

Contents

Abstract	1
1 Introduction	2
2 Properties of Optical Fibers	4
2.1 Physical Properties	4
2.2 Attenuation	6
2.3 Snell's Law and Total Internal Reflection	6
2.4 Goss Hanchen Shift	10
3 Introducing Maxwell's Equations	12
3.1 Gauss's Law for Electrostatic Flux	12
3.2 Polarization and Displacement Fields	14
3.3 Electric Current	18
3.4 Magnetic Induction	20
3.5 Electromagnetic Induction	23
3.6 Ampere-Maxwell Law (Generalization of Ampere's Law)	24
3.7 Maxwell's Wave Equation	25
4 The Second Order Wave Equation	26
4.1 Hyperbolic PDEs	26
4.2 Second Order Wave Equation	28
4.3 1-Dimensional Wave Equation	29
4.4 Plane Waves	32
5 Dispersion	36
5.1 Introduction to Dispersive Waves	36
5.2 Linear Dispersive Waves	38

5.3	Group Velocity for Linear PDEs	41
6	Guided Modes in Cylindrical Waveguides	44
6.1	Guided Rays of Light and Dispersion	44
6.2	Characteristic Equations for Guided Modes	45
7	Conclusions	55
A	Appendix: Vector Calculus Results	58
A.1	The Gradient	58
A.2	Divergence of a Field	58
A.3	Divergence Theorem	59
A.4	Curl and Circulation of a Vector Field	59
A.5	Stoke's Theorem	59
B	Appendix: Differential Equations	60
B.1	Solutions to ODEs Around A Regular Singular Point	60
B.2	Bessel's Equation	63
B.3	The Modified Bessel's Equation	69
B.4	Defining PDEs and Initial Boundary Value Problems	74
B.5	Method of Characteristics	78
C	Appendix: Cylindrical Coordinates	83
	References	85

Abstract

Optical fibers are used primarily for transmitting information. The theory of optical fibers provides methods for transferring more information over a longer distance than traditional data transmission methods. This is due to the dispersive properties of light propagating through an optical fiber. In this report, the fundamental theory of linear dispersion over short distances in optical fibers is established. The differential equations and boundary conditions permitting light propagation in an optical fiber are derived. The theory of linear dispersive waves and group velocity are introduced. The method of characteristics is used to determine the characteristic equations which possible solutions must satisfy. We discuss the dispersion and group velocity behaviour of possible solutions satisfying the wave equation and characteristic equations. The theory of linear dispersion in optical fibers over short distances is useful for practical applications involving transmission of information. The theory of linear dispersive waves in optical fibers also leads to the study of non-linear effects in optical fibers.

1 Introduction

This report describes the propagation of light waves in an optical fiber. First, the topics that are discussed in this report are described and then an overview is given of the sections where these topics are described. The part of the optical fiber where wave propagation occurs is within its cylindrical glass core. Waves describing oscillations in space are represented by oscillatory functions of both space and time. In general, an oscillatory function represents a waveform that has a wavenumber that depends on position and a frequency that depends on time. An oscillatory function also has a phase which propagates with a phase velocity determined by the frequency and wavenumber.

Light propagation occurs in the optical fiber by the nature of the electric and magnetic fields, which determine how the positions of electrons may change with respect to changes time. Maxwell's equations govern the electric and magnetic fields in any material based on specifically and experimentally derived electric and magnetic field properties in the material. When light propagates through the fiber, the electric and magnetic fields describing the light behaves according to a wave equation, called Maxwell's wave equation, which has oscillatory solutions.

Maxwell's wave equation is a second order partial differential equation (PDE) with respect to three dimensional position and time. In general, Maxwell's wave equation is a nonlinear PDE. The intensity of light propagating through the core corresponds to the varying amplitudes of the oscillatory solution to Maxwell's wave equation. Maxwell's wave equation takes the form of a linear PDE for sufficiently small intensities of light or for sufficiently small variations in the electric field distribution. These assumptions are necessary for finding possible linear dispersive solutions to Maxwell's wave equation using the method of characteristics.

A dispersive solution to a PDE is called hyperbolic at points where solutions exist and are uniquely determined by families of paths in the solution space. In

general, hyperbolic PDEs are analyzed using the method of characteristics.

The second order wave equation is defined as a hyperbolic PDE. This is used to define Maxwell's wave equation as a second order wave equation with 4 independent variables. The method of characteristics may be applied in general to find solutions to the 1-dimensional wave equation to interpret the possible solutions to Maxwell's wave equation, which takes the form of a second order wave equation. Solutions to Maxwell's wave equation may take the form of a dispersive plane wave.

The definition of a dispersive equation and dispersive solutions are derived for general linear PDEs. When a differential equation has dispersive solutions, the solutions have oscillatory behaviour. We require a more general notion of phase, since it is changing as the solution oscillates. The relationship between wavenumber and frequency is generalized further by deriving the group velocity. Dispersion only occurs if the phase velocity is not the same as the group velocity. In other words, the average and local changes in phase are not the same.

Solutions to Maxwell's wave equation are derived by choosing boundary conditions that necessarily and sufficiently confine rays of light in the core. The characteristic equations corresponding to the paths of light rays in the solution of Maxwell's wave equation are derived from parametrizing light with linear rays. These characteristic equations represent differing modalities of light ray distribution within the core. The full dispersive solution for Maxwell's wave equation is a superposition of these modes. The example of a single type of light ray is presented as one of the possible modes satisfying the characteristic equation in section.

Now, an overview of the topics mentioned above is given. The physical properties of an optical fiber and the properties of light within its glass core are described in section 2. The electric and magnetic fields and Maxwell's equations

which describe them are defined in section 3. Before being able to define dispersion, the second order wave equation is defined as a hyperbolic PDE in section 4.2, where hyperbolic PDEs are discussed as well. This is required for existence of solutions to Maxwell's wave equation.

The method of characteristics is derived in appendix B.5 and used in subsection 4.4 for the example of a 1-dimensional wave. Doing so leads to the definition of a plane wave in section 4.2, which is required in section 6 for finding solutions that occur in this form. Before combining ideas from sections 1 to 4 to find solutions, dispersion, phase velocity and group velocity are defined in section 5.

Dispersion occurs only if the phase velocity is not the same as the group velocity. The dispersion is observed by finding solutions to Maxwell's wave equation in section 6 satisfying the linearization assumption, meaning that light is parametrized by rays. This only occurs for short distances in the optical fiber, since non-linear effects may emerge over a longer distance as it is inherent of the material properties. The approximation of the relative electric permittivity in subsection 3.2 allows for the linearization assumption.

2 Properties of Optical Fibers

2.1 Physical Properties

These subsections follow [3], [8], [10] and [9]. There are several physical properties that an optical fiber possesses which must be considered. The isotropic, birefringent and homogeneous properties of the core are defined in this subsection. The main objective of this section is to define the specific properties of an optical fiber such that attenuation may be controlled by conditions at the core-cladding boundary. Attenuation is defined in subsection 2.2.

First, the microscopic and molecular structure of the glass core is character-

ized. An isotropic material is a material in which the physical properties do not depend on orientation, i.e. uniform structure. Glass is isotropic because the silica (SiO_2) molecules are arranged uniformly in a lattice that can be oriented symmetrically in any direction. It is possible, however, to manufacture optical fibers that include other compounds inducing a non-isotropic state in the glass.

Assume that there are no imperfections in the symmetry and no birefringence. Birefringence describes the affects of warping the cylindrical structure and is assumed to be neglected.

A homogeneous material has physical properties which do not depend on position. This report analyzes dispersion in a homogeneous optical fiber. It is possible, however, to manufacture optical fibers with physical properties that vary radially. The core of the optical fiber is covered by a material referred to as the cladding. The physical characteristics of the cladding are selected such that the attenuation in core may be controlled.

Refer to figure 1 for an illustration of the optical fiber structure being considered. Note that the coating has no affect on the light propagation in the core and does not contribute to the boundary conditions at the core-cladding interface which we consider.

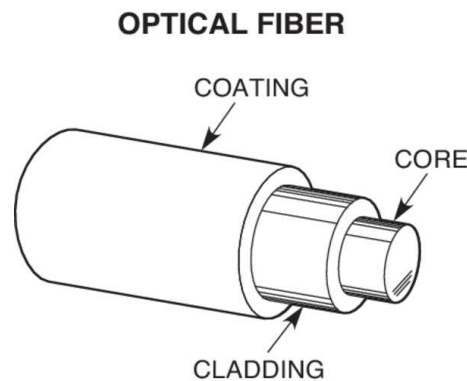


Figure 1: Optical fiber, [14].

2.2 Attenuation

The losses in an optical fiber refer to the forms in which loss of electric charge occurs when light propagates through the core. Generally, the attenuation in an optical fiber is derived from a reduction in the flux of the electric and magnetic field in a medium occurring. The electric and magnetic fields are defined in section 3. Molecules in the core reorient charges such that electromagnetic waves may propagate through it. In general, a loss of power occurs as electromagnetic waves propagate in the core and the wave becomes dissipated. A lossless medium is defined as one that does not suffer *any* attenuation.

When light passes through any medium, the light is either reflected or absorbed according to the photoelectric effect. Otherwise, the light is reflected or refracted at the boundaries of the material according to Snell's law. The photoelectric effect and Snell's law accounts for all possible forms of attenuation in optical fibers.

The photoelectric effect is defined by the emission of electrons from a material when light is shined on it. This occurs in the valence orbital of atoms, which is not discussed further. Glass is a neutral material. This means valence orbitals of every atom are full as a consequence of SiO_2 molecules being bonded together in a lattice with shared valence electrons. Therefore, no electrons are reflected or absorbed due to the photoelectric effect. Moreover, the only form of attenuation occurring is refraction at the core-cladding boundary.

2.3 Snell's Law and Total Internal Reflection

Snell's law is the law governing refraction of light rays between materials, each having an associated index of refraction. Refraction refers to the path a ray of light takes when passing through a boundary between differing media. Before stating Snell's law, the definitions for a light ray and refractive index are required.

Snell's law is used in this section to deduce the condition for total internal reflection. Total internal reflection refers to the phenomena of a light ray reflecting back into the incident media without refracting. Refer to figure 3.

A light ray is defined by Descarte's law, which states that light travels along the path of least time. It may be proved, using variational calculus, that such a path is a straight line with respect to small variations on the proposed path [11]. A very crucial part of this proof is that it assumes a difference in time to occur if and only if there is at least one propagation direction. We do not show this proof; however, we assume such small variations and apply this as a necessary assumption in section 6. Snell's law follows from the variational proof and the fact that an extremum for the optical path is an extremum for the time taken as well.

The assumption of small variations in the light path corresponds to the linear approximation of permittivity, (defined in section 3), and implies that the refractive index in the glass core is constant. A materials refractive index is defined as the ratio of the speed of light in vacuum with the speed of light in the medium *for small variations in the light path*. The speed of light in vacuum is given by

$$c = \sqrt{\frac{1}{\varepsilon_0 \mu_0}},$$

where the constants of permittivity, ε_0 , and permeability, μ_0 , are defined in section 3. Refer to figure 2 for the definition of Snell's Law. It states that

$$\frac{\sin \theta_2}{\sin \theta_1} = \frac{v_2}{v_1} = \frac{n_1}{n_0},$$

where v_1, v_2 is the phase velocity of light and n_1, n_0 is the refractive index in the media, respectively.

From Snell's law, we may analyze the boundary conditions associated with solutions to the wave equation governing light propagation in the core. Light rays bend either away or toward the normal, relative to the incident ray direction,

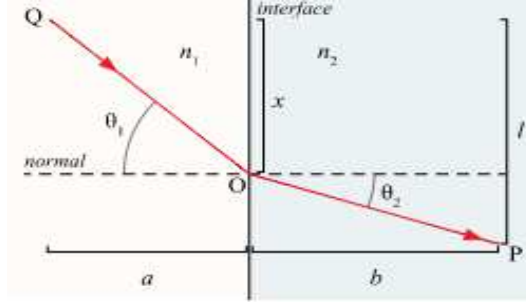


Figure 2: Snell's law, [16].

depending on which media has a higher refractive index. When rays move from higher to lower refractive index regions then light rays bend away from the normal. This is illustrated by figure 3 for arbitrary refractive indices n_1 and n_0 .

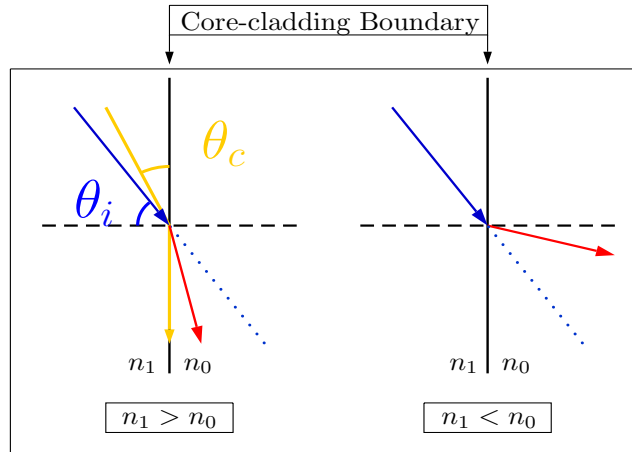


Figure 3: Refraction for relative differences in refractive index.

In order for rays with incident angles $0 < \theta < \pi/2$ to be reflected back into the core, the cladding must have a refractive index that is lower than the refractive index for glass. Let n_1 and n_0 be the refractive index for the core and cladding, respectively. We choose a material for the cladding such that $n_1 > n_0$. Since the light rays bend away from the normal, there exists some angle where light begins to be reflected back into the incident media. This is called the critical angle. It is

possible to use Snell's law to relate the critical angle, θ_c , to the refractive indices by

$$\sin \theta_c = \frac{n_0}{n_1}. \quad (2.1)$$

Therefore, any angle larger than the the critical angle, given by (2.1), results in a refracted light ray. Now we check the maximum acceptance angle for totally reflected rays at the end face of the optical fiber based on the critical angle. Refer to figure 4.

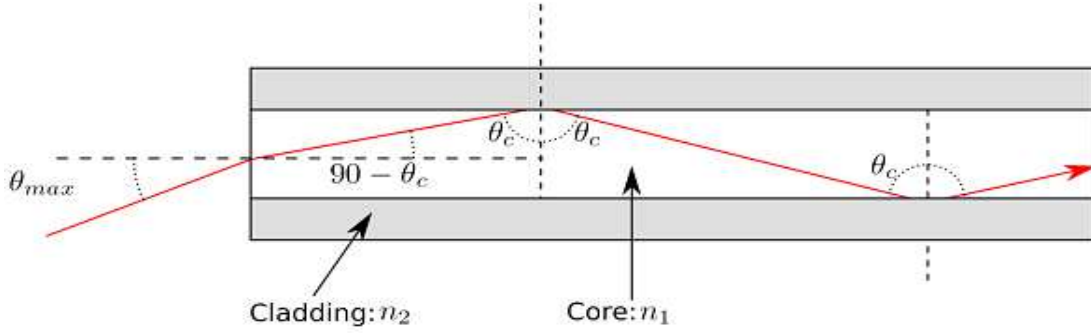


Figure 4: Maximum acceptance angle, [15].

Observe that the incident angle for rays within the core coming from the end right end face have an angle $\pi/2 - \theta_c$. Applying the critical angle condition from equation (2.1) with some trigonometric identities for angles $0 < \theta_{max} < \pi/2$ gives

$$\theta_{max} \leq \sin^{-1} \sqrt{n_1^2 + n_0^2}.$$

Due to the cylindrical symmetry, this is the acceptance angle for which light rays are totally reflected in the core. Any light rays entering outside of the cone determined by the acceptance angle are refracted upon the first encounter with the core-cladding boundary and have no affect on the wave propagation in the core.

2.4 Goss Hanchen Shift

The only remaining physical phenomenon in optical fibers that needs to be accounted for as a boundary condition is a lateral shift called the Goss-Hanchen shift. It is observed experimentally and can be deduced mathematically by the derivations in [3] and [8]. Light rays suffer a lateral shift upon reflection at the boundary. This lateral shift is due to a phase shift of the light ray whenever it is reflected at least twice. The following summarizes the arguments in [8] and states the phase-matching condition at the core-cladding boundary.

The phase matching condition refers to the phase of a light ray propagating along the length of the core with an arbitrary planar phase surface. A phase surface is discussed in section 5.2. A ray of light is necessarily governed by Maxwell's equation described in section 3.

Due to the assumption of small variations in the path of light, (which defines a ray of light), different points on the same planar phase surfaces may have to traverse longer geometrical distance in the same time. These distances are the supposed paths of light rays. The light rays cannot bend while propagating in the core alone, but when reflected from the boundary, the magnetic field at every point of the ray rotates an angle $\pi/2$.

The resulting light ray suffers a complex phase shift according to the angle of incidence. The phase difference may be interpreted by considering the reflected magnetic field components of the light ray in superposition with the reflected electrical field components. The electric and magnetic field are defined in section 3. The phase shift of the magnetic field is illustrated in figure 5 and 6. We now state the phase matching condition according to figure 5 and figure 6. Let kn_1 be the wavenumber for the phase of incident ray having angle ϕ and let β be the component of the wavenumber in the z direction as in figure 5. The phase

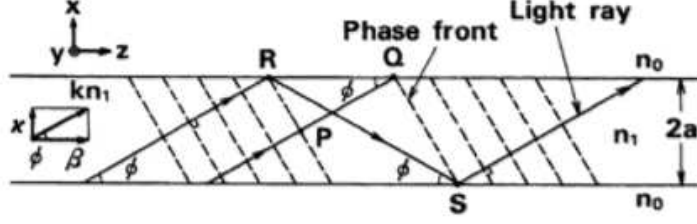


Figure 5: Light ray propagation for planar phase surfaces, [8], page 2.

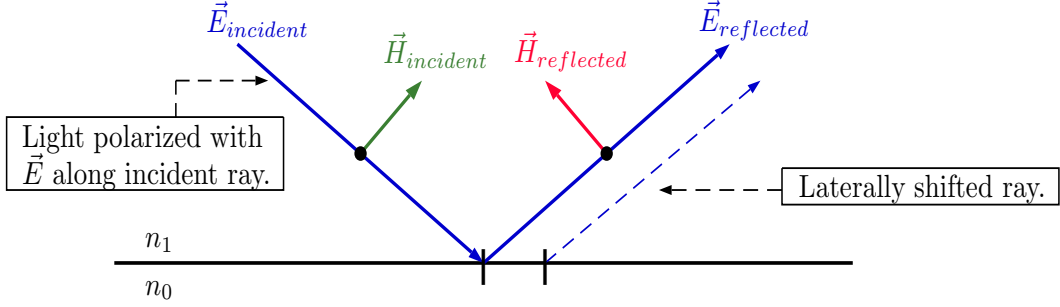


Figure 6: Demonstrating the need for phase matching condition.

matching condition is given as

$$\tan \left(kn_1 a \sin \phi - \frac{m\pi}{2} \right) = \sqrt{\frac{2 \left(\frac{n_1^2 - n_0^2}{2n_1^2} \right)}{\sin^2 \phi} - 1}, \quad (2.2)$$

where phase fronts are perpendicular to the path of a light ray propagating with wavelength and wavenumber λ/n_1 and $kn_1 = 2\pi n_1/\lambda$, respectively. Note that λ denotes the wavelength of light in vacuum. It is important to point out that a light ray propagates on a planar phase front along the z direction with respect to the equation

$$\beta = kn_1 \cos \phi.$$

The phase matching condition implies that light ray propagation is discrete and is determined by the radius, refractive indices, and wavelength λ . In order for light rays to be confined within the core, the phase matching condition must be satisfied.

3 Introducing Maxwell's Equations

The subsections here mainly follow [10] with some interpretation from [5]. Maxwell's equations are a set of four differential equations. The objective of section 3 is to derive Maxwell's equations from various experimental and mathematical results. The four equations will be derived and stated individually throughout the following subsections. The equations describe the local behaviour of time dependent electric and magnetic fields in 3-dimensional space.

Let S be a surface which encloses an arbitrary volume in 3-dimensional space that includes the origin. Let the set of points contained in the volume and on S be denoted by V . The electric and magnetic fields are vector fields associated with the points in V . The electric field is defined in section 3.1 and the magnetic field is defined in section 3.4.

Maxwell's equations describe these two vector fields and provides relations involving the electric and magnetic fields with respect to position in V and time, t . The resulting equations are used to derive Maxwell's wave equation in section 3.7.

3.1 Gauss's Law for Electrostatic Flux

This subsection follows the arguments in [10] to derive Maxwell's first equation. This is done by establishing certain properties of stationary electrical charges. Let q_0, q_j be point charges in V and \vec{r}_0, \vec{r}_j be the respective positions of q_0, q_1 from the origin. Assume that a 'segment' of charge may be infinitely subdivided and that a continuous charge distribution exists in V . The volume charge density measures the amount of charge per unit volume and is defined pointwise for charges q in V by

$$\rho = \lim_{\Delta V' \rightarrow 0} \frac{\Delta q}{\Delta V'}, \quad (3.1)$$

where Δq is total change in charge and $\Delta V'$ is total change in V' for an arbitrary open ball $V' \subset V$. Coulomb's law states that the force between two point charges q and q_0 is proportional to their product. Coulomb's law for the force \vec{F}_0 on charge q_0 relative to q_1 is derived using quantum mechanics and is given mathematically by

$$F_0 = \frac{1}{4\pi\epsilon_0} \frac{q_0 q_j (\vec{r}_0 - \vec{r}_j)}{|\vec{r}_0 - \vec{r}_j|^3}, \quad (3.2)$$

where the constant of proportionality ϵ_0 is obtained experimentally and referred to as the permittivity of vacuum. The electric field is defined pointwise by

$$\vec{E} = \lim_{q_0 \rightarrow 0} \frac{\vec{F}_0}{q_0}.$$

Consider the case of a 'test' charge, q_0 , at the origin being the only charge present in V . Suppose another point charge q_j is in V . The electric field, $\vec{E}(\vec{r}_j)$, at the point determined by \vec{r}_j is given by

$$\vec{E}(\vec{r}_j) = \frac{q_0 \vec{r}_j}{4\pi\epsilon_0 |\vec{r}_j|^3} \quad (3.3)$$

We are only interested in points that are in V . Let da be a differential area element of $S \subset V$ with normal unit vector \hat{n} .

Definition 3.1. *The flux of a vector field \vec{A} over any surface S is defined by*

$$\int_S \vec{A} \circ \hat{n} da,$$

where \hat{n} is a unit vector perpendicular to S , and da is a differential area element on S defined by \hat{n} at some point on S .

The flux of $\vec{E}(\vec{r}_j)$ is

$$\oint_S \vec{E}(\vec{r}_j) \circ \hat{n} da = \frac{q_0}{4\pi\epsilon_0} \oint_S \frac{\vec{r}_j \circ \hat{n}}{r_j^3} da \quad (3.4)$$

Observe that

$$\frac{\vec{r}_j \circ \hat{n} da}{|\vec{r}_j|} \perp \vec{r}_j$$

and projects da perpendicularly to \vec{r}_j . Project da onto the differential area element, da' , of the surface S'_j centered at the origin with radius defined by \vec{r}'_j satisfying

$$\oint_S \frac{\vec{r}_j \circ \hat{n}}{|\vec{r}_j|^3} da = \oint_{S'_j} \frac{\vec{r}'_j \circ \hat{n}}{|\vec{r}'_j|^3} da' = 4\pi.$$

Then equation (3.4) may be written as

$$\oint_S \vec{E}(\vec{r}_j) \circ \hat{n} da = \frac{q_0}{\epsilon_0}.$$

Consider the case of multiple point charges in V . Each charge q_j for $j = 1, \dots, n$ has radius \vec{r}_j that defines an arbitrary area element. The argument above may be applied in series to obtain the flux of a total electric field, \vec{E} , as

$$\oint_S \vec{E} \circ \hat{n} da = \frac{1}{\epsilon_0} \sum_{j=1}^n q_0 = \frac{1}{\epsilon_0} \int_V \rho dV$$

where $\sum_{j=1}^n q_0 = q_{enc}$ is the total enclosed charge and each point charge contributes $\rho dv/\epsilon_0$ to the surface integral and since a continuous charge distribution corresponds to $n \rightarrow \infty$ so that the sum becomes an integral. Thus, Gauss' law for electrostatic fields is given as

$$\oint_S \vec{E} \circ \hat{n} da = \frac{1}{\epsilon_0} \int_V \rho dV. \quad (3.5)$$

The divergence theorem is found in appendix A and is applied to (3.5) in order to obtain the differential form of Gauss' law. Gauss' law follows, after applying divergence theorem and then differentiating with respect to V , and is stated as

$$\vec{\nabla} \circ \vec{E} = \frac{\rho}{\epsilon_0}. \quad (3.6)$$

3.2 Polarization and Displacement Fields

In this subsection the electric field theory is extended to the case of dielectric materials following [10]. A dielectric material is one which has charges embedded

in it. Denote the total charge embedded within an open volume, V , bounded by hypothetical surface S , by the variable Q . Without loss of generality, consider at least three charge conductors embedded in a dielectric with surfaces S_1, S_2, S_3 . Let the charges be denoted by q_1, q_2 and q_3 , respectively, such that $Q = q_1 + q_2 + q_3$.

The macroscopic field, \vec{D} , must be derived from the electrostatic field, \vec{E} by applying Gauss' law to the charge conductors in V . Before doing so, the polarization field, \vec{P} , of a dielectric must be defined.

Polarization occurs in dielectrics and accounts for the response of a dielectric material to the electric field, \vec{E} . The polarization field occurs as an effect of embedded charge conductors and is represented by the so-called constitutive equation, $\vec{P}(\vec{E})$, that is derived experimentally. The constitutive relationship depends on the properties of molecules in a dielectric, in general. The experimental quantity that defines the constitutive equation is the electric susceptibility, $\chi(E)$, satisfying the constitutive equation

$$\vec{P} = \chi(E)\vec{E}. \quad (3.7)$$

The electric susceptibility is generally expressed using a Taylor series expansion. This generalization of the electric susceptibility is not discussed here since it is only necessary for observing non-linear dispersive waves propagating in glass. The first order linear approximation of the electric susceptibility with respect to the Taylor expansion is sufficient for linear dispersive waves to propagate in glass. In this case the electric susceptibility, $\chi(E)$, is a constant.

Let da be the differential area element and let \hat{n} be its normal unit vector for an arbitrary surface. Let dv be the differential volume element of V . Define the net polarization charge as

$$Q_p = \int_S \vec{P} \circ \hat{n} da + \int_V \left(-\vec{\nabla} \circ \vec{P} \right) dv = \int_{S_1+S_2+S_3} \vec{P} \circ \hat{n} da + \int_V \left(-\vec{\nabla} \circ \vec{P} \right) dv. \quad (3.8)$$

The right hand side follows since integrating over the hypothetical surface S is equivalent to integrating over the surfaces S_j enclosing the total embedded charge, Q .

The integral form of Gauss' law, (3.5), is applied to account for all charges in V and gives

$$\oint_S \vec{E} \circ \hat{n} da = \frac{1}{\varepsilon_0} (Q + Q_p), \quad (3.9)$$

according to the definition, (3.1), of the charge density, ρ . By applying the divergence theorem from appendix A to equation (3.9) we obtain

$$\int_V \left(-\vec{\nabla} \circ \vec{P} \right) dv = \oint_{S \cup S_1 \cup S_2 \cup S_3} \left(-\vec{\nabla} \circ \vec{P} \right) dv = - \oint_S \vec{P} \circ \hat{n} da - \int_{S_1 + S_2 + S_3} \vec{P} \circ \hat{n} da,$$

by considering all contributions S , S_1 , S_2 and S_3 bounding the polarization field in V .

This implies that equation (3.8) becomes

$$Q_p = - \oint_S \vec{P} \circ \hat{n} da$$

and consequently equation (3.9) becomes

$$\oint_S \left(\varepsilon_0 \vec{E} + \vec{P} \right) \circ da = Q.$$

From this last equation, the electric displacement field is defined as

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P}.$$

Observe from the definition of charge density, given by equation (3.1), that

$$\oint_S \vec{D} \circ \hat{n} da = Q = \rho \Delta V, \quad (3.10)$$

by realizing Q is a change in total charge according to the integral. It follows that

$$\lim_{\Delta V \rightarrow 0} \frac{1}{\Delta V} \oint_S \vec{D} \circ \hat{n} da = \rho.$$

According to the definition of divergence from appendix A that involves the integral, we obtain the form of Gauss' law with respect to the electric displacement field given by

$$\vec{\nabla} \circ \vec{D} = \rho. \quad (3.11)$$

Equation (3.11) is Maxwell's first equation.

Before proceeding with the next section, it is necessary to derive the constitutive relationship, $\vec{P} = \chi(\vec{E})\vec{E}$, with respect to \vec{D} . Combining equations (3.7) and (3.12) implies that

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P} = \varepsilon_0 \vec{E} + \chi(E)\vec{E} = [\varepsilon_0 + \chi(E)] \vec{E}.$$

Therefore, it follows that

$$\vec{D} = \varepsilon(E)\vec{E}, \quad (3.12)$$

where $\varepsilon(E) = \varepsilon_0 + \chi(E)$ is defined as the relative electric permittivity that is specifically derived for each material according to the experimentally obtained relation $\chi(E)$.

In general, the electric susceptibility of glass is a function of source frequency, ω , and is expressed as a Taylor expansion with respect to the variable ω . For considering nonlinear effects in optical fibers, the nonlinear terms of the expansion are required. In this report we consider sufficiently small frequency and distances, so only the linear terms of the expansion are required. This results in a semi-linear wave equation and linearly approximated dispersive solutions. Refer to appendix B.4 for the definition of a semi-linear PDE. The wave equation is defined at the end of section 3, which requires the linearly approximated relative permittivity, ε .

3.3 Electric Current

The electric current must be defined to describe the effect of varying charge displaced within V with respect to variations in time. The definitions and properties of electric current and electric current density are given in this subsection according to [10].

The electric current is defined for steady state conditions as

$$I = \text{rate of charge transfer through surface of } V = \frac{\partial Q(t)}{\partial t}, \quad (3.13)$$

where $Q(t)$ is the net charge transported in time, t , for a steady flow of charge, i.e., steady current.

Assumptions must be made for defining a current density through the surface of V . Current density is defined similarly to how the charge density is defined in section 3.1. Note that these are not independent.

Assume the following:

1. There is one type of charge carrier, namely, the electron with charge q .
2. N = number of charge carriers per unit volume.
3. Ignore effects of random thermal motion.
4. Every charge carrier has equal drift velocity, \vec{v} .

The current is carried entirely by valence electrons. The charge density, denoted by \vec{J} , is computed as the current passing through a differential area element, da , for steady state currents following the assumptions above. Denote the amount of charge crossing da , over an interval of time ∂t , by ∂Q . By the assumptions above it follows that for each charge carrier, j ,

$$\partial Q_j = q_j N_j \circ \hat{n} \partial t da.$$

Then, the amount of current passing through da is

$$\partial I = \frac{\partial Q}{\partial t} = \sum_{j=1}^{\infty} q_j N_j \vec{v}_j \circ \hat{n} da = \left[\sum_{j=1}^{\infty} q_j N_j \vec{v}_j \right] \circ \hat{n} da.$$

We define the current density as

$$\vec{J} = \sum_{j=1}^{\infty} q_j N_j \vec{v}_j. \quad (3.14)$$

It follows from taking the integral of dI , that the current is given by

$$\oint_S \vec{J} \circ \hat{n} da. \quad (3.15)$$

Refer to appendix A for the definition of divergence and divergence theorem. By assumption 4 above, steady currents experience no change in velocity at different positions in the volume, V . This implies that steady state currents have zero divergence with respect to current density. Therefore,

$$\vec{\nabla} \circ \vec{J} = 0. \quad (3.16)$$

We require a derivation of the continuity equation, which relates charge density and current density as previously noted, and it is derived next. It is a consequence of the conservation of energy and the divergence theorem from appendix A. It is required in section 3.6 for deriving Maxwell's fourth equation.

From the definition of current given by (3.15), the divergence theorem is applied for an outward normal unit vector \hat{n} and it follows that

$$I = - \oint_S \vec{J} \circ \hat{n} da = - \int \int \int_V \vec{\nabla} \circ \vec{J} dv, \quad (3.17)$$

where dv is a differential volume element.

Recall the original definition of current given by (3.13). The change in total charge distributed in the volume, V , with respect to changes in time must be equal to the change in total charge with respect to time. It follows that

$$\frac{\partial Q}{\partial t} = \frac{\partial}{\partial t} \int \int \int_V \rho dv. \quad (3.18)$$

Combining equations (3.17) and (3.18) results in the continuity equation for current density given by

$$\int \int \int_V \left(\frac{\partial \rho}{\partial t} + \vec{\nabla} \circ \vec{J} \right) dv = 0 \quad \Rightarrow \quad \frac{\partial \rho}{\partial t} + \vec{\nabla} \circ \vec{J} = 0.$$

3.4 Magnetic Induction

This subsection derives Maxwell's second equation. The same assumptions as in [10] are considered in the following derivations. Recall that the force experienced by the electric field at a particular point in V is given by equation (3.2). The magnetic force is defined as the force produced by charge carriers that are in motion with a velocity \vec{v} . It is defined as

$$\vec{F}_m = \frac{\mu_0}{4\pi} \frac{qq_1}{r} \vec{v} \times \left(\vec{v}_1 \times \frac{\vec{r}}{r} \right),$$

where μ_0 is the constant of permeability that is defined according to experimental laws for the equation to be compatible with the mks units. The magnetic induction field is defined from this equation as

$$\vec{B} = \frac{\mu_0}{4\pi} \frac{q_1}{r^2} \vec{v}_1 \times \frac{\vec{r}}{r}.$$

It follows that the magnetic force, \vec{F}_m , may be written as

$$\vec{F}_m = q\vec{v} \times \vec{B}. \quad (3.19)$$

The **Lorentz force** is the total force experienced by a charge carrier in the magnetic induction field, \vec{B} , and is defined as

$$\vec{F} = q \left(\vec{E} + \vec{v} \times \vec{B} \right). \quad (3.20)$$

The force on a differential line element $d\vec{l}$ on a path C for steady current is given by

$$d\vec{F} = Nq|\vec{v}|A d\vec{l} \times \vec{B} = I d\vec{l} \times \vec{B},$$

since each carrier has current $Nq|\vec{v}|$ over cross-sectional area A of a steady current wire. This implies

$$\vec{F} = \oint_C I d\vec{l} \times \vec{B}. \quad (3.21)$$

The law of Biot and Savart gives the force, \vec{F}_2 exerted on a path C_2 from the influence of \vec{F}_1 coming from another path C_1 as

$$\vec{F}_2 = \frac{\mu_0}{4\pi} I_1 I_2 \oint_{C_1} \oint_{C_2} \frac{d\vec{l} \times (d\vec{l} \times (\vec{r}_2 - \vec{r}_1))}{|\vec{r}_2 - \vec{r}_1|^3}. \quad (3.22)$$

Combining equations (3.21) and (3.22) gives

$$\vec{B}(\vec{r}_2) = \frac{\mu_0}{4\pi} \int_V \frac{\vec{J}(\vec{r}_1) \times (\vec{r}_2 - \vec{r}_1)}{|\vec{r}_2 - \vec{r}_1|^3} dv. \quad (3.23)$$

Experimental observations based on the Biot Savart law imply that the all magnetic induction fields can be described in terms of a current distribution by equation (3.23). This observation implies that the divergence of a magnetic induction field, \vec{B} , for steady state currents is necessarily zero. In mathematical terms this is given by

$$\nabla \circ \vec{B} = 0, \quad (3.24)$$

which defines Maxwell's second equation.

Ampere's circuital law for steady currents is derived using equation (3.23) throughout chapter 8 of [10]. Ampere's law determines the behaviour of the magnetic induction field, \vec{B} , for steady currents. The mathematical formulation of this law is given by

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J}. \quad (3.25)$$

This means that the magnetic induction field lines curl in the direction of the current density. Therefore, magnetic induction fields are composed of charge carriers with position along closed paths. Each point on the path has a corresponding drift velocity \vec{v} . Ampere's law only applies for steady currents and the law is generalized in section 3.6 to obtain Maxwell's fourth equation.

Some materials experience electron spin in its physical configuration. This produces an ‘embedded current’ that magnetizes the material. The magnetic induction field, \vec{B} , corresponds to the part of the total magnetic field that is present for steady currents. The magnetic intensity field, denoted by \vec{H} , is defined to account for the total magnetic field relative to the magnetization of a material. It is defined as

$$\vec{H} = \frac{1}{\mu_0} \vec{B} + \vec{M}, \quad (3.26)$$

where \vec{M} is the magnetic field due to the magnetization of the material.

The magnetization of a material is experimentally derived for specific materials. Typically, materials exhibit a linear relationship with the magnetic induction field determined by the experimental constant, χ_m called the magnetic susceptibility. This linear relationship is given by $\vec{M} = \chi_m \vec{H}$. From the definition of the magnetic intensity field, (3.26), we obtain

$$\vec{H} = \frac{1}{\mu_0} \vec{B} + \chi_m \vec{H} \quad \Rightarrow \quad \vec{H} = \left(\frac{1 - \chi_m}{\mu_0} \right) \vec{B}. \quad (3.27)$$

The quantity

$$\frac{(1 - \chi_m)}{\mu_0}$$

is called the relative magnetic permeability and is defined by experiment for specific materials.

Let \vec{J}_M denote the current density of the magnetization field for a particular material, such that it satisfies Ampere’s circuital law, (3.25) with respect to its magnetization, \vec{M} . It follows from this law that

$$\vec{\nabla} \times \vec{M} = \vec{J}_M \quad \text{and} \quad \vec{\nabla} \times \vec{B} = \mu_0 (\vec{J} + \vec{J}_M).$$

This implies

$$\vec{\nabla} \times \vec{H} = \vec{\nabla} \times \left(\frac{1}{\mu_0} \vec{B} - \vec{M} \right) = \vec{\nabla} \times \frac{1}{\mu_0} \vec{B} - \vec{\nabla} \times \vec{M} = \left(\vec{J} + \vec{J}_M \right) - \vec{J}_M = \vec{J}, \quad (3.28)$$

which is only valid for steady currents.

Recall the definition of steady current with respect to charge density given by equation (3.15). Stoke's theorem from appendix A is applied with respect to (3.28) and a steady current, I , to obtain

$$\oint_S \vec{\nabla} \times \vec{H} \circ \hat{n} da = \oint_C \vec{H} \circ d\vec{l} = \int_S \vec{J} \circ \hat{n} da = I \quad (3.29)$$

3.5 Electromagnetic Induction

This subsection derives Maxwell's second equation and follows the same assumptions from [10]. Define the electromotive force around a circuit as

$$\mathcal{E} = \oint_C \vec{E} \circ d\vec{l}.$$

Recall the Lorentz force, (3.20) is the force on any test charge q .

A large number of experiments may be summarized with the equation

$$\mathcal{E} = -\frac{\partial \Phi}{\partial t},$$

where

$$\Phi = \int_S \vec{B} \circ \hat{n} da$$

is the flux of the magnetic induction field. Then it follows that

$$\int_S (\vec{\nabla} \times \vec{E}) \circ \hat{n} da = \oint_C \vec{E} \circ d\vec{l},$$

by Stoke's theorem in appendix A. Combining these equations gives

$$-\frac{\partial}{\partial t} \int_S \vec{B} \circ \hat{n} da = - \int_S \frac{\partial \vec{B}}{\partial t} \circ \hat{n} da. \quad (3.30)$$

Differentiating the left hand side and right hand side of equation (3.30) gives Maxwell's third equation,

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}.$$

It is rewritten using equation (3.27) as

$$\vec{\nabla} \times \vec{E} = -\mu_0 \frac{\partial \vec{H}}{\partial t}. \quad (3.31)$$

3.6 Ampere-Maxwell Law (Generalization of Ampere's Law)

For this subsection refer to figure 7 for justification of the arguments to be stated following from [10]. Recall equation (3.29) from section 3.4. Let the surfaces S_1

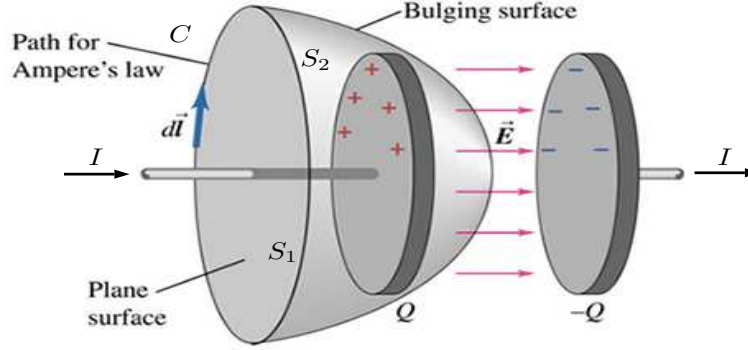


Figure 7: Inconsistency of Ampere's law for a charged capacitor, [17]

and S_2 correspond to the plane surface (disc) and the bulging surface (elliptic paraboloid), respectively. Let C be the path that Ampere's law is applied to according to figure 7. Applying Ampere's law for the magnetic intensity field from equation (3.30) to each surface in figure 7 implies that

$$\oint_C \vec{H} \circ d\vec{l} = \int_C \vec{J} \circ \hat{n} da = I \quad \text{and} \quad \oint_{S_1} \vec{H} \circ d\vec{l} = \int_{S_2} \vec{J} \circ \hat{n} da = 0. \quad (3.32)$$

At this point in developing these equations for figure 7, Ampere's law implies that $I = 0$, but it is observed that a non-zero current is present in the wire. The electric field existing between the plates must be accounted for.

The inconsistency in Ampere's law is resolved using the differential form of the law from equation (3.28). Observe that the divergence of any curl is necessarily zero for any vector field. This implies that the continuity equation for current density, given by (3.18), is also inconsistent because steady current satisfies equation (3.15).

Recall the definition of Maxwell's first equation, (3.6), and apply this to the

continuity equation to obtain

$$\vec{\nabla} \left[\vec{j} + \frac{\partial \vec{D}}{\partial t} \right] = 0, \quad (3.33)$$

for a sufficiently continuous displacement field, \vec{D} , as a function of space and time coordinates. Redefining Ampere's law as

$$\vec{\nabla} \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \quad (3.34)$$

removes the inconsistency and is the definition of Maxwell's fourth equation.

3.7 Maxwell's Wave Equation

Refer to [10] and [3] in this subsection. This subsection combines Maxwell's equations (3.11), (3.24), (3.31) and (3.34) to derive Maxwell's wave equation.

The glass core of the optical fiber being considered is non-conductive by virtue of how it is defined in section 2 to be isotropic, homogeneous, non-birefringent and lossless. This implies that the current density, \vec{J} , and charge density, ρ , is zero. In other words, the core is charge neutral with respect to Maxwell's equations and the transfer of charges within the core is unimpeded. Moreover, Maxwell's equations are independent of \vec{J} and ρ according to the derivations throughout section 3. Also note that the magnetization field, \vec{M} , is zero because charges embedded in the core are statically and dynamically neutral.

The first and fourth of Maxwell's equations reduce to

$$\vec{\nabla} \circ \vec{D} \quad \text{and} \quad \vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t}.$$

Computing the curl of equation (3.31) and (3.34) obtains

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -mu_0 \frac{\partial \vec{\nabla} \times \vec{D}}{\partial t} \quad \text{and} \quad \vec{\nabla} \times \vec{\nabla} \times \vec{H} = \frac{\partial \vec{\nabla} \times \vec{D}}{\partial t}. \quad (3.35)$$

It is necessary to rewrite the second wave equation with respect to \vec{E} since \vec{D} describes the charges embedded in the core, which behave neutrally. At the

same time, the equation (3.12) is applied to the equation on the right hand side of (3.35) to obtain

$$\vec{\nabla} \times \vec{\nabla} \times \vec{H} = \varepsilon \frac{\partial \vec{\nabla} \times \vec{E}}{\partial t}.$$

Observe that for any vector field, \vec{A} , the equation

$$\vec{\nabla} \times \vec{\nabla} \times \vec{A} = -\nabla^2 \vec{A} + \vec{\nabla} (\vec{\nabla} \circ \vec{A})$$

is satisfied by applying the definitions and theorems in appendix A. Applying this to the equations (3.35) gives

$$\nabla^2 \vec{E} - \mu_0 \varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} = \vec{\nabla} (\vec{\nabla} \circ \vec{E}) \quad \text{and} \quad \nabla^2 \vec{H} + \varepsilon \frac{\partial^2 \vec{H}}{\partial t^2} = \vec{\nabla} (\vec{\nabla} \circ \vec{H}) \quad (3.36)$$

From equation (3.11) and (3.7), the dot product is directly applied to obtain

$$\vec{\nabla} \circ \vec{D} = \varepsilon \vec{\nabla} \circ \vec{E} + \varepsilon \vec{E} \circ \vec{\nabla} \varepsilon = 0.$$

According to subsection 3.2, ε is a constant. Therefore, the gradient is zero, which implies that

$$\vec{\nabla} \circ \vec{E} = 0. \quad (3.37)$$

Applying (3.24) and (3.37) to (3.36) gives Maxwell's wave equation with respect to either \vec{E} or \vec{H} and is defined by the equations

$$\nabla^2 \vec{E} = \mu_0 \varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} \quad \text{and} \quad \nabla^2 \vec{H} = -\varepsilon \frac{\partial^2 \vec{H}}{\partial t^2}. \quad (3.38)$$

4 The Second Order Wave Equation

4.1 Hyperbolic PDEs

This section mainly follows [6] and course notes in partial differential equations. The method of characteristics is used to define hyperbolic partial differential equations (PDEs). Refer to appendices B.4 and B.5 for a discussion on PDEs and

the method of characteristics, respectively. We restrict our attention to second order semi-linear PDEs with 4-independent variables, since this is the nature of Maxwell's wave equation. The motivation for defining hyperbolic second order PDEs is to define Maxwell's wave equation as a hyperbolic PDE and present an example of the general wave equation in 1-dimensional space. The example illustrates the existence of unique solutions for 1-dimensional wave equation and what possible solutions may look like for light propagation in optical fibers. The example also illustrates a plane wave which is defined in subsection 4.4

First, we define hyperbolic PDEs. The general wave equation which represents Maxwell's wave equation is discussed in subsection . Then, the method of characteristics is applied to the 1-dimensional wave equation.

Definition 4.1. (*Hyperbolic PDE*) *In general, a PDE is defined as hyperbolic at particular sets of points in the domain, Ω , of the solution. It is hyperbolic at points where the Cauchy problem for the PDE, (having a hyper-surface solution at the point of interest), has a unique solution in the neighbourhood of the point on any non-characteristic hyper-plane passing through the point. This definition is interpreted from the results in appendix B.5.*

Semi-linear PDEs are classified as hyperbolic according to the principal part of the PDE since it determines the dominant behaviour of changes in the hyper-surface solution.

Definition 4.2. (*Restriction of Semi-Linear PDEs*) *We may write a second order semi-linear PDE with 4 independent variables in the form*

$$\begin{aligned} & a_1(x, y, z, t)u_{xx} + a_2(x, y, z, t)u_{xy} + a_3(x, y, z, t)u_{xz} + a_4(x, y, z, t)u_{xt} \\ & + a_5(x, y, z, t)u_{yy} + a_6(x, y, z, t)u_{yz} + a_7(x, y, z, t)u_{yt} + a_8(x, y, z, t)u_{zz} \\ & + a_9(x, y, z, t)u_{zt} + a_{10}(x, y, z, t)u_{tt} + b(x, y, z, t, u, u_x, u_y, u_z, u_t) = 0. \end{aligned}$$

The principal part of this semi-linear PDE is defined by

$$\begin{aligned} & a_1(x, y, z, t)u_{xx} + a_2(x, y, z, t)u_{xy} + a_3(x, y, z, t)u_{xz} + a_4(x, y, z, t)u_{xt} \\ & + a_5(x, y, z, t)u_{yy} + a_6(x, y, z, t)u_{yz} + a_7(x, y, z, t)u_{yt} + a_8(x, y, z, t)u_{zz} \\ & + a_9(x, y, z, t)u_{zt} + a_{10}(x, y, z, t)u_{tt} = 0, \end{aligned} \quad (4.1)$$

and corresponds to the highest derivatives present in the PDE, in general. The method of characteristics is applied to the principal part of a semi-linear PDE.

4.2 Second Order Wave Equation

The wave equation is most generally a second order linear PDE. It may be defined in up to n -dimensions. The wave equation is referred to as an $(n - 1)$ -dimensional wave equation to account for spatial dimensions whenever it describes oscillatory phenomenon in time and space. Let c be some constant and $(x, y, z, t) \in \Omega$ be the domain of the solution to the PDE for the 3-dimensional wave equation. Then it is defined by

$$u_{tt}(x, y, z, t) - c^2 \nabla^2 u(x, y, z, t) = 0, \quad (4.2)$$

where ∇^2 is the Laplacian, which satisfies

$$\nabla^2 u = u_{xx} + u_{yy} + u_{zz}$$

for $(x, y, z) \in \mathbb{R}^3$. Assuming sufficient continuity given by the Lipschitz condition, seeking points in the wave equation that are hyperbolic is equivalent to finding a transformation with respect to the method of characteristics such that characteristics intersect the solution uniquely in Ω .

The method of characteristics differs as the dimensions of the wave equation increase. The characteristics for the n -dimensional wave equation are n -dimensional hyper-planes. This can be shown by a generalization of the method from section 4.1, which shows this for the 3-dimensional wave equation. The

parametrization is found implicitly by analyzing initial and boundary conditions that are required to form a dispersive solution in section 6. The resulting characteristic equations describe light rays propagating in the core and depend on conditions at the core-cladding boundary.

4.3 1-Dimensional Wave Equation

As an example, consider the 1-dimensional wave equation,

$$u_{tt}(x, t) - c^2 u_{xx}(x, t) = 0, \quad x \in [0, L] \subset \mathbb{R}, \quad t \in [0, \infty) \subset \mathbb{R} \quad (4.3)$$

with initial conditions

$$u(x, 0) = f(x), \quad u_t(x, 0) = g(x)$$

and boundary conditions

$$u(0, t) = 0, \quad u(L, t) = 0.$$

First we use the method of characteristics for a general second order semi-linear PDE with two independent variables, (x, t) . The results are applied to the 1-dimensional wave equation in order to show it is hyperbolic for a particular transformation. In addition, it is shown that there exists a transformation that is independent of any initial or boundary conditions. Refer to appendix B.5 for applying the method of characteristics.

Define an arbitrary second order semi-linear PDE with two independent variables as

$$\mathcal{L}[u] = a(x, t)u_{xx} + b(x, t)u_{tt} + c(x, t)u_{xt} + \text{lower order terms} = 0. \quad (4.4)$$

Define the transformation $(x, t) \mapsto (\xi, \eta)$ and use the chain rule to substitute $U_{\xi\eta}$, $U_{\xi\xi}$, $U_{\eta\eta}$, U_ξ , U_η and $U(\xi, \eta) = u(x, t)$ into the PDE given by equation (4.1).

It follows that

$$\begin{aligned}\mathcal{L}[U] = & (a\xi_x^2 + b\xi_x\xi_t + c\xi_t^2)U_{\xi\xi} + (2a\xi_x\eta_x + b\xi_x\eta_t + b\xi_t\eta_x + 2c\xi_x\eta_t)U_{\xi\eta} \\ & + (a\eta_x^2 + b\eta_x\eta_t + \eta_t^2)U_{\eta\eta} + \text{lower order terms} = 0 \quad (4.5)\end{aligned}$$

Define,

$$\begin{aligned}a\xi_x^2 + b\xi_x\xi_t + c\xi_t^2 &= \alpha, \\ 2a\xi_x\eta_x + b\xi_x\eta_t + b\xi_t\eta_x + 2c\xi_x\eta_t &= \beta, \\ a\eta_x^2 + b\eta_x\eta_t + \eta_t^2 &= \gamma.\end{aligned}$$

Then, it follows that

$$\beta^2 - 4\alpha\gamma = (b^2 - 4ac)(\xi_x\eta_t - \xi_t\eta_x)^2,$$

where, $\xi_x\eta_t - \xi_t\eta_x$, is the Jacobian of the transformation. This Jacobian is necessarily non-zero and finite according to appendix B.5.

Moreover, the discriminant,

$$b^2 - 4ac, \quad (4.6)$$

of the PDE does not change sign after the transformation, $(x, t) \mapsto (\xi, \eta)$. This implies that, in general, the form in which the principle part of semi-linear equations are written is independent with respect to the sign of the discriminant. Consequently, we define and analyze points in the solution space for second order semi-linear PDEs independently based on the sign of the discriminant, (4.6). In particular, hyperbolic second order PDEs with 2 independent variables are those for which $b^2 - 4ac > 0$.

Upon applying a transformation, the PDE may be rewritten by substituting the transformed variables directly into the PDE. These are called canonical forms. Using a particular choice of transformation, we may represent any principle part

of a semi-linear PDE in exactly three ways, each corresponding to the sign of the discriminant. We may define a transformation satisfying

$$\alpha = a\xi_x^2 + b\xi_x\xi_y + c\xi_y^2 = b\xi_x\xi_y = 0 \quad (4.7)$$

$$\gamma = a\eta_x^2 + b\eta_x\eta_y + c\eta_y^2 = b\eta_x\eta_y = 0 \quad (4.8)$$

If these are true, suppose a and c are not both zero then

$$\begin{aligned} a \left(\frac{\xi_x}{\xi_y} \right)^2 + b \left(\frac{\xi_x}{\xi_y} \right) + c &= 0 \\ a \left(\frac{\eta_x}{\eta_y} \right)^2 + b \left(\frac{\eta_x}{\eta_y} \right) + c &= 0 \\ a + b \left(\frac{\eta_y}{\eta_x} \right) + c \left(\frac{\eta_y}{\eta_x} \right)^2 &= 0 \\ a + b \left(\frac{\xi_y}{\xi_x} \right) + c \left(\frac{\xi_y}{\xi_x} \right)^2 &= 0 \end{aligned} \quad (4.9)$$

These contradict the assumption that a, c not both zero. Then the transformation which satisfies (4.8) and (4.7) necessarily satisfies

$$\frac{\eta_x}{\eta_y} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad \frac{\xi_x}{\xi_y} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

For a valid transformation, the variables ξ and η are independent. Without loss of generality choose

$$\frac{\eta_x}{\eta_y} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad \frac{\xi_x}{\xi_y} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

Along a characteristic curve, one of the transformed variables are constant. Along a curve satisfying either $\xi(x, t) = \text{constant}$ or $\eta(x, t) = \text{constant}$, we obtain the differential equations

$$\frac{\partial \xi}{\partial x} + \frac{\partial \xi}{\partial y} \frac{dy}{dx} = 0 \quad \text{or} \quad \frac{\partial \eta}{\partial x} + \frac{\partial \eta}{\partial y} \frac{dy}{dx} = 0, \quad (4.10)$$

which imply that

$$\frac{dy}{dx} = -\frac{\xi_x}{\xi_y} = \frac{b - \sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad \frac{dy}{dx} = -\frac{\eta_x}{\eta_y} = \frac{b + \sqrt{b^2 - 4ac}}{2a}. \quad (4.11)$$

Therefore, the transformation $(x, t) \mapsto (\xi, \eta)$ is defined by these relations and also defines all characteristic curves for the PDE.

This transformation is now applied to the 1-dimensional wave equation from (4.3). From the 1-dimensional wave equation it follows that

$$a(x, t) = -c^2, \quad b(x, t) = 0, \quad c(x, t) = 1.$$

Then the equation is completely hyperbolic since $b^2 - 4ac = 4c^2 > 0$. The characteristic equations defining the characteristic curves follow from equations (4.11) and the values of a, b, c , where

$$\frac{dt}{dx} = \frac{1}{c}, \quad \text{or,} \quad \frac{dt}{dx} = \frac{-1}{c}, \quad (4.12)$$

implies that

$$x - ct = \eta(x, t) \quad \text{and} \quad x + ct = \xi(x, t) \quad (4.13)$$

are the two families of characteristic curves that describe the solution.

4.4 Plane Waves

A plane wave solution, known as the d'Alembert solution, is now derived for the 1-dimensional wave equation. The plane wave is defined after the example involving the d'Alembert solution. The canonical form for the 1-dimensional wave equation gives a plane wave solution by substituting transformations (4.13) into the wave equation. It follows that $\xi_x = 1$, $\xi_t = c$, $\eta_x = 1$ and $\eta_t = -c$ which implies

$$\begin{aligned} u_{tt} &= c^2 \left(\frac{\partial}{\partial \xi} - \frac{\partial}{\partial \eta} \right)^2 U = c^2 U_{\xi\xi} - 2U_{\xi\eta} + U_{\eta\eta} \\ u_{xx} &= \left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta} \right)^2 U = c^2 U_{\xi\xi} + 2U_{\xi\eta} + U_{\eta\eta} \end{aligned} \quad (4.14)$$

by applying the chain rule several times.

Inputting these into the 1-dimensional wave equation gives $-4c^2 U_{\xi\eta}$. Integrating this directly gives

$$u(x, t) = p(x + ct) - q(x - ct) = \quad (4.15)$$

for some functions p and q . We may write $u_t = cp' - cq'$ and apply the initial conditions $u_t(x, 0) = g(x)$ and treat this as an initial value problem independent of the boundary conditions, $u(0, t) = 0$ and $u(L, t) = 0$. The result is

$$\begin{aligned} p'(x) = \frac{1}{2c} (g(x) + cf'(x)) &\Rightarrow p(x) = \frac{1}{2}f(x) + \frac{1}{2c} \int_0^x g(s)ds + c_0, \quad c_0 = \text{const.} \\ &\Rightarrow p(x + ct) = \frac{1}{2}f(x + ct) + \frac{1}{2c} \int_0^{x+ct} g(s)ds + c_0. \end{aligned}$$

Similarly,

$$q(x - ct) = \frac{1}{2}f(x - ct) - \frac{1}{2c} \int_0^{x-ct} g(s)ds - c_0.$$

Inputting these into the solution from (4.15) gives the d'Alembert solution

$$\begin{aligned} u(x, t) &= \frac{1}{2} \left(f(x + ct) + f(x - ct) \right) + \frac{1}{2c} \left[\int_0^{x+ct} g(s)ds - \int_0^{x-ct} g(s)ds \right] \\ &= \frac{1}{2} \left(f(x + ct) + f(x - ct) \right) + \frac{1}{2c} \int_{x-ct}^{x+ct} g(s)ds \end{aligned}$$

The boundary conditions $u(0, t) = 0$ and $u(L, t) = 0$ allows us to determine the solution at every point in Ω using the transformation $(x, t) \mapsto (\xi, \eta)$. In general, by choosing any constants $(\eta_0, 0)$ and $(\xi, 0)$, the hyperbolic region bounded by non-intersecting characteristics is illustrated in figure 8 as the green region. This is called the domain of influence.

Solutions are defined along either of the characteristic lines $\eta = x - ct$ and $\xi = x + ct$ for any $\xi, \eta \in (\eta_0, \xi_0) \subset \mathbb{R}$. The solution is unique in the domain of influence by the method of characteristics, since neither characteristics intersect there.

The domain of dependence is defined as the interval that initial conditions are given on. In general this may be a set rather than an interval. Conditions at

this boundary completely determine the behaviour of the solution in the domain of influence. This is valid for any pair of points $(\eta_0, 0)$ and $(\xi_0, 0)$ in the initial value problem.

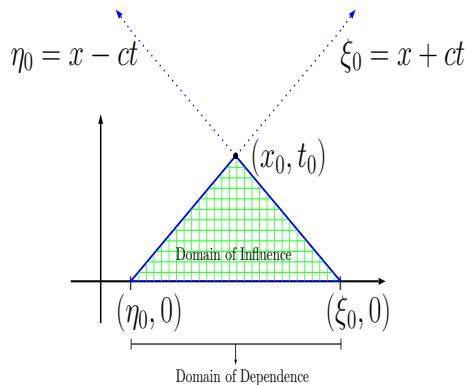


Figure 8: Region for the d'Alembert solution.

Now we may apply the boundary conditions

$$u(0, t) = 0, u(L, t) = 0.$$

This restricts the initial value problem to a fixed domain of dependence of length L , but the characteristic equations do not change. The result is shown in figure 9.

The domain of dependence is now the interval $(0, L)$. The d'Alembert solution may be integrated over the region bounded by any unique non-intersecting characteristic lines to find the solution there. By defining the odd periodic extensions of $f(x) = \bar{f}(x)$ and $\bar{g}(x)$ on the interval $(-L, L)$, we may write the d'Alembert solution as

$$u(x, t) = \frac{1}{2} \left(\bar{f}(x + ct) + \bar{f}(x - ct) \right) + \int_{x-ct}^{x+ct} \bar{g}(s) ds. \quad (4.16)$$

Doing this allows pairs of characteristics to be defined arbitrarily and the d'Alembert solution is transformed along the corresponding domain of influence

to give solutions at those points. This is done canonically to define the solution at any point in the column illustrated in the figure.

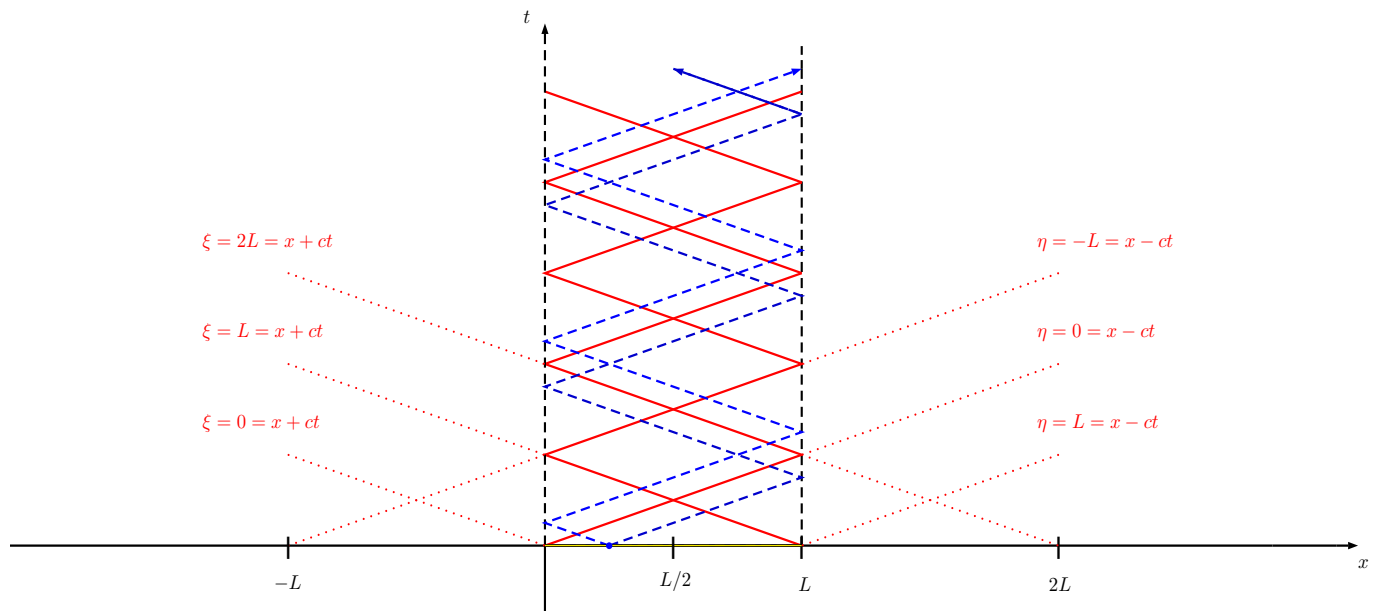


Figure 9: Solution to IBVP for the 1-dimensional wave equation.

This is valid for any $0 < x < L$ and any $t > 0$. Moreover, we may visualize the characteristics behaving like reflections between the boundaries $u(0, t) = 0$ and $u(L, t) = 0$. The solution at any point on the dashed blue line is determined entirely by the domain of dependence, (highlighted yellow), and the characteristic lines. In other words, characteristic lines are independent of the placement of the boundary conditions $u(0, t) = 0$ and $u(L, t) = 0$. The boundary conditions are only indicating where the solution is bounded on the x -axis and what its value is on that boundary.

The d'Alembert solution to the IBVP for the 1-dimensional wave equation is called a plane wave solution. The planes being referred to in the example here are defined by the lines $x + ct$ and $x - ct$. It turns out that for the 1-dimensional wave equation example, these are points rather than planes. In general, plane wave

solutions represent a type of solution for PDEs with arbitrarily many independent variables.

Plane waves are functions describing oscillations in space and time. In general, a plane wave solution to a hyperbolic PDE may be written as

$$u(x_1, \dots, x_n, t) = f(a\vec{x} - bt),$$

for position \vec{x} and time t . Let c be a constant. For a value of the solution $u(\vec{x}_0, t_0) = c$, the plane

$$a\vec{x}_0 - bt = c$$

propagates its points with a speed a/b at any given moment in time. This is developed further when defining dispersion in section 5.1. Plane wave solutions propagate in the domain of influence based entirely on the conditions placed on a domain of dependence from which the plane wave originates.

We seek solutions of this form for Maxwell's wave equation. In an optical fiber, a plane wave solution corresponds to oscillatory solutions in the core determined strictly by the initial beam of light entering the core through one of its circular faces and by characteristic equations. The characteristic equations are independent of the core-cladding boundary conditions since rays of light are parametrized and form characteristics.

5 Dispersion

5.1 Introduction to Dispersive Waves

In this section, the notion of dispersion is introduced, following [13]. Refer to the definitions in appendix B.4 and consider a PDE given by semi-linear differential operator $\mathcal{L}[\varphi]$, with solution u . Note that in this introductory subsection one

dimension of space is assumed, but in general, multiple dimensions of space may occur and are considered in the subsections which follow here.

The solution, φ , is dispersive whenever it is an oscillatory solution of the form,

$$\varphi = a \cos(\kappa x - \omega t), \quad (5.1)$$

where the frequency, ω , is a real function of the wavenumber, κ , and the function $\omega(\kappa)$ is determined from the particular PDE being considered. Note that this is a plane wave solution.

When solutions to a PDE are dispersive, the dispersion is characterized by the propagation of wavenumbers. The phase speed in the oscillatory solution, (5.1) is given by

$$\frac{\omega(\kappa)}{\kappa}.$$

The solution is said to be dispersive if this phase speed is not a constant and depends on κ . The term dispersion refers to the fact that a more general solution consists of the superposition of more than one oscillatory solution, like (5.1), each corresponding to different possible values of κ . This implies that different modes, (corresponding to possible values of κ), may propagate the respective wavenumbers, (determined by κ), with a different velocity. Plane wave solutions *may* exist when they are oscillatory and could be ‘changing’ with respect to time according to the dispersion. The plane wave solution is discussed further in section 6.

First, in subsection 5.2, the dispersion is defined with respect to linear PDEs. In subsection 5.3, the notion of group velocity is defined and corresponds to the possibly varying phase velocities of frequency modes $\omega(\kappa)$.

5.2 Linear Dispersive Waves

This section defines dispersion, so that we may find possible dispersive solutions to Maxwell's wave equation following [13]. The classification of dispersive waves is made on the type of solution rather than on the differential equation. Let A be a complex constant. Let ω be a real constant. Let $\vec{\kappa}$ and \vec{x} be real valued vectors in cylindrical coordinates, where $\vec{\kappa}$ is constant. Dispersive waves are recognized by wave train solutions of the form

$$\phi(\vec{x}, t) = Ae^{i\vec{\kappa} \circ \vec{x} - i\omega t}. \quad (5.2)$$

Suppose such a solution needs to satisfy a *linear* partial differential equation in cylindrical coordinates. The differential equation gives a relation

$$F\left(r, \theta, z, t, \phi, \frac{\partial \phi}{\partial r}, \frac{\partial \phi}{\partial \theta}, \frac{\partial \phi}{\partial z}, \frac{\partial \phi}{\partial t}, \frac{\partial^2 \phi}{\partial r \partial \theta}, \frac{\partial^2 \phi}{\partial r \partial z}, \frac{\partial^2 \phi}{\partial r \partial t}, \frac{\partial^2 \phi}{\partial \theta \partial z}, \frac{\partial^2 \phi}{\partial \theta \partial t}, \frac{\partial^2 \phi}{\partial z \partial t}, \frac{\partial^2 \phi}{\partial r^2}, \frac{\partial^2 \phi}{\partial \theta^2}, \frac{\partial^2 \phi}{\partial z^2}, \frac{\partial^2 \phi}{\partial t^2}\right) = 0,$$

which is a linear function of $\phi(\vec{x}, t)$ and its partial derivatives. Note that $\phi(\vec{x}, t)$ is separable with respect to x_r, x_θ, x_z, t since $\vec{\kappa} \circ \vec{x}$ is a linear combination of x_r, x_θ, x_z . This implies that mixed partial derivatives are equal in F . Taking partial derivatives with respect to either t or \vec{x} will produce factors of either ω or simultaneously x_r, x_θ, x_z , respectively.

The relation $F(\vec{*}) = 0$ reduces to

$$G(\omega, \kappa_r, \kappa_\theta, \kappa_z) = 0 \quad (5.3)$$

by factoring out $\phi(\vec{x}, t)$. By construction of G , it is a linear equation with respect to $\omega, \kappa_r, \kappa_\theta, \kappa_z$ and is called the dispersion relation. The roots of G with respect to ω are given by relations

$$\omega_j = W_j(\vec{\kappa}) = W_j(\kappa_r, \kappa_\theta, \kappa_z),$$

where ω_j is a distinct root with corresponding relation W_j .

Let ω represent an arbitrary root ω_j . To obtain real solutions,

$$Re(\phi) = |A| \cos(\vec{\kappa} \circ \vec{x} - \omega t + \arg A) = \varphi \quad (5.4)$$

is an oscillatory wave with respect to time and space. Let $\vartheta = \vec{\kappa} \circ \vec{x} - \omega t$, where $\vec{x} = x_1 \hat{r} + x_2 \hat{\theta} + x_3 \hat{z}$ is the position of a point in 3-dimensional cylindrical coordinate space. Then ϑ accounts for any change in spatial position with respect to time in φ as it oscillates, since $\arg A$ is constant. By definition, the variable ϑ is the phase of φ . Fixing $\vartheta = \text{constant}$ defines a spatial plane that changes with respect to time. These are called phase surfaces.

Next, the oscillatory behaviour of φ in the spatial and temporal contexts with respect to phase surfaces will be considered. This makes sense because we may separate $Re(\phi)$ as a product of a time dependent oscillatory functions and spatially dependent oscillatory functions by using the multiplicative properties of the exponential function prior to taking the real part.

When dealing with spatial oscillations in radial units, we refer to one cycle of the wave as its wavelength and the number of cycles per 2π degrees as its wavenumber. Similarly, one cycle of a temporal wave is the period and the number of cycles per 2π radians is the frequency. The spatial gradient of a phase surface is the gradient of the plane

$$\vartheta|_{t=0} = \text{constant},$$

for some constant.

For any constant defining a phase surface and any fixed point in time, the points of the surface are linearly shifted by the constant and the time dependent part of ϑ . Therefore, the position of an arbitrary point of the surface for some instant in time is determined by $\vec{\kappa} \circ \vec{x} = C$ for any constant C . Hence, we may equivalently let $C = 0$ to define the position of points for any phase surface at an instant in time for some dispersion mode, ω . The spatial gradient of the plane $\vartheta = C$ is given by the vector $\vec{\kappa}$ by definition of the gradient for the plane

$\vec{\kappa} \circ \vec{x} = 0$ and by observing it is the same for any constant C . Therefore, the spatial gradient of ϑ is always $\vec{\kappa}$ and has magnitude $\kappa = |\vec{\kappa}|$ for some dispersion mode ω . Similarly, we may take the gradient with respect to t . The gradient simplifies as ϑ_t . By construction of $\vec{\nabla}\vartheta$ and by definition of the gradient, κ is the average number of crests for φ per 3-dimensional unit distance.

Also, $-\vartheta_t$ is the average number of crests in the real solution, φ , per unit time in the direction of κ . Using radial units we may define the wavenumber and frequency of a phase surface to be κ and ω respectively. Moreover, we may define wavelength and period as $\lambda = 2\pi/\kappa$ and $\tau = 2\pi/\omega$ respectively. In conclusion, each wavenumber ω has a continuum of phase surfaces $\vartheta = \text{constant}$.

Consider the velocity of an arbitrary 3-dimensional point on a phase surface, $\vartheta = C$, for any constant C . The velocity of such a point is *in the direction that is perpendicular to the phase surface* and has a *magnitude* of wavelength per period given by ω/κ . Therefore, the phase velocity is defined as

$$\vec{c} = \frac{\omega}{\kappa} \hat{\kappa},$$

where $\hat{\kappa} = \vec{\kappa}/\kappa$ is the unit vector in the direction of $\vec{\kappa}$. Note that the phase velocity only accounts for the spatial variation in the phase surfaces with respect to time in the direction of $\hat{\kappa}$ for a particular frequency mode ω .

In [13], the author restricts the definition of dispersion to accommodate the following two cases. The first case may be observed by the heat equation $\varphi_t = \nabla^2 \varphi$ giving complex modes of dispersion of the form $\omega = -i\kappa^2$ when a solution of the form (5.2) is applied. It follows that the real part of the solution becomes

$$\text{Re}(\varphi) = \kappa|A| \cos(\vec{\kappa} \circ \vec{x} + \arg A)$$

which is oscillatory only with respect to space. To maintain wavelike solutions in the first case, the class of linear dispersion is restricted to equations with disper-

sion relations that have only real solutions. This is equivalent to the restriction

$$W(\vec{\kappa}) \neq 0.$$

The second case is observed whenever the group velocity is constant. The group velocity is defined in section 5.3. The restriction for the second case is only stated here and the justification can be deduced from the arguments in section 5.3. The second restriction is

$$\det \left| \frac{\partial^2 W}{\partial \kappa_i \partial \kappa_j} \right| \neq 0. \quad (5.5)$$

One may also refer to section 11.1 of [13] to see the correspondence between the dispersion relation and their *linear* differential equations is one-to-one. This means that one may be obtained from the other and vice versa. Therefore, the properties derived from a linear dispersive relation correspond uniquely to a linear differential equation and vice versa.

5.3 Group Velocity for Linear PDEs

This subsection also follows [13]. We define the group velocity because a solution can only be dispersive if it admits group velocity behaviour. The group velocity behaviour may possibly be trivial for the simplest cases of a dispersive solution, which is shown in section 6.

First, the group velocity is derived in 1-dimensional space and then extended to 3-dimensional space. Recall the solution of the form of equation (5.2) with real part as in (5.4) such that one dimension of space is assumed. Assume a phase function $\vartheta(x_1, t)$ exists and describes the *local* action of points in a phase surface with respect to time. By local, this means at the level of continuous dependence of the phase ϑ on points, x_1 , and time, t .

The local wavenumber and local frequency are defined by

$$\kappa = \vartheta_{x_1} \quad (5.6)$$

and

$$\omega = -\vartheta_t. \quad (5.7)$$

Let an arbitrary dispersion relation corresponding to a linear differential equation exist. Let the corresponding dispersion relation be given as

$$\omega = W(\kappa).$$

Now we determine an equation that allows us to solve for $\kappa(x, t)$. It follows from differentiating (5.6) and (5.7) that

$$\frac{\partial \kappa}{\partial t} = \vartheta_{x_1 t}$$

and

$$\frac{\partial \omega}{\partial x_1} = -\vartheta_{tx_1},$$

where mixed partial derivatives are equal.

$$\frac{\partial \kappa}{\partial t} + \frac{\partial \omega}{\partial x_1} = \frac{\partial \kappa}{\partial t} + \frac{\partial W(\kappa)}{\partial x_1} = \frac{\partial \kappa}{\partial t} + \frac{dW(\kappa)}{d\kappa} \frac{\partial \kappa}{\partial x_1} = 0, \quad (5.8)$$

where

$$C(\kappa) = W'(\kappa)$$

is the propagation velocity of the wavenumber $\kappa(x_1, t)$. This propagation velocity defines the group velocity. Equation (5.8) is a non-linear version of the simple wave equation. Now the group velocity is extended to the case of 3-dimensional space in cylindrical coordinates. The assumed phase function becomes a function $\vartheta(x_1, x_2, x_3, t)$, where $\vec{x} = x_1 \hat{r} + x_2 \hat{\theta} + x_3 \hat{z}$. The local frequency ω and local wavenumber $\vec{\kappa} = \kappa_1 \hat{r} + \kappa_2 \hat{\theta} + \kappa_3 \hat{z}$ are defined by

$$\omega = -\frac{\partial \vartheta}{\partial t} \quad (5.9)$$

and

$$\kappa_i = \frac{\partial \vartheta}{\partial x_i},$$

for $i = 1, 2, 3$. In the case of 3-dimensional space, the dispersion relation for some linear differential equation becomes

$$\omega = W(\kappa_1, \kappa_2, \kappa_3, x_1, x_2, x_3, t).$$

Although, different dispersion modes may have different group velocities and phase speeds, a mode itself is dispersive only if W is a nonlinear function of \vec{k}_j .

Following a similar argument as for the 1-dimensional case, we may differentiate ω with respect to x_1, x_2, x_3 and differentiate k_i with respect to k_j , for $i, j \in \{1, 2, 3\}$ and $i \neq j$. For each $i \in \{1, 2, 3\}$ we obtain a pair of equations

$$\frac{\partial \kappa_i}{\partial t} + \frac{\partial \omega}{\partial x_i} = 0 \quad (5.10)$$

and

$$\frac{\partial \kappa_i}{\partial x_j} - \frac{\partial \kappa_j}{\partial x_i} = 0 \quad (5.11)$$

for any $j \in \{1, 2, 3 \mid i \neq j\}$. Combining (5.9) and (5.10) gives

$$\frac{\partial \kappa_i}{\partial t} + \frac{\partial W}{\partial x_i} = 0 = \frac{\partial \kappa_i}{\partial t} + \frac{\partial W}{\partial \kappa_j} \frac{\partial \kappa_j}{\partial x_i} + \frac{\partial W}{\partial x_i},$$

for any $i \neq j$. This may be written as

$$\frac{\partial \kappa_i}{\partial t} + \frac{\partial W}{\partial \kappa_j} \frac{\partial \kappa_j}{\partial x_i} = -\frac{\partial W}{\partial x_i}$$

and (5.11) may be applied to obtain

$$\frac{\partial \kappa_i}{\partial t} + \frac{\partial W}{\partial \kappa_i} \frac{\partial \kappa_j}{\partial x_i} = -\frac{\partial W}{\partial x_i}. \quad (5.12)$$

In 3-dimensional space, group velocity is defined by determining $\vec{\kappa}$ using equation (5.12) for each $i \in \{1, 2, 3\}$. In doing so, the components

$$C_j(\kappa_1, \kappa_2, \kappa_3, x_1, x_2, x_3, t) = \frac{\partial W(\kappa_1, \kappa_2, \kappa_3, x_1, x_2, x_3, t)}{\partial \kappa_j}$$

define the group velocity \vec{C} . Solutions are dispersive only if the group velocity is not the same as the phase velocity, i.e. $\vec{c} \neq \vec{C}$.

6 Guided Modes in Cylindrical Waveguides

6.1 Guided Rays of Light and Dispersion

This section is dedicated to finding the characteristic equations describing the distribution of light rays propagating through the core of an optical fiber. This section mainly follows [3], [8] and [2] in order of most to least relevant according to the arguments presented here. Recall the discussion of light ray propagation from section 2.

In general, a guided mode refers to a particular dispersive solution governing the paths of light rays propagating along phase surfaces. In other words, a mode describes a particular path that a ray or families of rays may take. The dispersion modes described in section 5.2 correspond to a whole range of light ray modes such that groups of light rays may belong to the same phase surface and propagate with a group velocity.

Suppose a beam of light is coupled to the circular face of the optical fiber and is the domain of dependence for a plane wave solution. The possible dispersive solutions to the wave equation inside the core are found by tracing the paths of individual rays of light. By parametrizing rays of light, the initial and boundary conditions required for producing valid characteristic equations can be found. These characteristics, which describe the paths of each individual ray of light, form a unique dispersive solution based only on the light entering at the domain of dependence. The domain of dependence is the circular face of the core where light enters in the form of rays at any angle smaller than the acceptance angle.

We assume small distances and consequently short periods of time in order to make the assumption of linearity. The short distance assumption is required because the non-linearity is inherent of the material and occurs over long distances as solutions on the paths of light rays ‘disperse’. It is inherited by the

dependence of the relative permittivity on frequency, where the assumption of linearity was made in section 3.2. Therefore, non-linearity cannot be neglected over long distances and so the assumption of short distances is required.

6.2 Characteristic Equations for Guided Modes

This subsection derives the characteristic equations that correspond to the paths of light rays. We continue to follow [8] and [3]. In the case of linear dispersive waves, it is possible to work with uniform plane wave solutions and derive boundary conditions at the core-cladding interface satisfying a particular dispersive solution with respect to planar phase surfaces. We assume the solution is separable with respect to coordinates r , θ , z and t .

Monochromatic waves are defined as those which have all fields \vec{E} , \vec{B} , \vec{D} , \vec{H} , \vec{P} and \vec{M} characterized by a single frequency ω . Monochromatic waves are represented in cylindrical coordinates by dispersive solutions of the form

$$\vec{E}(r, \theta, z, t) = \vec{E}(r, \theta, z)e^{-i\omega t} \quad \text{and} \quad \vec{H}(r, \theta, z, t) = \vec{H}(r, \theta, z)e^{-i\omega t}$$

whenever the time dependence of the field is taken as $e^{-i\omega t}$. Monochromatic waves are assumed so that dispersive solutions may be found whenever such solutions exist for a single frequency of light. In general, many frequencies can occur when multiple sources of light are present, but we consider the case of one source representing a monochromatic wave.

Assuming that the wave in the core is monochromatic, the solutions from (6.1) are inputted into Maxwell's third and fourth equations (3.31) and (3.34), respectively. Then,

$$\vec{\nabla} \times \vec{E} = -\mu_0 \frac{\partial \vec{H}}{\partial t} \quad \text{and} \quad \vec{\nabla} \times \vec{H} = \varepsilon \frac{\partial \vec{E}}{\partial t}, \quad (6.1)$$

where $\varepsilon = \varepsilon_0 n^2$ for the refractive index, n .

The mean Poynting vector is derived in [3] and is shown there to be along the z -axis for electric and magnetic fields that are polarized perpendicularly to it along the (r, θ) -plane. It represents the direction in which electromagnetic waves propagate when the electric and magnetic fields are polarized along the transverse directions r and θ .

The following derivation uses the assumption that the light is polarizable in these directions in order to assume plane wave solutions to Maxwell's wave equation of the form

$$\vec{E}(r, \theta, z, t) = \vec{E}(r, \theta)e^{i(\omega t - \beta z)} \quad \text{and} \quad \vec{H}(r, \theta, z, t) = \vec{H}(r, \theta)e^{i(\omega t - \beta z)} . \quad (6.2)$$

We now follow the arguments from [3], [8] and [2] to find the characteristic equations describing the modal distribution of light rays. These are exactly the characteristic equations that correspond to the dispersive solutions within the core satisfying prescribed boundary conditions. The use of boundary conditions for interpreting group velocity behaviour of solutions is shown at the end of this section. The validity of solutions is also discussed further in this section.

The cylindrical component version of the gradient, derived in appendix C, must be applied to the equations (6.1). In the following derivations, the details related to the first equation from (6.1) are shown and results related to the second equation from (6.1) are stated, since they are similar and can be found in [3] and

[8]. Applying the cylindrical gradient gives

$$\begin{aligned}
\vec{\nabla} \times \vec{E} e^{i(\omega t - \beta z)} &= \left[\frac{1}{r} \frac{\partial \vec{e}_z(t)}{\partial \theta} - \frac{\partial \vec{e}_\theta(t)}{\partial z} \right] \hat{r} + \left[\frac{\partial \vec{e}_r(t)}{\partial z} - \frac{\partial \vec{e}_z(t)}{\partial r} \right] \hat{\theta} \\
&\quad + \left[\frac{1}{r} \frac{\partial \vec{e}_r(t)}{\partial \theta} + \frac{\vec{e}_\theta(t)}{r} - \frac{\partial \vec{e}_\theta(t)}{\partial r} \right] \hat{z} \\
&= \left[\frac{1}{r} \frac{\partial E_z}{\partial \theta} e^{i(\omega t - \beta z)} + i\beta \frac{\partial E_\theta}{\partial z} e^{i(\omega t - \beta z)} \right] \hat{r} \\
&\quad + \left[-i\beta \frac{\partial E_r}{\partial z} e^{i(\omega t - \beta z)} - \frac{\partial E_z}{\partial r} e^{i(\omega t - \beta z)} \right] \hat{\theta} \\
&\quad + \left[\frac{1}{r} \frac{\partial E_r}{\partial \theta} e^{i(\omega t - \beta z)} + \frac{E_\theta}{r} e^{i(\omega t - \beta z)} - \frac{\partial E_\theta}{\partial r} e^{i(\omega t - \beta z)} \right] \hat{z} \quad (6.3)
\end{aligned}$$

$$\begin{aligned}
-\mu \frac{\partial \vec{h}(t)}{\partial t} &= -\mu \frac{\partial (\vec{H} e^{i(\omega t - \beta z)})}{\partial t} = -\mu \left[i\omega H_r e^{i(\omega t - \beta z)} \right] \hat{r} - \mu \left[i\omega H_\theta e^{i(\omega t - \beta z)} \right] \hat{\theta} \\
&\quad - \mu \left[i\omega H_z e^{i(\omega t - \beta z)} \right] \hat{z} \quad (6.4)
\end{aligned}$$

Observe that $\partial E_\theta / \partial z = E_\theta$ and $\partial E_r / \partial z = E_r$, since propagation is assumed in the z direction. Also, observe by the chain rule that

$$\frac{E_\theta}{r} - \frac{\partial E_\theta}{\partial r} = \frac{1}{r} \left(\frac{\partial(r E_\theta)}{\partial r} - \frac{\partial E_r}{\partial \theta} \right)$$

By applying these last two observations, factoring out $e^{i(\omega t - \beta z)}$ and equating the components in equations (6.3) and (6.4) we obtain the equations

$$\begin{aligned}
\frac{1}{r} \frac{\partial E_z}{\partial \theta} + i\beta \frac{\partial E_\theta}{\partial z} &= \frac{1}{r} \frac{\partial E_z}{\partial \theta} + i\beta E_\theta = -\mu_0 i\omega H_r, \\
-i\beta \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} &= -i\beta E_r - \frac{\partial E_z}{\partial r} = -\mu_0 i\omega H_\theta, \\
\frac{\partial E_\theta}{\partial r} + \frac{E_\theta}{r} - \frac{1}{r} \frac{\partial E_r}{\partial \theta} &= \frac{1}{r} \left(\frac{\partial(r E_\theta)}{\partial r} - \frac{\partial E_r}{\partial \theta} \right) = -\mu_0 i\omega H_z.
\end{aligned} \quad (6.5)$$

A similar procedure is followed for the curl corresponding to the second equa-

tion from (6.1) and results in the equations

$$\begin{aligned}\frac{1}{r} \frac{\partial H_z}{\partial \theta} + i\beta H_\theta &= i\varepsilon_0 \omega n^2 E_r, \\ -i\beta H_r - \frac{\partial H_z}{\partial r} &= i\varepsilon_0 \omega n^2 E_\theta, \\ \frac{1}{r} \left(\frac{\partial(rH_\theta)}{\partial r} - \frac{\partial H_r}{\partial \theta} \right) &= i\varepsilon_0 \omega n^2 E_z.\end{aligned}\tag{6.6}$$

The boundary conditions for Maxwell's wave equations are such that the respective electric and magnetic fields are continuous at the core-cladding boundary. In other words, the tangential components of the electric and magnetic fields are equal between boundaries. In mathematical terms this means

$$E_t^{(core)} = E_t^{(cladding)} \quad \text{and} \quad H_t^{(core)} = H_t^{(cladding)}, \quad \forall t \in \{\theta, z\}, r = a. \tag{6.7}$$

The equations for curl are derived above since they are required for applying the boundary conditions (6.7).

The system of equations from (6.5) and (6.6) can be combined and rewritten with respect to the transversal components r, θ of the electric and magnetic fields. It follows that

$$\begin{aligned}E_\theta &= \frac{-i}{\beta} \left(-i\mu_0 \omega H_r - \frac{1}{r} \frac{\partial E_z}{\partial \theta} \right) \\ E_r &= \frac{i}{\beta} \left(-i\mu_0 \omega H_\theta + \frac{\partial E_z}{\partial r} \right)\end{aligned}$$

and

$$\begin{aligned}H_\theta &= \frac{-i}{\beta} \left(i\varepsilon_0 \omega n^2 E_r - \frac{1}{r} \frac{\partial H_z}{\partial \theta} \right) \\ H_r &= \frac{i}{\beta} \left(i\varepsilon_0 \omega n^2 E_\theta + \frac{\partial H_z}{\partial r} \right).\end{aligned}$$

These are now combined to obtain equations for E_θ, E_r or H_θ, H_r that only depend on components H_z and E_z . Inputting the second pair of equations into

the first gives

$$\begin{aligned} E_\theta &= \frac{-i}{\beta} \left(-i\mu_0\omega \left(\frac{i}{\beta} \left(i\varepsilon_0\omega n^2 E_\theta + \frac{\partial H_z}{\partial r} \right) \right) - \frac{1}{r} \frac{\partial E_z}{\partial \theta} \right) \\ &= \frac{-i}{\beta} \left(\frac{i\mu_0\varepsilon_0\omega^2 n^2}{\beta} E_\theta + \frac{\mu_0\omega}{\beta} \frac{\partial H_z}{\partial r} - \frac{1}{r} \frac{\partial E_z}{\partial \theta} \right) \end{aligned}$$

and

$$\begin{aligned} E_r &= \frac{i}{\beta} \left(-i\mu_0\omega \left(\frac{-i}{\beta} \left(i\varepsilon_0\omega n^2 E_r - \frac{1}{r} \frac{\partial H_z}{\partial \theta} \right) \right) + \frac{\partial E_z}{\partial r} \right) \\ &= \frac{i}{\beta} \left(\frac{-i\mu_0\varepsilon_0\omega^2 n^2}{\beta} E_r + \frac{\mu_0\omega}{r} \frac{\partial H_z}{\partial \theta} + \frac{\partial E_z}{\partial r} \right). \end{aligned}$$

These two equations imply that

$$\left(1 - \frac{\mu_0\varepsilon_0\omega^2 n^2}{\beta^2} \right) E_\theta = \frac{-i}{\beta} \left(\frac{\mu_0\omega}{\beta} \frac{\partial H_z}{\partial r} - \frac{1}{r} \frac{\partial E_z}{\partial \theta} \right)$$

and

$$\left(1 - \frac{\mu_0\varepsilon_0\omega^2 n^2}{\beta^2} \right) E_r = \frac{i}{\beta} \left(\frac{\mu_0\omega}{r\beta} \frac{\partial H_z}{\partial \theta} + \frac{\partial E_z}{\partial r} \right).$$

These may be rewritten as

$$E_\theta = \frac{-i}{(k^2 n^2 - \beta^2)} \left(\frac{\beta}{r} \frac{\partial E_z}{\partial \theta} - \omega\mu_0 \frac{\partial H_z}{\partial r} \right) \quad (6.8)$$

and

$$E_r = \frac{-i}{(k^2 n^2 - \beta^2)} \left(\beta \frac{\partial E_z}{\partial r} + \frac{\omega\mu_0}{r} \frac{\partial H_z}{\partial \theta} \right). \quad (6.9)$$

Similarly,

$$H_\theta = \frac{-i}{(k^2 n^2 - \beta^2)} \left(\frac{\beta}{r} \frac{\partial H_z}{\partial \theta} + \omega\varepsilon_0 n^2 \frac{\partial E_z}{\partial r} \right) \quad (6.10)$$

and

$$H_r = \frac{-i}{(k^2 n^2 - \beta^2)} \left(\beta \frac{\partial H_z}{\partial r} - \frac{\omega\varepsilon_0 n^2}{r} \frac{\partial E_z}{\partial \theta} \right), \quad (6.11)$$

where $k = \omega\sqrt{\mu_0\varepsilon_0}$ and we define $\gamma^2 = n^2 k^2 - \beta^2$.

The components E_z and H_z of the electric and magnetic fields each satisfy Maxwell's wave equation. Inputting the solutions of the form (6.2) gives the

following equations from the z components;

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r} \frac{\partial E_z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 E_z}{\partial \theta^2} + (k^2 n(r, \theta)^2 - \beta^2) E_z = 0 \quad (6.12)$$

$$\frac{\partial^2 H_z}{\partial r^2} + \frac{1}{r} \frac{\partial H_z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 H_z}{\partial \theta^2} + (k^2 n(r, \theta)^2 - \beta^2) H_z = 0 \quad (6.13)$$

In the physics literature on the subject of optical fibers, the transverse electric (TE) and transverse magnetic (TM) modes correspond to the assumptions $E_z = 0$ and $H_z = 0$, respectively. The Hybrid modes correspond to the assumption that $H_z \neq 0$ and $E_z \neq 0$.

In order for solutions to the wave equation to exist, they must depend continuously on the data and be separable. Hence, the electric and magnetic fields along this direction are described by

$$E_z(r, \theta) = R(r)\Theta(\theta) = H_z(r, \theta).$$

This splits the wave equation along the z direction from (6.13) into two the equations

$$\frac{d^2 \Theta}{d\theta^2} = -m^2 \Theta \quad (6.14)$$

and

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} + \left(\gamma^2 - \frac{m}{r^2} \right) R = 0. \quad (6.15)$$

Note that the boundary conditions derived from total internal reflection and Goss-Hanchen shift are independent of position which allows separation of variables to take place with consistently defined boundary conditions.

Cylindrical symmetry implies that we require Θ to be an oscillatory function in order to preserve continuous dependence. Hence, we may write

$$\Theta(\theta) = C_1 e^{i(m\theta + \alpha)},$$

for some constants C_1, m, α . The real part is taken when solutions are required, but in general the electromagnetic phase is complex valued. For continuous dependence to be preserved, it is necessary that $m \in \mathbb{Z}$. The two families of real

solutions for Θ are out of phase by $\pi/2$. This also occurs with the electric and magnetic fields. So, without loss of generality let

$$E_z = AR(r) \cos(m\theta + \alpha) \quad \text{and} \quad H_z = BR(r) \sin(m\theta + \alpha),$$

for some constants A and B .

Equation (6.15) is either Bessel's equation or the modified Bessel equation depending on whether or not r is real or complex valued, respectively. The complex value of r comes from the fact that the electric and magnetic fields are perpendicular to one another. So, whenever it is complex valued it has no real part and vice versa. Bessel's equation and the modified Bessel's equation is discussed in appendix B.2 and B.3. The real and complex values of r correspond to the Bessel's equation and the modified Bessel's equation respectively.

The solutions to the Bessel equations are derived with convergent series solutions in appendix B.2 and B.3. Refer to the graphs of the Bessel's functions and modified Bessel's functions in figure 10.

It is known that the Bessel functions of the first and second kind behave like functions of sine and cosine. Hence, these are candidates for light propagation in the core in the form of a plane wave solution. The logarithmic part of the series solution for Y_m indicates that the solution goes to $-\infty$ as it approaches zero. This could not possibly describe a finite light beam entering the core. Therefore, the only possible solution in the core is J_m and has real solutions whenever $\beta^2 < n_1^2 k^2$. Recall that the refractive index in the core and cladding are given by n_1 and n_0 respectively. Define

$$\begin{aligned} E_z &= AJ_m(ur) \cos(m\theta + \alpha), \\ H_z &= BJ_m(ur) \sin(m\theta + \alpha), \\ u^2 &= n_1^2 k^2 - \beta. \end{aligned} \tag{6.16}$$

In the cladding, we require solutions to approach zero as $r \rightarrow \infty$. This is only satisfied by the modified Bessel function of the second kind $K_m(r)$ since the

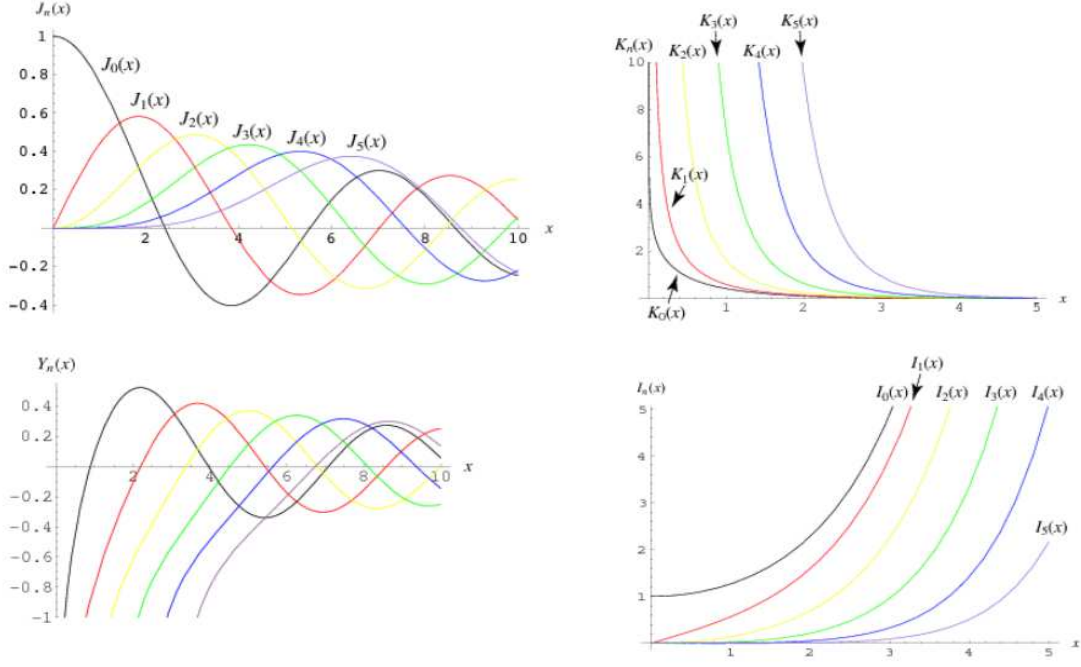


Figure 10: Bessel's functions and modified Bessel's functions, [18], [19]. [20]. [21].

modified Bessel function of the first kind goes to infinity as $r \rightarrow \infty$. For real solutions in the cladding, it is necessary that $\beta^2 > n_0^2 k^2$. Define

$$\begin{aligned} E_z &= CK_m(wr) \cos(m\theta + \alpha), \\ H_z &= DK_m(wr) \sin(m\theta + \alpha), \\ w^2 &= \beta - n_0^2 k^2. \end{aligned} \tag{6.17}$$

We may now use these to define the components of the electric and magnetic field in the directions of r and θ in equations (6.8)-(6.11). It follows that

$$\begin{aligned} E_\theta^{(core)} &= \frac{-i}{u^2} \left(Am \frac{\beta}{r} J_m(ur) + B\omega\mu_0 \frac{\partial J_m(ur)}{\partial r} \right) \sin(m\theta + \alpha), \\ E_\theta^{(cladding)} &= \frac{-i}{w^2} \left(Cm\beta \frac{\partial K_m(wr)}{\partial r} + D \frac{\omega\mu_0}{r} \frac{\partial K_m(wr)}{\partial \theta} \right) \sin(m\theta + \alpha), \\ H_\theta^{(core)} &= \frac{-i}{u^2} \left(Am \frac{\beta}{r} J_m(ur) + B\omega\epsilon_0 n_1^2 \frac{\partial K_m(wr)}{\partial r} \right) \cos(m\theta + \alpha), \\ H_\theta^{(cladding)} &= \frac{-i}{w^2} \left(Cm\beta J_m(wr) + D \frac{\omega\epsilon_0 n_0^2}{r} \frac{\partial K_m(wr)}{\partial \theta} \right) \cos(m\theta + \alpha). \end{aligned}$$

Recall from subsection 6.1 that light rays are confined in the core. Total internal reflection only occurs if the refractive index of the core and cladding satisfy $n_1 > n_0$, respectively, and only depends on the choice of radius, $r = a$, where the core-cladding boundary is. The boundary conditions for each of the components of the electric and magnetic fields given by equations (6.7) are applied to the equations above involving Bessel functions. The system of equations,

$$\begin{pmatrix} \frac{m\beta}{u^2a^2}J_m(ua) & \frac{k\omega\mu_0}{ua}J'_m(ua) & \frac{m\beta}{w^2a}K_m(wa) & -\frac{k\omega\mu_0}{w^2a}K'_m(wa) \\ \frac{kn_1^2}{ua}J'_m(ua) & \frac{\omega\mu_0m}{u^2a^2}J_m(ua) & \frac{k^2n_0^2}{wa}K'_m(wa) & \frac{\omega\mu_0m\beta}{w^2a^2}K_m(wa) \\ J_m(ua) & 0 & -K_m(wa) & 0 \\ 0 & J_m(ua) & 0 & K_m(wa) \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

is obtained.

The characteristic equations are a result of finding the determinant of the system of equations and equating it to zero. These are the characteristics corresponding to the solution of Maxwell's wave equation described by the Bessel functions. Characteristics follow the path of light rays and are distributed according to the plane wave solution of m propagation modes. The larger the value of m , the more dispersed the modes become and more rays of an initial arbitrary beam are considered.

The characteristic equations also give the propagation constant β and amplitudes $E_z, E_r, E_\theta, H_z, H_r, H_\theta$ whenever $n_1, n_0, a =$ 'core radius' and normalized frequency

$$V = \sqrt{(ua)^2 + (wa)^2} = \left[a\sqrt{\mu_0\epsilon_0(n_1^2 - n_0^2)} \right] \omega$$

are given. Knowing the propagation constant and amplitudes allows us to analyze the dispersive solutions and group velocity behaviour according to the definitions and discussion in section 5.2.

The normalized group velocity can be found and we may check which modes have non-zero group velocity for a fixed value of m and some value of the nor-

malized frequency, V , which in turn depends on some value of the frequency, ω . The normalized group velocity is given as the ratio of the speed of light, c , with the actual group velocity, v_g , by rewriting β and k with respect to the normalized frequency, V , and applying the product rule;

$$\frac{c}{v_g} = \frac{d\beta}{dk} = \frac{a\sqrt{n_1^2 - n_0^2}}{a\sqrt{n_1^2 - n_0^2}} \frac{d(\frac{V\beta}{k})}{dV} = V \frac{d(\frac{\beta}{k})}{dV} + \frac{\beta}{k}.$$

If a normalized group velocity is zero for a mode m , then there is no dispersion occurring with respect to that mode.

The characteristic equations from the determinant are given as

$$\begin{aligned} \left(\frac{J'_m(ua)}{uaJ_m(ua)} + \frac{K'_m(wa)}{waK_m(wa)} \right) \left(\frac{J'_m(ua)}{\kappa J_m(ua)} + \frac{n_0^2 K'_m(wa)}{n_1^2 waK_m(wa)} \right) \\ = m^2 \left(\frac{1}{(ua)^2} + \frac{1}{(wa)^2} \right) \left(\frac{n_1^2}{n_0^2} \frac{1}{(ua)^2} + \frac{1}{(wa)^2} \right). \end{aligned} \quad (6.18)$$

In general this eigenvalue problem may have several solutions for each integer value of m . It is customary to express these solutions by β_{mn} where both m and n are integers. An eigenvalue, β_{mn} , corresponds to a mode which the fiber can support. Since we may solve the wave equation for either \vec{E} or \vec{H} then these are designated HE_{mn} and EH_{mn} , representing solving the wave equation with respect to \vec{H} and \vec{E} respectively.

Now we consider the most fundamental solutions to the characteristic equations. Let $m = 0$, then the characteristic equation, (6.18), becomes

$$\left(\frac{J'_m(ua)}{uaJ_m(ua)} + \frac{K'_m(wa)}{waK_m(wa)} \right) \left(\frac{J'_m(ua)}{\kappa J_m(ua)} + \frac{n_0^2 K'_m(wa)}{n_1^2 waK_m(wa)} \right) = 0.$$

Then, necessarily

$$\left(\frac{J'_m(ua)}{uaJ_m(ua)} + \frac{K'_m(wa)}{waK_m(wa)} \right) = 0 \quad (6.19)$$

and

$$\left(\frac{J'_m(ua)}{\kappa J_m(ua)} + \frac{n_0^2 K'_m(wa)}{n_1^2 waK_m(wa)} \right) = 0. \quad (6.20)$$

Equations (6.19) and (6.20) are the characteristic equations for the TE and TM modes, respectively. Both of these cannot be simultaneously zero unless describing the trivial solution. Notice that whenever (6.19) is satisfied, then $A = C = 0$ in the system of equations given by the boundary conditions. If $A = C = 0$ then $E_z = 0$ and we conclude that the electric field does not propagate in the z direction. A similar argument holds for the TM mode using equation 6.20.

For a given frequency, ω , the normalized group velocities corresponding to values of β are obtained from characteristic equations for integer values of m . When comparing these group velocities for different values of ω , it is known according to [3], [8] and [2] that the group velocity is zero for all other modes except the TE mode whenever $V_0 < 2.4$, where the number 2.4 is rounded down from its further approximated decimal expansion. Hence, the TE mode is the only mode that disperses for the corresponding frequency, ω_0 , determined by radius a and refractive indices n_1 and n_0 .

7 Conclusions

To conclude this report, a summary of what has been discussed is taken up first. Then, ways in which solutions can be further interpreted are discussed. Lastly we state how to begin analyzing non-linear effects in optical fibers.

The goal of this report was to describe light propagation in optical fibers. In order to do this, a description of optical fibers was discussed so that a simple optical fiber could be defined. Doing so also made it possible to interpret the boundary conditions that are involved in a given problem concerning light propagation.

After defining the fixed physical properties of an optical fiber, the general properties of light within glass were discussed. A parametrization of light in the

form of rays is given and already simplifies the problem. It was also observed that light takes the form of electromagnetic energy. The equations governing the dynamic properties of electric and magnetic fields were required to describe the light propagating in the core of the fiber.

Maxwell's equations were used to derive Maxwell's wave equation. This is the wave equation which governs any electromagnetic waves in an optical fiber and includes the governing of light propagation. Once the necessary equations and boundary conditions have been established, a method of finding solutions was required.

Hyperbolic partial differential equations were introduced and are necessary for defining dispersion in a linear medium. The method of characteristics was used to illustrate an example of a hyperbolic wave. This demonstrated the ray tracing technique that is used when finding solutions to Maxwell's wave equation for optical fibers. This lead to the definition of dispersion, which arose in the plane wave solutions to the initial boundary value problem for the 1-dimensional wave equation.

By defining dispersion, it became necessary to define group velocity. The group velocity is responsible for the broadening of light and is what causes dispersion to take place only if the phase velocity differs from the group velocity. By deriving the group velocity, we could interpret possible solutions more carefully.

The assumptions, definitions and derivations from sections 1 to 5 were gathered to come up with a solution. The assumption of monochromatic plane wave solutions was made. It also needed to be assumed that only short distances could be considered since the wave equations are non-linear in general. A linear approximation assumption that only describes effects over short distances and small duration in time was made. It was shown that it is possible to use Bessel functions and modified Bessel functions to compute characteristic equations. These

equations describe the distribution of propagation modes.

The modal distribution is further interpreted by analyzing the group velocity behaviour. This is done by comparing the normalized frequency V with the normalized propagation constant β/k . The graph of these two related quantities have zeros corresponding to the zeros of the Bessel function J_m and modified Bessel function K_m .

The frequencies where these zeros occur correspond to frequencies where modes are ‘cut off’. For a particular mode m_0 to propagate in the core, it must have a sufficiently large normalized frequency, $V_0 = \sqrt{(wa)^2 + (ua)^2}$, such that $V_0 > 2.4$. A larger frequency implies that more modes are being dispersed in the fiber since the corresponding group velocities are non-zero. Therefore, the dispersion increases with increasing frequency, since more modes exist and travel at differing group velocities.

The group velocity behaviour can then be analyzed when the refractive indices are independent of frequency. This depends on the linearization assumption that precluded the definition of light rays and linear dispersive solutions. Non-linear effects may also emerge for a sufficiently large amplitude and short durations of time over long distances. Therefore, sufficiently small amplitudes of light waves entering the core must also be assumed for the linearization assumption to be valid.

Recall that light had been parametrized linearly using the assumption of a local solution in section 2. This was equivalent to the assumption of a linearly approximated index of refraction in glass that does not depend on the frequency.

The refractive index of glass appears in Maxwell’s wave equation and, in general, depends on frequency when considering non-linear effects, but this report was only concerned with the linear approximation of Maxwell’s wave equation.

The non-linear theory of optical fibers is required to find solutions occur-

ring over long distances, long durations of time and for larger amplitudes. Non-linearity is a consequence of the properties embedded in the material, namely, glass. The assumption of linearity was made in this report by linearly approximating the electric susceptibility, $\chi(E)$, in the corresponding Taylor expansion representing $\chi(E)$.

In general, the non-linear effects grow over long distances or over long durations of time or for large amplitudes. Using the linear approximation of permittivity allowed us to linearize the problem and solutions obtained in this report are only valid for short distances and short durations of time and small amplitudes.

A Appendix: Vector Calculus Results

A.1 The Gradient

Definition A.1. The **gradient** is a vector valued differential operator which generalizes the derivative for multivariate functions. It is defined by,

$$\vec{\nabla} \equiv \vec{i} \frac{\partial}{\partial x} + \vec{j} \frac{\partial}{\partial y} + \vec{k} \frac{\partial}{\partial z} \quad (\text{A.1})$$

and has the property that

$$du = \vec{\nabla} u \circ d\vec{r} \quad (\text{A.2})$$

for an arbitrary function $u(\vec{r})$.

A.2 Divergence of a Field

Definition A.2. The **divergence** of a field \vec{A} is defined by the projection of the field onto the gradient for each infinitesimal point in space. This is equivalent to the dot product of the gradient with \vec{A} given by,

$$\text{div}(\vec{A}) = \vec{\nabla} \circ \vec{A} = \lim_{\Delta V \rightarrow 0} \frac{1}{\Delta V} \oint_S \vec{A} \circ \hat{n} da \quad (\text{A.3})$$

A.3 Divergence Theorem

Theorem A.1. *The flux of a vector field through a closed surface S is equal to the integral of the divergence of the field over a volume V for which S is a boundary. This may be written for some field, \vec{A} , as*

$$\oint_S \vec{A} \circ \hat{n} da = \int_V (\vec{\nabla} \circ \vec{A}) dV \quad (\text{A.4})$$

A.4 Curl and Circulation of a Vector Field

The curl and the circulation of a vector field is defined so that Stokes' theorem may be interpreted.

Definition A.3. *The **curl** of a vector field \vec{A} is a “measure of the vector field's tendency to circulate about a point”(p.75)[5]. It is defined,*

$$\text{curl}(\vec{A}) = \vec{\nabla} \times \vec{A} = \lim_{\Delta S \rightarrow 0} \frac{1}{\Delta S} \oint_C \vec{A} \circ d\vec{l} \quad (\text{A.5})$$

Definition A.4. *The **circulation** of a vector field \vec{A} is the magnitude of cumulative force of the vector field along the direction and length of a closed path C . Formally this is may be written as,*

$$\text{Circulation}(\vec{A}) = \oint_C \vec{A} \circ d\vec{l} \quad (\text{A.6})$$

A.5 Stoke's Theorem

Theorem A.2. *The circulation of a vector field over a closed path C is equal to the integral of the normal component of the curl of that field over a surface S for which C is boundary. This may be expressed by the equation,*

$$\oint_C \vec{A} \circ d\vec{l} = \int_S (\vec{\nabla} \times \vec{A}) \circ \hat{n} da \quad (\text{A.7})$$

Remark. The proofs for the Divergence and Stokes' theorems and the derivations for the gradient, divergence, and curl can be found in [10].

B Appendix: Differential Equations

B.1 Solutions to ODEs Around A Regular Singular Point

This section of the appendix follows [4] to find solutions to Bessel's equation. Bessel's equation is an ODE that is defined in subsection B.2. First, theorems and definitions for existence and uniqueness of solutions to ODEs around regular singular points are stated. The proof of the theorems can be found in [4].

The method of Frobenius is applied to Bessel's equation in subsection B.2. The method of Frobenius is a method of obtaining solutions by supposing a particular series solution exists. The supposed series is inputted into the ODE to obtain a recurrence relation for coefficients in the solution. We seek convergent series solutions by applying the method of Frobenius. The details of this method can be found in [4].

Theorem B.1. (*Existence of Series Solutions for Analytic Coefficients*)

Let $x_0 \in \mathbb{R}$ and suppose the coefficients a_1, \dots, a_n in

$$L(y) = y^{(n)} + a_1(x)y^{(n-1)} + \dots + a_n(x)y \quad (\text{B.1})$$

have convergent power series expansions in powers of $(x - x_0)$ on an interval $|x - x_0| < r_0$ for $r_0 > 0$. If $\gamma_1, \dots, \gamma_n$ are any n constants, there exists a solution ϕ of the problem

$$L(y) = 0, \quad y(x_0) = \gamma_1, \quad \dots, \quad y^{(n-1)}(x_0) = \gamma_n,$$

with a power series expansion

$$\phi(x) = \sum_{k=0}^{\infty} c_k (x - x_0)^k \quad (\text{B.2})$$

convergent for $|x - x_0| < r_0$. We have

$$c_k k! = \gamma_k + 1, \quad (k = 0, 1, \dots, n - 1),$$

and c_k for $k \geq n$ may be computed in terms of c_0, c_1, \dots, c_{n-1} by substituting (B.2) into $L(y) = 0$.

Theorem B.2. (*Uniqueness of Solutions*) Let $x_0 \in I \subset \mathbb{R}$, and let $\gamma_1, \dots, \gamma_n$ be any n constants. There is at most n solutions ϕ of $L(y) = 0$ on I satisfying

$$\phi(x_0) = \gamma_1, \phi'(x_0) = \gamma_2, \dots, \phi^{(n-1)}(x_0) = \gamma_n,$$

where $L(y)$ is the linear differential operator for some ODE applied to the solution.

Definition B.1. (*Regular Singular Point*) Let $x_0 \in \mathbb{R}$ and consider the linear differential equation with variable coefficients

$$a_0(x)y^{(n)} + a_1(x)y^{(n-1)} + \dots + a_n(x)y = 0. \quad (\text{B.3})$$

Then x_0 is a regular singular point for (B.3) if the equation can be written in the form

$$(x - x_0)^n y^{(n)} + b_1(x)(x - x_0)^{(n-1)} y^{(n-1)} + \dots + b_n(x)y = 0$$

near x_0 , where the functions b_1, \dots, b_n are analytic at x_0 .

Theorem B.3. (*Series Solutions Near a Regular Singular Point*) Consider the equation

$$x^2 y'' + a(x)xy' + b(x)y = 0$$

where a, b have convergent power expansions for $|x| < r_0$ and $r_0 > 0$. Let r_1, r_2 satisfy $\text{Re}(r_1) \geq \text{Re}(r_2)$ and be the roots of the indicial polynomial

$$q(r) = r(r - 1) + a(0)r + b(0).$$

For $0 < |x| < r_0$ there is a solution ϕ_1 of the form

$$\phi_1(x) = |x|^{r_1} \sum_{k=1}^{\infty} c_k x^k, \quad (c_0 = 1),$$

where the series converges for $|x| < r_0$. If $(r_1 - r_2) \neq 0$ or $(r_1 - r_2) = N \in \mathbb{Z}^+$, there is a second solution $\phi_2(x)$ for $0 < |x| < r_0$ of the form

$$\phi_2(x) = |x|_{r_2} \sum_{k=0}^{\infty} \tilde{c}_k x^k, \quad (\tilde{c}_0 = 1),$$

where the series converges for $|x| < r_0$. The coefficients c_k and \tilde{c}_k can be obtained by substitution of the solutions into the differential equation.

Remark. The **method of Frobenius** refers to the method of obtaining the indicial polynomial by substituting the series solution into the given differential equation. The Frobenius method is used to find values of r for which the series solution proposed in the theorem satisfies the differential equation being considered.

The second solution in the theorem above is known to exist for $r_1 \neq r_2$ by the existence and uniqueness theorems and the fact that ϕ_1, ϕ_2 are linearly independent for non-trivial solutions. There are two exceptional cases for determining the second solution in the theorem above. First, if $(r_1 - r_2) = 0$ then the solutions ϕ_1 and ϕ_2 are no longer linearly independent and another solution is constructed using the method in section 4.6 of [4].

The author of [4] also derives a second solution for a second exceptional case, that is, $(r_1 - r_2) = N \in \mathbb{Z}^+$. Defining the coefficients of the second series solution, in this case, may result in a denominator of zero for a coefficient \tilde{c}_k , for some integer $k \geq 1$. Hence, another linearly independent series solution is constructed such that coefficients cannot have zero in the denominator. The next theorem summarizes the results of these special cases.

Theorem B.4. (*Special Cases for Solutions Near a Regular Singular Point*) Consider the same assumptions as theorem B.3, except that $(r_1 - r_2)$ satisfies either $(r_1 - r_2) = 0$ or $(r_1 - r_2) = N \in \mathbb{Z}^+$. In the first case, there are

two linearly independent solutions ϕ_1, ϕ_2 for $0 < |x| < r_0$ of the form

$$\phi_1(x) = |x|^{r_1} \sigma_1(x), \quad \phi_2(x) = |x|^{r_1} \sigma_2(x) + (\log|x|) \phi_1(x),$$

where σ_1, σ_2 have power series expansions which are convergent for $|x| < r_0$, and $\sigma_1(0) \neq 0$. In the second case, there are two linearly independent solutions ϕ_1, ϕ_2 for $0 < |x| < r_0$ of the form

$$\phi_1(x) = |x|^{r_1} \sigma_1(x), \quad \phi_2(x) = |x|^{r_2} \sigma_3(x) + A \phi_1(x) \log|x|,$$

where $\sigma_1, \sigma_2, \sigma_3$ have power series expansions which are convergent for $|x| < r_0$, $\sigma_1(0) \neq 0$, $\sigma_2(0) \neq 0$, and c is a constant that may be zero.

B.2 Bessel's Equation

Bessel's equation of order α is a second order linear ordinary differential equation. Let $\alpha \in \mathbb{C}$ be a constant such that $\operatorname{Re}(\alpha) \geq 0$. Bessel's equation of order α is given as

$$\frac{d^2 y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + \left(1 - \alpha^2\right) y = 0 \tag{B.4}$$

The theorems from subsection B.1 are now applied to Bessel's equation of order α to find possible solutions using the method of Frobenius, which is defined in [4]. The method of Frobenius is a general method used to prove theorem B.3 and is not discussed here in full detail. It will, however, be referred to for some justification in applying theorems B.3 and B.4 as needed.

Recall Bessel's equation, (B.4), and apply the definition of a regular singular point by writing the equation as

$$x^2 y'' + x b_1(x) y' + b_2(x) y = 0,$$

where $b_1(x) = 1$ and $b_2(x) = (x^2 - \alpha^2)$ are analytic at $x = 0$. Theorem B.3 is applied to find the first solution for integer values of α . This gives the indicial

equation satisfying

$$q(r) = r(r-1) + b_1(0)r + b_2(0) = r^2 - \alpha^2 = (r_1 - \alpha)(r_2 + \alpha) = 0.$$

It follows that $r = \pm\alpha$. The indicial equation is obtained, in general, by inputting a solution of the form

$$\sum_{k=0}^{\infty} c_k x^{k+r}$$

into the differential equation. Doing so follows from the method of Frobenius.

Inputting this solution into Bessel's equation gives,

$$\begin{aligned} 0 &= x^2 \sum_{k=0}^{\infty} c_k (k+r)(k+r-1) x^{k+r-2} + x \sum_{k=0}^{\infty} c_k (k+r) x^{k+r-1} + (x^2 - \alpha^2) \sum_{k=0}^{\infty} c_k x^{k+r} \\ &= \sum_{k=0}^{\infty} c_k \left((k+r)(k+r-1) + (k+r) - \alpha^2 \right) x^{k+r} + \sum_{k=0}^{\infty} c_{k-2} x^{k+r} \\ &= q(r) c_0 x^r + q(r+1) c_1 x^{r+1} + \sum_{k=2}^{\infty} \left(q(r+k) c_k + c_{k-2} \right) x^{k+r}. \end{aligned} \quad (\text{B.5})$$

When $k = k_0$ we recover the terms proportional to x^{r+k_0} . It follows by the method of Frobenius that $k = 0$ gives

$$0 = q(r) c_0 x^r.$$

The method also implies $c_0 \neq 0$, since otherwise $c_k = 0$, $\forall k \geq 0$, and gives a trivial solution in defining a recurrence relation. Also, $x^r \neq 0$, since $x > 0$. Therefore, the indicial polynomial, $q(r)$, is necessarily zero to obtain non-trivial solutions. It is also necessary that $c_1 = 0$ by linear independence with respect to powers of x , and since $x^{r+1} \neq 0$ and inputting $r = \pm\alpha$ gives

$$q(r+1) = (1 \pm \alpha + \alpha)(1 \pm \alpha - \alpha) = \begin{cases} 1 + 2\alpha & \text{for } r = \alpha \\ 1 - 2\alpha & \text{for } r = -\alpha \end{cases} \neq 0, \quad \forall \alpha \in \mathbb{R}.$$

It follows from (B.5) that coefficients of the infinite series are necessarily zero.

Then, a recurrence relation may be defined as

$$C_k(r) = \frac{-c_{k-2}}{(k+r-\alpha)(k+r+\alpha)}. \quad (\text{B.6})$$

The recurrence relation gives coefficients of every other integer value subscript. Now, let $r = r_1 = \alpha$ to obtain the even and odd coefficients, computed from c_0 and c_1 , respectively, of the series solution with $r = \alpha$. It follows that

$$C_k(0) = c_k = \frac{-c_{k-2}}{k(2\alpha + k)},$$

by inputting $r = \alpha$. This relation is used to compute even coefficients c_{2m} for integers $m > 0$. All odd coefficients are zero because $c_1 = 0$. By computing the first few values of the recurrence relation,

$$c_{2m} = \frac{(-1)^m c_0}{m! 2^{2m} (\alpha + m) \cdot \dots \cdot (\alpha + 1)}, \quad (\text{B.7})$$

is obtained. Consider $\alpha \in \{0, 1, 2, 3, \dots\}$ and recall that c_0 is a constant that can be chosen arbitrarily. If $\alpha = 0$ we can choose $c_0 = 1$ so that (B.7) can be written as

$$c_{2m} = \frac{(-1)^m c_0}{2^{2m} (m!)^2}.$$

Therefore, by theorem B.3, the solution to Bessel's equation of order $\alpha = 0$ is

$$J_0(x) = \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{2^{2m} (m!)^2} \quad (\text{B.8})$$

For integers $\alpha = n \geq 1$, we may choose $C_0(n) = 1/(2^n n!)$ to obtain

$$C_{2m}(n) = \frac{(-1)^m}{2^{2m+n} m! (m+n)!}. \quad (\text{B.9})$$

These coefficients give solutions to Bessel's equation of order $\alpha = n$, for integers $n > 0$. The choice of $C_0(n)$ guarantees the series converges. Such solutions are given by

$$J_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m+n}}{2^{2m+n} m! (m+n)!}. \quad (\text{B.10})$$

Next, we introduce the gamma function and its properties. We use it to define a particular $C_0(\alpha)$. The motivation of defining $C_0(\alpha)$ using the gamma function, is to find a series solution to Bessel's equation of order α for any non-negative value of α such that the solution converges.

Definition B.2. (*Gamma Function*) Let $z \in \mathbb{C}$ satisfy $\operatorname{Re}(z) \geq 0$. The gamma function is defined by the integral

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

The gamma function satisfies the following properties:

1. $\Gamma(z+1) = z\Gamma(z)$.
2. $\Gamma(1) = 1$.
3. $\Gamma(n) = n\Gamma(n-1) = \dots = (n-1)!\Gamma(1) = (n-1)!$.

Observe that choosing

$$C_0(\alpha) = \frac{1}{2^\alpha \Gamma(\alpha+1)}$$

implies that the recurrence can be written

$$C_{2m}(\alpha) = \frac{(-1)^m \alpha!}{2^{2m+\alpha} m! (m+\alpha)! \Gamma(\alpha+1)} = \frac{(-1)^m}{2^{2m+\alpha} m! \Gamma(m+\alpha+1)}, \quad (\text{B.11})$$

which gives a convergent series solution to Bessel's equation of arbitrary order α , called a Bessel function, given by

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m+\alpha}}{2^{2m+\alpha} m! \Gamma(m+\alpha+1)}. \quad (\text{B.12})$$

Now that we have defined the first solution to Bessel's equation of order α , we must find a second linearly independent solution. Whenever $(r_1 - r_2) \neq 0$ and $(r_1 - r_2) \neq N \in \mathbb{Z}^+$ we may apply theorem B.3 directly and obtain the second linearly independent solution, $J_{-\alpha}(x)$. This may be done by following the proof of convergence for the series solution found on page 160-162 of [?], but the details will not be pursued here. $J_\alpha(x)$ and $J_{-\alpha}(x)$ are two necessarily different solutions. Hence, whenever $r_1 - r_2 \neq 0$, then $J_\alpha(x)$ and $J_{-\alpha}(x)$ are the two

linearly independent solutions corresponding to the series with coefficients $C_k(r_1)$ and $C_k(r_2)$, respectively.

Consider either of the special cases $(r_1 - r_2) = 0$ or $(r_1 - r_2) = N \in \mathbb{Z}^+$. Theorem B.4 is applied to each case, following the method of Frobenius, to get the second solution as required. Recall theorem B.4. The arguments involved in finding a second solution for these special cases, that are indeed linearly independent from the first, is not discussed here. Those details are given in [4]. Define $(r - r_2)C_k(r) = \tilde{C}_k(r)$. In summarizing the arguments mentioned above, we obtain

$$\sigma_1(x) = \sum_{k=0}^{\infty} C_k(r_1)x^k, \quad \sigma_2(x) = \sum_{k=0}^{\infty} x^k \left[\frac{d}{dr} C_k(r) \right] \Big|_{r=r_1}, \quad (\text{B.13})$$

$$\sigma_3(x) = \sum_{k=0}^{\infty} x^k \left[\frac{d}{dr} (r - r_2) C_k(r) \right] \Big|_{r=r_2}, \quad A = (r - r_2) C_N(r) \Big|_{r=r_2}. \quad (\text{B.14})$$

Now, we can find another pair of solutions to Bessel's equation of order α using these equations; by solving for the coefficients

$$C_k(r_1), \quad \frac{d}{dr} C_k(r_1), \quad \frac{d}{dr} \tilde{C}_k(r_2) = \left[\frac{d}{dr} (r - r_2) C_k(r) \right] \Big|_{r=r_2}, \quad A = \tilde{C}_N(r_2),$$

according to theorem B.4.

In order to take these derivatives, we must obtain an equation for $C_k(r)$, given by (B.6), with respect to only r and k . We can find $C_k(r)$ as required by computing the first few terms. It follows that

$$\begin{aligned} C_0(r) &= C_0, \quad (\text{for some constant } C_0), \\ C_2(r) &= \frac{-C_0}{(r+2+\alpha)(r+2-\alpha)}, \\ C_4(r) &= \frac{-C_2(r)}{(r+4+\alpha)(r+4-\alpha)} = \frac{C_0}{(r+2+\alpha)(r+2-\alpha)(r+4+\alpha)(r+4-\alpha)}, \\ C_6(r) &= \frac{-C_4(r)}{(r+6+\alpha)(r+6-\alpha)} = \frac{-C_0}{\prod_{k=1}^3 (r+2k+\alpha)(r+2k-\alpha)}, \dots \end{aligned}$$

which gives

$$C_{2m}(r) = \frac{C_0(-1)^m}{\prod_{k=1}^m (r+2k+\alpha)(r+2k-\alpha)}. \quad (\text{B.15})$$

In order to apply the product rule in differentiating $\tilde{C}_{2m}(r)$, we require

$$C_{2m}(r)' = [C_0(-1)^m d_{2m}(r)]',$$

where

$$d_{2m}(r) = \frac{1}{\prod_{k=1}^m (r+2k+\alpha)(r+2k-\alpha)}.$$

Observe that

$$\frac{d}{dr} \log d_{2m}(r) = \frac{\frac{d}{dr} d_{2m}(r)}{d_{2m}(r)},$$

by applying the chain rule. It follows that

$$\frac{d}{dr} \log d_{2m}(r) = \left(- \sum_{k=1}^m \frac{d}{dr} (r+2k+\alpha)(r+2k-\alpha) \right) = -2 \sum_{k=1}^m (r+2k).$$

Therefore,

$$C'_{2m}(r) = \left[\frac{-2C_0(-1)^m \sum_{k=1}^m (r+2k)}{\prod_{k=1}^m (r+2k+\alpha)(r+2k-\alpha)} \right]. \quad (\text{B.16})$$

Moreover, we may now differentiate $\tilde{C}_{2m}(r)$ using this equation and the product rule. It follows that

$$\frac{d}{dr} \tilde{C}_{2m}(r) = (r-r_2)C'_{2m}(r) + C_{2m}(r). \quad (\text{B.17})$$

The dependence on α is removed in (B.16) above for the special case of $(r_1 - r_2) = 0$ from theorem B.4, since we have a condition $(r_1 - r_2) = 2\alpha$. For the case where $(r_1 - r_2) = 0 = 2\alpha$, we obtain $\alpha = 0$. Then according to theorem B.4 and equations (B.15), (B.16) and (B.8) we obtain the first solution, $J_0(x)$, and the second linearly independent solution,

$$B(x) = J_0(x) \log |x| + \sum_{m=0}^{\infty} \left(\frac{-2C_0(-1)^m \sum_{k=1}^m (r+2k)}{\prod_{k=1}^m (r+2k)^2} \right) x^{2m}.$$

Then, according to theorem B.2, the solution to Bessel's equation of order zero is

$$c_2 J_0(x) + c_3 B(x) \quad (\text{B.18})$$

for constants c_2, c_3 satisfying some initial conditions.

Similar to the first case, we obtain a second linearly independent solution for the case $(r_1 - r_2) = 2\alpha = N \in \mathbb{Z}^+$ using theorem B.4, except it is for general α . Recall that $r_2 = -\alpha$. Then by equation (B.15) and the equation $2m = N = 2\alpha$, we obtain

$$A = \left[(r + \alpha) \frac{C_0(-1)^\alpha}{\prod_{k=1}^\alpha (r + 2k + \alpha)(r + 2k - \alpha)} \right] \Big|_{r=-\alpha} = \frac{C_0(-1)^{-\alpha}}{(2\alpha) \prod_{k=1}^{\alpha-1} 4k(k - \alpha)}. \quad (\text{B.19})$$

According to theorem B.4 and equations (B.12), (B.16), (B.17) and (B.19) we obtain the first solution, $J_\alpha(x)$, and the second linearly independent solution,

$$|x|^\alpha \sum_{m=0}^{\infty} \left((r + \alpha) C_{2m}(r)' + C_{2m}(r) \right) x^{2m} + \left[\frac{C_0(-1)^{-\alpha}}{(2\alpha) \prod_{k=1}^{\alpha-1} 4k(k - \alpha)} \right] J_\alpha(x) \log |x|,$$

denoted by $Y_\alpha(x)$. Note that $Y_\alpha(x)$ is called the Bessel function of the second kind.

Then, according to theorem B.2, Bessel's equation of order α has the solution

$$c_4 J_\alpha(x) + c_5 Y_\alpha(x), \quad (\text{B.20})$$

whenever $\alpha = N/2 \in \mathbb{Z}^+$.

B.3 The Modified Bessel's Equation

For applying Bessel's equation in section 6, we must consider complex values of x which have no real part in Bessel's equation, (B.4). In this case, the Bessel functions *may* not be solutions to Bessel's equation because the way that Bessel's equation is written becomes different. This is relevant for finding solutions according to the theorems in appendix B.1. It is possible to modify Bessel's equation to be suitable for $ix \in \mathbb{C}$ satisfying $\text{Re}(ix) = 0$. This subappendix deals with

finding real solutions with respect to $x \in \mathbb{R}$ that are also solutions with respect to the variable $ix \in \mathbb{C}$ for a modified version of Bessel's equation.

The transformation $x \mapsto ix$ is applied to Bessel's equation and gives an equation called the modified Bessel's equation. The same method of finding solutions as for Bessel's equation is used to find real series solutions centered at $x = 0$, $x \in \mathbb{R}$ corresponding to the imaginary variable ix .

The modified Bessel's equation is given by

$$(ix)^2 y'' + ixy' + ((ix)^2 + \alpha^2)y = 0.$$

Applying the method of Frobenius in the same way as for Bessel's equation gives the required condition for the indicial equation by the relation

$$0 = q(r)c_0(ix)^r = (r - \alpha)(r + \alpha)c_0 i^r x^r. \quad (\text{B.21})$$

Following the same argument as for Bessel's equation, we have $i^r \neq 0$ for any value of r . Hence, the indicial equations for both regular and modified Bessel's equation are the same and given by $q(r)$ above. The series solution,

$$\sum_{k=0}^{\infty} C_k(r)(ix)^{r+k}, \quad (\text{B.22})$$

can be written with respect to powers of x by factoring out i^r in (B.21). Doing so means that the series must be written as

$$\sum_{k=0}^{\infty} C_k(r)x^r(ix)^k. \quad (\text{B.23})$$

Whenever coefficients are real, the series solution is real only for even values of k . This also introduces a factor of -1 for each coefficient. It is now shown by substituting the series solution, (B.22), into the modified Bessel's equation that, indeed, only even values of k occur in the series solution and the coefficients are

real. Substituting the series solution gives

$$\begin{aligned}
0 &= (ix)^2 \sum_{k=0}^{\infty} C_k(r)(k+r)(k+r-1)(ix)^{k+r-2} + ix \sum_{k=0}^{\infty} C_k(r)(k+r)(ix)^{k+r-1} \\
&\quad + ((ix)^2 - \alpha^2) \sum_{k=0}^{\infty} C_k(r)(ix)^{k+r} \\
&= \sum_{k=0}^{\infty} C_k(r) \left((k+r)(k+r-1) + (k+r) - \alpha^2 \right) (ix)^{k+r} - \sum_{k=0}^{\infty} C_{k-2}(r)(ix)^{k+r} \\
&= q(r)C_0(ix)^r + q(r+1)C_1(r)(ix)^{r+1} + \sum_{k=2}^{\infty} \left(q(r+k)C_k(r) - C_{k-2}(r) \right) (ix)^{k+r}.
\end{aligned} \tag{B.24}$$

By the same arguments as for the Bessel's equation, we have that $c_1 = 0$. Hence, the recurrence is given only for even values of k , since the coefficients of the series satisfy

$$q(r+k)C_k(r) - C_{k-2}(r) = 0.$$

The factor of -1 must also be introduced in the term,

$$\sum_{k=2}^{\infty} \left(q(r+k)C_k(r) - C_{k-2}(r) \right) (ix)^{k+r},$$

above. In doing so, it may be rewritten as

$$\sum_{k=2}^{\infty} \left(C_{k-2}(r) - q(r+k)C_k(r) \right) (ix)^r x^k,$$

and satisfies

$$0 = \sum_{k=2}^{\infty} \left(C_{k-2}(r) - q(r+k)C_k(r) \right) x^{k+r},$$

since $q(r) = 0$ and $c_1 = 0$ in equation (B.24).

This gives the recurrence relation

$$C_k(r) = \frac{C_{k-2}(r)}{(r+k+\alpha)(r+k-\alpha)} \tag{B.25}$$

This recurrence relation is like the one from equation (B.6) for Bessel's equation, except with a factor of -1 . Therefore, the procedure for finding first and second linearly independent solutions corresponding to the cases in theorems B.3 and B.4 is practically the same for the modified Bessel equation as it is for Bessel's equation. The coefficients of these series solutions for the modified Bessel's equation,

$$C_k^{mod}(r_1), \quad \frac{d}{dr}C_k^{mod}(r_1), \quad \frac{d}{dr}\tilde{C}_k^{mod}(r_2) = \left[\frac{d}{dr}(r-r_2)C_k^{mod}(r) \right] \Big|_{r=r_2}, \quad A = \tilde{C}_N^{mod}(r_2),$$

can be found by 'modifying' the coefficients of the solutions to Bessel's equation according to the transformation

$$\frac{d}{dr}C_{2m}^{mod}(r) = \frac{1}{(-1)^m} \frac{d}{dr}C_{2m}(r), \quad (\text{B.26})$$

where $C_{2m}(r)$ is defined by (B.15).

First, the two linearly independent solutions in theorem B.3 are found for the modified Bessel equation, (i.e., for the case $r_1 - r_2 \neq 0$). Applying the transformation (B.26) to each coefficient for Bessel's functions obtains the coefficients for the modified Bessel functions. The result of this is given by

$$C_{2m}^{mod}(r) = \frac{C_0(r)}{2^{2m}(m!)^2}.$$

In order to ensure the coefficients provide a convergent series, let

$$C_0^{mod}(n) = \frac{1}{2^n n!} \quad \text{and} \quad C_0^{mod}(\alpha) = \frac{1}{2^\alpha \Gamma(1 + \alpha)},$$

since these may be chosen arbitrarily. Then the coefficients corresponding to the series solutions of the modified Bessel's equation, according to the assumptions of theorem B.3, are given by

$$C_{2m}^{mod}(n) = \frac{1}{2^{2m+n} m! (m+n)!} \quad (\text{B.27})$$

and

$$C_{2m}^{mod}(\alpha) = \frac{\alpha!}{2^{2m+\alpha} m! (m+\alpha)! \Gamma(1+\alpha)}. \quad (\text{B.28})$$

Therefore, the modified Bessel's function of the first kind can be defined for either integer values of $\alpha = n$ or arbitrary values of α , and is given by

$$I_n(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{2m+n}}{m! (m+n)!} \quad (\text{B.29})$$

or

$$I_{\alpha}(x) = \frac{(x/2)^{2m+\alpha}}{m! \Gamma(m+\alpha+1)}, \quad (\text{B.30})$$

respectively. The two solutions $I_{\alpha}(x)$ and $I_{-\alpha}(x)$ form a basis for any other solution of the modified Bessel's equation for the case that $(r_1 - r_2) \neq 0$. It remains to find the second linearly independent solutions for the two special cases in theorem B.4.

When $r_1 - r_2 = 0$, note that $(r_1 - r_2)/2 = \alpha = 0$ gives the first solution $I_0(x)$. For the second solution we require

$$\frac{d}{dr} C_{2m}^{mod}(r).$$

It is obtained by applying (B.16) to the transformation (B.26). Doing so obtains

$$\frac{d}{dr} C_{2m}^{mod}(-\alpha) = \frac{-2m(m+1)}{\prod_{k=1}^m (2k)^2} = \frac{-1}{2^{2m-1}} \left[\frac{1}{(m-1)!} + \frac{1}{m!(m-1)!} \right]. \quad (\text{B.31})$$

Therefore, the second series solution is given by

$$M(x) = I_0(x) \log |x| + \sum_{m=0}^{\infty} \frac{-2m(m+1)}{\prod_{k=1}^m (2k)^2} x^{2m} \quad (\text{B.32})$$

Now, consider the case $r_1 - r_2 = N \in \mathbb{Z}^+$. The coefficients

$$\left[\frac{d}{dr} (r - r_2) C_k^{mod}(r) \right] \Big|_{r=r_2}, \quad A = \tilde{C}_N^{mod}(r_2),$$

are given by the transformation, (B.26), and equations (B.17) and (B.19). By assumption, $N = 2\alpha = 2m_0$ for some integer $m_0 > 0$. It follows that the second

solution, called the MBF of the second kind, is given by

$$K_\alpha(x) = \frac{C_0 I_\alpha(x) \log |x|}{2\alpha 4^{\alpha-1} (\alpha-1)! \prod_{k=1}^{\alpha-1} (k-\alpha)} + |x|^{-\alpha} \sum_{m=0}^{\infty} \frac{C_0 \left(1 - 4\alpha \left(\frac{2k(2k+1)}{2} - m\alpha\right)\right)}{4^m m! \prod_{k=1}^m (k-\alpha)}. \quad (\text{B.33})$$

Note that this does converge by inspecting the coefficients in the series. Therefore, the full solution to the modified Bessel's equation for the case $r_1 - r_2 = N \in \mathbb{Z}^+$ is given by

$$\phi_2(x) = c_6 K_\alpha(x) + c_7 I_\alpha(x),$$

for some constants c_6, c_7 .

B.4 Defining PDEs and Initial Boundary Value Problems

In this section, definitions, theorems and other statements involving differential equations are provided. These statements mainly come from course notes and the preliminary chapters in [4].

Definition B.3. (*Partial Derivative*) If $u : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function of n variables x_1, \dots, x_n , then the partial derivative of u with respect to some variable x_i , $i = 1, \dots, n$ at some point $\vec{x}_0 = (x_{(1,0)}, x_{(2,0)}, \dots, x_{(n,0)}) \in \mathbb{R}^2$ is defined as

$$u_{x_i}(\vec{x}_0) = \frac{\partial u}{\partial x_i} = \lim_{h \rightarrow 0} \frac{u(x_{(1,0)}, \dots, x_{(i,0)} + h, \dots, x_{(n,0)}) - u(x_{(1,0)}, \dots, x_{(n,0)})}{h}.$$

For higher derivatives this may be generalized. For example,

$$u_{x_i x_j}(\vec{x}_0) = \frac{\partial^2 u}{\partial x_i \partial x_j} = \lim_{h \rightarrow 0} \frac{u(x_{(1,0)}, \dots, x_{(i,0)} + h, \dots, x_{(j,0)} + h, \dots, x_{(n,0)}) - u(x_{(1,0)}, \dots, x_{(n,0)})}{h}.$$

Definition B.4. (*Partial Differential Equation*) A partial differential equation (PDE) for a function u of the independent variables x_1, \dots, x_n is an equation of the form

$$F(x_1, \dots, x_n, \dots, u, u_{x_1}, \dots, u_{x_n}, u_{x_1 x_2}, \dots, u_{x_1 x_n}, \dots),$$

where F is a given function of the independent variables x_1, \dots, x_n of the unknown function u and of its partial derivations. The order of a PDE is the order of the highest partial derivative that appears in F .

Definition B.5. (System of PDEs) A system of PDEs is defined by one or more unknown functions each corresponding to its own PDE. The order of a system of PDEs is the order of the highest derivative that appears in any of the individual PDEs.

We restrict statements about PDEs from this point on to those that are no more than second order with four independent variables x_1, x_2, x_3, x_4 . This makes definitions and theorems directly applicable to the differential equations involved in this report. The following statements may be generalized for higher orders and for more independent variables.

Whenever mixed partial derivatives are equal there is usually simplifications which occur in analyzing PDEs. We state the following theorem to get a general condition that implies the equality of mixed partial derivatives.

Theorem B.5. (Clairaut's Theorem) If u is defined on the closed ball B that contains the point $\vec{x}_0 = (x_{(1,0)}, x_{(2,0)}, x_{(3,0)}, x_{(4,0)})$ and mixed partial derivatives $u_{x_i x_j}, u_{x_j x_i}$ are continuous on B , then

$$u_{x_i x_j}(\vec{x}_0) = u_{x_j x_i}(\vec{x}_0).$$

Definition B.6. (Solution to PDE) A function $u(x_1, x_2, x_3, x_4)$ is a solution to the PDE,

$$F(x_1, \dots, x_4, u_{x_1}, \dots, u_{x_4}, u_{x_1 x_2}, \dots) = 0, \quad (\text{B.34})$$

whenever substituting u and its partial derivatives into (B.34), it is satisfied identically in some region $\Omega \subset \mathbb{R}^4$

Remark. If the independent variables exist in \mathbb{R}^4 for some PDE, then all of the partial derivatives are continuous. This assumption implies that mixed partial derivatives are equal by Clairaut's theorem above.

Definition B.7. (*Non-linear, Linear, Semi-Linear and Quasi-Linear PDEs*) Consider a PDE with unknown function $u(x_1, x_2, x_3, x_4)$. The PDE is linear in its definition without any additional assumptions on dependent or independent variables in F . The PDE is linear if it is linear in the unknown function u and its derivatives. The PDE is semi-linear if it is linear with respect to the highest derivatives, but not necessarily with respect to lower order terms. The PDE is quasi-linear if it is linear only with respect to its highest derivatives.

Remark. Let the sets N, L, S, Q be the set of non-linear, linear, semi-linear, and quasi-linear PDEs, respectively. Then, $N \subset L \subset S \subset Q$.

Definition B.8. (*Linear Differential Operator*) A linear differential operator is a transformation of one function to another. Let $\Omega \subset \mathbb{R}^4$ and $\vec{x} \in \Omega$. A linear differential operator is a linear operator which maps $u(\vec{x}) \mapsto v(\vec{x})$, where the process mapping u to v is partial differentiation. We represent PDEs with linear differential operators. For example,

$$u(\vec{x}) \xrightarrow{\mathcal{L}} \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = v(\vec{x}).$$

In general, we may represent PDEs with linear differential operators. Note that 'linear', here, only refers to linearity of the operator. An operator, \mathcal{O} , is linear if for some functions, u, v , and constant γ , it satisfies

$$\mathcal{O}[\gamma u + v] = \gamma \mathcal{O}[u] + \mathcal{O}[v].$$

Consider the domain $\Omega \subset \mathbb{R}^4$ with points, (x, y, z, t) , representing independent variables of a PDE. Let u be the unknown function for the PDE given by

$$\mathcal{L}[u] = H(x, y, z, t), \tag{B.35}$$

for some function $H(x, y, z, t)$. This is called a non-homogeneous PDE and it can be shown, in general, that any non-homogeneous PDE can be transformed into a homogeneous PDE with a possibly different unknown function \bar{u} and transformed function $f_i(x, y, z, t) \mapsto g_i(x, y, z, t)$ for some functions $g_i(x, y, z, t)$.

Definition B.9. (*Initial Boundary Value Problem*) Let the operator (B.35) define a PDE. The PDE defines an initial boundary value problem (IBVP). An IBVP for the PDE is given by the system of PDEs,

$$\mathcal{L}[u] = 0, \tag{B.36}$$

$$\mathcal{L}_1[u] = f_1(x, y, z, t),$$

$$\dots,$$

$$\mathcal{L}_k[u] = f_k(x, y, z, t),$$

where \mathcal{L}_i are other linear differential operators, also having unknown function u .

Definition B.10. (*Well-Posed IBVP*) An IBVP is well-posed if it satisfies the following:

1. The solution, u , depends continuously on the initial condition and boundary conditions.
2. The solution, u , is unique.
3. The solution, u , to the PDE exists in Ω .

The uniqueness and continuous dependence of solutions both follow as a direct consequence of the linearity of the differential operator. The method of characteristics is required to discuss existence. This is done in appendix B.5. Note that uniqueness is affected by the method of characteristics.

B.5 Method of Characteristics

This section mainly follows [6] and notes from courses in both partial and ordinary differential equations. The method of characteristics is a method of transformation of the n independent variables for some Quasi-Linear PDE. In particular, the method of characteristics is used to find unique solutions to a given Cauchy problem.

Definition B.11. (*Cauchy Problem*) *A Cauchy problem is generally an n^{th} order PDE on some domain $\Omega \in \mathbb{R}^n$ with specified conditions, called Cauchy data, given on some curve $\Gamma \subset \Omega$.*

The method of characteristics allows for the analysis of solutions on $(n - 1)$ -dimensional hyper-planes specified by initial conditions, such that the solution to the Cauchy problem given by the PDE and $(n - 1)$ -dimensional hyper-plane has a unique solution. In section 4.2, the 3-dimensional wave equation is defined as a second order semi-linear PDE with 4 independent variables, (including time). Therefore, we restrict our attention to second order equations with the independent variables $(x, y, z, t) \in \Omega$ for defining the method of characteristics.

The procedure for applying the transformation which parametrizes u on the curve, Γ , is called the method of characteristics. It can be summarized by the following:

1. Parametrizing the curve, $\Gamma : (x_0(\tau), y_0(\tau), z_0(\tau), t_0(\tau), u_0(\tau))$, that defines the Cauchy data from initial conditions.
2. The hyper-surface solution is given by $(x, y, z, t) \in \Omega$ and is determined by hyper-planes $\zeta = u(x, y, z, t)$, for real values of ζ . The normal vector to the solution surface is

$$\left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}, \frac{\partial u}{\partial t}, -1 \right).$$

Define $F(x, y, z, t, \zeta) = u(x, y, z, t) - \zeta = 0$ in order to obtain

$$\vec{\nabla}F = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}, \frac{\partial F}{\partial t}, \frac{\partial F}{\partial \zeta} \right) = \vec{0} = \vec{\nabla}u(x, y, z, t).$$

The equation for the tangent plane at some point given by a position vector, $\vec{x}_1 = (x_1, y_1, z_1, t_1, u_1)$, is

$$(\vec{x} - \vec{x}_1) \vec{\nabla}F(\vec{x}_1) = 0.$$

In other words, we can define the normal vector at any point by $\vec{\nabla}F(\vec{x}_1)$ and it is equal to the normal vector shared by the solution surface at any point.

3. The PDE can be written in vector form as a dot product of cartesian vectors,

$$(a, b, c, d, g) \circ (u_x, u_y, u_z, u_t, -1) = 0. \quad (\text{B.37})$$

The hyper-plane defined by this dot product is contained in the hyper-surface solution parametrized by τ .

4. Consider an arbitrary parametric hyper-plane $(X(s), Y(s), Z(s), T(s), U(s))$, where $x = X(s), y = Y(s), z = Z(s), t = T(s), u = U(s)$. We may construct a family of hyper-planes given parametrically by

$$(X(s, \tau), Y(s, \tau), Z(s, \tau), t(s, \tau), U(s, \tau)), \quad (\text{B.38})$$

and find conditions to ensure this hyper-plane lies in the hyper-surface solution parametrized by s , for all s , and passes through Γ at some point τ .

5. The tangent to the hyper-surface solution, $(X(s), Y(s), Z(s), T(s), U(s))$, is the hyper-plane

$$\left(\frac{dX}{ds}, \frac{dY}{ds}, \frac{dZ}{ds}, \frac{dT}{ds}, \frac{dU}{ds} \right).$$

This must be normal to the hyper-surface solution for any s . It must also be normal to the hyper-planes for the family of characteristic hyper-planes,

each determined by a fixed value of τ , and implies that it is also normal to the hyper-surface solution. Hence, it satisfies

$$\left(\frac{dX}{ds}, \frac{dY}{ds}, \frac{dZ}{ds}, \frac{dT}{ds}, \frac{dU}{ds}\right) \circ (u_x, u_y, u_z, u_t, -1) = 0. \quad (\text{B.39})$$

Definition B.12. (*Characteristic Equations*) From the listed summary above and equations (B.37) and (B.39), the vectors

$$\left(\frac{dX}{ds}, \frac{dY}{ds}, \frac{dZ}{ds}, \frac{dT}{ds}, \frac{dU}{ds}\right) \quad \text{and} \quad (a, b, c, d, g)$$

are mutually orthogonal. Then, the following system of ODEs must be solved for a characteristic hyper-plane to be defined on the hyper-surface solution at any point $(x, y, z, t) \in \Omega$:

$$X_s = a(X(s, \tau), Y(s, \tau), Z(s, \tau), T(s, \tau), U(s, \tau)), \quad (\text{B.40})$$

$$Y_s = b(X(s, \tau), Y(s, \tau), Z(s, \tau), T(s, \tau), U(s, \tau)), \quad (\text{B.41})$$

$$Z_s = c(X(s, \tau), Y(s, \tau), Z(s, \tau), T(s, \tau), U(s, \tau)), \quad (\text{B.42})$$

$$T_s = d(X(s, \tau), Y(s, \tau), Z(s, \tau), T(s, \tau), U(s, \tau)), \quad (\text{B.43})$$

$$U_s = g(X(s, \tau), Y(s, \tau), Z(s, \tau), T(s, \tau), U(s, \tau)). \quad (\text{B.44})$$

These are called the characteristic equations.

Remark. The variable τ is fixed when defining the characteristic equation, but it represents any point on the hyper-surface solution. So, s is fixed with respect to the solution and is arbitrary. Therefore, the Cauchy data is applied for any fixed s with respect to the parametrized hyper-plane at some specified point τ corresponding to the solution. It is assumed here that the independent variables corresponding to the components of 3-dimensional space can each be parametrized by time, t , which is generally the case. Otherwise, this procedure enumerated above would be repeated for each fixed transformed independent variable and would produce a larger family of characteristic equations.

The point $(x_0(\tau), y_0(\tau), z_0(\tau), t_0(\tau), u_0(\tau))$ is a reference point shared between the hyper-surface solution and some characteristic hyper-plane projected onto the hyper-surface. The reference point is also an initial condition for each of the characteristic equations it describes.

By the inverse function theorem, the Jacobian of the transformation must be finite and non-zero to maintain uniqueness of the hyper-plane defined by the transformation. Suppose, as an example, that there is a unique inverse transformation $(x, y, z, t) \rightarrow (s, v, \rho, \tau)$, then we may write

$$u(x, y, z, t) = U\left(s(x, y, z, t), v(x, y, z, t), \rho(x, y, z, t), \tau(x, y, z, t)\right).$$

This is equivalent to the statement of the inverse function theorem, which implies

$$\begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} & \frac{\partial z}{\partial s} & \frac{\partial t}{\partial s} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} & \frac{\partial t}{\partial v} \\ \frac{\partial x}{\partial \rho} & \frac{\partial y}{\partial \rho} & \frac{\partial z}{\partial \rho} & \frac{\partial t}{\partial \rho} \\ \frac{\partial x}{\partial \tau} & \frac{\partial y}{\partial \tau} & \frac{\partial z}{\partial \tau} & \frac{\partial t}{\partial \tau} \end{vmatrix}^T \neq 0, \quad \text{or} \quad \infty. \quad (\text{B.45})$$

Therefore, this must hold for characteristic curves to be uniquely defined paths on the hyper-surface solution. To describe a path on the hyper-surface solution with a characteristic curve parametrised by, say s , the characteristic cannot intersect with another characteristic defined by any other fixed value of s, v, ρ, τ .

If we project characteristic curves onto the hyper-surface solution determines by real values $u(x, y, z, t)$, then regions where characteristics do not intersect correspond to regions where the Jacobian of the transformation, $(x, y, z, t) \rightarrow (s, v, \rho, \tau)$, is nonzero and finite. A characteristic curve, Γ , is unique whenever it exists and does not intersect non-trivially with any other characteristic curves and describes the solution surface on the path determined by Γ .

On the other hand, the existence of non-intersecting characteristic curves depends on the existence and uniqueness of solutions to the characteristic equations, which are ODEs. Recall the conditions for existence of solutions for ODEs. A solution to the characteristic equations exists and is necessarily unique if it satisfies the Lipschitz condition. The Lipschitz condition is satisfied whenever du/ds is bounded. This is true if each of its partial derivatives are bounded. Suppose the functions a, b, c are bounded by

$$|a(x_1, y_1, z_1, t_1, u_1) - a(x_2, y_2, z_2, t_2, u_2)| \leq K_x |x_1 - x_2|, \quad (\text{B.46})$$

$$|b(x_1, y_1, z_1, t_1, u_1) - b(x_2, y_2, z_2, t_2, u_2)| \leq K_y |y_1 - y_2|, \quad (\text{B.47})$$

$$|c(x_1, y_1, z_1, t_1, u_1) - c(x_2, y_2, z_2, t_2, u_2)| \leq K_z |z_1 - z_2|, \quad (\text{B.48})$$

$$|d(x_1, y_1, z_1, t_1, u_1) - d(x_2, y_2, z_2, t_2, u_2)| \leq K_t |t_1 - t_2|, \quad (\text{B.49})$$

$$|g(x_1, y_1, z_1, t_1, u_1) - g(x_2, y_2, z_2, t_2, u_2)| \leq K_u |u_1 - u_2|, \quad (\text{B.50})$$

then the Lipschitz condition is satisfied for,

$$\vec{f} = (a, b, c, d, g) = \frac{du}{ds} = \vec{f}(s, x, y, z, t, u),$$

which represents the solution surface for any initial conditions determined by s with respect to (x, y, z, t, u) and for Lipschitz constant, $\max\{K_x, K_y, K_z, K_t, K_u\}$.

Therefore, for characteristic curves to exist and be unique, the Jacobian of the transformation must satisfy

$$\det \begin{vmatrix} X_s & X_v & X_\rho & X_\tau \\ Y_s & Y_v & Y_\rho & Y_\tau \\ V_s & V_v & V_\rho & V_\tau \\ P_s & P_v & P_\rho & P_\tau \end{vmatrix} \neq 0 \quad \text{or} \quad \infty \quad (\text{B.51})$$

and at least each of a, b, c, d, g must be bounded. The Lipschitz condition merely specifies points where we may employ the method of characteristics and preserves continuity.

C Appendix: Cylindrical Coordinates

Observe that the geometry of an optical fiber is symmetrical with a cylindrical coordinate system. This section gives a derivation of the curl of a vector field in cylindrical coordinates to be used particularly in section 3 for interpreting and applying Maxwell's equations.

Begin by defining $x = r \cos\theta$, $y = r \sin\theta$, $z = z$. Then it follows that, $x^2 + y^2 = r^2(\cos^2\theta + \sin^2\theta) = r^2$ and $y/x = \tan\theta$, which implies $\sqrt{x^2 + y^2} = r$ and $\tan^{-1}(y/x) = \theta$. A cartesian vector of unital magnitude \vec{u} is given in cylindrical coordinates by,

$$\vec{u} = \begin{pmatrix} r \cos\theta \\ r \sin\theta \\ z \end{pmatrix} \quad (\text{C.1})$$

and so the cylindrical unit vectors may be derived for each direction \hat{r} , $\hat{\theta}$, \hat{z} by applying the operation

$$\hat{v} = \frac{d\vec{u}/dv}{|d\vec{u}/dv|}$$

to $v = r, \theta, z$ respectively. Then

$$\hat{r} = \frac{d\vec{u}}{dr} \bigg/ \left| \frac{d\vec{u}}{dr} \right| = (\sqrt{\cos^2\theta + \sin^2\theta})^{-1} \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix},$$

$$\hat{\theta} = \frac{d\vec{u}}{d\theta} \bigg/ \left| \frac{d\vec{u}}{d\theta} \right| = (r\sqrt{\cos^2\theta + \sin^2\theta})^{-1} \begin{pmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{pmatrix},$$

and

$$\hat{z} = \frac{d\vec{u}}{dz} \bigg/ \left| \frac{d\vec{u}}{dz} \right| = (\sqrt{1^2})^{-1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are the unit vectors associated with cylindrical coordinates. In order to define the curl in cylindrical coordinates, we first need to define the operation of the gradient in cylindrical coordinates. Recall, the gradient must satisfy

$$du = \vec{\nabla} u \circ d\vec{r}$$

where $u(r, \theta, z)$ is a static scalar field and the operation, ‘ d ’, is the total derivative and \vec{r} is a cartesian vector. By direct computation it follows that

$$du = \frac{\partial u}{\partial r} dr + \frac{\partial u}{\partial \theta} d\theta + \frac{\partial u}{\partial z} dz, \quad (\text{C.2})$$

$$\vec{\nabla} u \circ d\vec{r} = (\vec{\nabla} u)_r (d\vec{r})_r + (\vec{\nabla} u)_\theta (d\vec{r})_\theta + (\vec{\nabla} u)_z (d\vec{r})_z, \quad (\text{C.3})$$

$$d\vec{r} = d[r\hat{r} + z\hat{z}] = \hat{r} dr + r d\hat{r} + \hat{z} dz + z d\hat{z}, \quad (\text{C.4})$$

and

$$\begin{aligned} \frac{\partial \hat{r}}{\partial r} &= \vec{0}, & \frac{\partial \hat{r}}{\partial \theta} &= \hat{\theta}, & \frac{\partial \hat{r}}{\partial z} &= \vec{0}, \\ \frac{\partial \hat{\theta}}{\partial r} &= \vec{0}, & \frac{\partial \hat{\theta}}{\partial \theta} &= -\hat{r}, & \frac{\partial \hat{\theta}}{\partial z} &= \vec{0}, \\ \frac{\partial \hat{z}}{\partial r} &= \vec{0}, & \frac{\partial \hat{z}}{\partial \theta} &= \vec{0}, & \frac{\partial \hat{z}}{\partial z} &= \vec{0}. \end{aligned} \quad (\text{C.5})$$

The relation $\vec{r} = r\hat{r} + z\hat{z}$ in (C.4) is obtained by observing that any cartesian vector can be represented in this way for some r and z that are real constants. Applying (C.5) and (C.2) to (C.4) gives, $d\vec{r} = \hat{r}dr + \hat{\theta}r d\theta + \hat{z}dz$. Applying (C.4), in its new terms, to (C.3) gives the relation,

$$\vec{\nabla} u \circ d\vec{r} = (\vec{\nabla} u)_r (dr) + (\vec{\nabla} u)_\theta (r d\theta) + (\vec{\nabla} u)_z (dz) \quad (\text{C.6})$$

which is applied with (C.2) to the original assumption, (A.2), to obtain

$$\vec{\nabla} \equiv \hat{r} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{z} \frac{\partial}{\partial z} \quad (\text{C.7})$$

after equating the differential components of du with those of $\vec{\nabla}u \circ d\vec{r}$. Now we may define the curl in general by taking the cross product of $\vec{\nabla}$ with a vector field $\vec{A} = \vec{A}_r\hat{r} + \vec{A}_\theta\hat{\theta} + \vec{A}_z\hat{z}$. The curl in cylindrical coordinates is defined as

$$\begin{aligned}\vec{\nabla} \times \vec{A} &= \left(\hat{r} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{z} \frac{\partial}{\partial z} \right) \times \vec{A} \\ &= \left(\hat{r} \frac{\partial}{\partial r} \times \vec{A} \right) + \left(\hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} \times \vec{A} \right) + \left(\hat{z} \frac{\partial}{\partial z} \times \vec{A} \right) \\ &= \left(\hat{r} \times \frac{\partial \vec{A}}{\partial r} \right) + \left(\hat{\theta} \frac{1}{r} \times \frac{\partial \vec{A}}{\partial \theta} \right) + \left(\hat{z} \times \frac{\partial \vec{A}}{\partial z} \right),\end{aligned}$$

which can be expanded as

$$\begin{aligned}\vec{\nabla} \times \vec{A} &= \hat{r} \times \left(\frac{\partial(A_r\hat{r} + A_\theta\hat{\theta} + A_z\hat{z})}{\partial r} \right) + \frac{\hat{\theta}}{r} \times \left(\frac{\partial(A_r\hat{r} + A_\theta\hat{\theta} + A_z\hat{z})}{\partial \theta} \right) \\ &\quad + \hat{z} \times \left(\frac{\partial(A_r\hat{r} + A_\theta\hat{\theta} + A_z\hat{z})}{\partial z} \right). \quad (\text{C.8})\end{aligned}$$

Equation (C.9) simplifies by writing the cross products in terms of the unit vectors \hat{r} , $\hat{\theta}$ and \hat{z} , since cross products of unit vectors result in a unit vector by orthogonality. By doing this, we obtain the equation for curl in cylindrical coordinates,

$$\vec{\nabla} \times \vec{A} = \hat{r} \left(\frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z} \right) + \hat{\theta} \left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \right) + \hat{z} \left(\frac{1}{r} \frac{\partial A_r}{\partial \theta} + \frac{A_\theta}{r} - \frac{\partial A_\theta}{\partial r} \right). \quad (\text{C.9})$$

References

- [1] Abramowitz, M., Stegun, I. A., 1970, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications.
- [2] Agrawal, G. P., 1989, *Nonlinear Fiber Optics*, Academic Press.
- [3] Belanger, P. A., 1993, *Optical Fiber Theory*, World Scientific.

- [4] Coddington E. A., 1961, *An Introduction to Ordinary Differential Equations*, Dover Publications.
- [5] Fleisch, D. A., 2008, *A Student's Guide to Maxwell's Equations*, Cambridge University Press.
- [6] Kevorkian, J., 2001, *Patial Differential Equations: Analytic Solution Techniques*, Springer.
- [7] Knobel, R., 2000, *An Introduction to the Mathematical Theory of Waves*, The American Mathematical Society.
- [8] Okamoto, K., 2000, *Fundamentals of Optical Waveguides*, Academic Press.
- [9] Banerjee P. P., 1991, *Principles of Applied Optics*, Asken Associates Incorporated Publishers.
- [10] Reitz, John R., Milford, F. J., Christy, R. W., 2009, *Foundations of Electromagnetic Theory, 4th Ed.*, Pearson.
- [11] Sagan, H., 1992, *Introduction to the Calculus of Variations*, Dover Publications.
- [12] Weinberger H. F., 1965, *A First Course in Partial Differential Equations With Complex Variables and Transform Methods*, Dover Publications.
- [13] Whitham, G. B., 1999, *Linear and Non-Linear Waves*, Wiley.
- [14] <https://www.newport.com/t/fiber-optic-basics>
- [15] <https://zoomapps.files.wordpress.com/2015/02/total-internal-reflection-fibre-optic-cable.png>
- [16] <https://commons.wikimedia.org/wiki/User:Smedlib>

- [17] <https://s3-us-west-1.amazonaws.com/foscoshopify/graphics/H4P9b.png>
- [18] <http://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html>
- [19] <http://mathworld.wolfram.com/BesselFunctionoftheSecondKind.html>
- [20] <http://mathworld.wolfram.com/ModifiedBesselFunctionoftheSecondKind.html>
- [21] <http://mathworld.wolfram.com/ModifiedBesselFunctionoftheFirstKind.html>