1 THE BASICS (MATH, WAVES, etc.) FOR THE NONPHYSICAL SCIENTIST

Chapter Contents

1.1 General Properties of Waves and Coordinates
1.2 Trigonometry
1.3 Imaginary Numbers
1.4 The Exponential e
1.5 Infinite Series
1.6 Differentiation
1.7 Integration
1.8 Differential Equations
1.9 Vectors and Scalars
1.10 Unit Vector
1.11 Scalar and Vector Fields
1.12 Matrices and Linear Algebra
1.13 Waves
1.14 Combining Waves of Two Wavelengths
1.15 Fourier Series and Integrals
   1.15.1 Fourier Series
   1.15.2 Fourier Integral
In spite of the large number of equations in this text, for the most part, the central points can be understood without following the equations in detail. The equations are predominately provided for those with a background in the field. This introductory chapter is not for the purpose of teaching the nonphysical scientist how to solve math problems with these tools discussed. Instead, the purpose is to allow the reader to have a sense of the concepts that are discussed with these derivations. For those interested in a more detailed understanding of the mathematics, reference texts are listed at the end of each chapter. Discussed in this chapter are the basic topics of general waves, trigonometry, exponentials, imaginary numbers, differentiation, integration, differential equations, vectors, waves, and Fourier mathematics. These topics are ultimately found throughout other chapters of the text.

1.1 GENERAL PROPERTIES OF WAVES AND COORDINATES

A light wave of a single wavelength can be thought of as a fluctuation or cyclical phenomenon that is constantly changing the electromagnetic field amplitude (defined in Chapter 2) between crests and troughs. It can therefore be described by cyclical mathematical functions, such as the trigonometric functions of cosine and sine or the exponential $e^{ix}$, which will also be explained in more detail shortly. Only a few trigonometric functions will be used in this text, to simplify the understanding of OCT. But before discussing trigonometric functions, exponentials, or waves, it is important to understand both the concept of a coordinate system (polar vs. rectangular coordinates) and how the position of the wave in the coordinate system is quantified (degrees vs. radians).

In Figure 1.1A and B, the principles of rectangular and polar coordinates are illustrated for what is known as a transverse wave, where the amplitude (oscillation) is perpendicular to the direction of propagation. The relative position in the cycle and amplitude can be quantified in either of these coordinate systems. Figure 1.1A is a rectangular coordinate system where one axis ($x$) is measuring a distance or time interval while the $y$ axis measures the amplitude ($a$) of the wave. Light in a vacuum is a transverse wave. For the polar coordinates system in Figure 1.1B, amplitude is mapped as a function of angle or relative position in the cycle. Besides the coordinate system, the positions in the cycle are represented in Figure 1.1A and B, which are quantified in either units of degrees or radian. One complete cycle, where the wave returns to the same phase, can be described as either 360 degrees or $2\pi$ radians, which is seen in Figure 1.1A and B. The term for one complete cycle is the
Figure 1.1  (A) This figure is a rectangular coordinate system where one axis (x) is measuring a distance or time interval while the y axis measures the amplitude (a) of the wave. The positions in the cycle are also represented here. The two waves are at different relative places in their cycle, or "out of phase." 1A (1) represents the sine function while 1A (2) is the cosine function. The value of the cosine function is one at zero degrees while the sine function is zero at zero degrees. (B) This is a polar coordinate system. Here amplitude is mapped as a function of angle or relative position in the cycle. The positions in the cycle are also represented. (C) This image demonstrates a propagating longitudinal wave along a spring, which is distinct from the transverse wave in 1A.
wavelength and is denoted by the symbol \( \lambda \). The rate of wave fluctuation can also be described by the frequency \( (f) \), which is the number of complete cycles in a given time period. In this book, radians are generally used because they are easier to manipulate. It should also be noted that some waves move. Those waves have a speed that is determined by how quickly a given position on a wave moves in space or time. For example, the speed of light in a vacuum is \( 3 \times 10^8 \) m/sec. It should be noted that along with transverse waves, there are also longitudinal waves, of which sound (in most media of interest) or vibrations of a spring are examples. Figure 1.1C demonstrates a propagating longitudinal wave along a spring. With a longitudinal wave, oscillations (amplitude) are in the same direction as the direction of propagation of the wave. In this text we will focus primarily on transverse waves, which is how light behaves in a vacuum.

When more than one wave is present, the position of each wave in the cycle can be different relative to the other. Figure 1.1A shows two waves that are at different relative places in their cycle or out of phase. In the example in Figure 1.1A, the peak of one wave is misaligned with the trough of the second wave, meaning the waves are 90 degrees or \( \pi/2 \) out of phase. If the waves had been aligned, they would have been described as in phase.

### 1.2 TRIGONOMETRY

Trigonometric functions are one way of describing cyclical functions in numerical terms, which is therefore important in the understanding of optical coherence tomography (OCT). Sine and cosine functions are repeating functions that differ by 90 degrees in the phase, which can be seen in Figure 1.1A, where 1.1A (1) is the representation of the sine function and 1.1A (2) is the cosine function. The value of the cosine function is one at zero degrees while the sine function is zero at zero degrees. These functions will be important in describing the properties of light and OCT.

As stated, trigonometric functions will primarily be used to describe cyclical functions, but it should be noted that they can be and will be used in analyzing certain other circumstances, such as the angles between vectors (described below) and lengths within geometric structures (equivalent to evaluating the lengths of the sides of a triangle) as shown in Figure 1.2. This type of application should become more clear in later chapters.

There are many rules for manipulating and using trigonometric functions, but for the most part they do not need to be discussed unless brought up specifically in the text. Three important examples will be listed here. First, the cosine is symmetrical about the axis (if you flip the wave about the axis, it looks the same) while the sine is not. Therefore, \( \cos(-x) = \cos(x) \) {symmetric} and \( \sin(x) = -\sin(-x) \). Second, since they are cyclical functions, \( \cos(x + 2\pi) = \cos(x) \) and \( \sin(x + 2\pi) = \sin(x) \). Third, \( \sin^2 \phi + \cos^2 \phi = 1 \). Besides the sine and cosine, additional trigonometric functions are listed in Table 1.1.
1.3 IMAGINARY NUMBERS

Imaginary numbers will be used frequently throughout the text. To many people, it is concerning that numbers exist that are called imaginary. However, from a mathematical standpoint, it has been stated by many people that the concept of imaginary numbers is no stranger than negative numbers. In the early part of the last millennium, many had difficulty with negative numbers because the meaning of a negative object was not easy to comprehend. But today, few have difficulty comprehending what a negative balance on a credit card bill means, which is of course a negative number. Similarly, the imaginary number \( i \) (sometimes written as \( j \)) is just a mathematical tool to represent the square root of \(-1\), which has no other method of description. Imaginary numbers are useful for manipulating many of the

\[ Z \cos \theta = X \]
\[ Z \sin \theta = Y \]

**Figure 1.2** This figure shows the lengths within geometric structures (equivalent to evaluating the lengths of the sides of a triangle). It is a simple illustration of functions of trigonometry, sine, and cosine.

<table>
<thead>
<tr>
<th>Table 1.1 Trigonometry Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reciprocal Identities</strong></td>
</tr>
<tr>
<td>( \sin \theta = \frac{1}{\csc \theta} )</td>
</tr>
<tr>
<td>( \csc \theta = \frac{1}{\sin \theta} )</td>
</tr>
<tr>
<td><strong>Quotient Identities</strong></td>
</tr>
<tr>
<td>( \tan \theta = \frac{\sin \theta}{\cos \theta} )</td>
</tr>
<tr>
<td><strong>Pythagorean Identities</strong></td>
</tr>
<tr>
<td>( \sin^2 \theta + \cos^2 \theta = 1 )</td>
</tr>
</tbody>
</table>
equations in physics. For our purposes, which will be clearer later, imaginary numbers are very useful for representing both cyclical phenomena and the phase of a wave(s).

A complex number is one that contains both real and imaginary components (the value of either component may be zero) and can be written in the form \( z = x + iy \). In this equation, \( iy \) represents the term in the complex or imaginary domain while \( x \) is the “real” term. In general, these numbers may undergo the same operations as real numbers, with the example of addition shown here:

\[
z = (2 + 4i) + (3 + 5i) = 5 + 9i = x + iy
\]

An important concept with complex or imaginary numbers is the complex conjugate. If \( y = (2 + 4i) \), then \( y^* = (2 - 4i) \) is the complex conjugate. The multiplication of \( y \) by \( y^* \) yields a real rather than imaginary number \((2 + 8i - 8i + 16) = 18\). This operation will be performed throughout the text to generate real from imaginary numbers.

### 1.4 THE EXPONENTIAL \( e \)

The number \( e \) is a powerful and important number in science and in describing the principles behind OCT. It is used extensively throughout this text predominantly because it simplifies so many mathematical manipulations. In a large number of cases, it is used to replace the sine and cosine functions because in some cases it is easier for describing repetitive functions. While not particularly relevant to OCT physics, the actual value of \( e \) to eight significant digits is 2.7182818.

There are two reasons why \( e \) is useful to our study of OCT. First, it can be used to describe exponential growth or decay. In the formula, \( y = e^{ax} \), if \( a \) is positive, it is exponential growth; if \( a \) is negative, it is exponential decay. As light penetrates through tissue, in general, it roughly undergoes exponential decay that needs to be compensated for in OCT image processing. Figure 1.3 demonstrates an example of exponential growth.

![Figure 1.3](image) This demonstrates an example of exponential growth.
The second use of $e$, that is probably the most important with respect to this text, is its application in conjunction with imaginary numbers or more specifically the function $e^{ix}$. This is a function that, like the sine and cosine, is cyclical in nature. Using radians for the value of $x$, $e^{\pi i} = -1$, $e^0 = e^{2\pi i} = 1$, and $e^{\pi i/2} = \pm i$. This function can be plotted in polar coordinates as in Figure 1.4, so this again is a cyclic function and can be used similarly to the sine and cosine functions.

![Figure 1.4](image)

The second use of $e$, that is probably the most important with respect to this text, is its application in conjunction with imaginary numbers or more specifically the function $e^{ix}$. This is a function that, like the sine and cosine, is cyclical in nature. Using radians for the value of $x$, $e^{\pi i} = -1$, $e^0 = e^{2\pi i} = 1$, and $e^{\pi i/2} = \pm i$. This function can be plotted in polar coordinates as in Figure 1.4, so this again is a cyclic function and can be used similarly to the sine and cosine functions.

A similar manipulation with trigonometric functions would be significantly more difficult. Not surprisingly, since the function $e^{ix}$ is cyclical it can be related to the sine and cosine functions. One way they are related is through Euler’s formula, $e^{i\Theta} = \cos \Theta + \sin \Theta$. This relationship is derived in Appendix 1-1. This appendix utilizes a power series, discussed in the next section, so that the reader may wait before reading Appendix 1-1. Through algebraic manipulation, Euler’s equation may be rewritten in terms of the $\cos \Theta$ and $\sin \Theta$:

$$
\cos \Theta = \frac{(e^{i\Theta} + e^{-i\Theta})}{2} \quad \sin \Theta = \frac{(e^{i\Theta} - e^{-i\Theta})}{2i}
$$

(1.2)
1.5 INFINITE SERIES

Infinite series are frequently used in the physical sciences as well as pure mathematics and are literally the summation of an infinite number of terms. They can be used to evaluate functions including Fourier series, integrals, and differential equations. They have the general form of:

\[ \sum_{n=0}^{\infty} c_n x^n = f(x) \]

A power series is perhaps the most important infinite series. With this series, the power sequentially increases with each variable and the equation has the general form:

\[ \sum c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots = f(x) \]

A power series may converge (approach a specific value) or diverge with any value of \( x \). The domain of the function \( f(x) \) is all \( x \).

Example:

If \( c_n = 1 \).

\[ \sum x^n = 1 + x + x^2 + x^3 + \cdots = 1/(1 - x) = f(x) \]

This function converges when \(-1 < x < 1\) and diverges when \( |x| \geq 1 \). This allows the \( f(x) \) to be described in terms of a power series.

More generally, a power series has the form:

\[ \sum c_n (x - a)^n = c_0 + c_1 (x - a)^1 + c_2 (x - a)^2 + c_3 (x - a)^3 + \cdots \]

This is a power series in \( (x - a) \) or a power series about \( a \). Here, the power series specifies the value of a function at one point, \( x \), in terms of the value of the function relative to a reference point. It is essentially the expansion in powers of the change in a variable \( (\Delta x = x - a) \). When \( x = a \), all terms are equal to zero for \( n \geq 1 \) and so the power series always converges when \( x = a \).

In many cases, the power series does not completely describe the function and the remainder term is required such that:

\[ f(x) = P_n(x) + R_{N+1}(x) \]

This allows power series representation to be developed for a wide range of functions, where \( P_n(x) \) is the power series and \( R_{N+1}(x) \) is the remainder term. The remainder term may be represented by:

\[ R_{N+1}(x) = (f^{n+1}(c)x^{n+1})/(n + 1)! \]

where \( c \) has a value between 0 and \( x \).
A Taylor series expansion, a type of power series, is very useful for defining and manipulating many functions. If \( f \) has a power series representation (expansion) at \( a \), that is, if
\[
f(x) = \sum c_n(x-a)^n \quad |x-a| < R
\]
then its coefficients can be found by the formula:
\[
c_n = \frac{f^{(n)}(a)}{n!}
\]
Here \( f^{(n)}(a) \) represents the \( n \)th derivative. The power series expansion can now be rewritten as:
\[
f(x) = \sum \left( \frac{f^{(n)}(a)/n!}{(x-a)^n} \right) = f(a) + \left( \frac{f^1(a)}{1!}(x-a)^1 + \frac{f^2(a)}{2!}(x-a)^2 + \cdots \right)
\]
This is the Taylor series expansion of the function \( f \) centered about \( a \). Each coefficient can then be determined when each derivative has a value of \( x = a \) or \( f^{(n)}(0) \). This is best illustrated with the Taylor series expansion of \( e^x \). The value of \( a \) here is equal to 0 for simplicity. Then:
\[
\begin{align*}
f(0) &= e^0 = 1 \\
f^1(0) &= xe^0 = x \\
f^2(0) &= x^2e^0 = x^2 \ldots \text{etc.}
\end{align*}
\]
Then, substituting these coefficients we obtain:
\[
e^x = 1 + x + x^2/2! + \cdots + x^{n-1}/(n-1)!
\]
A wide range of functions can be represented by the power series.

**Calculus**

A large percentage of the mathematics used in this textbook involve integration or differentiation, which is the basis of calculus. The next two sections will try to give the nonphysical scientist a feel for the heart of calculus and the operations of differentiation and integration.

### 1.6 DIFFERENTIATION

Derivatives are very important in physics and in the understanding of the physical principles behind OCT. A derivative, for the purpose of this discussion, is a mathematical device for measuring the rate of change. But before going directly to the definition of a derivative, the concepts of a function and an incremental change must be discussed.

The term function will be used frequently in this text. In a relatively simple sense, a function can be described as a formula that tells you: If this is the value of one variable, how will the value of another variable be affected? In the following example, the variables \( x \) and \( y \) will represent distances.
\[
y = 2x
\]  
(1.3)
For this function, when $x$ is 2 meters (m), then it tells us that the position in the $y$ direction is 4 m. When $x$ is 3 m, $y$ is 6 m. Throughout the text, we will face more complex functions, but the overall principles behind a function are the same.

A function does not need to contain single numbers but may also contain variables of change, including incremental change. By incremental change, we mean a change in a variable, such as position in the $x$ direction ($\Delta x$), which is very small. So now, our function in Eq. 1.3 is looking at a change in each variable rather than identifying a specific position:

$$\Delta y = 2\Delta x$$  \hfill (1.4)

So if the position in the $x$ direction is changed by 3 m, the position in the $y$ is changed by 6 m. Now, we will define the concept of the average rate of change of a variable, using here the change in the position ($x$) with time ($t$), which is expressed as:

$$\frac{\Delta x}{\Delta t} = \text{change in } x/\text{change in } t$$

Speed, a relatively familiar concept, is an example of a measure of rate of change. It is a measure of the change in distance with respect to time. So if an airplane has traveled 100 miles in 1 hour, the rate of change ($\Delta$) or speed is 100 miles/hr.

The derivative is a special measure of rate of change. It is actually just the instantaneous (rather than incremental) rate of change of a variable over an infinitely small increment, rather than over a finite variable such as a minute.

Knowing that a derivative is related to an infinitely small rate of change, to more easily understand the derivative, the concept of limits is introduced. As an example such as $y = 1/x$, the function approaches a limit when $x$ approaches infinity ($\alpha$). The limit, the value of $y$, is 0 because $1/\alpha$ is zero.

This process can be generalized. If we use change in position ($\Delta x$) with respect to time as our variables, the limit as the time goes to zero ($\Delta t$) is:

$$\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

The term $dx/dt$ is referred to as the derivative. This is more clearly explained in Figure 1.5. The value for the incremental or average change in region A is an average rate of change in that region. If we make A smaller and smaller until it becomes the infinitely small point B, we now have the instantaneous rate of change at that point or the derivative. The derivative can be expressed in a variety of ways, such as $d f(x)/dx$, $y'$, $f'(x)$, and $dy/dx$.

One of the difficult aspects of the derivative is that there are a multitude of rules for taking the derivative of functions. However, since the reader is not required to derive anything in this book, these rules will not be covered and can be found in any calculus textbook. We will, though, give an example of a derivative of a function for illustrative purposes.

In this example function, an object is moving in such a way that the relationship between the time the object travels ($t$) and position of the object ($y$) is:

$$y = t^2$$  \hfill (1.5)
Then the velocity \( (v) \) or rate of change (derivative) of position is:

\[
\frac{dy}{dt} = 2t
\]

This result is a rule of calculus where in this example, the derivative of a square is 2 times that of the variable. Because the resulting velocity here is not constant, but changes with time (it is a function of \( t \) rather than a constant), the object is therefore accelerating or increasing its velocity (a positive value) with time.

The derivative can be taken more than once. Therefore, if a second derivative is obtained, what is produced is the rate of change of the original rate of change. Among the common ways or symbols of expressing the second order derivative are:

\[
\frac{d^2y}{dx^2}, f''(x), y'', \frac{d^2f(x)}{dx^2}
\]

Going back to Eq. 1.5 for our moving object and taking the second derivative, we get:

\[
\frac{d^2y}{dt^2} = 2
\]
In this case, the rate of change of the velocity in Eq. 1.6 is the acceleration. Because there are no variables left, just the constant 2, the object is moving at constant acceleration (but the velocity is not constant). If the second derivative had contained a variable, then the acceleration would not have been constant.

Again, there are a range of rules for performing derivatives, but they will not be discussed here, just the general principles. However, because the exponential will be used extensively throughout the text, an interesting point can be brought up with respect to this function. The exponential is the only function whose derivative is itself:

\[ y = e^x \text{ then } \frac{dy}{dx} = e^x \] (1.8)

### 1.7 INTEGRATION

Integration, to some degree, can be viewed as the inverse of differentiation. Although the integrals we will deal with are generally of a class known as definite, a brief discussion of indefinite integrals is in order as a foundation for definite integrals. In the previous discussion of differentiation, it was noted that if \( y = x^2 \), then \( \frac{dy}{dx} = 2x \). The indefinite integration of \( 2x \) is:

\[ y = \int 2x \, dx = x^2 + C \] (1.9)

When the conditions of integration, the region over which the integration is performed, are not defined (indefinite integral), we end up with the reverse of differentiation plus a constant \( C \), so it is not exactly a reverse of differentiation.

When the conditions of integrations are defined, for example between the physical points \( a \) and \( b \), it is referred to as a definite integral. This is shown in Figure 1.6 and results in the loss of the constant \( C \):

\[ y = \int_a^b 2x \, dx = \left[ b^2 + C \right] - \left[ a^2 + C \right] = b^2 - a^2 \] (1.10)

This, as it turns out, is the area under the curve between \( a \) and \( b \) in the Figure 1.6.

In general, with single integration, the results represent the area under a curve or line. However, unlike differentiation where the results generally always represent a rate of change, the results of integration are not always area (or volume), as we will see in an example with vector calculus. For example, we will see it used for measuring work performed along a given length later in the text.

Just like multiple differentiation, multiple integrations can be performed on a given function. An example of a double integral is given by the formula:

\[ z = \int_c^d \int_a^b \frac{1}{xy} \, dx \, dy \] (1.11)

This particular formula is measuring the volume under a surface rather than area under a curve.
The equations used in the text typically are more complex than these simple derivatives or integrals, but the principles are generally the same. Predominately, we are dealing with equations that contain differentials of multiple variables rather than single variables. They are referred to as differential equations. A differential equation is one that contains one or more derivatives, which will be more obvious in the following paragraphs. They are a very important class of equations that will be used extensively throughout the book and, for the most part, are used extensively throughout physics. A discussion on solving differential equations is well beyond the scope of this book, but this section is intended to give the reader a feel for these equations, making it easier to follow sections of this book. An example of a differential equation is the wave equation, which we will see again and which governs all kinds of physical phenomena including light waves, the motion of a spring, and the propagation of sound. This equation is:

\[ \frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2} \]  

The equation contains two partial derivatives or different types of derivatives, \( \frac{\partial^2 y}{\partial x^2} \) and \( \frac{\partial^2 y}{\partial t^2} \). In other words, these are functions that are differentiated with respect to different variables, \( x \) or \( t \), unlike the simple equations discussed up to this point. Classifying differential equations turns out to be very important in handling and understanding these equations. These equations are generally classified by three criteria: type, order, and degree.

The equation type is classified as either ordinary or partial. An ordinary equation contains only two variables, like \( x \) and \( y \). A partial differential equation contains
three or more variables. Most equations we will deal with are partial differential equations. Partial differential equations do not use the $dx$ symbol but a $\partial x$ symbol instead, which means that differentiation is occurring with respect to more than one variable. The wave equation shown above is an example of a partial differential equation, with derivatives both with respect to $t$ and $x$.

The order of a differential equation means how many times the function is differentiated. The highest order term, say a second order differentiation, is used to classify the equation. In the wave equation, both terms are second order. This means that any given term is differentiated twice.

Some equations contain exponential terms such as squares or cubes. When describing the degree of a differential equation, we are defining it by the highest exponential term. If the equation contains a square as its highest exponential, then the degree is two. Because the wave equation has no exponential terms higher than 1, its degree is one.

An important term used in describing differential equations is the linearity. If the degree of each term, in other words its exponential, is either zero or one, the equation is said to be linear. A more detailed explanation of linearity is found in the references listed at the end of this chapter. The vast majority of equations dealt with in this book will be linear. This is an important point because nonlinear equations can be difficult, if not impossible, to solve.

The wave equation is again a good example for understanding how differential equations are used. Classical waves such as a sound wave in air or ripples on water can be represented mathematically. These waves in general must satisfy or be solutions of the wave equation. The process of satisfying the wave equation is best given by the equation for a standing or stationary wave on a string of fixed length [2]:

$$y_n = (A_n \cos \omega_n t + B_n \cos \omega_n t) \sin \frac{\omega_n x}{c}$$  \hspace{1cm} (1.13)

The actual meaning of the terms is not critical, but the coefficients in the formula are determined by the specific experimental setup (length of the string, tension, etc.).

Determining if Eq. 1.13 is a solution to the wave equation is as follows. First, we perform a second order differentiation with respect to $x (\partial^2 y/\partial x^2)$ (left side of the wave equation) on Eq. 1.13 and plug it into the wave equation. The second step is to perform second order differentiation with respect to time ($\partial^2 y/\partial t^2$) and plug it into the right side of the wave equation. The two sides will be equal if Eq. 1.13 satisfies the wave equation. To study phenomena related to OCT, we frequently need to find solutions to the differential equations. The solutions will be stated, but from this section, hopefully the reader will now have an idea what satisfying a differential equation means.

### 1.9 VECTORS AND SCALARS

Throughout the text, the terms vector and scalar will be used. A scalar is a quantity that can be described with one number, the magnitude, such as speed, mass, or temperature. An example is a car with a speed of 5 miles per hour.
A vector is slightly more complex and requires defining both the magnitude and an addition quantity, usually direction. An example of a vector is velocity, as opposed to speed that is a scalar quantity. If we describe space in terms of $x$, $y$, and $z$ components, then the velocity, a vector, has a magnitude, but a component or percentage of the total magnitude is found within each of the $x$, $y$, and $z$ directions. This is shown in Figure 1.7. If the particle is moving at 10 m/sec at a 45° angle in the $x$–$y$ plane, its velocity in the $z$ direction ($v_z$) is zero, while its velocity in the $x$ and $y$ directions will be 7.07 m/sec (using Pythagorean’s theorem). If the particle is moving purely in the $x$ direction, its velocities in the $y$ and $z$ directions will be zero, while its velocity in the $x$ direction will be 10 m/sec. So once again, a vector does not just define a magnitude, but something in addition to the magnitude. In this case, it is the distribution in space of the magnitude.

1.10 UNIT VECTOR

An important tool for handling vectors is the unit vector. A unit vector is a vector that has unit magnitude or a value of one on each axis. This is easier to explain with the rectangular unit vectors $i$, $j$, and $k$. In Figure 1.7, we see these unit vectors plotted where each have a value of one. If we want to represent a given vector $F$, we can do it as a multiple of the individual unit vectors. Therefore:

$$F = a_1i + a_2j + a_3k$$  \hspace{1cm} (1.14)

Here the $a$'s in front of the unit vectors are scalar terms by which the unit vector is multiplied. Together, they fully describe the vector. The value of the use of unit vectors may not be obvious at this point, but it will be seen that this formalism allows
vectors in many instances to be more easily manipulated. The magnitude of vector $\mathbf{F}$ is given by:

$$\mathbf{F} = |\mathbf{F}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$$  \hspace{1cm} (1.15)

### 1.11 SCALAR AND VECTOR FIELDS

A field is an abstract concept that is easier to explain with an example. If we plot the temperature over say the United States, at any given time we are assigning a number value (i.e., temperature) for each location. We are therefore plotting temperature as a function of its $x$ and $y$ position. This type of field is a special type known as a scalar field because at any given time, the field only has one value at any given location. Now, if we plotted the velocity of snow in an avalanche, each point has velocity, a vector quantity. In other words, at any given time, each point in the avalanche has defined both magnitude and direction. This type of field is known as a vector field. We will learn later that electric and magnetic fields are vector fields.

### 1.12 MATRICES AND LINEAR ALGEBRA

Again, linear equations are special equations that are dealt with throughout this text. As stated, these equations contain no squares or higher powers (for example, $x^3$). These equations have two important properties. First, the addition of two linear equations results in a new equation that is linear. Second, when each term is multiplied by a constant, the final sum is a multiple of the same constant:

$$z = x + y \text{ then } az = ax + ay$$  \hspace{1cm} (1.16)

Very often we will be dealing with several of these linear equations when we analyze a system. An example is the equations that describe how light propagates through a series of lenses (or mirror or grating), assuming the components are linear. As light propagates through the series of lenses, each lens can be represented by a given linear equation. Matrices allow these multiple equations to be analyzed all at once. If, for example, we have the following linear equations:

$$a = bx + cy$$

$$d = ex + fy$$  \hspace{1cm} (1.17)

then in matrix form, the equations can be represented by:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} bc \\ ef \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$  \hspace{1cm} (1.18)
As with integration and differentiation, the rules that guide the handling of these matrices are well described, but will not be discussed here. These matrix equations will be seen intermittently throughout the text, but the mechanisms by which they are solved is not necessary to understand. Just an understanding as to what a matrix represents should be sufficient.

1.13 WAVES

Light has an inherent wave nature, as discussed. Therefore, it is necessary to discuss waves to understand even the simplest properties of light. Waves are introduced here rather than the next chapter, even though this chapter focuses on mathematics. This is because an understanding of waves is needed to complete our introduction of the relevant OCT mathematics, particularly the concept of Fourier mathematics. One way to look at a classical wave is to view it as a disturbance which is self-sustained and can move through a medium, transmitting energy and momentum. Although this definition is sufficient for this book, the reader should understand that it is a simplified definition that is not applicable to all situations, particularly quantum mechanics. In the next chapter, we will learn that in a vacuum distant from the source, light is a transverse electromagnetic wave. As stated previously, transverse wave means that the medium (or in the case of light, the electromagnetic field) is displaced perpendicular to the direction of motion. Sound in air, on the other hand, is a longitudinal wave. This means the displacement is in the direction of motion as shown in Figure 1.1C.

An important point is that waves are not limited to pure sinusoidal or harmonic functions. A pulse can be considered a wave although we will see that it too can be represented as the summation of specific pure sinusoidal functions.

A one-dimension wave can be described by the formula \( \psi(x, t) = F(x - vt) \), where \( \psi \) is a common symbol for a wave function, \( t \) is time, and \( v \) is velocity. \( F(x - vt) \) is a general function that tells us the shape of the wave while \( x \) may represent the position in one dimension. A common example function of \( \psi \), which is important in OCT imaging, is a Gaussian or bell-shape function \( \psi(x, t) = e^{-(x-\nu t)^2} \). This bell-shape disturbance is propagating at a velocity \( \nu \).

To begin initially examining waves quantitatively, we will choose to start with sinusoidal or simple harmonic functions. For now, we will begin with the sine function. Amplitude, phase, frequency, velocity, and position need to be defined for the wave function, but these components will be introduced separately. The initial wave function will only be a function of position and will have the form:

\[
\psi(x) = A \sin kx
\]

This function satisfies the wave equation. In the equation, \( A \) is the amplitude (the maximum) and \( k \) is the propagation constant. The propagation constant is related to the wavelength and its units are such that the product of \( kx \) has values of either degrees or radians. The spatial period is known as the wavelength and one complete
cycle is $2\pi$. Therefore, the propagation constant must have a value such that

$$k\lambda = 2\pi \quad \text{or} \quad k = \frac{2\pi}{\lambda}$$  \hspace{1cm} (1.20)

The wave is periodic in space. It should be noted that the wave we are describing currently is monochromatic, meaning that it consists of one wavelength and frequency. Soon, we will be introducing the mathematics to describe waves of multiple wavelengths. In reality, there are essentially no true monochromatic waves in the physical world. Even light from a laser contains a small range of wavelengths and this type of light generally is referred to as quasimonochromatic. The particular wave in Eq. 1.19 is a standing or stationary wave; it is not moving. To introduce the concept of movement, the equation is modified to include velocity:

$$\psi(x, t) = A \sin(kx + vt)$$  \hspace{1cm} (1.21)

So this wave is periodic in space and in time. The temporal period ($T$), the time for a complete cycle to pass a certain point, is described by:

$$kvT = 2\pi \quad \text{or} \quad T = \frac{\lambda}{v}$$  \hspace{1cm} (1.22)

The equation can also be modified to use a term that is often more convenient, the angular frequency ($\omega$). This is the number of radians per unit time and is defined as $2\pi/T$. It is the velocity multiplied by the propagation constant. To incorporate $\omega$, the wave function can be modified to:

$$\psi(x, t) = A \sin(kx + \omega t)$$  \hspace{1cm} (1.23)

Equations 1.21 and 1.23 are of course equivalent. In the description of waves so far, we have ignored the possibility of a phase shift or taken into account where the wave begins. When talking about sine and cosine functions, it was stated that they were 90 degrees out of phase. This means their crests and troughs do not overlap. From Figure 1.1A, the sine wave is zero at position zero. This does not need to be the case if we introduce a phase factor, as a shift in the phase of the wave in either space or time can be introduced. The equation that includes phase is:

$$\psi(x, t) = A \sin(kx + \omega t + \varepsilon)$$  \hspace{1cm} (1.24)

If our phase factor is $-90^\circ$ or $\pi/2$, the function will be equal to a cosine function, consistent with the definitions described above.

$$\psi(x, t) = A \sin(kx + \omega t - \pi/2) = A \cos(kx + \omega t)$$  \hspace{1cm} (1.25)

We have now introduced terms into the sine function to control the cyclical waves place as a function of time or space and velocity, in addition to taking into account any phase shift. All the terms in the parentheses now give the position of the wave in the cycle at a specific location and time, which we will describe as the phase factor ($\varphi(x, t)$) (not to be confused with the phase).

$$\varphi(x, t) = kx + \omega t + \varepsilon$$  \hspace{1cm} (1.26)

Now that we have defined the properties of a monochromatic wave, before we move on to polychromatic waves (waves of multiple wavelengths), several different
derivatives will be introduced. The values of these derivatives for OCT will become more obvious as we move to polychromatic waves, because rate of change of some variables has significance to the technology. If we take the derivative of the phase factor \( (\varphi(x, t)) \) with respect to time at constant position, as we learned in the derivative section, we get:

\[
|\left(\frac{\partial \varphi}{\partial t}\right)_x| = \omega
\] (1.27)

A derivative, as stated, is a measure of the rate of change. This is the rate of change of phase factor with time at fixed position, which is the angular momentum \( (\omega) \). Similarly, if we take the derivative with respect to position at fixed time, we get:

\[
|\left(\frac{\partial \varphi}{\partial x}\right)_t| = k
\] (1.28)

Therefore, the rate of change of phase factor with position is the propagation constant. With these two derivatives, a very important parameter can be defined, the phase velocity. Velocity is the change of position with time \( (\frac{\partial x}{\partial t})_x \). From this, using straightforward algebra, the phase velocity is derived:

\[
\frac{\partial x}{\partial t}_x = \frac{(\frac{\partial \varphi}{\partial t})_x}{(\frac{\partial \varphi}{\partial x})_t} = +\omega/k = +v
\] (1.29)

This is the speed of propagation of constant phase factor (for example, a peak) or the phase velocity, and is the angular velocity divided by the propagation constant.

Throughout the text, our representation of waves will not be limited to trigonometric functions. The exponential function, as stated, can also be used as a simple harmonic function taking into account position, velocity, wavelength, and initial phase. The complex representation, using an exponential function, is often easier to handle mathematically. The complex exponential representation can have the form of:

\[
\psi(x, t) = e^{-(kx + \omega t + \varepsilon)}
\] (1.30)

It should be noted that so far the monochromatic waves described have been one dimensional. However, most waves are really three dimensional, and we will represent them with the formula:

\[
\psi(\mathbf{r}, t) = A \sin (\mathbf{k} \cdot \mathbf{r} + \omega t + \varepsilon)
\] (1.31)

where \( \mathbf{r} \) now replaces \( x \) and represents the vector for the \( x \), \( y \), and \( z \) components. Therefore, we use \( \mathbf{r} \) to represent a vector that takes into account the three different spatial components of the wave. A more detailed understanding of \( \mathbf{r} \) is not necessary for the purposes of this text. The \( \mathbf{r} \) is in bold because this is a standard representation for vectors (containing the three spatial components).

1.14 COMBINING WAVES OF TWO WAVELENGTHS

To this point, we have only been discussing monochromatic waves. All light contains more than one wavelength and is therefore polychromatic. OCT imaging uses a very
wide range of wavelengths and is therefore often referred to as broadband light, a type of polychromatic wave. To study polychromatic waves, the discussion will begin by looking at light containing two different frequencies. Because the waves are linear, the waves can be added together and any combination of waves will be both linear and another solution to the wave equation. This is known as the principle of superposition.

We begin with two waves with different frequencies but the same amplitude and phase. The two waves are described by the wave functions $\psi_1(x, t)$ and $\psi_2(x, t)$, which have the formulas and summation described by:

$$\psi_1(x, t) = A \cos (k_1 x + \omega_1 t + \varepsilon_1)$$

$$\psi_2(x, t) = A \cos (k_2 x + \omega_2 t + \varepsilon_1)$$

$$\psi(x, t) = \psi_1(x, t) + \psi_2(x, t) = A \cos(k_1 x + \omega_1 t) + A \cos (k_2 x + \omega_2 t)$$  \hspace{1cm} (1.34)

The cosine representation is used here rather than the sine term for convenience because it allows the use of a simple trigonometric entity. The rule is that:

$$\cos \Theta + \cos \alpha = 2 \cos \left( \frac{\Theta - \alpha}{2} \right) \cos \left( \frac{\Theta + \alpha}{2} \right)$$  \hspace{1cm} (1.35)

Applying this rule to Eq. 1.34, we get:

$$\psi(x, t) = 2A \cos[(\omega_1 - \omega_2)t/2 - (k_1 - k_2)x/2]$$

$$\cos[(\omega_1 + \omega_2) t/2 - (k_1 + k_2)x/2]$$  \hspace{1cm} (1.36)

Rearranging the equation in this fashion gives us a better understanding as to what happens when the two waves are combined. We will begin with the assumption that the frequencies of the two waves are similar but not exactly the same. The reason for selecting waves of similar frequencies will be discussed later. The combined wave function in Eq. 1.36 consists of two components, a slowly varying envelope, which is the first cosine term, and a rapidly varying component, the second cosine term. A plot of this function is shown in Figure 1.8. What the first term represents is amplitude modulation or modulation of the total intensity. It is created because the crests and troughs of the waves at two different frequencies are adding at some points and canceling out at other points.

In addition to the slow amplitude modulation, there is a rapid oscillation in the signal, which is the second term in Eq. 1.36. Both the slow oscillation and the finer vibrations are seen in Figure 1.8.

Ultimately, with OCT, we will be concerned about the velocity of both the envelope (group velocity or beat velocity) and the velocity of the fine vibrations (phase velocities).

The phase velocity for the combination of two waves is:

$$\nu_p = \sigma/k = [1/2](\omega_1 + \omega_2)/(k_1 + k_2) = \frac{(\omega_1 + \omega_2)}{(k_1 + k_2)}$$  \hspace{1cm} (1.37)
When the frequencies are almost equal, \( \omega_1/k_1 \approx \omega_2/k_2 \approx v_p \). Therefore, through straightforward algebra:

\[
\omega_1 + \omega_2/k_1 + k_2 = v_p(k_1 + k_2)/k_1 + k_2 = v_p
\]

The group velocity is given by:

\[
v_g = 1/2(\omega_1 - \omega_2)/1/2(k_1 - k_2) = (\omega_1 - \omega_2)/(k_1 - k_2) = \frac{\Delta \omega}{\Delta k}
\]

The group velocity can be thought of as the velocity of the maximum amplitude of the group, so that it is the velocity with which the energy in the group is transmitted. The group velocity may be faster or slower than the phase velocity, which is a topic covered in Chapter 5.

### 1.15 FOURIER SERIES AND INTEGRALS

Now that the concept of polychromatic waves has been introduced, we will discuss the mathematics used to handle them, specifically Fourier mathematics. This is probably the most important area of mathematics used to describe the features of OCT.

In the previous section, we dealt with the addition of waves of two different frequencies that resulted in the formation of an AC component. To handle waves of many different frequencies and amplitudes, the analysis is not quite as simple and Fourier techniques are needed. This is because the crests and troughs add in a complex way. We begin with the example function, which is seen in Figure 1.9 and represents an infinite series of square pulses:

\[
\psi(x) = A(\sin x - 1/3 \sin 3x + 1/5 \sin 5x - 1/7 \sin 7x \ldots)
\]
For this function, we are adding different sinusoidal functions with different wavelengths and different weighting factors in front of the sine term to produce the pulse series. Figure 1.9 demonstrates how adding the factors individually progressively results in a function that more closely resembles the series of square pulses. Therefore, Eq. 1.40 can be viewed as consisting of a spectrum or series of different frequencies, which when added together form a specific pattern or function. In the case of Eq. 1.40, a repeating pattern of pulses is produced.

One way to describe these results is that a Fourier relationship exists between the amplitude distribution in the time domain (Figure 1.9) and the frequency distribution in the frequency domain (Eq. 1.40). This is actually a Fourier series that is discussed in detail in the next section. Fourier mathematics is describing how multiple frequencies at different amplitudes add up to give specific intensity shapes, again like a series of pulses.

Two different types of Fourier approaches exist, the Fourier series and the Fourier integral. Fourier series are used with periodic functions like the one in Figure 1.10. Fourier integrals are used with non-periodic functions, like a single pulse. Again, Fourier relationships are central to the understanding of the principles behind OCT, so we will discuss both.

### 1.15.1 Fourier Series

A statement of and the formula for a Fourier series (more fundamental than Eq. 1.40) will be presented but not proved. The statement is that virtually all periodic
Figure 1.10  This figure demonstrates a series of square pulses and the corresponding frequency distribution in the Fourier domain. We see that as the frequencies within the beam get closer and closer, the square pulses get farther and farther apart. Therefore, if the frequencies summed up were spaced an infinitely small distance apart (i.e., no detectable separation), then we would ultimately get a single pulse. Of course, summing over an infinitely small distance is integration and is the basis of the Fourier integral. Courtesy of Hecht Optics, Fourth Edition, Figure 7.21, page 304 [2].
functions, like the one in Figure 1.9, for the purposes of this text, can be represented by the Fourier series:

$$
\psi(x) = a_0/2 + a_n \cos nkx + b_n \sin nkx
$$

(1.41)

Here $a_0$, $a_n$, and $b_n$ are coefficients which weigh the individual terms in the series that vary depending on the periodic function. The coefficients are appropriately weighed to produce the periodic function, some of which may have a value of zero. We see that Eq. 1.40 is a version of Eq. 1.41, where the value is zero for $a_0/2$, all the coefficients of the cosine terms, and every other coefficient of the sine terms.

In Figure 1.10, we see a series of square pulses and the corresponding frequency distribution. We see that as the frequencies within the beam get closer and closer, the square pulses get farther and farther apart. Therefore, if the frequencies summed were spaced an infinitely small distance apart (i.e., no detectable separation), then we would ultimately get a single pulse. Of course, summing over an infinitely small distance is integration and is the basis of the Fourier integral. Therefore, to describe in the frequency domain a non-repeating structure, such as a Gaussian pulse, integration is required.

### 1.15.2 Fourier Integral

The Fourier integral can be represented in a variety of ways. If you are looking at a pulse energy amplitude as a function of time, such as a femtosecond pulse, this is called analyzing in the time domain. If you are looking at the frequencies that add up to make the pulse, it is stated that you are analyzing in the frequency domain. Again, Fourier integrals will allow us to relate the frequency and time domains. The actual Fourier integrals can be represented in a range of ways. Different texts represent the integral in other ways, which are all equivalent. The representation used here, which will not be proved, is:

$$
\begin{align*}
    f(t) &= \int_{-\infty}^{\infty} F(v)e^{i2\pi vt} dt \\
    F(v) &= \int_{-\infty}^{\infty} f(t)e^{-i2\pi vt} dt
\end{align*}
$$

$F(v)$ is the frequency distribution (i.e., amplitude of different frequencies) of a non-periodic function (for example, Gaussian pulse), whereas $f(t)$ is the time distribution of the amplitude of a non-periodic function or shape. In the first equation, $f(t)$ is referred to as the Fourier transform of $F(v)$. In the second equation, $F(v)$ is referred to as the inverse Fourier transform of $f(t)$. What the equation is saying is that the distribution in time of a non-periodic function $f(t)$ can be found from the integral of the various frequencies multiplied by an exponential. The second equation is basically the reverse process. Again, it should be reinforced that Eqs. 1.42 and 1.43 are written in slightly different ways in different textbooks and sometimes even in
This text. This is because the \( \pi \) term and negative sign can be placed in a range of positions. All work equally as well, as long as the changes made in the two equations are consistent.

Why is Fourier math so important in understanding OCT? As an example, the sources used with OCT need to be broadband or contain a large number of frequencies. The reason is that the size of an appropriately shaped frequency distribution is proportional to the resolution of the system. This will be covered later in much greater detail.

### 1.16 DIRAC DELTA FUNCTION

An important function that we will be using throughout the text is the Dirac delta function. At first glance it will seem like a trivial function, but its significance becomes more obvious as it is actually put into use. The delta function is defined as:

\[
\delta(a - a_0) = 0 \quad \text{when } a \neq a_0
\]

So this is a function that has a nonzero value only at one value, \( a_0 \). This will turn out to be a very useful mathematical device. The value of its area at \( a_0 \) is 1, while its height is infinite. One of its greatest features is the sifting property, where it is a method for assigning a number to a function at given points.

\[
\int_{-\infty}^{\infty} f(a) \delta(a - a_0) = f(a_0)
\]  

This is important, for example, in the process of digitizing or dividing continuous data into segments. When a continuous signal is digitized, it is sampled or broken up at regular intervals. The mathematical procedure for doing this utilizes a series of Dirac delta functions.

### 1.17 MISCELLANEOUS BASIC PHYSICS

Before concluding this section, in addition to reviewing the relevant mathematics, some basic terms used in physics need to be defined including momentum, force, energy, power, work, and decibels. Momentum is the mass of an object times the velocity. Force, which has units here of Newtons, is defined in two ways. One is the change in momentum per unit time. In other words, an object’s momentum remains constant unless acted on by a force. The second definition is that it is mass times the acceleration. They are equivalent definitions. Work can be defined as the force times the distance over which something is moved. Power is the change in work per unit time or how fast work is being done. Its units are watts or joules/second. So power is the rate at which work is being performed.

Energy is a little more difficult to define than power, work, or momentum and is one of the few things that is easier to envision on a quantum level than a classical one. The units for energy are joules or Newtons times meters. While momentum,
for example, is a measurement of a changeable parameter, energy, like mass, is something that simply exists in the universe. Energy is something, in combination with mass (equivalent to energy through \( e = mc^2 \)), which is conserved so that the total sum of mass and energy for the universe never changes even though it may change for a given system. Particles and fields possess energy (and can gain or lose it) in the form of either stored (potential) or active (kinetic) energy. The most concrete explanation for energy is that energy is what changes when work is performed on a system. (It should be noted that the concept that the mass and energy of the universe is constant is a belief but when taking into account the entire universe, this may not be completely true. Some theories describing the universe actually suggest it may not be constant. However, in our everyday world, it can be considered constant.) In subsequent chapters we will be concerned about the irradiance, which is the rate of flow of energy and is proportional to the square of the amplitude of the electric field.

With OCT, the maximal and minimal signal intensity measured is very large so that the decibel system, a method of reducing the number of digits used to describe data, is used for quantitating signal intensity. For an optical field, a decibel is:

\[
\text{dB} = 10 \log \left( \frac{I}{I_0} \right)
\]

Here, \( I_0 \) is the minimal signal detected by the system. For those who do not remember a log (base 10) from high school, it is the power 10 needs to be raised in order to get a certain number. So the log of 10 is 1, the log of 100 is 2. With OCT, the range of signal that can be detected by OCT is approximately 110 dB. This means that \( I \) is \( 10^{11} \) higher than \( I_0 \). This will be dealt with in more detail in Chapter 7.

**REFERENCES**

DERIVING EULER’S FORMULA

A power series will be used to prove the relationship between the imaginary exponential and the sine/cosine terms through Euler’s formula. A power series is a special type of infinite series. An infinite series is a common device in science and mathematics and is literally a summation of an infinite number of terms. The power series has the form:

\[ \sum c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n \ldots \]

Functions like cosine and sine can be broken down into a power series or infinite series of numbers that have a specific relationship to one another. The power series representation of \( e \), the sine function, and the cosine function are well known and can be found in most calculus textbooks. The power series for \( e^x \) is:

\[ e^x = 1 + x + x^2/2! + x^3/3! + x^4/4! \ldots \]

The symbol ! means factorial and can best be described by an example: the term \( 4! = 4 \times 3 \times 2 \times 1 = 24 \). The power series for \( e^{ix} \) is:

\[ e^{ix} = 1 + ix - x^2/2! - ix^3/3! + x^4/4! \ldots \]

Notice that not all terms are imaginary and the signs are no longer all positive. This is because, for example, \( -x^2/2! \), the square of \( i \) is \( -1 \). For \( -ix^3/3! \), the cube of \( i \) is \( -i \).

Now the imaginary and real terms can be grouped as follows:

\[ e^{ix} = (1 - x^2/2! + x^4/4! \ldots) + i(x - x^3/3! + x^5/5! \ldots) \]

The power series in the first parentheses is that for the cosine function while the one in the second parentheses is the sine function multiplied by \( i \). Therefore:

\[ e^{ix} = \cos x + i \sin x \]