Chapter 1

Waves in Two and Three Dimensions

In this chapter we extend the ideas of the previous chapter to the case of waves in more than one dimension. The extension of the sine wave to higher dimensions is the *plane wave*. Wave packets in two and three dimensions arise when plane waves moving in different directions are superimposed.

Diffraction results from the disruption of a wave which is impingent upon an object. Those parts of the wave front hitting the object are scattered, modified, or destroyed. The resulting *diffraction pattern* comes from the subsequent interference of the various pieces of the modified wave. A knowledge of diffraction is necessary to understand the behavior and limitations of optical instruments such as telescopes.

Diffraction and interference in two and three dimensions can be manipulated to produce useful devices such as the *diffraction grating*.

1.1 Math Tutorial — Vectors

Before we can proceed further we need to explore the idea of a *vector*. A vector is a quantity which expresses both magnitude and direction. Graphically we represent a vector as an arrow. In typeset notation a vector is represented by a boldface character, while in handwriting an arrow is drawn over the character representing the vector.

Figure 1.1 shows some examples of *displacement vectors*, i. e., vectors which represent the displacement of one object from another, and introduces

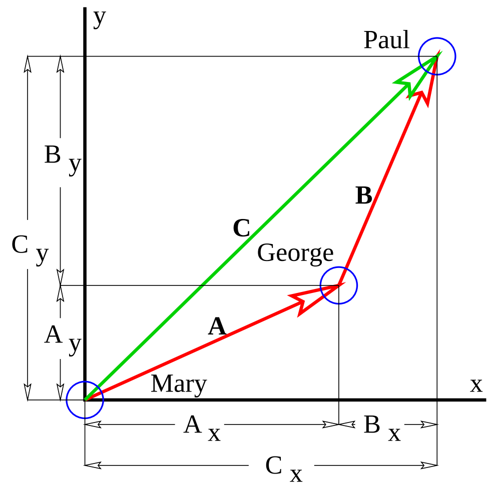

Figure 1.1: Displacement vectors in a plane. Vector **A** represents the displacement of George from Mary, while vector **B** represents the displacement of Paul from George. Vector **C** represents the displacement of Paul from Mary and $\mathbf{C} = \mathbf{A} + \mathbf{B}$. The quantities A_x, A_y, etc., represent the Cartesian components of the vectors.

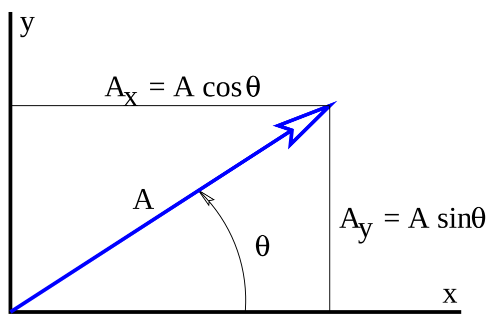

Figure 1.2: Definition sketch for the angle θ representing the orientation of a two dimensional vector.

the idea of vector addition. The tail of vector **B** is collocated with the head of vector **A**, and the vector which stretches from the tail of **A** to the head of **B** is the sum of **A** and **B**, called **C** in figure 1.1.

The quantities A_x, A_y, etc., represent the *Cartesian components* of the vectors in figure 1.1. A vector can be represented either by its Cartesian components, which are just the projections of the vector onto the Cartesian coordinate axes, or by its direction and magnitude. The direction of a vector in two dimensions is generally represented by the counterclockwise angle of the vector relative to the x axis, as shown in figure 1.2. Conversion from one form to the other is given by the equations

$$A = (A_x^2 + A_y^2)^{1/2} \qquad \theta = \tan^{-1}(A_y/A_x), \tag{1.1}$$

$$A_x = A\cos(\theta) \qquad A_y = A\sin(\theta), \tag{1.2}$$

where A is the magnitude of the vector. A vector magnitude is sometimes represented by absolute value notation: $A \equiv |\mathbf{A}|$.

Notice that the inverse tangent gives a result which is ambiguous relative to adding or subtracting integer multiples of π. Thus the quadrant in which the angle lies must be resolved by independently examining the signs of A_x and A_y and choosing the appropriate value of θ.

To add two vectors, **A** and **B**, it is easiest to convert them to Cartesian component form. The components of the sum $\mathbf{C} = \mathbf{A} + \mathbf{B}$ are then just the sums of the components:

$$C_x = A_x + B_x \qquad C_y = A_y + B_y. \tag{1.3}$$

Subtraction of vectors is done similarly, e. g., if $\mathbf{A} = \mathbf{C} - \mathbf{B}$, then

$$A_x = C_x - B_x \qquad A_y = C_y - B_y. \tag{1.4}$$

A *unit vector* is a vector of unit length. One can always construct a unit vector from an ordinary vector by dividing the vector by its length: $\mathbf{n} = \mathbf{A}/|\mathbf{A}|$. This division operation is carried out by dividing each of the vector components by the number in the denominator. Alternatively, if the vector is expressed in terms of length and direction, the magnitude of the vector is divided by the denominator and the direction is unchanged.

Unit vectors can be used to define a Cartesian coordinate system. Conventionally, **i**, **j**, and **k** indicate the x, y, and z axes of such a system. Note that **i**, **j**, and **k** are mutually perpendicular. Any vector can be represented in terms

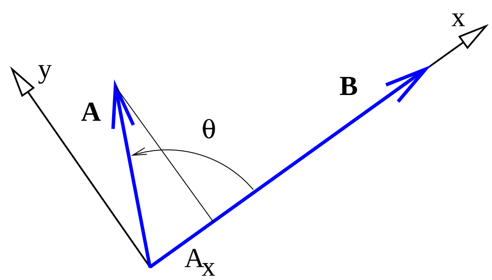

Figure 1.3: Definition sketch for dot product.

of unit vectors and its Cartesian components: $\mathbf{A} = A_x\mathbf{i}+A_y\mathbf{j}+A_z\mathbf{k}$. An alternate way to represent a vector is as a list of components: $\mathbf{A} = (A_x, A_y, A_z)$. We tend to use the latter representation since it is somewhat more economical notation.

There are two ways to multiply two vectors, yielding respectively what are known as the *dot product* and the *cross product*. The cross product yields another vector while the dot product yields a number. Here we will discuss only the dot product. The cross product will be presented later when it is needed.

Given vectors $\mathbf{A}$ and $\mathbf{B}$, the dot product of the two is defined as

$$\mathbf{A} \cdot \mathbf{B} \equiv |\mathbf{A}||\mathbf{B}| \cos\theta, \tag{1.5}$$

where θ is the angle between the two vectors. In two dimensions an alternate expression for the dot product exists in terms of the Cartesian components of the vectors:

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y. \tag{1.6}$$

It is easy to show that this is equivalent to the cosine form of the dot product when the x axis lies along one of the vectors, as in figure 1.3. Notice in particular that $A_x = |\mathbf{A}|\cos\theta$, while $B_x = |\mathbf{B}|$ and $B_y = 0$. Thus, $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}|\cos\theta|\mathbf{B}|$ in this case, which is identical to the form given in equation (1.5).

All that remains to be proven for equation (1.6) to hold in general is to show that it yields the same answer regardless of how the Cartesian coordinate system is oriented relative to the vectors. To do this, we must show that $A_x B_x + A_y B_y = A'_x B'_x + A'_y B'_y$, where the primes indicate components in a coordinate system rotated from the original coordinate system.

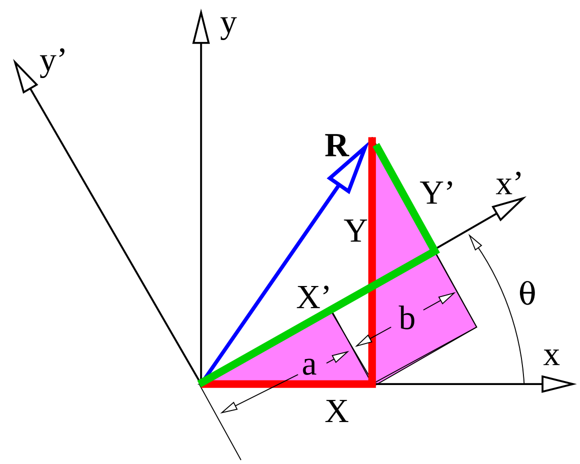

Figure 1.4: Definition figure for rotated coordinate system. The vector **R** has components X and Y in the unprimed coordinate system and components X' and Y' in the primed coordinate system.

Figure 1.4 shows the vector **R** resolved in two coordinate systems rotated with respect to each other. From this figure it is clear that $X' = a + b$. Focusing on the shaded triangles, we see that $a = X\cos\theta$ and $b = Y\sin\theta$. Thus, we find $X' = X\cos\theta + Y\sin\theta$. Similar reasoning shows that $Y' = -X\sin\theta + Y\cos\theta$. Substituting these and using the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$ results in

$$\begin{aligned} A'_x B'_x + A'_y B'_y &= (A_x\cos\theta + A_y\sin\theta)(B_x\cos\theta + B_y\sin\theta) \\ &+ (-A_x\sin\theta + A_y\cos\theta)(-B_x\sin\theta + B_y\cos\theta) \\ &= A_x B_x + A_y B_y \end{aligned} \tag{1.7}$$

thus proving the complete equivalence of the two forms of the dot product as given by equations (1.5) and (1.6). Multiply out the above expression to verify this.

A numerical quantity that doesn't depend on which coordinate system is being used is called a *scalar*. The dot product of two vectors is a scalar. However, the components of a vector, taken individually, are not scalars, since the components change as the coordinate system changes. Since the laws of physics cannot depend on the choice of coordinate system being used,

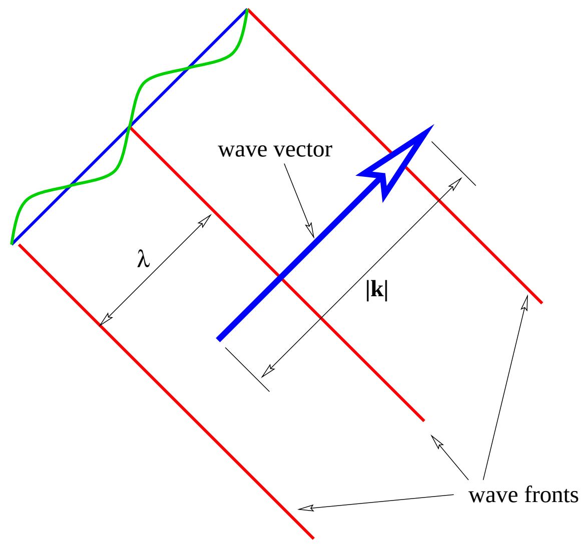

Figure 1.5: Definition sketch for a plane sine wave in two dimensions. The wave fronts are constant phase surfaces separated by one wavelength. The wave vector is normal to the wave fronts and its length is the wavenumber.

---

we insist that physical laws be expressed in terms of scalars and vectors, but not in terms of the components of vectors.

In three dimensions the cosine form of the dot product remains the same, while the component form is

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z. \tag{1.8}$$

1.2 Plane Waves

A plane wave in two or three dimensions is like a sine wave in one dimension except that crests and troughs aren't points, but form lines (2-D) or planes (3-D) perpendicular to the direction of wave propagation. Figure 1.5 shows a plane sine wave in two dimensions. The large arrow is a vector called the *wave vector*, which defines (1) the direction of wave propagation by its orientation perpendicular to the wave fronts, and (2) the wavenumber by its length. We can think of a wave front as a line along the crest of the wave.

The equation for the displacement associated with a plane sine wave in three dimensions at some instant in time is

$$h(x, y, z) = \sin(\mathbf{k} \cdot \mathbf{x}) = \sin(k_x x + k_y y + k_z z). \tag{1.9}$$

Since wave fronts are lines or surfaces of constant phase, the equation defining a wave front is simply $\mathbf{k} \cdot \mathbf{x} = const$.

In the two dimensional case we simply set $k_z = 0$. Therefore, a wave front, or line of constant phase ϕ in two dimensions is defined by the equation

$$\mathbf{k} \cdot \mathbf{x} = k_x x + k_y y = \phi \quad \text{(two dimensions)}. \tag{1.10}$$

This can be easily solved for y to obtain the slope and intercept of the wave front in two dimensions.

As for one dimensional waves, the time evolution of the wave is obtained by adding a term $-\omega t$ to the phase of the wave. In three dimensions the wave displacement as a function of both space and time is given by

$$h(x, y, z, t) = \sin(k_x x + k_y y + k_z z - \omega t). \tag{1.11}$$

The frequency depends in general on all three components of the wave vector. The form of this function, $\omega = \omega(k_x, k_y, k_z)$, which as in the one dimensional case is called the *dispersion relation*, contains information about the physical behavior of the wave.

Some examples of dispersion relations for waves in two dimensions are as follows:

- Light waves in a vacuum in two dimensions obey

  $$\omega = c(k_x^2 + k_y^2)^{1/2} \quad \text{(light)}, \tag{1.12}$$

  where c is the speed of light in a vacuum.

- Deep water ocean waves in two dimensions obey

  $$\omega = g^{1/2}(k_x^2 + k_y^2)^{1/4} \quad \text{(ocean waves)}, \tag{1.13}$$

  where g is the strength of the Earth's gravitational field as before.

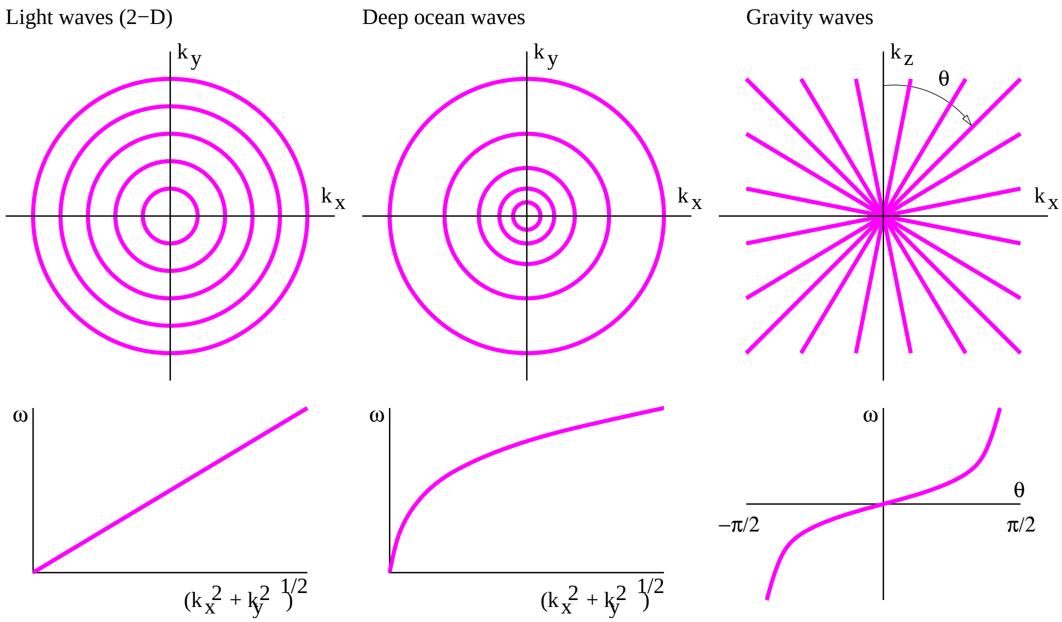

Figure 1.6: Contour plots of the dispersion relations for three kinds of waves in two dimensions. In the upper panels the curves show lines or contours along which the frequency ω takes on constant values. Contours are drawn for equally spaced values of ω. For light and ocean waves the frequency depends only on the magnitude of the wave vector, whereas for gravity waves it depends only on the wave vector's direction, as defined by the angle θ in the upper right panel. These dependences for each wave type are illustrated in the lower panels.

- Certain kinds of atmospheric waves confined to a vertical $x-z$ plane called *gravity waves* (not to be confused with the gravitational waves of general relativity)[1] obey

$$\omega = \frac{N k_x}{k_z} \quad \text{(gravity waves)}, \tag{1.14}$$

where N is a constant with the dimensions of inverse time called the

[1]Gravity waves in the atmosphere are vertical or slantwise oscillations of air parcels produced by buoyancy forces which push parcels back toward their original elevation after a vertical displacement.

Brunt-Väisälä frequency.

Contour plots of these dispersion relations are plotted in the upper panels of figure 1.6. These plots are to be interpreted like topographic maps, where the lines represent contours of constant elevation. In the case of figure 1.6, constant values of frequency are represented instead. For simplicity, the actual values of frequency are not labeled on the contour plots, but are represented in the graphs in the lower panels. This is possible because frequency depends only on wave vector magnitude $(k_x^2 + k_y^2)^{1/2}$ for the first two examples, and only on wave vector direction θ for the third.

1.3 Superposition of Plane Waves

We now study wave packets in two dimensions by asking what the superposition of two plane sine waves looks like. If the two waves have different wavenumbers, but their wave vectors point in the same direction, the results are identical to those presented in the previous chapter, except that the wave packets are indefinitely elongated without change in form in the direction perpendicular to the wave vector. The wave packets produced in this case move in the direction of the wave vectors and thus appear to a stationary observer like a series of passing *pulses* with broad lateral extent.

Superimposing two plane waves which have the same frequency results in a stationary wave packet through which the individual wave fronts pass. This wave packet is also elongated indefinitely in some direction, but the direction of elongation depends on the dispersion relation for the waves being considered. These wave packets are in the form of steady *beams*, which guide the individual phase waves in some direction, but don't themselves change with time. By superimposing multiple plane waves, all with the same frequency, one can actually produce a single stationary beam, just as one can produce an isolated pulse by superimposing multiple waves with wave vectors pointing in the same direction.

If the frequency of a wave depends on the magnitude of the wave vector, but not on its direction, the wave's dispersion relation is called *isotropic*; otherwise it is *anisotropic*. In the isotropic case, two waves have the same frequency only if the lengths of their wave vectors, and hence their wavelengths, are the same. The first two examples in figure 1.6 satisfy this condition, while the last example is anisotropic.

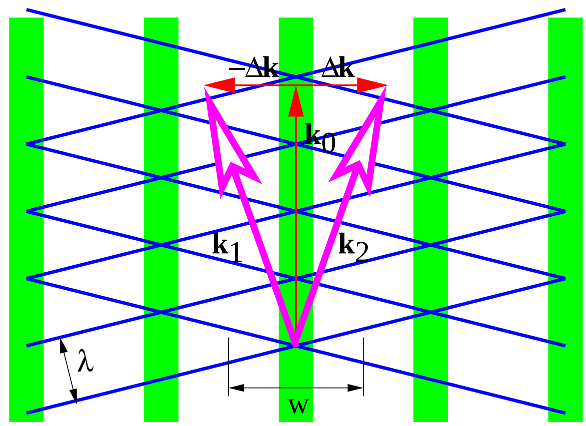

Figure 1.7: Wave fronts and wave vectors ($\mathbf{k}_1$ and $\mathbf{k}_2$) of two plane waves with the same wavelength but oriented in different directions. The vertical bands show regions of constructive interference where wave fronts coincide. The vertical regions in between have destructive interference, and hence define the lateral boundaries of the beams produced by the superposition. The quantities $\mathbf{k}_0$ and $\Delta\mathbf{k}$ are also shown.

We now use the language of vectors to investigate the superposition of two plane waves with wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$:

$$h = \sin(\mathbf{k}_1 \cdot \mathbf{x} - \omega t) + \sin(\mathbf{k}_2 \cdot \mathbf{x} - \omega t). \tag{1.15}$$

Applying the trigonometric identity for the sine of the sum of two angles (as we have done previously), equation (1.15) can be reduced to

$$h = 2\sin(\mathbf{k}_0 \cdot \mathbf{x} - \omega t)\cos(\Delta\mathbf{k} \cdot x) \tag{1.16}$$

where

$$\mathbf{k}_0 = (\mathbf{k}_1 + \mathbf{k}_2)/2 \qquad \Delta\mathbf{k} = (\mathbf{k}_2 - \mathbf{k}_1)/2. \tag{1.17}$$

This is in the form of a sine wave moving in the $\mathbf{k}_0$ direction with phase speed $c_{phase} = \omega/|\mathbf{k}_0|$ and wavenumber $|\mathbf{k}_0|$, modulated in the $\Delta\mathbf{k}$ direction by a cosine function. The lines of destructive interference are normal to $\Delta\mathbf{k}$. The distance w between lines of destructive interference is the distance between successive zeros of the cosine function in equation (1.16), implying

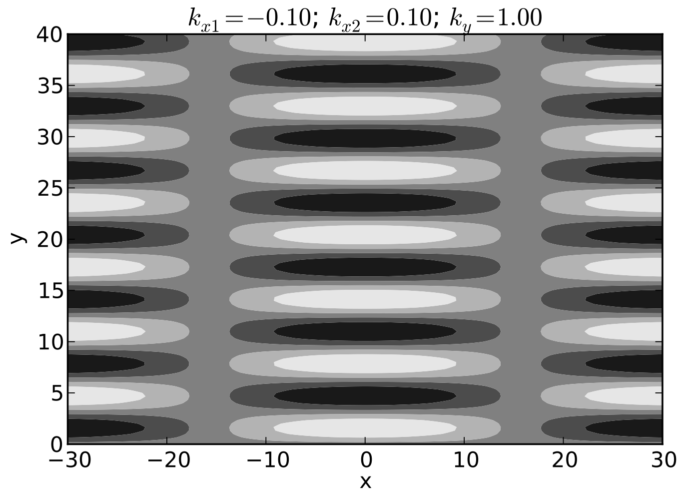

Figure 1.8: Example of beams produced by two plane waves with the same wavelength moving in different directions. The wave vectors of the two waves are $\mathbf{k} = (\pm 0.1, 1.0)$. Regions of positive displacement are lighter, while regions of negative displacement are darker.

---

that $|\Delta\mathbf{k}|w = \pi$, which leads to

$$w = \pi/|\Delta\mathbf{k}|. \tag{1.18}$$

Thus, the smaller $|\Delta\mathbf{k}|$, the greater is the beam diameter.

1.3.1 Two Waves of Identical Wavelength

In this section we investigate the beams produced by superimposing isotropic waves of the same frequency. Figure 1.7 illustrates what happens in such a superposition. Vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ of equal length give rise to a mean wave vector $\mathbf{k}_0$ and half the difference, $\Delta\mathbf{k}$. As illustrated, the lines of constructive and destructive interference are perpendicular to $\Delta\mathbf{k}$. Figure 1.8 shows a concrete example of the beams produced by superposition of two plane waves of equal wavelength oriented as in figure 1.7. The beams are aligned vertically, since $\Delta\mathbf{k}$ is horizontal, with the lines of destructive interference separating the beams located near $x = \pm 16$. The transverse width of the beams of

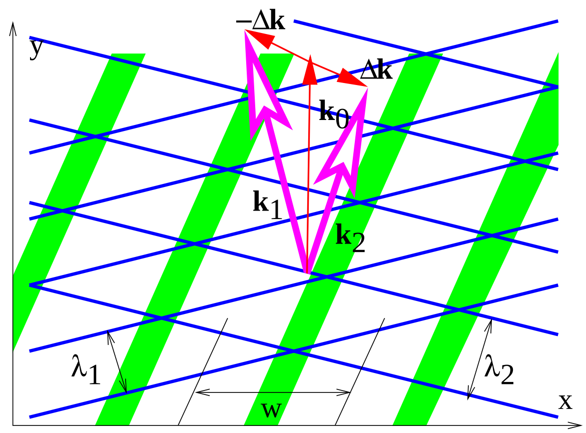

Figure 1.9: Wave fronts and wave vectors ($\mathbf{k}_1$ and $\mathbf{k}_2$) of two plane waves with different wavelengths oriented in different directions. The slanted bands show regions of constructive interference where wave fronts coincide. The slanted regions in between have destructive interference, and as previously, define the lateral limits of the beams produced by the superposition. The quantities $\mathbf{k}_0$ and $\Delta\mathbf{k}$ are also shown.

≈ 32 satisfies equation (1.18) with $|\Delta\mathbf{k}| = 0.1$. Each beam is made up of vertically propagating phase waves, with the crests and troughs indicated by the regions of white and black.

1.3.2 Two Waves of Differing Wavelength

In the third example of figure 1.6, the frequency of the wave depends only on the direction of the wave vector, independent of its magnitude, which is the reverse of the case for an isotropic dispersion relation. In this highly anisotropic case, different plane waves with the same frequency have wave vectors which point in the same direction, but have different lengths.

More generally, one might have waves for which the frequency depends on *both* the direction and magnitude of the wave vector. In this case, two different plane waves with the same frequency would typically have wave vectors which differed both in direction and magnitude. Such an example is illustrated in figures 1.9 and 1.10.

Figure 1.11 summarizes what we have learned about adding plane waves

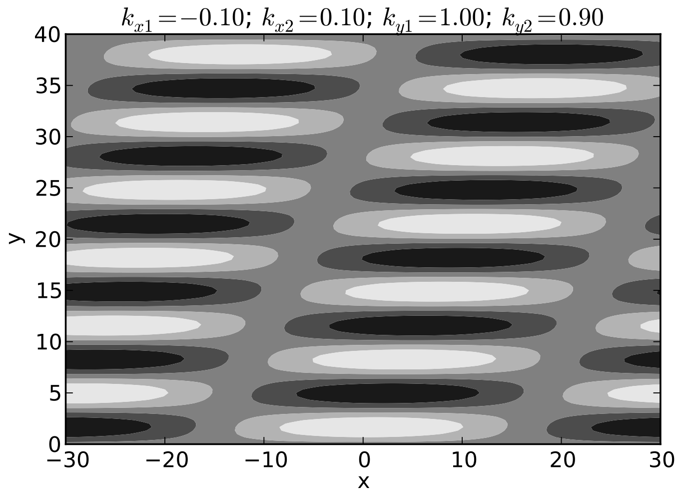

Figure 1.10: Example of beams produced by two plane waves with wave vectors differing in both direction and magnitude. The wave vectors of the two waves are $\mathbf{k}_1 = (-0.1, 1.0)$ and $\mathbf{k}_2 = (0.1, 0.9)$. Regions of positive displacement are lighter, while regions of negative displacement darker.

---

with the same frequency. In general, the beam orientation (and the lines of constructive interference) are not perpendicular to the wave fronts. This only occurs when the wave frequency is independent of wave vector direction.

1.3.3 Many Waves with the Same Wavelength

As with wave packets in one dimension, we can add together more than two waves to produce an isolated wave packet. We will confine our attention here to the case of an isotropic dispersion relation in which all the wave vectors for a given frequency are of the same length.

Figure 1.12 shows an example of this in which wave vectors of the same wavelength but different directions are added together. Defining α_i as the angle of the ith wave vector clockwise from the vertical, as illustrated in figure 1.12, we could write the superposition of these waves at time $t = 0$ as

$$h = \sum_i h_i \sin(k_{xi}x + k_{yi}y)$$

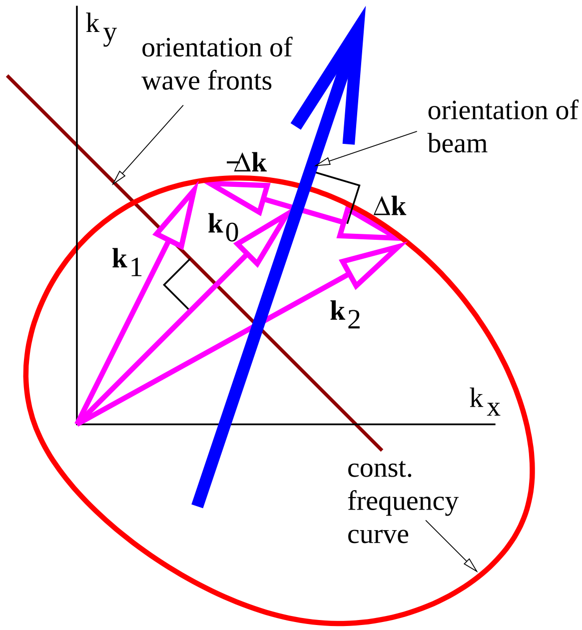

Figure 1.11: Illustration of factors entering the addition of two plane waves with the same frequency. The wave fronts are perpendicular to the vector average of the two wave vectors, $\mathbf{k}_0 = (\mathbf{k}_1 + \mathbf{k}_2)/2$, while the lines of constructive interference, which define the beam orientation, are oriented perpendicular to the difference between these two vectors, $\Delta\mathbf{k} = (\mathbf{k}_2 - \mathbf{k}_1)/2$.

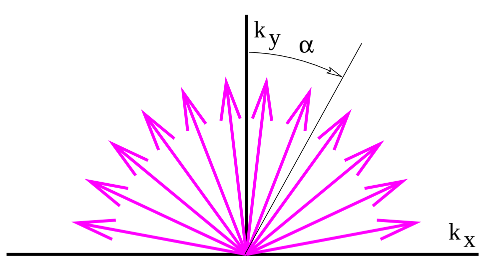

Figure 1.12: Illustration of wave vectors of plane waves which might be added together.

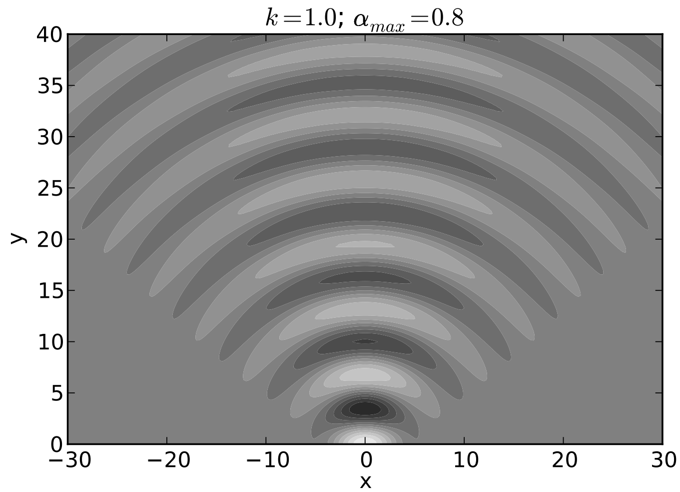

Figure 1.13: Plot of the displacement field $h(x, y)$ from equation (1.19) for $\alpha_{max} = 0.8$ and $k = 1$.

$$= \sum_i h_i \sin[kx \sin(\alpha_i) + ky \cos(\alpha_i)] \tag{1.19}$$

where we have assumed that $k_{xi} = k \sin(\alpha_i)$ and $k_{yi} = k \cos(\alpha_i)$. The parameter $k = |\mathbf{k}|$ is the magnitude of the wave vector and is the same for all the waves. Let us also assume in this example that the amplitude of each wave component decreases with increasing $|\alpha_i|$:

$$h_i = \exp[-(\alpha_i/\alpha_{max})^2]. \tag{1.20}$$

The exponential function decreases rapidly as its argument becomes more negative, and for practical purposes, only wave vectors with $|\alpha_i| \leq \alpha_{max}$ contribute significantly to the sum. We call α_{max} the *spreading angle*.

Figure 1.13 shows what $h(x, y)$ looks like when $\alpha_{max} = 0.8$ radians and $k = 1$. Notice that for $y = 0$ the wave amplitude is only large for a small region in the range $-4 < x < 4$. However, for $y > 0$ the wave spreads into a broad, semicircular pattern.

Figure 1.14 shows the computed pattern of $h(x, y)$ when the spreading angle $\alpha_{max} = 0.2$ radians. The wave amplitude is large for a much broader range of x at $y = 0$ in this case, roughly $-12 < x < 12$. On the other hand,

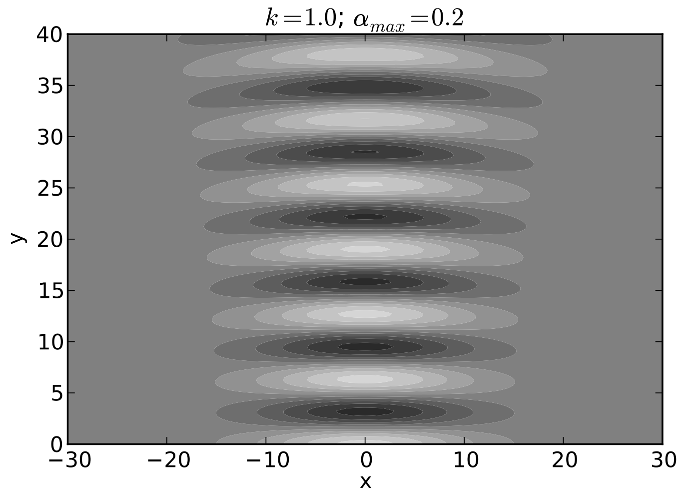

Figure 1.14: Plot of the displacement field $h(x, y)$ from equation (1.19) for $\alpha_{max} = 0.2$ and $k = 1$.

---

the subsequent spread of the wave is much smaller than in the case of figure 1.13.

We conclude that a superposition of plane waves with wave vectors spread narrowly about a central wave vector which points in the y direction (as in figure 1.14) produces a beam which is initially broad in x but for which the breadth increases only slightly with increasing y. However, a superposition of plane waves with wave vectors spread more broadly (as in figure 1.13) produces a beam which is initially narrow in x but which rapidly increases in width as y increases.

The relationship between the spreading angle α_{max} and the initial breadth of the beam is made more understandable by comparison with the results for the two-wave superposition discussed at the beginning of this section. As indicated by equation (1.18), large values of k_x, and hence α, are associated with small wave packet dimensions in the x direction and vice versa. The superposition of two waves doesn't capture the subsequent spread of the beam which occurs when many waves are superimposed, but it does lead to a rough quantitative relationship between α_{max} (which is just $\tan^{-1}(k_x/k_y)$ in the two wave case) and the initial breadth of the beam. If we invoke the small

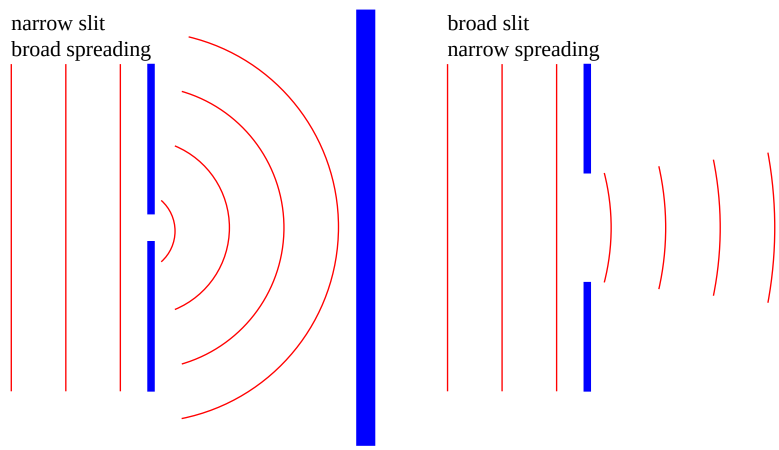

Figure 1.15: Schematic behavior when a plane wave impinges on a narrow slit and a broad slit.

---

angle approximation for $\alpha = \alpha_{max}$ so that $\alpha_{max} = \tan^{-1}(k_x/k_y) \approx k_x/k_y \approx k_x/k$, then $k_x \approx k\alpha_{max}$ and equation (1.18) can be written $w = \pi/k_x \approx \pi/(k\alpha_{max}) = \lambda/(2\alpha_{max})$. Thus, we can find the approximate spreading angle from the wavelength of the wave λ and the initial breadth of the beam w:

$$\alpha_{max} \approx \lambda/(2w) \quad \text{(single slit spreading angle).} \tag{1.21}$$

1.4 Diffraction Through a Single Slit

How does all of this apply to the passage of waves through a slit? Imagine a plane wave of wavelength λ impingent on a barrier with a slit. The barrier transforms the plane wave with infinite extent in the lateral direction into a beam with initial transverse dimensions equal to the width of the slit. The subsequent development of the beam is illustrated in figures 1.13 and 1.14, and schematically in figure 1.15. In particular, if the slit width is comparable to the wavelength, the beam spreads broadly as in figure 1.13. If the slit width is large compared to the wavelength, the beam doesn't spread as much, as figure 1.14 illustrates. Equation (1.21) gives us an approximate quantitative result for the spreading angle if w is interpreted as the width of the slit.

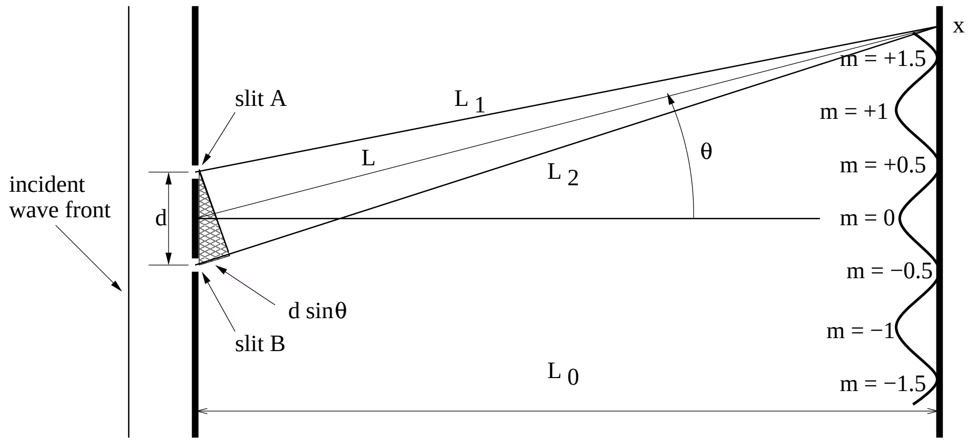

Figure 1.16: Definition sketch for the double slit. Light passing through slit B travels an extra distance to the screen equal to $d \sin\theta$ compared to light passing through slit A.

One use of the above equation is in determining the maximum angular resolution of optical instruments such as telescopes. The primary lens or mirror can be thought of as a rather large "slit". Light from a distant point source is essentially in the form of a plane wave when it arrives at the telescope. However, the light passed by the telescope is no longer a plane wave, but is a beam with a tendency to spread. The spreading angle α_{max} is given by equation (1.21), and the telescope cannot resolve objects with an angular separation less than α_{max}. Replacing w with the diameter of the lens or mirror in equation (1.21) thus yields the telescope's angular resolution. For instance, a moderate sized telescope with aperture 1 m observing red light with $\lambda \approx 6 \times 10^{-7}$ m has a maximum angular resolution of about 3×10^{-7} radians.

1.5 Two Slits

Let us now imagine a plane sine wave normally impingent on a screen with two narrow slits spaced by a distance d, as shown in figure 1.16. Since the slits are narrow relative to the wavelength of the wave impingent on them, the spreading angle of the beams is large and the diffraction pattern from each slit individually is a cylindrical wave spreading out in all directions, as illustrated in figure 1.13. The cylindrical waves from the two slits interfere, resulting in oscillations in wave intensity at the screen on the right side of figure 1.16.

Constructive interference occurs when the difference in the paths traveled by the two waves from their originating slits to the screen, $L_2 - L_1$, is an integer multiple m of the wavelength λ: $L_2 - L_1 = m\lambda$. If $L_0 \gg d$, the lines L_1 and L_2 are nearly parallel, which means that the narrow end of the dark triangle in figure 1.16 has an opening angle of θ. Thus, the path difference between the beams from the two slits is $L_2 - L_1 = d \sin\theta$. Substitution of this into the above equation shows that constructive interference occurs when

$$d \sin\theta = m\lambda, \quad m = 0, \pm 1, \pm 2, \ldots \qquad \text{(two slit interference)}. \tag{1.22}$$

Destructive interference occurs when m is an integer plus $1/2$. The integer m is called the *interference order* and is the number of wavelengths by which the two paths differ.

1.6 Diffraction Gratings

Since the angular spacing $\Delta\theta$ of interference peaks in the two slit case depends on the wavelength of the incident wave, the two slit system can be used as a crude device to distinguish between the wavelengths of different components of a non-sinusoidal wave impingent on the slits. However, if more slits are added, maintaining a uniform spacing d between slits, we obtain a more sophisticated device for distinguishing beam components. This is called a *diffraction grating*.

Figures 1.17-1.19 show the intensity of the diffraction pattern as a function of position x on the display screen (see figure 1.16) for gratings with 2, 4, and 16 slits respectively, with the same slit spacing. Notice how the interference peaks remain in the same place but increase in sharpness as the number of slits increases.

The width of the peaks is actually related to the overall width of the grating, $w = nd$, where n is the number of slits. Thinking of this width as the dimension of large single slit, the single slit equation, $\alpha_{max} = \lambda/(2w)$, tells us the angular width of the peaks.[2]

[2] Note that for this type of grating to work, the width of the grating has to be much less

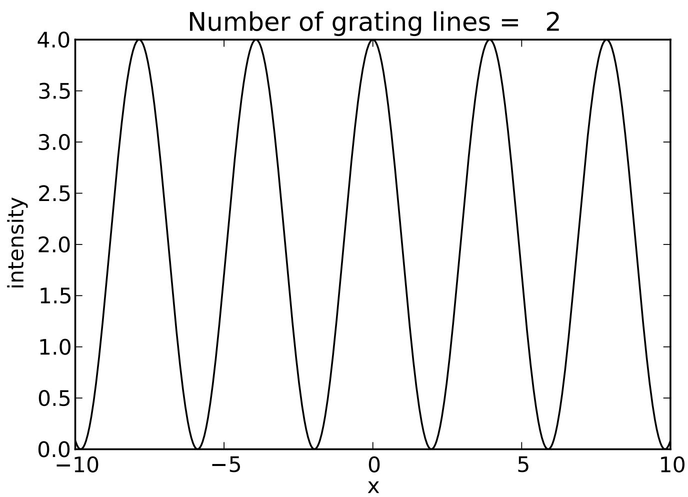

Figure 1.17: Intensity of interference pattern from a diffraction grating with 2 slits on the screen in figure 1.16. The position x on the screen is proportional to the angle θ in the small angle approximation.

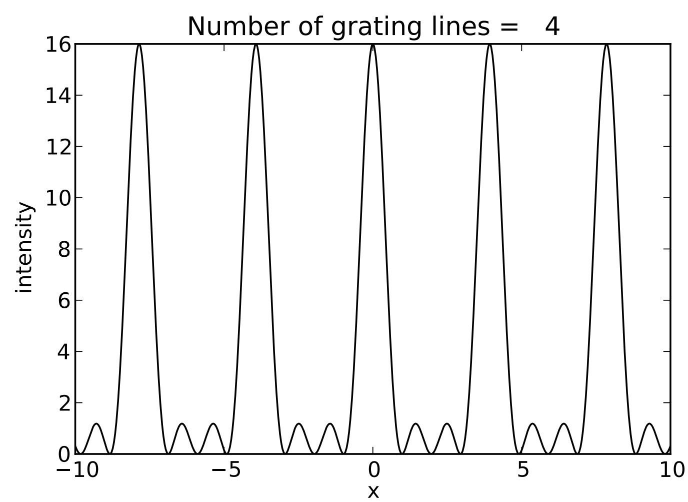

Figure 1.18: Intensity of interference pattern from a diffraction grating with 4 slits.

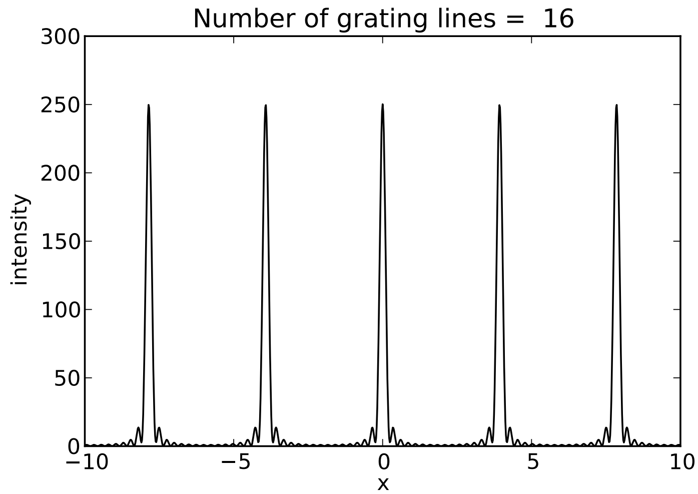

Figure 1.19: Intensity of interference pattern from a diffraction grating with 16 slits.

Whereas the angular width of the interference peaks is governed by the single slit equation, their angular positions are governed by the two slit equation. Let us assume for simplicity that $|\theta| \ll 1$ so that we can make the small angle approximation to the two slit equation, $m\lambda = d \sin\theta \approx d\theta$, and ask the following question: How different do two wavelengths differing by $\Delta\lambda$ have to be in order that the interference peaks from the two waves not overlap? In order for the peaks to be distinguishable, they should be separated in θ by an angle $\Delta\theta = m\Delta\lambda/d$, which is greater than the angular width of each peak, α_{max}:

$$\Delta\theta > \alpha_{max}. \tag{1.23}$$

Substituting in the above expressions for $\Delta\theta$ and α_{max} and solving for $\Delta\lambda$, we get $\Delta\lambda > \lambda/(2mn)$, where λ is the average of the two wavelengths and $n = w/d$ is the number of slits in the diffraction grating. Thus, the fractional difference between wavelengths which can be distinguished by a diffraction grating depends solely on the interference order m and the number of slits n

---

than the width of the interference peaks on the display screen. This is a severe limitation. Real diffraction grating spectrometers use a lens to focus the diffraction pattern on the screen, and are not subject to this limitation.

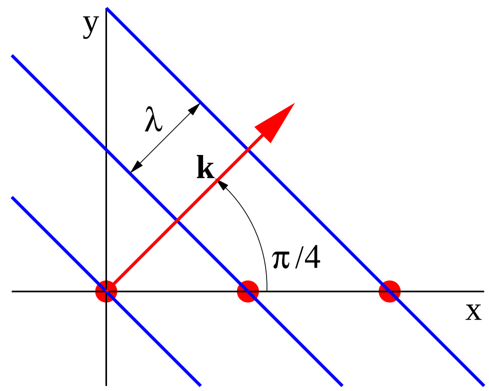

Figure 1.20: Sketch of wave moving at 45° to the x-axis.

---

in the grating:

$$\frac{\Delta\lambda}{\lambda} > \frac{1}{2mn}. \tag{1.24}$$

1.7 Problems

1. Point A is at the origin. Point B is 3 m distant from A at 30° counterclockwise from the x axis. Point C is 2 m from point A at 100° counterclockwise from the x axis.

   (a) Obtain the Cartesian components of the vector $\mathbf{D}_1$ which goes from A to B and the vector $\mathbf{D}_2$ which goes from A to C.

   (b) Find the Cartesian components of the vector $\mathbf{D}_3$ which goes from B to C.

   (c) Find the direction and magnitude of $\mathbf{D}_3$.

2. For the vectors in the previous problem, find $\mathbf{D}_1 \cdot \mathbf{D}_2$ using both the cosine form of the dot product and the Cartesian form. Check to see if the two answers are the same.

3. Show graphically or otherwise that $|\mathbf{A} + \mathbf{B}| \neq |\mathbf{A}| + |\mathbf{B}|$ except when the vectors $\mathbf{A}$ and $\mathbf{B}$ are parallel.

4. A wave in the x-y plane is defined by $h = h_0 \sin(\mathbf{k} \cdot \mathbf{x})$ where $\mathbf{k} = (1, 2)$ cm^{-1}.
   (a) On a piece of graph paper draw x and y axes and then plot a line passing through the origin which is parallel to the vector $\mathbf{k}$.
   (b) On the same graph plot the line defined by $\mathbf{k} \cdot \mathbf{x} = k_x x + k_y y = 0$, $\mathbf{k} \cdot \mathbf{x} = \pi$, and $\mathbf{k} \cdot \mathbf{x} = 2\pi$. Check to see if these lines are perpendicular to $\mathbf{k}$.
5. A plane wave in two dimensions in the $x-y$ plane moves in the direction 45° counterclockwise from the x-axis as shown in figure 1.20. Determine how fast the intersection between a wave front and the x-axis moves to the right in terms of the phase speed c of the wave. Hint: What is the distance between wave fronts along the x-axis compared to the wavelength?
6. Two deep plane ocean waves with the same frequency ω are moving approximately to the east. However, one wave is oriented a small angle β north of east and the other is oriented β south of east.
   (a) Determine the orientation of lines of constructive interference between these two waves.
   (b) Determine the spacing between lines of constructive interference.
7. An example of waves with a dispersion relation in which the frequency is a function of both wave vector magnitude and direction is shown graphically in figure 1.21.
   (a) What is the phase speed of the waves for each of the three wave vectors? Hint: You may wish to obtain the length of each wave vector graphically.
   (b) For each of the wave vectors, what is the orientation of the wave fronts?
   (c) For each of the illustrated wave vectors, sketch two other wave vectors whose average value is approximately the illustrated vector, and whose tips lie on the same frequency contour line. Determine the orientation of lines of constructive interference produced by the superimposing pairs of plane waves for which each of the vector pairs are the wave vectors.

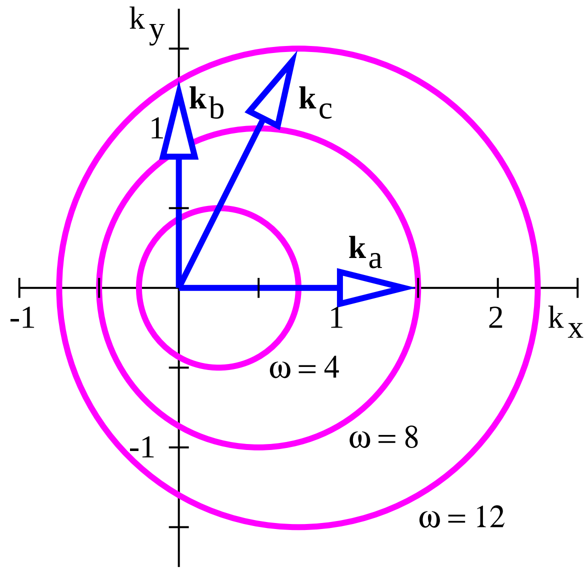

Figure 1.21: Graphical representation of the dispersion relation for shallow water waves in a river flowing in the x direction. Units of frequency are hertz, units of wavenumber are inverse meters.

---

8. Two gravity waves have the same frequency, but slightly different wavelengths.

   (a) If one wave has an orientation angle $\theta = \pi/4$ radians, what is the orientation angle of the other? (See figure 1.6.)

   (b) Determine the orientation of lines of constructive interference between these two waves.

9. A plane wave impinges on a single slit, spreading out a half-angle α after the slit. If the whole apparatus is submerged in a liquid with index of refraction $n = 1.5$, how does the spreading angle of the light change? (Hint: Recall that the index of refraction in a transparent medium is the ratio of the speed of light in a vacuum to the speed in the medium. Furthermore, when light goes from a vacuum to a transparent medium, the light frequency doesn't change. Therefore, how does the wavelength of the light change?)

10. Determine the diameter of the telescope needed to resolve a planet 2×10^8 km from a star which is 6 light years from the earth. (Assume blue light which has a wavelength $\lambda \approx 4 \times 10^{-7}$ m $= 400$ nm. Also, don't worry about the great difference in brightness between the two for the purposes of this problem.)

11. A laser beam from a laser on the earth is bounced back to the earth by a corner reflector on the moon.

    (a) Engineers find that the returned signal is stronger if the laser beam is initially spread out by the beam expander shown in figure 1.22. Explain why this is so.

    (b) The beam has a diameter of 1 m leaving the earth. How broad is it when it reaches the moon, which is 4×10^5 km away? Assume the wavelength of the light to be 5×10^{-7} m.

    (c) How broad would the laser beam be at the moon if it weren't initially passed through the beam expander? Assume its initial diameter to be 1 cm.

12. Suppose that a plane wave impinges on two slits in a barrier at an angle, such that the phase of the wave at one slit lags the phase at the other slit by half a wavelength. How does the resulting interference pattern change from the case in which there is no lag?

13. Suppose that a thin piece of glass of index of refraction $n = 1.33$ is placed in front of one slit of a two slit diffraction setup.

    (a) How thick does the glass have to be to slow down the incoming wave so that it lags the wave going through the other slit by a phase difference of π? Take the wavelength of the light to be $\lambda = 6 \times 10^{-7}$ m.

    (b) For the above situation, describe qualitatively how the diffraction pattern changes from the case in which there is no glass in front of one of the slits. Explain your results.

14. A light source produces two wavelengths, $\lambda_1 = 400$ nm (blue) and $\lambda_2 = 600$ nm (red).

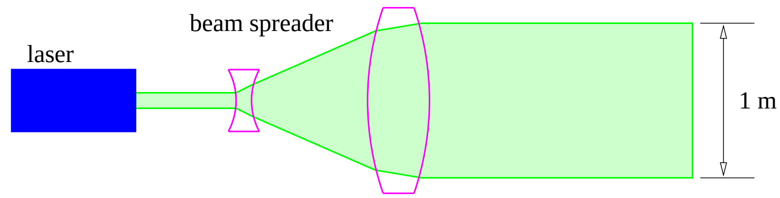

Figure 1.22: Sketch of a beam expander for a laser.

(a) Qualitatively sketch the two slit diffraction pattern from this source. Sketch the pattern for each wavelength separately.

(b) Qualitatively sketch the 16 slit diffraction pattern from this source, where the slit spacing is the same as in the two slit case.

15. A light source produces two wavelengths, $\lambda_1 = 631$ nm and $\lambda_2 = 635$ nm. What is the minimum number of slits needed in a grating spectrometer to resolve the two wavelengths? (Assume that you are looking at the first order diffraction peak.) Sketch the diffraction peak from each wavelength and indicate how narrow the peaks must be to resolve them.