomenclature, we can rewrite Equation 3.4.3 as

$$f_1 = f(x_1, y_1) = [\boldsymbol{X}]_1 [\boldsymbol{M}]_1{}^{-1} [\boldsymbol{L}]_1 [\boldsymbol{P}]_1 \vec{f}\,, \tag{3.4.5}$$

where subscripts have been appended in order to identify the end point in the interpolation. The value for f at p_2 can be obtained in a similar manner using the formula

$$f_2 = f(x_2, y_2) = [\boldsymbol{X}]_2 [\boldsymbol{M}]_2{}^{-1} [\boldsymbol{L}]_2 \begin{bmatrix} f_{1,2} \\ f_{2,2} \\ f_{3,2} \\ f_{4,2} \end{bmatrix} = [\boldsymbol{X}]_2 [\boldsymbol{M}]_2{}^{-1} [\boldsymbol{L}]_2 [\boldsymbol{P}]_2 \vec{f}\,, \tag{3.4.6}$$

where $[\boldsymbol{X}]_2 = [{x_2}^3 \quad {x_2}^2 \quad x_2 \quad 1]$ since x is the independent parameter for this spline. It is apparent from the expression that the second row of vertices in Figure 3.4.3 pertains to the interpolation of f_2. Also, matrices $[\boldsymbol{M}]_2$, $[\boldsymbol{L}]_2$ and $[\boldsymbol{P}]_2$ are different from their analogous counterparts in Equation 3.4.5 and must be redefined for each end point. Under special circumstances, the number of vertices may be reduced to three (see Appendix 3), for which case $[\boldsymbol{L}]$ and $[\boldsymbol{P}]$ also need to be modified to accommodate the reduction.

Applying the same principles to the intermediate points shown in Figure 3.4.2, we use the interpolated values of f at the end points as spline knots for generating a new spline that lies on the trajectory of the path integral, with arc length s along the trajectory as the new independent parameter. In addition to the interpolated values at p_1 and p_2, we also need the interpolated values of f at p_0 and p_3 to use as spline knots. After obtaining these values in similar fashion, we combine the linear expressions for f_i (where $i = 0,1,2,3$) into a single concatenated matrix equation,

$$\begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} [\boldsymbol{X}]_0 [\boldsymbol{M}]_0{}^{-1} [\boldsymbol{L}]_0 [\boldsymbol{P}]_0 \\ [\boldsymbol{X}]_1 [\boldsymbol{M}]_1{}^{-1} [\boldsymbol{L}]_1 [\boldsymbol{P}]_1 \\ [\boldsymbol{X}]_2 [\boldsymbol{M}]_2{}^{-1} [\boldsymbol{L}]_2 [\boldsymbol{P}]_2 \\ [\boldsymbol{X}]_3 [\boldsymbol{M}]_3{}^{-1} [\boldsymbol{L}]_3 [\boldsymbol{P}]_3 \end{bmatrix} \vec{f}\,. \tag{3.4.7}$$

Using this method of bookkeeping helps organize the information we have collected for spline interpolation. From the values of spline knots evaluated in the above expression, the polynomial coefficients for the spline along the path segment are given by

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = [\boldsymbol{M}]^{-1}[\boldsymbol{L}] \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix}. \tag{3.4.8}$$

Upon obtaining these coefficients, the values for f at all intermediate points can then be evaluated simultaneously by defining a new matrix analogous to the $[\boldsymbol{X}]$ matrix in Equation 3.4.3 as follows:

$$\vec{f}(s_i) = \begin{bmatrix} f(s_1) \\ f(s_2) \\ \vdots \\ f(s_N) \end{bmatrix} = \begin{bmatrix} {s_1}^3 & {s_1}^2 & s_1 & 1 \\ {s_2}^3 & {s_2}^2 & s_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ {s_N}^3 & {s_N}^2 & s_N & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}, \tag{3.4.9}$$

where s_i denotes the arc length at the i^{th} intermediate point and N is the total number of points on the path segment, encompassing all intermediate points as well as the two end points.

The next logical step is to evaluate the quantities ds and θ in Equation 3.4.1 at all N points along the path segment. The former becomes trivial if the chosen set of points is uniformly spaced in arc length along the path segment. The latter, on the other hand, must be interpolated from the trajectory itself, which we denote by $y(x)$. Since $y(x)$ is assumed to be independent of the field quantity f in our model, interpolating θ does not involve the system's DoFs and can be implemented straightforwardly using the equations developed in Appendix 3, for instance. A row vector $\vec{R}$ is then constructed as follows,

$$\vec{R} = \Delta s[cos\,\theta(s_1) \quad cos\,\theta(s_2) \quad \cdots \quad cos\,\theta(s_N)]\,, \tag{3.4.10}$$

where Δs is the uniform separation in arc length between adjacent points along the segment, so that multiplying $\vec{R}$ with $\vec{f}(s_i)$ yields the integral $\int_{p_1}^{p_2} ds \cdot cos[\theta(s)]f(s)$. Of course, the product $\vec{R}\vec{f}(s_i)$ yields a numerical value for the integral only if $\vec{f}$ is given. However, since $\vec{f}$ is unknown in the inverse problem, we take the row vector resulting from the multiplication of all matrices to the left of vector $\vec{f}$ in $\vec{R}\vec{f}(s_i)$ and call it $\vec{V}$, such that $\vec{R}\vec{f}(s_i) = \vec{V}\vec{f}$, *i.e.*,

$$\vec{V} = \Delta s[cos\,\theta(s_1) \quad \cdots \quad cos\,\theta(s_N)] \begin{bmatrix} {s_1}^3 & {s_1}^2 & s_1 & 1 \\ {s_2}^3 & {s_2}^2 & s_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ {s_N}^3 & {s_N}^2 & s_N & 1 \end{bmatrix} [\boldsymbol{M}]^{-1}[\boldsymbol{L}] \begin{bmatrix} [\boldsymbol{X}]_0[\boldsymbol{M}]_0{}^{-1}[\boldsymbol{L}]_0[\boldsymbol{P}]_0 \\ [\boldsymbol{X}]_1[\boldsymbol{M}]_1{}^{-1}[\boldsymbol{L}]_1[\boldsymbol{P}]_1 \\ [\boldsymbol{X}]_2[\boldsymbol{M}]_2{}^{-1}[\boldsymbol{L}]_2[\boldsymbol{P}]_2 \\ [\boldsymbol{X}]_3[\boldsymbol{M}]_3{}^{-1}[\boldsymbol{L}]_3[\boldsymbol{P}]_3 \end{bmatrix}, \tag{3.4.11}$$

Hence, $\vec{V}$ contains all coefficients to be multiplied by the system's DoFs in evaluating Equation 3.4.1 over a particular path segment. Constructing $\vec{V}$ for all the partitioned path segments along the trajectory of the overall path integral and summing the resulting $\vec{V}$ vectors yields a new row vector containing all the coefficients to be multiplied with the system's DoFs for evaluating Equation 3.4.1 over the entire trajectory of the path integral,

$$\vec{V}_{path} = \sum_{j=1}^{M} \vec{V}_j \,, \tag{3.4.12}$$

where M is the total number of path segments along the path integral and $\vec{V}_i$ corresponds to the i^{th} segment. As before, if the field quantity vector $\vec{f}$ is given, then multiplying $\vec{V}_{path}$ with $\vec{f}$ yields the numerically integrated value for Equation 3.4.1. We hasten to add that the row vector in Equation 3.4.9 can be modified to improve the accuracy of numerical integration, *e.g.*, by incorporating the trapezoidal rule.

We take a brief moment to address the issue of degeneracy for spline knots in the interpolation process. Spline knots within the immediate vicinity of one another can lead to ill-conditioning and numerical instability during the computation of linear constrained cubic splines and are hence termed *degenerate*. It is important to check the distance between adjacent end points along the trajectory of the path integral (as seen in Figure 3.4.3) to ensure that all end points are sufficiently separated. One way to avoid conditioning issues is to discard one end point whenever degenerate pairs are identified. The choice is arbitrary with the exception that spline knots at the ends of the path integral are always maintained. We illustrate the process of eliminating degeneracy in Figure 3.4.4, where the number of end points has been reduced from 7 to 4 while maintaining the end points of the overall trajectory. In the figure, Δs_{min} corresponds to the minimum separation between non-degenerate end points and is used as a threshold for identifying degeneracy.

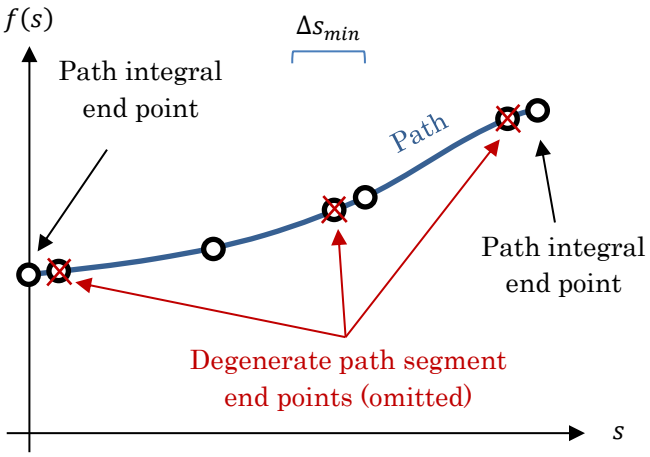


Figure 3.4.4. Eliminating degenerate end points from the path integral.

We now return to the discretization of the deflectometry path integral to complete our improved linear system model. It is apparent that Equation 3.4.1 only represents one portion of the overall deflectometry path integral upon inspection of Equation 3.3.7. The analogous counterpart is therefore

$$\delta' = -\int_a^b ds \cdot sin(\theta)\, f(x,y) \,. \tag{3.4.13}$$

We employ the same interpolation process in numerically evaluating this expression; the only change needed is the row vector in Equation 3.4.11, which becomes

$$\vec{R} = -\Delta s[sin\,\theta(s_1) \quad sin\,\theta(s_2) \quad \cdots \quad sin\,\theta(s_N)] \tag{3.4.14}$$

in the new integral. Consequently, the coefficient row vector in Equation 3.4.11 changes to

$$\vec{U} = -\Delta s[sin\,\theta(s_1) \quad \cdots \quad sin\,\theta(s_N)] \begin{bmatrix} {s_1}^3 & {s_1}^2 & s_1 & 1 \\ {s_2}^3 & {s_2}^2 & s_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ {s_N}^3 & {s_N}^2 & s_N & 1 \end{bmatrix} [\boldsymbol{M}]^{-1}[\boldsymbol{L}] \begin{bmatrix} [\boldsymbol{X}]_0[\boldsymbol{M}]_0{}^{-1}[\boldsymbol{L}]_0[\boldsymbol{P}]_0 \\ [\boldsymbol{X}]_1[\boldsymbol{M}]_1{}^{-1}[\boldsymbol{L}]_1[\boldsymbol{P}]_1 \\ [\boldsymbol{X}]_2[\boldsymbol{M}]_2{}^{-1}[\boldsymbol{L}]_2[\boldsymbol{P}]_2 \\ [\boldsymbol{X}]_3[\boldsymbol{M}]_3{}^{-1}[\boldsymbol{L}]_3[\boldsymbol{P}]_3 \end{bmatrix}, \tag{3.4.15}$$

and we have

$$\vec{U}_{path} = \sum_{j=1}^{M} \vec{U}_i \tag{3.4.16}$$

in lieu of Equation 3.4.12, where $\vec{U}_i$ corresponds to the coefficient row vector for the i^{th} path segment. A closer examination of Equation 3.3.7 reveals that if we were to replace the field quantity in Equation 3.4.13 with the discrete partial derivate $\left.\frac{\partial w}{\partial x}\right|_{l,k}$ and replace the field quantity in Equation 3.4.1 with $\left.\frac{\partial w}{\partial y}\right|_{l,k}$, then the overall angular deflection can be expressed as

$$\Delta\theta = \sigma' + \sigma = \vec{U}_{path}\vec{f} + \vec{V}_{path}\vec{g}\,, \tag{3.4.17}$$

where $\vec{f}$ contains all values of $\left.\frac{\partial w}{\partial x}\right|_{l,k}$ lexicographically scanned into a single column vector and $\vec{g}$ contains all values of $\left.\frac{\partial w}{\partial y}\right|_{l,k}$ arranged in similar fashion. Equation 3.4.17 describes the numerical evaluation of the overall deflectometry path integral in Equation 3.3.7 and can be rewritten in the convenient form

$$\Delta\theta = \left[\vec{U}_{path},\, \vec{V}_{path}\right]\vec{\delta}\,, \tag{3.4.18}$$

where $\vec{\delta}$ is the concatenation of column vectors $\vec{f}$ and $\vec{g}$ and contains all unknown values of both discrete partial derivative fields. This expression is a vast improvement over the discrete representation used in Equation 3.3.9 and will be used for discretizing deflectometry path integrals for the remainder of this chapter.

A full numerical implementation for the improved discrete representation of the deflectometry path integral is shown in Figure 3.4.5, where the trajectory is a cubic polynomial constructed from the specified boundary conditions. The units along x and y are arbitrary, provided that the same unit of measure is used for both spatial variables. We defined the partial derivative fields $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$ on a 11×11 uniform rectangular grid and partitioned the 2-D simulation space into rectangular regions in the same

manner as Figure 3.4.1. The trajectory was then partitioned into individual path segments, similar to Figures 3.4.1 & 3.4.3.

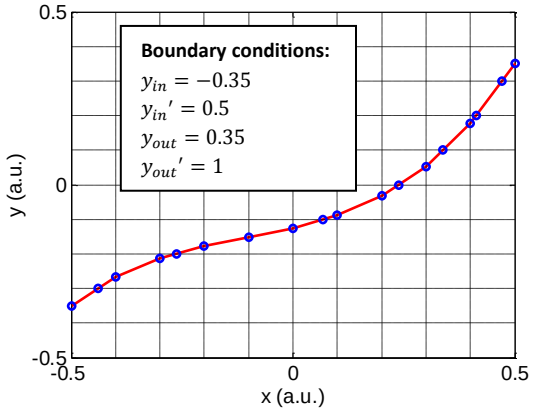


Figure 3.4.5. Improved discretization of a deflectometry path integral with a cubic polynomial trajectory.

With the help of Equations 3.4.15 & 3.4.16, we obtained $\vec{U}_{path}$ for the depicted trajectory and arranged the coefficients in the row vector into matrix form to visualize the DoFs involved in evaluating the deflectometry path integral. The coefficient matrix is imaged in Figure 3.4.6a, where darker shades correspond to increasingly negative values and all values are relative to the minimum value in $\vec{U}_{path}$. Note that since $0 < \theta \leq \frac{\pi}{2}$ for this particular trajectory, $\vec{U}_{path}$ contains mostly negative values due to the $-sin(\theta)$ term in Equation 3.4.13. Similarly, we obtained $\vec{V}_{path}$ from Equations 3.4.11 & 3.4.12 and imaged its coefficients in similar fashion in Figure 3.4.6b.

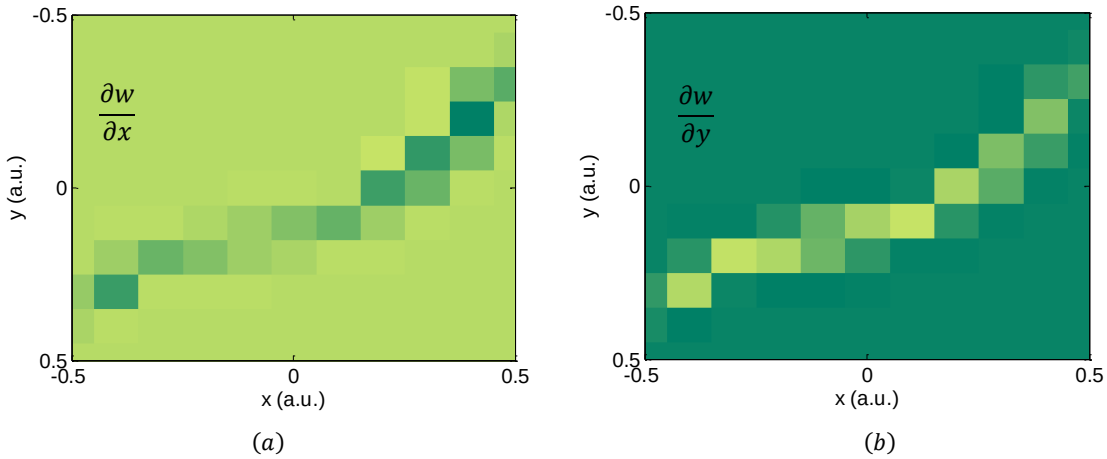


Figure 3.4.6. Visualizing the coefficients to be multiplied with (a) $\frac{\partial w}{\partial x}\Big|_{l,k}$ and (b) $\frac{\partial w}{\partial y}\Big|_{l,k}$.

3.5 Recovering the Index Field

In this section, we will demonstrate the inversion of the primary system of equations described in Section 3.3 and solve for the partial derivative fields from computer-generated deflection data. We will then integrate the partial derivatives to reconstruct the unknown index field. The deflectometry data used to establish the primary system equation will be generated through a ray trace inside a virtual test index field to simulate measurements obtained in an actual experiment. This has the advantage of allowing us to measure the intrinsic accuracy of the numerical method separate from the noise associated with experimental measurements. An outline of the overall procedure for recovering the index distribution from measured deflectometry data is provided below, where it is understood that the index field and its gradient have been discretized on a uniform rectangular grid, and the family of light rays used to interrogate the medium cover all grid points. Each step in the procedure will be covered in detail in the ensuing discussion.

1. Measure the index of refraction along the surface using, *e.g.*, a refractometer.
2. Measure the exit position and slope for a family of interrogating rays launched at specific entrance positions and slopes.
3. Ascertain internal ray slopes along surface of the medium using Snell's law.
4. Construct cubic polynomials that satisfy internal boundary values to approximate the unknown trajectories of the interrogating rays.
5. For each ray trajectory, derive the corresponding deflectometry path integral as an algebraic equation in $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$.
6. Assemble the algebraic equations into a linear system and augment the system with irrotational constraints to enforce a conservative gradient field.
7. Invert the overall system equation to solve for ∇w and integrate ∇w to compute the refractive index n.

Our simulations reveal that an accurate reconstruction of the rectangular index field is possible using strictly interrogating rays that propagate between opposite boundaries of the GRIN medium if the internal boundary ray slopes are known or can be measured indirectly. This helps reduce the complexity of real metrology systems designed to acquire the deflectometry data. In practice, the internal boundary ray slopes can be obtained through Snell's law if the surface refractive index is known, which we accomplish in step 1. In our computer-based demonstration, step 1 is replaced by our knowledge of the surface index for the predetermined test index field. It follows that the internal ray slopes can be determined without knowledge of the surface index using an iterative approach, which we will defer to Section 3.8.

For the sake of analysis and ease of implementation, let us divide the measurements in step 2 into two categories so that the light rays propagating strictly between the left and right boundaries are grouped together and similarly for the rays propagating between the top and bottom boundaries. To help shed light on the various numerical issues that one may encounter during the retrieval process, suppose that, for the

time being, we were to interrogate the index field using only the first group of rays. Under these circumstances, only a small set of measurements are able to sample the system's DoFs near the top and bottom surfaces of the index field; this typically leads to extremely high error sensitivity in these regions due to the numerical conditioning of the system (which we will discuss in detail in Section 3.7). Moreover, we showed in Section 3.3 that the index gradient component lying normal to the local ray vector is responsible for its deflection. Based on this principle, it is clear that the interrogating rays in the first group do not provide adequate sampling of $\partial w/\partial x$ because their direction of propagation is predominantly horizontal. Attempting to solve for $\partial w/\partial x$ will produce highly inaccurate results. Likewise, the reconstruction of $\partial w/\partial y$ will also contain large errors if we restrict ourselves to light rays propagating strictly between the top and bottom boundaries of the index field. Hence, both groups of measurements are needed to accurately invert the deflectometry system.

We propose to interrogate the index field using light rays that connect all possible pairings of predetermined entry and exit points positioned uniformly along opposite surfaces of the rectangular index field. This provides a systematic, albeit uneven, sampling of all the DoFs in the system. To contrive such a family of interrogating rays, we start with a sweep of ray traces originating at an entry point a_i and launch the light rays at various angles, as shown in Figure 3.5.1. An optimal ray is then chosen from the traced trajectories based on the ray's proximity to the target point b_j upon reaching the exit boundary, which we highlight in the figure as a solid curve.

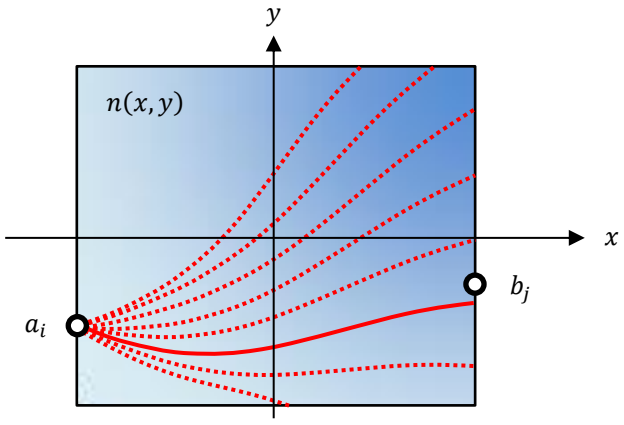


Figure 3.5.1. Numerical sweep of traced light rays used to identify the optimal trajectory connecting a pair of entry and exit points.

In similar fashion, the same set of traced trajectories from the sweep can be used to identify an entire set of optimal light rays connecting a particular entry point to each of the specified exit points, as seen in Figure 3.5.2 as solid curves (note that these rays do not necessarily need to exit at the target locations); the minor discrepancies in the rays' exit positions has very little impact on the reconstruction accuracy of the index field, provided that no degenerate trajectories result from the selection process. Using

M entry points and an equal number of exit points, this process is repeated for each entry point until we obtain M^2 interrogating rays propagating between the left and right boundaries. Another set of M^2 rays propagating between the top and bottom boundaries are generated in the same manner, bringing the total number of rays used in the interrogation to $2M^2$. Once we have acquired both sets of measurements, the boundary values of position and slope for the individual rays are stored as data for establishing the primary system equation, and the traced trajectories are discarded since our model does not assume knowledge of this information. Of course, application of Snell's law is implied at the surface of the GRIN element to compute the corresponding external ray slopes during the ray trace. If the internal ray slope exceeds the critical value for total internal reflection at either end of the interrogating ray, the corresponding measurement cannot be used to invert the deflectometry system (since the ray cannot be measured externally) and is discarded from the test data.

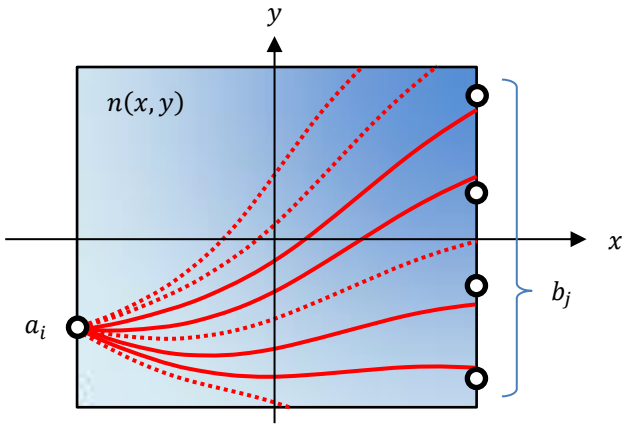


Figure 3.5.2. Light rays assigned to pairings between a particular entry point and all exit points.

Step 3 then consists of applying Snell's law at the entrance and exit boundaries (in reverse) to compute a set of internal boundary ray slopes. The principles behind the subsequent steps leading up to step 7 have already been discussed at length in the preceding sections of this chapter, so we will not reiterate them here. In short, after defining the partial derivatives in question on a discrete rectangular grid, the tentative approximations to the unknown trajectories of the interrogating rays obtained in step 4 enable us to derive linear algebraic equations describing the deflectometry path integral associated with each interrogating ray in step 5, for which we employ the advanced discretization techniques developed in Section 3.4. In step 6, irrotational constraints applied to the discrete points on the sampling grid are appended to the primary system of equations to enforce a conservative gradient field (see Section 3.3). Upon obtaining an overall system equation, we arrive at step 7 in the procedure, whose mathematical operations are highly numerical in nature and warrant an in-depth discussion.

In theory, computing the inverse of $[S]$ (or $([S]^T \cdot [S])^{-1}$ if $[S]$ is not a square matrix) in the system equation allows us to solve for the partial derivative fields. In practice, however, we require

methods of inversion with better numerically stability in order to reduce the system's sensitivity to error. Due to the similarity in the coefficients produced by similar ray trajectories, $[S]$ is inherently ill-conditioned. As a result, any error in the deflectometry data can be greatly amplified in the solution. In addition, $[S]$ is a sparse matrix because only a small set of DoFs pertain to the individual deflectometry path integrals. These factors generally make the direct inversion of $[S]$ numerically unstable. The only numerical technique we have found to be reliable in inverting the primary system equation is the least squares QR factorization (LSQR) method. While similar to other iterative inversion techniques, the LSQR method has been shown to exhibit superior numerical stability when the system is ill-conditioned [Barrett, *et al.*, 1994; Paige & Saunders, 1982]. Moreover, the method generates a least squares solution for overdetermined systems (which are common in our formulation).

In order to quantify inversion accuracy, it is useful to define a metric based on the RMS value of the system discrepancy vector given by

$$\vec{\varepsilon} \equiv [S] \cdot \vec{\delta}_{sol} - \vec{P} \,, \tag{3.5.1}$$

where $\vec{\delta}_{sol}$ is the computed solution to $\vec{\delta}$ in the primary system equation. We use ε_{RMS} to quantify the calculated solution's ability to reproduce the input deflectometry data. In our search for a suitable inversion technique, we tested several alternatives to the LSQR method including the biconjugate gradients method [Barrett, *et al.*, 1994], the symmetric LQ method [Paige & Saunders, 1975], the generalized minimum residual method [Youcef & Schultz, 1986], the conjugate gradients squared method [Barrett, *et al.*, 1994; Sonneveld, 1989], and the quasi-minimal residual method [Freund & Nachtigal, 1991]. In all test cases, the LSQR method consistently provided the most reliable solution with the lowest ε_{RMS}. Given this result, we will employ the LSQR method exclusively for inverting the primary system equation as well as any variant derived from it.

Upon obtaining the partial derivatives from the system inversion, we integrate the resulting 2-D gradient vector field specified by $\vec{\delta}_{sol}$ by inverting the discrete operator associated with the computation of partial derivatives for the potential function we seek. To find the inverse operator, let us examine the discrete algebraic equations used to compute the partial derivatives at each grid point. Specifically, the central difference is used to compute the partial derivatives at interior grid points, *i.e.*,

$$\left.\frac{\partial w}{\partial x}\right|_{i,j} = \frac{1}{2\Delta_x}\left(w_{i+1,j} - w_{i-1,j}\right) , \tag{3.5.2a}$$

$$\left.\frac{\partial w}{\partial y}\right|_{i,j} = \frac{1}{2\Delta_y}\left(w_{i,j+1} - w_{i,j-1}\right) , \tag{3.5.2b}$$

where Δ_x is the separation between grid points along x and Δ_y is the separation in y. For leading-edge grid points, we take the forward difference

$$\left.\frac{\partial w}{\partial x}\right|_{i_{min},j} = \frac{1}{\Delta_x}\left(w_{i_{min}+1,j} - w_{i_{min},j}\right) , \tag{3.5.3a}$$

$$\left.\frac{\partial w}{\partial y}\right|_{i,j_{min}} = \frac{1}{\Delta_y}\left(w_{i,j_{min}+1} - w_{i,j_{min}}\right), \tag{3.5.3b}$$

whereas the backward difference is used for trailing-edge grid points,

$$\left.\frac{\partial w}{\partial x}\right|_{i_{max},j} = \frac{1}{\Delta_x}\left(w_{i_{max},j} - w_{i_{max}-1,j}\right), \tag{3.5.4a}$$

$$\left.\frac{\partial w}{\partial y}\right|_{i,j_{max}} = \frac{1}{\Delta_y}\left(w_{i,j_{max}} - w_{i,j_{max}-1}\right). \tag{3.5.4b}$$

These linear algebraic equations in $w_{i,j}$ are assembled into a sparse system matrix describing the discrete gradient operator, whose inverse can be obtained using standard numerical methods. Letting the potential function $w(x,y) = w_1(x,y) + w_0$ where w_0 is some additive constant such that $\nabla w = \nabla w_1$, we use the inverse gradient operator to compute the discrete potential function $w_1|_{l,k}$ from the recovered partial derivative fields,

$$[\nabla]^{-1} \cdot \vec{\delta}_{sol} = \vec{w}_1 , \tag{3.5.5}$$

where $[\nabla]$ is the discrete gradient operator, $\vec{\delta}_{sol}$ is the computed solution to the deflectometry system equation, and $\vec{w}_1$ contains all the discrete values of $w_1|_{l,k}$ lexicographically ordered into a column vector. Generally speaking, inverting the primary system equation will not produce a conservative gradient field due to noise sources present in the system. The inconsistency manifests as non-zero values corresponding to the augmented irrotational constraints in the system discrepancy vector. If the partial derivative fields are inconsistent, *i.e.*, they do not satisfy the irrotational constraint, then a least squares solution is generated using the inverted gradient operator. We hasten to add that even with the augmented irrotational constraints in the deflectometry system equation, the inversion of an overdetermined system (as is generally the case) always generates a least squares solution. Hence, any noise present in the system will likely result in a reconstructed gradient field with non-zero curl. Finally, the additive constant w_0 in the integration can be determined with additional information such as an independent measurement of the refractive index at some position, and the index distribution of the GRIN medium is given by $n(x,y) = e^{w(x,y)}$.

We are now equipped with the necessary tools to demonstrate the efficacy of our method for measuring 2-D index fields using deflectometry. In the following simulation, we used 800 light rays ($M = 20$) to interrogate a hypothetical test index field given by

$$n(x,y) = n_0 + \Delta n \cdot sin\left[\pi\left(x + \frac{L}{2}\right)\right] \cdot cos(\pi y) , \tag{3.5.6}$$

where L is the length of the GRIN medium along x, n_0 is a constant offset for the index, and Δn is its modulation amplitude. Figure 3.5.3 depicts the index distribution in 2-D space, where the units along x and y are completely arbitrary provided that we use the same unit of measure for both spatial variables.

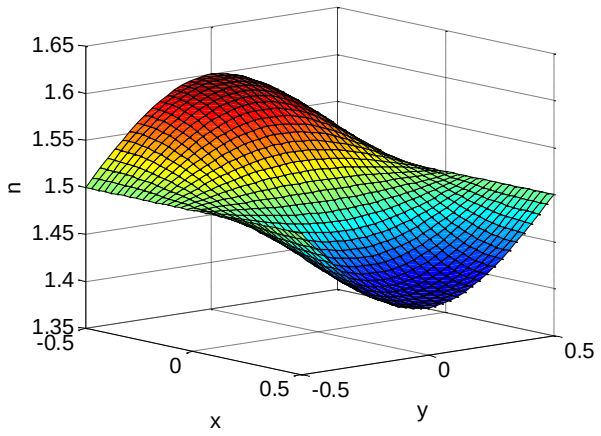


Figure 3.5.3. 3-D plot of the test index field used to demonstrate our deflectometry technique.

In the reconstruction process, we defined the partial derivative fields on a 15×15 uniform rectangular grid, implying a total of 225 DoFs in the system describing the discrete partial derivative $\frac{\partial w}{\partial x}\Big|_{l,k}$ and another 225 for $\frac{\partial w}{\partial y}\Big|_{l,k}$. We then used cubic functions to approximate the unknown ray trajectories using the deflectometry data generated from the initial ray trace and derived the overall system equation from the deflectometry path integrals. Upon inverting the system equation, the partial derivative fields were extracted from the solution $\vec{\delta}_{sol}$ and rearranged into matrix form for visualization purposes, which we show in Figures 3.5.4a & 3.5.4b. The vector gradient field $\nabla w = \frac{\partial w}{\partial x}\hat{\imath} + \frac{\partial w}{\partial y}\hat{\jmath}$ inferred from the partial derivatives is seen in Figure 3.5.4c. Figure 3.5.4d shows the system discrepancy vector $\vec{\varepsilon}$ defined in Equation 3.5.1, where the first 800 values indicate the agreement between the deflectometry data generated from the initial ray trace and the deflection of light rays caused by the reconstructed partial derivative fields; the remaining 196 values correspond to discrepancies from vanishing curls in the augmented irrotational constraints. Upon closer examination, it is evident that the largest index errors tend to occur in the corner regions of the sampling grid. This is primarily due to the reduced quality of interpolation near boundary grid points, which is exacerbated at these locations (see Appendix 3).

Since we have assumed knowledge of the surface index for the test index field (allowing us to forego step 1 in the retrieval algorithm), the integration constant w_0 is known, and the integration of ∇w yields the logarithmic refractive index $w(x, y)$ and taking its exponential recovers the refractive index distribution $n(x, y)$. The reconstructed index field, obtained from integrating the partial derivative fields shown in Figure 3.5.4, is shown in Figure 3.5.5a. In order to compute the reconstruction error, the test index field from Figure 3.5.3 was redefined by evaluating Equation 3.5.2 on the same 15×15 rectangular grid and its difference from the reconstructed index field is shown in Figure 3.5.5b.

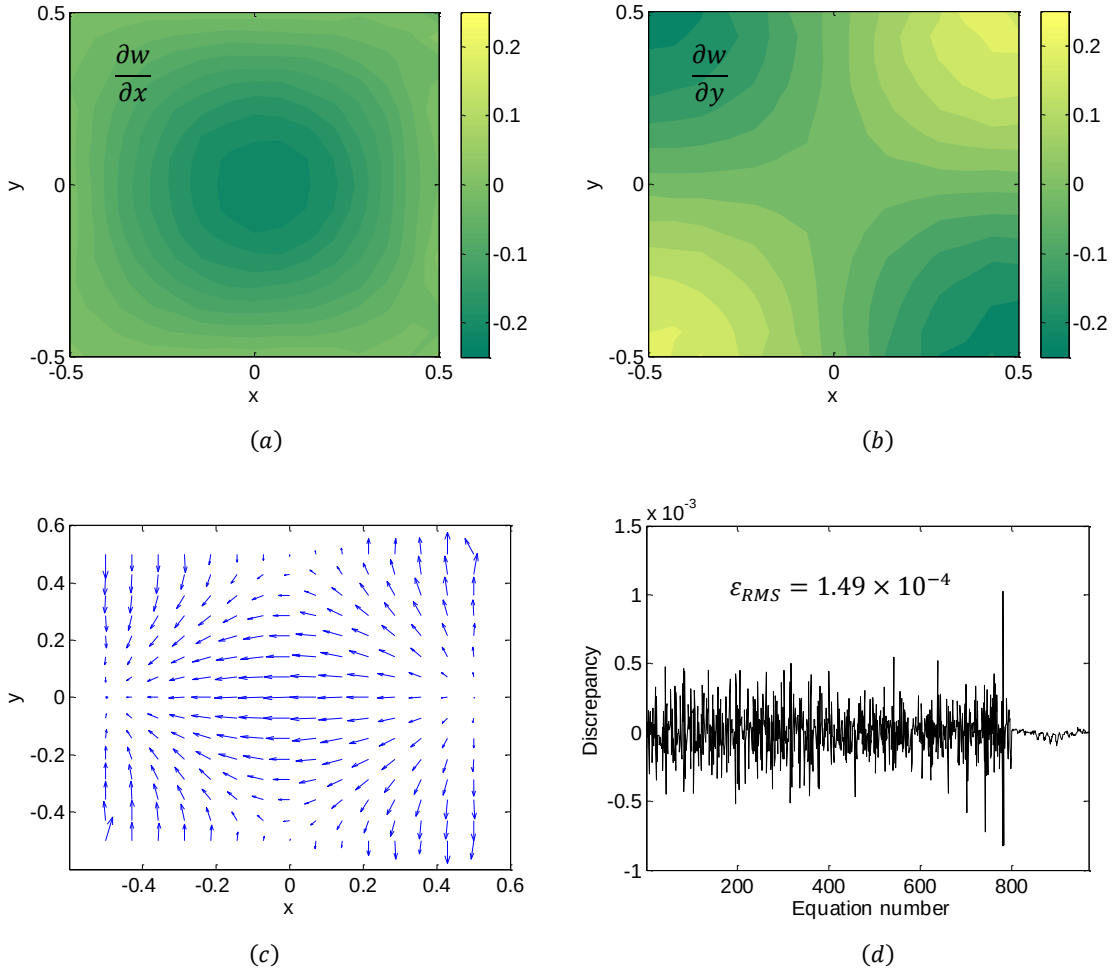
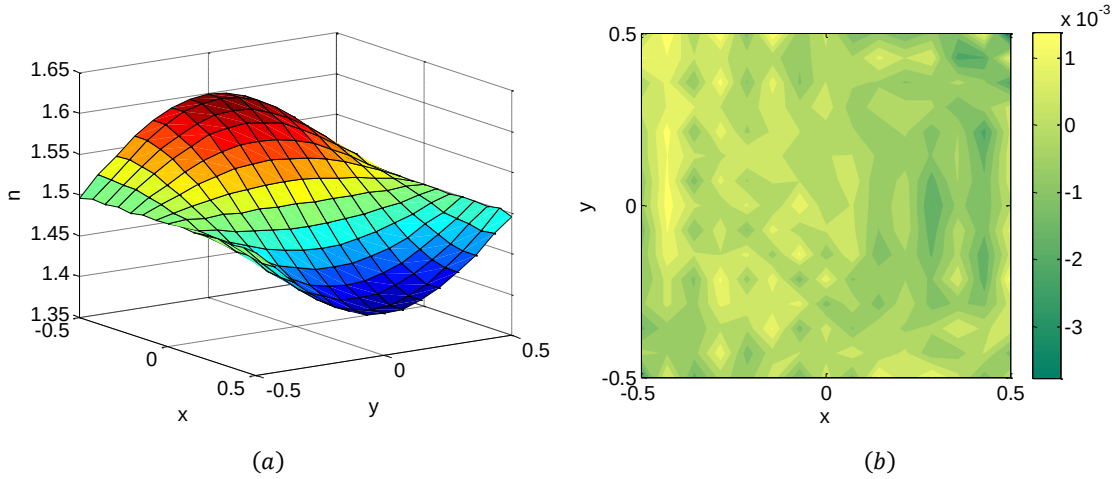


Figure 3.5.4. Partial derivative fields (a) $\frac{\partial w}{\partial x}$ and (b) $\frac{\partial w}{\partial y}$ reconstructed from deflectometry data after inverting the system equation, (c) the resulting vector gradient field ∇w, and (d) the system discrepancy vector $\vec{\varepsilon}$.

The strong agreement between the reconstruction and the test index field suggests that the reconstruction process is relatively insensitive to small variations in the trajectories of the interrogating rays; the fact that we used cubic functions to approximate the ray trajectories did not result in large errors in the recovered index field. In the case of slowly varying index fields, it stands to reason that the coefficients contained in $[S]$ remain relatively static when the trajectories are perturbed. This also makes sense from Fermat's principle. Since the optical path lengths traveled by the interrogating rays are the extrema in the optimization, variations in the solution due to small deviations from these extrema are inherently small since derivative of the optical path length vanishes around an extremum. This feature of the linear system can be exploited to obtain the actual ray trajectories using the cubic functions as a starting point, which we will demonstrate in the next section.



(a) (b)

Figure 3.5.5. (a) Reconstruction of test index field after integrating the recovered partial derivative fields and (b) index error relative to the test index field, with an RMS value of 8.86×10^{-4} *RIU.*

3.6 Correcting Ray Trajectories

Despite the strong agreement we were able to achieve in retrieving the test index field in the previous section, the solution is still fundamentally flawed as it is based on ray trajectories that are inconsistent with geometrical optics. We did not enforce the trajectory's dependence on the index field as mandated by the ray equation because relaxing this constraint enabled us to formulate the inverse problem as a linear system in Section 3.2. In this section, we will exploit the system's resilience to perturbations in the trajectories of the interrogating rays and develop a successive refinement process for improving our tentative approximations to these trajectories, thereby improving our mathematical description of the deflectometry system. Furthermore, we will show that the optimized reconstruction obtained through the procedure yields ray trajectories that are consistent with the ray equation.

In Section 3.5, we constructed the deflectometry system equation using cubic functions to approximate the ray trajectories in order to retrieve the index field. Using this as our starting point, two separate ray traces are performed inside the computed index field for each interrogating ray: In one instance, the ray is launched with initial conditions associated with its entry point and propagated the ray to its exit point; in the other, the ray is launched in reverse with the initial conditions associated with the exit point and propagated to its entry point in similar fashion. A weighted sum of the traced trajectories from both instances is then used to update our tentative approximation to the ray's trajectory in order to ensure that the boundary conditions are not violated in the refined approximation. The weighting coefficient starts at unity at the launch point of the ray trace and decays to zero as the trace approaches its exit point. Figure 3.6.1 illustrates three candidate functions for the weighting coefficient. Of the three functions, the sinusoid introduces the least amount of change in the boundary ray slopes due to its vanishing derivative (with respect to arc length) at the end points and is selected for our implementation. We hasten to add that a

numerical ray trace is possible based on the sole knowledge of the partial derivatives of the index field using the principles developed in Section 3.3. This means that the corrective procedure does not require knowledge of the integration constant used to reconstruct the index field from its partial derivatives.

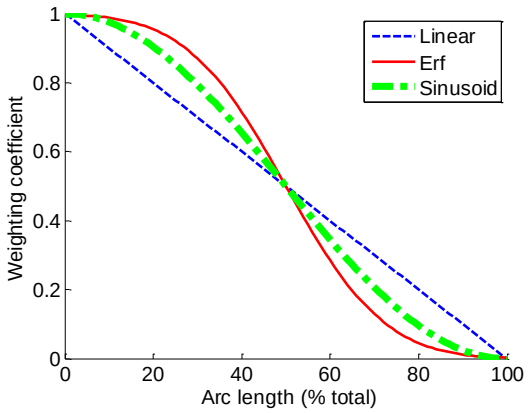


Figure 3.6.1. Candidate weighting functions for combining traced ray trajectories launched from both ends of an interrogating ray.

Upon obtaining the refined ray trajectories, we need to update the coefficients in $[S]$ and a new solution can be computed in the usual manner. This process can be repeated until agreement is achieved between the assumed approximate trajectories and the ray trace trajectories. In order to quantify the inconsistencies in the solution and monitor convergence of the algorithm, it is useful to define an error metric between the assumed trajectories and the traced trajectories; a natural choice is the integrated absolute difference between the two, *i.e.*,

$$\Delta = \int_a^b |y_{trace}(x) - y_{initial}(x)| dx \,, \tag{3.6.1}$$

where $y_{trace}(x)$ is the trajectory obtained from numerically tracing a ray launched at the entry point of the interrogating ray with the appropriate initial conditions and $y_{initial}(x)$ is the approximation used for the trajectory in the computation of $[S]$ prior to system inversion during each iteration.

To incorporate the corrective ray trace algorithm we have just described into the overall inversion procedure outlined at the beginning of Section 3.5, we augment it with the following steps:

8. Use a ray trace to compute the trajectory of each interrogating ray inside n.
9. Refine the approximate trajectories based on ray trace results while preserving boundary values.
10. Repeat steps 5-9 until the error metrics corresponding to each interrogating ray converge.

Using this procedure to refine the initial cubic polynomial trajectories used to reconstruct the index field in Section 3.5, we show the progressive improvement in the overall consistency of the solution after each

iteration in Figure 3.6.2a, where a strong agreement between approximated and traced trajectories is reached after just three iterations. In the figure, the interrogating rays are sorted in ascending order based on their error metric in order to prevent coincidental patterns from developing in the plot.

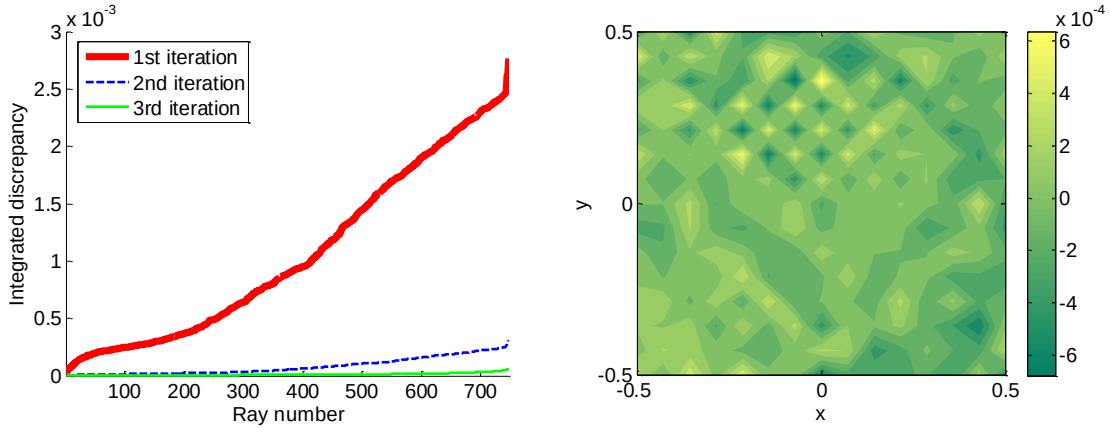


Figure 3.6.2. (a) Progressive improvement in the error metric evaluated for individual interrogating rays after three iterations of refining ray trajectories and (b) difference between the initial and final reconstructions of the index field, with an RMS value of 1.96×10^{-4} *RIU.*

The overall RMS index error relative to the test index field did not show a discernible improvement after making corrections to the ray trajectories in this particular simulation. However, we did observe a significant difference between the reconstructed index fields in the first and last iterations of the corrective procedure, which we illustrate in Figure 3.6.2b. This result confirms that the inversion process is extremely robust when the deflectometry system is subjected to perturbations in its ray trajectories, as we alluded to at the end of Section 5. Furthermore, it serves as indication that only a small portion of the reconstruction error observed in Section 3.5 can be attributed to the error associated with approximate ray trajectories. This justifies our strategy in establishing a linear system model of the inverse problem by initially ignoring the refractive effects of the index field and using the measured boundary conditions to generate approximate trajectories for the interrogating rays.

3.7 Reconstruction Error Analysis

Although the aim of the numerical demonstration in the previous section was to model exact experimental measurements, the deflectometry data used in the inversion procedure was generated from a numerical ray trace. The ray trace itself is, of course, susceptible to quantization effects and introduces random noise to the data, contributing to the observed RMS error in the reconstruction. Specifically, this noise can be attributed to the discretization error associated with Euler's method (see Appendix 1) as well

as the error associated with interpolating $n(x, y)$ and its gradient from its discrete sample values $n_{l,k}$. To help reduce the quantization noise in the generated data, a higher resolution was used to represent the test index field during the initial ray trace and the step size used in Euler's method was optimized. When we apply the data to the system inversion procedure, the quantization noise resulting from the initial ray trace is analogous to measurement error in an actual experiment as they directly influence the boundary values used to recover the index field. Of course, any noise associated with the ray trace is not present in an actual experiment, and the error in the measured deflectometry data is entirely dictated by measurement accuracy.

In addition to the quantization error associated with the ray trace data, the discretization of the deflectometry path integral is also susceptible to quantization effects and introduces noise to the computed coefficients in $[S]$. This error contribution acts independently of error contained in the deflectometry data and depends solely on the quality of interpolation used to obtain the quantities in the path integral and the quality of the numerical integration technique used in the discretization process (see Section 3.4). Finally, the inconsistency associated with unphysical gradient vector fields computed from the inversion procedure makes a contribution to the reconstruction error as well. In summary, the index error observed in the demonstrated reconstruction from Section 5 can be attributed to three sources of error: the quantization effects inherent in the initial ray trace used to generate the deflectometry data, the quantization effects associated with the discretization of the deflectometry path integrals, and the inconsistencies arising from the numerical inversion of the system. With this understanding of the base-line error produced during the reconstruction of the index field, we will examine the impact of uncertainty in boundary value measurements on reconstruction accuracy in the following simulations by manually introducing white Gaussian noise to the deflectometry data prior to system inversion.

Functionally speaking, the LSQR method used to invert the primary system equation is analogous to computing $[S]^{-1}$ in Equation 3.3.13, or $([S]^T \cdot [S])^{-1}$ if $[S]$ is overdetermined. Therefore, we expect the inversion process to be susceptible to the same base-line inversion sensitivity that is characteristic of any linear system, *e.g.*,

$$[A]\vec{x} = \vec{y}\,, \tag{3.7.1}$$

in which we seek the value of $\vec{x}$ using the inverse formula

$$\vec{x} = [A]^{-1}\vec{y}\,. \tag{3.7.2}$$

The system's sensitivity to error, also referred to as its conditioning, is characterized by the change in its output $\vec{x}$ due to small changes in the input $\vec{y}$. One way to quantify this is

$$\vec{\vartheta} = [A]^{-1}(\vec{y} + \vec{\epsilon}) - \vec{x}\,, \tag{3.7.3}$$

where the input vector $\vec{y}$ is contaminated with a noise vector $\vec{\epsilon}$ and $\vec{x}$ is computed from Equation 3.7.2. By inspection, the output from inverting the 2-D deflectometry system in Equation 3.3.13 is the partial

derivative vector $\vec{\delta}$, and the input consists of measured deflection values in $\vec{P}$ since the irrotational constraints used to augment the deflectometry path integrals are unaffected by measurement error. Using the same simulation parameters from our numerical implementation in Section 3.5, white Gaussian noise is introduced to the measured boundary values of ray position and ray slope on top of the base-line quantization noise present in the numerical ray trace prior to system inversion.

In order to isolate the base-line sensitivity of the inversion process, we first computed $[S]$ based on cubic polynomial approximations to the ray trajectories that fit the noise-free deflectometry data. We then held the coefficients in $[S]$ constant, whereas the deflection values in $\vec{P}$ in the system equation were contaminated with the white Gaussian noise, as implied by Equation 3.7.3. The resulting error in the index field reconstructed from the same set of interrogating rays used in Section 3.5 at various noise levels is depicted in Figure 3.7.1, where each data point represents an ensemble average over 100 trials of the simulation. As expected, the change in the output of the system scales linearly with small changes in its input, and the non-zero value at the very left of the plot corresponds to the reconstruction error observed in Section 3.5.

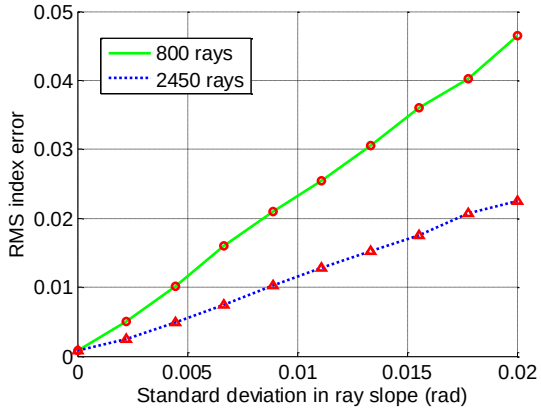


Figure 3.7.1. Index error resulting from contaminated deflection values using system coefficients calculated from noise-free data.

It follows that the base-line inversion sensitivity of the system is related to the number of interrogating rays used in the inversion. In theory, the minimum number of equations needed to fully specify a solution to a linear system is equal to the number of DoFs. In practice, however, any error in the input of the system can be greatly amplified in the inversion process if the system is ill-conditioned, as is generally the case for our 2-D deflectometry system. One way to mitigate the effects of ill-conditioning and reduce the system's sensitivity to error is to use redundant measurements. When used in conjunction with inversion methods that optimize the solution in the least squares sense, the redundant measurements help suppress error contributions from individual measurements by averaging over more samples, provided that the errors from different measurements are uncorrelated [Sweeney & Vest, 1973]. Recalling that the partial

derivative fields were defined on a 15×15 grid for the demonstrated reconstruction in Section 3.5 and were interrogated using 800 individual measurements (rays), and accounting for the augmented irrotational constraints, the overall deflectometry system consisted of roughly 1×10^3 algebraic equations and 450 unknowns, which leaves approximately 500 redundant measurements. In a similar test, we increased the number of interrogating rays from 800 to 2450 and illustrate the reduction in the system's error sensitivity in Figure 3.7.1. A comparison between the two plots reveals that the linear scaling factor between noise level and the resulting RMS index error varies inversely with the amount of data redundancy used in the reconstruction. This result is consistent with elementary statistics, where the averaging effect from redundant data is expected to scale as the square root of the sample count. Running a linear regression for the data sets depicted in the figure confirms that the ratio of their respective slopes is in close agreement with the expected value of $\sqrt{800}:\sqrt{2450}$. We hasten to add that the base-line inversion sensitivity depicted in the figure is specific to the LSQR method. While it is not within the scope of this thesis to investigate the error sensitivity of other numerical inversion techniques mentioned in Section 3.5, it is reasonable to assume that the characteristic slopes corresponding to their base-line inversion sensitivities will be significantly steeper relative to that of the LSQR method under identical conditions.

In a more realistic model of the deflectometry system, the coefficients in $[S]$ must be modified in accordance with the contaminated boundary values; these coefficients are generated from deflectometry path integrals whose trajectories must satisfy the measured (or generated) boundary values. As before, the deflection values in $\vec{P}$ are also subjected to the contamination. The scenario we have just described is analogous to perturbing y in Equation 3.7.3 and inducing changes in the system matrix $[A]$. In order to account for these changes, we incorporated the path dependence of $[S]$ into the inversion process by recalculating its coefficients from modified ray trajectories that reflect the changes in contaminated boundary values of ray slope. The results, as seen in Figure 3.7.2, reveal this dependence to be super-linear when the index error is plotted against the artificially introduced noise. Once again, each data point represents an ensemble average over 100 trials of the simulation.

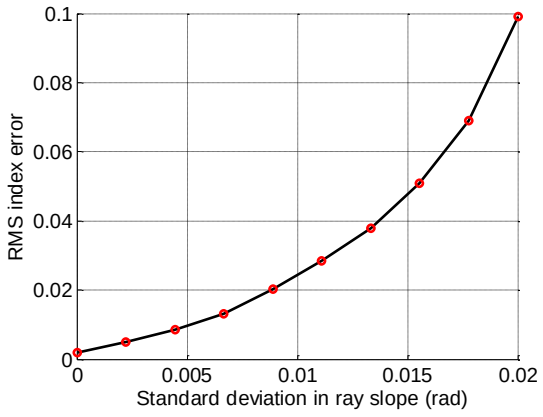


Figure 3.7.2. Reconstruction error due to contaminated boundary values of ray slope.

For the purposes of modeling the measurements taken in an actual experiment, the boundary ray slopes in the data were calculated from ray position measurements taken in two parallel planes separated by one arbitrary unit in length, as shown in Figure 3.7.3, where the arbitrary unit is common to all spatial quantities. In order to isolate the contamination of boundary ray slopes from the uncertainty in boundary ray positions for the trial simulations shown in Figure 3.7.2, x_1 in the figure was shifted to coincide with the exit surface of the GRIN element while maintaining the unit distance separation between x_1 and x_2. Subsequently, the white Gaussian noise was applied exclusively to the ray position measurements at x_2. The described geometry conforms to the definition of a radian for small exit angles relative to the x-axis, which was used as the unit of measure in Figure 3.7.2. In practice, uncertainty in ray slope is a consequence of measurement error in the ray positions in both planes. The calculated ray slopes are then extrapolated from x_1 to the output surface of the GRIN element to determine y_o, whose measurement error scales directly with the extrapolation distance. The values for y_i are obtained in the same manner.

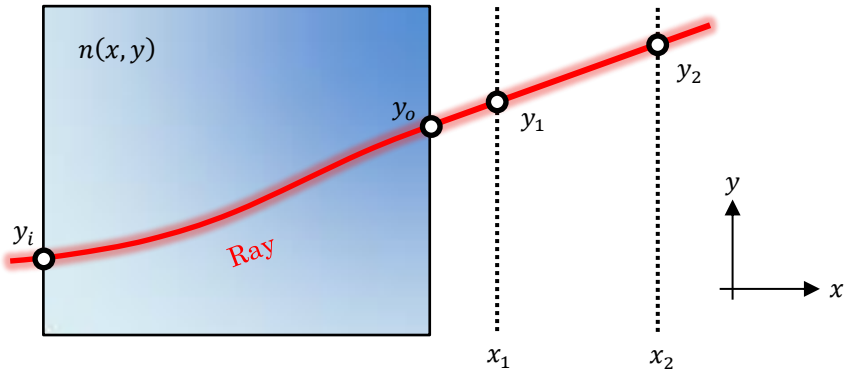


Figure 3.7.3. Geometry for calculating external boundary ray slopes from position measurements.

A similar test was performed to isolate the reconstruction error due to contaminated boundary ray positions from the effects of contaminated ray slopes, which we show in Figure 3.7.4. This was achieved by first calculating boundary ray slopes from noise-free ray positions at the measurement planes shown in Figure 3.7.3 and then applying white Gaussian noise to the position measurements at x_1. Each data point in the figure represents an ensemble average over 100 trials of the simulation. As before, we shifted the near plane (x_1) to coincide with the exit surface of the GRIN element to avoid the arbitrary scaling of measurement error in y_i and y_o during the extrapolation process. Unlike boundary ray slope values, the noise associated with boundary ray positions does not have any impact on $\vec{P}$ in Equation 3.3.13; only the coefficients in $[S]$ are affected in the system equation. Consequently, its contribution to the reconstruction error is much smaller than the error resulting from contaminated ray slopes. This is evident in the

comparison between Figures 3.7.2 & 3.7.4, where the reconstruction error in the latter figure remains close to the RMS index error observed in Section 3.5.

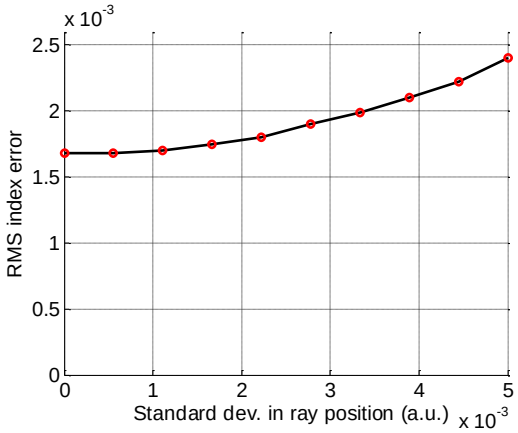


Figure 3.7.4. Error in reconstructed index field due to contaminated boundary values of ray position.

Combining the test results we have obtained thus far, Figure 3.7.5 illustrates the regimes in measurement noise where one of the three identified error mechanisms is dominant in contributing to the overall reconstruction error. To summarize, the quantization effects associated with the interpolation scheme used in discretizing the deflectometry path integrals are dominant at extremely low levels of measurement noise. In theory, this contribution can be reduced to arbitrarily low values by increasing the grid resolution used in the discrete representations of the field quantities, with the caveat of introducing additional DoFs to the system. At moderate noise levels, the base-line inversion sensitivity of the linear system is the primary contributor, whose linear scaling factor varies inversely with the amount of data redundancy used in the reconstruction. At extremely high measurement noise levels, the super-linear contribution from the path dependence of $[S]$ becomes much more significant.

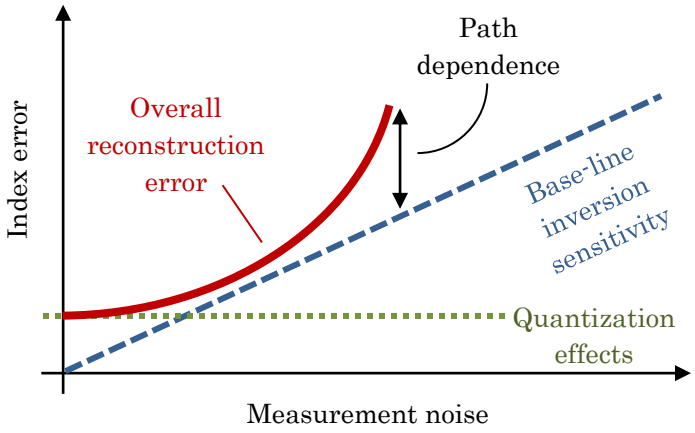


Figure 3.7.5. Regimes of primary error sources in the reconstruction of the index field.

In order to gain a better understanding of the LSQR method's limitations in inverting the primary system equation, it is useful to examine the numerical aspects of the inversion process through its conditioning. We can quantify the system's conditioning by, for example, dividing the variance of $\vec{\vartheta}$ by the variance of the input noise $\vec{\epsilon}$ in Equation 3.7.3. We have already shown in the comparison between the simulation data sets in Figure 3.7.1 that the use of data redundancy improves the conditioning of the system matrix. A more thorough characterization of the impact on reconstruction accuracy due to redundant data is shown in Figure 3.7.6, where we used varying amounts of interrogating rays to reconstruct the index field without introducing artificial noise to the simulated deflectometry data.

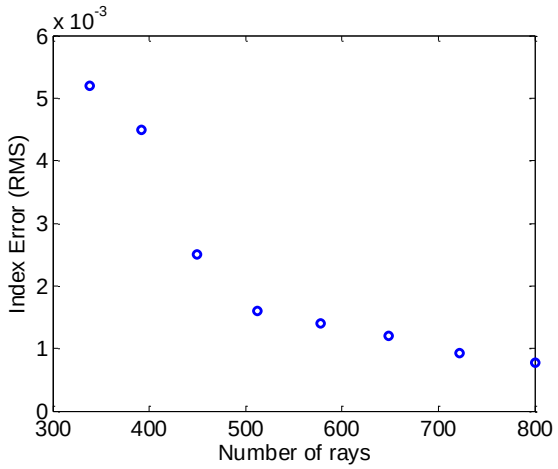


Figure 3.7.6. Reduced reconstruction accuracy resulting from reduced data redundancy.

Another important factor contributing to the deflectometry system's conditioning is the angular coverage of interrogating rays. The impact on reconstruction accuracy due to a reduction in angular coverage is shown in Figure 3.7.7, where the interrogating rays whose internal angles at the boundaries exceed a threshold value were omitted in the reconstruction process. Once again, no artificial noise was introduced to the deflectometry data obtained from the initial ray trace. The result explains the high degree of accuracy we were able to achieve in Section 3.5 using the chosen family of interrogating rays. There are no improvements beyond 45° because this is the maximum angle allowed in the interrogation scheme due to the geometry of the GRIN element. The relationship between angular coverage and the deflectometry system's conditioning holds implications for the potential reconstruction accuracy one can achieve in interrogating a GRIN element based on its geometry. In particular, the large angles required for accurate reconstructions may be difficult to achieve in cases of large aspect ratios. Furthermore, one must keep in mind the restrictions imposed by total internal reflection at the boundaries of the GRIN element. Aside from the possibility of total internal reflection, GRIN elements that exhibit more complex geometries may also impose restrictions on the range of boundary ray slope values that are accessible through external measurements.

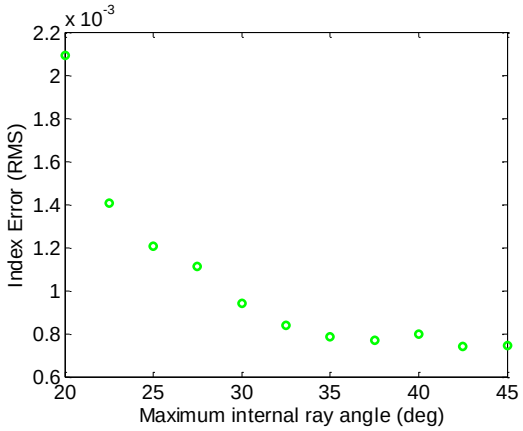


Figure 3.7.7. Reduced reconstruction accuracy due to limited angular coverage of interrogating rays.

The coverage of sample values representing the discrete partial derivative fields also plays a role in the conditioning of the deflectometry system. Recall that in generating the deflectometry data for a rectangular GRIN element, we employed two sets of entry and exit points uniformly distributed along opposite boundaries to establish the family of interrogating rays (see Section 3.5). Under this interrogation scheme, the spatial extent spanned by the entry or exit points along one boundary relative to the total length of the boundary provides a specific metric for the system's conditioning. In addition, this metric is a tunable parameter that we can vary to test its impact on the system's conditioning. We defined this metric as the *coverage* quantity labeled in Figure 3.7.8 divided by the *boundary length*.

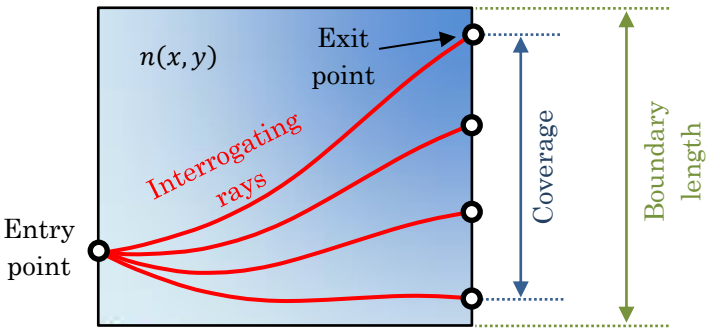


Figure 3.7.8. Definition of coverage for interrogating rays propagating between opposite boundaries.

We illustrate the impact of coverage on reconstruction accuracy in Figure 3.7.9. As before, no artificial noise was introduced to the deflectometry data in these simulations. This particular test provided a more quantitative look at the distribution of sampling for the system's DoFs provided by the chosen family of interrogating rays, whereas before we simply assumed that the interrogating rays covered all grid points in the discrete partial derivative fields.

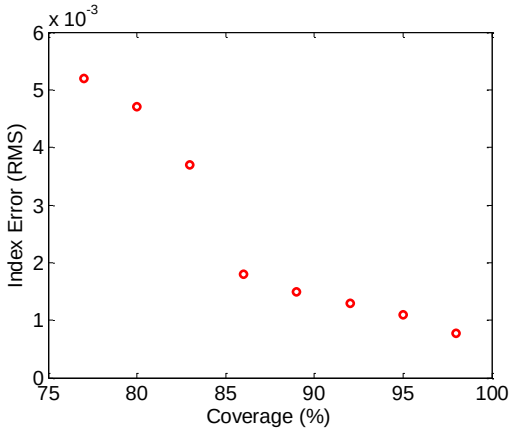


Figure 3.7.9. Reduced reconstruction accuracy due to limited spatial coverage spanned by entry and exit points of interrogating rays.

It is no surprised that Figures 3.7.6, 3.7.7, & 3.7.9 all indicate threshold values along the horizontal axis below which a significant increase in reconstruction error occurs. This increase can be attributed to the numerical instability of the LSQR method when it is presented with an extremely ill-conditioned system. We hasten to add that the characterization plots shown in this section are specific to the test index field shown in Figure 3.5.3. In general, the system's sensitivity to errors in the measured deflectometry data will depend on the complexity of the index field under investigation. To provide a specific example, an index distribution whose gradient field is aligned predominantly in one direction does not require broad angular coverage, provided that the direction of propagation for the interrogating rays is predominantly normal to the overall gradient.

3.8 Inversion from External Ray Slopes

In Section 3.5, we used prior knowledge of the surface refractive index to calculate the internal boundary ray slopes for each interrogating ray. In this section, we will address the more general scenario where the surface index is not available from measurements. In such cases, the internal boundary ray slopes cannot be determined using Snell's law in step 3 of the retrieval algorithm outlined in Section 3.5. Our formulation of the inverse problem then becomes a chicken-and-egg problem, where determining the internal boundary ray slopes require knowledge of the surface index and recovering the surface index requires knowledge of the internal boundary ray slopes.

In order to ascertain the internal boundary ray slope, we employ an iterative approach similar to the principles used in Section 3.7. Our strategy is to continually refine the calculated surface index using tentative estimates of the index field. Each inversion cycle establishes a set of surface index values based on internal boundary ray slopes computed from the most recent reconstruction. Starting with an assumed index distribution as our tentative solution (*e.g.*, a homogeneous index field), the surface index is used to

compute the internal boundary ray slopes from their external values via Snell's law, and a tentative description of the deflectometry system is established based on these boundary values. The resulting system equation is then inverted to obtain a new solution with presumably more accurate refractive index values along the boundaries. During the next inversion cycle, we recalculate the internal ray slopes based on the newly calculated index field and repeat the inversion process. In this manner, a 2-D deflectometry system can be inverted directly from the external ray slopes measured in the ambient medium to recover the index field. We outline the steps needed to implement our proposed iterative approach in the ensuing discussion.

Our ultimate goal in retrieving the index field directly from external ray slopes is to obtain a solution whose surface index is consistent with the internal and external ray slopes. Specifically, application of Snell's law to the external ray slopes using the surface index of the solution should yield the internal ray slopes that correspond to the ray trajectories used for constructing $[S]$ in the primary system equation. Furthermore, these ray trajectories should be consistent with the calculated index field, *i.e.*, they obey the ray equation (see Section 3.6). The challenge in modeling such a system is incorporating the phenomenon of surface refraction into the system of equations established in step 6 of the retrieval algorithm from Section 3.5. This is not straightforward because the refractive index appears explicitly in Snell's law, whereas the linear algebraic equations inferred from the deflectometry path integral are cast in terms of its partial derivatives. To reconcile this difference, we propose to differentiate the refractive index along the surface of the tentative solution. This will generate a set of conditions that specify the partial derivatives along the surface in the next inversion cycle. These conditions can be expressed as linear algebraic equations in the discrete values of the partial derivative fields and are used to augment the primary system equation, similar to the application of irrotational constraints described in Section 3.3.

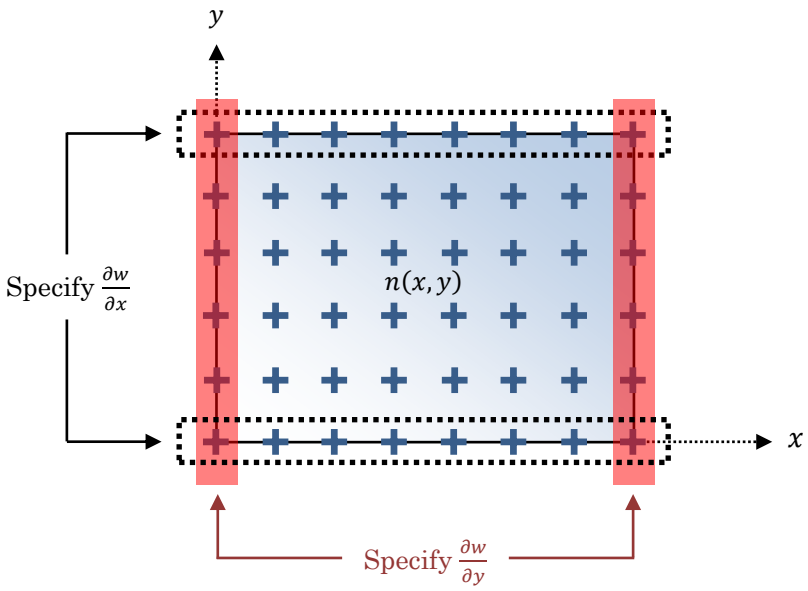


Figure 3.8.1. Constrained values for partial derivative sample points generated from boundary index values.

For instance, the partial derivatives along the top and bottom boundaries of the GRIN element shown in Figure 3.8.1 can be expressed in the form

$$f_{l,k} = \frac{\partial w}{\partial x}(x,y)\Big|_{y=y_{max}}, \tag{3.8.1a}$$

$$f_{l,k} = \frac{\partial w}{\partial x}(x,y)\Big|_{y=y_{min}}, \tag{3.8.1b}$$

respectively, where $f_{l,k}$ is the nomenclature used for the discrete sample values of $\frac{\partial w}{\partial x}$ from Section 3.3. Similarly, partial derivative constraints along the left and right boundaries in the figure can be expressed as

$$g_{l,k} = \frac{\partial w}{\partial y}(x,y)\Big|_{x=x_{min}}, \tag{3.8.1a}$$

$$g_{l,k} = \frac{\partial w}{\partial y}(x,y)\Big|_{x=x_{max}}, \tag{3.8.1b}$$

where $g_{l,k}$ refers to the discrete sample values of $\frac{\partial w}{\partial y}$.

It is important to keep in mind that the inversion of an overdetermined system produces a least squares solution. Consequently, the sample point corresponding to a particular DoF specified by the above conditions may take on a different value in the current inversion cycle. For this reason, we also refer to the expressions in Equations 3.8.1 & 3.8.2 as *soft* constraints. On the other hand, if we wish to introduce *rigid* constraints for the partial derivative fields, the associated DoFs must be eliminated from the system equation entirely, *i.e.*, they must be removed from the $\vec{\delta}$ vector and their contributions to the path integrals as well as the augmented equations must be incorporated into the right-hand side of the primary system equation. Rigid contraints are useful in cases where the surface index is measured directly and we seek a solution that is in total agreement with the measured values, for instance.

The purpose of the soft constraints is to provide an indirect connection between Snell's law and discrete partial derivative values at the surface. Preliminary tests have shown that the augmented this connection is crucial for facilitating convergence in our iterative approach; without it, it is possible to obtain an inconsistent solution (from the retrieval algorithm outlined in Section 3.5) whose surface index does not comploy with Snell's law, and yet its index gradient is perfectly capable of reproducing the angular deflection values specified by the deflectometry data. Because the conditions in Equation 3.8.1 are based on assumptions that are likely to change as we continue to refine our solution, we refer to the overall system equation employing these conditions as its *dynamic* variant, denoted by the subscript *dyn* in the expression

$$[S]_{dyn}\vec{\delta} = \vec{P}_{dyn}\,. \tag{3.8.3}$$

We emphasize that, with the exception of the rows corresponding to the augmented irrotational constraints, all coefficients in $[S]_{dyn}$ and values in $\vec{P}_{dyn}$ are derived from tentative internal boundary ray slopes and are hence subject to change from one inversion cycle to the next. In order to distinguish Equation 3.8.3 from Equation 3.3.13, we will refer to the latter as the *regular* deflectometry system equation.

Since the actual internal boundary ray slopes are unknown and their assumed values obtained from tentative reconstructions are (presumably) incorrect, the angular deflection values in $\vec{P}_{dyn}$ in Equation 3.8.3 are prone to error. As a result, inconsistencies are inadvertently introduced into the deflectometry system equation. Provided that the dynamic system equation is overdetermined, the least squares solution generated in the numerical inversion automatically optimizes the discrepancy between the deflection values specified in $\vec{P}_{dyn}$ and those obtained from evaluating the deflectometry system equation

$$[S]_{dyn}\vec{\delta}_{sol} = \vec{P}_{sol}\ , \tag{3.8.4}$$

where $\vec{\delta}_{sol}$ is the tentative solution to $\vec{\delta}$ in Equation 3.8.3 for the current inversion cycle. As usual, the partial derivative fields are extracted from $\vec{\delta}_{sol}$ and integrated (see Section 3.5) to reconstruct the index field. In our new inversion procedure, however, the integration constant is chosen to minimize the discrepancy between the surface index of the tentative solution and that of the calculated solution. The calculated solution is then used to compute a new set of internal ray slopes from the external data during the next inversion cycle. We emphasize that the newly calculated internal ray slopes will not be in agreement with the ray trajectories used to construct the dynamic system in the current inversion cycle unless convergence in the algorithm has already been achieved. Hence, $[S]_{dyn}$ and $\vec{P}_{dyn}$ must be recalculated based on a new set of internal boundary values established during the next inversion cycle. This process is repeated until the tentative solutions obtained from successive reconstructions are in agreement.

In order to monitor convergence of the algorithm outlined above, it is useful to define a metric based on the RMS difference between internal ray slopes computed prior to and after system inversion during each inversion cycle:

$$\nu = \sqrt{\frac{1}{2N}\sum_{i=1}^{2N}\left(\left[\overrightarrow{y'}_{initial}\right]_i - \left[\overrightarrow{y'}_{final}\right]_i\right)^2}\ , \tag{3.8.5}$$

where subscript i denotes the i^{th} boundary ray slope and N is the number of interrogating rays. There are $2N$ terms in the summation because each ray is associated with an entry and exit slope. $\overrightarrow{y'}_{initial}$ consists of internal ray slopes computed from the tentative solution obtained in the previous inversion cycle (or the initially assumed index field for the first inversion cycle) and $\overrightarrow{y'}_{final}$ refers to the respective slopes computed from the recovered index field after solving the dynamic system equation in the current inversion cycle.

The overall inversion procedure for reconstructing the index field from external ray slopes is outlined below. In the general case, we start by assuming a homogeneous field of bulk index value n_{bulk} consistent with available *a priori* knowledge of the optical medium as our initial guess. This assumption allows us to establish an initial set of approximate internal ray slopes for constructing the dynamic system equation.

1. Start with a homogeneous medium as the tentative solution and pick a bulk index value n_{bulk}.
2. Compute the internal ray slopes from the externally measured values using the surface index of the tentative solution via Snell's law.
3. Construct the regular deflectometry system equation as outlined in Section 3.5.
4. Compute the partial derivatives along the surface from the tentative solution and apply conditions to the regular deflectometry system to construct the dynamic system equation.
5. Invert the dynamic system equation to solve for ∇w and integrate to compute n. Update the tentative solution.
6. Repeat steps 2-5 until the metric in Equation 3.8.5 converges.

Generally speaking, using this procedure to solve the dynamic system equation will yield a solution that differs from the actual index field. Although the conditions specifying the partial derivatives along the surface of the solution are useful in guiding the algorithm to the general area of the solution space, they contain inherent inconsistencies because they are derived from tentative estimates of the refractive index. If no additional measures are taken, these inconsistencies will typically persist and propagate through all subsequent reconstructions of the index field. Through additional testing, removing these conditions has been shown to be useful in further refining the solution. This can be attributed to the fact that the inversion of the deflectometry system is relatively insensitive to disturbances in the surface index when compared to the influence exerted by partial derivative values specified by these conditions. This is because the partial derivatives of the refractive index appear explicitly in the deflectometry path integral (and are directly responsible for the deflection of interrogating rays), whereas the surface index itself does not (and only pertains to the phenomenon of surface refraction. Therefore, upon reaching convergence in the above procedure, we proceed as follows:

7. Repeat steps 2 and 3. Invert the regular system equation to solve for ∇w and integrate to compute n. Update the tentative solution.
8. Repeat steps 2-7 until successive reconstructions obtained in step 7 are in agreement.
9. Correct approximate ray trajectories by ray tracing through the calculated solution until trajectories are consistent with $n(x, y)$.

It may be useful to interpret the inconsistencies embedded in these conditions as local minima in an optimization routine. Removing these conditions in step 7 frees the solution from the local minima and allows the algorithm to seek the true global minimum.

There remains one more issue that requires our attention before we can demonstrate our new inversion procedure in a computer-based simulation. Preliminary tests have shown that achieving convergence in our proposed algorithm requires significantly better conditioning than that required when the internal boundary ray slopes are known. In fact, the family of interrogating rays used to demonstrate our reconstruction procedure in Section 3.5 will not suffice for solving the dynamic system equation reliably. The additional degrees of freedom associated with the unknown internal boundary ray slopes permit tentative solutions that do not lead to convergence in the retrieval algorithm. An increase in the angular coverage of interrogating rays is needed in order to eliminate these extraneous degrees of freedom. To achieve this, we employ a new measurement scheme in which a fan of rays is launched at entry points along the surface to interrogate the index field. The measurement geometry for a single entry point is illustrated in Figure 3.8.2, where the rays emanating from the entry point are uniformly separated in launch angle and span the forward half-plane. Any interrogating ray whose boundary slope exceeds the critical angle for total internal reflection cannot be measured externally and hence cannot be used in inverting the deflectometry system. We repeat the depicted interrogation for an entire set of entry points uniformly distributed across the entire surface (all four sides) of the rectangular GRIN element. The deflectometry data associated with these measurements, consisting of boundary ray positions and external ray slopes, are then used to establish a sufficiently conditioned dynamic system equation.

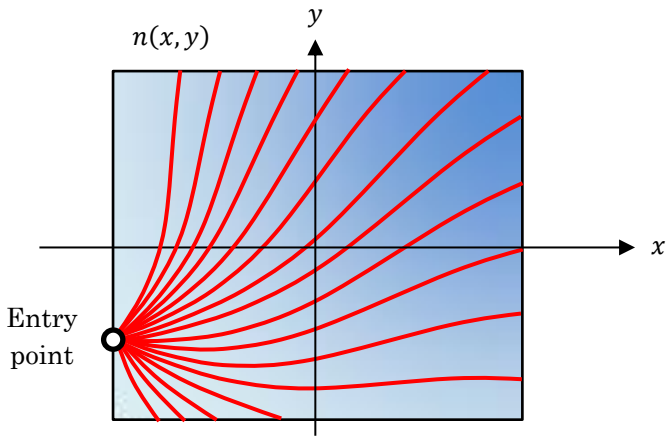
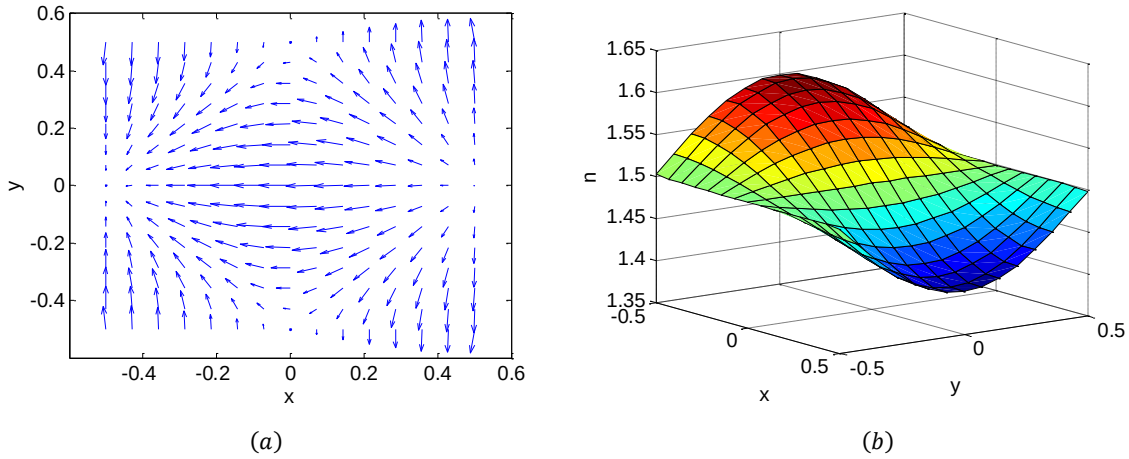


Figure 3.8.2. Interrogation geometry for increasing angular coverage of interrogating rays.

We are now ready to demonstrate the proposed inversion procedure by reconstructing the test index field shown in Figure 3.5.3 directly from computer-generated external boundary value data. In the simulation, the traced deflectometry data encompassed 160×35 (entry locations $\times$ launch angles) rays used to interrogate the GRIN element. For a direct comparison with the reconstruction error observed in Section 3.5, we defined the partial derivative fields on the same rectangular grid (15×15). Furthermore, the regular deflectometry system equation was first constructed assuming knowledge of the internal boundary ray slopes so that its inversion could be used to establish a benchmark value in the reconstruction

accuracy of the new measurement geometry. The resulting gradient and index fields obtained from the inversion are shown in Figure 3.8.3, with an RMS error of 7.29×10^{-4} RIU in the latter. A comparison of Figures 3.8.3a & 3.5.6a shows a substantial increase in the reconstruction accuracy of the gradient field in the corner regions of the GRIN element for the new measurement scheme. This result can be attributed to the improved conditioning of the regular system equation constructed from the new set of deflectometry data.



(a) (b)

Figure 3.8.3. (a) Vector gradient field ∇w and (b) index field $n(x, y)$ reconstructed from internal ray slopes for benchmarking reconstruction accuracy.

Following the benchmark reconstruction, the internal ray slopes were discarded along with the index values along the boundaries. Starting with the initial value $n_{bulk} = 1.5$ RIU in step 1 and restricting ourselves to the nested loop implied in steps 2-6 of the new inversion procedure, we depict several tentative reconstructions of the index field in Figure 3.8.4. Figure 3.8.4d shows the earliest tentative solution to satisfy the exit condition for step 6, where it is apparent that the solution still exhibits a noisy index profile along the boundaries of the GRIN element due to the inherent inconsistencies embedded in the partial derivative values specified along the surface in the dynamic system equation. Using the surface index in from this particular solution to construct a regular deflectometry system equation in step 7, an improved estimate of the solution is shown in Figure 3.8.5a, where the reduction of noise in the refractive index from Figure 3.8.4c is apparent. Similar reconstructions obtained from subsequent iterations of step 7 show comparable improvements in the calculated index field, as can be see in Figures 3.8.5b & 3.8.5c.

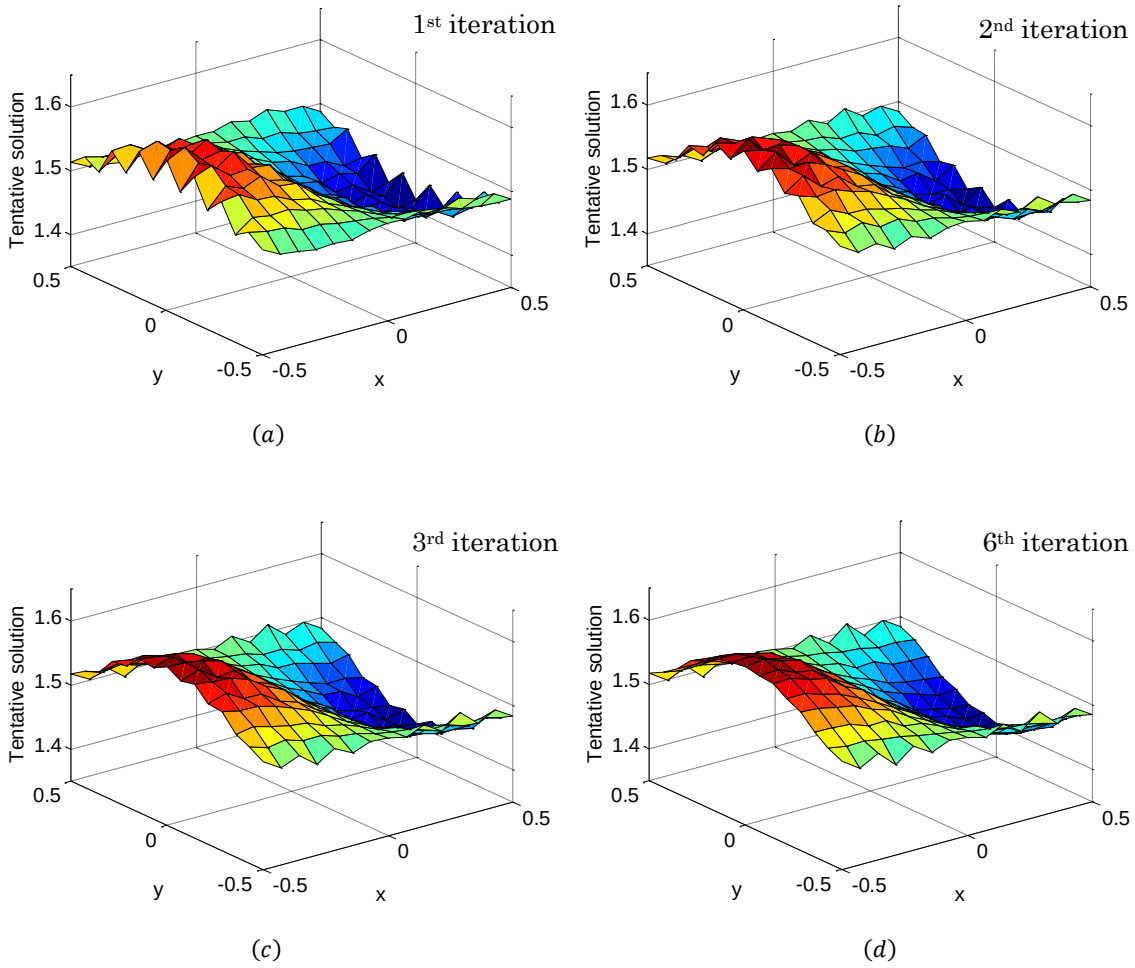


Figure 3.8.4. Tentative solutions obtained from following steps 1-6 in the overall procedure.

Figure 3.8.5*c* depicts the final result after three iterations of step 7, with an RMS index error of 1.80×10^{-3} RIU relative to the test index field from Figure 3.5.3 (after it was redefined on the same 15×15 rectangular grid). The increase in reconstruction error from its benchmark value is a consequence of the minor errors that persist in the calculated internal boundary ray slopes, whose contribution to the overall reconstruction error was discussed in detail in Section 3.7 (refer to Figure 3.7.2 in particular). We emphasize that the final values of the internal boundary ray slopes are the culmination of the algorithm's optimization path inside the solution space. This is responsible for the existence of local minima within the immediate vicinity of the global minimum. We hasten to add that excessive iterations inside the nested loop implied by step 6 as well as the overall loop implied by step 8 can lead to divergent scenarios due to error accumulation. Such scenarios are easily identified through the convergence metric in the iterative process and can be avoided by relaxing the exit conditions in step 6 and step 8.

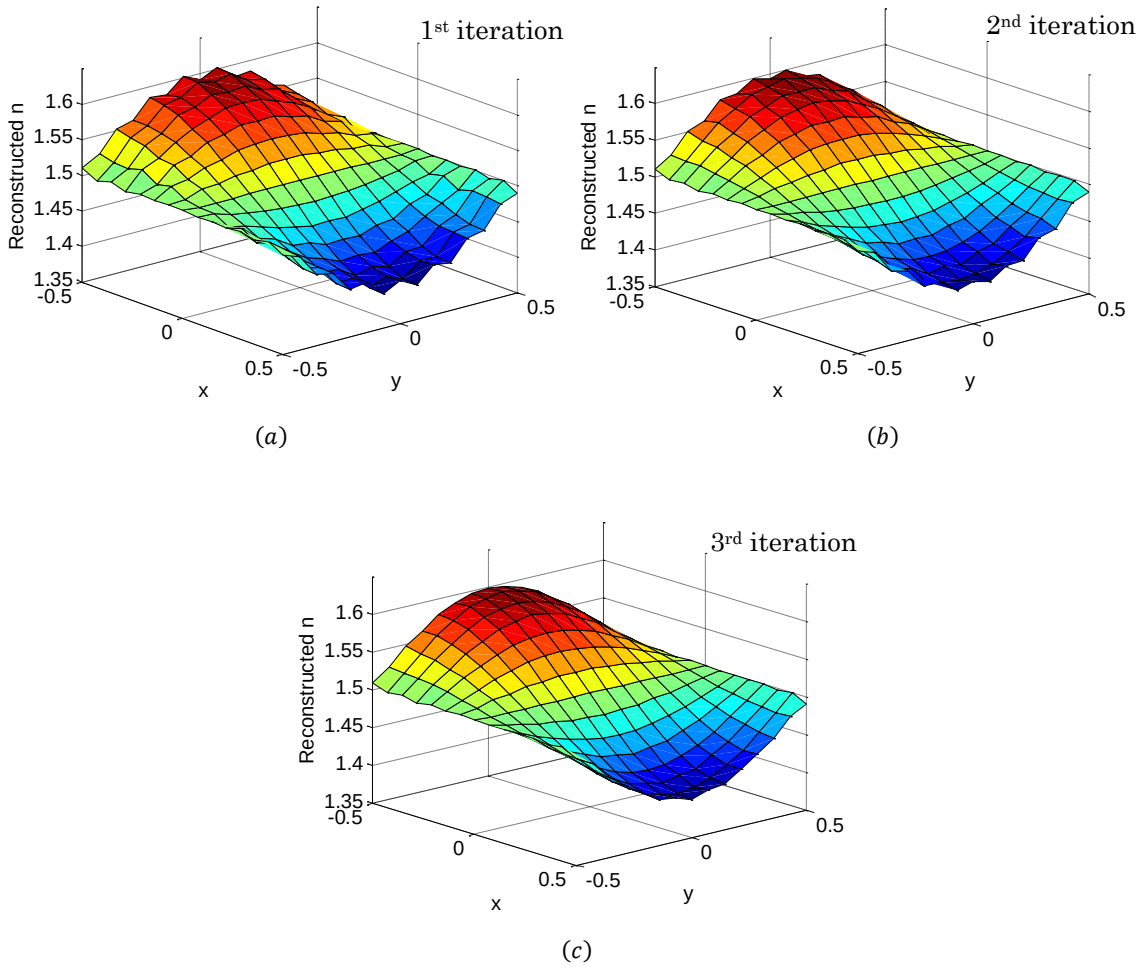


Figure 3.8.5. Tentative solutions obtained during step 7 at various stages in the overall procedure. The iteration count strictly refers to inversions of the regular deflectometry system equation in step 7.

In the previous example, the initially assumed value of n_{bulk} in step 1 was conveninently selected to be the median value of the actual result. Additional tests have shown that this initial value drives the recovery algorithm towards a particular solution. For instance, Figure 3.8.6 illustrates an alternate solution obtained from the same procedure starting with $n_{bulk} = 1.9$ RIU in step 1. Although the reconstruction in the figure does not appear as smooth as the solution shown in Figure 3.8.5*c*, the index field is equally valid in solving the dynamic system equation. We used the convergence metric defined in Equation 3.8.5 and the system discrepancy vector defined in Equation 3.5.1 to evaluate the solution's consistency and the respective quantities were roughly on the same order for the reconstructions shown in Figures 3.8.5*c* & 3.8.6. In practice, this difference cannot be used to select an optimal solution when the deflectometry data contains measurement noise.

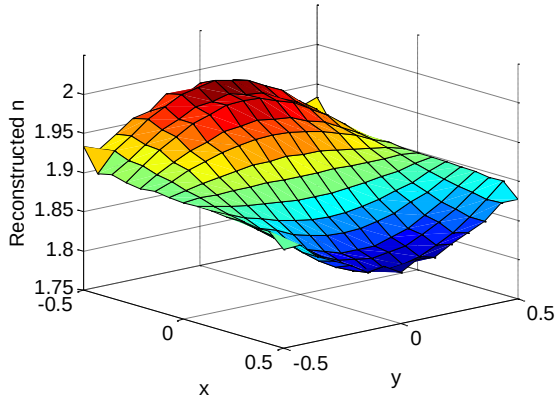


Figure 3.8.6. Optimal index field reconstructed from external boundary values after changing the initial assumption in step 1 to $n_{bulk} = 1.9$ *RIU.*

Interestingly, a comparison between Figures 3.8.5*c* & 3.8.6 reveals that in cases where the internal boundary values are not available from measurement, the gradient field ∇w is no longer uniquely specified by the deflectometry data due to the additional degrees of freedom associated with unknown boundary values. This implies that without additional information, multiple variations of the dynamic system equation can result from the external deflectometry data, each of which has a self-consistent solution. As with the inversion of the regular deflectometry system equation, the solution becomes fully specified (both the gradient field and the index field) with additional constraints such as knowledge of the index at a single position, or knowledge of the optical path length for a single interrogating ray. Under this scenario, the numerical integration in steps 5 & 7 in the inversion procedure can be modified to satisfy the additional constraints.

3.9 Extension to Three Dimensions

In this section, we show that the principles developed in Section 3.3 are fully generalizable to deflectometry path integrals for 3-D index distributions. We note that it is outside the scope of this dissertation to describe a 3-D deflectometry system in its entirety or demonstrate its numerical implementation, so a significant portion of the mathematical rigor will be overlooked.

Assuming the trajectories inside the index distribution are of the form $\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} y(x) \\ z(x) \end{bmatrix}$, Equation 3.3.1 can be expressed in Cartesian coordinates as follows. The first derivative with respect to arc length variable s in Cartesian coordinates is

$$\frac{d}{ds} = \frac{dx}{ds}\frac{d}{dx} = f\frac{d}{dx}, \tag{3.9.1}$$

where $f = \frac{dx}{ds} = \frac{dx}{\sqrt{dx^2+dy^2+dz^2}} = \frac{1}{\sqrt{1+\left(\frac{dy}{dx}\right)^2+\left(\frac{dz}{dx}\right)^2}}$ using the fact that $ds^2 = dx^2 + dy^2 + dz^2$ in three dimensions. The second derivative is then

$$\frac{d^2}{ds^2} = f\frac{d}{dx}\left(f\frac{d}{dx}\right) = f\frac{df}{dx}\frac{d}{dx} + f^2\frac{d^2}{dx^2}. \tag{3.9.2}$$

Rewriting Equation 3.9.1 in scalar form for each Cartesian vector component, we have

$$\frac{dn}{ds}\frac{dx}{ds} + n\frac{d^2x}{ds^2} = \frac{\partial n}{\partial x}, \tag{3.9.3a}$$

$$\frac{dn}{ds}\frac{dy}{ds} + n\frac{d^2y}{ds^2} = \frac{\partial n}{\partial y}. \tag{3.9.3b}$$

$$\frac{dn}{ds}\frac{dz}{ds} + n\frac{d^2z}{ds^2} = \frac{\partial n}{\partial z}. \tag{3.9.3c}$$

With the help of Equations 3.9.1 & 3.9.2, Equation 3.9.3a becomes

$$\left(f\frac{dn}{dx}\right)\left(f\frac{dx}{dx}\right) + n\left(f\frac{df}{dx}\frac{dx}{dx} + f^2\frac{d^2x}{dx^2}\right) = \frac{\partial n}{\partial x},$$

$$f^2\frac{dn}{dx} = \frac{\partial n}{\partial x} - nf\frac{df}{dx}, \tag{3.9.4}$$

where we have used $\frac{dx}{dx} = 1$ and $\frac{d^2x}{dx^2} = 0$. Similarly, Equation 3.9.3b becomes

$$\left(f\frac{dn}{dx}\right)\left(f\frac{dy}{dx}\right) + n\left(f\frac{df}{dx}\frac{dy}{dx} + f^2\frac{d^2y}{dx^2}\right) = \frac{\partial n}{\partial y},$$

$$\frac{dy}{dx}f^2\frac{dn}{dx} + n\left(f\frac{df}{dx}\frac{dy}{dx} + f^2\frac{d^2y}{dx^2}\right) = \frac{\partial n}{\partial y}. \tag{3.9.5}$$

Substituting Equation 3.9.4 for $f^2\frac{dn}{dx}$ in Equation 3.9.5, canceling terms, and dividing by n yields

$$\frac{dy}{dx}\frac{\partial n}{\partial x} - \frac{dy}{dx}nf\frac{df}{dx} + nf\frac{df}{dx}\frac{dy}{dx} + nf^2\frac{d^2y}{dx^2} = \frac{\partial n}{\partial y},$$

$$\frac{\partial w}{\partial y} - \frac{\partial w}{\partial x}y' = \frac{y''}{1+(y')^2+(z')^2}, \tag{3.9.6}$$

where $w = \ln(n)$ is the logarithmic refractive index. Similarly, Equation 3.9.3c can be written as

$$\left(f\frac{dn}{dx}\right)\left(f\frac{dz}{dx}\right) + n\left(f\frac{df}{dx}\frac{dz}{dx} + f^2\frac{d^2z}{dx^2}\right) = \frac{\partial n}{\partial z}. \tag{3.9.7}$$

Upon inspection, Equation 3.9.7 is completely analogous to Equation 3.9.5, with the spatial variable y replaced by z. Hence, we have

$$\frac{\partial w}{\partial y} - \frac{\partial w}{\partial x} z' = \frac{z''}{1+(y')^2+(z')^2} \,. \tag{3.9.8}$$

Paralleling our derivation in Section 3.3, we switch to relative coordinates to further simplify the expressions in Equations 3.9.6 & 3.9.8 and illustrate the fundamental relationship between the local index gradient and the deflection of the light ray. Defining the x-axis parallel to the ray vector at a particular location results in $y' = 0$ and $z' = 0$ for both equations. Consequently, the y and z components of the ray vector effectively become decoupled under these conditions, and the equations are reduced to

$$\frac{\partial w}{\partial y} = y'' \,, \tag{3.9.9a}$$

$$\frac{\partial w}{\partial z} = z'' \,. \tag{3.9.9b}$$

The definition of the y and z axes in our relative coordinate system are arbitrary provided that they are orthogonal to the x-axis and to each other. In other words, they can be rotated freely about the relative x-axis. In our treatment of 3-D deflectometry path integrals, we choose to use slopes in Cartesian coordinates for expressing directionality in lieu of angles; the latter typically involves more complex calculations that may lead to nonlinear equations in ∇w. Invoking the small angle approximation for slow varying index distributions, we have $\Delta x \approx \Delta s$ and rewrite Equation 3.9.9 in the differential form

$$\Delta y' \approx \frac{\partial w}{\partial y} \Delta s = \nabla w \cdot \hat{u} \, \Delta s \,, \tag{3.9.10a}$$

$$\Delta z' \approx \frac{\partial w}{\partial z} \Delta s = \nabla w \cdot \hat{v} \, \Delta s \,, \tag{3.9.10b}$$

where $\hat{u}$ and $\hat{v}$ are unit vectors pointing along the y- and z-axes in the relative coordinate system, respectively. Although the geometry of the inverse problem is more complex due to the increased dimensionality, the fundamental principles of deflectometry remain the same: The local index gradient projected along any direction perpendicular to the local ray vector is responsible for the ray's deflection in that direction, as is apparent from Equation 3.9.10. The problem can be further simplified, for instance, by defining the relative coordinate system such that the gradient vector is completely contained in the xy-plane, in which case $\frac{\partial w}{\partial z} = 0$ and Equation 3.9.10b vanishes.

As with 2-D index fields, the deflectometry path integral inside 3-D index distributions is more conveniently expressed in the laboratory coordinate system. To achieve this, we need to rotate the deflected ray vector (resulting from the differentials in Equation 3.9.10) back to laboratory coordinates in a manner that maintains linearity in ∇w. One possibility is to compute the position of the deflected ray in laboratory coordinates using a rotation matrix and then differentiate with respect to x to compute the change incurred in the y- and z-slopes in laboratory coordinates. The cumulative sum of these differential contributions to the deflection of the light ray over its entire trajectory then describes a deflectometry path integral for 3-D index distributions that is analogous to Equation 3.3.7 and is linear in ∇w.

One notable difference for 3-D index distributions is that the deflections in ray slope can be separated into their respective components along the y and z laboratory coordinates. Hence, each deflectometry path integral can be split into two individual scalar equations describing the accumulated deflections along two dissimilar vector components. As with the 2-D case, the index gradient ∇w is fully specified by the deflectometry data if the internal boundary ray slopes can be measured or calculated. The index distribution itself, on the other hand, is only specified up to an unknown constant which can be identified by knowledge of the index at some position. Because the existence of the underlying potential function $w(x, y, z)$ is implied, the partial derivative fields $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$ and $\frac{\partial w}{\partial z}$ must satisfy the irrotational constraint

$$\nabla \times \nabla w = \left(\frac{\partial}{\partial y}\frac{\partial w}{\partial z} - \frac{\partial}{\partial z}\frac{\partial w}{\partial y}\right)\hat{\imath} + \left(\frac{\partial}{\partial z}\frac{\partial w}{\partial x} - \frac{\partial}{\partial x}\frac{\partial w}{\partial z}\right)\hat{\jmath} + \left(\frac{\partial}{\partial x}\frac{\partial w}{\partial y} - \frac{\partial}{\partial y}\frac{\partial w}{\partial x}\right)\hat{k} = 0\ . \qquad (3.9.11)$$

In order to satisfy the above expression, the individual vector components must vanish simultaneously. Hence, the irrotational constraint applied to each point on the sampling grid provides three individual constraints that can be appended to the overall system as linear algebraic equations. The resulting deflectometry system equation retains the same form as Equation 3.3.13, where $[S]$ and $\vec{P}$ now contain $2N + 3(L-1)(K-1)(J-1)$ rows. Here, N is the number of measured path integrals and $L \times K \times J$ is equal to the number of points used to sample the partial derivative fields on a uniform 3-D rectangular grid. The top $2N$ rows in these quantities correspond to the Cartesian vector components of the deflectometry path integrals, whereas the remaining rows correspond to the vanishing vector components in Equation 3.9.11. Assuming that there are enough equations to invert the system, we can solve for the partial derivative fields in the usual manner. The partial derivatives can then be integrated, with knowledge of the integration constant, to obtain the 3-D logarithmic index distribution $w(x, y, z)$.

4. Analysis and Design of GRIN Resonators

4.1 Background Material

Temporal (or spectral) coherence and spatial coherence are properties often used to distinguish laser beams from ordinary light sources. Temporal coherence refers to a strong correlation between the amplitude and phase of the output electromagnetic field at some location with its amplitude and phase at earlier or later times, whereas spatial coherence refers to a high degree of correlation between the instantaneous phase angles of the wavefront at any two points across the output beam. Ideal single-mode oscillation inside the laser cavity will exhibit very desirable coherence properties in the output beam, both temporal and spatial; non-ideal multimode and multifrequency oscillation will cause the degree of correlation to deteriorate. In the non-ideal case, the output beam is often described as the summation of several independent oscillation frequencies and may have substantial random variations in the amplitude and frequency for each of the constituent oscillations. As a result, the superposition of the oscillation frequencies produces an output beam that exhibits characteristics resembling random noise according to various statistical and spectral measures. Thus, a high-fidelity output signal produced from single-mode oscillation inside the laser cavity is generally desired.

For a given laser cavity, one can always find a discrete set of transverse eigenmodes, or distinct amplitude and phase patterns for the circulating beam, which will reproduce themselves in form after one round trip. In a conventional Fabry-Perot resonator, the amplitude and phase patterns of these self-reproducing eigenmodes are solely determined by the specific curvature and shape of the mirrors at the ends of the laser cavity, and diffraction losses for a particular mode are caused by whatever apertures are present inside the cavity (which may be introduced artificially to model the finite extent of the end mirrors). If for any particular eigenmode, the gain medium is able to maintain sufficient amplification to overcome all losses in one round trip through the resonator, then that eigenmode will be one possible steady-state oscillation. Provided that the laser cavity exhibits different round-trip losses for the distinct eigenmodes, one simply needs to maintain the round-trip gain above the lasing threshold of the transverse spatial mode with the lowest round-trip losses (*i.e.*, its fundamental mode) but below the lasing thresholds of all competing modes to ensure single-mode operation. While this is not particularly difficult to achieve, the second lowest lasing threshold imposes a severe limitation on the amount of power contained in the single-mode oscillation output. Thus, the potential single-mode output power from a laser cavity is directly related to the difference between the two lowest lasing thresholds, *i.e.*, its modal discrimination. This difference arises from the varying tendencies of the transverse eigenmodes to diffract, or spread out in the transverse direction, as they propagate inside the resonator. The higher-order transverse modes generally have more complex transverse amplitude and phase variations. As a result, their transverse spread is typically larger, which makes their round-trip diffraction losses greater than those of the fundamental mode. An inherent

disadvantage to these Fabry-Perot resonators is that they require long cavity lengths and small-diameter modes as well as mode-selecting apertures in order to achieve adequate modal discrimination, resulting in limited interaction with the gain medium and thereby reducing the amount of power that can be extracted by the fundamental mode under single-mode oscillation conditions.

It is well known that the lower-order transverse modes in almost any type of resonator will exhibit an uncanny ability to shape their amplitude and phase patterns in ways that minimize their diffraction losses on each round trip. Several strategies have been described in the literature over the years for improving the modal properties of optical resonators [Oron, *et al.*, 2001]. For instance, many of the problems associated with Fabry-Perot configurations can be circumvented by the use of unstable resonator geometries [Siegman, 1986]; these configurations can support a large-diameter fundamental mode while simultaneously providing superior modal discrimination. However, the unstable geometry produces an inherently lossy fundamental mode and cannot be used in lasers systems with insufficient gain. To mitigate the high losses, variable reflectivity mirrors have been used in unstable configurations to help reduce the diffraction effects associated with hard edges in the reflected beam [Lavigne, *et al.*, 1985]. This idea can also be extended to graded-phase mirrors to shape the fundamental mode [Paré & Bélanger, 1992]. The same beam shaping principles can be applied by placing a spatial filter in the form of a grid of opaque strips inside the laser cavity [Kermene, *et al.*, 1992]. Simiarly, diffractive mode-selecting mirrors [Leger, *et al.*, 1995] and phase-conjugate mirrors [Yang & Leger, 2004] have also been shown to enhance discrimination against higher-order modes. In addition, the use of intracavity optical elements to enhance the modal properties of lasers has also been studied. For instance, the shaping of resonator eigenmodes can be achieved by including arbitrary assortments of lenses, or lens-like media, inside the laser cavity [Casperson, 1975]. This idea has since been extended to cascades of hard-edge apertures [Otis, *et al.*, 1979; Cagniot, *et al.*, 2008] and cascades of π-phase plates [Cagniot, *et al.*, 2010]. Intracavity circular hollow glass waveguides have been used to enhance transverse modal selection as well [Zhou, *et al.*, 1990]. Of course, combinations of the approaches mentioned above have also been explored. More recently, near-optimal modal properties have been described, where gain-guiding principles are applied to unstable refractive index profiles in fiber lasers [Siegman, 2007].

In this chapter, we will explore the application of intracavity GRIN elements for enhancing the modal properties of laser cavities and describe a method for computing the spatial beam profiles of the transverse eigenmodes produced in these resonator configurations. The underlying mathematics behind wave propagation in GRIN media is first developed in Section 4.2 based on the Fresnel-Kirchoff formulation. We then outline a straightforward implementation of our propagation method in Section 4.3 and show agreement between the propagated fields and known analytical solutions in various optical media. In Section 4.4, we resolve numerous practical issues associated with characterizing general GRIN media for wave propagation and demonstrate the diffraction effects of Gaussian beams that transit a Luneburg lens. We then proceed to compute the transverse eigenmodes of resonators employing intracavity GRIN elements in Section 4.5. In Section 4.6, we outline a general numerical procedure for calculating the

index profile of a GRIN mode converter based on our previous work [Lin & Leger, 2012]. Finally, a practical laser cavity design that employs Gaussian-to-flat-top mode conversion using a GRIN element is examined in Section 4.7.

4.2 The Fresnel-Kirchoff Integral

We will start by developing the mathematics for wave propagation in GRIN media following a Green's function approach [Lutomirski & Yura, 1971]. The derivation is analogous to the integral theorem of Helmholtz and Kirchoff used for calculating free space diffraction [Born & Wolf, 1965]. To begin, let us suppose that the GRIN medium occupies a volume V and is bounded by a closed surface S. Within the enclosed volume, let the Green's function $G(\vec{r},\vec{p})$ represent the field disturbance at $\vec{r}$ due to a point source at $\vec{p}$. As usual, G must satisfy the scalar wave equation

$$(\nabla^2 + k^2 n^2) G(\vec{r},\vec{p}) = -4\pi\delta(|\vec{r}-\vec{p}|)\ , \tag{4.2.1}$$

where $k = \frac{2\pi}{\lambda_0}$ is the vacuum wavenumber, n is the refractive index, and δ is the Dirac delta function. Next, let $\psi(\vec{r})$ be the field disturbance at $\vec{r}$ due to an incident wave, and let us assume that it has continuous first-order and second-order partial derivatives within and on the closed surface S. Naturally, ψ must satisfy the source-free scalar wave equation

$$(\nabla^2 + k^2 n^2)\psi(\vec{r}) = 0\ . \tag{4.2.2}$$

Multiplying Equation 4.2.1 by $\psi(\vec{r})$ and Equation 4.2.2 by $G(\vec{r},\vec{p})$, and then subtracting, we have

$$G(\vec{r},\vec{p})\nabla^2\psi(\vec{r}) - \psi(\vec{r})\nabla^2 G(\vec{r},\vec{p}) = 4\pi\delta(|\vec{r}-\vec{p}|)\psi(\vec{r})\ . \tag{4.2.3}$$

Integrating the above expression over the enclosed volume V and dividing by 4π yields

$$\psi(\vec{p}) = \frac{1}{4\pi}\int_V [G(\vec{r},\vec{p})\nabla^2\psi(\vec{r}) - \psi(\vec{r})\nabla^2 G(\vec{r},\vec{p})]dV\ . \tag{4.2.4}$$

Applying Green's theorem [Courant & Hilbert, 1991] allows us to express this volume integral over V as a surface integral over S,

$$\psi(\vec{p}) = \frac{1}{4\pi}\oint_S [\psi(\vec{r})\nabla G(\vec{r},\vec{p}) - G(\vec{r},\vec{p})\nabla\psi(\vec{r})]\cdot d\vec{S}\ , \tag{4.2.5}$$

where $d\vec{S}$ is normal to the enclosing surface S and directed into the enclosed volume V. We note that the integrand in Equation 4.2.5 is the negative of the integrand in Equation 4.2.4 due to our choice of direction for $d\vec{S}$. The significance of Equation 4.2.5 is self-evident because it enables us to compute the propagated wave ψ at $\vec{p}$ from boundary conditions along the enclosing surface S.

Let us now suppose that the enclosed volume is filled with a GRIN medium whose refractive index distribution is given by

$$n(\vec{r}) = n_0 + n_1(\vec{r}) \,, \tag{4.2.6}$$

where n_0 is some constant bulk value. Inside the medium, the Green's function $G(\vec{r},\vec{p})$ can be expressed as a perturbed spherical wavefront emanating from the point source at $\vec{p}$,

$$G(\vec{r},\vec{p}) \to G(s) = \frac{1}{s} e^{ikn_0 s + \varphi} \,, \tag{4.2.7}$$

where $s = |\vec{r} - \vec{p}|$ is the distance from the point source, and φ is the phase contribution due to the spatially-varying component of the refractive index. For instance, in a weakly refracting medium where rays travel in approximately linear trajectories, this phase contribution can be estimated as

$$\varphi \approx ik \int_{\vec{p}}^{\vec{r}} n_1 ds \,. \tag{4.2.8}$$

We note that the field amplitude of the spherical wavelet described in Equation 4.2.7 only depends on s, implying that the irradiance is constant everywhere on the wavefront. This only holds true under the assumption that rays travel in approximately straight lines inside the GRIN material. We will address the implications of curved ray trajectories in Section 4.3.

In evaluating the integral in Equation 4.2.5, we invoke the Kirchoff boundary conditions [Elmore & Heald, 1985] depicted in Figure 4.2.1, where the point $\vec{p}$ is surrounded by an opaque screen S except for an aperture A of finite extent. Along the opaque screen, we have the vanishing boundary condition $\psi = \nabla\psi \cdot d\vec{S} = 0$; within the aperture A, ψ and $\nabla\psi \cdot d\vec{S}$ have unperturbed values and behave as if the screen were not present. We treat the field disturbance at the aperture as a local planar wavefront and assume it has a partial derivative

$$\frac{\partial\psi}{\partial z} \approx ik\psi(\vec{r})cos\big(\theta_\psi\big) \tag{4.2.9}$$

with respect to z, where θ_ψ is the angle between the z-axis and the direction normal to the local wavefront specified by the phase profile of ψ at $\vec{r}$ as seen in Figure 4.2.1. Since $d\vec{S}$ points in the $+z$ direction at the aperture, we have

$$\nabla\psi(\vec{r}) \cdot d\vec{S} = ik\psi(\vec{r})cos\big(\theta_\psi\big)dS \,. \tag{4.2.10}$$

Continuing with the evaluation of Equation 4.2.5, let us consider the differential $\nabla G(\vec{r},\vec{p}) \cdot d\vec{S}$, which we rewrite as

$$\nabla G(\vec{r},\vec{p}) \cdot d\vec{S} = \frac{\partial G}{\partial z} dS = \frac{\partial G}{\partial s}\frac{ds}{dz} dS \,. \tag{4.2.11}$$

Substituting in G from Equation 4.2.7 yields

$$\frac{\partial G}{\partial s} = \frac{1}{s^2} e^{ikn_0 s+\varphi} \left(iks + \frac{\partial \varphi}{\partial s} s + 1 \right). \tag{4.2.12}$$

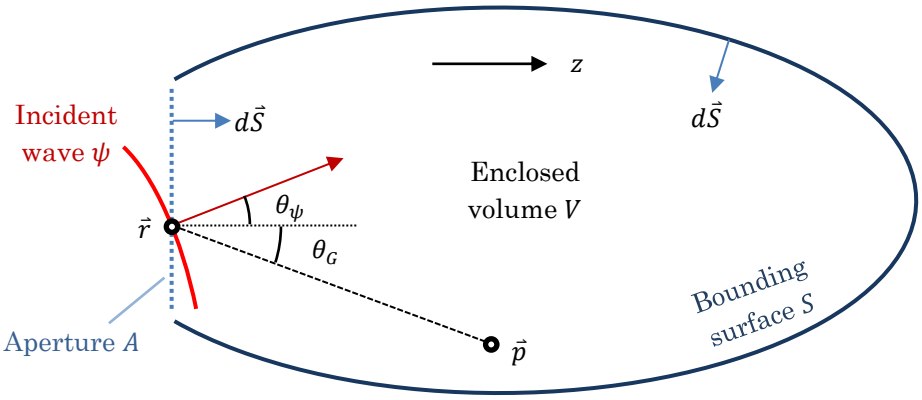


Figure 4.2.1.Geometry involved in the Kirchoff boundary conditions.

Assuming that $ks \gg 1$ at optical frequencies and that $\left|\frac{\partial \varphi}{\partial s}\right|$ is on the order of $|n_1|$, the iks term dominates the expression inside the parentheses, and Equation 4.2.11 simplifies to

$$\nabla G(\vec{r},\vec{p}) \cdot d\vec{S} \approx -\frac{ik}{s} e^{ikn_0 s+\varphi} cos(\theta_G)\,, \tag{4.2.13}$$

where we have used the fact that $\frac{ds}{dz} = -cos(\theta_G)$ in transit from $\vec{p}$ to $\vec{r}$. Finally, substituting Equations 4.2.10 & 4.2.13 into Equation 4.2.5, applying the Kirchoff boundary conditions and using Equation 4.2.7, we find that

$$\psi(\vec{p}) = -\frac{ik}{4\pi} \oint_A \psi(\vec{r})\, G(\vec{r},\vec{p}) \left[cos\left(\theta_\psi\right) + cos(\theta_G) \right] d^2\vec{r}\,, \tag{4.2.14}$$

where A denotes the evaluation over the region of S inside the aperture. It follows that $G(\vec{r},\vec{p})$ obeys reciprocity in a GRIN medium [Lutomirkski & Yura, 1971], hence the field and source points may be interchanged, *i.e.*, $G(\vec{r},\vec{p}) = G(\vec{p},\vec{r})$. With this modification to the expression, Equation 4.2.14 can be interpreted as an extension of the Huygens-Fresnel principle to wave propagation inside a GRIN medium: The field disturbance at $\vec{r}$ inside the aperture is treated as a secondary source of strength $\psi(\vec{r})$ and its contribution to the field disturbance at $\vec{p}$ is equal to $\psi(\vec{r})G(\vec{r},\vec{p})\left[cos\left(\theta_\psi\right) + cos(\theta_G)\right]d^2\vec{r}$; the overall field disturbance at $\vec{p}$ is then given by the superposition of all secondary wavelets reaching $\vec{p}$ from the aperture, where the wavelets in the GRIN medium are determined by the propagation of spherical wavefronts from

the secondary sources to point $\vec{p}$. Note that the Green's function in Equation 4.2.7 reduces to a spherical wavefront in the immediate vicinity of the secondary source, *i.e.*,

$$\lim_{s\to 0} G(s) = \frac{1}{s} e^{ikn_0 s}\,, \tag{4.2.15}$$

since the phase perturbation φ from Equation 4.2.8 vanishes.

Therefore, if we know how a spherical wave propagates in the GRIN medium, we can determine the propagated field from any arbitrary disturbance in the aperture. It is straightforward to show that when $n_0 = 1$ and $n_1 = 0$, *i.e.*, the medium is free space, $G(\vec{r}, \vec{p})$ is an ideal spherical wavefront everywhere in V, and Equation 4.2.13 reduces to the Fresnel-Kirchoff diffraction formula.

4.3 Wave Propagation in GRIN Media

In this section, we will outline a numerical procedure for computing the Green's function for a discrete set of secondary sources representing the input field distribution; we will then use the results of this characterization to compute the propagated field distribution in the output plane. In addition, we will demonstrate the efficacy of our propagation method by comparing the calculated field distributions with analytical results in free space as well as inside quadratic ducts.

Recall from Section 2.4 that the field description of an optical beam (or any incident electromagnetic wave) in a given plane can be converted into an equivalent ray distribution based on energy conservation laws. In the same manner, a secondary wavelet emitted by a secondary source can be represented as a fan of rays emanating from a point residing in the input plane, which we illustrate in Figure 4.3.1.

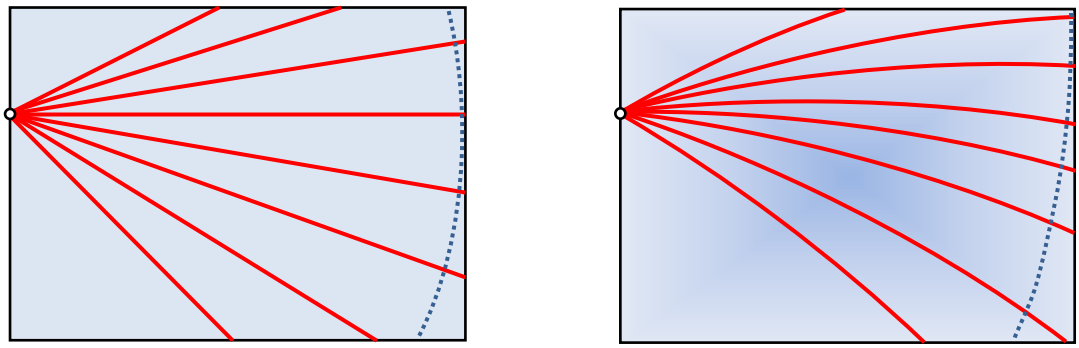


Figure 4.3.1. Secondary wavelet propagating through (a) a homogeneous medium and (b) a GRIN medium.

If the individual rays represent equal amounts of radiant energy, then the separation in angle is uniform amongst the constituent rays within the immediate vicinity of the secondary source. In a homogeneous

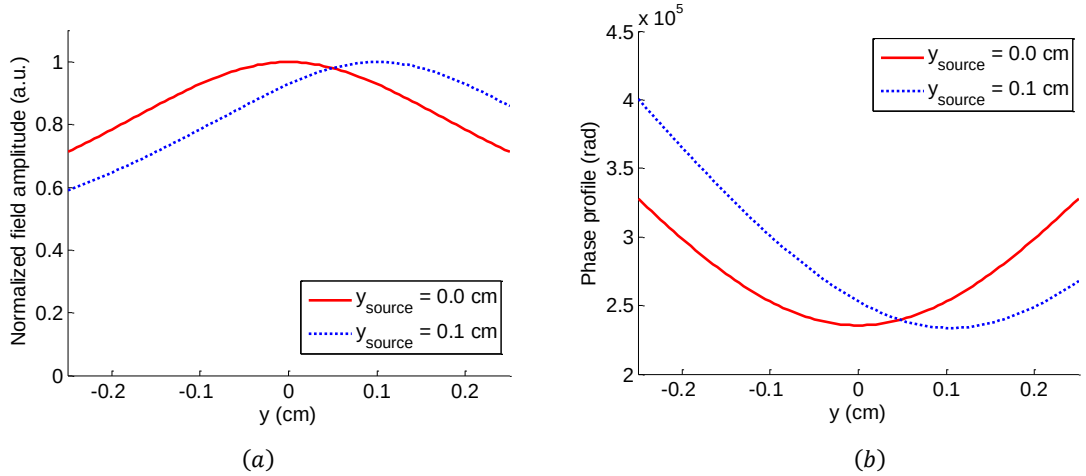
medium, the rays follow linear trajectories, and the secondary wavelet remains spherical (as is the case in the formulation of the Fresnel-Huygens principle). On the other hand, the rays follow curved trajectories inside a GRIN medium, as shown in Figure 4.3.1b, according to the ray equation. It is apparent that within the immediate vicinity of the secondary source where the propagation distance approaches zero, the secondary wavelets in both cases are spherical. However, as the wavelet propagates farther into the GRIN medium, the index distribution begins to redirect the direction of travel for the energy radiating outward, distorting the wavelet's irradiance profile as well as the shape of its wavefront.

Equation 4.2.14 implies that the superposition of Green's functions in the output plane yields the propagated field distribution. Hence, the constituent rays of the secondary wavelets are traced numerically inside the GRIN medium (see Appendix 1) to evaluate the corresponding Green's functions in that plane, where we use the principles developed in Section 2.4 to convert the traced ray distribution back into an equivalent field description. For this conversion, we require positions of the traced ray trajectories in the output plane as well as the total OPL accumulated by the rays in propagating from the secondary source to the output plane. From the ray positions, the irradiance profile for the wavelet is obtained using the conversion scheme detailed in Section 2.4 *in reverse*. Specifically, an integrated irradiance function is constructed along the transverse axis by *counting* the discrete contributions of radiant energy from individual rays, similar to the stepwise approximation in Figure 2.4.3. Differentiating this function yields sample values for the irradiance profile of the secondary wavelet in the output plane. We then use constrained cubic splines [Kruger, 2002] to interpolate between the sample values in order to provide a full description of the irradiance profile in the output plane along the transverse axis. Finally, we take the square root of the irradiance profile to obtain the field amplitude of the Green's function. Similarly, multiplying the total OPLs for each traced ray by the vacuum wavenumber $k = \frac{2\pi}{\lambda_0}$ and taking modulo-2π of the product provides a set of sample values for the phase profile of the secondary wavelet in the output plane. Once again, we employ constrained cubic splines to provide a full description of the phase profile along the transverse axis. Because the Fresnel-Kirchoff integral describes the coherent interference of secondary wavelets, we cannot simply construct the wavefront based on ray angles in the output plane, as suggested by the conversion scheme detailed in Section 2.4; such a geometrical construction would only specify the Eikonal for the propagating wavefront up to an arbitrary additive constant. Utilizing the OPL in computing the phase profile ensures that the additive constants in the Eikonals of the secondary wavelets are common to all secondary sources and that the interference between the secondary wavelets are calculated properly when we substitute the computed Green's functions into Equation 4.2.14. Finally, multiplying the field amplitude with the phase of the secondary wavelet yields the complex Green's function in the output plane.

Figure 4.3.2 illustrates the computed Green's function for a 2-D quadratic duct for a pair of secondary sources placed at $y = 0$ cm and $y = 0.1$ cm. The length of the duct is equal to 0.25 cm, and its transverse index profile is given by

$$n(y) = n_0\sqrt{1-\kappa^2 y^2}\,, \tag{4.3.1}$$

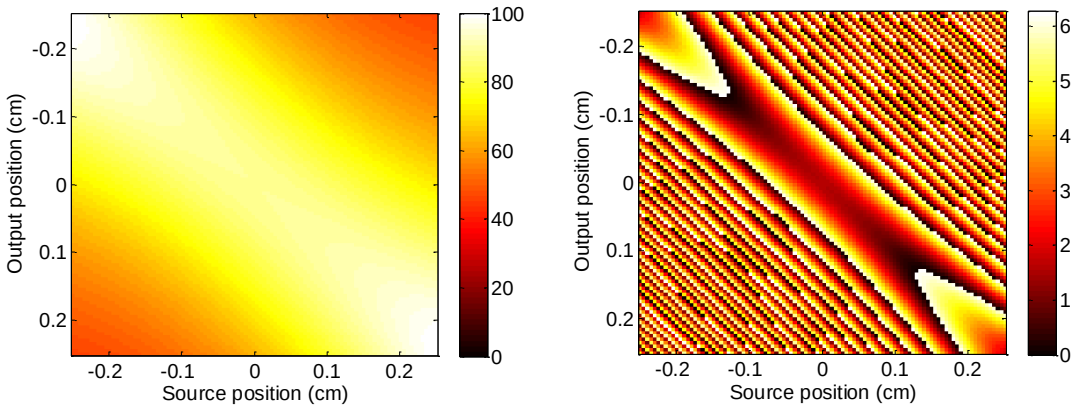
with $n_0 = 1.5$ RIU and $\kappa = 2\pi/5$ cm^{-1}.



(a) (b)

Figure 4.3.2. (a) Field amplitude and (b) phase profile (prior to modulo-2π) of the Green's function for two secondary sources in a quadratic duct.

The same calculation is performed for a set of uniformly-spaced secondary sources in the input plane for the same quadratic duct, and the corresponding Green's functions are imaged in Figure 4.3.3. In the figure, the Green's functions evaluated along the transverse axis in the output plane are plotted along the vertical axis, while the horizontal axis indicates the location of the corresponding secondary source in the input plane. This convention is especially useful for both organizing data and computing the Fresnel-Kirchoff integral.



(a) (b)

Figure 4.3.3. (a) Normalized field amplitude and (b) phase profile (in radians) of Green's functions used for wave propagation in a quadratic duct.

In particular, the Green's functions for all the secondary sources in the input plane are stored in a matrix $[G]$ with dimensionality $N \times N$, where N is the number of transverse grid points. If we assume a hypothetical, normally incident plane wave so that all the secondary sources exhibit identical strength and phase, then each column in $[G]$ describes the Green's function in the output plane corresponding to a particular secondary source in the input plane, whereas each row can be interpreted as the contributions from individual Green's functions to the field disturbance at a particular location in the output plane. Sorting the Green's functions according to their respective source positions in ascending order and imaging the matrix produces the plots shown in Figure 4.3.3. We remark that the Green's function is generally not shift-invariant in inhomogeneous media, as is apparent from the figure.

Next, suppose that a field vector $\vec{u}_1$ describes the field disturbance in the input plane. In addition, let us suppose that a secondary source is placed at each transverse grid point such that the complex value of $\vec{u}_1$ at each grid point constitutes the initial strength and phase of the secondary wavelet emitted from that particular source. While it is well known that secondary sources irradiate wavelets with a $-\frac{\pi}{2}$ phase shift due to the $-i$ factor in Equation 4.2.14, we can ignore this phase term in our computations (or any constant phase term across the transverse axis in general) because the phase will be common to all related field disturbances in a particular transverse plane. Since $\vec{u}_1$ describes the initial strength and phase of all secondary wavelets, we use its complex values to scale the contributions from the individual secondary wavelets. A straightforward implementation of this is achieved by computing the vector product between a particular row of $[G]$ and $\vec{u}_1$, which yields the field disturbance at the output position associated with that row if we ignore the effects of the obliquity factor $cos(\theta_\psi) + cos(\theta_G)$ in Equation 4.2.14. This vector product is a direct analog of computing the Fresnel-Kirchoff integral at a single point $\vec{p}$ in the output plane. Extending this principle to all grid points in the output plane, we have the formula

$$\alpha[G] \cdot \vec{u}_1 = \vec{u}_2 \,, \tag{4.3.2}$$

where $\vec{u}_2$ is the propagated field in the output plane, and α is a real-valued normalization constant. Assuming energy conservation is observed between the input and output planes, α is chosen such that the total irradiance in the input field is equal to that of the propagated field, *i.e.*,

$$\|\vec{u}_1\|^2 = \|\vec{u}_2\|^2 \,, \tag{4.3.3}$$

where $\|\vec{u}\|^2 = \sum_i u_i^2$ is the vector's L^2-norm.

The obliquity factor $cos(\theta_\psi) + cos(\theta_G)$ must be specified in a separate matrix $[O]$ if $[G]$ is to remain general to arbitrary field disturbances in the input plane. This new matrix specifies the obliquity factor for all elements in $[G]$; it is a function of $\vec{r}$ and $\vec{p}$ from Equation 4.2.14 as well as the input field $\vec{u}_1$. To incorporate the obliquity factor into our calculations, we modify Equation 4.3.2 as follows:

$$\alpha([O] * [G]) \cdot \vec{u}_1 = \vec{u}_2 \,, \tag{4.3.4}$$

where the symbol $*$ refers to element-wise multiplication between the matrices. If we make the approximation $cos(\theta_\psi) + cos(\theta_G) \approx 1 + cos(\theta_G)$, then the dependence of $[O]$ on $\vec{u}_1$ is effectively removed. This approximation enables us to designate a new characteristic matrix, $[\tilde{G}] = [O] * [G]$, so that Equation 4.3.4 is reduced to

$$\alpha[\tilde{G}] \cdot \vec{u}_1 = \vec{u}_2 \, . \tag{4.3.5}$$

We used the following simulation to check the agreement between the propagated field obtained through the application of Equation 4.3.4 and the diffracted field computed from the angular plane wave spectrum (APWS) method [Goodman, 2005] for calculating free space diffraction. In the APWS method, the input field disturbance is decomposed into a complete orthonormal set of 2-D basis functions using a 2-D Fourier transform. In a physical context, these basis functions can be interpreted as plane waves propagating in all possible directions in the forward half-plane. It is well known that plane waves are eigenfunctions of the scalar wave equation in a homogeneous medium. Hence, if we take the complex phase incurred by each plane wave between the input and output planes to be the eigenvalue corresponding to that particular direction, then the superposition of the propagated eigenfunctions yields the field disturbance in the output plane. In the simulation, a Gaussian beam whose beam waist coincides with the input plane was propagated over a distance of $z = 0.25$ cm. The beam width at the beam waist was set to $\sigma = 3 \times 10^{-2}$ cm, and the wavelength of the field was set to $\lambda_0 = 2 \times 10^{-2}$ cm. We chose this wavelength in order to reduce the computational resources needed to demonstrate our beam propagation method while avoiding aliasing effects that can arise due to the sampling of secondary sources in the input plane. We will defer the treatment of aliasing to Section 4.3. In this simulation, 100 uniformly-spaced transverse grid points were used to represent the complex field in both planes, which implies that 100 secondary sources were used in computing Equation 4.3.4.

The propagated fields obtained from both methods are depicted in Figure 4.3.4 and show strong agreement with one another. Under most practical scenarios, it can be shown that even moderate discrepancies in the obliquity factor or the field amplitude in the Green's function bear no real consequence in computing the propagated field. Hence, for our intents and purposes, the propagated field disturbance can be said to be solely dictated by the phase profile of the Green's function (even though such a statement is not rigorously true). This characteristic feature of the Fresnel-Kirchoff integral reconciles the fact that no obliquity factor appears in the APWS method and explains the strong agreement we were able to achieve using these methods. Moreover, it enables us to use the approximate expression in Equation 4.3.5 for practically all input field disturbances. We hasten to add that the APWS method is only consistent with the Fresnel-Huygens construction under paraxial conditions and requires shift-invariance in the Green's functions [Goodman, 2005].

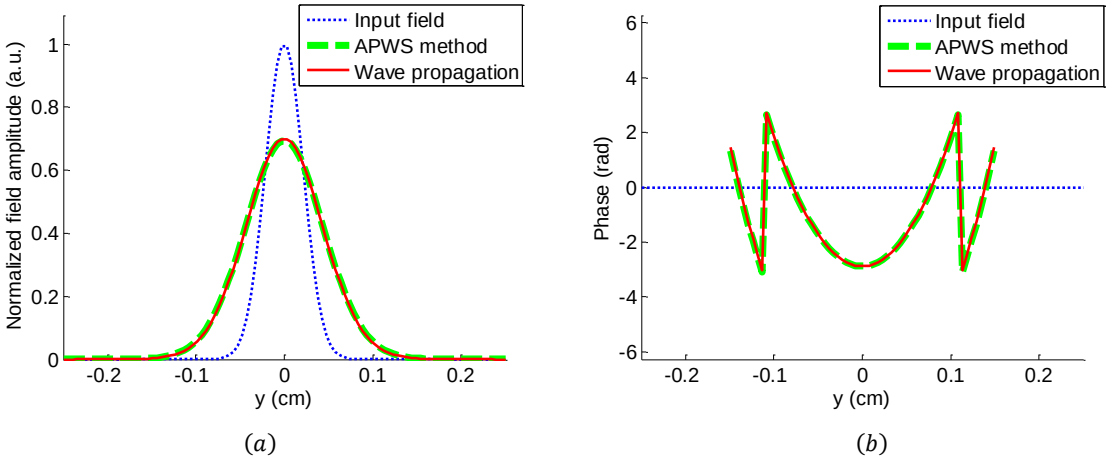


Figure 4.3.4. Comparison of (a) the field amplitudes and (b) phase profiles of a Gaussian beam propagated over a set distance in free space using the Fresnel-Kirchoff integral and the APWS method.

In a similar test, we examined the propagation of Gaussian beams inside a quadratic duct, whose transverse index profile is given by Equation 4.3.1. Substituting Equation 4.3.1 into the scalar wave equation, one obtains analytical solutions of the form

$$\psi(y,z) = H_m\left(\sqrt{2}\frac{y}{w_0}\right)e^{-\left(\frac{y}{w_0}\right)^2}e^{-i\beta_m z}\,, \tag{4.3.6}$$

where $w_0 = \sqrt{\frac{\lambda_0}{\pi\kappa n_0}}$, β_m is the propagation constant given by

$$\beta_m = \frac{2\pi}{\lambda_0}\sqrt{1-\frac{\lambda_0\kappa}{\pi}(m+1)}\,, \tag{4.3.7}$$

and $H_m(x)$ is the Hermite polynomial of order m. It is clear from Equation 4.3.6 that

$$\psi_m(y) = H_m\left(\sqrt{2}\frac{y}{w_0}\right)e^{-\frac{x^2+y^2}{{w_0}^2}} \tag{4.3.8}$$

are the eigenmodes of propagation inside the quadratic duct, and $e^{-i\beta_m z}$ is the corresponding eigenvalue after propagating distance z. Employing the same principles as the APWS method for calculating free space diffraction, any arbitrary input field $u_1(y)$ can be decomposed into the eigenmodes of propagation using the formula

$$\alpha_m = \int_{-\infty}^{\infty} u_1(y)\psi_m^*(y)dy\,, \tag{4.3.9}$$

where α_m is the projection of $u_1(y)$ onto $\psi_m(y)$, so that

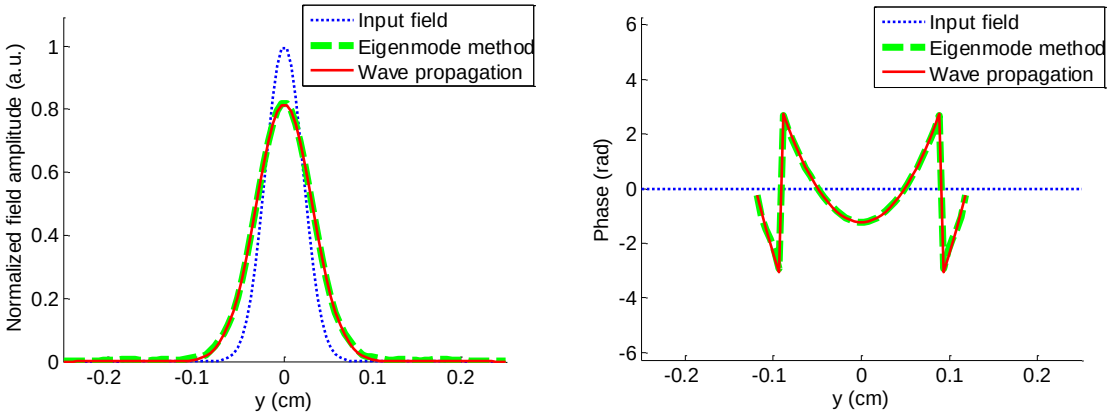
$$u_1(y) = \sum_m \alpha_m \psi_m(y) \, . \tag{4.3.10}$$

In principle, the integer m runs from zero to infinity, implying the existence of an infinite number of eigenmodes in the quadratic duct, all of which are needed to form a complete set of orthonormal basis functions. This, of course, makes the modal decomposition of arbitrarily complex input fields infeasible. To make the computation tractable, we restrict ourselves to simple input fields where $u_1(y)$ can be faithfully represented when m is bounded. In other words, we assume the modal coefficients α_m to be negligible for $m > m_{max}$. Under this assumption, the propagated field in the output plane is given by the superposition of the propagated eigenmodes,

$$u_2(y) = \sum_m \alpha_m e^{-i\beta_m z} \psi_m(y) \, . \tag{4.3.11}$$

where β_m is given by Equation 4.3.7.

Figure 4.3.5 shows a comparison of the propagated field calculated using the Fresnel-Kirchoff integral in Equation 4.3.5 as well as the modal decomposition method for a propagation distance of $z = 0.25$ cm inside the quadratic duct. In the modal decomposition method, we truncated the infinite series at $m_{max} = 11$. Note that since the wavefront is planar and parallel to the transverse axis in the input plane, we have $cos(\theta_\psi) = 1$ for all $\vec{r}$. Consequently, the expressions in Equations 4.3.4 & 4.3.5 are completely equivalent.



(a) (b)

Figure 4.3.5. Comparison of (a) the field amplitudes and (b) phase profiles of a Gaussian beam propagated over a set distance in a quadratic duct using the Fresnel-Kirchoff integral and the eigenmode method.

The comparison shows strong agreement between the propagated field amplitude and phase profile in both methods. The focusing action of the quadratic duct becomes apparent in the comparison between the propagated fields in Figures 4.3.4 & 4.3.5, where the transverse index variation inside the quadratic duct is seen to counteract the natural tendency for the confined beam to diffract and expand. To see what happens

to the Gaussian beam as it continues down the duct, we repeatedly applied Equation 4.3.5 to the propagated field using the iterative formula

$$\vec{u}_{i+1} = \alpha\left[\tilde{G}\right] \cdot \vec{u}_i \,, \tag{4.3.12}$$

where the subscript indicates the iteration count of propagation cycles, starting at $i = 1$. The initial input field was propagated to integer multiples of distance z in this manner, and its evolution over a total propagation distance of 5 cm is illustrated in Figure 4.3.6, where an interpolant based on the triangulation of scattered data was used to fill in the intermediate field distributions [Choi, *et al.*, 1988].

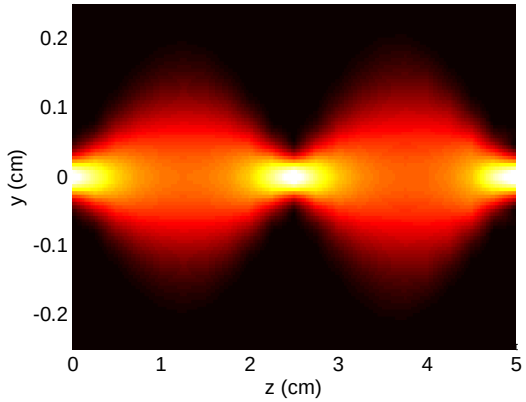


Figure 4.3.6. Scalloping behavior of a Gaussian beam propagating through a quadratic duct.

The apparent scalloping behavior of the propagating Gaussian beam is well known. In the framework of geometrical optics, it can be shown through ray transfer matrices that paraxial rays travel in sinusoidal trajectories inside the quadratic duct with a periodicity (or pitch) of $\Lambda = \frac{2\pi}{\kappa}$ [Yariv & Yeh, 2007]. Based on the quadratic duct parameters defined in Equation 4.3.1, we have $\Lambda = 5$ cm, which is in agreement with the above plot. The scalloping behavior can also be interpreted as the characteristic beating between propagating eigenmodes due to different propagation constants, analogous to the slow beat one hears when two different musical notes at extremely similar frequencies are played simultaneously.

4.4 Practical Considerations

In this section, we address several practical aspects of computing wave propagation inside GRIN media using the propagation method we have just developed. We begin with the phenomenon of phase aliasing, which can result from the interference of a discrete set of secondary wavelets. Phase aliasing can be interpreted as the imperfect cancellation of contributions from secondary sources to the field disturbance at some position in the output plane. To illustrate this principle, let us consider a discrete set of wavelets

emitted by secondary sources at locations $\vec{r}_i$ in the input plane, as depicted in Figure 4.4.1a. According to Equation 4.2.14, the superposition, or the interference, of these wavelets determines the field disturbance at point $\vec{p}$. As one possible demonstration of the aliasing effect, we consider the case where all the secondary wavelets in the figure have the same field amplitude after propagating to $\vec{p}$. In addition, their phases are uniformly distributed along the unit circle on the complex plane, as seen in Figure 4.4.1b. Furthermore, let us assume that the obliquity factor $cos(\theta_\psi) + cos(\theta_G)$ is approximately constant amongst the secondary sources. Under this scenario, the superposition of these secondary wavelets results in zero net field amplitude at $\vec{p}$ due to the perfect cancellation of each secondary wavelet with another at the opposite side of the unit circle on the complex plane.

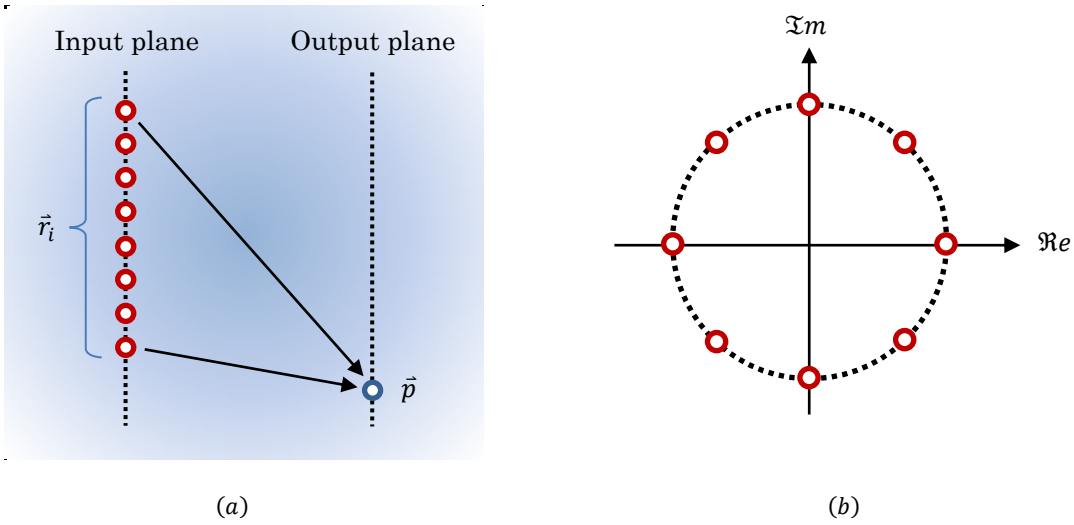


Figure 4.4.1. Contributions to the field disturbance at point $\vec{p}$ depicted from (a) a geometrical perspective showing distance traveled from secondary sources in the input plane and (b) a phasor perspective on the complex plane.

Recalling that our implementation in computing the Fresnel-Kirchoff integral (see Section 4.3) establishes a secondary source at each transverse grid point in the input plane, the separation between neighboring secondary sources is dictated by the transverse grid spacing, which we denote by Δ. Let us suppose that the value of Δ is chosen such that only the top and bottom secondary sources depicted in Figure 4.4.1a are used in evaluating Equation 4.2.14. Under this scenario, the superposition integral corresponds to the sum of two adjacent wavelet contributions from Figure 4.4.1b, which results in a constructive interference of the wavelets and produces a large error in the calculated field disturbance at $\vec{p}$. Based on the conditions responsible for this error, we reason that the onset of phase aliasing occurs when the phase difference between the wavelets emitted from neighboring secondary sources approaches π, or equivalently, when the OPL difference between the paths from adjacent secondary sources to point $\vec{p}$ approaches roughly half a wavelength inside the medium. To express the onset of phase aliasing

mathematically, let us consider the geometry depicted in Figure 4.4.2 and determine the maximum allowed transverse grid spacing Δ. Here, h denotes the transverse width of the medium.

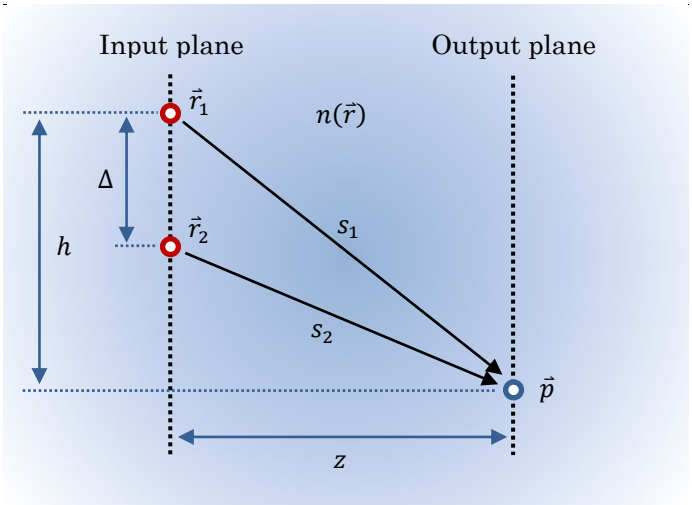


Figure 4.4.2. Geometry of phase aliasing in the interference of wavelets emitted from neighboring secondary sources.

Starting with the index distribution of the GRIN medium given by

$$n(\vec{r}) = n_0 + n_1(\vec{r}) \,, \tag{4.4.1}$$

where $\vec{r}$ is the position vector, we assume n_1 to be small so that n_0 times the arc length gives an adequate estimate of the OPL traveled. To further simplify the analysis, let us assume that rays travel in approximately linear trajectories inside the medium, for which case the distance traveled from the secondary sources to point $\vec{p}$ can be approximated as

$$s_1 \approx \sqrt{h^2 + z^2} \,, \tag{4.4.2a}$$

$$s_2 \approx \sqrt{(h - \Delta)^2 + z^2} \,. \tag{4.4.2b}$$

For practical values of Δ, Equation 4.4.1b can be rewritten as

$$s_2 \approx \sqrt{h^2 + z^2}\left[1 - \frac{\Delta(2h-\Delta)}{2(h^2+z^2)}\right], \tag{4.4.3}$$

where we have used $h \gg \Delta$. Hence, the difference in the distance traveled from the secondary sources to point $\vec{p}$ is given by

$$s_2 - s_1 \approx \frac{\Delta(2h-\Delta)}{2\sqrt{h^2+z^2}} \,. \tag{4.4.4}$$

In our foregoing discussion, we reasoned that the onset of phase aliasing can be expressed as the condition

$$n_0(s_2 - s_1) \rightarrow \frac{\lambda_0}{2}. \tag{4.4.5}$$

Therefore, phase aliasing can effectively be avoided if the following condition is satisfied:

$$\frac{\Delta(2h-\Delta)}{2\sqrt{h^2+z^2}} \leq \frac{\lambda}{2}, \tag{4.4.6}$$

where $\lambda = \lambda_0/n_0$ is roughly one wavelength inside the medium. It is clear from the inequality that smaller values of Δ are needed for smaller propagation distances z. In the worst case scenario where $z \ll h$ and $h \gg \Delta$, the above condition reduces to

$$\Delta \leq \frac{\lambda}{2}. \tag{4.4.7}$$

This upper bound on transverse grid spacing effectively prevents errors from developing in the output field due to the imperfect cancellation of secondary wavelets for arbitrary h and z. It states that we need at least one secondary source for every half-wavelength in the input plane. In other words, the minimum number of grid points needed to describe the input field is roughly equal to twice the number of wavelengths across the transverse plane.

Next, let us examine the practical issues associated with interpolation in evaluating the Green's function at the output plane. Because interpolation error generally increases with distance to the nearest sample, we are particularly concerned with interpolation errors outside the outer-most (traced) ray positions in the output plane. Outside these boundary sample values, spline interpolation typically enforces continuity in either the first or second derivative computed from the nearest spline. If a significant portion of the output plane in the figure becomes devoid of any rays, as seen in the ray trace depicted in Figure 4.4.3, the Green's function can effectively be considered as *clipped* or truncated.

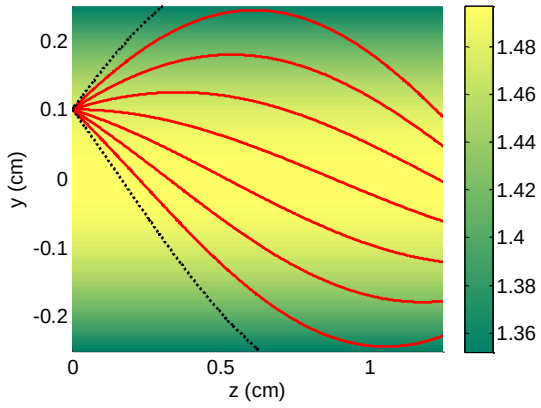


Figure 4.4.3. Clipped Green's functions under long propagation distances in a focusing medium.

In such cases, the rays that exit through the top and bottom boundaries of the simulation space are not redirected back into the GRIN medium and are therefore discarded. The resulting boundary regions devoid of any rays leave too much ambiguity for interpolation to produce reliable results beyond the outer-most rays, and the Green's function is set to zero outside the outer-most ray positions. Using these clipped Green's functions in the computation of the Fresnel-Kirchoff integral leads to contradictions in the propagated field disturbance. Specifically, the propagation over some distance $z = \sum_i z_i$ would differ from a series of propagations over the constituent distances z_i. This is especially problematic for a focusing medium, *e.g.*, a quadratic duct, if the propagation distance is on the order of one focal length, as shown in Figure 4.4.3. One way to remedy this issue is to decompose the overall propagation distance into shorter segments in order to eliminate the possibility of clipping in the computed Green's functions. We tested this partitioning scheme for a hypothetical Luneburg lens, whose index distribution is given by

$$n(r) = \begin{cases} \sqrt{2 - \left(\frac{r}{R}\right)^2} & , \quad r \leq R \\ 1 & , \quad r > R \end{cases}, \tag{4.4.7}$$

where $R = 0.25$ cm is the radius of the lens and $r = \sqrt{z^2 + y^2}$ is the radial distance from its center. In the simulation, the Luneburg lens was decomposed into 10 segments along the z-axis for characterization, similar to the partitions depicted in Figure 4.4.4a. The behavior of the lens is well known in the geometrical optics regime, which we illustrate using a ray trace in Figure 4.4.4b: Inside the Luneburg lens, all trajectories corresponding to a set of normally incident rays converge at the rear edge of the lens sphere. It is important to note that as the rays converge to a point and the amplitude profile of the field disturbance is reduced to dimensions on the order of one wavelength, the approximations in geometrical optics break down. Specifically, Equation 2.1.7 cannot be used to approximate Equation 2.1.6a under such conditions (see Section 2.1).

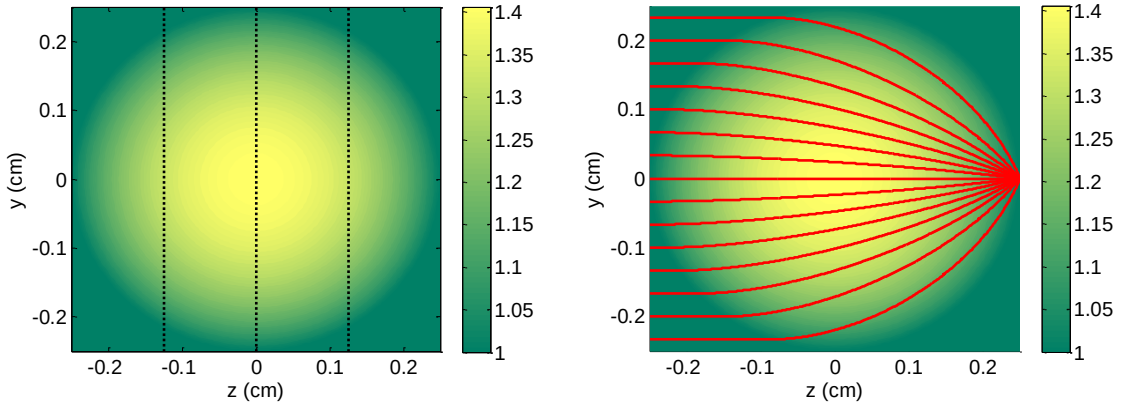


Figure 4.4.4. (a) Partitioning of a Luneburg lens for characterization and wave propagation and (b) focusing of parallel incident rays to a point by the lens under geometric optics.

Wave propagation inside an arbitrary index distribution $n(y,z)$ that has been partitioned along z can be calculated through the tandem application of Equation 4.3.4. The same principle was employed in Section 4.3, where we used Equation 4.3.12 to compute successive stages of beam propagation inside a quadratic duct and demonstrated the scalloping of a Gaussian beam inside the medium. Of course, the iterative expression in Equation 4.3.12 is only valid for GRIN ducts where the index distribution is of the form $n(y,z) = n(y)$. In the general case where the refractive index distribution differs for each segment in the partition, the Green's functions for each segment must be characterized separately. In other words, $[G]$ in Equation 4.3.4 is specific to each segment. Using the computed Green's functions for each segment of the Luneburg lens, the following simulation demonstrates the propagation of a confined optical beam through the lens using our beam propagation method. The optical beam used in the simulation was a Gaussian beam of beam width $\sigma = 0.1$ cm, with wavelength $\lambda_0 = 0.01$ cm, whose beam waist occurs at $z = -0.25$ cm. Figure 4.4.5b illustrates the evolution of the optical beam inside the Luneburg lens as it was focused into a "point" at the rear edge of the lens sphere. In generating the plot, we used an interpolant based on the triangulation of scattered data to provide intermediate fields between the propagated planes [Choi, *et al.*, 1988]. Due to the natural tendency of the confined beam to expand, the Gaussian beam could only be squeezed to dimensions of roughly one wavelength inside the medium. Aside from this diffraction effect, however, the simulation shows strong agreement with the ray trace depicted in Figure 4.4.4b.

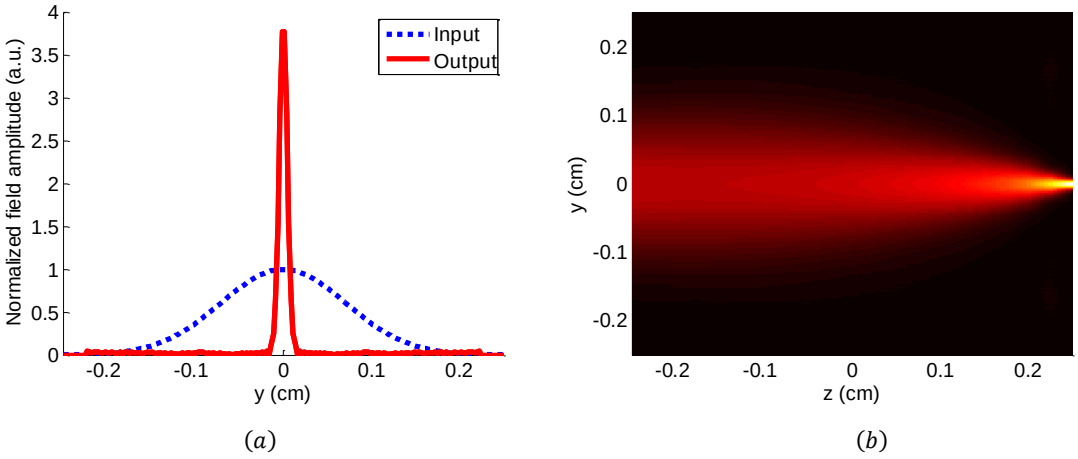


Figure 4.4.5. (a) Field amplitudes at the front and rear of the Luneburg lens and (b) evolution of the incident Gaussian beam as it propagates through the lens.

It is known that diffraction effects cause the focal point of the Gaussian beam to occur slightly behind the geometrical focal point at $z = 0.25$ cm [Gao, *et al.*, 2011]. Rather than demonstrating the focal shift explicitly by propagating the optical beam beyond the edge of the lens sphere, we plot the optical beam's phase profile along select transverse planes at $z = -0.25$ cm, $z = 0$ cm and $z = 0.25$ cm in Figure 4.4.6a to show that the beam has not yet reached its focal point and is still converging upon reaching the rear edge

of the lens sphere. We also intensified the diffraction effects by increasing the wavelength of the beam to $\lambda_0 = 0.02$ cm in this simulation.

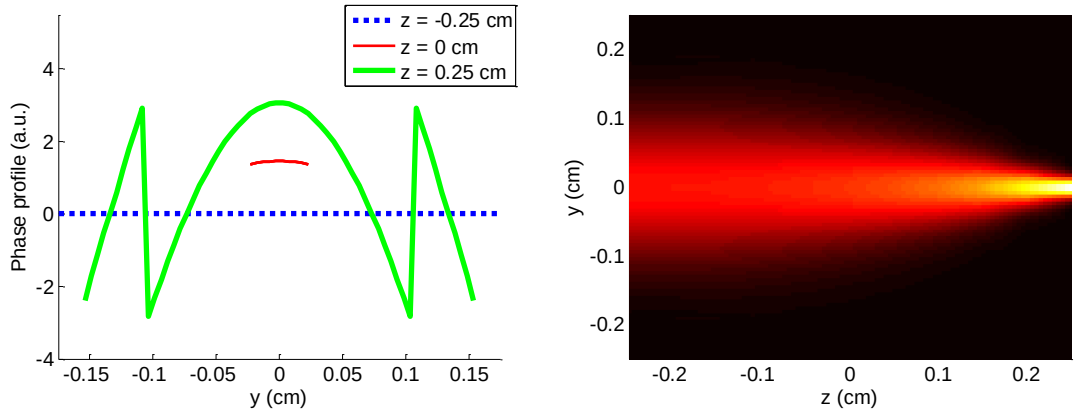


Figure 4.4.6. (a) Phase profiles along select transverse planes inside the Luneburg lens and (b) evolution of the incident Gaussian beam as it propagates through the lens.

An alternative for demonstrating the effects of diffraction inside the Luneburg lens is to reduce the initial beam width of the incident Gaussian beam. As is apparent from the simulations in Figure 4.4.7, the tendency of the beam to diffract can be made to completely counteract the focusing action of the lens for all z by increasing the initial confinement of the optical beam. All other parameters were unchanged in these simulation from those used for generating Figure 4.4.6.

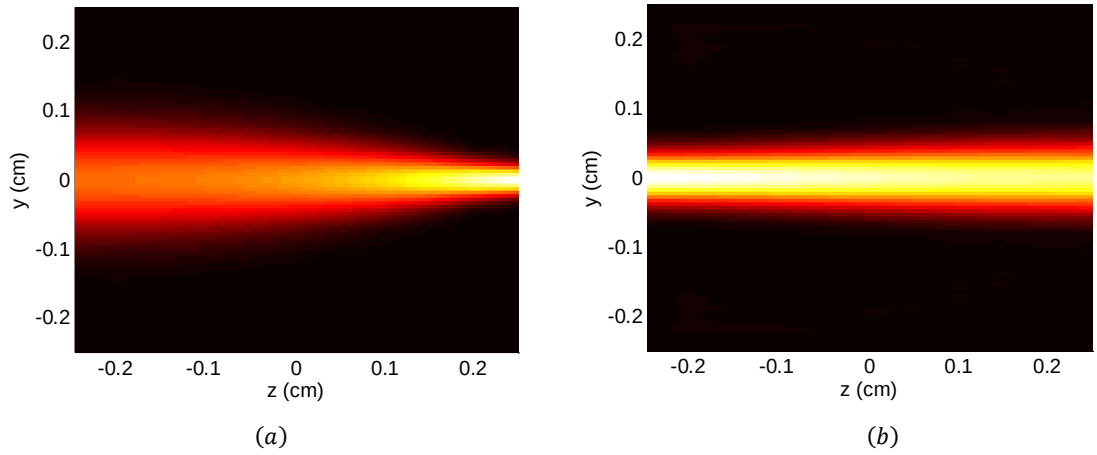


Figure 4.4.7. Field amplitude of a Gaussian beam with an initial beam width of (a) $\sigma = 0.75$ cm and (b) $\sigma = 0.03$ cm as it propagates through a Luneburg lens.

It is worth mentioning that our cascaded propagation technique bears some similarity to the split-step beam propagation method in GRIN media [Poon & Kim, 2006]. In the latter method, the index distribution is partitioned along the z-axis such that the phase incurred over extremely short propagation distances in the medium is approximated by transmission through a thin phase plate. The method then incorporates diffractive effects by using the Fresnel diffraction integral to propagate the transmitted beam through a homogeneous medium between successive phase plates. This simplification is, of course, based on approximations that do not reflect reality. One way to illustrate a fundamental flaw in the split-step beam propagation method is to consider the transmission of a normally incident collimated beam through a linear index gradient in y of the form $n(y) = n_0 + n_1 y$, where n_0 and n_1 are both real constants. If the index distribution over a short propagation distance is compressed into a phase plate, then the transmitted wave is a tilted collimated beam. However, it can be shown through a ray trace that normally incident rays separated in y take on different slope values after propagating a short distance through such a medium. This inconsistency can also be reasoned from Equation 3.3.1, where it is clear that only a linear gradient in the logarithmic refractive index $w = \ln(n)$ will yield parallel rays after a set propagation distance inside the medium.

4.5 Computing Resonator Modes

In the preceding sections, we demonstrated the efficacy of the Fresnel-Kirchoff integral for beam propagation through various GRIN media and tackled the practical issues that can lead to large computational errors. In this section, we will demonstrate the application of our beam propagation technique in computing the eigenmodes of optical resonators incorporating intracavity GRIN elements.

We begin our analysis with the round-trip propagation equation in an optical resonator,

$$v(x, y) = \boldsymbol{F}\{u(x, y)\}\,, \tag{4.5.1}$$

where $\boldsymbol{F}$ is the round-trip operator, and $v(x, y)$ is the field disturbance, which can be defined in any arbitrary transverse plane inside the resonator, after propagating the field $u(x, y)$ (defined in the same transverse plane) one round trip through the resonator. Given $\boldsymbol{F}$, one can always find a discrete of transverse eigenmodes, or eigenfunctions of the operator, that will reproduce themselves in form after one round-trip, *i.e.*,

$$\boldsymbol{F}\{\psi_m(x, y)\} = \beta_m \psi_m(x, y)\,, \tag{4.5.2}$$

where β_m is the complex eigenvalue corresponding to m^{th} eigenmode $\psi_m(x, y)$. The Fox-Li method, also referred to as the power method or the ring-down method, is perhaps the most commonly used technique for calculating the resonator's transverse eigenmodes. In this method, successive analytical or computer calculations are used to determine changes in the transverse field pattern of the optical beam after each round trip through the resonator [Fox & Li, 1961]. One typically starts from an arbitrary field disturbance

$u_0(x,y)$ in an arbitrary transverse plane inside the resonator, which can presumably be expressed as a linear combination of the cavity's eigenmodes,

$$u_0(x,y) = \sum_m \alpha_m \psi_m(x,y) \,, \tag{4.5.3}$$

where α_m is the component of $u_0(x,y)$ along the m^{th} eigenmode,

$$\alpha_m = \int_{-\infty}^{\infty} u_0(x,y)\psi_m^*(x,y)d^2r \,, \tag{4.5.4}$$

where d^2r denotes integration over the transverse plane. After one round trip through the cavity, the field disturbance in the initial transverse plane becomes

$$u_1(x,y) = \boldsymbol{F}\{u_0(x,y)\} = \sum_m \boldsymbol{F}\{\alpha_m \psi_m(x,y)\} = \sum_m \beta_m \alpha_m \psi_m(x,y) \,. \tag{4.5.5}$$

Likewise, the field disturbance after N round trips through the cavity will be given by

$$u_N(x,y) = \boldsymbol{F}^N\{u_0(x,y)\} = \sum_m (\beta_m)^N \alpha_m \psi_m(x,y) \,. \tag{4.5.6}$$

In passive systems where $|\beta_m| \leq 1$ for all m, the round trip power loss for the m^{th} eigenmode is equal to $1 - |\beta_m|^2$.

For the ensuing analysis, let us suppose that the eigenmodes are ordered such that $|\beta_i| > |\beta_j|$ for $0 \leq i < j$ and that there is no degeneracy in the eigenvalues. It is then apparent from Equation 4.5.6 that the term with the largest value for $|\beta_m|$ will dominate the entire expression as N approaches a large value, while all the other terms eventually become negligible, *i.e.*,

$$\lim_{N\to\infty} u_N(x,y) = (\beta_0)^N \alpha_0 \psi_0(x,y) \,. \tag{4.5.7}$$

This limiting expression describes, in principle, how one can obtain the eigenmode with the lowest round-trip losses in the Fox-Li method. In practice, the value of N needed in the ring-down process depends on the modal discrimination, or the difference, between $|\beta_0|$ and $|\beta_1|$. For instance, a larger difference between the eigenvalues would require a smaller N for computing $\psi_0(x,y)$.

When the eigenmodes form an orthonormal set, *i.e.*,

$$\int_{-\infty}^{\infty} \psi_i(x,y)\psi_j^*(x,y)d^2r = \delta_{ij} \,, \tag{4.5.8}$$

where δ_{ij} is the Kronecker delta function, the Fox-Li method can be used to obtain higher-order eigenmodes ($m > 0$) through a technique known as Gram-Schmidt orthogonalization. In this method, we choose an initial field disturbance $u_0(x,y)$ that is orthogonal to $\psi_0(x,y)$ by removing its component along the lowest order eigenmode via

$$u_0(x,y) = u(x,y) - \psi_0(x,y)\left[\int_{-\infty}^{\infty} u(x,y)\psi_0^*(x,y)d^2r\right] \,, \tag{4.5.9}$$

where $u(x,y)$ can be any arbitrary field disturbance in the transverse plane. Using Equation 4.5.9 as the initial field disturbance in the Fox-Li method removes the $m = 0$ term from the expressions in Equations 4.5.3, 4.5.5 & 4.5.6. Consequently, Equation 4.5.7 now yields

$$\lim_{N\to\infty} u_N(x,y) = (\beta_1)^N \alpha_1 \psi_1(x,y)\,. \tag{4.5.10}$$

Extending the same principle to the computation of any eigenmode of order M by the Fox-Li method, we require an initial field disturbance

$$u_0(x,y) = u(x,y) - \sum_{m=0}^{M-1} \psi_m(x,y)\left[\int_{-\infty}^{\infty} u(x,y)\psi_m^*(x,y)d^2r\right], \tag{4.5.11}$$

where $u(x,y)$ is, once again, any arbitrary field disturbance in the transverse plane.

In practice, it is not computationally efficient to enforce the limiting expression in Equations 4.5.7 & 4.5.10 in computing the spatial eigenmodes. Instead, the similarity between field disturbances in a chosen transverse plane from back-to-back round trips through the resonator is evaluated as

$$\tilde{\beta}(x,y) = \frac{u_k(x,y)}{u_{k-1}(x,y)}, \tag{4.5.12}$$

where the integer k denotes the number of round trips through the resonator. Once the overall deviation in $\tilde{\beta}(x,y)$ from its mean value $\bar{\beta}$ has fallen below some threshold value Δ_{th}, *i.e.*,

$$\int\left|\tilde{\beta}(x,y) - \bar{\beta}\right| d^2r < \Delta_{th}\,, \tag{4.5.13}$$

the algorithm is said to have reached steady state. Under this condition, it is reasonable to assume that $u_k(x,y)$ has been reduced to the eigenmode of interest and $\bar{\beta}$ is equal to its eigenvalue.

The Fox-Li method can be extended to GRIN resonators, *i.e.*, optical resonators that employ GRIN media, by incorporating our beam propagation method described in the preceding sections into the round-trip operator $\boldsymbol{F}$. To demonstrate this, suppose that we have a 2-D optical resonator with perfectly reflecting flat mirrors ($d = 0.5$ cm across) at both ends and a thin absorber with amplitude transmittance

$$t(y) = e^{-\left(\frac{y}{\rho}\right)^2}\,, \tag{4.5.14}$$

where $\rho = 0.1$ cm, is placed in the center of the cavity, as seen in Figure 4.5.1a. Unfolding the resonator structure to eliminate the end mirrors yields an equivalent periodic structure depicted in Figure 4.5.1b. From the latter figure, it becomes apparent that the round-trip operator consists of propagating the field disturbance inside the GRIN medium over a distance L, where L is the cavity length of the resonator, and then multiplying the propagated field disturbance by $t(y)$. If the cavity of the optical resonator is filled with free space, *i.e.*, $n(y) = 1$, then the propagation over distance L can be computed using the Fresnel diffraction integral or the APWS method [Goodman, 2005].

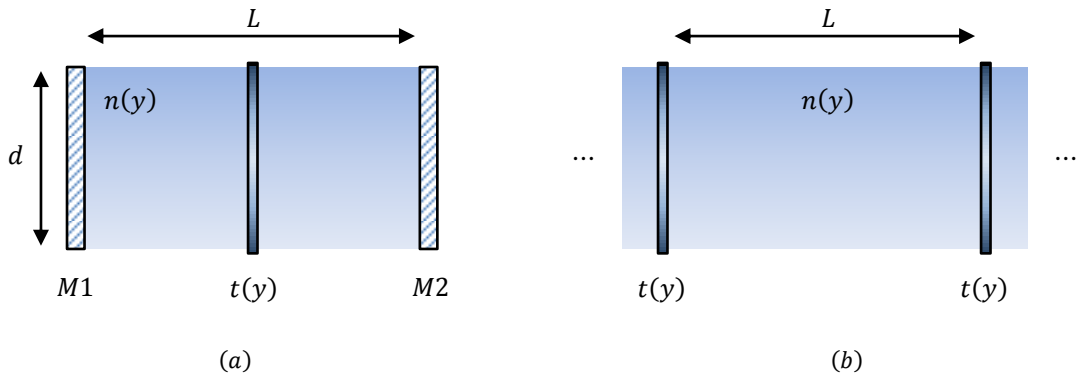


Figure 4.5.1. (a) Hypothetical resonator configuration and (b) an equivalent periodic structure.

If the cavity is instead filled by a quadratic medium whose index distribution is of the form $n(y) = n_0\sqrt{1-\kappa^2 y^2}$, then analytical solutions to the eigenmodes of the medium can be used to propagate the field disturbance over distance L. Of course, the operator corresponding to wave propagation in both cases can be replaced by our beam propagation method as well. However, verifying the calculated eigenmodes using our beam propagation method holds little merit; we have already shown strong agreement with these analytical methods in Section 4.3 for both test cases. Furthermore, it follows that the optical resonator does not produce particularly interesting eigenmodes under these circumstances. As one might expect, the eigenmodes are very similar to Hermite-Gaussians with perturbed phase profiles. In order to make the resonator more interesting, let us fill the cavity of the resonator with a cubic medium whose index distribution is given by

$$n(y) = n_0(1 - \kappa|y|^3)\,, \tag{4.5.15}$$

where $n_0 = 1.5$ RIU and $\kappa = 10$ cm^{-3}.

Using the tandem propagation technique described in Section 4.4 to compute wave propagation over a distance of 3.5 cm inside the GRIN material, we illustrate the evolution of the propagated field amplitude in Figure 4.5.2 for two Gaussian beams of different initial beam widths. In generating the plots in the figure, 400 transverse grid points were used to describe the field disturbance, and the propagated field was calculated every 0.1 cm along z. The wavelength of the optical beam was set to $\lambda_0 = 1 \times 10^{-2}$ cm. Figure 4.5.2a suggests that the cubic GRIN duct essentially behaves like an aberrated focusing medium that can counteract the beam's natural tendency to diffract if the field disturbance is confined to a local region near the axis of symmetry ($y = 0$). Figure 4.5.2b, on the other hand, shows very peculiar transverse fields if the optical beam is not well confined inside the medium. In particular, nulls (or zero crossings) in the field amplitude can be observed at select propagation distances.

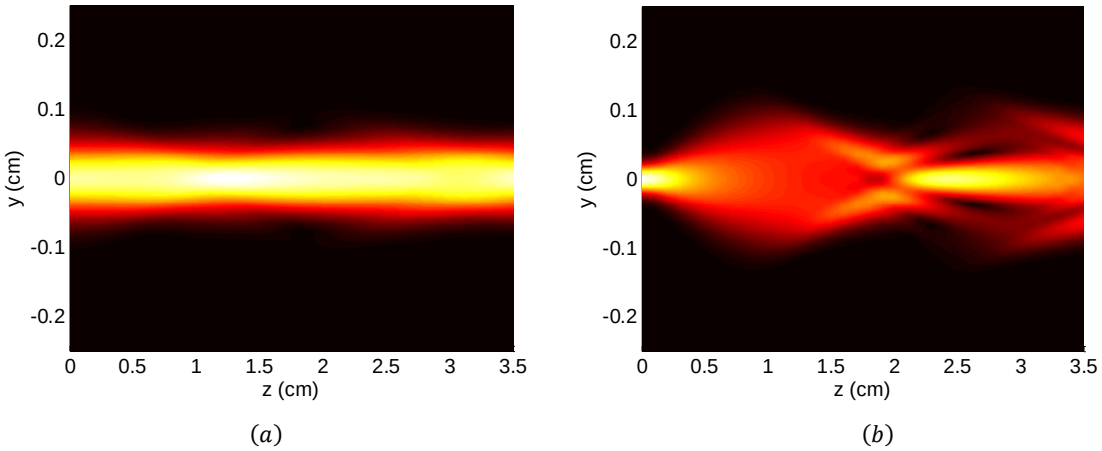


Figure 4.5.2. Field amplitude of a Gaussian beam with an initial beam width of (a) $\sigma = 0.05$ cm and (b) $\sigma = 0.02$ cm as it propagates through the cubic medium.

We demonstrate the ramifications of this wave propagation behavior by using the Fox-Li method to compute the eigenmodes of the resonator. Setting the round-trip distance to $L = 2.0$ cm for the cubic GRIN duct resonator, the first two eigenmodes, $\psi_0(y)$ and $\psi_1(y)$, are shown in Figure 4.5.3, whose round-trip losses are 23.80% and 26.01%, respectively. The resulting discrimination between the two lowest-order eigenmodes is inferior compared to that of conventional Fabry-Perot laser cavities with a similar Fresnel number. However, the number of zero crossings exhibited by these modes is of particular interest and warrants further analysis of the resonator's behavior.

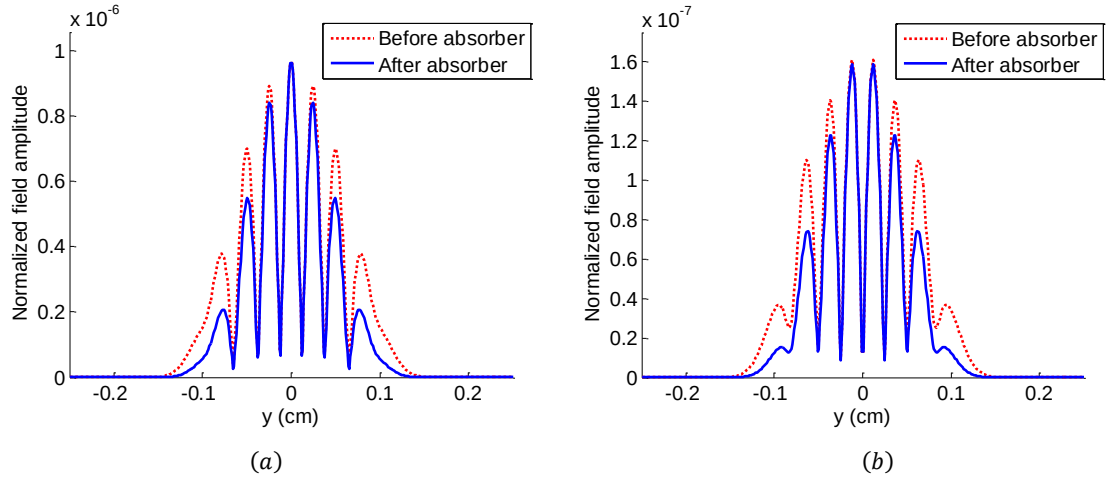


Figure 4.5.3. Field amplitudes for eigenmodes (a) $\psi_0(y)$ and (b) $\psi_1(y)$ inside the cubic GRIN duct resonator.

The most plausible explanation for these peculiar eigenmodes is that a mode with a simpler field amplitude profile simply cannot reproduce itself after one round trip through the resonator. This is, of course, a fundamental requirement for any field disturbance to qualify as an eigenmode. Figure 4.5.4 illustrates how the first eigenmode $\psi_0(y)$ reproduces itself after one round trip, where the evolution of its field amplitude is plotted over the propagation distance L inside the GRIN medium.

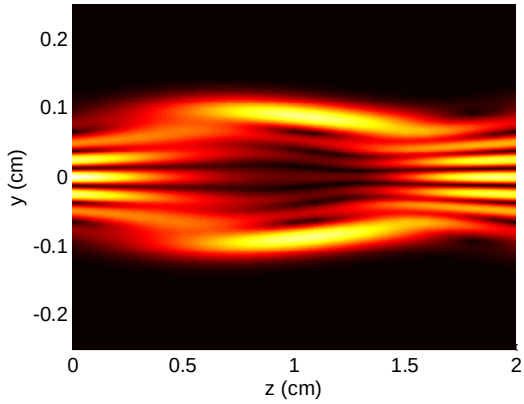


Figure 4.5.4. Propagation of eigenmode $\psi_0(y)$ through the cubic GRIN duct.

It follows that the number of zero crossings in these eigenmodes changes with the round trip propagation distance. When this distance falls below 1.0 cm, however, transmission through the Gaussian absorber occurs with such frequency that the diffraction of the transmitted optical beam essentially dictates the resulting eigenmodes. Consequently, the resonator produces far less interesting eigenmodes resembling Hermite-Gaussians, which we depict in Figure 4.5.5.

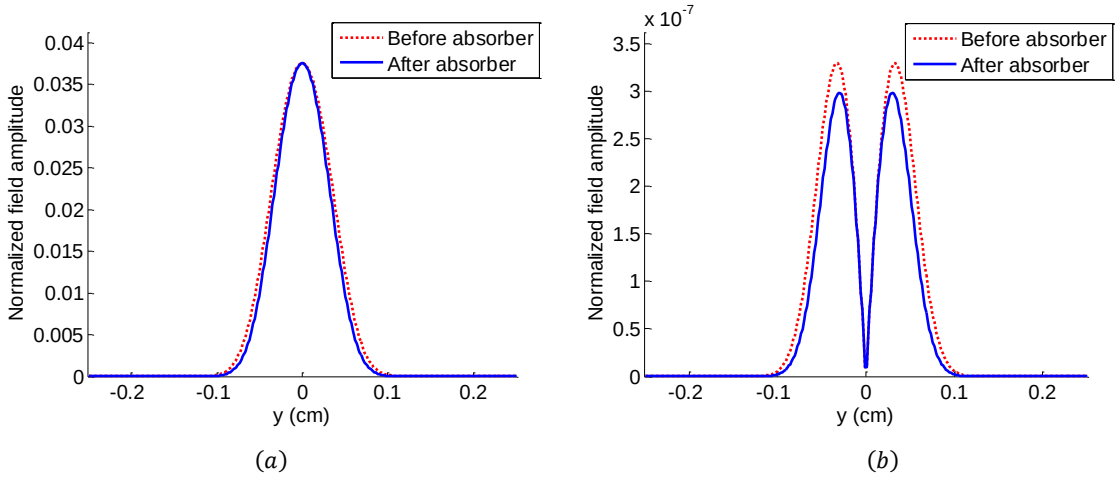


Figure 4.5.5. Field amplitudes for eigenmodes (a) $\psi_0(y)$ and (b) $\psi_1(y)$ inside the cubic GRIN duct resonator after reducing the round-trip propagation length.

For these modes, the round-trip losses are 9.78% and 24.03%, and the mode number m is equal to the number of zero-crossings in the field amplitude. This discrimination value is very typical for these types of resonator modes for the given Fresnel number.

We conclude this section with a precautionary measure for the Fox-Li method regarding the application of Gram-Schmidt orthogonalization. During the ring-down process, computational errors associated with the round-trip operator $\boldsymbol{F}$, *e.g.* in the computation of the Fresnel-Kirchoff integral, can lead to scattering between the eigenmode components of the field disturbance. For instance, a field disturbance with no initial component along $\psi_0(x,y)$ can develop a small, non-zero component along the eigenmode following the application of $\boldsymbol{F}$. For this reason, we recommend applying the orthogonalization in Equation 4.5.11 at the beginning of each round trip in the ring-down process. Otherwise, the method will tend to yield the fundamental eigenmode $\psi_0(x,y)$ regardless of the chosen initial field disturbance $u_0(x,y)$. Finally, we note that the symmetry of the field disturbance can be exploited to achieve orthogonality. For instance, a Gaussian beam with beam width $\sigma = 0.1$ cm was used as the initial field disturbance in obtaining the eigenmode $\psi_0(y)$ from Figure 4.5.3*a*, *i.e.*,

$$u_0(y) = e^{-\left(\frac{y}{\sigma}\right)^2}\ , \tag{4.5.16}$$

under the expectation that $\psi_0(y)$ is an even function in y. Likewise, we expected $\psi_1(y)$ to be an odd function and changed $u_0(y)$ to

$$u_0(y) = ye^{-\left(\frac{y}{\sigma}\right)^2} \tag{4.5.17}$$

accordingly for computing the second eigenmode. Based on the symmetry of $\psi_0(y)$ from Figure 4.5.3*a*, it is clear that the initial field disturbance in Eq. (4.5.17) does not contain any components along $\psi_0(y)$, *i.e.*,

$$\alpha_0 = \int_{-\infty}^{\infty} u_0(y)\psi_0^*(y)dy = 0\ . \tag{4.5.18}$$

Furthermore, if the field disturbance remains an odd function during the ring-down process, Equation 4.5.11 becomes entirely unnecessary for computing $\psi_1(y)$. We hasten to add, however, that the Gram-Schmidt orthogonalization process remains necessary if we are interested in higher order modes, *e.g.*, $\psi_2(y)$, that exhibit the same symmetry as any of the previously calculated eigenmodes.

4.6 Mode Converter Design

Our goal by the end of this chapter is develop the mechanics for analyzing optical resonators that incorporate more complex GRIN elements to control the propagation of light inside the cavity. In this regard, GRIN elements that are tailored to perform a specific coherent mode conversion are of particular interest to us. In this section, we will demonstrate a procedure for designing such a GRIN element, whose

index distribution can be any arbitrary, smooth function of the spatial variables. Furthermore, we will show that arbitrary coherent mode conversion can be treated as the inverse design problem of finding an index distribution to effect a particular desired transformation in the ray distribution representing the optical beam (see Section 2.4).

We start by defining the input and output modes of the desired conversion: Suppose that we wish to convert a collimated Gaussian beam of beam width $\sigma = 0.55$ cm to a collimated order-6 super-Gaussian beam whose field amplitude falls to e^{-1} times its peak value at $y = 0.8$ cm. We chose a super-Gaussian output beam profile [Palma & Bagini, 1994] in lieu of a true flat-top beam because the infinitely sharp drop-offs in the beam's irradiance profile (in the latter case) typically give rise to high-frequency edge effects during beam propagation. To this end, we could have also chosen flattened Gaussian beams [Gori, 1994], which exhibit similar propagation behavior to super-Gaussian beams. Figure 4.6.1a illustrates the desired coherent mode conversion, where it is understood that the phase profile of the optical beam is flat in both the input and output planes. In assigning a ray distribution to represent the optical beam, the conservation of energy mandates that each ray in the input plane has a counterpart in the output plane with an equal amount of radiant energy. Hence, the discrete levels in the stepwise approximation to the integrated irradiance function (see Section 2.4) must be matched in energy values between the input and output ray representations, as seen in Figure 4.6.1b. This also ensures that the number of boundary values (for position and slope) corresponding to the rays in the distribution is matched between the input and output planes. In the figure, we used 10 rays containing equal amounts of radiant energy to represent the optical beam; their boundary positions in the input and output planes are indicated by the position of the circular marks (along the respective integrated irradiance functions) on the horizontal axis.

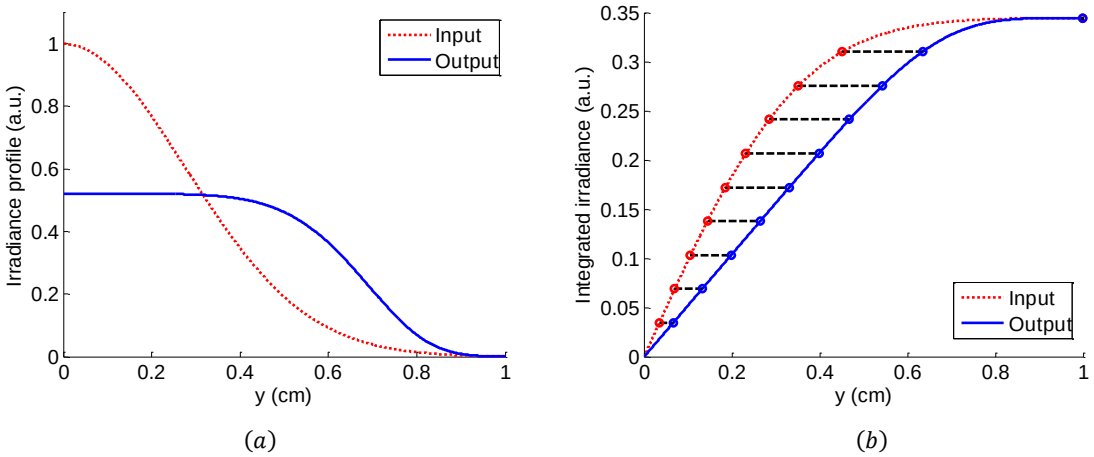


Figure 4.6.1. (a) Input and output beam irradiance profiles for the desired mode conversion and (b) establishing equal amounts of radiant energy for corresponding rays in the input and output distributions.

Upon establishing the input and output ray distributions, a set of trajectories $Y_i(x)$ is then used to direct ray propagation between the boundary planes, where evaluating the ray family $Y_i(x)$ at a particular

value of i singles out a particular ray path. The choice of trajectories is completely arbitrary, provided that they satisfy the boundary values of ray position and slope. These trajectories will ultimately determine the characteristics of the index distribution used to effect the desired mode conversion (see Appendix 4). Figure 4.6.2 illustrates one possible choice of $Y_i(x)$, where cubic polynomials are employed to connect corresponding boundary ray positions. We note that the boundary slopes for all rays in the input and output distributions are set to zero to reflect the shape of the wavefront for the collimated beams in both planes.

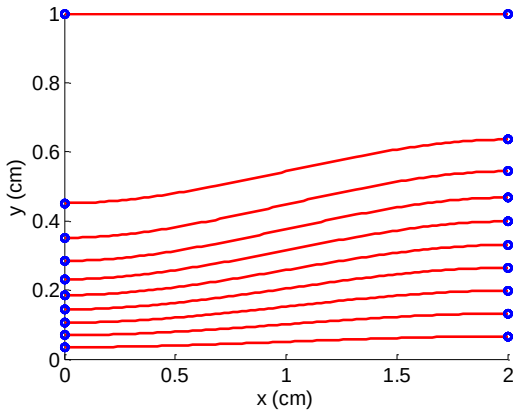


Figure 4.6.2. Connecting corresponding boundary ray positions with cubic polynomials.

For a mode converter of length L, the cubic polynomial for a given trajectory in the ray family is given by

$$Y_i(x) = \frac{\tau_i}{3L}x^3 + \frac{\tau_i}{2}x^2 + y_{in,i}\ , \tag{4.6.1}$$

where $\tau_i = \frac{6}{L^2}\left(y_{out,i} - y_{in,i}\right)$, $y_{in,i}$ is the ray's input position, and $y_{out,i}$ is its output position. Accounting for technological constraints in fabricating GRIN materials, we generally seek trajectories that minimize the absolute peak value of the second derivative with respect to x, which in turn minimizes the maximum index gradient required according to Equation 3.3.2. In addition, we avoid ray crossings in $Y_i(x)$ whenever possible in assigning ray trajectories because the assumptions of geometrical optics break down at these locations. Moreover, ray crossings restrict our freedom in the design of the mode converter because the index gradient is fully determined at these locations (see Appendix 2). It stands to reason that a family of trajectories following a similar functional form tends to optimize smoothness in the resulting refractive index distribution. For example, if one trajectory in $Y_i(x)$ follows a cubic polynomial in x, then all other trajectories should conform to cubic polynomials to obtain a smooth refractive index profile. If there exist trajectories in $Y_i(x)$ that differ significantly in functional form, a gradual transition between the two types of trajectories should be considered across index i in $Y_i(x)$.

Once $Y_i(x)$ has been determined, the index profile along each wavefront is specified up to an arbitrary constant, provided that $Y_i(x)$ satisfies a particular set of conditions [Beliakov & Chan, 1997].

Rather than solving a system of partial differential equations as Beliakov and Chan did in their analysis, we present a far more intuitive approach based on Equation 2.1.8. Although the latter approach does not specifically require ray trajectories to be single-valued functions of x, we assume this to be the case for ease of notation without the loss of generality. Equation 2.1.8 is a fundamental expression of the relationship between the Eikonal S and the differential OPL traveled by a ray inside the medium. Rewriting the equation in differential form along $\hat{s}$ yields

$$dS = nds\,, \tag{4.6.2}$$

where n is the refractive index, and ds is the arc length. In this form, it is clear that the product between the refractive index and the arc length traveled is responsible for the incremental increase in S. Along the ray trajectories, dS in the differential can be replaced by $d\phi$ where ϕ is the OPL traveled by the ray, provided that the rays have identical ϕ along geometrical wavefronts described by constant contours of S. Using this fact, solving for the refractive index in Equation 4.6.2 becomes straightforward once these wavefronts have been determined.

Given a set of ray trajectories $Y_i(x)$, the geometric wavefronts are given by the characteristic lines

$$\frac{dx}{dy} = \frac{dY_i}{dx}\,, \tag{4.6.3}$$

where we have used the fact that the wavefronts are always perpendicular to the local ray vector $\hat{s}$. In our implementation, we employ linear extrapolation in calculating the ray positions for all trajectories in $Y_i(x)$ along a particular wavefront. We denote these positions by $X_j(y)$, where evaluation at a particular value of j isolates the crossing points between the j^{th} wavefront and all ray trajectories in $Y_i(x)$.

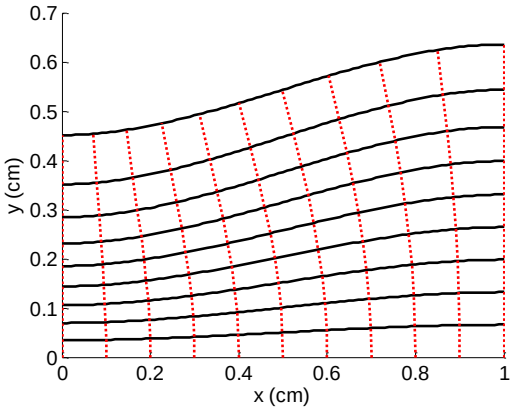


Figure 4.6.3. Sampling wavefronts derived from sampling rays in the mode converter design.

In principle, $Y_i(x)$ can encompass a continuum of ray trajectories if the difference between the radiant energy levels in Figure 4.6.1b is infinitesimally small. Similarly, $X_j(y)$ can also represent a continuum of

wavefronts with infinitesimal increments in ϕ. For practical reasons, our implementation uses a discrete set of *sampling* wavefronts derived from a set of *sampling* rays to characterize the mode converter, as seen in Figure 4.6.3, where the solid lines correspond to the trajectories of sampling rays and the dotted lines trace the sampling wavefronts.

In extrapolating the geometrical wavefronts, significant walk-off errors can develop if the sampling rays are sparse or if the change in slope between adjacent sampling rays, *i.e.*, $\frac{d}{di}\left(\frac{dY_i}{dx}\right)$, is large. The latter depends on the redistribution of irradiance in the desired mode conversion and on the choice of $Y_i(x)$; hence, it generally does not cause issues if all the ray trajectories in $Y_i(x)$ exhibit similar form. However, a large contrast in the beam's irradiance profile often produces sparse regions in the ray distribution if the rays used to represent the optical beam contain equal or similar amounts of radiant energy, as is evident from Figure 4.6.2. A simple solution to this problem is to selectively reduce the step size between the radiant energy levels in assigning boundary positions as well as the amount of radiant energy contained in each ray. For instance, if the separation between neighboring boundary ray positions is too large, the difference in radiant energy corresponding to the separation can be divided into two equal partitions along the vertical axis so that *two* distinct rays contribute to the energy difference, with each ray containing *half* the amount of radiant energy as the original ray. This process provides a systematic method for inserting additional rays into the input and output ray distributions; it can be applied iteratively until the maximum separation (in y) between boundary ray positions is below a threshold value. Figure 4.6.4 illustrates a family of sampling ray trajectories incorporating variable amounts of radiant energy for the desired mode conversion specified in Figure 4.6.1a, where each ray is color-coded based on its radiant energy and ϵ is the base (constant) value used in Figure 4.6.2. In addition, the switch from solid lines to dotted lines (and vice versa) in the figure indicates a reduction in the amount of radiant energy (by a factor of 2) contained in the sampling ray as we move away from the axis of symmetry ($y = 0$).

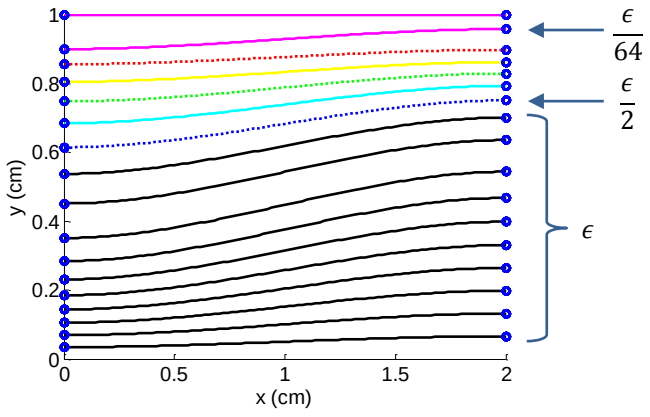


Figure 4.6.4. Redefining boundary ray distributions using variable radiant energies.

To extrapolate the sampling wavefronts from the sampling ray trajectories, we start at a given z position along $y = 0$, which we identify as the position of the wavefront along a virtual sampling ray traveling along the axis of symmetry in this particular design. From this starting position, linear extrapolation is used to locate the intersection between the wavefront and all the ray trajectories in $Y_i(x)$ in order to construct the characteristic line in Equation 4.6.3 as we move away from the axis of symmetry. To reduce walk-off errors, we incorporate a correction term during the extrapolation process, which we illustrate in Figure 4.6.5. Using the depicted sampling rays as an example, let us suppose that we know the wavefront's location on ray 1, marked by point A, and we are extrapolating upward to find the wavefront's position on ray 2. Computing the tangent of the wavefront from the slope of ray 1 at point A and extrapolating to ray 2 yields the wavefront location marked by point B in the figure. We then use the wavefront's tangent obtained from the slope of ray 2 at point B to extrapolate from the mid-point between A and B, denoted by point C, to ray 2. Doing so brings us to point D, which we take to be the wavefront's position on ray 2. We use this extrapolation technique to derive each sampling wavefront in $X_j(y)$ from the sampling trajectories in $Y_i(x)$.

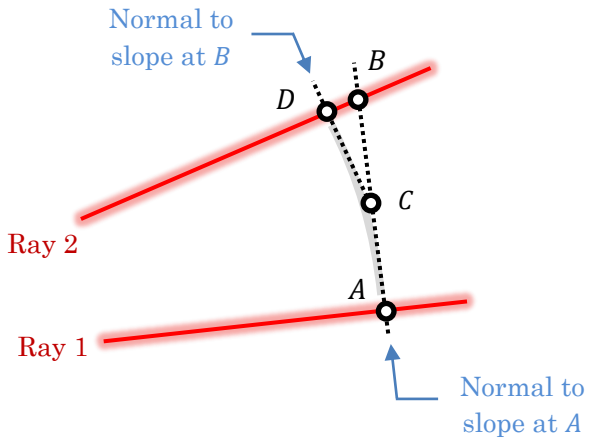


Figure 4.6.5. Correction to linear extrapolation of geometrical wavefronts.

Upon fully specifying $X_j(y)$, there remains one more piece of information we need to solve for the refractive index distribution. In order to narrow the design down to a single solution, OPL values must be assigned to each wavefront in $X_j(y)$ so that the quantity $d\phi$ or dS is specified in Equation 4.6.2. These values can be chosen arbitrarily provided that the OPL increases monotonically as we move in the $+x$ direction along any ray trajectory. In Appendix 4, we demonstrate how to optimize the individual OPL values for the purpose of minimizing the refractive index range of the calculated solution. For the time being, however, treating the OPL values as variable parameters in the design will suffice. Assuming that the arc length between adjacent sampling wavefronts along the ray trajectories is small, the index in Equation 4.6.2 can be approximated as

$$n(\vec{r}') = \frac{\phi_{i,j+1} - \phi_{i,j}}{\|\vec{r}_{i,j+1} - \vec{r}_{i,j}\|}, \tag{4.6.4}$$

where the Euclidean norm is taken in the denominator, and subscripts i, j denote the intersection between the i^{th} sampling ray and the j^{th} sampling wavefront. ϕ is the optical path length at the intersection, and $\vec{r}'$ is the intermediate point along the i^{th} sampling ray between adjacent wavefronts j and $j+1$ given by

$$\vec{r}' = \frac{1}{2}\left(\vec{r}_{i,j+1} + \vec{r}_{i,j}\right). \tag{4.6.5}$$

We compute Equation 4.6.4 for every intermediate point located between all pairs of adjacent wavefronts along each sampling ray. The refractive index values at the intermediate points are then used to interpolate the index distribution on a uniform rectangular grid, using the moving least squares approximation from Appendix 1, for instance. Recognizing that sparse ray distributions inevitably result in sparse sample points for constructing the interpolant, the characterization of the mode converter in Figure 4.6.4 is preferred over that of Figure 4.6.3, where the intermediate points provide a more uniform coverage of the index distribution in the former, thereby reducing errors resulting from the interpolation process. Figure 4.6.6 shows the resulting refractive index distribution for the mode conversion specified in Figure 4.6.1, where ϕ is chosen in the design such that the axial refractive index remains constant at $n(0) = 1.5$ RIU.

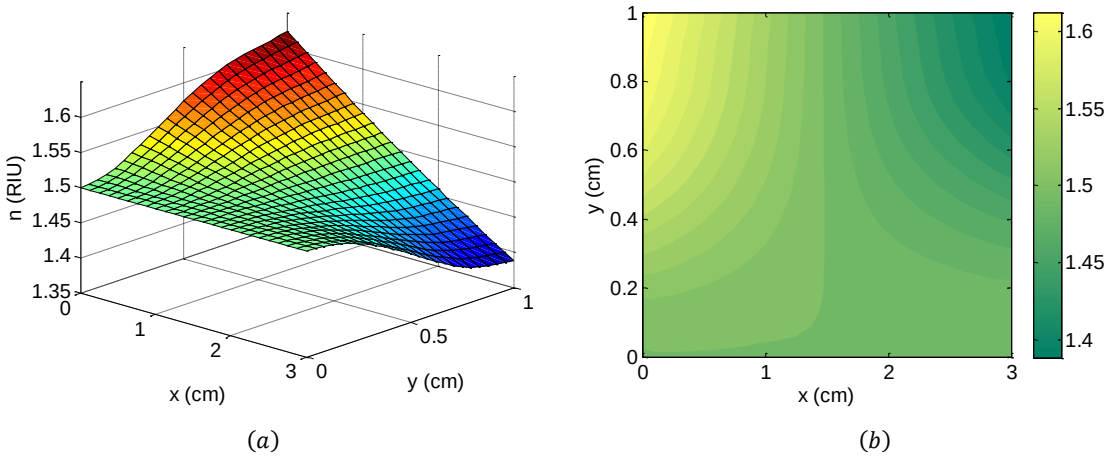


Figure 4.6.6. Index profile of the mode converter depicted as (a) a 3-D mesh plot and (b) a 2-D contour plot.

A ray trace was subsequently performed for the index distribution for an incident optical beam, as shown in Figure 4.6.7a. After assigning radiant energies to the individual rays to mimic the irradiance profile of the input Gaussian beam specified in the design, the traced ray distribution was converted into an output irradiance profile, which we plot in Figure 4.6.7b, to check agreement with the desired mode conversion. Note that in the figure, the launch positions of the ray trace are uniformly spaced along y in the

input plane. Therefore, the radiant energies of the individual rays in the input distribution is given as a Gaussian function of their initial distance from the central axis ($y = 0$) in the transverse plane.

The methodology we presented in this section is fully extendable to radial GRIN elements, where the refractive index varies as a function of the radial distance r from the central axis. The only difference lies in the coordinate transformation from a 2-D transverse profile to a 1-D radial profile. In other words, the optical beam's irradiance profile in radial coordinates must account for the coordinate transformation in order to satisfy energy conservation. Aside from this minor difference, the rest of the design procedure remains completely identical.

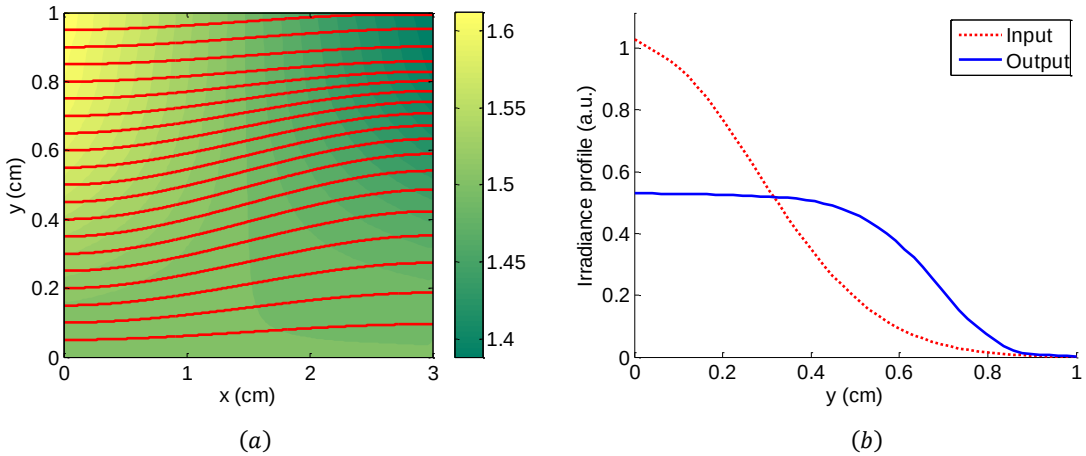


Figure 4.6.7. (a) Ray trace through the mode converter and (b) beam irradiance profiles calculated from input and output ray distributions.

4.7 Gaussian-to-Flat-Top GRIN Resonators

We took a detour in the previous section to develop a numerical procedure for designing GRIN mode converters. In this section, we will focus on analyzing the eigenmodes of a flat-mirror resonator containing these optical elements. It has been shown that a phase-conjugate etalon mirror placed at one end of the optical resonator can be used to shape its eigenmodes and promote spatial-mode selection in the resonator [Yang & Leger, 2004]. Naturally, an intracavity GRIN mode converter can be used to fill the same role. With or without the use of GRIN elements, Gaussian-to-flat-top mode conversion offers numerous advantages in resonator design [Litvin & Forbes, 2009]. For instance, placing the gain medium in the region of the resonator where the mode resembles a flat-top enables the mode to fill a larger mode volume due to the steep drop-off in irradiance at the edges of the optical beam; this allows for uniform gain saturation and high energy extraction of the gain medium. In addition, the mode's irradiance profile is more evenly distributed compared to a Gaussian beam of comparable size and equal energy, which helps mitigate practical issues such as thermally induced stress fracturing of optical materials and thermal aberrations. We propose to implement the mode converter inside the resonator using GRIN materials in lieu of diffractive

optical elements (DOEs), eliminating the need for complex phase functions and Fourier transform lenses (which were employed by Litvin and Forbes). Another advantage in using GRIN elements to perform the beam shaping is that they can do so over much shorter distances and allow for arbitrary mirror geometries in the resonator design.

We began our analysis by specifying the coherent mode conversion, which we illustrate in Figure 4.7.1, and calculating the index distribution following the methodology developed in Section 4.6. Note that the input and output irradiance profiles of the optical beam are more confined toward the central axis compared to the mode conversion specified in Figure 4.6.1a. The reason behind this is because our beam propagation method assumes energy conservation between the incident and propagated field disturbances. Hence, beam confinement in the desired mode conversion was chosen so that the diffraction of the resonator modes does not leak radiant energy outside the simulation space. The beam is collimated in both the input and output planes in the design; the initial beam width of the incident Gaussian beam was reduced to $\sigma = 0.4$ cm, and the field amplitude of the output order-6 super-Gaussian beam was chosen so that it falls to e^{-1} times its peak value at $y = 0.7$ cm. The calculated refractive index distribution of the mode converter is shown Figure 4.7.1b.

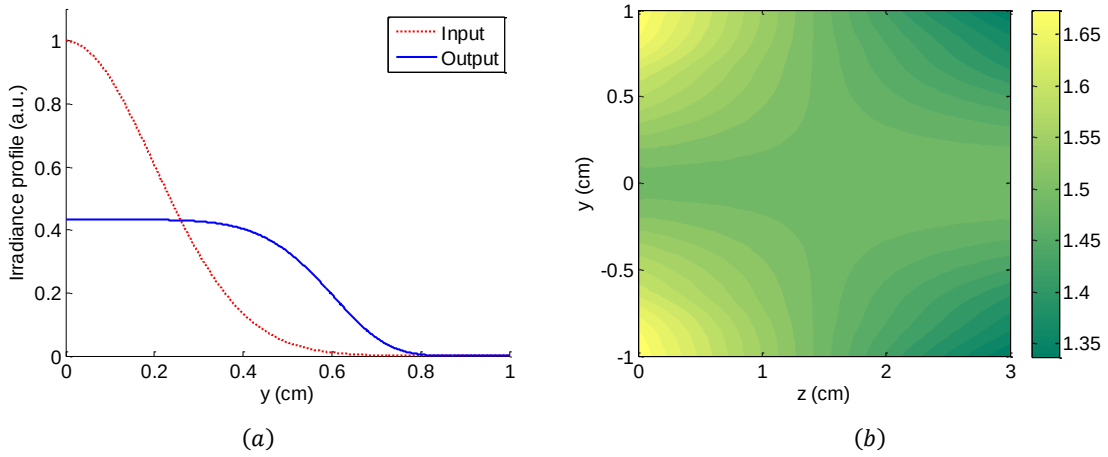


Figure 4.7.1. (a) Beam irradiance profiles for the desired mode conversion and (b) calculated refractive index distribution of the mode converter.

We then characterized the calculated refractive index distribution for forward beam propagation in 5 tandem partitions (see Section 4.4) for a vacuum wavelength of $\lambda_0 = 0.05$ cm. This particular value of λ_0 was chosen to increase diffractive effects in hopes of promoting mode-selection for the resonator while keeping the optical beam confined within the simulation space. Figure 4.7.2 shows the propagated irradiance profile at the output plane for an incident Gaussian beam whose beam waist coincided with the input plane, with a beam width of $\sigma = 0.4$ cm. The discrepancies between the propagated field disturbance and the desired mode at the output in the figure can be attributed to the diffraction of the optical beam.

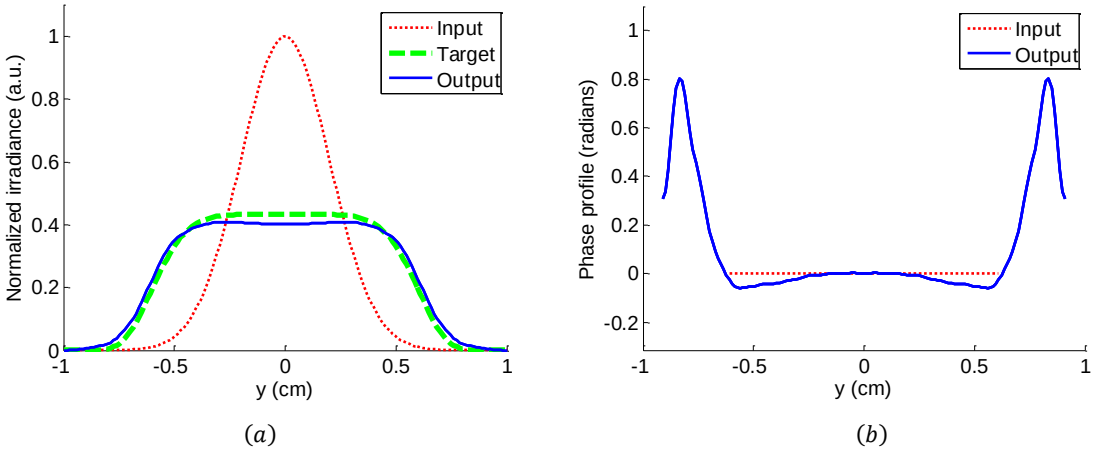


Figure 4.7.2. The optical beam's (a) irradiance and (b) phase profile for a Gaussian beam propagated from $z = 0$ *to* $z = 3$ *cm inside the mode converter.*

Because the calculated index distribution in Figure 4.7.1b exhibits an asymmetrical dependence on z, backward beam propagation through the GRIN element from the output plane to the input plane must be characterized separately. This is accomplished using an identical characterization procedure after flipping the index distribution horizontally along z. Figure 4.7.3 shows the propagated irradiance profile at the input plane for a super-Gaussian beam incident upon the output plane of the GRIN element and propagating in the $-z$ direction, whose initial beam profile matches the desired output mode. As before, discrepancies arise due to diffractive effects.

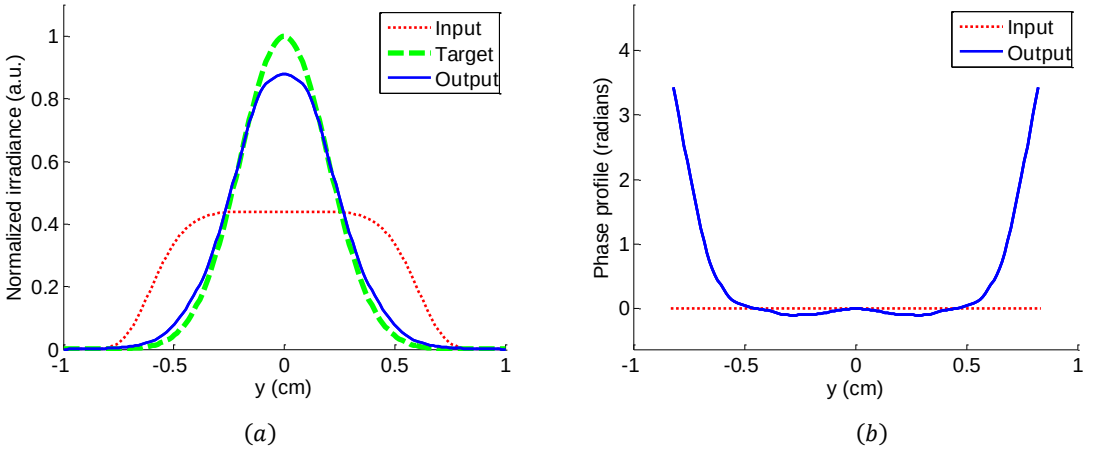


Figure 4.7.3. The optical beam's (a) irradiance and (b) phase profile for a super-Gaussian beam propagated from $z = 3$ *cm to* $z = 0$ *inside the mode converter.*

The deviations from a constant phase seen in the propagated beam's phase profiles in both figures are much less than $\lambda/4$ over the central region of the beam that contains substantial power. Hence, the wavefront at these locations can, for our intents and purposes, be considered to be approximately planar. In order to provide a better visualization of beam propagation inside the GRIN element, we plot the evolution of the optical beam's irradiance along z in Figure 4.7.4.

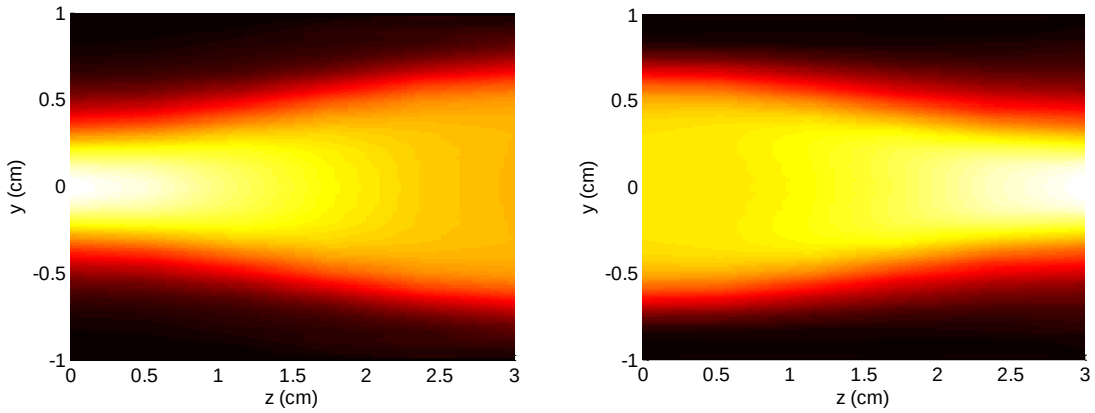


Figure 4.7.4. Evolution of the optical beam's irradiance profile in the (a) forward propagation and (b) backward propagation through the GRIN mode converter.

If we were to place perfectly reflecting flat mirrors at both ends of the GRIN element to make a resonator, then it is reasonable to expect one of the resonator's eigenmodes to bear a strong resemblance to the optical beam specified in the desired mode conversion. Provided that diffraction is negligible, it stands to reason that such a beam will be able to reproduce itself after one round trip through the resonator. Of course, the significant diffractive effects evident in the preceding figures will prevent this from happening. However, a similar beam that can reproduce itself in the resonator will take its place as an eigenmode. To calculate the resonator's eigenmodes, we employ the Fox-Li method detailed in Section 4.5. Because the method relies on the difference in the round-trip losses of the resonator's eigenmodes, a loss mechanism must be introduced to the resonator structure in order to provide modal selection. In our hypothetical resonator structure, we placed a thin absorber with amplitude transmittance

$$t(y) = e^{-\frac{1}{2}\left(\frac{y}{\rho}\right)^6}, \tag{4.7.1}$$

where $\rho = 0.6$ cm, against the output surface of the GRIN element, as shown in Figure 4.7.5. Since the defined simulation space only encompasses $|y| \leq 1$ cm and $0 \leq z \leq 3$ cm, it can be inferred that $d = 2$ cm and $L = 3$ cm in the figure.

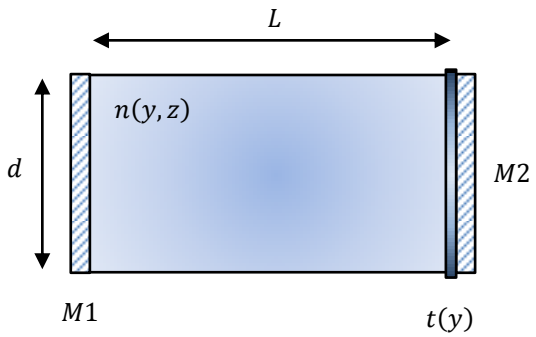


Figure 4.7.5. GRIN mode converter resonator structure.

By inspection, it is clear that any optical beam passes through the absorber twice during each round-trip through the resonator: once as the beam propagates toward mirror $M2$ and once more after it is reflected from $M2$. In our analysis, we will describe the eigenmodes of the resonator as field disturbances at the right end of the cavity ($z = 3$ cm), before and after it passes through the absorber; doing so has the advantage of allowing us to visualize the modal selection of the resonator due to the absorber. Hence, the round-trip operator $\boldsymbol{F}$ through the resonator entails the backward propagation (from right to left) of a field disturbance through the GRIN element, followed by its forward propagation (from left to right) through the GRIN element, and finally its (double) transmission through the absorber. We note that the reflections from the ideal flat mirrors can be ignored in the mathematics provided that we account for these reflections in the resonator model after it has been unfolded into an infinitely long periodic structure (similar to Figure 4.5.1).

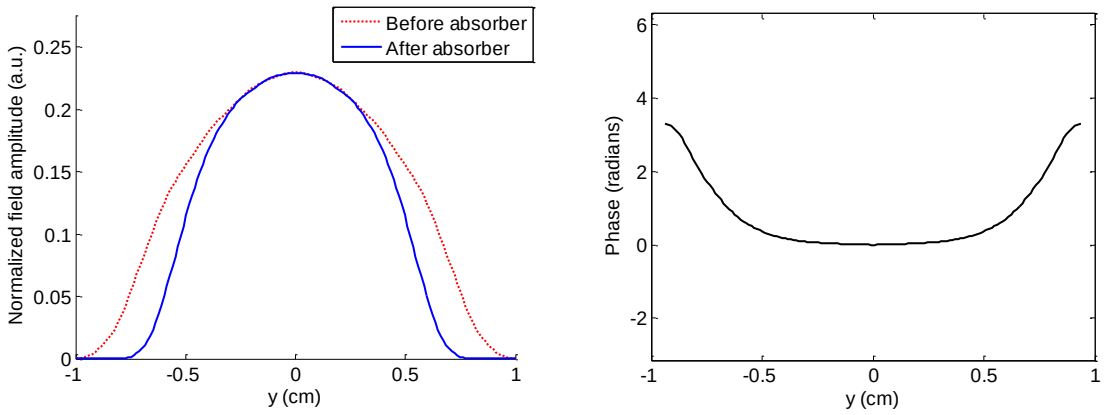
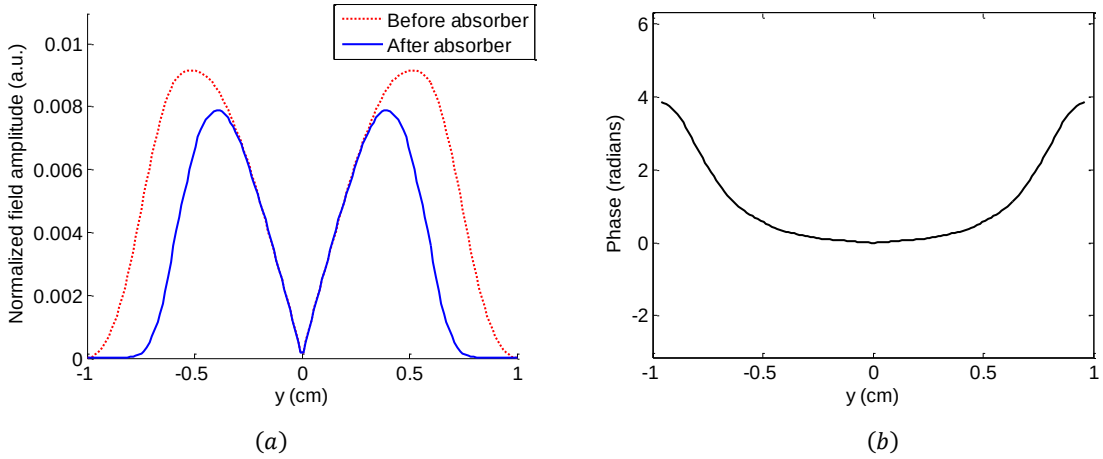


Figure 4.7.6. (a) Irradiance and (b) phase profile of the GRIN resonator's fundamental spatial eigenmode.

Applying the Fox-Li method to the resonator structure, we calculated the round-trip losses of its fundamental mode to be 18.97%. The spatial beam profile of the fundamental mode is depicted in Figure

4.7.6. We emphasize that although the refractive index distribution of the GRIN mode converter controls the propagation of light inside the resonator cavity, the resonator's eigenmodes are ultimately determined by the round-trip operator $\boldsymbol{F}$, which includes the thin absorber we placed inside the resonator. This explains why the fundamental mode is not in agreement with the desired output beam profile from Figure 4.7.1a. It also stands to reason that the latter cannot be an eigenfunction of $\boldsymbol{F}$ due to the beam shaping effects of the absorber. To calculate the next eigenmode of the resonator, we modified the initial field disturbance in the ring-down process to exhibit odd symmetry in y (*e.g.*, multiplying it by y). The calculated eigenmode is depicted in Figure 4.7.7 and has a round-trip loss equal to 53.89%.



(a) (b)

Figure 4.7.7. (a) Irradiance and (b) phase profile of the GRIN resonator's secondary spatial eigenmode.

We express the modal discrimination between the first two eigenmodes on a logarithmic scale as

$$\delta \equiv -10log_{10}\left(\frac{\gamma_1}{\gamma_2}\right), \tag{4.7.2}$$

where γ_1 and γ_2 are the round-trip losses for the fundamental mode shown in Figure 4.7.6 and the eigenmode in Figure 4.7.7, respectively. Based on this definition, the round-trip losses for the first two eigenmodes in our resonator correspond to a modal discrimination of 4.53 dB. In comparison, conventional Fabry-Perot resonators with comparable Fresnel numbers are able provide a modal discrimination of around 5.1 dB. Although our hypothetical resonator falls short in this regard, it still offers numerous advantages as a resonator that employs Gaussian-to-flat-top mode conversion, which we discussed at the beginning of this section. More importantly, we were able to develop a general methodology for analyzing resonator designs incorporating GRIN elements, whose refractive index distribution can be any arbitrary, smooth function of the spatial variables.

5. Conclusions & Future Work

5.1 Summary

In Chapter 2, we demonstrated a method for computing 1-D refractive index profiles in thick GRIN elements using deflectometry measurements. These measurements consisted of boundary values of ray position and slope outside the optical material. Our method was based on a fundamental relation between the ray slopes along the ray trajectory and the refractive indices at those locations, which follows directly from the ray equation of geometrical optics. The resulting expression contained no approximations apart from those necessary to neglect diffractive effects. In order to compute the refractive index profile from the deflectometry data under practical interrogation schemes, we introduced a bootstrap algorithm for recovering the index in successive regions of the overall index profile, starting from an initial location of known index. For comparison purposes, we used an existing method based on ray displacement measurements for calculating 1-D refractive index profiles in thick GRIN elements to compute the same index profiles as our ray deflection method. We improved upon the displacement method by introducing a correction term to the mathematical model and using an iterative ray trace algorithm to correct for intrinsic errors resulting from geometrical approximations. Both methods were used to demonstrate the experimental measurement of a thick GRIN element provided by the Institute of Optics at the University of Rochester, and the calculated refractive index profiles from the two methods showed good agreement, with a maximum discrepancy of 1.5×10^{-3} RIU. In our error analysis, we derived an analytical expression for estimating the accumulation of error in the bootstrap algorithm (applied to the ray deflection method) and verified the expression with computer simulation results. The effects of similar measurement error for the ray displacement method were also studied in our simulations, where the iterative ray trace algorithm was used to correct for inconsistencies in the calculated refractive index.

In Chapter 3, we demonstrated a method for recovering 2-D refractive index profiles in rectangular GRIN elements using boundary values of ray position and slope. Using a simplifying assumption that decouples the trajectories of interrogating rays from the index field, we formulated a linear system description of the inverse problem and developed a mathematical model assuming knowledge of the boundary index values, which were used to obtain the internal boundary ray slopes for the inversion process. In our model, the unknown ray trajectories inside the medium were approximated from the internal boundary values so that each deflectometry path integral could be expressed as a linear combination of the system's unknowns, *i.e.*, the Cartesian vector components of the index gradient field sampled on a uniform rectangular grid. We solved for these unknowns using a numerical inversion technique suited for solving sparse and ill-conditioned linear systems. Upon integrating the reconstructed gradient vector field, we were able to recover the refractive index field to within an arbitrary constant, which was determined through knowledge of the index along the boundaries. Using this approach, we demonstrated the reconstruction of a

2-D test index field from deflectometry data generated by a numerical ray trace and achieved RMS index errors below 0.5% of the index range. Taking advantage of the deflectometry system's resilience to slight variations in the ray trajectories, we developed an iterative ray trace procedure to improve the approximated trajectories in the inversion process and refine the mathematical description of the deflectometry system. The end result enforced the dependence of the ray trajectories on the index field so that the optimized solution is consistent with the deflectometry data (*i.e.*, the rays in the medium obey the ray equation). Using the same mathematical model, we were able to extend our computational method to external measurements of the boundary values using an iterative approach. However, knowledge of the index at some location is still needed to fully specify the index field. We hasten to add that the rectangular geometry of the test index field described in Chapter 3 was chosen simply to demonstrate the measurement procedure, and the method can readily be adapted to more complex geometries. In addition, we showed that the principles of our 2-D deflectometry method are fully extendable to the measurement of 3-D refractive index distributions. In order to assess the sensitivity of reconstruction process to measurement error in an experiment, we introduced artificial Gaussian noise to the simulated deflectometry data and observed three regimes in the noise level where a particular error mechanism was dominant in contributing to the overall reconstruction error. In addition, we examined the limitations of the numerical inversion technique by reconstructing the index field under varying amounts of data redundancy and coverage of the index gradient field provided by the interrogating rays. For known internal boundary ray slopes, the reconstruction error can be reduced to arbitrarily small values by increasing sampling resolution and/or using better interpolation techniques. The ultimate accuracy of the inversion technique is not fully understood if the internal boundary ray slopes are unknown due to the inherent inconsistencies that propagate through the retrieval algorithm and are manifested in the final result.

In Chapter 4, we developed a mathematical model based on the Fresnel-Kirchoff formulation for computing wave propagation in GRIN media. Using a numerical ray trace to model the propagation of spherical wavelets emitted by a discrete set of secondary sources representing the input field disturbance, we showed that the superposition of these wavelets could be used to determine the propagated field disturbance. We used this approach to calculate wave diffraction in free space and quadratic ducts and demonstrated a strong agreement between the propagated field disturbance obtained from our propagation method and well-known analytical solutions. Furthermore, we examined several numerical aspects of computation that give rise to practical issues in applying our method. We then developed a procedure for designing GRIN mode converters and placed a Gaussian-to-flat-top GRIN mode converter inside an optical resonator. Using our propagation method in conjunction with the Fox-Li method, we calculated the eigenmodes of the GRIN resonator and its modal discrimination. Although we failed to enhance the spatial mode-selection of the resonator with our design, the tools we developed in Chapter 4 are fully extendable to the analysis of optical resonators employing other intracavity GRIN elements.

5.2 Outlook for Future Work

A direct extension to our work on measuring 1-D refractive index profiles is to adapt the corrective ray trace algorithm used for the ray displacement method to the ray deflection method. Because the fundamental relation used in the ray deflection model is exact and does not contain any intrinsic errors associated with the geometrical assumptions made in the ray displacement model, we expect to achieve an even higher degree of accuracy in the reconstruction of the refractive index if the corrective procedure is applied to the former method.

There is still room for improvement in the mathematical model used to describe deflectometry path integrals inside 2-D refractive index distributions. In particular, it is not necessary to use the arc length ds as the variable of integration, as was needed for our analysis on a similar inverse problem based on OPL measurements. If we were to rewrite the expressions in terms of overall deflection in ray slope rather than ray angle and switch to x as the variable of integration, then the analog expressions for the deflectometry path integral become exact. To further reduce the quantization effects associated with the discrete representation of these path integrals, one can always use higher-order interpolation schemes to compute the linear coefficients associated with the system's unknowns for each interrogating ray. In specific cases where prior knowledge about the refractive index distribution can be exploited, Tikhonov regularization can be employed during the system inversion process to incorporate this information into our calculations.

We have yet to demonstrate an experimental measurement of a 2-D refractive index distribution. The need for a reliable method of automated data acquisition to meet the data redundancy requirements of accurate system inversion remains perhaps the biggest challenge in moving forward with an experimental demonstration. In principle, however, it is relatively straightforward to devise an experiment to verify our proposed computational method using the 1-D measurement methods from Chapter 2. For instance, we can make two different rectangular cuts from a large sample of GRIN material with a 1-D refractive index profile to obtain two smaller samples: one to use as a reference and the other for applying our 2-D deflectometry method. For the first cut, the boundary planes of the smaller sample are aligned either parallel or normal to the index gradient so that the resulting sample can be measured using the methods described in Chapter 2; the second cut is made at an angle with respect to the index gradient, as shown in Figure 5.1.1, so that the index gradient exhibits nonzero components along both Cartesian variables once the sample has been rotated to align its boundaries with the Cartesian axes. The resulting sample from the second cut can then be measured using our method described in Chapter 3, and a comparison with the 1-D index profile measured from the first sample can be made for verification purposes. Of course, practical issues associated with the experimental demonstration need to be considered. For instance, the restrictions imposed by total internal reflection at discontinuities in the refractive index along the boundaries of the GRIN element may require prism coupling or immersion in a fluid in order to gain access to specific internal angles. In addition, beam shaping effects are expected to introduce uncertainty to the measured boundary values of beam position and slope and thereby increase the index error in the reconstruction.

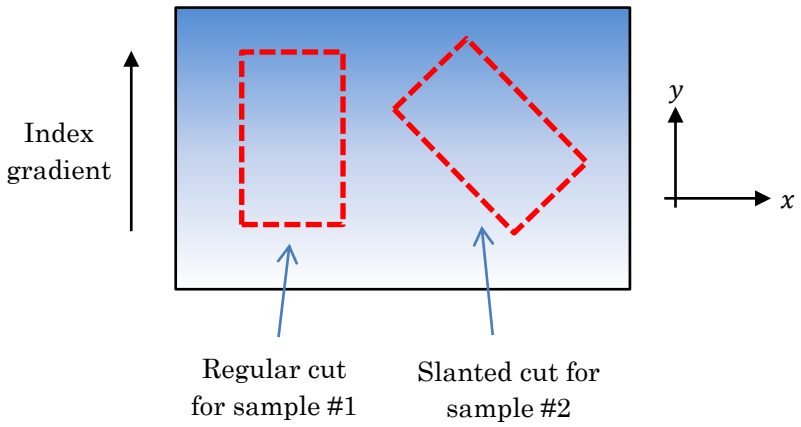


Figure 5.1.1. Alignment of cuts from a large GRIN sample with a 1-D index gradient.

An extension to our work on the application of GRIN elements inside optical resonators is the continued exploration of the resonator's spatial mode selection properties using different types of refractive index distributions for the intracavity GRIN element. The small number of refractive index distributions we have tested inside a flat-mirror resonator has been based on trial and error. Thus far, our designs have failed to enhance the resonator's modal discrimination between the first two eigenmodes. However, many aspects in the design of these GRIN resonators remain largely unexplored. For instance, the geometry of the end mirrors and the transmittance profile of the thin absorber can be altered to promote spatial mode selection. The spatial distribution of multiple thin absorbers placed inside the resonator cavity or even complex refractive index profiles that offer continuous, distributed loss mechanisms can also help in this regard. Furthermore, unstable GRIN resonator configurations that utilize gain-guiding principles are expected to achieve much better modal discrimination if high round-trip losses are acceptable. Another interesting, albeit more specific, extension to our work on GRIN resonators is to address the impact of design optimization for GRIN mode converters on the resonator's spatial mode-selection properties. Because the design procedure is based on geometrical optics, the modified index distribution after performing the optimization is equivalent to the original design under the ray approximation. However, the propagation of an optical beam is expected to behave differently in these designs when diffraction is taken into account. This information is especially useful when applied to the design of GRIN resonators, where subtle differences in the diffraction of the eigenmodes dictate the resonator's modal discrimination.

Finally, many computational aspects of our beam propagation method in GRIN media can still be improved. In particular, a large portion of the numerical artifacts that manifest in the round-trip operator are incurred from interpolation during the characterization process, *i.e.*, in computing a complete set of Green's functions at the propagated plane from ray trace results. Suppressing noise in the ring-down process is of paramount importance due to the exponential decay of signal amplitude in reaching a steady-state solution. The noise associated with numerical artifacts in the round-trip operator is repeatedly injected into the field

disturbance propagating inside the resonator and can easily overwhelm the diminishing signal. This is further exacerbated in cases where longer propagation distances are partitioned into shorter segments to avoid clipping in the characterized Green's functions. In order to mitigate error accumulation, large quantities of rays are needed to represent secondary wavelets in the characterization process; this improves the quality of interpolation by effectively reducing the separation between sample data points in the propagated plane. In addition, signal processing techniques can be utilized if there is no significant overlap between the noise and signal spectra; however, the application of signal processing requires due diligence since any spectral filtering performed during the ring-down process modifies the round-trip operator. Finally, improving the quality of the ray trace itself also warrants consideration. The combination of these improvements will likely render the characterization process extremely tedious and time-consuming, especially for GRIN elements whose transverse dimensions are much larger than the wavelength scale. Fortunately, the tremendous overhead cost associated with the characterization process can be reduced significantly by utilizing parallel computing with the help of graphics processors, for instance. Once the Green's functions have been determined, the remaining computations in our beam propagation method consist entirely of standard matrix multiplications.

6. References

Barnard, A. J. and Ahlborn, B. (1975), "Measurement of refractive index gradients by deflection of a laser beam," *Am. J. Phys.* **43**, 573-574.

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J. M., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C. and Van der Vorst, H. (1994), *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition* (Philadelphia, PA: SIAM), ch. 2.

Beadie, G., Fleet, E., Rosenberg, A., Lane, P. A., Shirk, J. S., Kamdar, A. R., Ponting, M., Hiltner, A. and Baer, E. (2008), "Gradient index polymer optics," *Proc. SPIE* **7061**, 706113.

Beliakov, G. and Buckley, S. (2003), "Reconstruction of the refractive index profile of planar waveguides using ray tracing analysis," *Proc. Optical Networking: Technologies, Traffic Engineering and Management,* Melbourne, Australia.

Beliakov, G. and Chan, D. Y. C. (1997), "Analysis of inhomogeneous optical systems by the use of ray tracing. I. Planar systems," *Appl. Opt.* **36**, 5304-5309.

Bodnar, Z. and Ratajczyk, F. (1965), "Some remarks concerning optical glass heterogeneity measurement with the help of the autocollimation method," *Appl. Opt.* **4**, 351-354.

Born, M. and Wolf, E. (1965), *Principles of Optics* (London, UK: Pergamon Press), ch. 8.

Cagniot, E., Derrar-Kaddour, Z., Fromager, M., Ait-Ameur, K. (2008), "Improving both transverse mode discrimination and diffraction losses in a plano-concave cavity," *Opt. Commun.* **281**, 4449-4454.

Cagniot, E., Fromager, M., Godin, T., Traïche, M., Passilly, N., Päivänranta, B. and Ait-Ameur, K. (2010), "Cascades of π-phase plates: a transparent diffractive focusing system," *J. Opt. Soc. Am. A* **27**, 1647-1654.

Carpenter, W. C. and Gill, P. A. T. (1973), "Automated solutions of time-dependent problems," in *The Mathematics of Finite Elements and Applications*, Ed. J. R. Whiteman, (London, UK: Academic Press), pp. 394-506.

Casperson, L. W. and Lunnam, S. D. (1975), "Gaussian modes in high loss laser resonators," *Appl. Opt.* **5**, 1193-1199.

Cha, S. and Vest, C. M. (1981), "Tomographic reconstruction of strongly refracting fields and its application to interferometric measurement of boundary layers," *Appl. Opt.* **20**, 2787-2794.

Choi, B. K., Shin, H. Y., Yoon, Y. I. and Lee, J. W. (1988), "Triangulation of scattered data in 3D space," *Comput. Aided Des.* **20**, 239-248.

Chu, P. L. (1977), "Nondestructive measurement of index profile of an optical-fibre preform," *Electron. Lett.* **13**, 736-738.

Courant, R. and Hilbert, D. (1991), *Methods of Mathematical Physics, Vol. 1* (New York, NY: Wiley & Sons), ch. 5.

Elmore, W. C. and Heald, M. A. (1985), *Physics of Waves* (Mineola, NY: Dover Publications, Inc., 1985), ch. 9.

Fox, A. G. and Li, T. (1961), "Resonant modes in a maser interferometer," *Bell Syst. Tech. J.* **40**, 453-488.

Freund, R. W. and Nachtigal, N. M. (1991) "QMR: A quasi-minimal residual method for non-Hermitian linear systems," *SIAM Journal: Numer. Math.* **60**, 315-339.

Gao, H., Takahashi, S., Tian, L. and Barbastathis, G. (2011), "Aperiodic subwavelength Luneburg lens with nonlinear Kerr effect compensation," *Opt. Express* **19**, 2257-2265.

Gasilov, S., Mittone, A., Brun, E., Bravin, A., Grandl, S., Mirone, A. and Coan, P. (2014), "Tomographic reconstruction of the refractive index with hard x-rays: an efficient method based on the gradient vector-field approach," *Opt. Express* **22**, 5216-5226.

Goodman, J. W. (2005), *Introduction to Fourier Optics* (Greenwood Village, CO: Roberts & Co. Publishers), ch. 4.

Gori, F. (1994), "Flattened Gaussian beams," *Opt. Commun.* **107**, 335-341.

Iwata, K. and Nagata, R. (1970) "Calculation of three-dimensional refractive index distribution from interferograms," *J. Opt. Soc. Am.* **60**, 133-135.

Ji, S., Yin, K., Mackey, M., Brister, A., Ponting, M. and Baer, E. (2013), "Polymeric nanolayered gradient refractive index lenses: technology review and introduction of spherical gradient refractive index ball lenses," *Opt. Eng.* **52**, 112105.

Junginger, H. G. and Van Haeringer, W. (1972), "Calculation of three-dimensional refractive-index field using phase integrals," *Opt. Commun.* **5**, 1-4.

Kermene, V., Saviot, A., Vampouille, M., Colombeau, B. and Froehly, C. (1992), "Flattening of the spatial laser beam profile with low losses and minimal beam divergence," *Opt. Lett.* **17**, 859-861.

Kovalskii, L. V. and Polanskii, V. K. (1966), "Application of the prism method for investigating transparent media with a refractive index gradient," *Opt. Spectrosc.* **20**, 408-409.

Kruger, C. J. C (2002), "Constrained cubic spline interpolation for chemical engineering applications," retrieved from http://www.korf.co.uk/spline.pdf.

Lavigne, P., McCarthy, N. and Demers, J.-G. (1985), "Design and characterization of complementary Gaussian reflectivity mirrors," *Appl. Opt.* **24**, 2581-2586.

Leger, J. R., Chen, D. and Mowry, G. (1995), "Design and performance of diffractive optics for custom laser resonators," *Appl. Opt.* **34**, 2498-2509.

Lin, D. and Leger, J. R. (2012), "Numerical gradient-index design for coherent mode conversion," *Adv. Opt. Technol.* **1**, 195-202.

Lin, D. Leger, J. R., Kunkel, M. and McCarthy, P. (2013), "One-dimensional gradient-index metrology based on ray slope measurements using a bootstrap algorithm," *Opt. Eng.* **52**, 112108.

Litvin, I. A. and Forbes, A. (2009), "Gaussian mode selection with intracavity diffractive optics," *Opt. Lett.* **34**, 2991-2993.

Liu, J. H., Yang, P. C. and Chiu, Y. H. (2006), "Fabrication of high-performance, gradient-refractive-index plastic rods with surfmer-cluster-stabilized nanoparticles," *J. Polym. Sci. A Polym. Chem.* **44**, 5933-5942.

Liu, J. H. and Chiu, Y. H. (2009), "Process equipped with a sloped UV lamp for the fabrication of gradient-refractive-index lenses," *Opt. Lett.* **34**, 1393-1395.

Lohmann, A. W., Dorsch, R. G., Mendlovic, D, Zalevsky, Z. and Ferreira, C. (1996), "Space-bandwidth product of optical signals and systems," *J. Opt. Soc. Am. A* **13**, 470-473.

Lutomirski, R. F. and H. T. Yura (1971), "Propagation of a finite optical beam in an inhomogeneous medium," *Appl. Opt.* **10**, 1652-1658.

Marchand, E. W. (1978), *Gradient Index Optics* (New York, NY: Academic Press), ch. 5.

McCarthy, P. (2015), "Gradient-index materials, designs, and metrology for broadband imaging systems," (Doctoral dissertation, University of Rochester), retrieved from http://hdl.handle.net/1802/29303.

Meemon, P., Yao, J., Lee, K.-S., Thompson, K. P., Ponting, M., Baer, E. and Rolland, J. P. (2013), "Optical coherence tomography enabling non-destructive metrology of layered polymeric GRIN material," *Sci. Rep.* **3**, 1709.

Messerschmidt, B., Possner, T. and Goering, R. (1995), "Colorless gradient-index cylindrical lenses with high numerical apertures produced by silver-ion exchange," *Appl. Opt.* **34**, 7825-7830.

Miller, M. and Malisek, V. (1968), "Study of the distribution of refractive index by the differential prism method," *Opt. Spectrosc.* **25**, 70-72.

Mohr, R. K., Wilder, J. A., Macedo, P. B. and Gupta, P. K. (1979), "Graded index lenses by the molecular stuffing process," paper WA-1, *Technical Digest on Gradient-index Optical Imaging Systems*, Rochester, New York.

Moore, D. T. and Ryan, D. P. (1978), "Measurement of the optical properties of gradient index materials," *J. Opt. Soc. Am.* **68**, 1157-1166.

Moore, D. T. (1980), "Gradient-index optics: a review," *Appl. Opt.* **19**, 1035-1038.

Nishidate, Y., Nagata, T., Morita, S.-Y. and Yamagata, Y. (2011), "Ray-tracing method of isotropic inhomogeneous refractive-index media from arbitrary discrete input," *Appl. Opt.* **50**, 5192-5199.

Nishijima, Y. and Oster, G. (1964), "Moiré patterns: their application to refractive index and refractive index gradient measurements," *J. Opt. Soc. Am.* **54**, 1-5.

Ohmi, S., Sakai, H., Asahara, Y., Nakayama, S., Yoneda, Y. and Izumitani, T. (1988), "Gradient-index rod lens made by a double ion-exchange process," *Appl. Opt.* **27**, 496-499.

Oron, R., Davidson, N., Friesem, A. A. and Hasman, E. (2001), *Progress in Optics Volume 42* (Minsk, BY: Elsevier Science B. V.), ch. 6.

Oster, G., Wasserman, M. and Zwerling, C. (1964), "Theoretical interpretation of moiré patterns," *J. Opt. Soc. Am.* **54**, 169-175.

Otis, G., Lachambre, J.-L. and Lavigne, P. (1979), "Focusing of laser beams by a sequence of irises," *Appl. Opt.* **18**, 875-883.

Paré, C. and Bélanger, P.-A. (1992), "Custom laser resonator using graded-phase mirrors," *IEEE J. Quantum Electron.* **28**, 355-362.

Paige, C. C. and Saunders, M. A. (1982), "LSQR: an algorithm for sparse linear equations and sparse least squares," *AMC Trans. Math. Soft.* **8**, 43-71.

Paige, C. C. and Saunders, M. A. (1975), "Solution of sparse indefinite systems of linear equations," *SIAM J. Numer. Anal.***12**, 617-629.

Palma, C. and Bagini, V. (1994), "Propagation of super-Gaussian beams," *Opt. Commun.* **111**, 6-10.

Pickering, M. A., Taylor, R. L. and Moore, D. T. (1986), "Gradient infrared optical material prepared by a chemical vapor deposition process," *Appl. Opt.* **25**, 3364-3372.

Poon, T.-C. and Kim, T. (2006), *Engineering Optics with Matlab* (Singapore, SG: World Scientific), ch. 1.

Rhodes, P. W. and Shealy, D. L. (1980), "Refractive optical systems for irradiance redistribution of collimated radiation: their design and analysis," *Appl. Opt.* **19**, 3545-3553.

Richerzhagen, B. (1996), "Finite element ray tracing: a new method for ray tracing in gradient-index media," *Appl. Opt.* **35**, 6186-6189.

Saunders, J. B. (1969), "An interferometer for measuring gradients in both refractive index and thickness of large or small optics," *J. Res. Nat. Bur. Stand.* **63**, 1-4.

Sharma, A., Kumar, D. V. and Ghatak, A. K. (1982), "Tracing rays through graded-index media: a new method," *Appl. Opt.* **21**, 984-987.

Siegman, A. E. (1986), *Lasers* (Mill Valley, CA: University Science Books), ch. 22.

Siegman, A. E. (2007), "Gain-guided, index-antiguided fiber lasers," *J. Opt. Soc. Am. B.* **24**, 1677-1682.

Sinai, P. (1971), "Correction of optical aberrations by neutron irradiation," *Appl. Opt.* **10**, 99-104.

Sonneveld, P. (1989), "CGS: A fast Lanczos-type solver for nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.***10**, 36-52.

Southwell, W. H. (1982), "Ray tracing in gradient-index media," *J. Opt. Soc. Am.* **72**, 908-911.

Sweeney, D. W. and Vest, C. M. (1973), "Reconstruction of three-dimensional refractive index fields from multidirectional interferometric data," *Appl. Opt.* **12**, 2649-2664.

Teichman, J., Holzer, J., Balko, B., Fisher, B. and Buckley, L. (2013) "Gradient index optics at DARPA," retrieved from https://www.ida.org/~/media/Corporate/Files/Publications/IDA_Documents/STD/D-5027-FINAL.pdf.

Teichman, J. (2013) "Measurement of gradient index materials by beam deflection, displacement or mode conversion," *Opt. Eng.* **52**, 112112.

Urness, A. C., Anderson, K., Ye, C., Wilson, W. L. and Mcleod, R. R. (2015), "Arbitrary GRIN component fabrication in optically driven diffusive photopolymers," *Opt. Express* **23**, 264-273.

Wang C. and Shealy, D. L. (1993), "Design of gradient-index lens systems for laser beam reshaping," *Appl. Opt.* **32**, 4763-4769.

Wiener, O. (1893), "Darstellung gekrummter Lichtstrahlen und Verwerthung derselben zur Untersuchung von Diffusion und Warmeleitung," *Ann. Phys. Chem.* **49**,105-149.

Wu, S. P., Nihei, E. and Koike, Y. (1996), "Large radial graded-index polymer," *Appl. Opt.* **35**, 28-32.

Yang, Z. H. and Leger, J. R. (2004), "Properties of a phase-conjugate etalon mirror and its application to laser resonator spatial-mode control," *Appl. Opt.* **43**, 4095-4099.

Yariv, A. and Yeh, P. (2007), *Photonics: Optical Electronics in Modern Communications* (New York, NY: Oxford University Press), ch. 2.

Youcef, S. and Schultz, M. H. (1986), "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.***7**, 856-869.

Yuasa, T., Maksimenko, A., Hashimoto, E., Sugiyama, H., Hyodo, K., Akatsuka, T. and Ando, M. (2006), "Hard-x-ray region tomographic reconstruction of refractive-index gradient vector field: imaging principles and comparisons with diffraction-enhanced-imaging-based computed tomography," *Opt. Lett.* **31**, 1818-1820.

Zahreddine, R. N., Lepkowicz, R. S., Bunch, R. M., Baer, E. and Hiltner, A. (2008), "Beam shaping system based on polymer spherical gradient refractive index lenses," *Proc. SPIE* **7062**, 706214.

Zhou, J. Y., Fu, C. H., Lu, Z. G., Li, Q. X. and Yu, Z. X. (1990), "Enhanced single transverse mode oscillation by an intracavity circular hollow glass waveguide," *Opt. Commun.* **81**, 385-390.

Appendix 1: Numerical Ray Tracing

We demonstrate in this appendix how to implement a numerical ray trace through a GRIN medium, given its refractive index distribution n. According to geometrical optics, ray trajectories inside a GRIN medium must satisfy Equation 2.2.1. It is possible to rewrite the expression as a second-order differential equation and decompose it into a pair of first-order differential equations; the first-order equations can then be solved numerically using Euler's method [Southwell, 1982]. These first-order differential equations are

$$\frac{d\vec{r}}{dt} = \vec{T}\,, \tag{A.1.1a}$$

$$\frac{d\vec{T}}{dt} = \frac{\nabla n}{n^3} - \frac{2}{n}\left(\vec{T}\cdot\nabla n\right)\vec{T}\,, \tag{A.1.1b}$$

where $dt = nds$ is the differential OPL, $\vec{T} = \frac{1}{n}[\hat{\imath}cos(\theta) + \hat{\jmath}sin(\theta)]$ is the ray's unit direction vector divided by the refractive index, and θ is the angle of the ray's direction vector with respect to the optical axis x. Let us suppose that the ray is launched at some location $\vec{r}_0$ with an initial direction specified by $\vec{T}_0$. After the ray travels an incremental value Δt in OPL, the new position $\vec{r}$ and the new $\vec{T}$ vector are given by

$$\vec{r}_1 = \vec{r}_0 + \Delta t\vec{T}_0\,, \tag{A.1.1a}$$

$$\vec{T}_1 = \vec{T}_0 + \Delta t\left[\frac{\nabla n}{n^3} - \frac{2}{n}\left(\vec{T}_0\cdot\nabla n\right)\vec{T}_0\right], \tag{A.1.2b}$$

where n and ∇n are both evaluated at $\vec{r}_0$. For all subsequent increments in the OPL along the ray trajectory, the position $\vec{r}$ and the $\vec{T}$ vector can be updated in similar fashion:

$$\vec{r}_{i+1} = \vec{r}_i + \Delta t\vec{T}_i\,, \tag{A.1.3a}$$

$$\vec{T}_{i+1} = \vec{T}_i + \Delta t\left[\frac{\nabla n}{n^3} - \frac{2}{n}\left(\vec{T}_i\cdot\nabla n\right)\vec{T}_i\right], \tag{A.1.3b}$$

where n and ∇n are both evaluated at $\vec{r}_i$. Equation A.1.3 provides a straightforward implementation for the ray trace based on Euler's method and produces accurate results if the size of the increment Δt is small. There are additional measures one can take to improve the accuracy of the ray trace, such as using Richardson's extrapolation [Carpenter & Gill, 1973] to optimize both numerical error and computational cost. The implementation expressed in Equation A.1.3 comes with the added benefit of providing automatic tracking for the total OPL traveled by the ray from its initial position. Of course, there are other methods of implementation that offer their own advantages in computing the ray trace [Sharma, *et al.*, 1982; Richerzhagen, 1996]. Figure A.1.1 illustrates the principle of Euler's method for numerical ray tracing in a

2-D refractive index distribution $n(x, y)$, which we represent in the figure by a discrete set of sample points $n_{l,k}$ on a uniform rectangular grid.

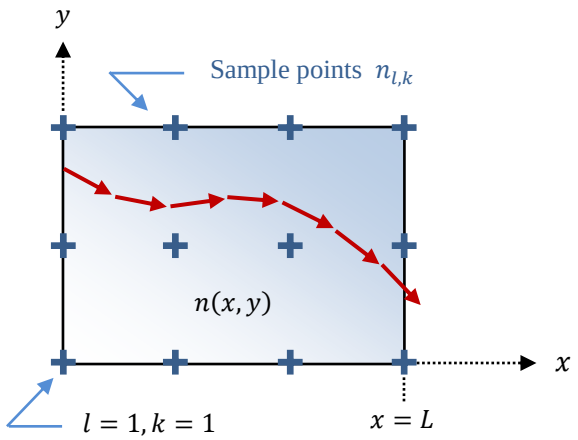


Figure A.1.1. Euler's method for computing a numerical ray trace.

To pinpoint the location where the ray makes contact with the exit plane, we compute the intersection between the last ray segment in the trace and the exit plane as follows: Approximating the ray trajectory between the calculated positions of successive increments in OPL as straight lines, as depicted in the figure, the contact point can be determined by first checking if the final ray position is out of bounds. If the ray trace terminates on $x = L$, there is no overshoot, and the final ray position is the contact point. If the trace ends beyond the $x = L$ plane, then the intersection is given by

$$\vec{r}_f = L\hat{\imath} + \left[y_{N-1} + \frac{L - x_{N-1}}{x_N - x_{N-1}}(y_N - y_i)\right]\hat{\jmath}\,, \tag{A.1.4}$$

where N is the number of steps in the ray trace, $\vec{r}_{N-1} = x_{N-1}\hat{\imath} + y_{N-1}\hat{\jmath}$ is the ray position at the end of the second-to-last step, and $\vec{r}_N = x_N\hat{\imath} + y_N\hat{\jmath}$ is the ray's final position. Likewise, similar formulae can be written for the contact points with the other boundary planes of the GRIN medium.

In practice, emphasis must be placed on the quality of interpolation used to obtain n and ∇n at $\vec{r}_i$ from the sample values $n_{l,k}$. In our implementation, we employ the moving least squares (MLS) approximation, which is a meshless technique that has been demonstrated in various GRIN profiles, *e.g.*, the Luneburg lens or a radial gradient optical fiber, to produce highly accurate raytrace results [Nishidata, *et al.*, 2011]. In the MLS approximation, the refractive index at position $\vec{r}$ is computed as

$$n(\vec{r}) = \vec{p}^{\,T}(\vec{r})\vec{a}(\vec{r})\,, \tag{A.1.5}$$

where $\vec{p}(\vec{r})$ is a column vector of basis functions and $\vec{a}(\vec{r})$ is the coefficient vector. There are many choices one can use for $\vec{p}(\vec{r})$. For instance, picking a two-dimensional quadratic basis yields

$$\vec{p}(\vec{r}) = \begin{bmatrix} 1 \\ x \\ y \\ x^2 \\ y^2 \\ xy \end{bmatrix}. \tag{A.1.6}$$

The coefficient vector $\vec{a}(\vec{r})$ is the solution to the following equation,

$$[M(\vec{r})]\vec{a}(\vec{r}) = [N(\vec{r})]\vec{n}\,, \tag{A.1.7}$$

where $\vec{n}$ is a column vector containing the refractive indices of the pertinent sample points. It is convenient to define a matrix $[P]$, with which we will use to compute the matrices $[M(\vec{r})]$ and $[N(\vec{r})]$ in Equation A.1.7,

$$[P] = \begin{bmatrix} \vec{p}^T(\vec{r}_1) \\ \vec{p}^T(\vec{r}_2) \\ \vdots \\ \vec{p}^T(\vec{r}_m) \end{bmatrix}. \tag{A.1.8}$$

where $\vec{p}(\vec{r}_j)$ is the vector of basis functions from Equation A.1.6 with the position parameter $\vec{r}$ replaced by the position of the j^{th} sample point $\vec{r}_j$, and m is the number of sample points contained within one radius of influence R_{inf} from the point of interest $\vec{r}$, which we illustrate in Figure A.1.2.

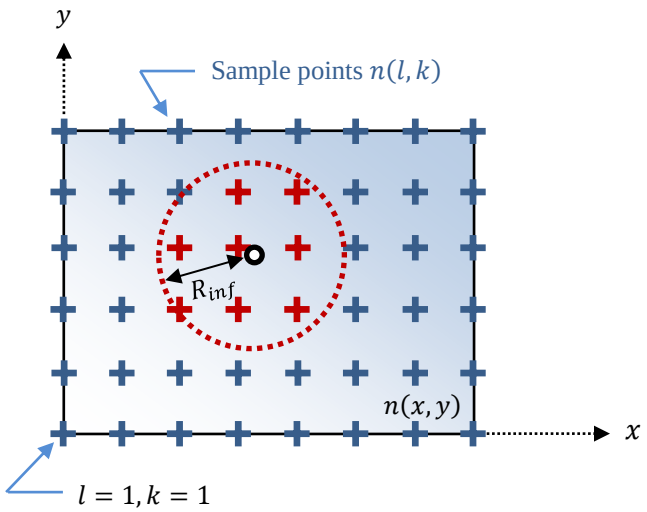


Figure A.1.2. Acquiring pertinent sample points contained within one radius of influence for MLS interpolation.

In addition, we need to define a diagonal matrix

$$[W(\vec{r})] = \begin{bmatrix} w(\vec{r}-\vec{r}_1) & 0 & \cdots & 0 \\ 0 & w(\vec{r}-\vec{r}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w(\vec{r}-\vec{r}_m) \end{bmatrix}. \tag{A.1.9}$$

where $w(\vec{r}-\vec{r}_j)$ is the weighting function. Like $\vec{p}(\vec{r}_j)$, there are many choices one can use for $w(\vec{r}-\vec{r}_j)$. One possibility is the quartic spline weight function

$$w(\vec{r}-\vec{r}_j) = 1 - \frac{6}{R_{inf}}(\vec{r}-\vec{r}_j)^2 + \frac{8}{R_{inf}}(\vec{r}-\vec{r}_j)^3 - \frac{3}{R_{inf}}(\vec{r}-\vec{r}_j)^4 , \tag{A.1.10}$$

where R_{inf} is the radius of influence depicted in Figure A.1.2. Finally, the matrices $[M(\vec{r})]$ and $[N(\vec{r})]$ in Equation A.1.7 are given as follows:

$$[M(\vec{r})] = [P]^T[W(\vec{r})][P] , \tag{A.1.11a}$$

$$[N(\vec{r})] = [P]^T[W(\vec{r})] . \tag{A.1.11b}$$

It is important to keep in mind that a poor choice of R_{inf} can lead to numerical instability during the MLS approximation. We find that values of $4\Delta < R_{inf} < 10\Delta$ are generally free of matrix conditioning issues in calculating $\vec{a}(\vec{r})$ in Equation A.1.7, where Δ is the grid spacing.

With accurate values of n and ∇n at $\vec{r}_i$ readily available from the MLS approximation, implementing the ray trace in Equation A.1.3 is straightforward. Starting with the initial launch conditions specified by $\vec{r}_0$ and $\vec{T}_0$, we proceed as follows:

1. Interpolate n and ∇n at the current ray position $\vec{r}$ using the MLS approximation.
2. Update the position vector $\vec{r}$ and the normalized direction vector $\vec{T}$.
3. If $\vec{r}$ is out of bounds, compute the contact point with boundary plane and terminate the ray trace. Otherwise, go back to from step 1.

To demonstrate the above procedure, we used it to perform a numerical ray trace through a simple radial refractive index distribution given by

$$n(r) = \sqrt{2.5 - r^2} , \tag{A.1.12}$$

where r is the radial distance from its center. Inside the medium, rays that are launched parallel to the optical axis z will follow a sinusoidal trajectory given by [Marchand, 1978]

$$r(z) = r_0 cos\left(\frac{z}{n_0}\right) , \tag{A.1.13}$$

where r_0 is the ray's initial radial position and n_0 is the refractive index at $r = r_0$. Refractive index distributions with analytical solutions for the ray trajectories, such as the one described in Equation A.1.12,

are especially useful for verifying the accuracy of a particular ray trace method. The ray trajectories for the given radial index distribution obtained from our implementation of the ray trace are shown in Figure A.1.3, where the periodicities $9.0mm$ (red), $9.5mm$ (magenta/dashed) and $9.75mm$ (blue/thick) correspond to the initial radial distances $0.6695mm$, $0.4625mm$ and $0.3034mm$, respectively. These trace values are in agreement with the analytical expression in Equation A.1.13.

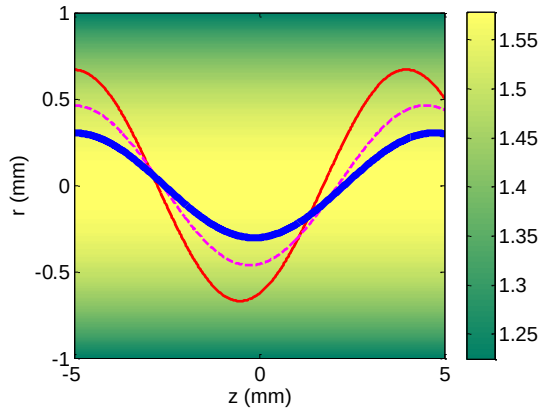


Figure A.1.3. Numerical ray traces through a simple radial refractive index profile with an analytical solution.

Appendix 2: Inversion from Known Trajectories

It is assumed in our methods for computing a 2-D refractive index distribution that the trajectories of the interrogating rays are not accessible by direct measurement. In this appendix, we describe a numerical method of computing the index distribution for known ray trajectories, *e.g.* from imaging scattered light in the medium. In addition, we provide a proof that shows the solution of the inverse problem to be non-unique for the same data set if the trajectories are not known.

To begin, suppose that we are given two distinct families of ray trajectories, $Y_1(x)$ and $Y_2(x)$, such that $Y_1(x) \neq Y_2(x)$. Provided that there exists a region of overlap where rays from one family intersect with rays from the other, the logarithmic index gradient components, $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$, are fully specified at the ray crossings. A simultaneous set of equations can be written for each ray crossing in the form

$$\begin{bmatrix} -y_1' & 1 \\ -y_2' & 1 \end{bmatrix} \begin{bmatrix} \partial w/\partial x \\ \partial w/\partial y \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \tag{A.2.1}$$

where $F_i = \frac{y_i''}{1+(y_i')^2}$ is termed the ray characteristic. Upon inverting the matrix on the left-hand side of the equation, the components of ∇w in the overlap region are given by

$$\begin{bmatrix} \partial w/\partial x \\ \partial w/\partial y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \tag{A.2.2}$$

where $\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{y_2'-y_1'} \begin{bmatrix} 1 & -1 \\ y_2' & -y_1' \end{bmatrix}$ is the inverse matrix. We note that the inverse matrix exists only if $y_2' \neq y_1'$; this is because the local ray curvature only yields information about the gradient component lying perpendicular to the ray trajectory, as is apparent from Equation 3.3.2. Therefore, a simple crossing between trajectories is needed to resolve dissimilar components of the logarithmic index gradient. Upon obtaining ∇w at the ray crossings between $Y_1(x)$ and $Y_2(x)$ using Equation A.2.2, the scattered gradient data can be used to interpolate ∇w onto a uniform rectangular grid so that 2-D integration can be performed to reconstruct the refractive index field. An integration technique for 2-D gradient fields specified on a rectangular grid is outlined in Section 3.5. Note that the overall integration constant in the reconstruction must be determined through additional measurements such as the refractive index at some location. An alternate method is to extrapolate the geometrical wavefronts for each ray family and reconstruct the index field for $Y_1(x)$ and $Y_2(x)$ individually. We provide a numerical procedure for computing an index distribution from a set of ray trajectories in Section 4.6. With minor adjustments, the same principles can be applied here. Rather than specifying the arbitrary constants along each wavefront in the design procedure, they can be optimized through the knowledge of ∇w. Although minor differences in the index will likely result within the overlapping region of the calculated index fields, they can be reconciled through

optimization of the arbitrary constants along each wavefront. Of course, the final solution obtained from this reconstruction method is still only specified up to an arbitrary constant, which can only be identified with additional measurements.

We will now provide a geometrical proof for non-uniqueness in the solution when the only information available consists of boundary value measurements of ray position and slope for $Y_1(x)$ and $Y_2(x)$. It follows that the question of uniqueness can be addressed through the geometrical principles developed in Section 4.6. Using a discrete set of geometrical wavefronts associated with each ray family, we show that it is possible for more than one refractive index distribution to produce the measured boundary values. The following proof is presented starting with a homogeneous 2-D index field for simplicity. However, the underlying principles are fully generalizable to any arbitrary slowly-varying index field. Let us consider a 2-D square element, with side lengths L, made of a homogeneous dielectric material. If we were to interrogate its refractive index using two sets of ray families that are normally incident upon adjacent sides of the square element, we would observe two sets of linear trajectories, as shown in Figure A.2.1. Note that the depicted interrogating geometry allows us to disregard the distinction between internal and external boundary values. Assuming the refractive index of the dielectric material is n_0, the OPL traveled by the interrogating rays across the length of the element in both directions is equal to $\varphi = n_0 L$.

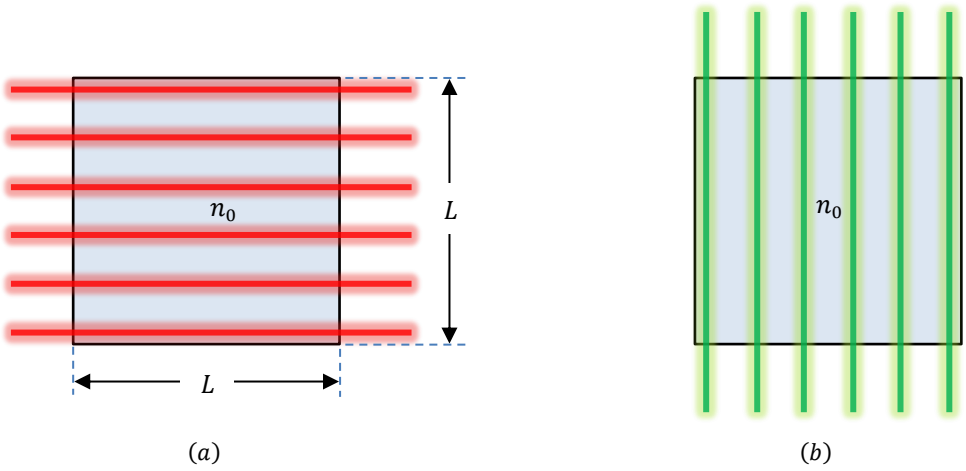


Figure A.2.1. Straight interrogating ray trajectories produced by a hypothetical homogeneous dielectric material for a family of (a) horizontally incident rays and (b) vertically incident rays.

Next, let us consider the geometrical wavefronts associated with the depicted sets of interrogating rays. Starting at $\varphi = 0$, we highlight the wavefronts whose OPLs correspond to integer multiples of $\Delta\varphi = \frac{1}{3} n_0 L$ and use these wavefronts to partition the homogeneous index field into thirds for each family of ray trajectories, as shown in Figure A.2.2.

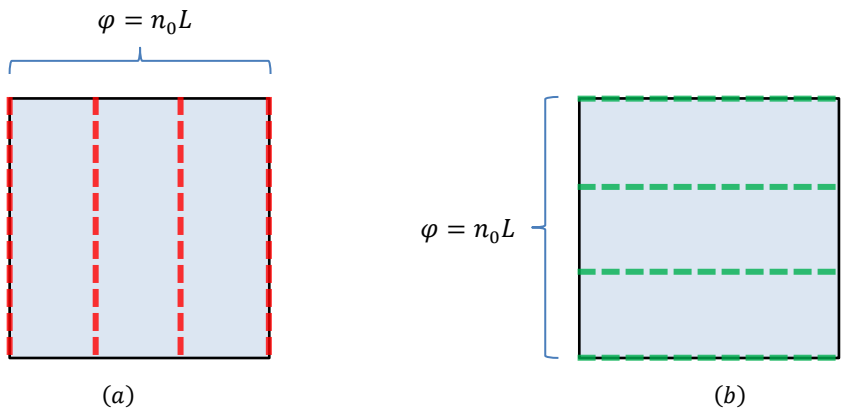


Figure A.2.2. Using geometrical wavefronts associated with (a) the horizontal interrogating rays and (b) the vertical interrogating rays to partition the index field into thirds.

Superimposing the horizontal and vertical partitions in the OPL, we effectively divide the index field into individual tiles, as shown in Figure A.2.3. If we treat the OPL difference $\Delta\varphi$ between any adjacent pair of parallel wavefronts as a differential increase in the Eikonal S and assume that the refractive index is constant within each tile, then the index responsible for the differential increase in S, or equivalently, the OPL difference $\Delta\varphi$, must satisfy Equation 2.1.8. Requiring the ray segments from both ray families inside each tile, depicted in Figure A.2.3, to satisfy this condition with matching refractive index values implies that the arc lengths along the ray segments between adjacent pairs of parallel wavefronts, labeled as Δs in the figure, must also be equal. This condition ensures consistency in the refractive index calculated for both the horizontal and vertical ray segments using Equation 2.1.8. This constraint applies to each individual tile and is satisfied independently of its surroundings under the given assumptions.

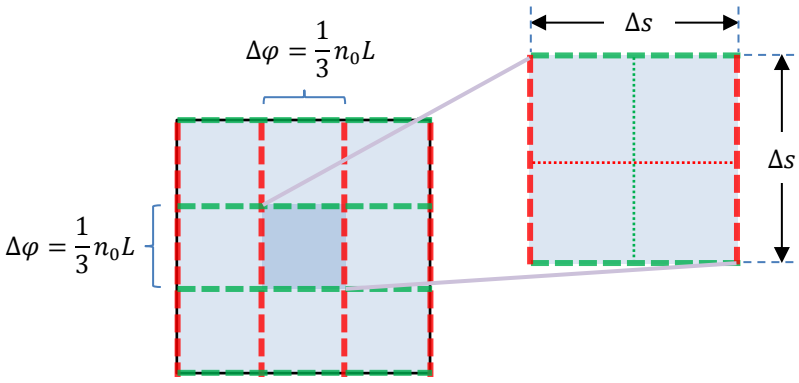


Figure A.2.3. Discrete tile representation of the differential Eikonal for both interrogating ray families, where the middle tile is depicted with horizontal and vertical ray segments passing through its center.

Next, suppose that the refractive index of the center tile in the figure is reduced by a small amount. Because the OPL difference across the tile has not changed, the arc length of the rays that transit the tile must increase to satisfy $\Delta\varphi = n\Delta s$. Hence, a perturbation in the refractive index at some location modifies the physical dimensions of the tile at that location. For the hypothetical reduction in the refractive index of the center tile depicted in Figure A.2.4a, the tile area must increase due to the required change in Δs, as shown in Figure A.2.4b.

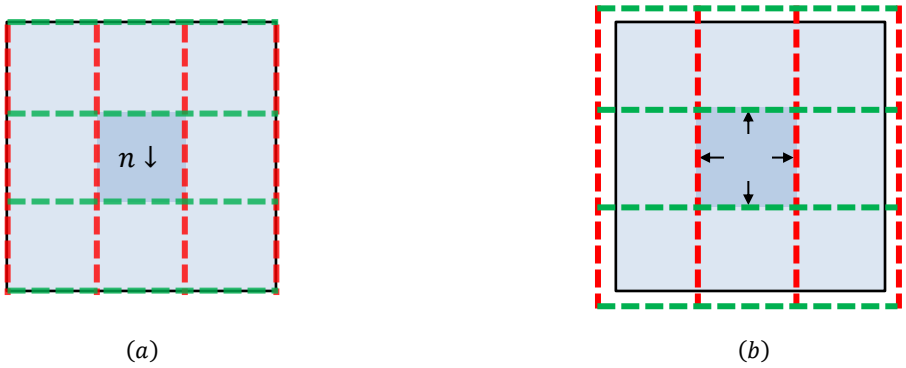


Figure A.2.4. (a) A hypothetical reduction in the index of the material inside the center tile induces (b) an increase in the tile's effective area.

In general, when the geometry of a particular tile changes, the four wavefronts that define its boundaries are necessarily affected due to continuity of the Eikonal S in a smooth refractive index field. Consequently, the geometry of the surrounding tiles whose boundaries are associated with the same wavefronts is also subject to change, as can be seen in Figure A.2.4b. The depicted change in the wavefront geometry does not correspond to different ray families. In fact, if we were to measure the material shown in Figure A.2.4b using the same sets of interrogating rays from Figure A.2.1, the boundary values of ray position and slope would be identical: The index distribution remains homogeneous and the rays still travel in linear trajectories inside the material. This result is sufficient in proving non-uniqueness in the index field reconstructed from the boundary values of two distinct ray families, but it does not eliminate the possibility of a non-unique index gradient. Specifically, a universal change in the index that we have just described can be attributed to the unknown integration constant in computing the index field from its gradient.

In order to show non-uniqueness in the index gradient for this particular set of measurements, let us suppose that the total OPL across the length of the material remains constant at $\varphi = n_0 L$. Under this assumption, the solution suggested in Figure A.2.4b becomes invalid because the depicted change in wavefront geometry requires a change in φ. Even with this additional constraint, the wavefront geometry in Figure A.2.4a can be manipulated to accommodate the hypothetical reduction of the refractive index in the center of the (initially) homogeneous material without producing any inconsistencies in the surrounding tiles. We illustrate one possibility in Figure A.2.5.

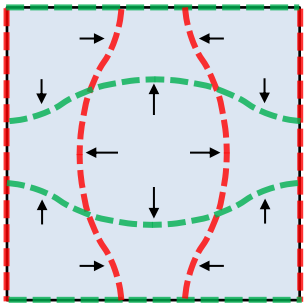


Figure A.2.5. Manipulation of the wavefront geometry to accommodate a hypothetical reduction in the index at the center of a homogenous material while maintaining the overall OPL across the length of the material.

A static OPL across the length of the material implies that the wavefronts along the boundaries of the index field are constrained, as can be seen in the figure. In response to an increase to the area of the center tile under this constraint, the neighboring tiles that share a wavefront boundary with the center tile must experience a decrease in its height accompanied by a simultaneous decrease in its width by the same amount. This ensures that the index remains consistent within each tile, *i.e.*, the OPL and the arc length across the each tile yield the same refractive index for both the vertical and horizontal ray segments depicted in Figure A.2.3. From the alternative wavefront geometry depicted in Figure A.2.5, it is apparent that all tiles still satisfy this condition. Furthermore, it is clear from symmetry that the normally incident rays shown in Figure A.2.1 do not experience a change in position or slope when they are measured on the opposite side of the material. Thus, we have proven that the index gradient is non-unique for the measured set of boundary values; the alternative wavefront geometry implies that curved ray trajectories may be used to satisfy these boundary values, which result in completely different index gradient fields that are capable of producing the same set of measurements. We hasten to add that an alternative wavefront geometry that matches the shape of the wavefront at the boundaries does not necessarily correspond to a valid solution in the general case. This is because the wavefronts are determined by ray slope and do not specify ray position. For the proof we presented in this appendix, however, we were able to exploit the symmetry of the alternative geometry to argue that the boundary values of ray position and slope are unaffected for both sets of interrogating rays from Figure A.2.1.

Appendix 3: Linear Constrained Cubic Splines

Cubic splines are often employed in numerical methods for interpolating values between 1-D sample data. In conventional splines, continuity is typically enforced in both slope and position at the end points, or the knots, of each spline. Imposing these constraints often leads to overshooting between the spline knots. To remedy this, we can specify the slope at the spline knots in *constrained* cubic splines, avoiding the possibility of overshooting while giving up smoothness [Kruger, 2002]. Due to its simplicity and flexibility in implementation, cubic spline interpolation is particularly useful in deriving discrete algebraic equations from deflectometry path integrals; it offers sufficient accuracy for smooth and slowly varying index fields and does not require vast amounts of computational resources (as opposed to the MLS approximation used Appendix 1). In this appendix, we will describe how to implement constrained cubic splines while maintaining a linear relationship between the spline coefficients and the spline knots.

We begin by developing the mathematics of 1-D linear constrained cubic splines for the general case where the spline knots are not uniformly spaced along the independent variable, as shown in Figure A.3.1.

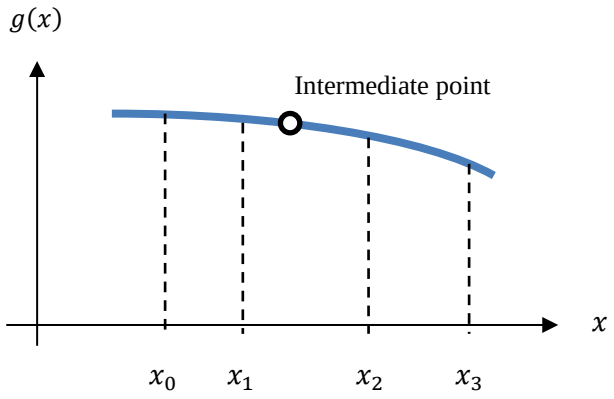


Figure A.3.1. Data points with unequal spacing in the independent variable x.

Let us suppose that $g(x)$ is known at the spline knots located at $x = x_0, x_1, x_2, x_3$ with the values g_0, g_1, g_2, g_3, respectively, and we wish to interpolate $g(x)$ within the interval $x_1 < x < x_2$. To construct the spline within the given interval, we start by computing the slopes at the center of each spline:

$$g'\left(\frac{x_0+x_1}{2}\right) = \frac{g_1-g_0}{x_1-x_0}, \tag{A.3.1a}$$

$$g'\left(\frac{x_1+x_2}{2}\right) = \frac{g_2-g_1}{x_2-x_1}, \tag{A.3.1b}$$

$$g'\left(\frac{x_2+x_3}{2}\right) = \frac{g_3-g_2}{x_3-x_2}\,, \tag{A.3.1c}$$

where the prime denotes differentiation with respect to x. These slopes can then be used to compute the slopes at $x = x_1$ and $x = x_2$ as follows: Using linear interpolation, the slope at the left end point $x = x_1$ is given by

$$g'(x_1) = \frac{x_2 - x_1}{x_2 - x_0} g'\left(\frac{x_0 + x_1}{2}\right) + \frac{x_1 - x_0}{x_2 - x_0} g'\left(\frac{x_1 + x_2}{2}\right)$$

$$= a g_0 + b g_1 + c g_2\,, \tag{A.3.2a}$$

where $= \frac{x_1-x_2}{(x_2-x_0)(x_1-x_0)}$, $b = \frac{x_2-x_1}{(x_2-x_0)(x_1-x_0)} + \frac{x_0-x_1}{(x_2-x_0)(x_2-x_1)}$, and $c = \frac{x_1-x_0}{(x_2-x_0)(x_2-x_1)}$; similarly, the slope at the right end point $x = x_2$ is given by

$$g'(x_2) = d g_1 + e g_2 + e g_2\,, \tag{A.3.2b}$$

where $= \frac{x_2-x_3}{(x_3-x_1)(x_2-x_1)}$, $e = \frac{x_3-x_2}{(x_3-x_1)(x_2-x_1)} + \frac{x_1-x_2}{(x_3-x_1)(x_3-x_2)}$, and $f = \frac{x_2-x_1}{(x_3-x_1)(x_3-x_2)}$. For convenience, it is useful to organize this information in matrix form as

$$\begin{bmatrix} g_1 \\ g_2 \\ g_1{}' \\ g_2{}' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a & b & c & 0 \\ 0 & d & e & f \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} \equiv [\boldsymbol{L}] \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.3}$$

where $g_i \equiv g(x_i)$ and $g_i{}' \equiv g'(x_i)$. Naturally, a cubic spline of the form $g(x) = Ax^3 + Bx^2 + Cx + D$ must satisfy the boundary conditions specified at the spline knots. These boundary conditions can also be expressed in matrix form as

$$\begin{bmatrix} g_1 \\ g_2 \\ g_1{}' \\ g_2{}' \end{bmatrix} = \begin{bmatrix} x_1{}^3 & x_1{}^2 & x_1 & 1 \\ x_2{}^3 & x_2{}^2 & x_2 & 1 \\ 3x_1{}^2 & 2x_1 & 1 & 0 \\ 3x_2{}^2 & 2x_2 & 1 & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \equiv [\boldsymbol{M}] \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}. \tag{A.3.4}$$

Hence, the polynomial coefficients of the cubic spline are given by

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = [\boldsymbol{M}]^{-1} [\boldsymbol{L}] \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.5}$$

Evaluating the spline yields an interpolated value for $g(x)$ for $x_1 < x < x_2$. We express this in matrix form as

$$g(x) = [\boldsymbol{X}]\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = [\boldsymbol{X}][\boldsymbol{M}]^{-1}[\boldsymbol{L}]\begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.6}$$

where $[\boldsymbol{X}] = [x^3 \quad x^2 \quad x \quad 1]$, and it has been assumed that $[\boldsymbol{M}]$ is not ill-conditioned.

Next, we turn our attention to the special case where g_1 is the left-most spline knot, and g_0 is not available as a data point in constructing a spline for $x_1 < x < x_2$. This requires a slight modification in the computation of g' at $x = x_1$. In our implementation, we choose to enforce smoothness in the vicinity of g_1 by maintaining a constant second derivative within the interval $x_1 < x < \frac{x_1+x_2}{2}$. The local second derivative is approximately given by

$$g'' \approx \frac{g'\left(\frac{x_2+x_3}{2}\right) - g'\left(\frac{x_1+x_2}{2}\right)}{\left(\frac{x_2+x_3}{2}\right) - \left(\frac{x_1+x_2}{2}\right)} = \frac{2}{x_3-x_1}\left(\frac{g_3-g_2}{x_3-x_2} - \frac{g_2-g_1}{x_2-x_1}\right). \tag{A.3.7}$$

Extrapolating the slope from $x = \frac{x_1+x_2}{2}$ to $x = x_1$ by maintaining a constant g'' across the interval yields

$$g'(x_1) = g'\left(\frac{x_1+x_2}{2}\right) + \left(x_1 - \frac{x_1+x_2}{2}\right)g''. \tag{A.3.8}$$

Consequently, Equation A.3.2a becomes

$$g'(x_1) = ag_0 + bg_1 + cg_2, \tag{A.3.9}$$

where $= -\frac{1}{x_2-x_1} + \frac{x_1-x_2}{(x_3-x_1)(x_2-x_1)}$, $b = \frac{1}{x_2-x_1} + \frac{x_1-x_2}{x_3-x_1}\left(-\frac{1}{x_3-x_2} - \frac{1}{x_2-x_1}\right)$ and $c = \frac{x_1-x_2}{(x_3-x_1)(x_3-x_2)}$, while Equation A.3.2b remains valid. Because there are only three known values that can be used to construct the spline, Equation A.3.3 must be modified to

$$\begin{bmatrix} g_1 \\ g_2 \\ g_1{}' \\ g_2{}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \\ d & e & f \end{bmatrix}\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \equiv [\boldsymbol{L}]\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.10}$$

where the coefficients a, b, c are given by Equation A.3.9, and the coefficients d, e, f are unchanged from Equation A.3.2b. Equation A.3.6 is also modified accordingly so that

$$g(x) = [\boldsymbol{X}][\boldsymbol{M}]^{-1}[\boldsymbol{L}]\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.11}$$

where $[\boldsymbol{X}]$ and $[\boldsymbol{M}]$ are unchanged from Equations A.3.4 & A.3.6. By analogy, the special case where g_2 is the right-most spline knot and g_3 is not available requires a similar treatment. The modified expression for Equation A.3.3 in this case is given by

$$\begin{bmatrix} g_1 \\ g_2 \\ {g_1}' \\ {g_2}' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} \equiv [\boldsymbol{L}] \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix}, \tag{A.3.12}$$

where $= \frac{x_2 - x_1}{(x_2 - x_0)(x_1 - x_0)}$, $e = -\frac{1}{x_2 - x_1} + \frac{x_2 - x_1}{x_2 - x_0}\left(-\frac{1}{x_2 - x_1} - \frac{1}{x_1 - x_0}\right)$, $f = \frac{1}{x_2 - x_1} + \frac{x_2 - x_1}{(x_2 - x_0)(x_2 - x_1)}$ and the coefficients a, b, c are unchanged from Equation A.3.2a.

The expressions simplify greatly when the known values g_i are equally spaced along the x-axis. For instance, Equation A.3.3 is reduced to

$$\begin{bmatrix} g_1 \\ g_2 \\ {g_1}' \\ {g_2}' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{1}{2\Delta} & 0 & \frac{1}{2\Delta} & 0 \\ 0 & -\frac{1}{2\Delta} & 0 & \frac{1}{2\Delta} \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} \equiv [\boldsymbol{L}] \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.13}$$

where $\Delta = x_1 - x_0 = x_2 - x_1 = x_3 - x_2$. Likewise, Equations A.3.10 & A.3.12 are simplified to

$$\begin{bmatrix} g_1 \\ g_2 \\ {g_1}' \\ {g_2}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{3}{2\Delta} & \frac{2}{\Delta} & -\frac{1}{2\Delta} \\ -\frac{1}{2\Delta} & 0 & \frac{1}{2\Delta} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \equiv [\boldsymbol{L}] \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{A.3.14a}$$

$$\begin{bmatrix} g_1 \\ g_2 \\ {g_1}' \\ {g_2}' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{1}{2\Delta} & 0 & \frac{1}{2\Delta} \\ \frac{1}{2\Delta} & -\frac{2}{\Delta} & \frac{3}{2\Delta} \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} \equiv [\boldsymbol{L}] \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix}, \tag{A.3.14b}$$

respectively.

Appendix 4: Mode Converter Design Optimization

Modern fabrication techniques for GRIN materials do not yet have the capability to produce arbitrary refractive index distributions $n(x, y, z)$. One of the primary limitations in fabrication is the dynamic range of the refractive index, defined as $\Delta n = n_{max} - n_{min}$. In this appendix, we describe several methods that can be used to reduce Δn in the design of GRIN mode converters.

We begin with the design specifications for a 3-D Gaussian-to-flat-top mode converter, which we illustrate in Figure A.4.1. As we briefly mentioned at the end of Section 4.6, the radial symmetry of the converter allows the design problem to be treated in one dimension using radial coordinates in the transverse plane, provided that the descriptions of the input and output irradiance profiles account for the coordinate transformation.

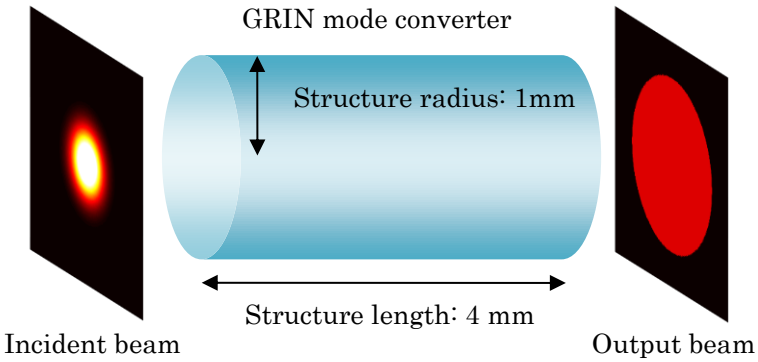


Figure A.4.1. Diagram of a Gaussian-to-flat-top mode converter. Incident Gaussian beam width is 0.354 *mm. Output beam radius is* 1 *mm.*

Following the design procedure described in Section 4.6, we employ half-sinusoidal trajectories to direct ray propagation between the input and output planes. The OPLs of the sampling wavefronts are chosen so that the refractive index remains constant along the axis of symmetry with a value of 1.6 RIU. The resulting index distribution that can be used to achieve the specified conversion is shown in Figure A.4.2. Using this refractive index distribution as our reference design, we outline several methods below that can be used to reduce Δn in the refractive index distribution.

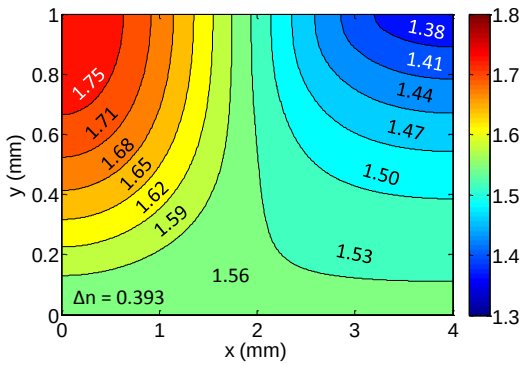


Figure A.4.2. Calculated refractive index distribution for the mode converter in Figure A.4.1, where the transverse variable y corresponds to radial distance from the axis of symmemtry.

One of the degrees of freedom we were allowed in the design procedure was the choice of OPLs along the sampling wavefronts. Assuming that the ray family $Y_i(x)$ has already been specified, the minimum Δn of the design can be determined from the maximum arc length ratio α_{max} given by

$$\alpha_{max} = \max_j \left(\frac{\max_i ds_{i,j}}{\min_i ds_{i,j}} \right), \tag{A.4.1}$$

where $ds_{i,j}$ is the arc length between sampling wavefronts $X_j(y)$ and $X_{j+1}(y)$ along the sampling ray corresponding to i in $Y_i(x)$. Since the OPL is constant between any pair of adjacent wavefronts (or more generally, any two wavefronts), the arc length ratio can also be expressed as

$$\alpha_{max} = \frac{n_{min}}{n_{max}}, \tag{A.4.2}$$

where we have made use of Equation 4.6.2 along the ray trajectories. The above expression states that the ratio of the maximum index n_{max} to the minimum index n_{min} is determined solely by the maximum ratio between the maximum and minimum arc lengths along all ray trajectories between any pair of adjacent wavefronts. There remains some freedom in choosing the OPL of the sampling wavefronts because α_{max} applies to the entire refractive index distribution, not to specific pairs of adjacent wavefronts. Hence, the entire dynamic range index Δn is only utilized for specific pairs of adjacent wavefronts, *i.e.*, pairs for which the ray segments between the wavefronts exhibit the maximum ratio of maximum-to-minimum arc length. Figure A.4.3 illustrates an alternative refractive index distribution to the reference design from Figure A.4.2 after the OPLs of the sampling wavefronts have been modified to achieve the minimum Δn. Based on the relationship expressed in Equation A.4.1, it stands to reason that the arc length ratio α_{max} can be scaled freely by stretching (or shrinking) the trajectories along the x-axis. This, however, implies a change to the specified length for the GRIN mode converter. In the extreme limit where the mode converter is infinitely long, α_{max} approaches unity, and Δn approaches zero.

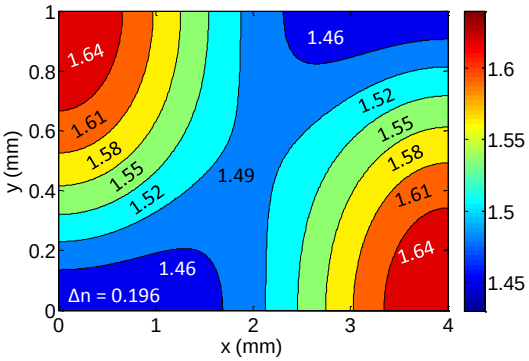
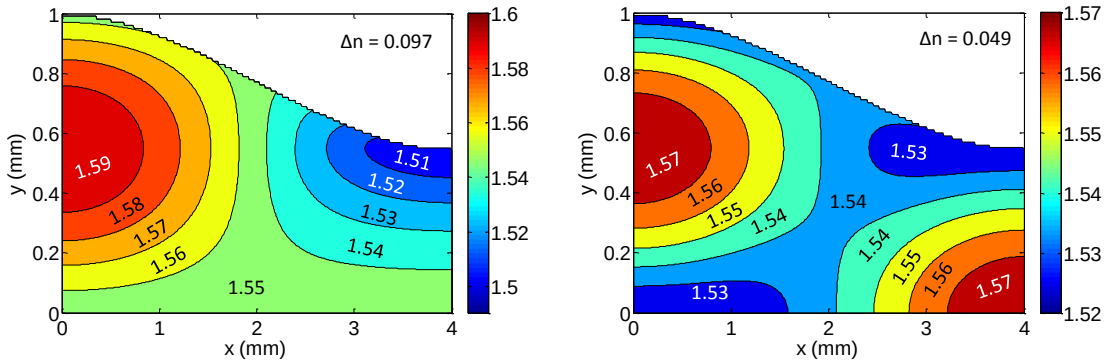


Figure A.4.3. Calculated refractive index distribution for the mode converter in Figure A.4.1 after modifying OPL values for sampling wavefronts to reduce Δn.

If we were to allow changes to the trajectories of the ray family $Y_i(x)$, more options become available for reducing Δn in the design. One possibility is to scale the size of the output mode. Figure A.4.4a shows the a newly calculated refractive index distribution for the desired mode conversion after changing the radius of the output beam to 0.55 cm while keeping the refractive index constant along the axis of symmetry. The design in Figure A.4.4b employs the same trajectories used in computing the index distribution in Figure A.4.4a but relaxes the constraint for the axial refractive index so that the OPL values of the sampling wavefronts can be modified to reduce Δn.



(a) (b)

Figure A.4.4. Calculated index distribution after (a) changing the radius of the output beam in the mode converter to 0.55 *cm and (b) relaxing the constraint on the axial index and optimizing the OPL values of the sampling wavefronts.*

An important observation for these new designs is that the refractive index profile in the radial direction is no longer monotonic. This is because the ray trajectories in the reference design were collectively being guided away from the axis of symmetry or toward it by the GRIN material for any value

of x. The new design after resizing the output mode does not have this feature and, thus, requires a non-monotonic index profile in the radial direction.

Another way to reduce Δn in the design, which we alluded to in Section 4.6, is to change the functional form of the ray trajectories in $Y_i(x)$. Figure A.4.5 shows a comparison between the reference design from Figure A.4.1 and an alternative design employing cubic polynomials to direct ray propagation between the input and output planes. For the specified mode conversion, the sinusoidal ray family produces a smaller Δn in the design, as is apparent from the comparison. We hasten to add, however, that we have not proven the optimality of sinusoidal trajectories. There may, in fact, be functions that are superior to sinusoids in minimizing Δn.

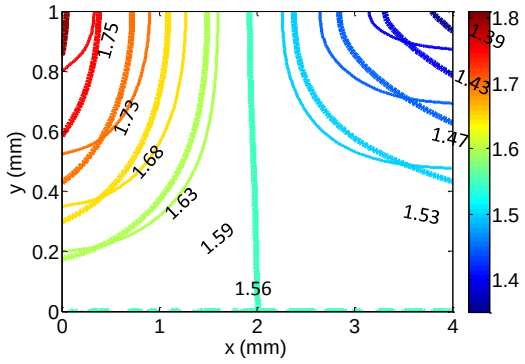


Figure A.4.5. Comparison between the calculated index distributions employing cubic polynomial trajectories (thick contour lines) and half-wave sinusoids (thin contour lines). The former produced a higher $\Delta n = 0.487$ *RIU compared to the design in Figure A.4.2.*