THE PHYSICS OF WAVES

HOWARD GEORGI

Harvard University

Originally published by

PRENTICE HALL

Englewood Cliffs, New Jersey 07632

Printed in the United States of America
10 9 8 7 6 5 4 3

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Contents

Preface

Waves are everywhere. Everything waves. There are familiar, everyday sorts of waves in water, ropes and springs. There are less visible but equally pervasive sound waves and electromagnetic waves. Even more important, though only touched on in this book, is the wave phenomenon of quantum mechanics, built into the fabric of our space and time. How can it make sense to use the same word — "wave" — for all these disparate phenomena? What is it that they all have in common?

The superficial answer lies in the mathematics of wave phenomena. Periodic behavior of any kind, one might argue, leads to similar mathematics. Perhaps this is the unifying principle.

In this book, I introduce you to a deeper, physical answer to the questions. The mathematics of waves is important, to be sure. Indeed, I devote much of the book to the mathematical formalism in which wave phenomena can be described most insightfully. But I use the mathematics only as a tool to formulate the underlying physical principles that tie together many different kinds of wave phenomena. There are three: linearity, translation invariance and local interactions. You will learn in detail what each of these means in the chapters to come. When all three are present, wave phenomena always occur. Furthermore, as you will see, these principles are a great practical help both in understanding particular wave phenomena and in solving problems. I hope to convert you to a way of thinking about waves that will permanently change the way you look at the world.

The organization of the book is designed to illustrate how wave phenomena arise in any system of coupled linear oscillators with translation invariance and local interactions. We begin with the single harmonic oscillator and work our way through standing wave normal modes in more and more interesting systems. Traveling waves appear only after a thorough exploration of one-dimensional standing waves. I hope to emphasize that the physics of standing waves is the same. Only the boundary conditions are different. When we finally get to traveling waves, well into the book, we will be able to get to interesting properties very quickly.

For similar reasons, the discussion of two- and three-dimensional waves occurs late in the book, after you have been exposed to all the tools required to deal with one-dimensional waves. This allows us at least to set up the problems of interference and diffraction in a simple way, and to solve the problems in some simple cases.

Waves move. Their motion is an integral part of their being. Illustrations on a printed page cannot do justice to this motion. For that reason, this book comes with moving illustrations, in the form of computer animations of various wave phenomena. These supplementary programs are an important part of the book. Looking at them and interacting with them, you will get a much more concrete understanding of wave phenomena than can be obtained from a book alone. I discuss the simple programs that produce the animations in more detail in Appendix A. Also in this appendix are instructions on the use of the supplementary program disk.

The subsections that are illustrated with computer animations are clearly labeled in the text by and the number of the program. I hope you will read these parts of the book while sitting at your computer screens.

The sections and problems marked with a $\ast$ can be skipped by instructors who wish to keep the mathematical level as low as possible.

Two other textbooks on the subject, **Waves**, by Crawford and **Optics** by Hecht, influenced me in writing this book. The strength of Crawford's book is the home experiments. These experiments are very useful additions to any course on wave phenomena. Hecht's book is an encyclopedic treatment of optics. In my own book, I try to steer a middle course between these two, with a better treatment of general wave phenomena than Hecht and a more appropriate mathematical level than Crawford. I believe that my text has many of the advantages of both books, but students may wish to use them as supplementary texts.

While the examples of waves phenomena that we discuss in this book will be chosen (mostly) from familiar waves, we also will be developing the mathematics of waves in such a way that it can be directly applied to quantum mechanics. Thus, while learning about waves in ropes and air and electromagnetic fields, you will be preparing to apply the same techniques to the study of the quantum mechanical world.

I am grateful to many people for their help in converting this material into a textbook. Adam Falk and David Griffiths made many detailed and invaluable suggestions for improvements in the presentation. Melissa Franklin, Geoff Georgi, Kevin Jones and Mark Heald, also had extremely useful suggestions. I am indebted to Nicholas Romanelli for copyediting and to Ray Henderson for orchestrating all of it. Finally, thanks go to the hundreds of students who took the waves course at Harvard in the last fifteen years. This book is as much the product of their hard work and enthusiasm, as my own.

Howard Georgi
Cambridge, MA

Preface to the online edition

As I prepared to teach the sophomore waves course at Harvard again after a break of over 10 years, I realized that I had accumulated a list of many things that I wanted to change in my waves text. And while I was very grateful to Prentice-Hall for all the help they gave me in turning my notes into a textbook, I felt that it was time to liberate the book from its paper straightjacket, and try to turn it into something more continuously evolving. Thus I asked Prentice-Hall to release the rights back to me, and they graciously agreed. My intention is to leave the textbook up on the web for students and teachers to use as they see fit, so long as they give me credit and do not use it for commercial purposes. I hope that readers will send suggestions for improvements. I will not have much time to think about these and implement them. But if I do incorporate something in the online version as the result of a suggestion, I will acknowledge the suggestion in a list of changes on my web page.

I have eliminated the table of contents from the online version and substituted hyperref hypertext instead. I hope that this will encourage people to use the text online and save trees.

Howard Georgi
Cambridge, MA
December, 2006

Chapter 1

Harmonic Oscillation

Oscillators are the basic building blocks of waves. We begin by discussing the harmonic oscillator. We will identify the general principles that make the harmonic oscillator so special and important. To make use of these principles, we must introduce the mathematical device of complex numbers. But the advantage of introducing this mathematics is that we can understand the solution to the harmonic oscillator problem in a new way. We show that the properties of linearity and time translation invariance lead to solutions that are complex exponential functions of time.

Preview

In this chapter, we discuss harmonic oscillation in systems with only one degree of freedom.

1. We begin with a review of the simple harmonic oscillator, noting that the equation of motion of a free oscillator is linear and invariant under time translation;

2. We discuss linearity in more detail, arguing that it is the generic situation for small oscillations about a point of stable equilibrium;

3. We discuss time translation invariance of the harmonic oscillator, and the connection between harmonic oscillation and uniform circular motion;

4. We introduce complex numbers, and discuss their arithmetic;

5. Using complex numbers, we find solutions to the equation of motion for the harmonic oscillator that behave as simply as possible under time translations. We call these solutions "irreducible." We show that they are actually complex exponentials.

6. We discuss an LC circuit and draw an analogy between it and a system of a mass and springs.

7. We discuss units.

8. We give one simple example of a nonlinear oscillator.

1.1 The Harmonic Oscillator

When you studied mechanics, you probably learned about the harmonic oscillator. We will begin our study of wave phenomena by reviewing this simple but important physical system. Consider a block with mass, m, free to slide on a frictionless air-track, but attached to a light[1] Hooke's law spring with its other end attached to a fixed wall. A cartoon representation of this physical system is shown in figure 1.1.

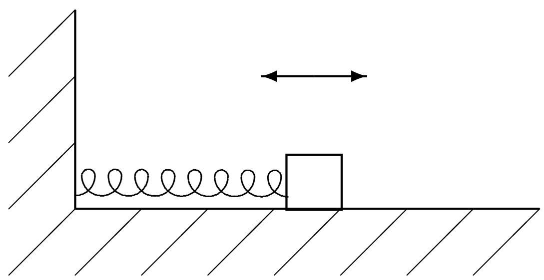


Figure 1.1: A mass on a spring.

This system has only one relevant degree of freedom. In general, the number of **degrees of freedom** of a system is the number of coordinates that must be specified in order to determine the configuration completely. In this case, because the spring is light, we can assume that it is uniformly stretched from the fixed wall to the block. Then the only important coordinate is the position of the block.

In this situation, gravity plays no role in the motion of the block. The gravitational force is canceled by a vertical force from the air track. The only relevant force that acts on the block comes from the stretching or compression of the spring. When the spring is relaxed, there is no force on the block and the system is in equilibrium. Hooke's law tells us that the force from the spring is given by a negative constant, $-K$, times the displacement of the block from its equilibrium position. Thus if the position of the block at some time is x and its equilibrium position is x_0, then the force on the block at that moment is

$$F = -K(x - x_0)\,. \tag{1.1}$$

[1] "Light" here means that the mass of the spring is small enough to be ignored in the analysis of the motion of the block. We will explain more precisely what this means in chapter 7 when we discuss waves in a massive spring.

The constant, K, is called the "spring constant." It has units of force per unit distance, or MT^{-2} in terms of M (the unit of mass), L (the unit of length) and T (the unit of time). We can always choose to measure the position, x, of the block with our origin at the equilibrium position. If we do this, then $x_0 = 0$ in (1.1) and the force on the block takes the simpler form

$$F = -Kx\,. \tag{1.2}$$

Harmonic oscillation results from the interplay between the Hooke's law force and Newton's law, $F = ma$. Let $x(t)$ be the displacement of the block as a function of time, t. Then Newton's law implies

$$m\,\frac{d^2}{dt^2}x(t) = -K\,x(t)\,. \tag{1.3}$$

An equation of this form, involving not only the function $x(t)$, but also its derivatives is called a "differential equation." The differential equation, (1.3), is the "equation of motion" for the system of figure 1.1. Because the system has only one degree of freedom, there is only one equation of motion. In general, there must be one equation of motion for each independent coordinate required to specify the configuration of the system.

The most general solution to the differential equation of motion, (1.3), is a sum of a constant times $\cos\omega t$ plus a constant times $\sin\omega t$,

$$x(t) = a\cos\omega t + b\sin\omega t\,, \tag{1.4}$$

where

$$\omega \equiv \sqrt{\frac{K}{m}} \tag{1.5}$$

is a constant with units of T^{-1} called the "angular frequency." The angular frequency will be a very important quantity in our study of wave phenomena. We will almost always denote it by the lower case Greek letter, ω (omega).

Because the equation involves a second time derivative but no higher derivatives, the most general solution involves two constants. This is just what we expect from the physics, because we can get a different solution for each value of the position and velocity of the block at the starting time. Generally, we will think about determining the solution in terms of the position and velocity of the block when we first get the motion started, at a time that we conventionally take to be $t = 0$. For this reason, the process of determining the solution in terms of the position and velocity at a given time is called the "initial value problem." The values of position and velocity at $t = 0$ are called initial conditions. For example, we can write the **most general solution**, (1.4), in terms of $x(0)$ and $x'(0)$, the displacement and velocity of the block at time $t = 0$. Setting $t = 0$ in (1.4) gives $a = x(0)$. Differentiating and then setting $t = 0$ gives $b = \omega\,x'(0)$. Thus

$$x(t) = x(0)\cos\omega t + \frac{1}{\omega}x'(0)\sin\omega t\,. \tag{1.6}$$

For example, suppose that the block has a mass of 1 kilogram and that the spring is 0.5 meters long[2] with a spring constant K of 100 newtons per meter. To get a sense of what this spring constant means, consider hanging the spring vertically (see problem (1.1)). The gravitational force on the block is

$$mg \approx 9.8 \text{ newtons}\,. \tag{1.7}$$

In equilibrium, the gravitational force cancels the force from the spring, thus the spring is stretched by

$$\frac{mg}{K} \approx 0.098 \text{ meters} = 9.8 \text{ centimeters}\,. \tag{1.8}$$

For this mass and spring constant, the angular frequency, ω, of the system in figure 1.1 is

$$\omega = \sqrt{\frac{K}{M}} = \sqrt{\frac{100\,\mathrm{N/m}}{1\,\mathrm{kg}}} = 10\frac{1}{\mathrm{s}}\,. \tag{1.9}$$

If, for example, the block is displaced by 0.01 m (1 cm) from its equilibrium position and released from rest at time, $t = 0$, the position at any later time t is given (in meters) by

$$x(t) = 0.01 \times \cos 10t\,. \tag{1.10}$$

The velocity (in meters per second) is

$$x'(t) = -0.1 \times \sin 10t\,. \tag{1.11}$$

The motion is periodic, in the sense that the system oscillates — it repeats the same motion over and over again indefinitely. After a time

$$\tau = \frac{2\pi}{\omega} \approx 0.628 \text{ s} \tag{1.12}$$

the system returns exactly to where it was at $t = 0$, with the block instantaneously at rest with displacement 0.01 meter. The time, τ (Greek letter tau) is called the "period" of the oscillation. However, the solution, (1.6), is more than just periodic. It is "simple harmonic" motion, which means that only a single frequency appears in the motion.

The angular frequency, ω, is the inverse of the time required for the phase of the wave to change by one radian. The "frequency", usually denoted by the Greek letter, ν (nu), is the inverse of the time required for the phase to change by one complete cycle, or 2π radians, and thus get back to its original state. The frequency is measured in hertz, or cycles/second. Thus the angular frequency is **larger** than the frequency by a factor of 2π,

$$\omega \text{ (in radians/second) } = 2\pi \text{ (radians/cycle) } \cdot \nu \text{ (cycles/second)}\,. \tag{1.13}$$

[2]The length of the spring plays no role in the equations below, but we include it to allow you to build a mental picture of the physical system.

The frequency, ν, is the inverse of the period, τ, of (1.12),

$$\nu = \frac{1}{\tau} \,. \tag{1.14}$$

Simple harmonic motion like (1.6) occurs in a very wide variety of physical systems. The question with which we will start our study of wave phenomena is the following: **Why do solutions of the form of (1.6) appear so ubiquitously in physics? What do harmonically oscillating systems have in common?** Of course, the mathematical answer to this question is that all of these systems have equations of motion of essentially the same form as (1.3). We will find a deeper and more physical answer that we will then be able to generalize to more complicated systems. The key features that all these systems have in common with the mass on the spring are (at least approximate) linearity and time translation invariance of the equations of motion. It is these two features that determine oscillatory behavior in systems from springs to inductors and capacitors.

Each of these two properties is interesting on its own, but together, they are much more powerful. They almost completely determine the form of the solutions. We will see that if the system is linear and time translation invariant, we can always write its motion as a sum of simple motions in which the time dependence is either harmonic oscillation or exponential decay (or growth).

1.2 Small Oscillations and Linearity

A system with one degree of freedom is **linear** if its equation of motion is a linear function of the coordinate, x, that specifies the system's configuration. In other words, the equation of motion must be a sum of terms each of which contains at most one power of x. The equation of motion involves a second derivative, but no higher derivatives, so a linear equation of motion has the general form:

$$\alpha \frac{d^2}{dt^2} x(t) + \beta \frac{d}{dt} x(t) + \gamma \, x(t) = f(t) \,. \tag{1.15}$$

If all of the terms involve exactly one power of x, the equation of motion is "homogeneous." Equation (1.15) is not homogeneous because of the term on the right-hand side. The "in-homogeneous" term, $f(t)$, represents an external force. The corresponding homogeneous equation would look like this:

$$\alpha \frac{d^2}{dt^2} x(t) + \beta \frac{d}{dt} x(t) + \gamma \, x(t) = 0 \,. \tag{1.16}$$

In general, α, β and γ as well as f could be functions of t. However, that would break the time translation invariance that we will discuss in more detail below and make the system much more complicated. We will almost always assume that α, β and γ are constants. The equation of motion for the mass on a spring, (1.3), is of this general form, but with β and f equal to zero. As we will see in chapter 2, we can include the effect of frictional forces by allowing nonzero β, and the effect of external forces by allowing nonzero f.

The linearity of the equation of motion, (1.15), implies that if $x_1(t)$ is a solution for external force $f_1(t)$,

$$\alpha \frac{d^2}{dt^2} x_1(t) + \beta \frac{d}{dt} x_1(t) + \gamma x_1(t) = f_1(t) \,, \tag{1.17}$$

and $x_2(t)$ is a solution for external force $f_2(t)$,

$$\alpha \frac{d^2}{dt^2} x_2(t) + \beta \frac{d}{dt} x_2(t) + \gamma x_2(t) = f_2(t) \,, \tag{1.18}$$

then the sum,

$$x_{12}(t) = A\, x_1(t) + B\, x_2(t) \,, \tag{1.19}$$

for constants A and B is a solution for external force $Af_1 + Bf_2$,

$$\alpha \frac{d^2}{dt^2} x_{12}(t) + \beta \frac{d}{dt} x_{12}(t) + \gamma x_{12}(t) = Af_1(t) + Bf_2(t) \,. \tag{1.20}$$

The sum $x_{12}(t)$ is called a "linear combination" of the two solutions, $x_1(t)$ and $x_2(t)$. In the case of "free" motion, which means motion with no external force, if $x_1(t)$ and $x_2(t)$ are solutions, then the sum, $A\, x_1(t) + B\, x_2(t)$ is also a solution.

The most general solution to any of these equations involves two constants that must be fixed by the initial conditions, for example, the initial position and velocity of the particle, as in (1.6). It follows from (1.20) that we can always write the most general solution for any external force, $f(t)$, as a sum of the "general solution" to the homogeneous equation, (1.16), and any "particular" solution to (1.15).

No system is exactly linear. "Linearity" is never exactly "true." Nevertheless, the idea of linearity is extremely important, because it is a useful approximation in a very large number of systems, for a very good physical reason. In almost any system in which the properties are smooth functions of the positions of the parts, the small displacements from equilibrium produce approximately linear restoring forces. The difference between something that is "true" and something that is a useful approximation is the essential difference between physics and mathematics. **In the real world, the questions are much too interesting to have answers that are exact. If you can understand the answer in a well-defined approximation, you have learned something important.**

To see the generic nature of linearity, consider a particle moving on the x-axis with potential energy, $V(x)$. The force on the particle at the point, x, is minus the derivative of the potential energy,

$$F = -\frac{d}{dx} V(x) \,. \tag{1.21}$$

A force that can be derived from a potential energy in this way is called a "conservative" force.

At a point of equilibrium, x_0, the force vanishes, and therefore the derivative of the potential energy vanishes:

$$F = -\frac{d}{dx} V(x)|_{x=x_0} = -V'(x_0) = 0\,. \tag{1.22}$$

We can describe the small oscillations of the system about equilibrium most simply if we redefine the origin so that $x_0 = 0$. Then the displacement from equilibrium is the coordinate x. We can expand the force in a Taylor series:

$$F(x) = -V'(x) = -V'(0) - x\,V''(0) - \frac{1}{2}x^2\,V'''(0) + \cdots \tag{1.23}$$

The first term in (1.23) vanishes because this system is in equilibrium at $x = 0$, from (1.22). The second term looks like Hooke's law with

$$K = V''(0)\,. \tag{1.24}$$

The equilibrium is stable if the second derivative of the potential energy is positive, so that $x = 0$ is a local minimum of the potential energy.

The important point is that for sufficiently small x, the third term in (1.23), and all subsequent terms will be much smaller than the second. The third term is negligible if

$$|x\,V'''(0)| \ll V''(0)\,. \tag{1.25}$$

Typically, each extra derivative will bring with it a factor of $1/L$, where L is the distance over which the potential energy changes by a large fraction. Then (1.25) becomes

$$x \ll L\,. \tag{1.26}$$

There are only two ways that a force derived from a potential energy can fail to be approximately linear for sufficiently small oscillations about stable equilibrium:

1. If the potential is not smooth so that the first or second derivative of the potential is not well defined at the equilibrium point, then we cannot do a Taylor expansion and the argument of (1.23) does not work. We will give an example of this kind at the end of this chapter.

2. Even if the derivatives exist at the equilibrium point, $x = 0$, it may happen that $V''(0) = 0$. In this case, to have a stable equilibrium, we must have $V'''(0) = 0$ as well, otherwise a small displacement in one direction or the other would grow with time. Then the next term in the Taylor expansion dominates at small x, giving a force proportional to x^3.

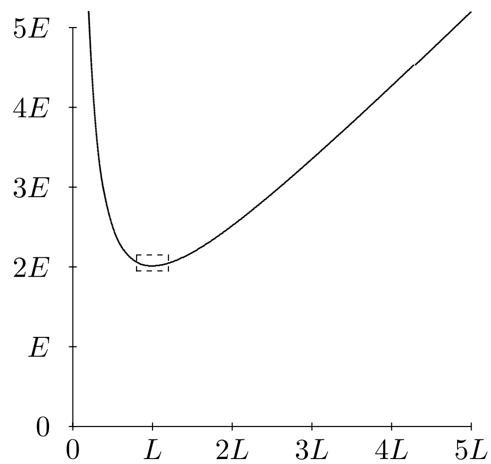


Figure 1.2: The potential energy of (1.27).

Both of these exceptional cases are very rare in nature. Usually, the potential energy is a smooth function of the displacement and there is no reason for $V''(0)$ to vanish. The generic situation is that small oscillations about stable equilibrium are linear.

An example may be helpful. Almost any potential energy function with a point of stable equilibrium will do, so long as it is smooth. For example, consider the following potential energy

$$V(x) = E\left(\frac{L}{x} + \frac{x}{L}\right) . \tag{1.27}$$

This is shown in figure 1.2. The minimum (at least for positive x) occurs at $x = L$, so we first redefine $x = X + L$, so that

$$V(X) = E\left(\frac{L}{X+L} + \frac{X+L}{L}\right) . \tag{1.28}$$

The corresponding force is

$$F(X) = E\left(\frac{L}{(X+L)^2} - \frac{1}{L}\right) . \tag{1.29}$$

we can look near $X = 0$ and expand in a Taylor series:

$$F(X) = -2\frac{E}{L}\left(\frac{X}{L}\right) + 3\frac{E}{L}\left(\frac{X}{L}\right)^2 + \cdots \tag{1.30}$$

Now, the ratio of the first nonlinear term to the linear term is

$$\frac{3X}{2L} , \tag{1.31}$$

which is small if $X \ll L$.

In other words, the closer you are to the equilibrium point, the closer the actual potential energy is to the parabola that we would expect from the potential energy for a linear, Hooke's law force. You can see this graphically by blowing up a small region around the equilibrium point. In figure 1.3, the dotted rectangle in figure 1.2 has been blown up into a square. Note that it looks much more like a parabola than figure 1.3. If we repeated the procedure and again expanded a small region about the equilibrium point, you would not be able to detect the cubic term by eye.

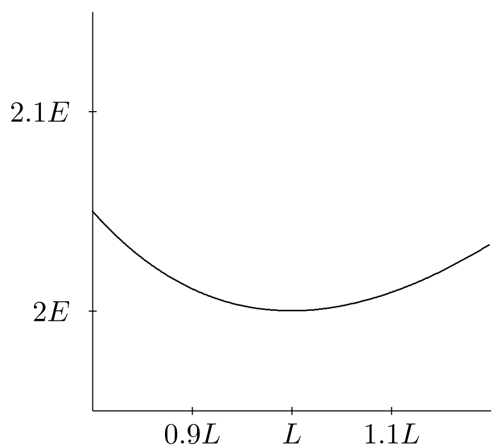


Figure 1.3: The small dashed rectangle in figure 1.2 expanded.

Often, the linear approximation is even better, because the term of order x^2 vanishes by symmetry. For example, when the system is symmetrical about $x = 0$, so that $V(x) = V(-x)$, the order x^3 term (and all x^n for n odd) in the potential energy vanishes, and then there is no order x^2 term in the force.

For a typical spring, linearity (Hooke's law) is an excellent approximation for small displacements. However, there are always nonlinear terms that become important if the displacements are large enough. Usually, in this book we will simply stick to small oscillations and assume that our systems are linear. However, you should not conclude that the subject of nonlinear systems is not interesting. In fact, it is a very active area of current research in physics.

1.3 Time Translation Invariance

1.3.1 Uniform Circular Motion

💻 1-1

When α, β and γ in (1.15) do not depend on the time, t, and in the absence of an external force, that is for free motion, time enters in (1.15) only through derivatives. Then the equation of motion has the form.

$$\alpha \frac{d^2}{dt^2} x(t) + \beta \frac{d}{dt} x(t) + \gamma x(t) = 0 . \tag{1.32}$$

The equation of motion for the undamped harmonic oscillator, (1.3), has this form with $\alpha = m$, $\beta = 0$ and $\gamma = K$. Solutions to (1.32) have the property that

$$\textbf{If } x(t) \textbf{ is a solution, } x(t+a) \textbf{ will be a solution also.} \tag{1.33}$$

Mathematically, this is true because the operations of differentiation with respect to time and replacing $t \to t + a$ can be done in either order because of the chain rule

$$\frac{d}{dt}x(t+a) = \left[\frac{d}{dt}(t+a)\right]\left[\frac{d}{dt'}x(t')\right]_{t'=t+a} = \left[\frac{d}{dt'}x(t')\right]_{t'=t+a} . \tag{1.34}$$

The physical reason for (1.33) is that we can change the initial setting on our clock and the physics will look the same. The solution $x(t + a)$ can be obtained from the solution $x(t)$ by changing the clock setting by a. The time label has been “translated” by a. We will refer to the property, (1.33), as **time translation invariance.**

Most physical systems that you can think of are time translation invariant in the absence of an external force. To get an oscillator without time translation invariance, you would have to do something rather bizarre, such as somehow making the spring constant depend on time.

For the free motion of the harmonic oscillator, although the equation of motion is certainly time translation invariant, the manifestation of time translation invariance on the solution, (1.6) is not as simple as it could be. The two parts of the solution, one proportional to $\cos \omega t$ and the other to $\sin \omega t$, get mixed up when the clock is reset. For example,

$$\cos\left[\omega(t+a)\right] = \cos \omega a \, \cos \omega t - \sin \omega a \, \sin \omega t . \tag{1.35}$$

It will be very useful to find another way of writing the solution that behaves more simply under resetting of the clocks. To do this, we will have to work with complex numbers.

To motivate the introduction of complex numbers, we will begin by exhibiting the relation between simple harmonic motion and uniform circular motion. Consider uniform circular motion in the x-y plane around a circle centered at the origin, $x = y = 0$, with radius R and with clockwise velocity $v = R\omega$. The x and y coordinates of the motion are

$$x(t) = R\cos(\omega t - \phi) , \quad y(t) = -R\sin(\omega t - \phi) , \tag{1.36}$$

where ϕ is the counterclockwise angle in radians of the position at $t = 0$ from the positive x axis. The $x(t)$ in (1.36) is identical to the $x(t)$ in (1.6) with

$$x(0) = R\cos\phi , \quad x'(0) = \omega R \sin\phi . \tag{1.37}$$

Simple harmonic motion is equivalent to **one component** of uniform circular motion. This relation is illustrated in figure 1.4 and in program 1-1 on the programs disk. As the point moves around the circle at constant velocity, $R\omega$, the x coordinate executes simple harmonic motion with angular velocity ω. If we wish, we can choose the two constants required to fix the solution of (1.3) to be R and ϕ, instead of $x(0)$ and $x'(0)$. In this language, the action of resetting of the clock is more transparent. Resetting the clock changes the value of ϕ without changing anything else.

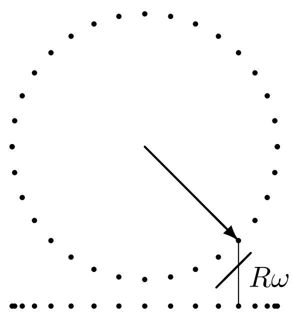


Figure 1.4: The relation between uniform circular motion and simple harmonic motion.

But we would like even more. The key idea is that linearity allows us considerable freedom. We can add solutions of the equations of motion together and multiply them by constants, and the result is still a solution. We would like to use this freedom to choose solutions that behave as simply as possible under time translations.

The simplest possible behavior for a solution $z(t)$ under time translation is

$$z(t + a) = h(a)\, z(t)\,. \tag{1.38}$$

That is, we would like find a solution that reproduces itself up to an overall constant, $h(a)$ when we reset our clocks by a. Because we are always free to multiply a solution of a homogeneous linear equation of motion by a constant, the change from $z(t)$ to $h(a)\, z(t)$ doesn't amount to much. We will call a solution satisfying (1.38) an "irreducible[3] solution" with respect to time translations, because its behavior under time translations (resettings of the clock) is as simple as it can possibly be.

It turns out that for systems whose equations of motion are linear and time translation invariant, as we will see in more detail below, we can always find irreducible solutions that

[3]The word "irreducible" is borrowed from the theory of group representations. In the language of group theory, the irreducible solution is an "irreducible representation of the translation group." It just means "as simple as possible."

have the property, (1.38). However, for simple harmonic motion, this requires complex numbers. You can see this by noting that changing the clock setting by π/ω just changes the sign of the solution with angular frequency ω, because both the $\cos$ and $\sin$ terms change sign:

$$\cos(\omega t + \pi) = -\cos\omega t\,, \quad \sin(\omega t + \pi) = -\sin\omega t\,. \tag{1.39}$$

But then from (1.38) and (1.39), we can write

$$\begin{aligned} -z(t) &= z(t + \pi/\omega) = z(t + \pi/2\omega + \pi/2\omega) \\ &= h(\pi/2\omega)\, z(t + \pi/2\omega) = h(\pi/2\omega)^2\, z(t)\,. \end{aligned} \tag{1.40}$$

Thus we cannot find such a solution unless $h(\pi/2\omega)$ has the property

$$[h(\pi/2\omega)]^2 = -1\,. \tag{1.41}$$

The square of $h(\pi/2\omega)$ is -1! Thus we are forced to consider complex numbers.[4] When we finish introducing complex numbers, we will come back to (1.38) and show that we can **always** find solutions of this form for systems that are linear and time translation invariant.

1.4 Complex Numbers

The square root of -1, called i, is important in physics and mathematics for many reasons. Measurable physical quantities can always be described by real numbers. You never get a reading of i meters on your meter stick. However, we will see that when i is included along with real numbers and the usual arithmetic operations (addition, subtraction, multiplication and division), then algebra, trigonometry and calculus all become simpler. While complex numbers are not necessary to describe wave phenomena, they will allow us to discuss them in a simpler and more insightful way.

1.4.1 Some Definitions

An imaginary number is a number of the form i times a real number.

A complex number, z, is a sum of a real number and an imaginary number: $z = a + ib$.

The real and "imaginary" parts, $\mathrm{Re}\,(z)$ and $\mathrm{Im}\,(z)$, of the complex number $z = a + ib$:

$$\mathrm{Re}\,(z) = a\,, \quad \mathrm{Im}\,(z) = b\,. \tag{1.42}$$

[4]The connection between complex numbers and uniform circular motion has been exploited by Richard Feynman in his beautiful little book, **QED**.

Note that the imaginary part is actually a real number, the real coefficient of i in $z = a + ib$.

The complex conjugate, z^*, of the complex number z, is obtained by changing the sign of i:

$$z^* = a - ib\,. \tag{1.43}$$

Note that $\mathrm{Re}\,(z) = (z + z^*)/2$ and $\mathrm{Im}\,(z) = (z - z^*)/2i$.

The complex plane: Because a complex number z is specified by two real numbers, it can be thought of as a two-dimensional vector, with components (a, b). The real part of z, $a = \mathrm{Re}\,(z)$, is the x component and the imaginary part of z, $b = \mathrm{Im}\,(z)$, is the y component. The diagrams in figures 1.5 and 1.6 show two vectors in the complex plane along with the corresponding complex numbers:

The absolute value, $|z|$, of z, is the length of the vector (a, b):

$$|z| = \sqrt{a^2 + b^2} = \sqrt{z^* z}\,. \tag{1.44}$$

The absolute value $|z|$ is always a real, non-negative number.

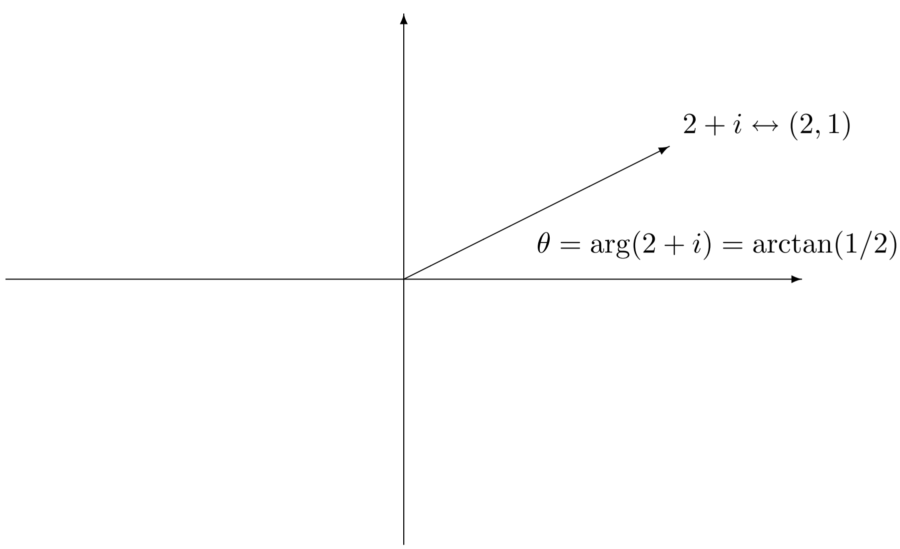


Figure 1.5: A vector with positive real part in the complex plane.

The argument or phase, arg(z), of a nonzero complex number z, is the angle, in radians, of the vector (a, b) counterclockwise from the x axis:

$$\arg(z) = \begin{cases} \arctan(b/a) \text{ for } a \geq 0\,, \\ \arctan(b/a) + \pi \text{ for } a < 0\,. \end{cases} \tag{1.45}$$

Like any angle, $\arg(z)$ can be redefined by adding a multiple of 2π radians or 360° (see figure 1.5 and 1.6).

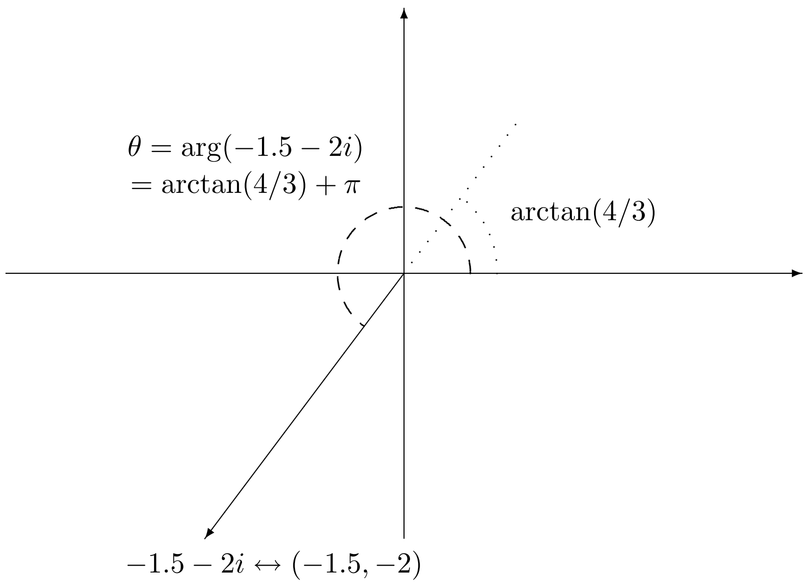


Figure 1.6: A vector with negative real part in the complex plane.

1.4.2 Arithmetic

1-2

The arithmetic operations addition, subtraction and multiplication on complex numbers are defined by just treating the i like a variable in algebra, using the distributive law and the relation $i^2 = -1$. Thus if $z = a + ib$ and $z' = a' + ib'$, then

$$\begin{aligned} z + z' &= (a + a') + i(b + b')\,, \\ z - z' &= (a - a') + i(b - b')\,, \\ zz' &= (aa' - bb') + i(ab' + ba')\,. \end{aligned} \tag{1.46}$$

For example:

$$(3 + 4i) + (-2 + 7i) = (3 - 2) + (4 + 7)i = 1 + 11i\,, \tag{1.47}$$

$$(3 + 4i) \cdot (5 + 7i) = (3 \cdot 5 - 4 \cdot 7) + (3 \cdot 7 + 4 \cdot 5)i = -13 + 41i\,. \tag{1.48}$$

It is worth playing with complex multiplication and getting to know the complex plane. At this point, you should check out program 1-2.

Division is more complicated. To divide a complex number z by a real number r is easy, just divide both the real and the imaginary parts by r to get $z/r = a/r + ib/r$. To divide by a complex number, z', we can use the fact that $z'^* z' = |z'|^2$ is real. If we multiply the numerator and the denominator of z/z' by z'^*, we can write:

$$z/z' = z'^* z/|z'|^2 = (aa' + bb')/(a'^2 + b'^2) + i(ba' - ab')/(a'^2 + b'^2) \,. \tag{1.49}$$

For example:

$$(3 + 4i)/(2 + i) = (3 + 4i) \cdot (2 - i)/5 = (10 + 5i)/5 = 2 + i \,. \tag{1.50}$$

With these definitions for the arithmetic operations, the absolute value behaves in a very simple way under multiplication and division. Under multiplication, the absolute value of a product of two complex numbers is the product of the absolute values:

$$|z\, z'| = |z|\, |z'| \,. \tag{1.51}$$

Division works the same way so long as you don't divide by zero:

$$|z/z'| = |z|/|z'| \quad \text{if} \quad z' \neq 0 \,. \tag{1.52}$$

Mathematicians call a set of objects on which addition and multiplication are defined and for which there is an absolute value satisfying (1.51) and (1.52) a division algebra. It is a peculiar (although irrelevant, for us) mathematical fact that the complex numbers are one of only four division algebras, the others being the real numbers and more bizarre things called quaternions and octonians obtained by relaxing the requirements of commutativity and associativity (respectively) of the multiplication laws.

The wonderful thing about the complex numbers from the point of view of algebra is that all polynomial equations have solutions. For example, the equation $x^2 - 2x + 5 = 0$ has no solutions in the real numbers, but has two complex solutions, $x = 1 \pm 2i$. In general, an equation of the form $p(x) = 0$, where $p(x)$ is a polynomial of degree n with complex (or real) coefficients has n solutions if complex numbers are allowed, but it may not have any if x is restricted to be real.

Note that the complex conjugate of any sum, product, etc, of complex numbers can be obtained simply by changing the sign of i wherever it appears. This implies that if the polynomial $p(z)$ has real coefficients, the solutions of $p(z) = 0$ come in complex conjugate pairs. That is, if $p(z) = 0$, then $p(z^*) = 0$ as well.

1.4.3 Complex Exponentials

Consider a complex number $z = a + ib$ with absolute value 1. Because $|z| = 1$ implies $a^2 + b^2 = 1$, we can write a and b as the cosine and sine of an angle θ.

$$z = \cos\theta + i \sin\theta \quad \text{for} \quad |z| = 1 \,. \tag{1.53}$$

Because

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{b}{a} \tag{1.54}$$

the angle θ is the argument of z:

$$\arg(\cos\theta + i\sin\theta) = \theta\,. \tag{1.55}$$

Let us think about z as a function of θ and consider the calculus. The derivative with respect to θ is:

$$\frac{\partial}{\partial\theta}(\cos\theta + i\sin\theta) = -\sin\theta + i\cos\theta = i(\cos\theta + i\sin\theta) \tag{1.56}$$

A function that goes into itself up to a constant under differentiation is an exponential. In particular, if we had a function of θ, $f(\theta)$, that satisfied $\frac{\partial}{\partial\theta}f(\theta) = kf(\theta)$ for real k, we would conclude that $f(\theta) = e^{k\theta}$. Thus if we want the calculus to work in the same way for complex numbers as for real numbers, we must conclude that

$$e^{i\theta} = \cos\theta + i\sin\theta\,. \tag{1.57}$$

We can check this relation by noting that the Taylor series expansions of the two sides are equal. The Taylor expansion of the exponential, $\cos$, and $\sin$ functions are:

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \\ \cos(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{4!} \cdots \\ \sin(x) &= x - \frac{x^3}{3!} + \cdots \end{aligned} \tag{1.58}$$

Thus the Taylor expansion of the left side of (1.57) is

$$1 + i\theta + (i\theta)^2/2 + (i\theta)^3/3! + \cdots \tag{1.59}$$

while the Taylor expansion of the right side is

$$(1 - \theta^2/2 + \cdots) + i(\theta - \theta^3/6 + \cdots) \tag{1.60}$$

The powers of i in (1.59) work in just the right way to reproduce the pattern of minus signs in (1.60).

Furthermore, the multiplication law works properly:

$$\begin{aligned} e^{i\theta}\,e^{i\theta'} &= (\cos\theta + i\sin\theta)(\cos\theta' + i\sin\theta') \\ &= (\cos\theta\,\cos\theta' - \sin\theta\,\sin\theta') + i(\sin\theta\,\cos\theta' + \cos\theta\,\sin\theta') \\ &= \cos(\theta+\theta') + i\sin(\theta+\theta') = e^{i(\theta+\theta')}\,. \end{aligned} \tag{1.61}$$

Thus (1.57) makes sense in all respects. This connection between complex exponentials and trigonometric functions is called Euler's Identity. It is extremely useful. For one thing, the logic can be reversed and the trigonometric functions can be "defined" algebraically in terms of complex exponentials:

$$\begin{aligned} \cos\theta &= \frac{e^{i\theta}+e^{-i\theta}}{2} \\ \sin\theta &= \frac{e^{i\theta}-e^{-i\theta}}{2i} = -i\frac{e^{i\theta}-e^{-i\theta}}{2}\,. \end{aligned} \tag{1.62}$$

Using (1.62), trigonometric identities can be derived very simply. For example:

$$\cos 3\theta = \mathrm{Re}\,(e^{3i\theta}) = \mathrm{Re}\,((e^{i\theta})^3) = \cos^3\theta - 3\cos\theta\,\sin^2\theta\,. \tag{1.63}$$

Another example that will be useful to us later is:

$$\begin{aligned} \cos(\theta+\theta')+\cos(\theta-\theta') &= (e^{i(\theta+\theta')}+e^{-i(\theta+\theta')}+e^{i(\theta-\theta')}+e^{-i(\theta-\theta')})/2 \\ &= (e^{i\theta}+e^{-i\theta})(e^{i\theta'}+e^{-i\theta'})/2 = 2\cos\theta\,\cos\theta'\,. \end{aligned} \tag{1.64}$$

Every nonzero complex number can be written as the product of a positive real number (its absolute value) and a complex number with absolute value 1. Thus

$$z = x+iy = R\,e^{i\theta} \quad \text{where} \quad R = |z|\,, \quad \text{and} \quad \theta = \arg(z)\,. \tag{1.65}$$

In the complex plane, (1.65) expresses the fact that a two-dimensional vector can be written either in Cartesian coordinates, (x, y), or in polar coordinates, (R, θ). For example, $\sqrt{3}+i = 2e^{i\pi/6}$; $1+i = \sqrt{2}\,e^{i\pi/4}$; $-8i = 8e^{3i\pi/2} = 8e^{-i\pi/2}$. Figure 1.7 shows the complex number $1+i = \sqrt{2}\,e^{i\pi/4}$.

The relation, (1.65), gives another useful way of thinking about multiplication of complex numbers. If

$$z_1 = R_1 e^{i\theta_1} \;\text{ and }\; z_2 = R_2 e^{i\theta_2}\,, \tag{1.66}$$

then

$$z_1 z_2 = R_1 R_2 e^{i(\theta_1+\theta_2)}\,. \tag{1.67}$$

In words, to multiply two complex numbers, you multiply the absolute values and add the arguments. You should now go back and play with program 1-2 with this relation in mind.

Equation (1.57) yields a number of relations that may seem surprising until you get used to them. For example: $e^{i\pi} = -1$; $e^{i\pi/2} = i$; $e^{2i\pi} = 1$. These have an interpretation in the complex plane where $e^{i\theta}$ is the unit vector $(\cos\theta, \sin\theta)$,

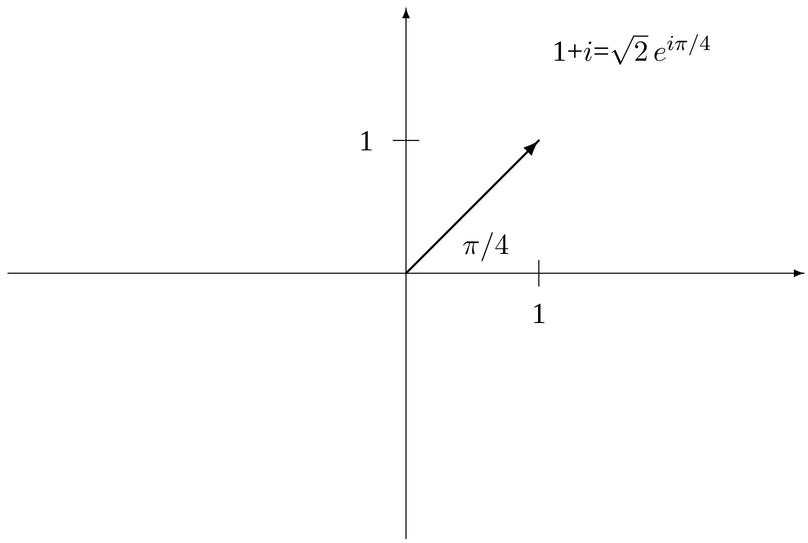


Figure 1.7: A complex number in two different forms.

which is at an angle θ measured counterclockwise from the x axis. Then -1 is 180° or π radians counterclockwise from the x axis, while i is along the y axis, 90° or $\pi/2$ radians from the x axis. 2π radians is 360°, and thus rotates us all the way back to the x axis. These relations are shown in figure 1.8.

1.4.4 Notation

It is not really necessary to have a notation that distinguishes between real numbers and complex numbers. The reason is that, as we have seen, the rules of arithmetic, algebra and calculus apply to real and complex numbers in exactly the same way. Nevertheless, some readers may find it helpful to be reminded when a quantity is complex. This is probably particularly useful for the quantities like x that represent physical coordinates. Therefore, at least for the first few chapters until the reader is thoroughly complexified, we will distinguish between real and complex "coordinates." If they are real, we will use letters x and y. If they are complex, we will use z and w.

1.5 Exponential Solutions

We are now ready to translate the conditions of linearity and time translation invariance into mathematics. What we will see is that the two properties of linearity and time translation invariance lead automatically to irreducible solutions satisfying (1.38), and furthermore that

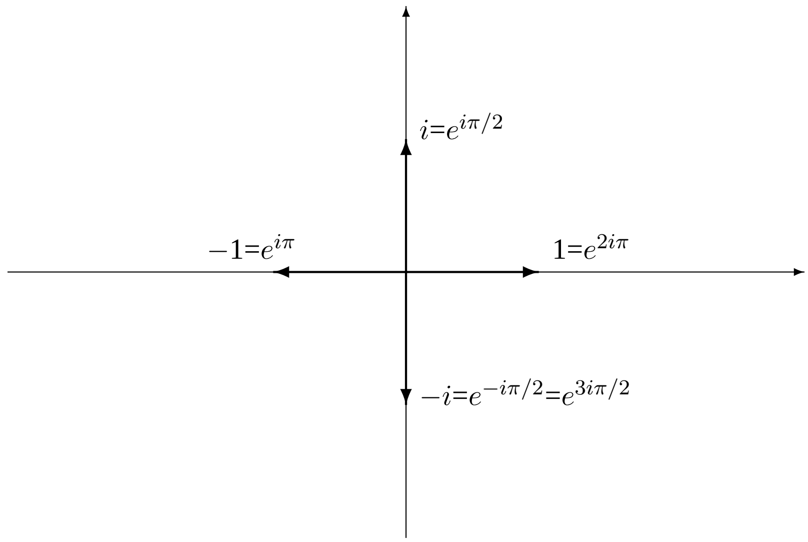


Figure 1.8: Some special complex exponentials in the complex plane.

these irreducible solutions are just exponentials. We do not need to use any other details about the equation of motion to get this result. Therefore our arguments will apply to much more complicated situations, in which there is damping or more degrees of freedom or both. **So long as the system has time translation invariance and linearity, the solutions will be sums of irreducible exponential solutions.**

We have seen that the solutions of homogeneous linear differential equations with constant coefficients, of the form,

$$\mathcal{M}\,\frac{d^2}{dt^2}\,x(t)+\mathcal{K}\,x(t)=0\,, \tag{1.68}$$

have the properties of linearity and time translation invariance. The equation of simple harmonic motion is of this form. The coordinates are real, and the constants $\mathcal{M}$ and $\mathcal{K}$ are real because they are physical things like masses and spring constants. However, we want to allow ourselves the luxury of considering complex solutions as well, so we consider the same equation with complex variables:

$$\mathcal{M}\,\frac{d^2}{dt^2}\,z(t)+\mathcal{K}\,z(t)=0\,. \tag{1.69}$$

Note the relation between the solutions to (1.68) and (1.69). Because the coefficients $\mathcal{M}$ and $\mathcal{K}$ are real, for every solution, $z(t)$, of (1.69), the complex conjugate, $z(t)^*$, is also a solution. The differential equation remains true when the signs of all the i's are changed.

From these two solutions, we can construct two real solutions:

$$\begin{aligned} x_1(t) &= \text{Re}\,(z(t)) = (z(t) + z(t)^*)\,/2\,; \\ x_2(t) &= \text{Im}\,(z(t)) = (z(t) - z(t)^*)\,/2i\,. \end{aligned} \tag{1.70}$$

All this is possible because of linearity, which allows us to go back and forth from real to complex solutions by forming linear combinations, as in (1.70). These are solutions of (1.68). Note that $x_1(t)$ and $x_2(t)$ are just the real and imaginary parts of $z(t)$. **The point is that you can always reconstruct the physical real solutions to the equation of motion from the complex solution. You can do all of the mathematics using complex variables, which makes it much easier. Then at the end you can get the physical solution of interest just by taking the real part of your complex solution.**

Now back to the solution to (1.69). What we want to show is that we are led to irreducible, exponential solutions for any system with time translation invariance and linearity! Thus we will understand why we can always find irreducible solutions, not only in (1.69), but in much more complicated situations with damping, or more degrees of freedom.

There are two crucial elements:

1. Time translation invariance, (1.33), which requires that $x(t + a)$ is a solution if $x(t)$ is a solution;
2. Linearity, which allows us to form linear combinations of solutions to get new solutions.

(1.71)

We will solve (1.68) using only these two elements. That will allow us to generalize our solution immediately to **any** system in which the properties, (1.71), are present.

One way of using linearity is to choose a "basis" set of solutions, $x_j(t)$ for $j = 1$ to n which is "complete" and "linearly independent." For the harmonic oscillator, two solutions are all we need, so $n = 2$. But our analysis will be much more general and will apply, for example, to linear systems with more degrees of freedom, so we will leave n free. What "complete" means is that any solution, $z(t)$, (which may be complex) can be expressed as a linear combination of the $x_j(t)$'s,

$$z(t) = \sum_{j=1}^{n} c_j x_j(t)\,. \tag{1.72}$$

What "linearly independent" means is that none of the $x_j(t)$'s can be expressed as a linear combination of the others, so that the only linear combination of the $x_j(t)$s that vanishes is the trivial combination, with only zero coefficients,

$$\sum_{j=1}^{n} c_j x_j(t) = 0 \Rightarrow c_j = 0\,. \tag{1.73}$$

Now let us see whether we can find an irreducible solution that behaves simply under a change in the initial clock setting, as in (1.38),

$$z(t+a) = h(a)\, z(t) \tag{1.74}$$

for some (possibly complex) function $h(a)$. In terms of the basis solutions, this is

$$z(t+a) = h(a) \sum_{k=1}^{n} c_k x_k(t)\,. \tag{1.75}$$

But each of the basis solutions also goes into a solution under a time translation, and each new solution can, in turn, be written as a linear combination of the basis solutions, as follows:

$$x_j(t+a) = \sum_{k=1}^{n} R_{jk}(a)\, x_k(t)\,. \tag{1.76}$$

Thus

$$z(t+a) = \sum_{j=1}^{n} c_j x_j(t+a) = \sum_{j,k=1}^{n} c_j R_{jk}(a)\, x_k(t)\,. \tag{1.77}$$

Comparing (1.75) and (1.77), and using (1.73), we see that we can find an irreducible solution if and only if

$$\sum_{j=1}^{n} c_j R_{jk}(a) = h(a)\, c_k \text{ for all } k. \tag{1.78}$$

This is called an "eigenvalue equation." We will have much more to say about eigenvalue equations in chapter 3, when we discuss matrix notation. For now, note that (1.78) is a set of n homogeneous simultaneous equations in the n unknown coefficients, c_j. We can rewrite it as

$$\sum_{j=1}^{n} c_j S_{jk}(a) = 0 \text{ for all } k, \tag{1.79}$$

where

$$S_{jk}(a) = \begin{cases} R_{jk}(a) \text{ for } j \neq k\,, \\ R_{jk}(a) - h(a) \text{ for } j = k\,. \end{cases} \tag{1.80}$$

We can find a solution to (1.78) if and only if there is a solution of the determinantal equation[5]

$$\det S_{jk}(a) = 0\,. \tag{1.81}$$

[5]We will discuss the determinant in detail in chapter 3, so if you have forgotten this result from algebra, don't worry about it for now.

(1.81) is an nth order equation in the variable $h(a)$. It may have no real solution, but it always has n complex solutions for $h(a)$ (although some of the $h(a)$ values may appear more than once). For each solution for $h(a)$, we can find a set of c_js satisfying (1.78). The different linear combinations, $z(t)$, constructed in this way will be a linearly independent set of irreducible solutions, each satisfying (1.74), for some $h(a)$. If there are n different $h(a)$s, the usual situation, they will be a complete set of irreducible solutions to the equations of motions. Then we may as well take our solutions to be irreducible, satisfying (1.74). We will see later what happens when some of the $h(a)$s appear more than once so that there are fewer than n different ones.

Now for each such irreducible solution, we can see what the functions $h(a)$ and $z(a)$ must be. If we differentiate both sides of (1.74) with respect to a, we obtain

$$z'(t+a) = h'(a)\, z(t)\,. \tag{1.82}$$

Setting $a = 0$ gives

$$z'(t) = H\, z(t) \tag{1.83}$$

where

$$H \equiv h'(0)\,. \tag{1.84}$$

This implies

$$z(t) \propto e^{Ht}\,. \tag{1.85}$$

Thus the irreducible solution is an exponential! **We have shown that (1.71) leads to irreducible, exponential solutions, without using any of details of the dynamics!**

1.5.1 * Building Up The Exponential

There is another way to see what (1.74) implies for the form of the irreducible solution that does not even involve solving the simple differential equation, (1.83). Begin by setting t=0 in (1.74). This gives

$$h(a) = z(a)/z(0)\,. \tag{1.86}$$

$h(a)$ is proportional to $z(a)$. This is particularly simple if we choose to multiply our irreducible solution by a constant so that $z(0) = 1$. Then (1.86) gives

$$h(a) = z(a) \tag{1.87}$$

and therefore

$$z(t+a) = z(t)\, z(a)\,. \tag{1.88}$$

Consider what happens for very small $t = \epsilon \ll 1$. Performing a Taylor expansion, we can write

$$z(\epsilon) = 1 + H\epsilon + O(\epsilon^2) \tag{1.89}$$

where $H = z'(0)$ from (1.84) and (1.87). Using (1.88), we can show that

$$z(N\epsilon) = [z(\epsilon)]^N \; . \tag{1.90}$$

Then for any t we can write (taking $t = N\epsilon$)

$$z(t) = \lim_{N\to\infty}[z(t/N)]^N = \lim_{N\to\infty}[1 + H(t/N)]^N = e^{Ht}\,. \tag{1.91}$$

Thus again, we see that the irreducible solution with respect to time translation invariance is just an exponential![6]

$$z(t) = e^{Ht}\,. \tag{1.92}$$

1.5.2 What is H?

When we put the irreducible solution, e^{Ht}, into (1.69), the derivatives just pull down powers of H so the equation becomes a purely algebraic equation (dropping an overall factor of e^{Ht})

$$\mathcal{M}\, H^2 + \mathcal{K} = 0\,. \tag{1.93}$$

Now, finally, we can see the relevance of complex numbers to the above discussion of time translation invariance. For positive $\mathcal{M}$ and $\mathcal{K}$, the equation (1.93) has no solutions at all if we restrict H to be real. We cannot find any real irreducible solutions. But there are always two solutions for H in the complex numbers. In this case, the solution is

$$H = \pm i\omega \quad \text{where} \quad \omega = \sqrt{\frac{\mathcal{K}}{\mathcal{M}}}\,. \tag{1.94}$$

It is only in this last step, where we actually compute H, that the details of (1.69) enter. Until (1.93), everything followed simply from the general principles, (1.71).

Now, as above, from these two solutions, we can construct two real solutions by taking the real and imaginary parts of $z(t) = e^{\pm i\omega t}$.

$$x_1(t) = \mathrm{Re}\;(z(t)) = \cos\omega t\,, \quad x_2(t) = \mathrm{Im}\;(z(t)) = \pm\sin\omega t\,. \tag{1.95}$$

Time translations mix up these two real solutions. That is why the irreducible complex exponential solutions are easier to work with. The quantity ω is the angular frequency that we saw in (1.5) in the solution of the equation of motion for the harmonic oscillator. Any linear

[6] For the mathematically sophisticated, what we have done here is to use the "group" structure of time translations to find the form of the solution. In words, we have built up an arbitrarily large time translation out of little ones.

combination of such solutions can be written in terms of an "amplitude" and a "phase" as follows: For real c and d

$$\begin{aligned} c\,\cos(\omega t) + d\,\sin(\omega t) &= c\,(e^{i\omega t} + e^{-i\omega t})/2 - id\,(e^{i\omega t} - e^{-i\omega t})/2 \\ &= \text{Re}\,\left((c+id)e^{-i\omega t}\right) = \text{Re}\,\left(A\,e^{i\theta}\,e^{-i\omega t}\right) \\ &= \text{Re}\,\left(A\,e^{-i(\omega t-\theta)}\right) = A\cos(\omega t-\theta)\,. \end{aligned} \tag{1.96}$$

where A is a positive real number called the **amplitude**,

$$A = \sqrt{c^2+d^2}\,, \tag{1.97}$$

and θ is an angle called the **phase**,

$$\theta = \arg(c+id)\,. \tag{1.98}$$

These relations are another example of the equivalence of Cartesian coordinates and polar coordinates, discussed after (1.65). The pair, c and d, are the Cartesian coordinates in the complex plane of the complex number, $c+id$. The amplitude, A, and phase, θ, are the polar coordinate representation of the same complex (1.96) shows that c and d are also the coefficients of $\cos\omega t$ and $\sin\omega t$ in the real part of the product of this complex number with $e^{-i\omega t}$. This relation is illustrated in figure 1.9 (note the relation to figure 1.4). As z moves clockwise with constant angular velocity, ω, around the circle, $|z| = A$, in the complex plane, the real part of z undergoes simple harmonic motion, $A\cos(\omega t-\theta)$.

Now that you know about complex numbers and complex exponentials, you should go back to the relation between simple harmonic motion and uniform circular motion illustrated in figure 1.4 and in supplementary program 1-1. The uniform circular motion can interpreted as a motion in the complex plane of the

$$z(t) = e^{-i\omega t}\,. \tag{1.99}$$

As t changes, $z(t)$ moves with constant clockwise velocity around the unit circle in the complex plane. This is the clockwise motion shown in program 1-1. The real part, $\cos\omega t$, executes simple harmonic motion.

Note that we could have just as easily taken our complex solution to be $e^{+i\omega t}$. This would correspond to counterclockwise motion in the complex plane, but the real part, which is all that matters physically, would be unchanged. It is **conventional** in physics to go to complex solutions proportional to $e^{-i\omega t}$. This is purely a convention. There is no physics in it. However, it is sufficiently universal in the physics literature that we will try to do it consistently here.

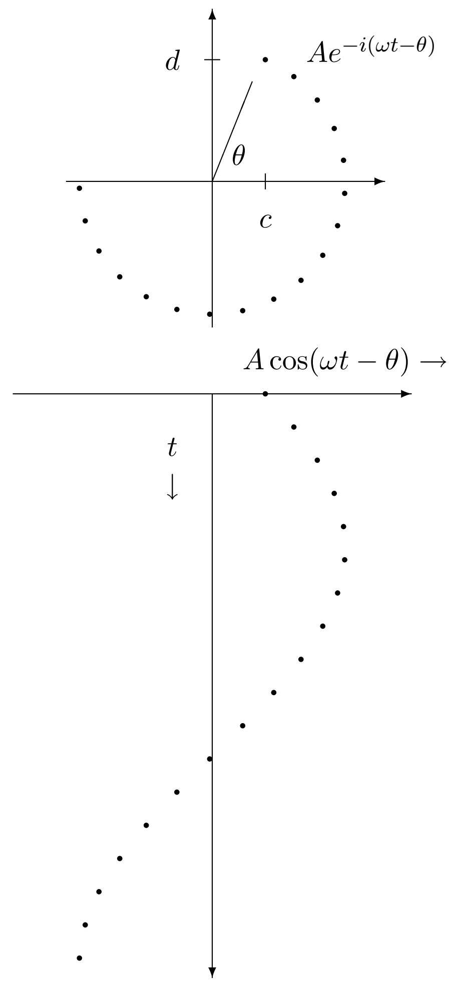


Figure 1.9: The relation (1.96) in the complex plane.

1.6 LC Circuits

One of the most important examples of an oscillating system is an LC circuit. You probably studied these in your course on electricity and magnetism. Like a Hooke's law spring, this system is linear, because the relations between charge, current, voltage, and the like for ideal inductors, capacitors and resistors are linear. Here we want to make explicit the analogy between a particular LC circuit and a system of a mass on a spring. The LC circuit with a resistanceless inductor with an inductance L and a capacitor of capacitance C is shown in figure 1.10. We might not ordinarily think of this as a circuit at all, because there is no battery or other source of electrical power. However, we could imagine, for example, that the capacitor was charged initially when the circuit was put together. Then current would flow when the circuit was completed. In fact, in the absence of resistance, the current would continue to oscillate forever. We shall see that this circuit is analogous to the combination of springs and a mass shown in figure 1.11. The oscillation frequency of the mechanical system is

$$\omega = \sqrt{\frac{K}{m}} \tag{1.100}$$

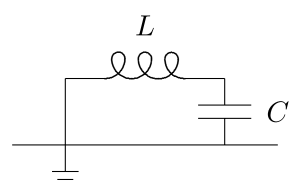


Figure 1.10: An LC circuit.

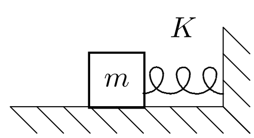


Figure 1.11: A system analogous to figure 1.10.

We can describe the configuration of the mechanical system of figure 1.10 in terms of x, the displacement of the block to the right. We can describe the configuration of the LC circuit of figure 1.10 in terms of Q, the charge that has been “displaced” through the inductor from the equilibrium situation with the capacitor uncharged. In this case, the charge displaced through the inductor goes entirely onto the capacitor because there is nowhere else for it to go, as shown in figure 1.12. The current through the inductor is the time derivative of the charge that has gone through,

$$I = \frac{dQ}{dt} \,. \tag{1.101}$$

To see how the LC circuit works, we can examine the voltages at various points in the system, as shown in figure 1.13. For an inductor, the voltage drop across it is the rate of

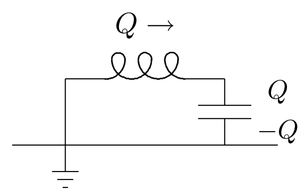


Figure 1.12: The charge moved through the inductor.

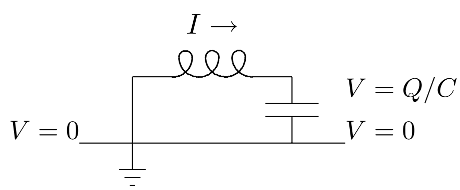


Figure 1.13: Voltage and current.

change of current through it, or

$$-L\,\frac{dI}{dt} = V\,. \tag{1.102}$$

For the capacitor, the stored charge is the voltage times the capacitance, or

$$V = Q/C\,. \tag{1.103}$$

Putting (1.101), (1.102) and (1.103) together gives

$$L\,\frac{dI}{dt} = L\,\frac{d^2Q}{dt^2} = -\frac{1}{C}\,Q\,. \tag{1.104}$$

The correspondence between the two systems is the following:

$$\begin{aligned} m &\leftrightarrow L \\ K &\leftrightarrow 1/C \\ x &\leftrightarrow Q \end{aligned} \tag{1.105}$$

When we make the substitutions in (1.105), the equation of motion, (1.3), of the mass on a spring goes into (1.104). Thus, knowing the solution, (1.6), for the mass on a spring, we can immediately conclude that the displaced charge in this LC circuit oscillates with frequency

$$\omega = \sqrt{\frac{1}{LC}}\,. \tag{1.106}$$

1.7 Units — Displacement and Energy

We have now seen two very different kinds of physical systems that exhibit simple harmonic oscillation. Others are possible as well, and we will give another example below. This is a good time to discuss the units of the equations of motions. The "generic" equation of motion for simple harmonic motion without damping looks like this

$$\mathcal{M}\,\frac{d^2\mathcal{X}}{dt^2} = -\mathcal{K}\,\mathcal{X} \tag{1.107}$$

where

$$\begin{aligned} &\mathcal{X} \text{ is the generalized coordinate,}\\ &\mathcal{M} \text{ is the generalized mass,}\\ &\mathcal{K} \text{ is the generalized spring constant.}\end{aligned} \tag{1.108}$$

In the simple harmonic motion of a point mass, $\mathcal{X}$ is just the displacement from equilibrium, x, $\mathcal{M}$ is the mass, m, and $\mathcal{K}$ is the spring constant, K.

The appropriate units for $\mathcal{M}$ and $\mathcal{K}$ depend on the units for $\mathcal{X}$. They are conventionally determined by the requirement that

$$\frac{1}{2}\mathcal{M}\left(\frac{d\mathcal{X}}{dt}\right)^2 \tag{1.109}$$

is the "kinetic" energy of the system arising from the change of the coordinate with time, and

$$\frac{1}{2}\mathcal{K}\mathcal{X}^2 \tag{1.110}$$

is the "potential" energy of the system, stored in the generalized spring.

It makes good physical sense to grant the energy a special status in these problems because in the absence of friction and external forces, the total energy, the sum of the kinetic energy in (1.109) and the potential energy in (1.110), is constant. In the oscillation, the energy is alternately stored in kinetic energy and potential energy. When the system is in its equilibrium configuration, but moving with its maximum velocity, the energy is all kinetic. When the system instantaneously comes to rest at its maximum displacement, all the energy is potential energy. In fact, it is sometimes easier to identify $\mathcal{M}$ and $\mathcal{K}$ by calculating the kinetic and potential energies than by finding the equation of motion directly. We will use this trick in chapter 11 to discuss water waves.

For example, in an LC circuit in SI units, we took our generalized coordinate to be a charge, Q, in Coulombs. Energy is measured in Joules or Volts×Coulombs. The generalized spring constant has units of

$$\frac{\text{Joules}}{\text{Coulombs}^2} = \frac{\text{Volts}}{\text{Coulombs}} \tag{1.111}$$

which is one over the unit of capacitance, Coulombs per Volt, or farads. The generalized mass has units of

$$\frac{\text{Joules} \times \text{seconds}^2}{\text{Coulombs}^2} = \frac{\text{Volts} \times \text{seconds}}{\text{Amperes}} \tag{1.112}$$

which is a unit of inductance (Henrys). This is what we used in our correspondence between the LC circuit and the mechanical oscillator, (1.105).

We can also add a generalized force to the right-hand side of (1.107). The generalized force has units of energy over generalized displacement. This is right because when the equation of motion is multiplied by the displacement, (1.109) and (1.110) imply that each of the terms has units of energy. Thus for example, in the LC circuit example, the generalized force is a voltage.

1.7.1 Constant Energy

The total energy is the sum of kinetic plus potential energy from (1.109) and (1.110),

$$E = \frac{1}{2}\mathcal{M}\left(\frac{d\mathcal{X}}{dt}\right)^2 + \frac{1}{2}\mathcal{K}\mathcal{X}^2\,. \tag{1.113}$$

If there are no external forces acting on the system, the total energy must be constant. You can see from (1.113) that the energy can be constant for an oscillating solution only if the angular frequency, ω, is $\sqrt{\mathcal{K}/\mathcal{M}}$. Suppose, for example, that the generalized displacement of the system has the form

$$\mathcal{X}(t) = \mathcal{A}\sin\omega t\,, \tag{1.114}$$

where $\mathcal{A}$ is an amplitude with the units of $\mathcal{X}$. Then the generalized velocity, is

$$\frac{d}{dt}\mathcal{X}(t) = \mathcal{A}\omega\cos\omega t\,. \tag{1.115}$$

To make the energy constant, we must have

$$\mathcal{K} = \omega^2\mathcal{M}\,. \tag{1.116}$$

Then, the total energy, from (1.109) and (1.110) is

$$\frac{1}{2}\mathcal{M}\omega^2\mathcal{A}^2\cos^2\omega t + \frac{1}{2}\mathcal{K}\mathcal{A}^2\sin^2\omega t = \frac{1}{2}\mathcal{K}\mathcal{A}^2\,. \tag{1.117}$$

1.7.2 The Torsion Pendulum

One more example may be useful. Let us consider the torsion pendulum, shown in figure 1.14.

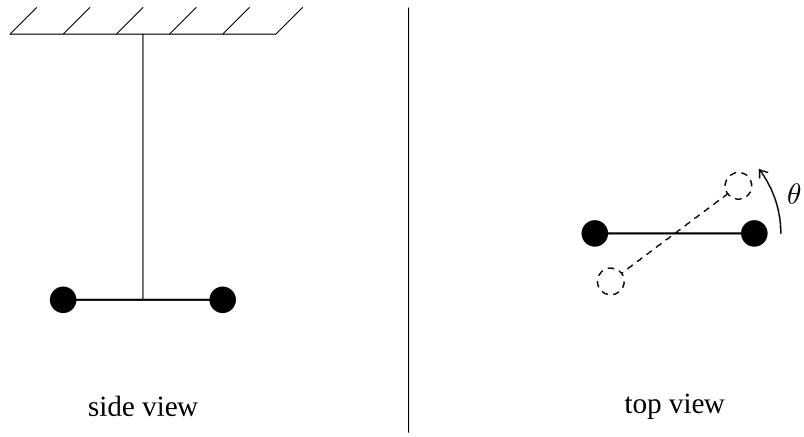


Figure 1.14: Two views of a torsion pendulum.

A torsion pendulum is a simple but very useful oscillator consisting of a dumbbell or rod supported at its center by a wire or fiber, hung from a support above. When the dumbbell is twisted by an angle θ, as shown in the top view in figure 1.14, the wire twists and provides a restoring torque on the dumbbell. For a suitable wire or fiber, this restoring torque is nearly linear even for rather large displacement angles. In this system, the natural variable to use for the displacement is the angle θ. Then the equation of motion is

$$I\,\frac{d^2\theta}{dt^2} = -\alpha\theta\,, \tag{1.118}$$

where I is the moment of inertia of the dumbbell about its center and $-\alpha\theta$ is the restoring force. Thus the generalized mass is the moment of inertia, I, with units of length squared times mass and the generalized spring constant is the constant α, with units of torque. As expected, from (1.109) and (1.110), the kinetic energy and potential energy are (respectively)

$$\frac{1}{2}I\left(\frac{d\theta}{dt}\right)^2 \text{ and } \frac{1}{2}\alpha\theta^2\,. \tag{1.119}$$

1.8 A Simple Nonlinear Oscillator

To illustrate some of the differences between linear and nonlinear oscillators, we will give one very simple example of a nonlinear oscillator. Consider the following nonlinear equation of motion:

$$m\frac{d^2}{dt^2}x = \begin{cases} -F_0 \ \text{ for } x > 0\,, \\ F_0 \ \text{ for } x < 0\,, \\ 0 \ \text{ for } x = 0\,. \end{cases} \tag{1.120}$$

This describes a particle with mass, m, that is subject to a force to the left, $-F_0$, when the particle is to the right of the origin ($x(t) > 0$), a force to the right, F_0, when the particle is to the left of the origin ($x(t) < 0$), and no force when the particle is sitting right on the origin.

The potential energy for this system grows linearly on both sides of $x = 0$. It cannot be differentiated at $x = 0$, because the derivative is not continuous there. Thus, we cannot expand the potential energy (or the force) in a Taylor series around the point $x = 0$, and the arguments of (1.21)-(1.24) do not apply.

It is easy to find a solution of (1.120). Suppose that at time, $t = 0$, the particle is at the origin but moving with positive velocity, v. The particle immediately moves to the right of the origin and decelerates with constant acceleration, $-F_0/m$, so that

$$x(t) = vt - \frac{F_0}{2m}t^2 \ \text{ for } t \leq \tau\,, \tag{1.121}$$

where

$$\tau = \frac{2mv}{F_0} \tag{1.122}$$

is the time required for the particle to turn around and get back to the origin. At time, $t = \tau$, the particle moves to the left of the origin. At this point it is moving with velocity, $-v$, the process is repeated for negative x and positive acceleration F_0/m. Then the solution continues in the form

$$x(t) = -v(t-\tau) + \frac{F_0}{2m}(t-\tau)^2 \ \text{ for } \tau \leq t \leq 2\tau\,. \tag{1.123}$$

Then the whole process repeats. The motion of the particle, shown in figure 1.15, looks superficially like harmonic oscillation, but the curve is a sequence of parabolas pasted together, instead of a sine wave.

The equation of motion, (1.120), is time translation invariant. Clearly, we can start the particle at the origin with velocity, v, at any time, t_0. The solution then looks like that shown in figure 1.15 but translated in time by t_0. The solution has the form

$$x_{t_0}(t) = x(t - t_0) \tag{1.124}$$

where $x(t)$ is the function described by (1.121), (1.123), etc. This shown in figure 1.16 for $t = t_0 = 3\tau/4$. The dotted curve corresponds to $t_0 = 0$

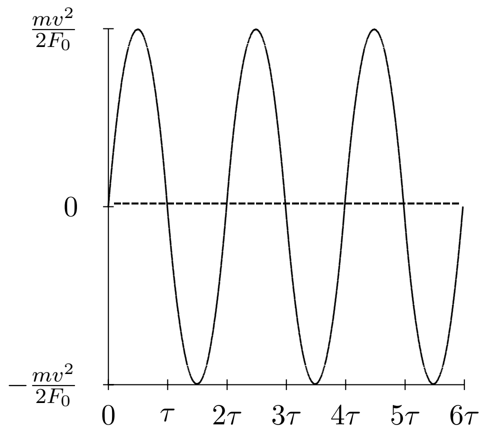


Figure 1.15: The motion of a particle with a nonlinear equation of motion.

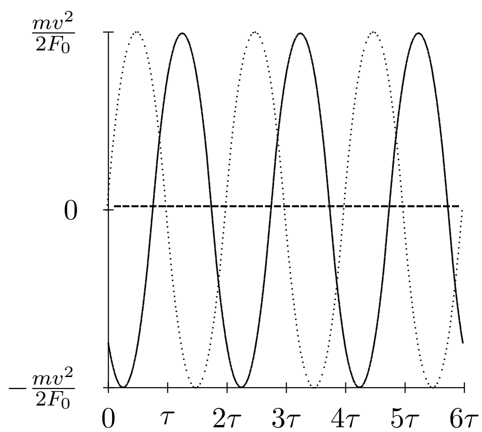


Figure 1.16: Motion started from the origin at $t = t_0 = 3\tau/4$.

Like the harmonic oscillator, this system oscillates regularly and indefinitely. However, in this case, the period of the oscillation, the time it takes to repeat, 2τ, depends on the amplitude of the oscillation, or equivalently, on the initial velocity, v. The period is proportional to v, from (1.122). The motion of the particle started from the origin at $t = t_0$, for an initial velocity $v/2$ is shown in figure 1.17. The dotted curve corresponds to an initial velocity, v.

While the nonlinear equation of motion, (1.120), is time translation invariant, the symmetry is much less useful because the system lacks linearity. From our point of view, the important thing about linearity (apart from the fact that it is a good approximation in so many important physical systems), is that it allows us to choose a convenient basis for the solutions to the equation of motion. We choose them to behave simply under time translations.

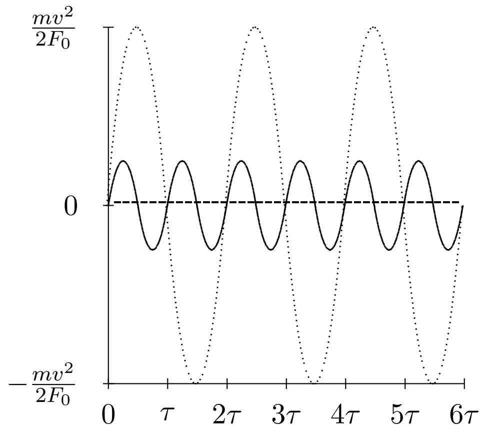


Figure 1.17: Initial velocity $v/2$.

Then, because of linearity, we can build up any solution as a linear combination of the basis solutions. In a situation like (1.120), we do not have this option.

Chapter Checklist

You should now be able to:

1. Analyze the physics of a harmonic oscillator, including finding the spring constant, setting up the equation of motion, solving it, and imposing initial conditions;

2. Find the approximate "spring constant" for the small oscillations about a point of equilibrium and estimate the displacement for which linearity breaks down;

3. Understand the connection between harmonic oscillation and uniform circular motion;

4. Use complex arithmetic and complex exponentials;

5. Solve homogeneous linear equations of motion using irreducible solutions that are complex exponentials;

6. Understand and explain the difference between frequency and angular frequency;

7. Analyze the oscillations of LC circuits;

8. Compute physical quantities for oscillating systems in SI units.

9. Understand time translation invariance in nonlinear systems.

Problems

1.1. For the mass and spring discussed (1.1)-(1.8), suppose that the system is hung vertically in the earth's gravitational field, with the top of the spring held fixed. Show that the frequency for vertical oscillations is given by (1.5). Explain why gravity has no effect on the angular frequency.

1.2a. Find an expression for $\cos 7\theta$ in terms of $\cos\theta$ and $\sin\theta$ by using complex exponentials and the binomial expansion.

b. Do the same for $\sin 5\theta$.

c. Use complex exponentials to find an expression for $\sin(\theta_1 + \theta_2 + \theta_3)$ in terms of the sines and cosines of the individual angles.

d. Do you remember the "half angle formula,"

$$\cos^2\frac{\theta}{2} = \frac{1}{2}(1+\cos\theta)\,?$$

Use complex exponentials to prove the "fifth angle formula,"

$$\cos^5\frac{\theta}{5} = \frac{10}{16}\cos\frac{\theta}{5} + \frac{5}{16}\cos\frac{3\theta}{5} + \frac{1}{16}\cos\theta\,.$$

e. Use complex exponentials to prove the identity

$$\sin 6x = \sin x\left(32\cos^5 x - 32\cos^3 x + 6\cos x\right)\,.$$

1.3a. Write $i+\sqrt{3}$ in the form $R\,e^{i\theta}$. Write θ as a rational number times π.

b. Do the same for $i-\sqrt{3}$.

c. Show that the two square roots of $R\,e^{i\theta}$ are $\pm\sqrt{R}\,e^{i\theta/2}$. **Hint:** This is easy! Don't work too hard.

d. Use the result of **c.** to find the square roots of $2i$ and $2+2i\sqrt{3}$.

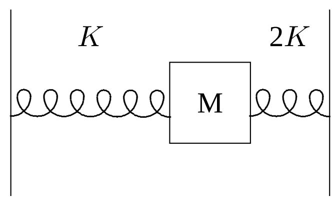
1.4. Find all six solutions to the equation $z^6 = 1$ and write each in the form $A+iB$ and plot them in the complex plane. **Hint:** write $z = Re^{i\theta}$ for R real and positive, and find R and θ.

1.5. Find three independent solutions to the differential equation

$$\frac{d^3}{dt^3}\, f(t) + f(t) = 0\,.$$

You should use complex exponentials to derive the solutions, but express the results in real form.

1.6. A block of mass M slides without friction between two springs of spring constant K and $2K$, as shown. The block is constrained to move only left and right on the paper, so the system has only one degree of freedom.



Calculate the oscillation angular frequency. If the velocity of the block when it is at its equilibrium position is v, calculate the amplitude of the oscillation.

1.7. A particle of mass m moves on the x axis with potential energy

$$V(x) = \frac{E_0}{a^4}\left(x^4 + 4ax^3 - 8a^2x^2\right)\,.$$

Find the positions at which the particle is in stable equilibrium. Find the angular frequency of small oscillations about each equilibrium position. What do you mean by small oscillations? Be quantitative and give a separate answer for each point of stable equilibrium.

1.8. For the torsion pendulum of figure 1.14, suppose that the pendulum consists of two 0.01 kg masses on a light rod of total length 0.1 m. If the generalized spring constant, α, is 5×10^{-7} N m. Find the angular frequency of the oscillator.

Chapter 2

Forced Oscillation and Resonance

The forced oscillation problem will be crucial to our understanding of wave phenomena. Complex exponentials are even more useful for the discussion of damping and forced oscillations. They will help us to discuss forced oscillations without getting lost in algebra.

Preview

In this chapter, we apply the tools of complex exponentials and time translation invariance to deal with damped oscillation and the important physical phenomenon of resonance in single oscillators.

1. We set up and solve (using complex exponentials) the equation of motion for a damped harmonic oscillator in the overdamped, underdamped and critically damped regions.

2. We set up the equation of motion for the damped and forced harmonic oscillator.

3. We study the solution, which exhibits a resonance when the forcing frequency equals the free oscillation frequency of the corresponding undamped oscillator.

4. We study in detail a specific system of a mass on a spring in a viscous fluid. We give a physical explanation of the phase relation between the forcing term and the damping.

2.1 Damped Oscillators

Consider first the free oscillation of a damped oscillator. This could be, for example, a system of a block attached to a spring, like that shown in figure 1.1, but with the whole system immersed in a viscous fluid. Then in addition to the restoring force from the spring, the block experiences a frictional force. For small velocities, the frictional force can be taken to have the form

$$-\,m\Gamma v\,, \tag{2.1}$$

where Γ is a constant. Notice that because we have extracted the factor of the mass of the block in (2.1), $1/\Gamma$ has the dimensions of time. We can write the equation of motion of the system as

$$\frac{d^2}{dt^2}\,x(t)+\Gamma\,\frac{d}{dt}\,x(t)+\omega_0^2\,x(t)=0\,, \tag{2.2}$$

where $\omega_0=\sqrt{K/m}$. This equation is linear and time translation invariant, like the undamped equation of motion. In fact, it is just the form that we analyzed in the previous chapter, in (1.16). As before, we allow for the possibility of complex solutions to the same equation,

$$\frac{d^2}{dt^2}\,z(t)+\Gamma\,\frac{d}{dt}\,z(t)+\omega_0^2\,z(t)=0\,. \tag{2.3}$$

Because (1.71) is satisfied, we know from the arguments of of chapter 1 that we can find irreducible solutions of the form

$$z(t)=e^{\alpha t}\,, \tag{2.4}$$

where α (Greek letter alpha) is a constant. Putting (2.4) into (2.2), we find

$$(\alpha^2+\Gamma\alpha+\omega_0^2)\,e^{\alpha t}=0\,. \tag{2.5}$$

Because the exponential never vanishes, the quantity in parentheses must be zero, thus

$$\alpha=-\frac{\Gamma}{2}\pm\sqrt{\frac{\Gamma^2}{4}-\omega_0^2}\,. \tag{2.6}$$

From (2.6), we see that there are three regions for Γ compared to ω_0 that lead to different physics.

2.1.1 Overdamped Oscillators

If $\Gamma/2>\omega_0$, both solutions for α are real and negative. The solution to (2.2) is a sum of decreasing exponentials. Any initial displacement of the system dies away with no oscillation. This is an **overdamped oscillator**.

The general solution in the overdamped case has the form,

$$x(t)=z(t)=A_+e^{-\Gamma_+t}+A_-e^{-\Gamma_-t}\,, \tag{2.7}$$

where

$$\Gamma_\pm=\frac{\Gamma}{2}\pm\sqrt{\frac{\Gamma^2}{4}-\omega_0^2}\,. \tag{2.8}$$

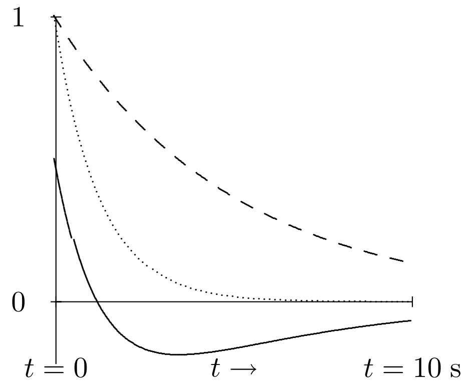


Figure 2.1: Solutions to the equation of motion for an overdamped oscillator.

An example is shown in figure 2.1. The dotted line is $e^{-\Gamma_+ t}$ for $\Gamma = 1$ s^{-1} and $\omega_0 = .4$ s^{-1}. The dashed line is $e^{-\Gamma_- t}$. The solid line is a linear combination, $e^{-\Gamma_+ t} - \frac{1}{2}e^{-\Gamma_- t}$.

In the overdamped situation, there is really no oscillation. If the mass is initially moving very fast toward the equilibrium position, it can overshoot, as shown in figure 2.1. However, it then moves exponentially back toward the equilibrium position, without ever crossing the equilibrium value of the displacement a second time. Thus in the free motion of an overdamped oscillator, the equilibrium position is crossed either zero or one times.

2.1.2 Underdamped Oscillators

If $\Gamma/2 < \omega_0$, the expression inside the square root is negative, and the solutions for α are a complex conjugate pair, with negative real part. Thus the solutions are products of a decreasing exponential, $e^{-\Gamma t/2}$, times complex exponentials (or sines and cosines) $e^{\pm i\omega t}$, where

$$\omega^2 = \omega_0^2 - \Gamma^2/4\,. \tag{2.9}$$

This is an **underdamped oscillator**.

Most of the systems that we think of as oscillators are underdamped. For example, a system of a child sitting still on a playground swing is an underdamped pendulum that can oscillate many times before frictional forces bring it to rest.

The decaying exponential $e^{-\Gamma t/2}e^{-i(\omega t-\theta)}$ spirals in toward the origin in the complex plane. Its real part, $e^{-\Gamma t/2}\cos(\omega t - \theta)$, describes a function that oscillates with decreasing amplitude. In real form, the general solution for the underdamped case has the form,

$$x(t) = A\, e^{-\Gamma t/2} \cos(\omega t - \theta)\,, \tag{2.10}$$

or

$$x(t) = e^{-\Gamma t/2}\left(c\,\cos(\omega t) + d\,\sin(\omega t)\right)\,, \tag{2.11}$$

where A and θ are related to c and d by (1.97) and (1.98). This is shown in figure 2.2 (to be compared with figure 1.9). The upper figure shows the complex plane with $e^{-\Gamma t/2}e^{-i(\omega t-\theta)}$ plotted for equally spaced values of t. The lower figure is the real part, $\cos(\omega t-\theta) \rightarrow$, for the same values of t plotted versus t. In the underdamped case, the equilibrium position is crossed an infinite number of times, although with exponentially decreasing amplitude!

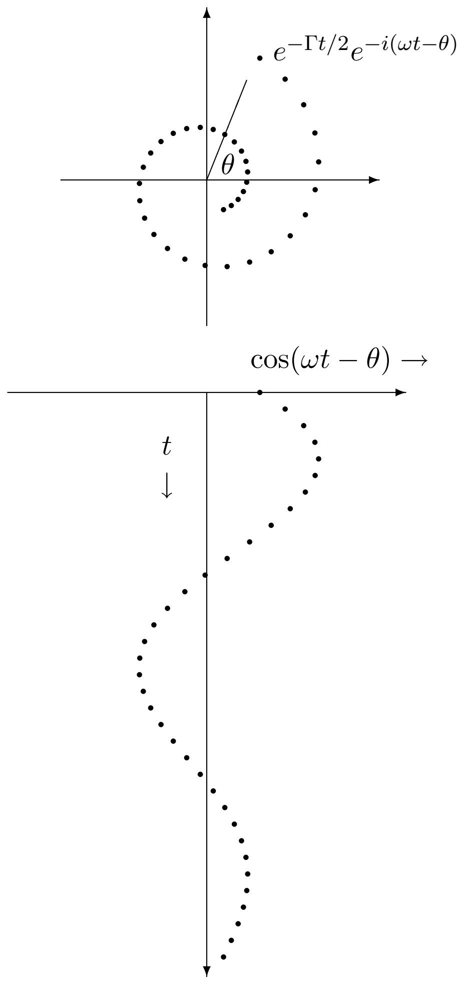


Figure 2.2: A damped complex exponential.

2.1.3 Critically Damped Oscillators

If $\Gamma/2 = \omega_0$, then (2.4), gives only one solution, $e^{-\Gamma t/2}$. We know that there will be two solutions to the second order differential equation, (2.2). One way to find the other solution is to approach this situation from the underdamped case as a limit. If we write the solutions to the underdamped case in real form, they are $e^{-\Gamma t/2}\cos\omega t$ and $e^{-\Gamma t/2}\sin\omega t$. Taking the limit of the first as $\omega \to 0$ gives $e^{-\Gamma t/2}$, the solution we already know. Taking the limit of the second gives 0. However, if we first divide the second solution by ω, it is still a solution because ω does not depend on t. Now we can get a nonzero limit:

$$\lim_{\omega\to 0}\frac{1}{\omega}e^{-\Gamma t/2}\sin\omega t = t\,e^{-\Gamma t/2}\,. \tag{2.12}$$

Thus $t\,e^{-\Gamma t/2}$ is also a solution. You can also check this explicitly, by inserting it back into (2.2). This is called the **critically damped** case because it is the boundary between overdamping and underdamping.

A familiar system that is close to critical damping is the combination of springs and shock absorbers in an automobile. Here the damping must be large enough to prevent the car from bouncing. But if the damping from the shocks is too high, the car will not be able to respond quickly to bumps and the ride will be rough.

The general solution in the critically damped case is thus

$$c\,e^{-\Gamma t/2} + d\,t\,e^{-\Gamma t/2}\,. \tag{2.13}$$

This is illustrated in figure 2.3. The dotted line is $e^{-\Gamma t}$ for $\Gamma = 1$ s^{-1}. The dashed line is $t\,e^{-\Gamma t}$. The solid line is a linear combination, $(1-t)\,e^{-\Gamma t}$.

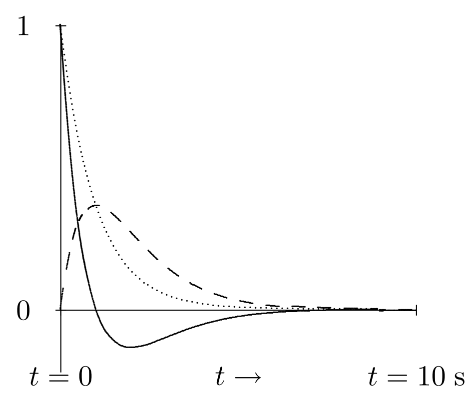


Figure 2.3: Solutions to the equation of motion for a critically damped oscillator.

As in the overdamped situation, there is no real oscillation for critical damping. However, again, the mass can overshoot and then go smoothly back toward the equilibrium position, without ever crossing the equilibrium value of the displacement a second time. As for over-damping, the equilibrium position is crossed either once or not at all.

2.2 Forced Oscillations

The damped oscillator with a harmonic driving force, has the equation of motion

$$\frac{d^2}{dt^2}\,x(t) + \Gamma\,\frac{d}{dt}\,x(t) + \omega_0^2\,x(t) = F(t)/m\,, \tag{2.14}$$

where the force is

$$F(t) = F_0 \cos\omega_d t\,. \tag{2.15}$$

The $\omega_d/2\pi$ is called the driving frequency. Notice that it is **not** necessarily the same as the natural frequency, $\omega_0/2\pi$, nor is it the oscillation frequency of the free system, (2.9). It is simply the frequency of the external force. It can be tuned completely independently of the other parameters of the system. It would be correct but awkward to refer to ω_d as the driving angular frequency. We will simply call it the driving frequency, ignoring its angular character.

The angular frequencies, ω_d and ω_0, appear in the equation of motion, (2.15), in completely different ways. You must keep the distinction in mind to understand forced oscillation. The natural angular frequency of the system, ω_0, is some combination of the masses and spring constants (or whatever relevant physical quantities determine the free oscillations). The angular frequency, ω_d, enters **only** through the time dependence of the driving force. This is the new aspect of forced oscillation. To exploit this new aspect fully, we will look for a solution to the equation of motion that oscillates with the same angular frequency, ω_d, as the driving force.

We can relate (2.14) to an equation of motion with a complex driving force

$$\frac{d^2}{dt^2}\,z(t) + \Gamma\,\frac{d}{dt}\,z(t) + \omega_0^2\,z(t) = \mathcal{F}(t)/m\,, \tag{2.16}$$

where

$$\mathcal{F}(t) = F_0 e^{-i\omega_d t}\,. \tag{2.17}$$

This works because the equation of motion, (2.14), does not involve i explicitly and because

$$\operatorname{Re}\mathcal{F}(t) = F(t)\,. \tag{2.18}$$

If $z(t)$ is a solution to (2.16), then you can prove that $x(t) = \operatorname{Re} z(t)$ is a solution (2.14) by taking the real part of both sides of (2.16).

The advantage to the complex exponential force, in (2.16), is that it is irreducible, it behaves simply under time translations. In particular, we can find a steady state solution proportional to the driving force, $e^{-i\omega_d t}$, whereas for the real driving force, the $\cos\omega_d t$ and $\sin\omega_d t$ forms get mixed up. That is, we look for a steady state solution of the form

$$z(t) = \mathcal{A}\, e^{-i\omega_d t}\,. \tag{2.19}$$

The steady state solution, (2.19), is a particular solution, not the most general solution to (2.16). As discussed in chapter 1, the most general solution of (2.16) is obtained by adding to the particular solution the most general solution for the free motion of the same oscillator (solutions of (2.3)). In general we will have to include these more general contributions to satisfy the initial conditions. However, as we have seen above, all of these solutions die away exponentially with time. They are what are called "transient" solutions. It is only the steady state solution that survives for a long time in the presence of damping. Unlike the solutions to the free equation of motion, the steady state solution has nothing to do with the initial values of the displacement and velocity. It is determined entirely by the driving force, (2.17). You will explore the transient solutions in problem (2.4).

Putting (2.19) and (2.17) into (2.16) and cancelling a factor of $e^{-i\omega_d t}$ from each side of the resulting equation, we get

$$(-\omega_d^2 - i\Gamma\omega_d + \omega_0^2)\,\mathcal{A} = \frac{F_0}{m}\,, \tag{2.20}$$

or

$$\mathcal{A} = \frac{F_0/m}{\omega_0^2 - i\Gamma\omega_d - \omega_d^2}\,. \tag{2.21}$$

Notice that we got the solution just using algebra. This is the advantage of starting with the irreducible solution, (2.19).

The amplitude, (2.21), of the displacement is proportional to the amplitude of the driving force. This is just what we expect from linearity (see problem (2.2)). But the coefficient of proportionality is complex. To see what it looks like explicitly, multiply the numerator and denominator of the right-hand side of (2.21) by $\omega_0^2 + i\Gamma\omega_d - \omega_d^2$, to get the complex numbers into the numerator

$$\mathcal{A} = \frac{(\omega_0^2 + i\Gamma\omega_d - \omega_d^2)\,F_0/m}{\left(\omega_0^2 - \omega_d^2\right)^2 + \Gamma^2\omega_d^2}\,. \tag{2.22}$$

The complex number $\mathcal{A}$ can be written as $A + iB$, with A and B real:

$$A = \frac{(\omega_0^2 - \omega_d^2)\,F_0/m}{\left(\omega_0^2 - \omega_d^2\right)^2 + \Gamma^2\omega_d^2}\,; \tag{2.23}$$

$$B = \frac{\Gamma\omega_d\,F_0/m}{\left(\omega_0^2 - \omega_d^2\right)^2 + \Gamma^2\omega_d^2}\,. \tag{2.24}$$

Then the solution to the equation of motion for the real driving force, (2.14), is

$$x(t) = \mathrm{Re}\, z(t) = \mathrm{Re}\left(\mathcal{A}e^{-i\omega_d t}\right) = A\cos\omega_d t + B\sin\omega_d t\,. \tag{2.25}$$

Thus the solution for the real force is a sum of two terms. The term proportional to A is in phase with the driving force (or 180° out of phase), while the term proportional to B is 90° out of phase. The advantage of going to the complex driving force is that it allows us to get both at once. The coefficients, A and B, are shown in the graph in figure 2.4 for $\Gamma = \omega_0/2$.

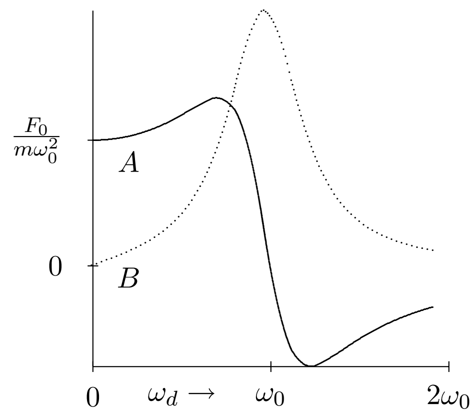


Figure 2.4: The elastic and absorptive amplitudes, plotted versus ω_d. The absorptive amplitude is the dotted line.

The real part of $\mathcal{A}$, $A = \mathrm{Re}\,\mathcal{A}$, is called the elastic amplitude and **the imaginary part of $\mathcal{A}$, $B = \mathrm{Im}\,\mathcal{A}$, is called the absorptive amplitude.** The reason for these names will become apparent below, when we consider the work done by the driving force.

2.3 Resonance

The $\left(\omega_0^2 - \omega_d^2\right)^2$ term in the denominator of (2.22) goes to zero for $\omega_d = \omega_0$. If the damping is small, this behavior of the denominator gives rise to a huge increase in the response of the system to the driving force at $\omega_d = \omega_0$. The phenomenon is called resonance. The angular frequency ω_0 is the resonant angular frequency. When $\omega_d = \omega_0$, the system is said to be "on resonance".

The phenomenon of resonance is both familiar and spectacularly important. It is familiar in situations as simple as building up a large amplitude in a child's swing by supplying a small force at the same time in each cycle. Yet simple as it is, it is crucial in many devices and many delicate experiments in physics. Resonance phenomena are used ubiquitously to build up a large, measurable response to a very small disturbance.

Very often, we will ignore damping in forced oscillations. Near a resonance, this is not a good idea, because the amplitude, (2.22), goes to infinity as $\Gamma \to 0$ for $\omega_d = \omega_0$. **Infinities are not physical.** This infinity never occurs in practice. One of two things happen before the amplitude blows up. Either the damping eventually cannot be ignored, so the response looks like (2.22) for nonzero Γ, or the amplitude gets so large that the nonlinearities in the system cannot be ignored, so the equation of motion no longer looks like (2.16).

2.3.1 Work

It is instructive to consider the work done by the external force in (2.16). **To do this we must use the real force, (2.14), and the real displacement (2.25), rather than their complex extensions, because, unlike almost everything else we talk about, the work is a nonlinear function of the force.** The power expended by the force is the product of the driving force and the velocity,

$$P(t) = F(t)\,\frac{\partial}{\partial t}x(t) = -F_0\omega_d A\cos\omega_d t\sin\omega_d t + F_0\omega_d B\cos^2\omega_d t\,. \tag{2.26}$$

The first term in (2.26) is proportional to $\sin 2\omega_d t$. Thus it is sometimes positive and sometimes negative. It averages to zero over any complete half-period of oscillation, a time π/ω_d, because

$$\int_{t_0}^{t_0+\pi/\omega_d} dt\,\sin 2\omega_d t = -\frac{1}{2}\,\cos 2\omega_d t|_{t_0}^{t_0+\pi/\omega_d} = 0\,. \tag{2.27}$$

This is why A is called the elastic amplitude. If A dominates, then energy fed into the system at one time is returned at a later time, as in an elastic collision in mechanics.

The second term in (2.26), on the other hand, is always positive. It averages to

$$P_{\text{average}} = \frac{1}{2}F_0\omega_d B\,. \tag{2.28}$$

This is why B is called the absorptive amplitude. It measures how fast energy is absorbed by the system. The absorbed power, P_{average}, reaches a maximum on resonance, at $\omega_0 = \omega_d$. This is a diagnostic that is often used to find resonances in experimental situations. Note that the dependence of B on ω_d looks qualitatively similar to that of P_{average}, which is shown in figure 2.5 for $\Gamma = \omega_0/2$. However, they differ by a factor of ω_d. In particular, the maximum of B occurs slightly below resonance.

2.3.2 Resonance Width and Lifetime

Both the height and the width of the resonance curve in figure 2.5 are determined by the frictional term, Γ, in the equation of motion. The maximum average power is inversely proportional to Γ,

$$\frac{F_0^2}{2m\Gamma}\,. \tag{2.29}$$

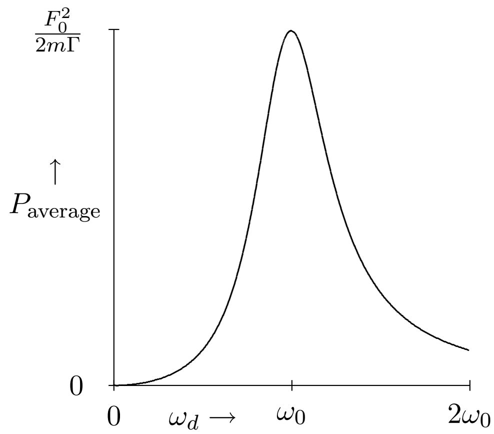


Figure 2.5: The average power lost to the frictional force as a function of ω_d for $\Gamma = \omega_0/2$.

The width (for fixed height) is determined by the ratio of Γ to ω_0. In fact, you can check that the values of ω_d for which the average power loss is half its maximum value are

$$\omega_{1/2} = \sqrt{\omega_0^2 + \frac{\Gamma^2}{4}} \pm \frac{\Gamma}{2}\,. \tag{2.30}$$

The Γ is the "full width at half-maximum" of the power curve. In figure 2.6 and figure 2.7, we show the average power as a function of ω_d for $\Gamma = \omega_0/4$ and $\Gamma = \omega_0$. The linear dependence of the width on Γ is clearly visible. The dotted lines show the position of half-maximum.

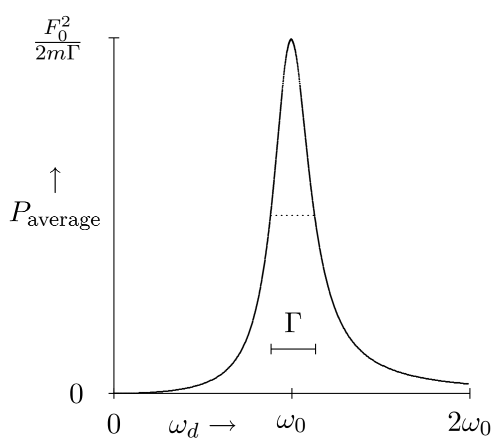


Figure 2.6: The average power lost to the frictional force as a function of ω_d for $\Gamma = \omega_0/4$.

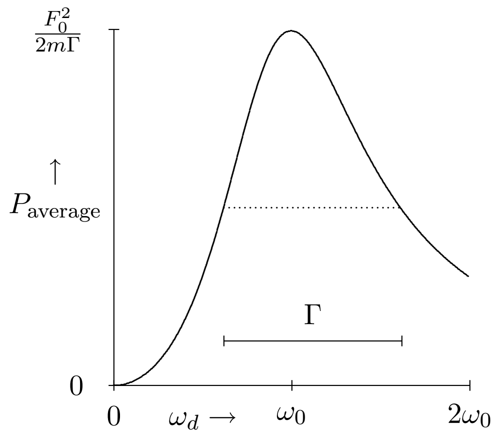


Figure 2.7: The average power lost to the frictional force as a function of ω_d for $\Gamma = \omega_0$.

This relation is even more interesting in view of the relationship between Γ and the time dependence of the free oscillation. The lifetime of the state in free oscillation is of order $1/\Gamma$. In other words, the width of the resonance peak in forced oscillation is inversely proportional to the lifetime of the corresponding normal mode of free oscillation. This inverse relation is important in many fields of physics. An extreme example is particle physics, where very short-lived particles can be described as resonances. The quantum mechanical waves associated with these particles have angular frequencies proportional to their energies,

$$E = \hbar\omega \tag{2.31}$$

where $\hbar$ is Planck's constant divided by 2π,

$$\hbar \approx 6.626 \times 10^{-34}\,\mathrm{J\,s}\,. \tag{2.32}$$

The lifetimes of these particles, some as short as 10^{-24} seconds, are far too short to measure directly. However, the short lifetime shows up in the large width of the distribution of energies of these states. That is how the lifetimes are actually inferred.

2.3.3 Phase Lag

We can also write (2.25) as

$$x(t) = R\cos(\omega_d t - \theta) \tag{2.33}$$

for

$$R = \sqrt{A^2 + B^2}\,, \quad \theta = \arg(A + iB)\,. \tag{2.34}$$

The phase angle, θ, measures the **phase lag** between the external force and the system's response. The actual time lag is θ/ω_d. The displacement reaches its maximum a time θ/ω_d **after** the force reaches its maximum.

Note that as the frequency increases, θ increases and the motion lags farther and farther behind the external force. The phase angle, θ, is determined by the relative importance of the restoring force and the inertia of the oscillator. At low frequencies (compared to ω_0), inertia (an imprecise word for the ma term in the equation of motion) is almost irrelevant because things are moving very slowly, and the motion is very nearly in phase with the force. Far beyond resonance, the inertia dominates. The mass can no longer keep up with the restoring force and the motion is nearly 180° out of phase with the force. We will work out a detailed example of this in the next section.

The phase lag goes through $\pi/2$ at resonance, as shown in the graph in figure 2.8 for $\Gamma = \omega_0/2$. A phase lag of $\pi/2$ is another frequently used diagnostic for resonance.

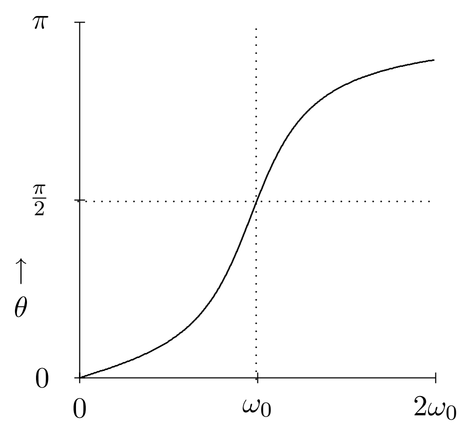


Figure 2.8: A plot of the phase lag versus frequency in a damped forced oscillator.

2.4 An Example

2.4.1 Feeling It In Your Bones

2-1

We will discuss the physics of forced oscillations further in the context of the simple system shown in figure 2.9. The block has mass m. The block moves in a viscous fluid that provides a frictional force. We will imagine that the fluid is something like a thick silicone oil, so that the steady state solution is reached very quickly. The block is attached to a cord that runs over a pulley and is attached to a spring, as shown. The spring has spring constant K. You

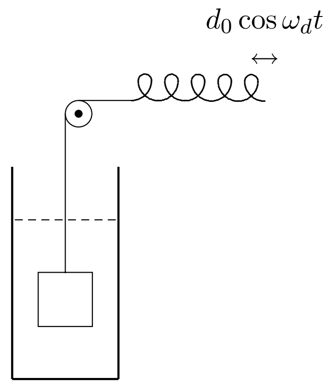


Figure 2.9: An oscillator that is damped by moving in a viscous fluid.

hold on to the other end of the spring and move it back and forth with displacement

$$d_0 \cos \omega_d t \, . \tag{2.35}$$

In this arrangement, you don't have to be in the viscous fluid with the block — this makes it a lot easier to breathe.

The question is, how does the block move? This system actually has exactly the equation of motion of the forced, damped oscillator. To see this, note that the change in the length of the spring from its equilibrium length is the difference,

$$x(t) - d_0 \cos \omega_d t \, . \tag{2.36}$$

Thus the equation motion looks like this:

$$m \frac{d^2}{dt^2} x(t) + m\Gamma \frac{d}{dt} x(t) = -K \left[x(t) - d_0 \cos \omega_d t \right] \, . \tag{2.37}$$

Dividing by m and rearranging terms, you can see that this is identical to (2.14) with

$$F_0/m = K \, d_0/m = \omega_0^2 d_0 \, . \tag{2.38}$$

Moving the other end of the spring sinusoidally effectively produces a sinusoidally varying force on the mass.

Now we will go over the solution again, stressing the physics of this system as we go. Try to imagine yourself actually doing the experiment! It will help to try to feel the forces involved in your bones. It may help to check out program 2-1 on the supplementary programs disk. This allows you to see the effect, but you should really try to **feel** it!

The first step is to go over to the complex force, as in (2.16). The result looks like

$$\overbrace{\frac{d^2}{dt^2}\,z(t)}^{\text{inertial}}+\overbrace{\Gamma\,\frac{d}{dt}\,z(t)}^{\text{frictional}}+\overbrace{\omega_0^2\,z(t)}^{\text{spring}}=\overbrace{\omega_0^2 d_0\,e^{-i\omega_d t}}^{\text{driving}}\,. \tag{2.39}$$

We have labeled the terms in (2.39) to remind you of their different physical origins.

The next step is to look for irreducible steady state solutions of the form of (2.19):

$$z(t)=\mathcal{A}\,e^{-i\omega_d t}\,. \tag{2.40}$$

Inserting (2.40) into (2.39), we get

$$\left[-\omega_d^2-i\Gamma\omega_d+\omega_0^2\right]\,\mathcal{A}\,e^{-i\omega_d t}=\omega_0^2 d_0\,e^{-i\omega_d t}\,. \tag{2.41}$$

What we will discuss in detail is the phase of the quantity in square brackets on the left-hand side of (2.41). Each of the three terms, inertial, frictional and spring, has a different phase. Each term also depends on the angular frequency, ω_d in a different way. The phase of $\mathcal{A}$ depends on which term dominates.

For very small ω_d, in particular for

$$\omega_d \ll \omega_0,\Gamma\,, \tag{2.42}$$

the spring term dominates the sum. Then $\mathcal{A}$ is in phase with the driving force. This has a simple physical interpretation. If you move the end of the spring slowly enough, both friction and inertia are irrelevant. When the block is moving very slowly, a vanishingly small force is required. The block just follows along with the displacement of the end of the spring, $\mathcal{A}\approx d_0$. You should be able to feel this dependence in your bones. If you move your hand very slowly, the mass has no trouble keeping up with you.

For very large ω_d, that is for

$$\omega_d \gg \omega_0,\Gamma\,, \tag{2.43}$$

the inertial term dominates the sum. The displacement is then 180° out of phase with the driving force. It also gets smaller and smaller as ω_d increases, going like

$$\mathcal{A}\approx-\frac{\omega_0^2}{\omega_d^2}\,d_0\,. \tag{2.44}$$

Again, this makes sense physically. When the angular frequency of the driving force gets very large, the mass just doesn't have time to move.

In between, at least two of the three terms on the left-hand side of (2.41) contribute significantly to the sum. At resonance, the inertial term exactly cancels the spring term, leaving only the frictional term, so that the displacement is 90° out of phase with the driving force. The size of the damping force determines how sharp the resonance is. If Γ is much smaller than ω_0, then the cancellation between the inertial and spring terms in (2.39) must be very precise in order for the frictional term to dominate. In this case, the resonance is very sharp. On the other hand, if $\Gamma \gg \omega_0$, the resonance is very broad, and the enhancement at resonance is not very large, because the frictional term dominates for a large range of ω_d around the point of resonance, $\omega_d = \omega_0$.

Try it! There is no substitute for actually doing this experiment. It will really give you a feel for what resonance is all about. Start by moving your hand at a very low frequency, so that the block stays in phase with the motion of your hand. Then very gradually increase the frequency. If you change the frequency slowly enough, the contributions from the transient free oscillation will be small, and you will stay near the steady state solution. As the frequency increases, you will first see that because of friction, the block starts to lag behind your hand. As you go through resonance, this lag will increase and go through 90°. Finally at very high frequency, the block will be 180° out of phase with your hand and its displacement (the amplitude of its motion) will be very small.

Chapter Checklist

You should now be able to:

1. Solve for the free motion of the damped harmonic oscillator by looking for the irreducible complex exponential solutions;

2. Find the steady state solution for the damped harmonic oscillator with a harmonic driving term by studying a corresponding problem with a complex exponential force and finding the irreducible complex exponential solution;

3. Calculate the power lost to frictional forces and the phase lag in the forced harmonic oscillator;

4. Feel it in your bones!

Problems

2.1. Prove that an overdamped oscillator can cross its equilibrium position at most once.

2.2. Prove, just using linearity, without using the explicit solution, that the steady state solution to (2.16) must be proportional to F_0.

2.3. For the system with equation of motion (2.14), suppose that the driving force has the form

$$f_0 \cos \omega_0 t \cos \delta t$$

where

$$\delta \ll \omega_0 \quad \text{and} \quad \Gamma = 0\,.$$

As $\delta \to 0$, this goes on resonance. What is the displacement for δ nonzero to **leading order in** δ/ω_0**?** Write the result in the form

$$\alpha(t)\ \cos \omega_0 t + \beta(t)\ \sin \omega_0 t$$

and find $\alpha(t)$ and $\beta(t)$. Discuss the physics of this result. **Hint:** First show that

$$\cos \omega_0 t \cos \delta t = \frac{1}{2} \operatorname{Re} \left(e^{-i(\omega_0+\delta)t} + e^{-i(\omega_0-\delta)t} \right) .$$

2.4. For the system shown in figure 2.9, suppose that the displacement of the end of the wire vanishes for $t < 0$, and has the form

$$d_0 \sin \omega_d t \quad \text{for} \quad t \geq 0\,.$$

a. Find the displacement of the block for $t > 0$. Write the solution as the real part of complex solution, by using a complex force and exponential solutions. Do not try to simplify the complex numbers. **Hint:** Use (2.23), (2.24) and (2.6). If you get confused, go on to part **b.**

b. Find the solution when $\Gamma \to 0$ and simplify the result. Even if you got confused by the complex numbers in **a.**, you should be able to find the solution in this limit. When there is no damping, the "transient" solutions do not die away with time!

2.5. For the LC circuit shown in figure 1.10, suppose that the inductor has nonzero resistance, R. Write down the equation of motion for this system and find the relation between friction term, $m\Gamma$, in the damped harmonic oscillator and the resistance, R, that completes the correspondence of (1.105). Suppose that the capacitors have capacitance, $C \approx 0.00667\mu F$, the inductor has inductance, $L \approx 150\mu H$ and the resistance, $R \approx 15\Omega$. Solve the equation of motion and evaluate the constants that appear in your solution in units of seconds.

Chapter 3

Normal Modes

Systems with several degrees of freedom appear to be much more complicated than the simple harmonic oscillator. What we will see in this chapter is that this is an illusion. When we look at it in the right way, we can see the simple oscillators inside the more complicated system.

Preview

In this chapter, we discuss harmonic oscillation in systems with more than one degree of freedom.

1. We will write down the equations of motion for a system of particles moving under general linear restoring forces without damping.

2. Next, we introduce matrices and matrix multiplication and show how they can be used to simplify the description of the equations of motion derived in the previous section.

3. We will then use time translation invariance and find the irreducible solutions to the equations of motion in matrix form. This will lead to the idea of "normal modes." We then show how to put the normal modes together to construct the general solution to the equations of motion.

4. ***** We will introduce the idea of "normal coordinates" and show how they can be used to automate the solution to the initial value problem.

5. ***** We will discuss damped forced oscillation in systems with many degrees of freedom.

3.1 More than One Degree of Freedom

In general, the number of degrees of freedom of a system is the number of independent coordinates required to specify the system's configuration. The more degrees of freedom the system has, the larger the number of independent ways that the system can move. The more possible motions, you might think, the more complicated the system will be to analyze. In fact, however, using the tools of linear algebra, we will see that we can deal with systems with many degrees of freedom in a straightforward way.

3.1.1 Two Coupled Oscillators

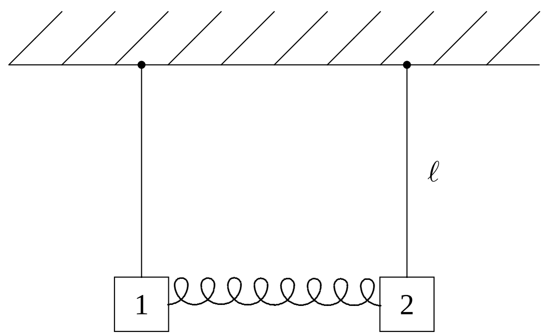


Figure 3.1: Two pendulums coupled by a spring.

Consider the system of two pendulums shown in figure 3.1. The pendulums consist of rigid rods pivoted at the top so they oscillate without friction in the plane of the paper. The masses at the ends of the rods are coupled by a spring. We will consider the free motion of the system, with no external forces other than gravity. This is a classic example of two "coupled oscillators." The spring that connects the two oscillators is the coupling. We will assume that the spring in figure 3.1 is unstretched when the two pendulums are hanging straight down, as shown. Then the equilibrium configuration is that shown in figure 3.1. This is an example of a system with two degrees of freedom, because two quantities, the displacements of each of the two blocks from equilibrium, are required to specify the configuration of the system. For example, if the oscillations are small, we can specify the configuration by giving the horizontal displacement of each of the two blocks from the equilibrium position.

Suppose that block 1 has mass m_1, block 2 has mass m_2, both pendulums have length ℓ and the spring constant is κ (Greek letter kappa). Label the (small) horizontal displacements of the blocks to the right, x_1 and x_2, as shown in figure 3.2. We could have called these

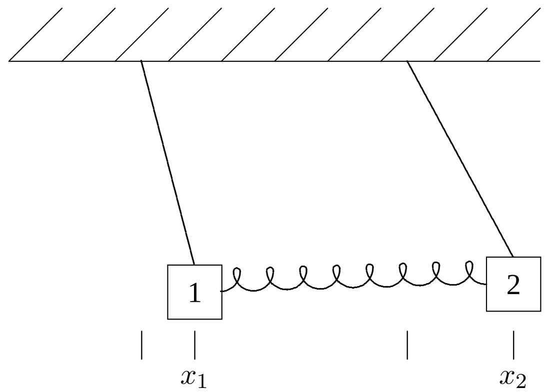


Figure 3.2: Two pendulums coupled by a spring displaced from their equilibrium positions.

masses and displacements anything, but it is very convenient to use the same symbol, x, with different subscripts. We can then write Newton's law, $F = ma$, in a compact and useful form.

$$m_j \frac{d^2}{dt^2} x_j = F_j \, , \tag{3.1}$$

for $j = 1$ to 2, where F_1 is the horizontal force on block 1 and F_2 is the horizontal force on block 2. Because there are two values of j, (3.1) is **two equations**; one for $j = 1$ and another for $j = 2$. These are the two equations of motion for the system with two degrees of freedom. We will often refer to all the masses, displacements or forces at once as m_j, x_j or F_j, respectively. For example, we will say that F_j is the horizontal force on the jth block. This is an example of the use of "indices" (j is an index) to simplify the description of a system with more than one degree of freedom.

When the blocks move horizontally, they will move vertically as well, because the length of the pendulums remains fixed. Because the vertical displacement is second order in the x_js,

$$y_j \approx \frac{x_j^2}{2} \, , \tag{3.2}$$

we can ignore it in thinking about the spring. The spring stays approximately horizontal for small oscillations.

To find the equation of motion for this system, we must find the forces, F_j, in terms of the displacements, x_j. It is the approximate linearity of the system that allows us to do this in a useful way. The forces produced by the Hooke's law spring, and the horizontal forces on the pendulums due to the tension in the string (which in turn is due to gravity) are both approximately linear functions of the displacements for small displacements. Furthermore, the forces vanish when both the displacements vanish, because the system is in equilibrium.

Thus each of the forces is some constant (different for each block) times x_1 plus some other constant times x_2. It is convenient to write this as follows:

$$F_1 = -K_{11}x_1 - K_{12}x_2\,, \quad F_2 = -K_{21}x_1 - K_{22}x_2\,, \tag{3.3}$$

or more compactly,

$$F_j = -\sum_{k=1}^{2} K_{jk}x_k \tag{3.4}$$

for $j = 1$ to 2. We have written the four constants as K_{11}, K_{12}, K_{21} and K_{22} in order to write the force in this compact way. Later, we will call these constants the matrix elements of the K matrix. In this notation, the equations of motion are

$$m_j \frac{d^2}{dt^2} x_j = -\sum_{k=1}^{2} K_{jk}x_k \tag{3.5}$$

for $j = 1$ to 2.

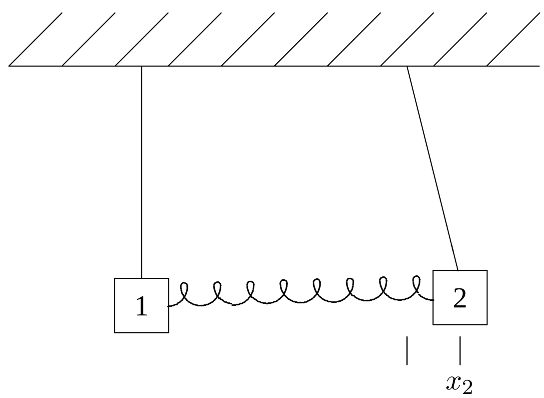


Figure 3.3: Two pendulums coupled by a spring with block 2 displaced from an equilibrium position.

Because of the linearity of the system, we can find the constants, K_{jk}, by considering the displacements of the blocks one at a time. Then we find the total force using (3.4). For example, suppose we displace block 2 with block 1 held fixed in its equilibrium position and look at the forces on both blocks. This will allow us to compute K_{12} and K_{22}. The system with block two displaced is shown in figure 3.3. The forces on the blocks are shown in figure 3.4, where T_j is the tension in the jth pendulum string. F_{12} is the force on block 1 due to the displacement of block 2. F_{22} is the force on block 2 due to the displacement of block 2. For small displacements, the restoring force from the spring is nearly horizontal and equal to

κx_2 on block 1 and $-\kappa x_2$ on block 2. Likewise, in the limit of small displacement, the vertical component of the force from the tension T_2 nearly cancels the gravitational force on block 2, $m_2 g$, so that the horizontal component of the tension gives a restoring force $-x_2 m_2 g/\ell$ on block 2. For block 1, the force from the tension T_1 just cancels the gravitational force $m_1 g$. Thus

$$F_{12} \approx \kappa x_2 \,, \quad F_{22} \approx -\frac{m_2 g x_2}{\ell} - \kappa x_2 \,, \tag{3.6}$$

and

$$K_{12} \approx -\kappa \,, \quad K_{22} \approx \frac{m_2 g}{\ell} + \kappa \,. \tag{3.7}$$

An analogous argument shows that

$$K_{21} \approx -\kappa \,, \quad K_{11} \approx \frac{m_1 g}{\ell} + \kappa \,. \tag{3.8}$$

Notice that

$$K_{12} = K_{21} \,. \tag{3.9}$$

We will see below that this is an example of a very general relation.

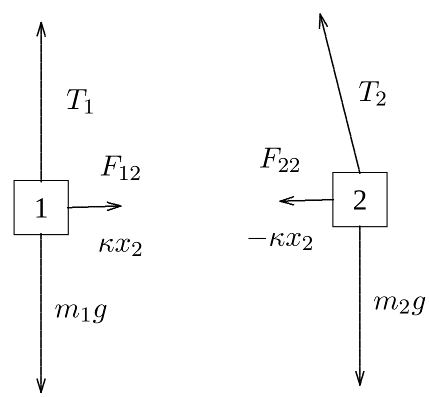


Figure 3.4: The forces on the two blocks in figure 3.3.

3.1.2 Linearity and Normal Modes

3-1

We will see in this chapter that the most general possible motion of this system, and of any such system of oscillators, can be decomposed into particularly simple solutions, in which all the degrees of freedom oscillate with the same frequency. These simple solutions are called "normal modes." The displacements for the most general motion can be written as sums of the simple solutions. We will study how this works in detail later, but it may be useful to see it first. A possible motion of the system of two coupled oscillators is animated in program 3-1. Below the actual motion, we show the two simple motions into which the more complicated motion can be decomposed. For this system, the normal mode with the lower frequency is one in which the displacements of the two blocks are the same:

$$x_1(t) = x_2(t) = b_1 \cos(\omega_1 t - \theta_1)\,. \tag{3.10}$$

The other normal mode is one in which the displacements of the two blocks are opposite

$$x_1(t) = -x_2(t) = b_2 \cos(\omega_2 t - \theta_2)\,. \tag{3.11}$$

The sum of these two simple motions gives the much more complicated motion shown in program 3-1.

3.1.3 n Coupled Oscillators

Before we try to solve the equations of motion, (3.5), let us generalize the discussion to systems with more degrees of freedom. Consider the oscillation of a system of n particles connected by various springs with no damping. Our analysis will be completely general, but for simplicity, we will talk about the particles as if they are constrained to move in the x direction, so that we can measure the displacement of the jth particle from equilibrium with the coordinate x_j. Then the equilibrium configuration is the one in which all the x_js are all zero.

Newton's law, $F = ma$, for the motion of the system gives

$$m_j \, \frac{d^2 x_j}{dt^2} = F_j \tag{3.12}$$

where m_j is the mass of the jth particle, F_j is the force on it. Because the system is linear, we expect that we can write the force as follows (as in (3.4)):

$$F_j = -\sum_{k=1}^{n} K_{jk}\, x_k \tag{3.13}$$

for $j = 1$ to n. The constant, $-K_{jk}$, is the force per unit displacement of the jth particle due to a displacement x_k of the kth particle. Note that all the F_js vanish at equilibrium when all the x_js are zero. Thus the equations of motion are

$$m_j \, \frac{d^2 x_j}{dt^2} = -\sum_{k} K_{jk}\, x_k \tag{3.14}$$

for $j = 1$ to n.

To measure K_{jk}, make a small displacement, x_k, of the kth particle, keeping all the other particles fixed at zero, assumed to be an equilibrium position. Then measure the force, F_{jk} on the jth particle with only the kth particle displaced. Since the system is linear (because it is made out of springs or in general, as long as the displacement is small enough), the force is proportional to the displacement, x_k. The ratio of F_{jk} to x_k is $-K_{jk}$:

$$K_{jk} = -F_{jk}/x_k \quad \text{when} \quad x_\ell = 0 \quad \text{for} \quad \ell \neq k \,. \tag{3.15}$$

Note that K_{jk} is defined with a $-$ sign, so that a positive K is a force that is opposite to the displacement, and therefore tends to return the system to equilibrium.

Because the system is linear, the total force due to an arbitrary displacement is the sum of the contributions from each displacement. Thus

$$F_j = \sum_k F_{jk} = -\sum_k K_{jk} x_k \,. \tag{3.16}$$

Let us now try to understand (3.9). If we consider systems with no damping, the forces can be derived from a potential energy,

$$F_j = -\frac{\partial V}{\partial x_j} \,. \tag{3.17}$$

But then by differentiating equation (3.16) we find that

$$K_{jk} = \frac{\partial^2 V}{\partial x_j \partial x_k} \,. \tag{3.18}$$

The partial differentiations commute with one another, thus equation (3.18) implies

$$K_{jk} = K_{kj} \,. \tag{3.19}$$

In words, the force on particle j due to a displacement of particle k is equal to the force on particle k due to the displacement of particle j.

3.2 Matrices

It is very useful to rewrite equation (3.14) in a matrix notation. Because of the linearity of the equations of motion for harmonic motion, it will be very useful to have the tools of linear algebra at hand for our study of wave phenomena. If you haven't studied linear algebra (or didn't understand much of it) in math courses, **DON'T PANIC**. We will start from scratch by describing the properties of matrices and matrix multiplication. The important thing to keep in mind is that matrices are nothing very deep or magical. They are just bookkeeping devices designed to make your life easier when you deal with more than one equation at a time.

A matrix is a rectangular array of numbers. An $N \times M$ matrix has N rows and M columns. Matrices can be added and subtracted simply by adding and subtracting each of the components. The difference comes in multiplication. It is very convenient to define a multiplication law that defines the product of an $N \times M$ matrix on the left with a $M \times L$ matrix on the right (the order is important!) to be an $N \times L$ matrix as follows:
Call the $N \times M$ matrix A and let A_{jk} be the number in the jth row and kth column for $1 \leq j \leq N$ and $1 \leq k \leq M$. These individual components of the matrix are called matrix elements. In terms of its matrix elements, the matrix A looks like:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NM} \end{pmatrix} . \tag{3.20}$$

Call the $M \times L$ matrix B with matrix elements B_{kl} for $1 \leq k \leq M$ and $1 \leq l \leq L$:

$$B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1L} \\ B_{21} & B_{22} & \cdots & B_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ B_{M1} & B_{M2} & \cdots & B_{ML} \end{pmatrix} . \tag{3.21}$$

Call the $N \times L$ matrix C with matrix elements C_{jl} for $1 \leq j \leq N$ and $1 \leq l \leq L$.

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1L} \\ C_{21} & C_{22} & \cdots & C_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \cdots & C_{NL} \end{pmatrix} . \tag{3.22}$$

Then the matrix C is defined to be the product matrix $A\,B$ if

$$C_{jl} = \sum_{k=1}^{M} A_{jk} \cdot B_{kl} \, . \tag{3.23}$$

Equation (3.23) is the algebraic statement of the “row-column” rule. To compute the $j\ell$ matrix element of the product matrix, AB, take the jth row of the matrix A and the ℓth column of the matrix B and form their dot-product (corresponding to the sum over k in (3.23)). This rule is illustrated below:

$$\begin{pmatrix} A_{11} & \cdots & A_{1k} & \cdots & A_{1M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{j1} & \cdots & A_{jk} & \cdots & A_{jM} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{Nk} & \cdots & A_{NM} \end{pmatrix} \begin{pmatrix} B_{11} & \cdots & B_{1\ell} & \cdots & B_{1L} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{k1} & \cdots & B_{k\ell} & \cdots & B_{kL} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{M1} & \cdots & B_{M\ell} & \cdots & B_{ML} \end{pmatrix}$$

$$= \begin{pmatrix} C_{11} & \cdots & C_{1\ell} & \cdots & C_{1L} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{j1} & \cdots & \boxed{C_{j\ell}} & \cdots & C_{jL} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{N\ell} & \cdots & C_{NL} \end{pmatrix} . \tag{3.24}$$

For example,

$$\begin{pmatrix} 2 & 3 \\ 0 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 13 \\ 0 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix} . \tag{3.25}$$

It is easy to check that the matrix product defined in this way is associative, $(AB)C = A(BC)$. However, in general, it is not commutative, $AB \neq BA$. In fact, if the matrices are not square, the product in the opposite order may not even make any sense! The matrix product AB only makes sense if the number of columns of A is the same as the number of rows of B. Beware!

Except for the fact that it is not commutative, matrix multiplication behaves very much like ordinary multiplication. For example, there are "identity" matrices. The $N \times N$ identity matrix, called I, has zeros everywhere except for 1's down the diagonal. For example, the 3×3 identity matrix is

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{3.26}$$

The $N \times N$ identity matrix satisfies

$$\begin{aligned} &IA = AI = A \text{ for any } N \times N \text{ matrix } A; \\ &IB = B \text{ for any } N \times M \text{ matrix } B; \\ &CI = C \text{ for any } M \times N \text{ matrix } C. \end{aligned} \tag{3.27}$$

We will be primarily concerned with "square" (that is $N \times N$) matrices.

Matrices allow us to deal with many linear equations at the same time.

An N dimensional column vector can be regarded as an $N \times 1$ matrix. We will call this object an "N-vector." It should not be confused with a coordinate vector in three-dimensional space. Likewise, we can think of an N dimensional row vector as a $1 \times N$ matrix. Matrix multiplication can also describe the product of a matrix with a vector to give a vector. The particularly important case that we will need in order to analyze wave phenomena involves square matrices. Consider an $N \times N$ matrix A multiplying an N-vector, X, to give another N-vector, F. The square matrix A has N^2 matrix elements, A_{jk} for j and k = 1 to N. The vectors X and F each have N matrix elements, just their components X_j and F_j for $j = 1$ to N. Then the matrix equation:

$$A\,X = F \tag{3.28}$$

actually stands for N equations:

$$\sum_{k=1}^{N} A_{jk} \cdot X_k = F_j \tag{3.29}$$

for j=1 to N. In other words, these are N simultaneous linear equations for the N X_j's. You all know, from your studies of algebra how to solve for the X_j's in terms of the F_j's and the A_{jk}'s but it is very useful to do it in matrix notation. Sometimes, we can find the "inverse" of the matrix A, A^{-1}, which has the property

$$A\,A^{-1} = A^{-1}A = I\,, \tag{3.30}$$

where I is the identity matrix discussed in (3.26) and (3.27). If we can find such a matrix, then the N simultaneous linear equations, (3.29), have a unique solution that we can write in a very compact form. Multiply both sides of (3.29) by A^{-1}. On the left-hand side, we can use (3.30) and (3.27) to get rid of the $A^{-1}A$ and write the solution as follows:

$$X = A^{-1}F\,. \tag{3.31}$$

3.2.1 * Inverse and Determinant

We can compute A^{-1} in terms of the "determinant" of A. The determinant of the matrix A is a sum of products of the matrix elements of A with the following properties:

- There are $N!$ terms in the sum;
- Each term in the sum is a product of N different matrix elements;
- In each product, every row number and every column number appears exactly once;
- Every such product can be obtained from the product of the diagonal elements, $A_{11}A_{22}\cdots A_{NN}$, by a sequence of interchanges of the column labels. For example, $A_{12}A_{21}A_{33}\cdots A_{NN}$ involves one interchange while $A_{12}A_{23}A_{31}A_{44}\cdots A_{NN}$ requires two.
- The coefficient of a product in the determinant is +1 if it involves an even number of interchanges and -1 if it involves an odd number of interchanges.

Thus the determinant of a 2×2 matrix, A is

$$\det A = A_{11}A_{22} - A_{12}A_{21}\,. \tag{3.32}$$

The determinant of a 3×3 matrix, A is

$$\begin{aligned}\det A = A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} \\ -A_{11}A_{23}A_{32} - A_{13}A_{22}A_{31} - A_{12}A_{21}A_{33}\,.\end{aligned} \tag{3.33}$$

Unless you are very unlucky, you will never have to compute the determinant of a matrix larger than 3×3 by hand. If you are so unlucky, it is best to use an inductive procedure that builds it up from the determinants of smaller submatrices. We will discuss this procedure below.

If $\det A = 0$, the matrix has no inverse. It is not “invertible.” In this case, the simultaneous linear equations have either no solution at all, or an infinite number of solutions. If $\det A \neq 0$, the inverse matrix exists and is uniquely given by

$$A^{-1} = \frac{\tilde{A}}{\det A} \tag{3.34}$$

where $\tilde{A}$ is the **cofactor** matrix defined by its matrix elements as follows:

$$(\tilde{A})_{jk} = \det A(jk) \tag{3.35}$$

with

$A(jk)_{lm} = 1$ if $m = j$ and $l = k$;
$A(jk)_{lm} = 0$ if $m = j$ and $l \neq k$;
$A(jk)_{lm} = 0$ if $m \neq j$ and $l = k$;
$A(jk)_{lm} = A_{lm}$ if $m \neq j$ and $l \neq k$.

In other words, $A(jk)$ is obtained from the matrix A by replacing the kj matrix element by 1 and all other matrix elements in row k or column j by 0. Thus if

$$A = \begin{pmatrix} A_{11} & \cdots & \boxed{A_{1j}} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kj} & \cdots & A_{kN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{Nj} & \cdots & A_{NN} \end{pmatrix}, \tag{3.36}$$

$$A(jk) = \begin{pmatrix} A_{11} & \cdots & \boxed{0} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{N1} & \cdots & 0 & \cdots & A_{NN} \end{pmatrix} . \tag{3.37}$$

Note the sneaky interchange of $j \leftrightarrow k$ in this definition, compared to (3.23).

For example if

$$A = \begin{pmatrix} 4 & 3 \\ 5 & 2 \end{pmatrix} \tag{3.38}$$

then

$$\begin{aligned} A(11) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad A(12) = \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix} \\ A(21) = \begin{pmatrix} 0 & 1 \\ 5 & 0 \end{pmatrix} \quad A(22) = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} . \end{aligned} \tag{3.39}$$

Thus,

$$\tilde{A} = \begin{pmatrix} 2 & -3 \\ -5 & 4 \end{pmatrix} \tag{3.40}$$

and since $\det A = 4 \cdot 2 - 5 \cdot 3 = -7$,

$$A^{-1} = \begin{pmatrix} -2/7 & 3/7 \\ 5/7 & -4/7 \end{pmatrix} . \tag{3.41}$$

A^{-1} satisfies $A\,A^{-1} = A^{-1}\,A = I$ where I is the identity matrix:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.42}$$

In terms of the submatrices, $A(jk)$, we can define the determinant inductively, as promised above. In fact, the reason that (3.30) works is that the determinant can be written as

$$\det A = \sum_{k=1}^{N} A_{1k} \det A(k1) . \tag{3.43}$$

Actually this is true for any row, not just $j = 1$. The relation, (3.30) can be rewritten as

$$\sum_{k=1}^{N} A_{jk} \det A(kj') = \begin{cases} \det A \text{ for } j = j' \\ 0 \text{ for } j \neq j' \end{cases} \tag{3.44}$$

The determinants of the submatrices, $\det A(k1)$, in (3.43) can, in turn, be computed by the same procedure. The result is a definition of the determinant that refers to itself. However, eventually, the process terminates because the matrices keep getting smaller and the determinant can always be computed in this way. The only problem with this procedure is that it is very tedious for a large matrix. For an $n \times n$ matrix, you end up computing $n!$ terms and adding them up. For large n, this is impractical. One of the nice features of the techniques that we will discuss in the coming chapters is that we will be able to avoid such calculations.

3.2.2 More Useful Facts about Matrices

Suppose that A and B are $N \times N$ matrices and v is an N-vector.

1. If you know the inverses of A and B, you can find the inverse of the product, AB, by multiplying the inverses in the reverse order:
$$(A\,B)^{-1} = B^{-1}\,A^{-1}\,. \tag{3.45}$$
2. The determinant of the product, AB, is the product of the determinants:
$$\det(A\,B) = \det A\;\det B\,, \tag{3.46}$$
thus if $\det(AB) = 0$, then either A or B has vanishing determinant.
3. A matrix multiplying a nonzero vector can give zero only if the determinant of the matrix vanishes:
$$A\,v = 0 \;\Rightarrow\; \det A = 0 \;\text{ or }\; v = 0\,. \tag{3.47}$$
This is the statement, in matrix language, that N homogeneous linear equations in N unknowns can have a nontrivial solution, $v \neq 0$, **only if** the determinant of the coefficients vanishes.
4. Similarly, if $\det A = 0$, there exists a nonzero vector, v, that is annihilated by A:
$$\det A = 0 \;\Rightarrow\; \exists\, v \neq 0 \;\text{ such that } A\,v = 0\,. \tag{3.48}$$
This is the statement, in matrix language, that N homogeneous linear equations in N unknowns **actually do** have a nontrivial solution, $v \neq 0$, if the determinant of the coefficients vanishes.
5. The transpose of an $N \times M$ matrix A, denoted by A^T, is the $M \times N$ matrix obtained by reflecting the matrix about a diagonal line through the upper left-hand corner. Thus if
$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NM} \end{pmatrix} \tag{3.49}$$

then

$$A^T = \begin{pmatrix} A_{11} & A_{21} & \cdots & \cdots & A_{N1} \\ A_{12} & A_{22} & \cdots & \cdots & A_{N2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ A_{1M} & A_{2M} & \cdots & \cdots & A_{NM} \end{pmatrix} . \tag{3.50}$$

Note that if $N \neq M$, the shape of the matrix is changed by transposition. Only for square matrices does the transpose give you back a matrix of the same kind. A square matrix that is equal to its transpose is called a "symmetric" matrix.

3.2.3 Eigenvalue Equations

We will make extensive use of the concept of an "eigenvalue equation." For an $N \times N$ matrix, R, the eigenvalue equation has the form:

$$R\,c = h\,c\,, \tag{3.51}$$

where c is a **nonzero** N-vector,[1] and h is a number. The idea is to find both the number, h, which is called the eigenvalue, and the vector, c, which is called the eigenvector. This is the problem we discussed in chapter 1 in (1.78) in connection with time translation invariance, but now written in matrix form.

A couple of examples may be in order. Suppose that R is a diagonal matrix, like

$$R = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.52}$$

Then the eigenvalues are just the diagonal elements, 2 and 1, and the eigenvectors are vectors in the coordinate directions,

$$R \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \quad R \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{3.53}$$

A less obvious example is

$$R = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} . \tag{3.54}$$

This time the eigenvalues are 3 and 1, and the eigenvectors are as shown below:

$$R \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad R \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} . \tag{3.55}$$

It may seem odd that in the eigenvalue equation, both the eigenvalue **and** the eigenvector are unknowns. The reason that it works is that for most values of h, the equation, (3.51), has

[1] $c = 0$ doesn't count, because the equation is satisfied trivially for any h. We are interested only in nontrivial solutions.

no solution. To see this, we write (3.51) as a set of homogeneous linear equations for the components of the eigenvector, c,

$$(R - hI)\ c = 0\,. \tag{3.56}$$

The set of equations, (3.56), has nonzero solutions for c only if the determinant of the coefficient matrix, $R - hI$, vanishes. But this will happen only for N values of h, because the condition

$$\det\,(R - hI) = 0 \tag{3.57}$$

is an Nth order equation for h. For each h that solves (3.57), we can find a solution for c.[2] We will give some examples of this procedure below.

3.2.4 The Matrix Equation of Motion

It is very useful to rewrite the equation of motion, (3.14), in a matrix notation. Define a column vector, X, whose jth row (from the top) is the coordinate x_j:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\,. \tag{3.58}$$

Define the "K matrix", an $n\times n$ matrix that has the coefficient K_{jk} in its jth row and kth column:

$$K = \begin{pmatrix} K_{11} & K_{12} & \cdots & K_{1n} \\ K_{21} & K_{22} & \cdots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & K_{nn} \end{pmatrix}\,. \tag{3.59}$$

K_{jk} is said to be the "jk matrix element" of the K matrix. Because of equation (3.19), the matrix K is symmetric, $K = K^T$.

Define the diagonal matrix M with m_j in the jth row and jth column and zeroes elsewhere

$$M = \begin{pmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n \end{pmatrix}\,. \tag{3.60}$$

M is called the "mass matrix."

[2] The situation is slightly more complicated when the solutions for h are degenerate. We discuss this in (3.117) below.

Using these definitions, we can rewrite (3.14) in matrix notation as follows:

$$M\,\frac{d^2X}{dt^2} = -K\,X\,. \tag{3.61}$$

There is nothing very fancy going on here. We have just used the matrix notation to get rid of the summation sign in (3.14). The sum is now implicit in the matrix multiplication in (3.61). This is useful because we can now use the properties of matrices and matrix multiplication discussed above to manipulate (3.61). For example, we can simplify (3.61) a bit by multiplying on the left by M^{-1} to get

$$\frac{d^2X}{dt^2} = -M^{-1}K\,X\,. \tag{3.62}$$

3.3 Normal Modes

If there is only one degree of freedom, then both X and M^{-1} are just numbers and the solutions to the equation of motion, (3.62), have the form of a constant amplitude times an exponential factor. In fact, we saw that this form is related to a very general fact about the physics – time translation invariance, (1.33). The arguments of chapter 1, (1.71)-(1.85), did not depend on the number of degrees of freedom. Thus they show that here again, we can find irreducible solutions, that go into themselves up to an overall constant when the clocks are reset. As in chapter 1, the first step is to allow the solutions to be complex. That is, we replace (3.62) by

$$\frac{d^2Z}{dt^2} = -M^{-1}K\,Z\,, \tag{3.63}$$

where Z is a complex n vector with components, z_j. The real parts of the components of Z are the components of a real solution satisfying (3.62),

$$x_j = \operatorname{Re} z_j\,. \tag{3.64}$$

We will say that the real vector, X, is the real part of the complex vector, Z,

$$X = \operatorname{Re} Z\,, \tag{3.65}$$

if (3.64) is satisfied.

Just as in chapter 1, we know that we can find irreducible solutions that have the same form up to an overall constant when the clocks are reset. We know from (1.85) that these have the form

$$Z(t) = A\,e^{-i\omega t} \tag{3.66}$$

where A is some constant n-vector and the angular frequency, ω, is still just a number. Now if $t \to t + a$,

$$Z(t) \to Z(t+a) = e^{-i\omega a}Z(t)\,. \tag{3.67}$$

While the irreducible form, (3.66), comes just from time translation invariance, we must still look at the equations of motion to determine the vector, A and the angular frequency, ω. Inserting (3.66) into (3.63), doing the differentiation and canceling the exponential factors from both sides, we find that (3.66) is a solution if

$$\omega^2 A = M^{-1} K A . \tag{3.68}$$

This matrix equation is an eigenvalue equation of the form that we discussed in (3.51)-(3.57). ω^2 is the eigenvalue of the matrix $M^{-1}K$ and A is the corresponding eigenvector. Let us see what it means physically.

The real part of the column vector Z specifies the displacement of each of the degrees of freedom of the system. The eigenvalue equation, (3.68), does not involve any complex numbers (because we have not put in any damping). Therefore (as we will see explicitly below), we can choose the solutions so that all the components of A are real. Then the real part of the complex solutions we seek in (3.66) is

$$X(t) = A \cos \omega t , \tag{3.69}$$

or in terms of the components of A,

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} , \tag{3.70}$$

$$x_1(t) = a_1 \cos \omega t , \quad x_2(t) = a_2 \cos \omega t , \quad \text{etc.} \tag{3.71}$$

Not only does everything move with the same frequency, but the **ratios** of displacements of the individual degrees of freedom are fixed. Everything oscillates in phase. The only difference between the motion of the different degrees of freedom is their different amplitudes from the different components of A.

The point is worth repeating. Time translation invariance and linearity imply that we can **always** find irreducible solutions, (3.67), in which all the degrees of freedom oscillate with the same frequency. The extra piece of information that leads to (3.69) is dynamical. If there is no damping, then all the components of A can be chosen to be real, and all the degrees of freedom oscillate not only with the same frequency, but also with the same phase.

If such a solution is to satisfy the equations of motion, then the acceleration must also be proportional to A, so that the individual displacements don't get out of synch. But that is what (3.68) is telling us. $-M^{-1}K$ is the matrix that, acting on the displacement, gives the acceleration. The eigenvalue equation (3.68) means that the acceleration is proportional to A again. The constant of proportionality, ω^2, is the return force per unit displacement per unit mass for the particular displacement specified by A.

We have already discussed the mathematical structure of the eigenvalue equation in (3.51)-(3.57). We will do it again, for emphasis, in the case of physical interest, (3.68). It should be clear that not every value of A and ω^2 gives a solution of (3.68). We will solve for the allowed values by first finding the possible values of ω^2 and then finding the corresponding values of A. To find the eigenvalues, note that (3.68) can be rewritten as

$$\left[M^{-1}K - \omega^2 I\right] A = 0\,, \tag{3.72}$$

where I is the $n\times n$ identity matrix. (3.72) is just a compact way of representing n homogeneous linear equations in the n components of A where the coefficients depend on ω^2. We saw in (3.47) and (3.48) that for systems of n homogeneous linear equations in n unknowns, a nonzero solution exists if and only if the determinant of the coefficient matrix vanishes. The reason is that if the determinant were nonzero, then the matrix, $M^{-1}K - \omega^2 I$, would have an inverse, and we could use (3.31) to conclude that the only solution for the vector, A, is $A = 0$. Thus to have a nonzero amplitude, A, we must have

$$\det\left[M^{-1}K - \omega^2 I\right] = 0\,. \tag{3.73}$$

(3.73) is a polynomial equation for ω^2. It is an equation of degree n in ω^2, because the term in the determinant from the product of all the diagonal elements of the matrix contains a piece that goes as $\left[\omega^2\right]^n$. All the coefficients in the polynomial are real. Physically, we expect all the solutions for ω^2 to be real and positive whenever the system is in stable equilibrium because we expect such systems to oscillate. Mathematically, we can show that ω^2 is always real, so long as all the masses are positive. We will do this below in (3.127)-(3.130).

Negative ω^2 are associated with unstable equilibrium. For example, consider a mass at the end of a rigid rod, free to swing in the earth's gravitational field in a vertical plane around a frictionless pivot, as shown in figure 3.5. The mass can move along the dotted line. The stable equilibrium position is indicated by the solid line. The unstable equilibrium position is indicated by the dashed line.

When the mass is at the unstable equilibrium point, the smallest disturbance will cause it to fall. Once away from equilibrium, the displacement increases exponentially until the angle from the vertical becomes so large that the nonlinearities in the equation of motion for this system take over. We will discuss this nonlinear oscillator further in appendix B.

Once we have found the possible values of ω^2, we can put each one back into (3.72) to get the corresponding A. Because (3.72) is homogeneous, the overall scale of A is not determined, **but all the ratios, a_j/a_k, are fixed for each ω^2.**

3.3.1 Normal Modes and Frequencies

The vector A is called the "normal mode" of the system associated with the frequency ω. Because A is real, in the absence of friction, the complex solutions, (3.66), can be put

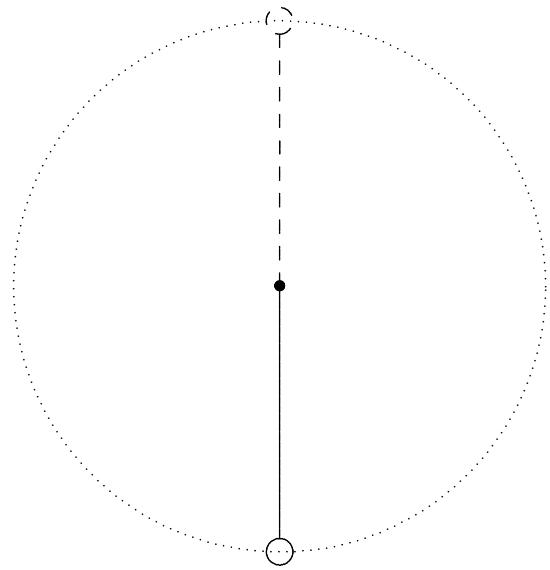


Figure 3.5: A mass on a rigid rod, free to swing in the earth's gravity in a vertical plane.

together into real solutions, like (3.69). The general real solution is of the form

$$\begin{aligned} X(t) = \text{Re}\,[(b+ic)Z(t)] = \\ b\,A\,\cos\omega t + c\,A\,\sin\omega t = d\,A\,\cos(\omega t - \theta) \end{aligned} \tag{3.74}$$

where b and c (or d and θ) are real numbers.

We can now construct the complete solution to the equation of motion. Because of linearity, we get it by adding together all the normal mode solutions with arbitrary coefficients that must be set by the initial conditions.

We can now see that the number of different normal modes is always equal to n, the number of degrees of freedom. Label the normal modes as A^α, where α is a label that (we will argue below) goes from 1 to n. Label the corresponding frequencies ω_α. Then the most general possible motion of the system is a sum of all the normal modes,

$$Z(t) = \sum_{\alpha=1}^{n} w_\alpha A^\alpha e^{-i\omega_\alpha t} \tag{3.75}$$

or in real form (with $w = b + ic$)

$$X(t) = \sum_{\alpha=1}^{n} \left[b_\alpha A^\alpha \cos(\omega_\alpha t) + c_\alpha A^\alpha \sin(\omega_\alpha t) \right] \\ = \sum_{\alpha=1}^{n} d_\alpha A^\alpha \cos(\omega_\alpha t - \theta_\alpha) \tag{3.76}$$

where b_α and c_α (or d_α and θ_α) are real numbers that must be determined from the initial conditions of the system. **Note that the set of all the normal mode vectors must be "complete," in the mathematical sense that any possible configuration of this system can be described as a linear combination of normal modes.** Otherwise, we could not satisfy arbitrary initial conditions with the solution, (3.76). This can be proved mathematically (because the matrix, K, is symmetric and the masses are positive), but the physical argument will be enough for us here. Likewise no normal mode can possibly be a linear combination of the other normal modes, because each corresponds to an independent possible motion of the physical system with its own frequency. The mathematical way of saying this is that the set of all the normal modes is "linearly independent."

> Because the set of normal modes must be both complete and linearly independent, there must be precisely n normal modes, where again, n is the number of degrees of freedom. (3.77)

If there were fewer than n normal modes, they could not possibly describe all possible configurations of the n degrees of freedom. If there were more than n, they could not be linearly independent n dimensional vectors. At least one of them could be written as a linear combination of the others. As we will see later, (3.77) is the physical principle behind Fourier analysis.

It is worth noting that solving the eigenvalue equation, (3.68), gets hard very rapidly as the number of degrees of freedom increases. First you have to compute the determinant of an $n \times n$ matrix. If all the entries are nonzero, this requires adding up $n!$ terms. Once you have finished that, you still have to solve a polynomial equation of degree n. For $n > 3$, this cannot be done analytically except in special cases.

On the other hand, it is always straightforward to check whether a given vector is an eigenvector of a given matrix and, if so, to compute the eigenvalue. We will use this fact in the problems at the end of the chapter.

3.3.2 Back to the 2×2 Example

Let us return to the example from the beginning of this chapter in the special case where the two pendulum blocks have the same mass, $m_1 = m_2 = m$. Simple as it is, this will be a very important system for our understanding of wave phenomena. Let us see how the techniques that we have developed allow us to solve for the allowed frequencies and the corresponding A vectors, the normal modes. From (3.7) and (3.8), the K matrix has the form

$$K = \begin{pmatrix} mg/\ell + \kappa & -\kappa \\ -\kappa & mg/\ell + \kappa \end{pmatrix} . \tag{3.78}$$

The M matrix is

$$M = \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix} . \tag{3.79}$$

Thus from (3.78) and (3.79),

$$M^{-1}K = \begin{pmatrix} g/\ell + \kappa/m & -\kappa/m \\ -\kappa/m & g/\ell + \kappa/m \end{pmatrix} . \tag{3.80}$$

The matrix $M^{-1}K - \omega^2 I$ is

$$M^{-1}K - \omega^2 I = \begin{pmatrix} g/\ell + \kappa/m - \omega^2 & -\kappa/m \\ -\kappa/m & g/\ell + \kappa/m - \omega^2 \end{pmatrix} . \tag{3.81}$$

To find the eigenvalues of $M^{-1}K$, we form the determinant

$$\begin{aligned} \det[M^{-1}K - \omega^2 I] &= \det\left[\begin{pmatrix} g/\ell + \kappa/m - \omega^2 & -\kappa/m \\ -\kappa/m & g/\ell + \kappa/m - \omega^2 \end{pmatrix}\right] \\ &= (g/\ell + \kappa/m - \omega^2)^2 - (\kappa/m)^2 \\ &= (\omega^2 - g/\ell)(\omega^2 - g/\ell - 2\kappa/m) = 0 . \end{aligned} \tag{3.82}$$

Thus the angular frequencies of the normal modes are

$$\omega_1^2 = g/\ell , \quad \omega_2^2 = g/\ell + 2\kappa/m . \tag{3.83}$$

To find the corresponding normal modes, we substitute these frequencies back into the eigenvalue equation. For ω_1^2, the normal mode vector, A^1,

$$A^1 = \begin{pmatrix} a_1^1 \\ a_2^1 \end{pmatrix} , \tag{3.84}$$

satisfies the matrix equation

$$[M^{-1}K - \omega_1^2 I]A^1 = 0 . \tag{3.85}$$

From (3.81) and (3.83),

$$M^{-1}K - \omega_1^2 I = \begin{pmatrix} \kappa/m & -\kappa/m \\ -\kappa/m & \kappa/m \end{pmatrix} . \tag{3.86}$$

Thus (3.85) becomes

$$\begin{aligned}&\begin{pmatrix}\kappa/m & -\kappa/m\\ -\kappa/m & \kappa/m\end{pmatrix}\begin{pmatrix}a_1^1\\ a_2^1\end{pmatrix}=0\\ &=\frac{\kappa}{m}\begin{pmatrix}a_1^1-a_2^1\\ -a_1^1+a_2^1\end{pmatrix}\Rightarrow a_1^1=a_2^1\,.\end{aligned} \tag{3.87}$$

We can take $a_1^1 = 1$ because we can multiply the normal mode vector by any number we like. Only the ratio a_1^1/a_2^1 matters. So, for example, we can take

$$A^1 = \begin{pmatrix}1\\ 1\end{pmatrix}\,. \tag{3.88}$$

This gives (3.10). The displacement in this normal mode is shown in figure 3.6.

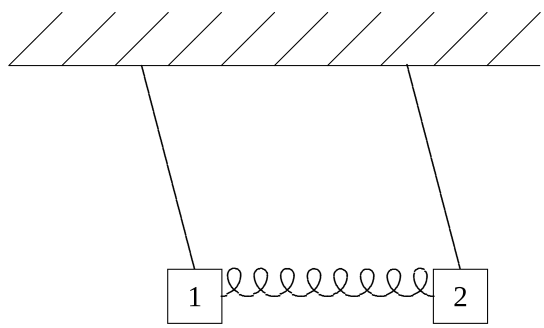


Figure 3.6: The displacement in the normal mode, A^1.

For ω_2^2, the normal mode vector, A^2,

$$A^2 = \begin{pmatrix}a_1^2\\ a_2^2\end{pmatrix}\,, \tag{3.89}$$

satisfies the matrix equation (where the identity matrix multiplying ω_2^2 is understood)[3]

$$[M^{-1}K - \omega_2^2]A^2 = 0\,. \tag{3.90}$$

[3]It is tiresome writing the identity matrix, I, everywhere. It is not really necessary because you can always tell from the context whether it belongs there or not. From now on, we will often leave it out. Thus, if you see something that looks like a number in a matrix equation, like the $-\omega_2^2$ in (3.90), you should mentally include a factor of I.

This time, (3.81) and (3.83) give

$$M^{-1}K - \omega_2^2 = \begin{pmatrix} -\kappa/m & -\kappa/m \\ -\kappa/m & -\kappa/m \end{pmatrix} . \tag{3.91}$$

Thus (3.90) becomes

$$\begin{aligned} &\begin{pmatrix} -\kappa/m & -\kappa/m \\ -\kappa/m & -\kappa/m \end{pmatrix} \begin{pmatrix} a_1^2 \\ a_2^2 \end{pmatrix} = 0 \\ &= -\frac{\kappa}{m} \begin{pmatrix} a_1^2 + a_2^2 \\ a_1^2 + a_2^2 \end{pmatrix} \Rightarrow a_1^2 = -a_2^2 \, . \end{aligned} \tag{3.92}$$

Again, only the ratio a_1^2/a_2^2 matters, so we can take

$$A^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} . \tag{3.93}$$

This gives (3.11). The displacement in this normal mode is shown in figure 3.7.

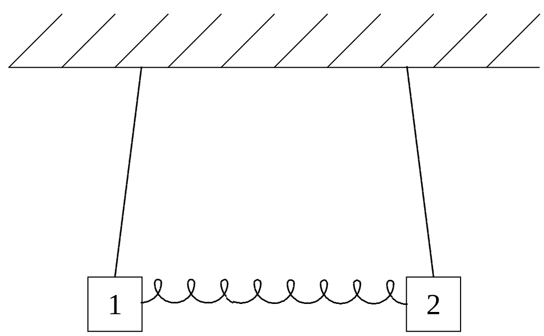


Figure 3.7: The displacement in the normal mode, A^2.

The physics of these modes is easy to understand. In mode 1, the blocks move together and the spring is never stretched from its equilibrium position. Thus the frequency is just g/ℓ, the same as an uncoupled pendulum. In mode 2, the blocks are moving in opposite directions, so the spring is stretched by twice the displacement of each block. Thus there is an additional restoring force of 2κ, and the square of the angular frequency is correspondingly larger.

3.3.3 n=2 — the General Case

Let us work out explicitly the case of $n = 2$ for an arbitrary K matrix,

$$M^{-1}K = \begin{pmatrix} K_{11}/m_1 & K_{12}/m_1 \\ K_{12}/m_2 & K_{22}/m_2 \end{pmatrix} , \tag{3.94}$$

where we have used $K_{21} = K_{12}$. Then (3.73) becomes

$$\left(\frac{K_{11}K_{22} - K_{12}^2}{m_1\, m_2}\right) - \left(\frac{K_{11}}{m_1} + \frac{K_{22}}{m_2}\right)\omega^2 + \omega^4 = 0\,, \tag{3.95}$$

with solutions

$$\omega^2 = \frac{1}{2}\left(\frac{K_{11}}{m_1} + \frac{K_{22}}{m_2}\right) \pm \sqrt{\frac{1}{4}\left(\frac{K_{11}}{m_1} - \frac{K_{22}}{m_2}\right)^2 + \frac{K_{12}^2}{m_1\, m_2}}\,. \tag{3.96}$$

For each ω^2, we can take $a_1 = 1$. Then

$$a_2 = \frac{m_1\,\omega^2 - K_{11}}{K_{12}}\,. \tag{3.97}$$

As we anticipated, the eigenvectors turned out to be real. This a general consequence of the reality of $M^{-1}K$ and ω^2. The argument is worth repeating. When all the elements of the matrix $M^{-1}K - \omega^2 I$ are real, the ratios, a_j/a_k are real (because they are obtained by solving a set of simultaneous linear equations with real coefficients). Thus if we choose one component of the vector A to be real (multiplying, if necessary, by a complex number), then all the components will be real. Physically, this means that for the solution, (3.66), all the different parts of the system are oscillating not only with the same frequency, but with the same phase up to a sign. This is true only because we have ignored damping. We will return to the question in the last section (an optional section that is not for the fainthearted).

3.3.4 The Initial Value Problem

Once you have solved for the normal modes and corresponding frequencies, it is straightforward to put them together into the most general solution to the equations of motion for the set of N coupled oscillators, (3.76). It is

$$X(t) = \sum_\alpha \left(b_\alpha A^\alpha \cos\omega_\alpha t + c_\alpha A^\alpha \sin\omega_\alpha t\right)\,. \tag{3.98}$$

The $2N$ constants b_α and c_α are determined by the initial conditions. The b_α are related to the initial displacements, $X(0)$:

$$X(0) = \sum_\alpha b_\alpha A^\alpha\,. \tag{3.99}$$

In words, b_α is the coefficient of the normal mode A^α in the initial displacement $X(0)$. The c_α are related to the initial velocities, $\left.\frac{dX(t)}{dt}\right|_{t=0}$:

$$\left.\frac{dX(t)}{dt}\right|_{t=0} = \sum_\alpha c_\alpha \omega_\alpha A^\alpha\,. \tag{3.100}$$

The equations, (3.99) and (3.100), are two sets of simultaneous linear equations for the b_α and c_α. They can be solved by hand. This is easy enough for a small number of degrees of freedom. We will see in the next section that we can also get the solutions directly with very little additional work by manipulating the normal modes.

Meanwhile, we should pause again to consider the physics of (3.98). This shows explicitly how the most general motion of the system can be decomposed into the simple motions associated with the normal modes. It is worth staring at an example (real, animated or preferably both) at this point. Try to construct the system in figure 3.1. Any two identical oscillators with a relatively weak spring connecting them will do. Convince yourself that the normal modes exist. If you start the system oscillating with the blocks moving the same way with the same amplitude, they will stay that way. If you get them started moving in opposite directions with the same amplitude, they will continue doing that. Now set up a random motion. See if you can understand how to take it apart into normal modes. It may help to stare again at program 3-1 on the program disk, in which this is done explicitly. In this animation, you see the two blocks of figure 3.1 and below, the two normal modes that must be added to produce the full solution.

3.4 * Normal Coordinates and Initial Values

There is another way of looking at the solutions of (3.14). We can find linear combinations of the original coordinates that oscillate only with a single frequency, no matter what else is going on. This construction is also useful. It allows us to use the form of the normal modes to simplify the solution to the initial value problem.

To see how this works, let us return to the simple example of two identical pendulums, (3.78)-(3.93). The most general possible motion of this system looks like

$$X(t) = bA^1\cos(\omega_1 t - \theta_1) + cA^2\cos(\omega_2 t - \theta_2)\,, \tag{3.101}$$

or, using (3.88) and (3.93)

$$\begin{aligned} x_1(t) &= b\cos(\omega_1 t - \theta_1) + c\cos(\omega_2 t - \theta_2)\,, \\ x_2(t) &= b\cos(\omega_1 t - \theta_1) - c\cos(\omega_2 t - \theta_2)\,. \end{aligned} \tag{3.102}$$

The motion of each block is nonharmonic, involving two different frequencies and four constants that must be determined by solving the initial value problem for both blocks.

But consider the linear combination

$$X^1(t) \equiv x_1(t) + x_2(t)\,. \tag{3.103}$$

In this combination, all dependence on c and θ_2 goes away,

$$X^1(t) = 2b\cos(\omega_1 t - \theta_1)\,. \tag{3.104}$$

This combination oscillates with the single frequency, ω_1, and depends on only two constants, b and θ_1, no matter what the initial conditions are. Likewise,

$$X^2(t) \equiv x_1(t) - x_2(t) \tag{3.105}$$

oscillates with the frequency, ω_2,

$$X^2(t) = 2c\cos(\omega_2 t - \theta_2)\,. \tag{3.106}$$

X^1 and X^2 are called "normal coordinates." We can just as well describe the motion of the system in terms of X^1 and X^2 as in terms of x_1 and x_2. We can go back and forth using the definitions, (3.103) and (3.105). While x_1 and x_2 are more natural from the point of view of the physical setup of the system, figure 3.1, X^1 and X^2 are more convenient for understanding the solution. As we will see below, by going back and forth from physical coordinates to normal coordinates, we can simplify the analysis of the initial value problem.

It turns out that it is possible to construct normal coordinates for any system of normal modes. Consider a normal mode A^α corresponding to a frequency ω_α. Construct the row vector

$$B^\alpha = A^{\alpha T}\, M \tag{3.107}$$

where $A^{\alpha T}$ is the transpose of A^α, a row vector with a_j^α in the jth column.

The row vector B^α is also an eigenvector of the matrix $M^{-1}K$, but this time from the left. That is

$$B^\alpha\, M^{-1}K = \omega_\alpha^2\, B^\alpha\,. \tag{3.108}$$

To derive (3.108), note that (3.68) can be transposed to give

$$A^{\alpha T}\, K\, M^{-1} = \omega_\alpha^2\, A^{\alpha T} \tag{3.109}$$

because M^{-1} and K are both symmetric (see (3.18) and notice that the order of M^{-1} and K are reversed by the transposition). Then

$$B^\alpha M^{-1}K = A^{\alpha T} M M^{-1} K = A^{\alpha T} K M^{-1} M \tag{3.110}$$

$$= \omega_\alpha^2 A^{\alpha T} M = \omega_\alpha^2 B^\alpha\,. \tag{3.111}$$

Given a row vector satisfying (3.108), we can form the linear combination of coordinates

$$X^\alpha = B^\alpha \cdot X = \sum_j b_j^\alpha x_j\,. \tag{3.112}$$

Then X^α is the normal coordinate that oscillates with angular frequency ω_α because

$$\frac{d^2X^\alpha}{dt^2} = B^\alpha \cdot \frac{d^2X}{dt^2} = -B^\alpha M^{-1}KX = -\omega_\alpha^2 B^\alpha \cdot X = -\omega_\alpha^2 X^\alpha\,. \tag{3.113}$$

Thus each normal coordinate behaves just like the coordinate in a system with only one degree of freedom. **The B^α vectors from which the normal coordinates are constructed carry the same amount of information as the normal modes. Indeed, we can go back and forth using (3.107).**

3.4.1 More on the Initial Value Problem

Here we show how to use normal modes and normal coordinates to simplify the solution of the initial value problem for systems of coupled oscillators. At the same time, we can use our physical insight to learn something about the mathematics of the eigenvalue problem. We would like to find the constants b_α and c_α determined by (3.99) and (3.100) without actually solving these linear equations. Indeed there is an easy way. We can make use of the special properties of the normal coordinates. Consider the combination

$$B^\beta A^\alpha \,. \tag{3.114}$$

This combination is just a number, because it is a row vector times a column vector on the right. We know, from (3.112), that $X^\beta = B^\beta X$ is the normal coordinate that oscillates with frequency ω_β, that is:

$$B^\beta \, X(t) \propto e^{\pm i\omega_\beta t} \,. \tag{3.115}$$

On the other hand, the only terms in (3.98) that oscillate with this frequency are those for which $\omega_\alpha = \omega_\beta$. Thus if ω_β is not equal to ω_α, then $B^\beta A^\alpha$ must vanish to give consistency with (3.115).

If the system has two or more normal modes with different A vectors, but the same frequency, we cannot use (3.115) to distinguish them. In this situation, we say that the modes are "degenerate." Suppose that A^1 and A^2 are two different modes with the same frequency,

$$M^{-1}K \, A^1 = \omega^2 \, A^1 \,, \quad M^{-1}K \, A^2 = \omega^2 \, A^2 \,. \tag{3.116}$$

Because the eigenvalues are the same, any linear combination of the two mode vectors is still a normal mode with the same frequency,

$$M^{-1}K \, \left(\beta_1 A^1 + \beta_2 A^2\right) = \omega^2 \, \left(\beta_1 A^1 + \beta_2 A^2\right) \,, \tag{3.117}$$

for any constants, β_1 and β_2.

Now if $A^{1T} M A^2 \neq 0$, we can use (3.117) to choose a new A^2 as follows:

$$A^2 \to A^2 - \frac{A^{1T} M A^2}{A^{1T} M A^1} \, A^1 \,. \tag{3.118}$$

This new normal mode satisfies

$$A^{1T} M A^2 = 0 \,. \tag{3.119}$$

The construction in (3.118) can be extended to any number of normal modes of the same frequency. Thus even if we have several normal modes with the same frequency, we can still use the linearity of the system to choose the normal modes to satisfy

$$B^\beta A^\alpha = A^{\beta^T} M A^\alpha = 0 \text{ for } \beta \neq \alpha \,. \tag{3.120}$$

We will almost always assume that we have done this.

We can use (3.120) to simplify the initial value problem. Consider (3.99). If we multiply this vector equation on both sides by the row vector B^β, we get

$$B^\beta X(0) = B^\beta \sum_\alpha b_\alpha A^\alpha = \sum_\alpha b_\alpha B^\beta A^\alpha = b_\beta B^\beta A^\beta \,. \tag{3.121}$$

where the last step follows because of (3.120), which implies that the sum over α only contributes for $\alpha = \beta$. Thus we can calculate b_α directly from the normal modes and $X(0)$,

$$b_\alpha = \frac{B^\alpha X(0)}{B^\alpha A^\alpha} \,. \tag{3.122}$$

Similarly

$$\omega_\alpha c_\alpha = \frac{1}{B^\alpha A^\alpha} B^\alpha \left. \frac{dX(t)}{dt} \right|_{t=0} \,. \tag{3.123}$$

The point is that we have already solved simultaneous linear equations like (3.99) in finding the eigenvectors of $M^{-1}K$ so it is not necessary to do it again in solving for b_α and c_α. Physically, we know that the normal coordinate X^α must be proportional to the coefficient of the normal mode A^α in the motion. The precise statement of this is (3.122).

3.4.2 * Matrices from Vectors

We can also use (3.120) and the physical requirement of linear independence of the normal modes to write $M^{-1}K$ and the identity matrix in terms of the normal modes.

First consider the identity matrix. One can think of the identity matrix as a machine that takes any vector and returns the same vector. But, using (3.120), we can construct such a machine out of the normal modes. Consider the matrix H, defined as follows:

$$H = \sum_\alpha \frac{A^\alpha B^\alpha}{B^\alpha A^\alpha} \,. \tag{3.124}$$

Note that H is a matrix because $A^\alpha B^\alpha$ in the numerator is the product of a column vector times a row vector **on the right**, rather than on the left. If we let H act on one of the normal mode vectors A^β, and use (3.120), it is easy to see that only the term $\alpha = \beta$ in the sum contributes and $H \cdot A^\beta = A^\beta$. But because the normal modes are a complete set of N linearly independent vectors, that implies that $H \cdot V = V$ for any vector, V. Thus H is the identity matrix,

$$H = I \,. \tag{3.125}$$

We can use this form for I to get an expression for $M^{-1}K$ in terms of a sum over normal modes. Consider the product $M^{-1}K \cdot H = M^{-1}K$, and use the eigenvalue condition

$M^{-1}KA^\alpha = \omega_\alpha^2 A^\alpha$ to obtain

$$M^{-1}K = \sum_\alpha \frac{\omega_\alpha^2\, A^\alpha B^\alpha}{B^\alpha A^\alpha}\,. \tag{3.126}$$

In mathematical language, what is going on in (3.124) and (3.126) is a change of the basis in which we describe the matrices acting on our vector space from the original basis of some obvious set of independent displacements of the degrees of freedom to the less obvious but more useful basis of the normal modes.

3.4.3 * ω^2 is Real

We can use (3.120) to show that all the eigenvalues of the $M^{-1}K$ are real. This is a particular example of an important general mathematical theorem. You will use it frequently when you study quantum mechanics. To prove it, let us assume the contrary and derive a contradiction. If ω^2 is a complex eigenvalue with eigenvector, A, then the complex conjugate, ω^{2*}, is also an eigenvalue with eigenvector, A^*. This must be so because the $M^{-1}K$ matrix is real, which implies that we can take the complex conjugate of the eigenvalue equation,

$$M^{-1}K\,A = \omega^2\,A\,, \tag{3.127}$$

to obtain

$$M^{-1}K\,A^* = \omega^{2*}\,A^*\,. \tag{3.128}$$

Then if ω^2 is complex, ω^2 and ω^{2*} are different and (3.120) implies

$$A^{*T}MA = 0\,. \tag{3.129}$$

But (3.129) is impossible unless $A = 0$ or at least one of the masses in M is negative. To see this, let us expand it in the components of A.

$$A^{*T}MA = \sum_{j=1}^{n} a_j^* m_j a_j = \sum_{j=1}^{n} m_j |a_j|^2\,. \tag{3.130}$$

Each of the terms in (3.130) is positive or zero. Thus the only solutions of the eigenvalue equation, (3.127), for complex ω^2 are the trivial ones in which $A = 0$ on both sides. All the normal modes have real ω^2.

Thus there are only three possibilities. $\omega^2 > 0$ corresponds to stable equilibrium and harmonic oscillation. $\omega^2 < 0$, in which case ω is pure imaginary, occurs when the equilibrium is unstable. $\omega^2 = 0$ is the situation in which the equilibrium is neutral and we can deform the system with no restoring force.

3.5 * Forced Oscillations and Resonance

One of the advantages of the matrix formalism that we have introduced is that in matrix language we can take over the above discussion of forced oscillation and resonance in chapter 2 almost unchanged to systems with more than one degree of freedom. **We simply have to replace numbers by appropriate vectors and matrices.** In particular, the force $F(t)$ in the equation of motion, (2.2), becomes a vector that describes the force on each of the degrees of freedom in the system. The only restriction here is that the frequency of oscillation is the same for each component of the force. The ω_0^2 in the equation of motion, (2.2), becomes the matrix $M^{-1}K$. The frictional term Γ becomes a matrix. In terms of the matrix Γ, the frictional force vector is $M\Gamma dZ/dt$ (compare (2.1)). Then we can look for an irreducible, steady state solution to the equation of motion of the form

$$Z(t) = We^{-i\omega t} \tag{3.131}$$

where W is a constant vector, which yields the matrix equation

$$\left[-\omega^2 - i\Gamma\omega + M^{-1}K\right] W = M^{-1}F_0 \,. \tag{3.132}$$

Formally, we can solve this by multiplying by the inverse matrix

$$W = \left[M^{-1}K - \omega^2 - i\Gamma\omega\right]^{-1} M^{-1}F_0 \,. \tag{3.133}$$

If Γ were zero in the matrix

$$\left[-\omega^2 - i\Gamma\omega + M^{-1}K\right] , \tag{3.134}$$

then we know that the inverse matrix would not exist for any value of ω corresponding to a free oscillation frequency of the system, ω_0, because the determinant of the $M^{-1}K - \omega_0^2$ matrix is zero. The amplitude W would go to ∞ in this limit, in the direction of the normal mode associated with the driving frequency, so long as the driving force has a component in the normal mode direction. **For ω close to ω_0, if there is no damping, the response amplitude is very large, proportional to $1/(\omega_0^2 - \omega^2)$, almost in the direction of the normal mode.** However, in the presence of damping, the response amplitude does not go to ∞ even for $\omega = \omega_0$, because the $i\Gamma\omega$ term is still nonvanishing.

We can see all this explicitly if the damping matrix Γ is proportional to the identity matrix,

$$\Gamma = \gamma \, I \,. \tag{3.135}$$

Then we can use (3.124)-(3.126) to write $\left[M^{-1}K - \omega^2 - i\Gamma\omega\right]$ as a sum over the normal modes, as follows:

$$\left[M^{-1}K - \omega^2 - i\Gamma\omega\right] = \sum_\alpha \left(\omega_\alpha^2 - \omega^2 - i\gamma\omega\right) \frac{A^\alpha B^\alpha}{B^\alpha A^\alpha} \,. \tag{3.136}$$

Then the inverse matrix can be constructed in a similar way, just by inverting the factor in the numerator:

$$\left[M^{-1}K-\omega^2-i\Gamma\omega\right]^{-1}=\sum_\alpha\left(\omega_\alpha^2-\omega^2-i\gamma\omega\right)^{-1}\frac{A^\alpha B^\alpha}{B^\alpha A^\alpha}\,. \tag{3.137}$$

Using (3.137), we can rewrite (3.133) as

$$W=\sum_\alpha\frac{A^\alpha}{\omega_\alpha^2-\omega^2-i\gamma\omega}\,\frac{B^\alpha M^{-1}F_0}{B^\alpha A^\alpha}\,. \tag{3.138}$$

This has a simple interpretation. The second factor on the right hand side of (3.138) is the coefficient of the normal mode A^α in the driving term, $M^{-1}F_0$. This coefficient is multiplied by the complex number

$$\left[\frac{1}{\omega_\alpha^2-\omega^2-i\gamma\omega}\right]\,, \tag{3.139}$$

which is exactly analogous to the factor in (2.21) in the one dimensional case. Thus if $\Gamma\propto I$, then, for each normal mode, the forced oscillation works just as it does for one degree of freedom. If Γ is not proportional to the identity matrix, the formulas are a bit more complicated, but the physics is qualitatively the same.

3.5.1 Example

We will illustrate these considerations with our favorite example, the system of two identical coupled oscillators, with $M^{-1}K$ matrix given by (3.80). We will imagine that the system is sitting in a viscous fluid that gives a uniform damping $\Gamma=\gamma I$, and that there is a periodic force that acts twice as strongly on block 1 as on block 2 (for example, we might give the blocks electric charge $2q$ and q and subject them to a periodic electric field), so that the force is

$$F(t)=\begin{pmatrix}2\\1\end{pmatrix}f_0\cos\omega t=\operatorname{Re}\left[\begin{pmatrix}2\\1\end{pmatrix}f_0e^{-i\omega t}\right]\,. \tag{3.140}$$

Thus

$$M^{-1}F_0=\begin{pmatrix}2\\1\end{pmatrix}\frac{f_0}{m}\,. \tag{3.141}$$

Now to use (3.133), we need only invert the matrix

$$[M^{-1}K-\omega^2-i\Gamma\omega]=\begin{pmatrix}\frac{g}{\ell}+\frac{\kappa}{m}-\omega^2-i\gamma\omega & -\frac{\kappa}{m}\\ -\frac{\kappa}{m} & \frac{g}{\ell}+\frac{\kappa}{m}-\omega^2-i\gamma\omega\end{pmatrix}\,. \tag{3.142}$$

This is simple enough to do by hand. We will do that first, and then compare the result with (3.137). The determinant is

$$\begin{aligned} &\left(\frac{g}{\ell}+\frac{\kappa}{m}-\omega^2-i\gamma\omega\right)^2-\left(\frac{\kappa}{m}\right)^2 \\ &=\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2-i\gamma\omega\right)\cdot\left(\frac{g}{\ell}-\omega^2-i\gamma\omega\right) . \end{aligned} \tag{3.143}$$

Applying (3.34), we find

$$\begin{aligned} &[M^{-1}K-\omega^2-i\Gamma\omega]^{-1} \\ &=\frac{1}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2-i\gamma\omega\right)\left(\frac{g}{\ell}-\omega^2-i\gamma\omega\right)} \\ &\cdot\begin{pmatrix} \frac{g}{\ell}+\frac{\kappa}{m}-\omega^2-i\gamma\omega & \frac{\kappa}{m} \\ \frac{\kappa}{m} & \frac{g}{\ell}+\frac{\kappa}{m}-\omega^2-i\gamma\omega \end{pmatrix} . \end{aligned} \tag{3.144}$$

If we isolate the contribution of the two zeros in the denominator of (3.144), we can write

$$\begin{aligned} &[M^{-1}K-\omega^2-i\Gamma\omega]^{-1} \\ &=\frac{1}{2}\frac{1}{\left(\frac{g}{\ell}-\omega^2-i\gamma\omega\right)}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ &+\frac{1}{2}\frac{1}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2-i\gamma\omega\right)}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{aligned} \tag{3.145}$$

which is just (3.137), as promised. Now substituting into (3.133), we find

$$\begin{aligned} W &= \frac{1}{2}\frac{1}{\left(\frac{g}{\ell}-\omega^2-i\gamma\omega\right)}\begin{pmatrix}3\\3\end{pmatrix}\frac{f_0}{m} \\ &+\frac{1}{2}\frac{1}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2-i\gamma\omega\right)}\begin{pmatrix}1\\-1\end{pmatrix}\frac{f_0}{m} \\ &= \frac{1}{2}\frac{\left(\frac{g}{\ell}-\omega^2+i\gamma\omega\right)}{\left(\frac{g}{\ell}-\omega^2\right)^2+(\gamma\omega)^2}\begin{pmatrix}3\\3\end{pmatrix}\frac{f_0}{m} \\ &+\frac{1}{2}\frac{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2+i\gamma\omega\right)}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2\right)^2+(\gamma\omega)^2}\begin{pmatrix}1\\-1\end{pmatrix}\frac{f_0}{m}, \end{aligned} \tag{3.146}$$

from which we can read off the final result:

$$X(t)=\mathrm{Re}\left(We^{-i\omega t}\right)=\begin{pmatrix}\alpha_1\cos\omega t+\beta_1\sin\omega t\\ \alpha_2\cos\omega t+\beta_2\sin\omega t\end{pmatrix} \tag{3.147}$$

where

$$\begin{aligned} \alpha_{1(2)} &= \frac{3}{2}\frac{\left(\frac{g}{\ell}-\omega^2\right)}{\left(\frac{g}{\ell}-\omega^2\right)^2+(\gamma\omega)^2}\frac{f_0}{m} \\ &\pm\frac{1}{2}\frac{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2\right)}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2\right)^2+(\gamma\omega)^2}\frac{f_0}{m} \end{aligned} \tag{3.148}$$

and

$$\begin{aligned} \beta_{1(2)} &= \frac{3}{2}\frac{\gamma\omega}{\left(\frac{g}{\ell}-\omega^2\right)^2+(\gamma\omega)^2}\frac{f_0}{m} \\ &\pm\frac{1}{2}\frac{\gamma\omega}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2\right)^2+(\gamma\omega)^2}\frac{f_0}{m}. \end{aligned} \tag{3.149}$$

The power expended by the external force is the sum over all the degrees of freedom of the force times the velocity. In matrix language, this can be written as

$$P(t)=F(t)^T\cdot\frac{dX(t)}{dt}. \tag{3.150}$$

The average power lost to the frictional force comes from the $\cos^2\omega t$ term in (3.150) and is

$$
\begin{aligned}
&= \frac{1}{\left(\frac{g}{\ell}-\omega^2\right)^2+(\gamma\omega)^2}\frac{9\gamma\omega^2 f_0^2}{4m} \\
&+ \frac{1}{\left(\frac{g}{\ell}+2\frac{\kappa}{m}-\omega^2\right)^2+(\gamma\omega)^2}\frac{\gamma\omega^2 f_0^2}{4m}
\end{aligned}
\tag{3.151}
$$

Figure 3.8 shows a graph of this (for $\kappa/m = 3g/2\ell$ and $\gamma^2 = g/4\ell$). There are two things to observe about figure 3.8. First note the two resonance peaks, at $\omega^2 = g/\ell$ and $\omega^2 = g/\ell + 2\kappa/m = 4g/\ell$. Secondly, note that the first peak is much more pronounced that the second. That is because the force is more in the direction of the normal mode with the lower frequency, thus it is more efficient in exciting this mode.

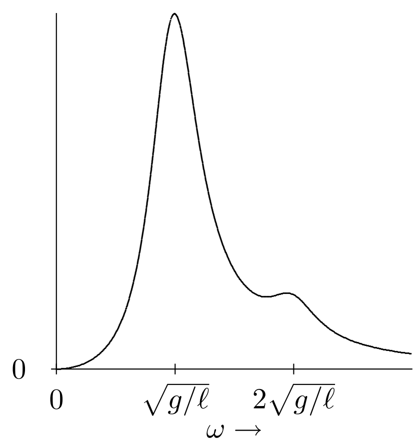


Figure 3.8: The average power lost to friction in the example of 3.140.

Chapter Checklist

You should now be able to:

1. Write down the equations of motion for a system with more than one degree of freedom in matrix form;

2. Find the M and K matrices from the physics;

3. Add, subtract and multiply matrices;

4. Find the determinant and inverse of 2×2 and 3×3 matrices;
5. Find normal modes and corresponding frequencies of a system with two degrees of freedom, which means finding the eigenvectors and eigenvalues of a 2×2 matrix;
6. Check whether a given vector is a normal mode of a system with more than two degrees of freedom, and if so, find the corresponding angular frequency;
7. Given the normal modes and corresponding frequencies and the initial positions and velocities of all the parts in any system, find the motion of all the parts at all subsequent times;
8. ***** Go back and forth from normal modes to normal coordinates;
9. ***** Reconstruct the $M^{-1}K$ matrix from the normal modes and normal coordinates;
10. ***** Explicitly solve for the free oscillations of system with two degrees of freedom with damping and be able to analyze systems with three or more degrees of freedom if you are given the eigenvectors;
11. ***** Explicitly solve forced oscillation problems with or without damping for systems with three or fewer degrees of freedom.

Problems

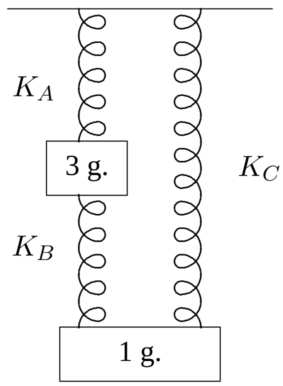
3.1. The 3 component column vector A, the 3 component row vector B and the 3×3 matrix C are defined as follows:

$$A = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & -2 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 1 \\ 2 & 2 & 0 \end{pmatrix}.$$

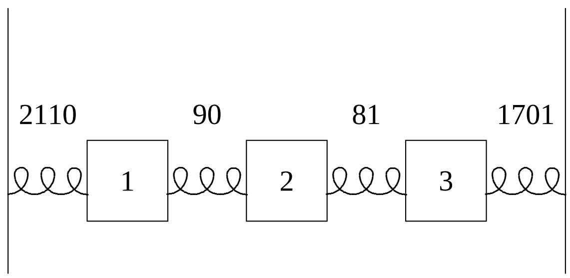
Compute the following objects:

$$BA, \quad BC, \quad AB.$$

3.2. Consider the vertical oscillation of the system of springs and masses shown below with the spring constants $K_A = 78$, $K_B = 15$ and $K_C = 6$ (all dynes/cm). Find the normal modes, normal coordinates and associated angular frequencies. If the 1 g. block is displaced up 1 cm from its equilibrium position with the 3 g block held at its equilibrium position and both blocks released from rest, describe the subsequent motion of both blocks.



3.3. Consider the system of springs and masses shown below:



with the spring constants in newtons/meter given above the springs and with $m_1 = 100$ kg, $m_2 = 9$ kg and $m_3 = 81$ kg.

a. Which of the following are normal modes of the system and what are the corresponding angular frequencies? Note that the $M^{-1}K$ matrix may look a little complicated.

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \\ 10 \end{pmatrix} \quad \begin{pmatrix} 9 \\ 60 \\ 10 \end{pmatrix} \quad \begin{pmatrix} 9 \\ -30 \\ 10 \end{pmatrix} \quad \begin{pmatrix} 9 \\ 30 \\ 10 \end{pmatrix} \quad \begin{pmatrix} 9 \\ 0 \\ -10 \end{pmatrix}$$

b. If the system is released from rest with an initial displacement as shown below (with the displacements measured in mm), how long does it take before it first returns to its initial configuration?

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \\ 10 \end{pmatrix}$$

3.4 * . A system of four masses connected by springs is described by a mass matrix,

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

and a K matrix

$$K = \begin{pmatrix} 29 & -10 & -4 & -2 \\ -10 & 58 & -14 & -2 \\ -4 & -14 & 31 & -26 \\ -2 & -2 & -26 & 74 \end{pmatrix}$$

a. Which of the following are normal modes?

$$\begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 1 \\ -1 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 4 \\ -3 \\ 0 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ -4 \\ 3 \end{pmatrix}$$

b. For each normal mode, find the corresponding angular frequency. **Hint:** this requires a little arithmetic. If you are lazy, you might want to use a programmable calculator or write a little computer program to check these for you. But the point of this problem is to show you that the amount of work required to check whether the vectors are normal modes is really tiny compared to the work involved in finding the modes from scratch.

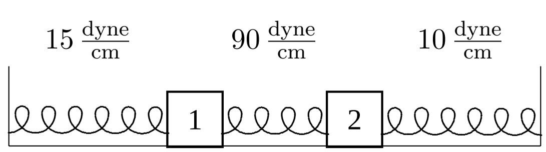
c. If blocks are released from rest from an initial displacement that is proportional to

$$\begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix},$$

which normal mode is not present in the subsequent motion?

d. Find the normal coordinates corresponding to each of the normal modes of the system.

3.5. Consider the longitudinal oscillations of the system shown below:



The blocks are free to slide horizontally without friction. The displacements of the blocks from equilibrium are both measured to the right. Block 1 has a mass of 15 grams and block 2 a mass of 10 grams. The spring constants of the springs are shown in dynes/cm.

a. Show that the $M^{-1}K$ matrix of this system is

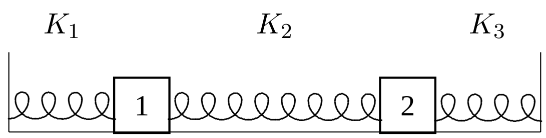
$$M^{-1}K = \begin{pmatrix} 7 & -6 \\ -9 & 10 \end{pmatrix} .$$

b. Show that the normal modes are

$$A^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad A^2 = \begin{pmatrix} 2 \\ -3 \end{pmatrix} .$$

Find the corresponding angular frequencies, ω_1 and ω_2.

3.6. Consider the longitudinal oscillations of the system shown below:



The blocks are free to slide horizontally without friction. The displacements of the blocks from equilibrium are both measured to the right. Block 1 has a mass of 15 grams and block 2 a mass of 10 grams. The spring constants of the springs are K_1, K_2 and K_3, as shown. The normal modes of this system are

$$A^1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} , \quad A^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

with corresponding frequencies

$$\omega_1 = 1\,\mathrm{s}^{-1} , \quad \omega_2 = 2\,\mathrm{s}^{-1} .$$

a. If the system is at rest at time $t = 0$ with displacements $x_1(0) = 5$ cm, $x_2(0) = 0$, or

$$X(0) = \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix} \mathrm{cm} .$$

Find the displacement of block 2 at time $t = \pi$ s.

b. Find K_1, K_2 and K_3.

3.7 * . In the system of problem (3.5), suppose we immerse the system in a damping fluid so that

$$\Gamma = \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix}$$

with $\gamma = 1$ s^{-1}, and that an external force of the following form is applied (in dynes):

$$F(t) = f \cos \omega t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cos \omega t \, .$$

Find and graph the average power lost to the frictional force as a function of ω from $\omega = 0$ to 10 s^{-1}.

Chapter 4

Symmetries

Symmetry is an important concept in physics and mathematics (and art!). In this chapter, we show how the mathematics of symmetry can be used to simplify the analysis of the normal modes of symmetrical systems.

Preview

In this chapter, we introduce the formal concept of symmetry or invariance.

1. We will work out some examples of the use of symmetry arguments to simplify the analysis of oscillating systems.

4.1 Symmetries

Let us return to the system of two identical pendulums coupled by a spring, discussed in chapter 3, in (3.78)-(3.93). This simple system has more to teach us. It is shown in figure 4.1. As in (3.78)-(3.93), both blocks have mass m, both pendulums have length ℓ and the spring constant is κ. Again we label the small displacements of the blocks to the right, x_1 and x_2.

We found the normal modes of this system in the last chapter. But in fact, we could have found them even more easily by making use of the symmetry of this system. If we reflect this system in a plane midway between the two blocks, we get back a completely equivalent system. We say that the system is "invariant" under reflections in the plane between the blocks. However, while the physics is unchanged by the reflection, our description of the system is affected. The coordinates get changed around. The reflected system is shown in figure 4.2. Comparing the two figures, we can describe the reflection in terms of its effect on the displacements,

$$x_1 \to -x_2 \,, \quad x_2 \to -x_1 \,. \tag{4.1}$$

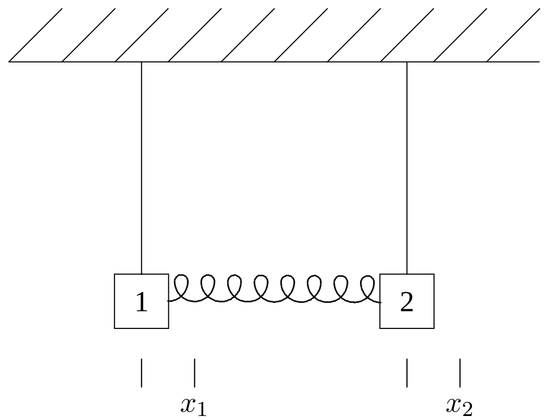


Figure 4.1: A system of coupled pendulums. Displacements are measured to the right, as shown.

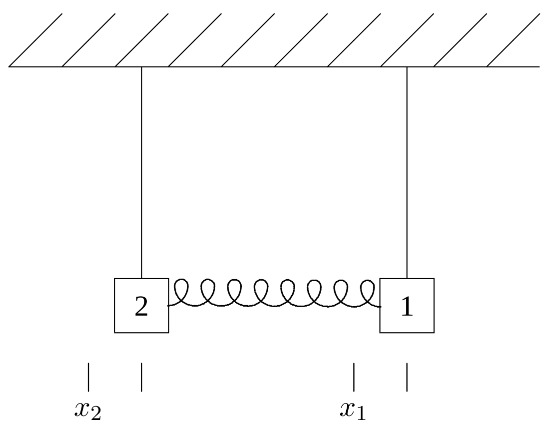


Figure 4.2: The system of coupled pendulums after reflection in the plane through between the two.

In particular, if

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \tag{4.2}$$

is a solution to the equations of motion for the system, then the reflected vector,

$$\tilde{X}(t) \equiv \begin{pmatrix} -x_2(t) \\ -x_1(t) \end{pmatrix} , \tag{4.3}$$

must also be a solution, because the reflected system is actually identical to the original. While this must be so from the physics, it is useful to understand how the math works. To see mathematically that (4.3) is a solution, define the symmetry matrix, S,

$$S \equiv \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} , \tag{4.4}$$

so that $\tilde{X}(t)$ is related to $X(t)$ by matrix multiplication:

$$\tilde{X}(t) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = S\,X(t) . \tag{4.5}$$

The mathematical statement of the symmetry is the following condition on the M and K matrices:[1]

$$M\,S = S\,M , \tag{4.6}$$

and

$$K\,S = S\,K . \tag{4.7}$$

You can check explicitly that (4.6) and (4.7) are true. From these equations, it follows that if $X(t)$ is a solution to the equation of motion,

$$M\,\frac{d^2}{dt^2}X(t) = -K\,X(t) , \tag{4.8}$$

then $\tilde{X}(t)$ is also. To see this explicitly, multiply both sides of (4.8) by S to get

$$S\,M\,\frac{d^2}{dt^2}X(t) = -S\,K\,X(t) . \tag{4.9}$$

Then using (4.6) and (4.7) in (4.9), we get

$$M\,S\frac{d^2}{dt^2}X(t) = -K\,SX(t) . \tag{4.10}$$

[1]Two matrices, A and B, that satisfy $AB = BA$ are said to "commute."

The matrix S is a constant, independent of time, thus we can move it through the time derivatives in (4.10) to get

$$M\,\frac{d^2}{dt^2}SX(t) = -K\,SX(t)\,. \tag{4.11}$$

But now using (4.5), this is the equation of motion for $\tilde{X}(t)$,

$$M\,\frac{d^2}{dt^2}\tilde{X}(t) = -K\,\tilde{X}(t)\,. \tag{4.12}$$

Thus, as promised, (4.6) and (4.7) are the mathematical statements of the reflection symmetry because they imply, as we have now seen explicitly, that if $X(t)$ is a solution, $\tilde{X}(t)$ is also.

Note that from (4.6), you can show that

$$M^{-1}\,S = S\,M^{-1} \tag{4.13}$$

by multiplying on both sides by M^{-1}. Then (4.13) can be combined with (4.7) to give

$$M^{-1}K\,S = S\,M^{-1}K\,. \tag{4.14}$$

We will use this later.

Now suppose that the system is in a normal mode, for example

$$X(t) = A^1\,\cos\omega_1 t\,. \tag{4.15}$$

Then $\tilde{X}(t)$ is another solution. But it has the same time dependence, and thus the same angular frequency. It must, therefore, be proportional to the same normal mode vector because we already know from our previous analysis that the two angular frequencies of the normal modes of the system are different, $\omega_1 \neq \omega_2$. Anything that oscillates with angular frequency, ω_1, must be proportional to the normal mode, A^1:

$$\tilde{X}(t) \propto A^1\,\cos\omega_1 t\,. \tag{4.16}$$

Thus the symmetry implies

$$S\,A^1 \propto A^1\,. \tag{4.17}$$

That is, we expect from the symmetry that the normal modes are also eigenvectors of S. This must be true whenever the angular frequencies are distinct. In fact, we can see by checking the solutions that this is true. The proportionality constant is just -1,

$$S\,A^1 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} A^1 = -A^1\,, \tag{4.18}$$

and similarly

$$S\,A^2 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} A^2 = A^2\,. \tag{4.19}$$

Furthermore, we can run the argument backwards. If A is an eigenvector of the symmetry matrix S, and if all the eigenvalues of S are different, then because of the symmetry, (4.13), A is a normal mode. To see this, consider the vector $M^{-1}KA$ and act on it with the matrix S. Using (4.14), we see that if

$$SA = \beta A \tag{4.20}$$

then

$$S\,M^{-1}KA = M^{-1}K\,SA = \beta M^{-1}KA\,. \tag{4.21}$$

In words, (4.21) means that $M^{-1}KA$ is an eigenvector of S with the same eigenvalue as A. But if the eigenvalues of S are all different, then $M^{-1}KA$ must be proportional to A, which means that A is a normal mode. Mathematically we could say it this way. If the eigenvectors of S are A^n with eigenvalues β_n, then

$$SA^n = \beta_n A^n\,, \text{ and } \beta_n \neq \beta_m \text{ for } n \neq m \Rightarrow A^n \text{ are normal modes.} \tag{4.22}$$

It turns out that for the symmetries we care about, the eigenvalues of S are always all different.[2]

Thus even if we had not known the solution, we could have used (4.20) to determine the normal modes without bothering to solve the eigenvalue problem for the $M^{-1}K$ matrix! Instead of solving the eigenvalue problem,

$$M^{-1}K\,A^n = \omega_n^2\,A^n\,, \tag{4.23}$$

we can instead solve the eigenvalue problem

$$S\,A^n = \beta_n\,A^n\,. \tag{4.24}$$

It might seem that we have just traded one eigenvalue problem for another. But in fact, (4.24) is easier to solve, because **we can use the symmetry to determine the eigenvalues, β_n, without ever computing a determinant.** The reflection symmetry has the nice property that if you do it twice, you get back to where you started. This is reflected in the property of the matrix S,

$$S^2 = I\,. \tag{4.25}$$

In words, this means that applying the matrix S twice gives you back exactly the vector that you started with. Multiplying both sides of the eigenvalue equation, (4.24), by S, we get

$$\begin{aligned} A^n = I\,A^n = S^2\,A^n = S\,\beta_n\,A^n \\ = \beta_n\,S\,A^n = \beta_n^2\,A^n\,, \end{aligned} \tag{4.26}$$

[2] See the discussion on page 103.

which implies

$$\beta_n^2 = 1 \quad \text{or} \quad \beta_n = \pm 1\,. \tag{4.27}$$

This saves some work. Once the eigenvalues of S are known, it is easier to find the eigenvectors of S. But because of the symmetry, we know that the eigenvectors of S will also be the normal modes, the eigenvectors of $M^{-1}K$. And once the normal modes are known, it is straightforward to find the angular frequency by acting on the normal mode eigenvectors with $M^{-1}K$.

What we have seen here, in a simple example, is how to use the symmetry of an oscillating system to determine the normal modes. In the remainder of this chapter we will generalize this technique to a much more interesting situation. The idea is always the same.

We can find the normal modes by solving the eigenvalue problem for the symmetry matrix, S, instead of $M^{-1}K$. And we can use the symmetry to determine the eigenvalues. (4.28)

4.1.1 Beats

4-1

The beginnings of wave phenomena can already be seen in this simple example. Suppose that we start the system oscillating by displacing block 1 an amount d with block 2 held fixed in its equilibrium position, and then releasing both blocks from rest at time $t = 0$. The general solution has the form

$$X(t) = A^1 \left(b_1 \cos\omega_1 t + c_1 \sin\omega_1 t\right) + A^2 \left(b_2 \cos\omega_2 t + c_2 \sin\omega_2 t\right). \tag{4.29}$$

The positions of the blocks at $t = 0$ gives the matrix equation:

$$X(0) = \begin{pmatrix} d \\ 0 \end{pmatrix} = A^1 b_1 + A^2 b_2\,, \tag{4.30}$$

or

$$\begin{array}{r} d = b_1 + b_2 \\ 0 = -b_1 + b_2 \end{array} \Rightarrow b_1 = b_2 = \frac{d}{2}\,. \tag{4.31}$$

Because both blocks are released from rest, we know that $c_1 = c_2 = 0$. We can see this in the same way by looking at the initial velocities of the blocks:

$$\dot{X}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \omega_1 A^1 c_1 + \omega_2 A^2 c_2\,, \tag{4.32}$$

or

$$\begin{array}{r} 0 = c_1 + c_2 \\ 0 = -c_1 + c_2 \end{array} \Rightarrow c_1 = c_2 = 0\,. \tag{4.33}$$

Thus

$$x_1(t) = \frac{d}{2}\left(\cos\omega_1 t + \cos\omega_2 t\right)$$
$$x_2(t) = \frac{d}{2}\left(\cos\omega_1 t - \cos\omega_2 t\right) \ . \tag{4.34}$$

The remarkable thing about this solution is the way in which the energy gets completely transferred from block 1 to block 2 and back again. To see this, we can rewrite (4.34) as (using (1.64) and another similar identity)

$$x_1(t) = d\cos\Omega t\cos\delta\omega t$$
$$x_2(t) = d\sin\Omega t\sin\delta\omega t \tag{4.35}$$

where

$$\Omega = \frac{\omega_1+\omega_2}{2}\,, \quad \delta\omega = \frac{\omega_2-\omega_1}{2}\,. \tag{4.36}$$

Each of the blocks exhibits "beats." They oscillate with the average angular frequency, Ω, but the amplitude of the oscillation changes with angular frequency $\delta\omega$. After a time $\frac{\pi}{2\delta\omega}$, the energy has been almost entirely transferred from block 1 to block 2. This behavior is shown in program 4-1 on your program disk. Note how the beats are produced by the interplay between the two normal modes. When the two modes are in phase for one of the blocks so that the block is moving with maximum amplitude, the modes are 180° out of phase for the other block, so the other block is almost still.

The complete transfer of energy back and forth from block 1 to block 2 is a feature both of our special initial condition, with block 2 at rest and in its equilibrium position, and of the special form of the normal modes that follows from the reflection symmetry. As we will see in more detail later, this is the same kind of energy transfer that takes place in wave phenomena.

4.1.2 A Less Trivial Example

4-2

Take a hacksaw blade, fix one end and attach a mass to the other. This makes a nice oscillator with essentially only one degree of freedom (because the hacksaw blade will only bend back and forth easily in one way). Now take six identical blades and fix one end of each at a single point so that the blades fan out at 60° angles from the center with their orientation such that they can bend back and forth **in the plane formed by the blades.** If you put a mass at the end of each, in a hexagonal pattern, you will have six uncoupled oscillators. But if instead you put identical magnets at the ends, the oscillators will be coupled together in some complicated way. You can see what the oscillations of this system look like in program 4-2 on the program

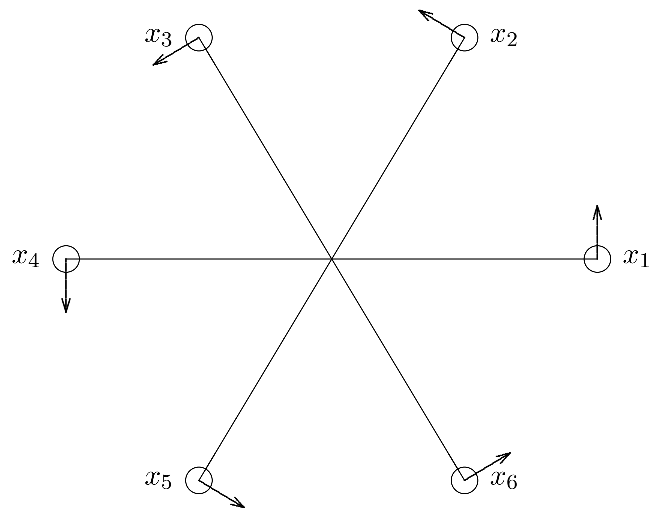


Figure 4.3: A system of six coupled hacksaw blade oscillators. The arrows indicate the directions in which the displacements are measured.

disk. If the displacements from the symmetrical equilibrium positions are small, the system is approximately linear. Despite the apparent complexity of this system, we can write down the normal modes and the corresponding angular frequencies with almost no work! The trick is to make clever use of the symmetry of this system.

This system looks exactly the same if we rotate it by 60° about its center. We should, therefore, take pains to analyze it in a manifestly symmetrical way. Let us label the masses 1 through 6 starting any place and going around counterclockwise. Let x_j be the counterclockwise displacement of the jth block from its equilibrium position. As usual, we will arrange these coordinates in a vector:[3]

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} . \tag{4.37}$$

The symmetry operation of rotation is implemented by the cyclic substitution

$$x_1 \to x_2 \to x_3 \to x_4 \to x_5 \to x_6 \to x_1 . \tag{4.38}$$

[3]From here on, we will assume that the reader is sufficiently used to complex numbers that it is not necessary to distinguish between a real coordinate and a complex coordinate.

This can be represented in a matrix notation as

$$X \to S\,X, \tag{4.39}$$

where the symmetry matrix, S, is

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.40}$$

Note that the 1s along the next-to-diagonal of the matrix, S, in (4.40) implement the substitutions

$$x_1 \to x_2 \to x_3 \to x_4 \to x_5 \to x_6\,, \tag{4.41}$$

while the 1 in the lower left-hand corner closes the circle with the substitution

$$x_6 \to x_1\,. \tag{4.42}$$

The symmetry requires that the K matrix for this system has the following form:

$$K = \begin{pmatrix} E & -B & -C & -D & -C & -B \\ -B & E & -B & -C & -D & -C \\ -C & -B & E & -B & -C & -D \\ -D & -C & -B & E & -B & -C \\ -C & -D & -C & -B & E & -B \\ -B & -C & -D & -C & -B & E \end{pmatrix}. \tag{4.43}$$

Notice that all the diagonal elements are the same (E), as they must be because of the symmetry. The jth diagonal element of the K matrix is minus the force per unit displacement on the jth mass due to its displacement. Because of the symmetry, each of the masses behaves in exactly the same way when it is displaced with all the other masses held fixed. Thus all the diagonal matrix elements of the K matrix, K_{jj}, are equal. Likewise, the symmetry ensures that the effect of the displacement of each block, j, on its neighbor, $j \pm 1$ ($j+1 \to 1$ if $j = 6$, $j-1 \to 6$ if $j = 1$ — see (4.42)), is exactly the same. Thus the matrix elements along the next-to-diagonal (B) are all the same, along with the Bs in the corners. And so on! The K matrix then satisfies (4.7),

$$S\,K = K\,S \tag{4.44}$$

which, as we saw in (4.13)-(4.12), is the mathematical statement of the symmetry. Indeed, we can go backwards and work out the most general symmetric matrix consistent with (4.44) and check that it must have the form, (4.43). You will do this in problem (4.4).

Because of the symmetry, we know that if a vector A is a normal mode, then the vector SA is also a normal mode with the same frequency. This is physically obvious. If the system oscillates with all its parts in step in a certain way, it can also oscillate with the parts rotated by 60°, but otherwise moving in the same way, and the frequency will be the same. This suggests that we look for normal modes that behave simply under the symmetry transformation S. In particular, if we find the eigenvectors of S and discover that the eigenvalues of S are all different, then we know that all the eigenvectors are normal modes, from (4.22). In the previous example, we found modes that went into themselves multiplied by ± 1 under the symmetry. In general, however, we should not expect the eigenvalues to be real because the modes can involve complex exponentials. In this case, we must look for modes that correspond to complex eigenvalues of S,[4]

$$SA = \beta\, A\,. \tag{4.45}$$

As above in (4.25)-(4.27), we can find the possible eigenvalues by using the symmetry. Note that because six 60° rotations get us back to the starting point, the matrix, S, satisfies

$$S^6 = I\,. \tag{4.46}$$

Because of (4.46), it follows that $\beta^6 = 1$. Thus β is a sixth root of one,

$$\beta = \beta_k = e^{2ik\pi/6} \text{ for } k = 0 \text{ to } 5\,. \tag{4.47}$$

Then for each k, there is a normal mode

$$S\, A^k = \beta_k\, A^k\,. \tag{4.48}$$

Explicitly,

$$SA^k = \begin{pmatrix} A_2^k \\ A_3^k \\ A_4^k \\ A_5^k \\ A_6^k \\ A_1^k \end{pmatrix} = \beta_k \cdot \begin{pmatrix} A_1^k \\ A_2^k \\ A_3^k \\ A_4^k \\ A_5^k \\ A_6^k \end{pmatrix}\,. \tag{4.49}$$

If we take $A_1^k = 1$, we can solve for all the other components,

$$A_j^k = (\beta_k)^{j-1}\,. \tag{4.50}$$

[4]Even this is not the most general possibility. In general, we might have to consider sets of modes that go into one another under matrix multiplication. That is not necessary here because the symmetry transformations all commute with one another.

Thus

$$\begin{pmatrix} A_1^k \\ A_2^k \\ A_3^k \\ A_4^k \\ A_5^k \\ A_6^k \end{pmatrix} = \begin{pmatrix} 1 \\ e^{2ik\pi/6} \\ e^{4ik\pi/6} \\ e^{6ik\pi/6} \\ e^{8ik\pi/6} \\ e^{10ik\pi/6} \end{pmatrix} . \tag{4.51}$$

Now to determine the angular frequencies corresponding to the normal modes, we have to evaluate

$$M^{-1}KA^k = \omega_k^2 A^k . \tag{4.52}$$

Since we already know the form of the normal modes, this is straightforward. For example, we can compare the first components of these two vectors:

$$\begin{aligned} \omega_k^2 &= \left(E - Be^{2ik\pi/6} - Ce^{4ik\pi/6} - De^{6ik\pi/6} - Ce^{8ik\pi/6} - Be^{10ik\pi/6}\right)/m \\ &= \frac{E}{m} - 2\frac{B}{m}\cos\frac{k\pi}{3} - 2\frac{C}{m}\cos\frac{2k\pi}{3} - (-1)^k\frac{D}{m} . \end{aligned} \tag{4.53}$$

Notice that $\omega_1^2 = \omega_5^2$ and $\omega_2^2 = \omega_4^2$. This had to be the case, because the corresponding normal modes are complex conjugate pairs,

$$A^5 = A^{1*}, \quad A^4 = A^{2*} . \tag{4.54}$$

Any complex normal mode must be part of a pair with its complex conjugate normal mode at the same frequency, so that we can make real normal modes out of them. This must be the case because the normal modes describe a real physical system whose displacements are real. The real modes are linear combinations (see (1.19)) of the complex modes,

$$A^k + A^{k*} \text{ and } (A^k - A^{k*})/i \text{ for } k = 1 \text{ or } 2 . \tag{4.55}$$

These modes can be seen in program 4-2 on the program disk. See appendix A and your program instruction manual for details.

Notice that the real solutions, (4.55), are not eigenvectors of the symmetry matrix, S. This is possible because the angular frequencies are not all different. However, the eigenvalues of S are all different, from (4.47). Thus even though we can construct normal modes that are not eigenvectors of S, it is still true that **all the eigenvectors of S are normal modes.** This is what we use in (4.48)-(4.50) to determine the A^n.

We note that (4.55) is another example of a very important principle of (3.117) that we will use many times in what follows:

> If A and A' are normal modes of a system **with the same angular frequency,** ω, then any linear combination, $bA + cA'$, is also a normal mode with the same angular frequency. (4.56)

Normal modes with the same frequency can be linearly combined to give new normal modes (see problem 4.3). On the other hand, a linear combination of two normal modes with **different** frequencies gives nothing very simple.

The techniques used here could have been used for any number of masses in a similar symmetrical arrangement. With N masses and symmetry under rotation of $2\pi/N$ radians, the Nth roots of 1 would replace the 6th roots of one in our example. Symmetry arguments can also be used to determine the normal modes in more interesting situations, for example when the masses are at the corners of a cube. But that case is more complicated than the one we have analyzed because the order of the symmetry transformations matters — the transformations do not commute with one another. You may want to look at it again after you have studied some group theory.

Chapter Checklist

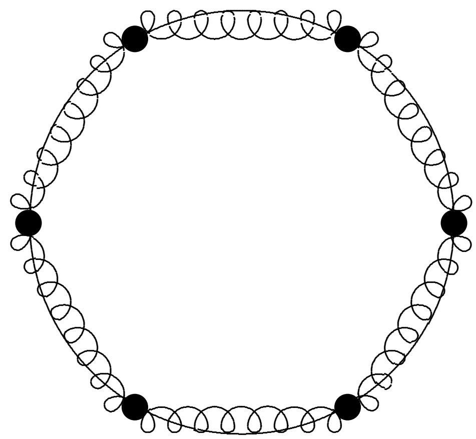
You should now be able to:

1. Apply symmetry arguments to find the normal modes of systems of coupled oscillators by finding the eigenvalues and eigenvectors of the symmetry matrix.

Problems

4.1. Show explicitly that (4.7) is true for the K matrix, (4.43), of system of figure 4.3 by finding SK and KS.

4.2. Consider a system of six identical masses that are free to slide without friction on a circular ring of radius R and each of which is connected to both its nearest neighbors by identical springs, shown below in equilibrium:



a. Analyze the possible motions of this system in the region in which it is linear (note that this is not quite just small oscillations). To do this, define appropriate displacement variables (so that you can use a symmetry argument), find the form of the K matrix and then follow the analysis in (4.37)-(4.55). If you have done this properly, you should find that one of the modes has zero frequency. Explain the physical significance of this mode. **Hint:** Do not attempt to find the form of the K matrix directly from the spring constants of the spring and the geometry. This is a mess. Instead, figure out what it has to look like on the basis of symmetry arguments. You may want to look at appendix c.

b. If at $t = 0$, the masses are evenly distributed around the circle, but every other mass is moving with (counterclockwise) velocity v while the remaining masses are at rest, find and describe in words the subsequent motion of the system.

4.3.

a. Prove (4.56).

b. Prove that if A and A' are normal modes corresponding to **different** angular frequencies, ω and ω' respectively, where $\omega^2 \neq \omega'^2$, then $bA + cA'$ is not a normal mode unless b or c is zero. **Hint:** You will need to use the fact that both A and A' are nonzero vectors.

4.4. Show that (4.43) is the most general symmetric 6×6 matrix satisfying (4.44).

Chapter 5

Waves

The climax of this book comes early. Here we identify the crucial features of a system that supports waves — **space translation invariance and local interactions.**

Preview

We identify the space translation invariance of the class of infinite systems in which wave phenomena take place.

1. Symmetry arguments cannot be directly applied to finite systems that support waves, such as a series of coupled pendulums. However, we show that if the couplings are only between neighboring blocks, the concept of symmetry can still be used to understand the oscillations. In this case we say that the interactions are "local." The idea is to take the physics apart into two different components: the physics of the interior; and the physics of the boundaries, which is incorporated in the form of boundary conditions. The interior can be regarded as part of an infinite system with space translation invariance, a symmetry under translations by some distance, a. In this case the normal modes are called standing waves.

2. We then introduce a notation designed to take maximum advantage of the space translation invariance of the infinite system. We introduce the angular wave number, k, which plays the role for the spatial dependence of the wave that the angular frequency, ω, plays for its time dependence.

3. We describe the normal modes of transverse oscillation of a beaded string. The modes are "wavy."

4. We study the normal modes of a finite beaded string with free ends as another example of boundary conditions.

5. We study a type of forced oscillation problem that is particularly important for translation invariant systems with local interactions. If the driving force acts only at the ends of the system, the solution can be found simply using boundary conditions.

6. We apply the idea of space translation invariance to a system of coupled LC circuits.

5.1 Space Translation Invariance

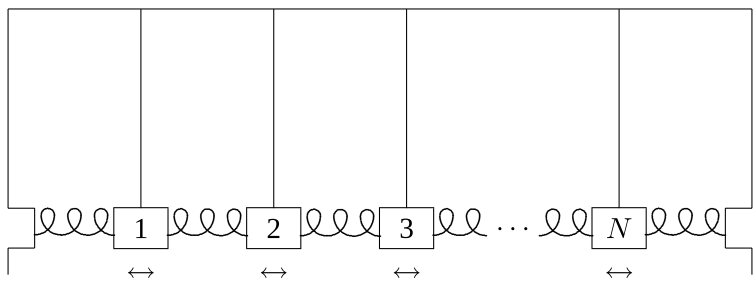


Figure 5.1: A finite system of coupled pendulums.

The typical system of coupled oscillators that supports waves is one like the system of N identical coupled pendulums shown in figure 5.1. This system is a generalization of the system of two coupled pendulums that we studied in chapters 3 and 4. Suppose that each pendulum bob has mass m, each pendulum has length ℓ, each spring has spring constant κ and the equilibrium separation between bobs is a. Suppose further that there is no friction and that the pendulums are constrained to oscillate only in the direction in which the springs are stretched. We are interested in the free oscillation of this system, with no external force. Such an oscillation, when the motion is parallel to the direction in which the system is stretched in space is called a "longitudinal oscillation". Call the longitudinal displacement of the jth bob from equilibrium ψ_j. We can organize the displacements into a vector, Ψ (for reasons that will become clear below, it would be confusing to use X, so we choose a different letter, the Greek letter psi, which looks like ψ in lower case and Ψ when capitalized):

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \\ \psi_N \end{pmatrix} . \tag{5.1}$$

Then the equations of motion (for small longitudinal oscillations) are

$$\frac{d^2\Psi}{dt^2} = -M^{-1}K\,\Psi \tag{5.2}$$

where M is the diagonal matrix with m's along the diagonal,

$$\begin{pmatrix} m & 0 & 0 & \cdots & 0 \\ 0 & m & 0 & \cdots & 0 \\ 0 & 0 & m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & m \end{pmatrix}, \tag{5.3}$$

and K has diagonal elements $(mg/\ell + 2\kappa)$, next-to-diagonal elements $-\kappa$, and zeroes elsewhere,

$$\begin{pmatrix} mg/\ell + 2\kappa & -\kappa & 0 & \cdots & 0 \\ -\kappa & mg/\ell + 2\kappa & -\kappa & \cdots & 0 \\ 0 & -\kappa & mg/\ell + 2\kappa & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & mg/\ell + 2\kappa \end{pmatrix}. \tag{5.4}$$

The $-\kappa$ in the next-to-diagonal elements has exactly the same origin as the $-\kappa$ in the 2×2 K matrix in (3.78). It describes the coupling of two neighboring blocks by the spring. The $(mg/\ell + 2\kappa)$ on the diagonal is analogous to the $(mg/\ell + \kappa)$ on the diagonal of (3.78). The difference in the factor of 2 in the coefficient of κ arises because there are two springs, one on each side, that contribute to the restoring force on each block in the system shown in figure 5.1, while there was only one in the system shown in figure 3.1. Thus $M^{-1}K$ has the form

$$\begin{pmatrix} 2B & -C & 0 & \cdots & 0 \\ -C & 2B & -C & \cdots & 0 \\ 0 & -C & 2B & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2B \end{pmatrix} \tag{5.5}$$

where

$$2B = g/\ell + 2\kappa/m\,, \quad C = \kappa/m\,. \tag{5.6}$$

It is interesting to compare the matrix, (5.5), with the matrix, (4.43), from the previous chapter. In both cases, the diagonal elements are all equal, because of the symmetry. The same goes for the next-to-diagonal elements. However, in (5.5), all the rest of the elements are zero because the interactions are only between nearest neighbor blocks. We call such interactions "local." In (4.43), on the other hand, each of the masses interacts with all the others. We will use the local nature of the interactions below.

We could try to find normal modes of this system directly by finding the eigenvectors of $M^{-1}K$, but there is a much easier and more generally useful technique. We can divide the physics of the system into two parts, the physics of the coupled pendulums, and the physics of the walls. To do this, **we first consider an infinite system with no walls at all.**

5.1.1 The Infinite System

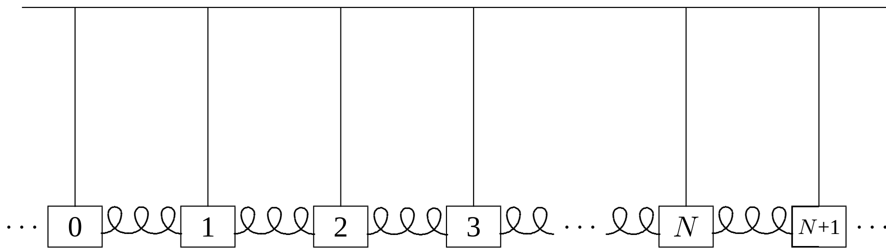


Figure 5.2: A piece of an infinite system of coupled pendulums.

Notice that in figure 5.2, we have not changed the interior of the system shown in figure 5.1 at all. We have just replaced the walls by a continuation of the interior.

Now we can find all the modes of the infinite system of figure 5.2 very easily, making use of a symmetry argument. **The infinite system of figure 5.2 looks the same if it is translated, moved to the left or the right by a multiple of the equilibrium separation, a. It has the property of "space translation invariance."** Space translation invariance is the symmetry of the infinite system under translations by multiples of a. In this example, because of the discrete blocks and finite length of the springs, the space translation invariance is "discrete." Only translation by integral multiples of a give the same physics. Later, we will discuss continuous systems that have continuous space translation invariance. However, we will see that such systems can be analyzed using the same techniques that we introduce in this chapter.

We can use the symmetry of space translation invariance, just as we used the reflection and rotation symmetries discussed in the previous chapter, to find the normal modes of the infinite system. **The discrete space translation invariance of the infinite system (the symmetry under translations by multiples of a) allows us to find the normal modes of the infinite system in a simple way.**

Most of the modes that we find using the space translation invariance of the infinite system of figure 5.2 will have nothing to do with the finite system shown in figure 5.1. **But if we can find linear combinations of the normal modes of the infinite system of figure 5.2 in which the 0th and N+1st blocks stay fixed, then they must be solutions to the equations of motion of the system shown in figure 5.1. The reason is that the interactions between the blocks are "local" — they occur only between nearest neighbor blocks.** Thus block 1 knows what block 0 is doing, but not what block -1 is doing. If block 0 is stationary it might as well be a wall because the blocks on the other side do not affect block 1 (or any of the blocks 1 to N) in any way. The local nature of the interaction allows us to put in the physics of the walls as a boundary condition after solving the infinite problem. This same trick will also enable us to solve many other problems.

Let us see how it works for the system shown in figure 5.1. First, we use the symmetry under translations to find the normal modes of the infinite system of figure 5.2. As in the previous two chapters, we describe the solutions in terms of a vector, A. But now A has an infinite number of components, A_j where the integer j runs from $-\infty$ to $+\infty$. It is a little inconvenient to write this infinite vector down, but we can represent a piece of it:

$$A = \begin{pmatrix} \vdots \\ A_0 \\ A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_N \\ A_{N+1} \\ \vdots \end{pmatrix} . \tag{5.7}$$

Likewise, the $M^{-1}K$ matrix for the system is an infinite matrix, not easily written down, but any piece of it (along the diagonal) looks like the interior of (5.5):

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 2B & -C & 0 & 0 & \cdots \\ \cdots & -C & 2B & -C & 0 & \cdots \\ \cdots & 0 & -C & 2B & -C & \cdots \\ \cdots & 0 & 0 & -C & 2B & \cdots \\ \ddots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} . \tag{5.8}$$

This system is "space translation invariant" because it looks the same if it is moved to the left a distance a. This moves block j+1 to where block j used to be, thus if there is a mode with components A_j, there must be another mode with the same frequency, represented by a vector, $A' = SA$, with components

$$A'_j = A_{j+1} \, . \tag{5.9}$$

The symmetry matrix, S, is an infinite matrix with 1s along the next-to-diagonal. These are analogous to the 1s along the next-to-diagonal in (4.40). Now, however, the transformation never closes on itself. There is no analog of the 1 in the lower left-hand corner of (4.40), because the infinite matrix has no corner. We want to find the eigenvalues and eigenvectors of the matrix S, satisfying

$$A' = S\,A = \beta\,A \tag{5.10}$$

or equivalently (from (5.9)), the modes in which A_j and A'_j are proportional:

$$A'_j = \beta A_j = A_{j+1} \tag{5.11}$$

where β is some nonzero constant.[1]

Equation (5.11) can be solved as follows: Choose $A_0 = 1$. Then $A_1 = \beta$, $A_2 = \beta^2$, etc., so that $A_j = (\beta)^j$ for all nonnegative j. We can also rewrite (5.11) as $A_{j-1} = \beta^{-1} A_j$, so that $A_{-1} = \beta^{-1}$, $A_{-2} = \beta^{-2}$, etc. Thus the solution is

$$A_j = (\beta)^j \tag{5.12}$$

for all j. Note that this solution works for any nonzero value of β, unlike the examples that we discussed in the previous chapter. The reason is that a translation by a, unlike the symmetries of reflection and rotation by 60° discussed in chapter 4, never gets you back to where you started no matter how many times you repeat it. Also, the infinite system, with an infinite number of degrees of freedom, has an infinite number of different normal modes corresponding to different values of β.

For each value of β, there is a unique (up to multiplication by an overall constant) eigenvector, A. We know that it is unique because we have explicitly constructed it in (5.12). Therefore, all of the eigenvalues of S are distinct. Thus from (4.22), we know that each of the eigenvectors is a normal mode of the infinite system. Because there is a one-to-one correspondence between nonzero numbers, β, and normal modes, we can (at least for now — we will find a better notation later), label the normal modes by the eigenvalue, β, of the symmetry matrix, S. We will call the corresponding eigenvector A^β, so that (5.12) can be written

$$A^\beta_j = \beta^j \, . \tag{5.13}$$

Now that we know the form of the normal modes, it is easy to get the corresponding frequencies by acting on (5.12) with the $M^{-1}K$ matrix, (5.8). This gives

$$\omega^2 \, A^\beta_j = 2B A^\beta_j - C A^\beta_{j+1} - C A^\beta_{j-1} \, , \tag{5.14}$$

or inserting (5.13),

$$\omega^2 \beta^j = 2B\beta^j - C\beta^{j+1} - C\beta^{j-1} = (2B - C\beta - C\beta^{-1})\beta^j \, . \tag{5.15}$$

This is true for all j, which shows that (5.13) is indeed an eigenvector (we already knew this from the symmetry argument, (4.22), but it is nice to check when possible), and the eigenvalue is

$$\omega^2 = 2B - C\beta - C\beta^{-1} \, . \tag{5.16}$$

[1] Zero does not work for β because the eigenvalue equation has no solution.

Notice that for almost every value of ω^2, there are two normal modes, because we can interchange β and β^{-1} without changing (5.16). The only exceptions are

$$\omega^2 = 2B \mp 2C \,, \tag{5.17}$$

corresponding to $\beta = \pm 1$. The fact that there are at most two normal modes for each value of ω^2 will have a dramatic consequence. It means that we only have to deal with two normal modes at a time to implement the physics of the boundary. This is a special feature of the one-dimensional system that is not shared by two- and three-dimensional systems. As we will see, it makes the one-dimensional system very easy to handle.

5.1.2 Boundary Conditions

5-1

We have now solved the problem of the oscillation of the infinite system. Armed with this result, we can put back in the physics of the walls. Any β (except $\beta = \pm 1$) gives a pair of normal modes for the infinite system of figure 5.2. But only special values of β will work for the finite system shown in figure 5.1. To find the normal modes of the system shown in figure 5.1, we use (4.56), the fact that **any linear combination of the two normal modes with the same angular frequency, ω, is also a normal mode.** If we can find a linear combination that vanishes for $j = 0$ and for $j = N + 1$, it will be a normal mode of the system shown in figure 5.1. It is the vanishing of the normal mode at $j = 0$ and $j = N + 1$ that are the "boundary conditions" for this particular finite system.

Let us begin by trying to satisfy the boundary condition at $j = 0$. For each possible value of ω^2, we have to worry about only two normal modes, the two solutions of (5.16) for β. So long as $\beta \neq \pm 1$, we can find a combination that vanishes at $j = 0$; just subtract the two modes A^β and $A^{\beta^{-1}}$ to get a vector

$$A = A^\beta - A^{\beta^{-1}} \,, \tag{5.18}$$

or in components

$$A_j \propto A_j^\beta - A_j^{\beta^{-1}} = \beta^j - \beta^{-j} \,. \tag{5.19}$$

The first thing to notice about (5.19) is that A^j cannot vanish for any $j \neq 0$ unless $|\beta| = 1$. Thus if we are to have any chance of satisfying the boundary condition at $j = N + 1$, we must assume that

$$\beta = e^{i\theta} \,. \tag{5.20}$$

Then from (5.19),

$$A_j \propto \sin j\theta \,. \tag{5.21}$$

Now we can satisfy the boundary condition at $j = N + 1$ by setting $A_{N+1} = 0$. This implies $\sin[(N+1)\theta] = 0$, or

$$\theta = n\pi/(N+1), \quad \text{for integer } n\,. \tag{5.22}$$

Thus the normal modes of the system shown in figure 5.1 are

$$A_j^n = \sin\left(\frac{jn\pi}{N+1}\right)\,, \quad \text{for} \quad n = 1\,,\ 2\,,\ \cdots\ N\,. \tag{5.23}$$

Other values of n do not lead to new modes, they just repeat the N modes already shown in (5.23). The corresponding frequencies are obtained by putting (5.20)-(5.21) into (5.16), to get

$$\omega^2 = 2B - 2C\cos\theta = 2B - 2C\cos\left(\frac{n\pi}{N+1}\right)\,. \tag{5.24}$$

From here on, the analysis of the motion of the system is the same as for any other system of coupled oscillators. As discussed in chapter 3, we can take a general motion apart and express it as a sum of the normal modes. This is illustrated for the system of coupled pendulums in program 5-1 on the program disk. The new thing about this system is the way in which we obtained the normal modes, and their peculiarly simple form, in terms of trigonometric functions. We will get more insight into the meaning of these modes in the next section. Meanwhile, note the way in which the simple modes can be combined into the very complicated motion of the full system.

5.2 k and Dispersion Relations

So far, the equilibrium separation between the blocks, a, has not appeared in the analysis. Everything we have said so far would be true even if the springs had random lengths, so long as all spring constants were the same. In such a case, the "space translation invariance" that we used to solve the problem would be a purely mathematical device, taking the original system into a different system with the same kind of small oscillations. Usually, however, in physical applications, the space translation invariance is real and all the inter-block distances are the same. Then it is very useful to **label the blocks by their equilibrium position.** Take $x = 0$ to be the position of the left wall (or the 0th block). Then the first block is at $x = a$, the second at $x = 2a$, etc., as shown in figure 5.3. We can describe the displacement of all the blocks by a function $\psi(x, t)$, where $\psi(ja, t)$ is the displacement of the jth block (the one with equilibrium position ja). Of course, this function is not very well defined because we only care about its values at a discrete set of points. Nevertheless, as we will see below when we discuss the beaded string, it will help us understand what is going on if we draw a smooth curve through these points.

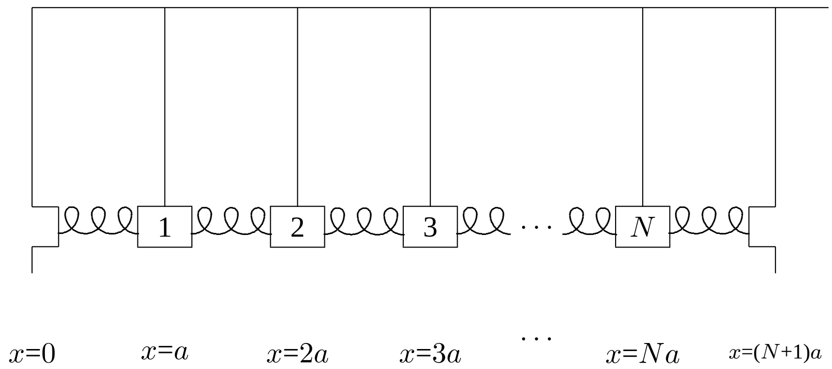


Figure 5.3: The coupled pendulums with blocks labeled by their equilibrium positions.

In the same way, we can describe a normal mode of the system shown in figure 5.1 (or the infinite system of figure 5.2) as a function $A(x)$ where

$$A(ja) = A_j \,. \tag{5.25}$$

In this language, space translation invariance, (5.11), becomes

$$A(x + a) = \beta A(x) \,. \tag{5.26}$$

It is conventional to write the constant β as an exponential

$$\beta = e^{ika} \,. \tag{5.27}$$

Any nonzero complex number can be written as a exponential in this way. In fact, we can change k by a multiple of $2\pi/a$ without changing β, thus we can choose the real part of k to be between $-\pi/a$ and π/a

$$-\frac{\pi}{a} < \mathrm{Re}\, k \leq \frac{\pi}{a} \,. \tag{5.28}$$

If we put (5.13) and (5.27) into (5.25), we get

$$A^\beta(ja) = e^{ikja} \,. \tag{5.29}$$

This suggests that we take the function describing the normal mode corresponding to (5.27) to be

$$A(x) = e^{ikx} \,. \tag{5.30}$$

The mode is determined by the number k satisfying (5.28).

The parameter k (when it is real) is called the angular wave number of the mode. It measures the waviness of the normal mode, in radians per unit distance. The "wavelength" of the mode is the smallest length, λ (the Greek letter lambda), such that a change of x by λ leaves the mode unchanged,

$$A(x+\lambda) = A(x)\,. \tag{5.31}$$

In other words, the wavelength is the length of a complete cycle of the wave, 2π radians. Thus the wavelength, λ, and the angular wave number, k, are inversely related, with a factor of 2π,

$$\lambda = \frac{2\pi}{k}\,. \tag{5.32}$$

In this language, the normal modes of the system shown in figure 5.1 are described by the functions

$$A^n(x) = \sin kx\,, \tag{5.33}$$

with

$$k = \frac{n\pi}{L}\,, \tag{5.34}$$

where $L = (N+1)a$ is the total length of the system. **The important thing about (5.33) and (5.34) is that they do not depend on the details of the system. They do not even depend on N.** The normal modes always have the same shape, when the system has length L. Of course, as N increases, the number of modes increases. For fixed L, this happens because $a = L/(N+1)$ decreases as N increases and thus the allowed range of k (remember (5.28)) increases.

The forms (5.33) for the normal modes of the space translation invariant system are called "standing waves." We will see in more detail below why the word "wave" is appropriate. The word "standing" refers to the fact that while the waves are changing with time, they do not appear to be moving in the x direction, unlike the "traveling waves" that we will discuss in chapter 8 and beyond.

5.2.1 The Dispersion Relation

In terms of the angular wave number k, the frequency of the mode is (from (5.16) and (5.27))

$$\omega^2 = 2B - 2C\cos ka\,. \tag{5.35}$$

Such a relation between k (actually k^2 because $\cos ka$ is an even function of k) **and** ω^2 **is called a "dispersion relation"** (we will learn later why the name is appropriate). The specific form (5.35) is a characteristic of the particular infinite system of figure 5.2. It depends on the masses and spring constants and pendulum lengths and separations.

But it does not depend on the boundary conditions. Indeed, we will see below that (5.35) will be useful for boundary conditions very different from those of the system shown in figure 5.1.

The dispersion relation depends only on the physics of the infinite system. (5.36)

Indeed, it is only through the dispersion relation that the details of the physics of the infinite system enters the problem. The form of the modes, $e^{\pm ikx}$, is already determined by the general properties of linearity and space translation invariance.

We will call (5.35) the dispersion relation for coupled pendulums. We have given it a special name because we will return to it many times in what follows. The essential physics is that there are two sources of restoring force: gravity, that tends to keep all the masses in equilibrium; and the coupling springs, that tend to keep the separations between the masses fixed, but are unaffected if all the masses are displaced by the same distance. In (5.35), the constants always satisfy $B \geq C$, as you see from (5.6).

The limit $B = C$ is especially interesting. This happens when there is no gravity (or $\ell \to \infty$). The dispersion relation is then

$$\omega^2 = 2B(1 - \cos ka) = 4B \sin^2 \frac{ka}{2} . \tag{5.37}$$

Note that the mode with $k = 0$ now has zero frequency, because all the masses can be displaced at once with no restoring force.[2]

5.3 Waves

5.3.1 The Beaded String

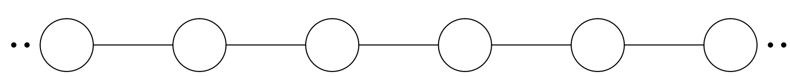


Figure 5.4: The beaded string in equilibrium.

Another instructive system is the beaded string, undergoing transverse oscillations. The oscillations are called "transverse" if the motion is perpendicular to the direction in which the system is stretched. Consider a massless string with tension T, to which identical beads

[2] See appendix C.

of mass m are attached at regular intervals, a. A portion of such a system in its equilibrium configuration is depicted in figure 5.4. The beads cannot oscillate longitudinally, because the string would break.[3] However, for small transverse oscillations, the stretching of the string is negligible, and the tension and the horizontal component of the force from the string are approximately constant. The horizontal component of the force on each block from the string on its right is canceled by the horizontal component from the string on the left. The total horizontal force on each block is zero (this must be, because the blocks do not move horizontally). But the strings produces a transverse restoring force when neighboring beads do not have the same transverse displacement, as illustrated in figure 5.5. The force of the string on bead 1 is shown, along with the transverse component. The dotted lines complete similar triangles, so that $F/T = (\psi_2 - \psi_1)/a$. You can see from figure 5.5 that the restoring force, F in the figure, for small transverse oscillations is linear, and corresponds to a spring constant T/a.

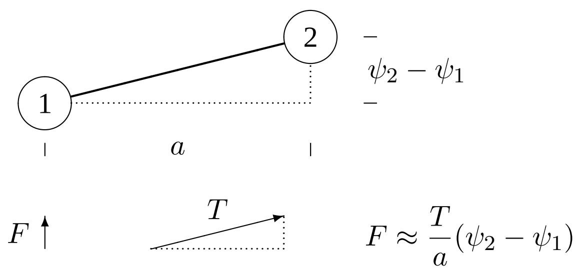


Figure 5.5: Two neighboring beads on a beaded string.

Thus (5.37) is also the dispersion relation for the small transverse oscillations of the beaded string with

$$B = \frac{T}{ma}\,, \tag{5.38}$$

where T is the string tension, m is the bead mass and a is the separation between beads. The dispersion relation for the beaded string can thus be written as

$$\omega^2 = \frac{4T}{ma}\sin^2\frac{ka}{2}\,. \tag{5.39}$$

This dispersion relation, (5.39), has the interesting property that $\omega \to 0$ as $k \to 0$. This is discussed from the point of view of symmetry in appendix C, where we discuss the

[3] More precisely, the string has a very large and nonlinear force constant for longitudinal stretching. The longitudinal oscillations have a much higher frequency and are much more strongly damped than the transverse oscillations, so we can ignore them in the frequency range of the transverse modes. See the discussion of the "light" massive spring in chapter 7.

connection of this dispersion relation with what are called "Goldstone bosons." Here we should discuss the special properties of the $k = 0$ mode with exactly zero angular frequency, $\omega = 0$. This is different from all other angular frequencies because we do not get a different time dependence by complex conjugating the irreducible complex exponential, $e^{-i\omega t}$. But we need two solutions in order to describe the possible initial conditions of the system, because we can specify both a displacement and a velocity for each bead. The resolution of this dilemma is similar to that discussed for critical damping in chapter 2 (see (2.12)). If we approach $\omega = 0$ from nonzero ω, we can form two independent solutions as follows:[4]

$$\lim_{\omega\to 0} \frac{e^{-i\omega t} + e^{i\omega t}}{2} = 1\,, \quad \lim_{\omega\to 0} \frac{e^{-i\omega t} - e^{i\omega t}}{-2i\omega} = t \tag{5.40}$$

The first, for $k = 0$, describes a situation in which all the beads are sitting at some fixed position. The second describes a situation in which all of the beads are moving together at constant velocity in the transverse direction.

Precisely analogous things can be said about the x dependence of the $k = 0$ mode. Again, approaching $k = 0$ from nonzero k, we can form two modes,

$$\lim_{k\to 0} \frac{e^{ikx} + e^{-ikx}}{2} = 1\,, \quad \lim_{k\to 0} \frac{e^{ikx} - e^{-ikx}}{2ik} = x \tag{5.41}$$

The second mode here describes a situation in which each subsequent bead is more displaced. The transverse force on each bead from the string on the left is canceled by the force from the string on the right.

5.3.2 Fixed Ends

5-2

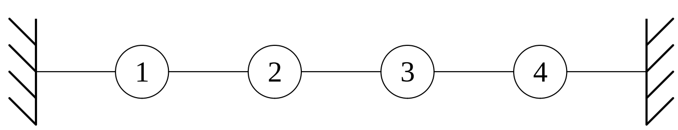


Figure 5.6: A beaded string with fixed ends.

Now suppose that we look at a **finite** beaded string with its ends fixed at $x = 0$ and $x = L = (N + 1)a$, as shown in figure 5.6. The analysis of the normal modes of this system is exactly the same as for the coupled pendulum problem at the beginning of the chapter. Once again, we imagine that the finite system is part of an infinite system with space

[4]You can evaluate the limits easily, using the Taylor series for $e^x = 1 + x + \cdots$.

translation invariance and look for linear combinations of modes such that the beads at $x = 0$ and $x = L$ are fixed. Again this leads to (5.33). The only differences are:

1. the frequencies of the modes are different because the dispersion relation is now given by (5.39);

2. (5.33) describes the **transverse** displacements of the beads.

This is a very nice example of the standing wave normal modes, (5.33), because you can see the shapes more easily than for longitudinal oscillations. For four beads ($N = 4$), the four independent normal modes are illustrated in figures 5.7-5.10, where we have made the coupling strings invisible for clarity. The fixed imaginary beads that play the role of the walls are shown (dashed) at $x = 0$ and $x = L$. Superimposed on the positions of the beads is the continuous function, $\sin kx$, for each k value, represented by a dotted line. Note that this function does **not** describe the positions of the coupling strings, which are stretched straight between neighboring beads.

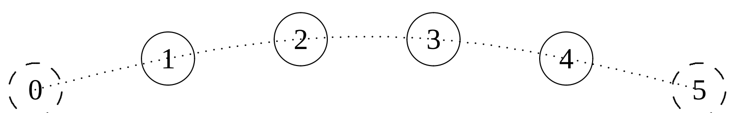


Figure 5.7: $n = 1$.

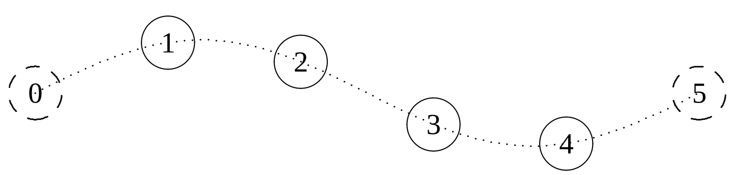


Figure 5.8: $n = 2$.

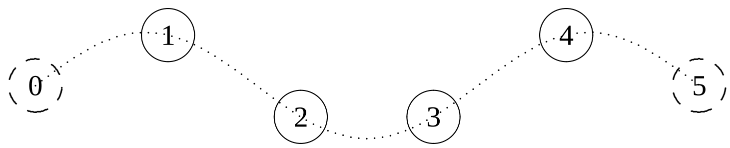


Figure 5.9: $n = 3$.

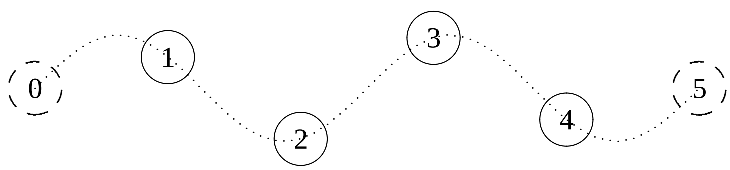


Figure 5.10: $n = 4$.

It is pictures like figures 5.7-5.10 that justify the word "wave" for these standing wave solutions. They are, frankly, wavy, exhibiting the sinusoidal space dependence that is the *sine qua non* of wave phenomena.

The transverse oscillation of a beaded string with both ends fixed is illustrated in program 5-2, where a general oscillation is shown along with the normal modes out of which it is built. Note the different frequencies of the different normal modes, with the frequency increasing as the modes get more wavy. We will often use the beaded string as an illustrative example because the modes are so easy to visualize.

5.4 Free Ends

Let us work out an example of forced oscillation with a different kind of boundary condition. Consider the transverse oscillations of a beaded string. For definiteness, we will take four beads so that this is a system of four coupled oscillators. However, instead of coupling the strings at the ends to fixed walls, we will attach them to massless rings that are free to slide in the transverse direction on frictionless rods. The string then is said to have its ends free (at least for transverse motion). Then the system looks like the diagram in figure 5.11, where the oscillators move up and down in the plane of the paper: Let us find its normal modes.

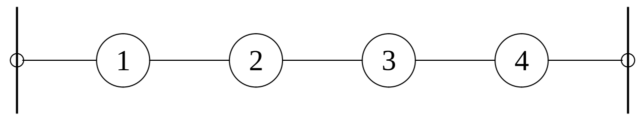


Figure 5.11: A beaded string with free ends.

5.4.1 Normal Modes for Free Ends

 5-3

As before, we imagine that this is part of an infinite system of beads with space translation invariance. This is shown in figure 5.12. Here, the massless rings sliding on frictionless rods have been replaced by the imaginary (dashed) beads, 0 and 5. The dispersion relation is just the same as for any other infinite beaded string, (5.39). The question is, then, what kind of boundary condition on the infinite system corresponds to the physical boundary condition, that the end beads are free on one side? The answer is that we must have the first imaginary bead on either side move up and down with the last real bead, so that the coupling string from bead 0 is horizontal and exerts no transverse restoring force on bead 1 and the coupling string from bead 5 is horizontal and exerts no transverse restoring force on bead 4:

$$A_0 = A_1 \,, \tag{5.42}$$

$$A_4 = A_5 \,; \tag{5.43}$$

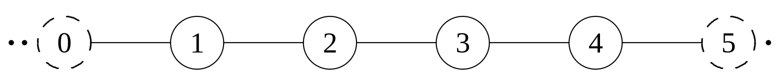


Figure 5.12: Satisfying the boundary conditions in the finite system.

We will work in the notation in which the beads are labeled by their equilibrium positions. The normal modes of the infinite system are then $e^{\pm ikx}$. **But we haven't yet had to decide where we will put the origin**. How do we form a linear combination of the complex exponential modes, $e^{\pm ikx}$, and choose k to be consistent with this boundary condition? Let us begin with (5.42). We can write the linear combination, whatever it is, in the form

$$\cos(kx - \theta) \,. \tag{5.44}$$

Any real linear combination of $e^{\pm ikx}$ can be written in this way up to an overall multiplicative constant (see (1.96)). Now if

$$\cos(kx_0 - \theta) = \cos(kx_1 - \theta)\,, \tag{5.45}$$

where x_j is the position of the jth block, then either

1. $\cos(kx - \theta)$ has a maximum or minimum at $\frac{x_0+x_1}{2}$, or
2. $kx_1 - kx_0$ is a multiple of 2π.

Let us consider case 1. We will see that case 2 does not give any additional modes. We will choose our coordinates so that the point $\frac{x_0+x_1}{2}$, midway between x_0 and x_1, is $x = 0$. We don't care about the overall normalization, so if the function has a minimum there, we will multiply it by -1, to make it a maximum. Thus in case 1, the function $\cos(kx - \theta)$ has a maximum at $x = 0$, which implies that we can take $\theta = 0$. Thus the function is simply $\cos kx$. The system with this labeling is shown in figure 5.13. The displacement of the jth bead is then

$$A_j = \cos[ka(j - 1/2)]\,. \tag{5.46}$$

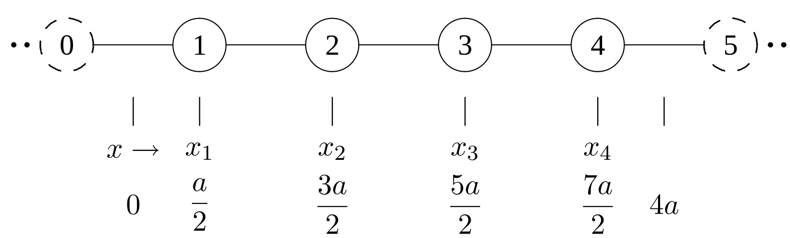


Figure 5.13: The same system of oscillators labeled more cleverly.

It should now be clear how to impose the boundary condition, (5.43), on the other end. We want to have a maximum or minimum midway between bead 4 and bead 5, at $x = 4a$. We get a maximum or minimum every time the argument of the cosine is an integral multiple of π. The argument of the cosine at $x = 4a$ is $4ka$, where k is the angular wave number. Thus the boundary condition will be satisfied if the mode has $4ka = n\pi$ for integer n. Then

$$\cos[ka(4 - 1/2)] = \cos[ka(5 - 1/2)] \Rightarrow ka = \frac{n\pi}{4}\,. \tag{5.47}$$

Thus the modes are

$$A_j = \cos[ka(j - 1/2)] \text{ with } k = \frac{n\pi}{4a} \text{ for } n = 0 \text{ to } 3\,. \tag{5.48}$$

For $n > 3$, the modes just repeat, because $k \geq \pi/a$.

In (5.48), $n = 0$ is the trivial mode in which all the beads move up and down together. This is possible because there is no restoring force at all when all the beads move together. As discussed above (see (5.40)) the beads can all move with a constant velocity because $\omega = 0$ for this mode. Note that case 2, above, gives the same mode, and nothing else, because if $kx_1 - kx_0 = 2n\pi$, then (5.44) has the same value for all x_j. The remaining modes are shown in figures 5.14-5.16. This system is illustrated in program 5-3 on the program disk.

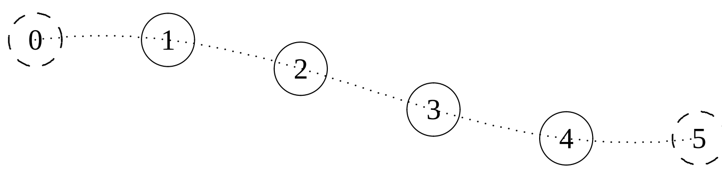


Figure 5.14: $n = 1$, $A_j = \cos[(j - 1/2)\pi/4]$.

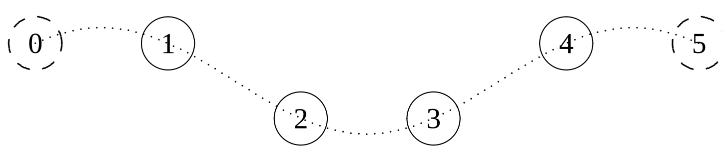


Figure 5.15: $n = 2$, $A_j = \cos[(j - 1/2)\, 2\pi/4]$.

5.5 Forced Oscillations and Boundary Conditions

Forced oscillations can be analyzed using the methods of chapter 3. This always works, even for a force that acts on each of the parts of the system independently. Very often, however, for a space translation invariant system, we are interested in a different sort of forced oscillation problem, one in which the external force acts only at one end (or both ends). In this case, we can solve the problem in a much simpler way using boundary conditions. An example of this sort is shown in figure 5.17.

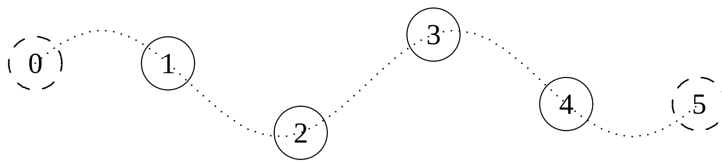


Figure 5.16: $n = 3$, $A_j = \cos[(j - 1/2)\, 3\pi/4]$.

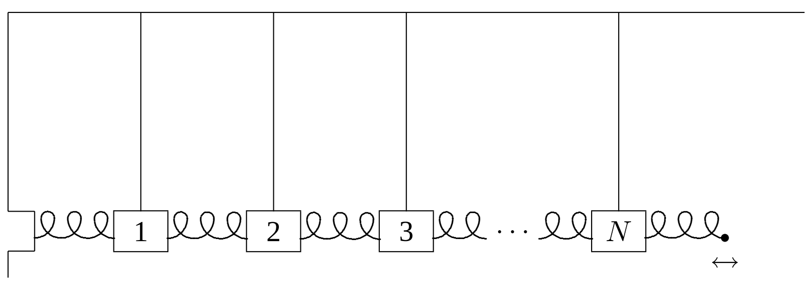


Figure 5.17: A forced oscillation problem in a space translation invariant system.

This is the system of (5.1), except that one wall has been removed and the end of the spring is constrained by some external agency to move back and forth with a displacement

$$z\ \cos\omega_d t\,. \tag{5.49}$$

As usual, in a forced oscillation problem, we first consider the driving term, in this case the fixed displacement of the $N + 1$st block, (5.49), to be the real part of a complex exponential driving term,

$$z\, e^{-i\omega_d t}\,. \tag{5.50}$$

Then we look for a steady state solution in which the entire system is oscillating with the driving frequency ω_d, with the irreducible time dependence, $e^{-i\omega_d t}$.

If there is damping from a frictional force, no matter how small, this will be the steady state solution that survives after all the free oscillations have decayed away. We can find such solutions by the same sort of trick that we used to find the modes of free oscillation of the system. We look for modes of the infinite system and put them together to satisfy boundary conditions.

This situation is different from the free oscillation problem. In a typical free oscillation problem, the boundary conditions fix k. Then we determine ω from the dispersion relation.

In this case, the boundary conditions determine ω_d instead. Now we must use the dispersion relation, (5.35), to find the wave number k.

Solving (5.35) gives

$$k = \frac{1}{a} \cos^{-1} \frac{2B - \omega_d^2}{2C} . \tag{5.51}$$

We must combine the modes of the infinite system, $e^{\pm ikx}$, to satisfy the boundary conditions at $x = 0$ and $x = (N+1)a = L$. As for the system (5.1), the condition that the system be stationary at $x = 0$ leads to a mode of the form

$$\psi(x,t) = y \, \sin kx \, e^{-i\omega_d t} \tag{5.52}$$

for some amplitude y. **But now the condition at $x = L = (N+1)a$ determines not the wave number (that is already fixed by the dispersion relation), but the amplitude y.**

$$\psi(L,t) = y \, \sin kL \, e^{-i\omega_d t} = z \, e^{-i\omega_d t} . \tag{5.53}$$

Thus

$$y = \frac{z}{\sin kL} . \tag{5.54}$$

Notice that if ω_d is a normal mode frequency of the system (5.1) with no damping, then (5.54) doesn't make sense because $\sin kL$ vanishes. That is as it should be. It corresponds to the infinite amplitude produced by a driving force on resonance with a normal frequency of a frictionless system. In the presence of damping, however, as we will discuss in chapter 8, the wave number k is complex because the dispersion relation is complex. We will see later that if k is complex, $\sin kL$ cannot vanish. Even if the damping is very small, of course, we do not get a real infinity in the amplitude as we go to the resonance. Eventually, nonlinear effects take over. Whether it is nonlinearity or the damping that is more important near any given resonance depends on the details of the physical system.[5]

5.5.1 Forced Oscillations with a Free End

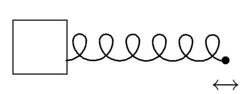


Figure 5.18: Forced oscillation of a mass on a spring.

[5]Note also that, when $\sin kL$ is complex, the parts of the system do not all oscillate in phase, even though all oscillate at the same frequency.

As another example, we will now discuss again the forced longitudinal oscillations of the simple system of a mass on a spring, shown in figure 5.18. The physics here is the same as that of the system in figure 2.9, except that to begin with, we will ignore damping. The block has mass m. The spring has spring constant K and equilibrium length a. To be specific, imagine that this block sits on a nearly frictionless table, and that you are holding onto the other end of the spring, moving it back and forth along the table, parallel to the direction of the spring, with displacement

$$d_0 \cos \omega_d t \,. \tag{5.55}$$

The question is, how does the block move? We already know how to solve this problem from chapter 2. Now we will do it in a different way, using space translation invariance, local interactions and boundary conditions. It may seem surprising that we can treat this problem using the techniques we have developed to deal with space translation invariant systems, because there is only one block. Nevertheless, that is what we are going to do. Certainly nothing prevents us from extending this system to an infinite system by repeating the block-spring combination. The infinite system then has the dispersion relation of the beaded string (or of the coupled pendulum for $\ell \to \infty$):

$$\omega_d^2 = \frac{4K}{m} \sin^2 \frac{ka}{2} \,. \tag{5.56}$$

The relevant part of the infinite system is shown in figure 5.19. The point is that we can impose boundary conditions on the infinite system, figure 5.19, that make it equivalent to figure 5.18.

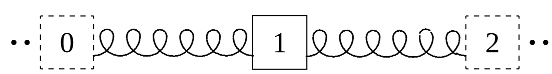


Figure 5.19: Part of the infinite system.

We begin by imagining that the displacement is complex, $d_0 \, e^{-i\omega_d t}$, so that at the end, we will take the real part to recover the real result of (5.55). Thus, we take

$$\psi_2(t) = d_0 \, e^{-i\omega_d t} \,. \tag{5.57}$$

Then to ensure that there is no force on block 1 from the imaginary spring on the left, we must take

$$\psi_0(t) = \psi_1(t) \,. \tag{5.58}$$

To satisfy (5.58), we can argue as in figure 5.13 that

$$\psi(x,t) = z(t) \cos kx \tag{5.59}$$

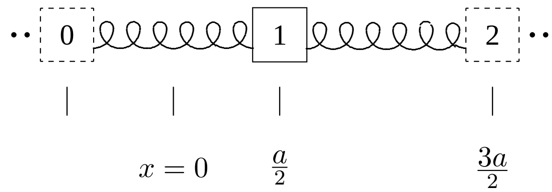


Figure 5.20: A better definition of the zero of x.

where x is defined as shown in figure 5.20.

Now since the equilibrium position of block 2 is $3a/2$, we substitute

$$\psi_2(t) = z(t)\cos\frac{3ka}{2} \tag{5.60}$$

into (5.57), to obtain

$$z(t) = \frac{d_0}{\cos\frac{3ka}{2}} e^{-i\omega_d t} \,. \tag{5.61}$$

Then the final result is

$$\psi_1(t) = \frac{\cos\frac{ka}{2}}{\cos\frac{3ka}{2}}\, d_0\, e^{-i\omega_d t} \tag{5.62}$$

or in real form

$$\psi_1(t) = \frac{\cos\frac{ka}{2}}{\cos\frac{3ka}{2}}\, d_0\, \cos\omega_d t \,. \tag{5.63}$$

We can now use the dispersion relation. First use trigonometry,

$$\cos 3y = \cos^3 y - 3\cos y \sin^2 y = \cos y\,\left(1 - 4\sin^2 y\right) \tag{5.64}$$

to write

$$\psi_1(t) = \frac{1}{1 - 4\sin^2\frac{ka}{2}}\, d_0\, \cos\omega_d t \tag{5.65}$$

or substituting (5.56),

$$\psi_1(t) = \frac{\omega_0^2}{\omega_0^2 - \omega_d^2}\, d_0\, \cos\omega_d t \,, \tag{5.66}$$

where ω_0 is the free oscillation frequency of the system,

$$\omega_0^2 = \frac{K}{m} \,. \tag{5.67}$$

This is exactly the same resonance formula that we got in chapter 2.

5.5.2 Generalization

The real advantage of the procedure we used to solve this problem is that it is easy to generalize it. For example, suppose we look at the system shown in figure 5.21.

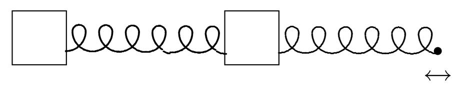


Figure 5.21: A system with two blocks.

Here we can go to the same infinite system and argue that the solution is proportional to $\cos kx$ where x is defined as shown in figure 5.22. Then the same argument leads to the result for the displacements of blocks 1 and 2:

$$\psi_1(t) = \frac{\cos\frac{ka}{2}}{\cos\frac{5ka}{2}}\, d_0\ \cos\omega_d t\,, \quad \psi_2(t) = \frac{\cos\frac{3ka}{2}}{\cos\frac{5ka}{2}}\, d_0\ \cos\omega_d t\,. \tag{5.68}$$

You should be able to generalize this to arbitrary numbers of blocks.

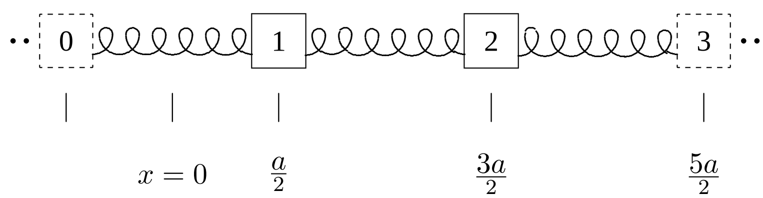


Figure 5.22: The infinite system.

5.6 Coupled LC Circuits

We saw in chapter 1 the analogy between the LC circuit in figure 1.10 and a corresponding system of a mass and springs in figure 1.11. In this section, we discuss what happens when we put LC circuits together into a space translation invariant system.

For example, consider an infinite space translation invariant circuit, a piece of which is shown in figure 5.23. One might guess, on the basis of the discussion in chapter 1, that the circuit in figure 5.23 is analogous to the combination of springs and masses shown in

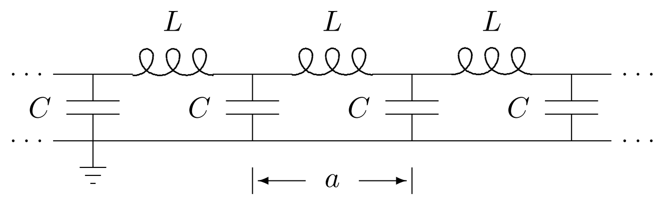


Figure 5.23: A an infinite system of coupled LC circuits.

figure 5.24, with the correspondence between the two systems being:

$$\begin{aligned} m &\leftrightarrow L \\ K &\leftrightarrow 1/C \\ x_j &\leftrightarrow Q_j \end{aligned} \tag{5.69}$$

where x_j is the displacement of the jth block to the right and Q_j is the charge that has been "displaced" through the jth inductor from the equilibrium situation with the capacitors uncharged. In fact, this is right, and we could use (5.69) to write down the dispersion relation for the figure 5.23. However, with our powerful tools of linearity and space translation invariance, we can solve the problem from scratch without too much effort. The strategy will be to write down what we know the solution has to look like, from space translation invariance, and then work backwards to find the dispersion relation.

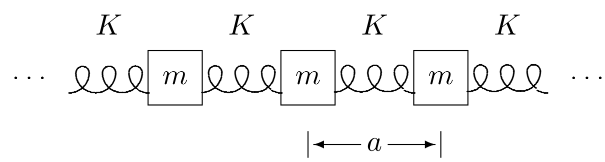


Figure 5.24: A mechanical system analogous to figure 5.23.

The starting point should be familiar by now. **Because the system is linear and space translation invariant, the modes of the infinite system are proportional to $e^{\pm ikx}$. Therefore all physical quantities in a mode, voltages, charges, currents, whatever, must also be proportional to $e^{\pm ikx}$.** In this case the variable, x, is really just a label. The electrical properties of the circuit do not depend very much on the disposition of the elements in space.[6]

[6] This is not exactly true, however. Relativity imposes constraints. See chapter 11.

The dispersion relation will depend only on ka, where a is the separation between the identical parts of the system (see (5.35)). However, it is easier to think about the system if it is physically laid out into a space translation invariant configuration, as shown in figure 5.23.

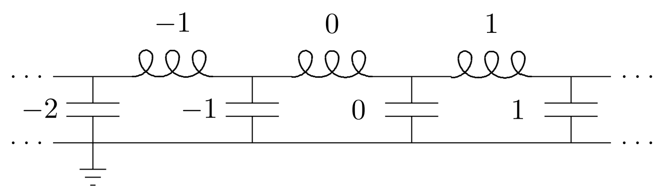


Figure 5.25: A labeling for the infinite system of coupled LC circuits.

In particular, let us label the inductors and capacitors as shown in figure 5.25. Then the charge displaced through the jth inductor in the mode with angular wave number, k, is

$$Q_j(t) = q\, e^{ijka} e^{-i\omega t} \tag{5.70}$$

for some constant charge, q. Note that we could just as well take the time dependence to be $\cos\omega t$, $\sin\omega t$, or $e^{i\omega t}$. It does not matter for the argument below. What matters is that when we differentiate $Q_j(t)$ twice with respect to time, we get $-\omega^2 Q_j(t)$. The current through the jth inductor is

$$I_j = \frac{d}{dt} Q_j(t) = -i\omega q\, e^{ijka} e^{-i\omega t}\,. \tag{5.71}$$

The charge on the jth capacitor, which we will call q_j, is also proportional to $e^{ijka}e^{-i\omega t}$, but in fact, we can also compute it directly. The charge, q_j, is just

$$q_j = Q_j - Q_{j+1} \tag{5.72}$$

because the charge displaced through the jth inductor must either flow onto the jth capacitor or be displaced through the j+1st inductor, so that $Q_j = q_j + Q_{j+1}$. Now we can compute the voltage, V_j, of each capacitor,

$$V_j = \frac{1}{C}\left(Q_j - Q_{j+1}\right) = \frac{q}{C}\left(1 - e^{ika}\right) e^{ijka} e^{-i\omega t}\,, \tag{5.73}$$

and then compute the voltage drop across the inductors,

$$L\frac{dI_j}{dt} = V_{j-1} - V_j\,, \tag{5.74}$$

inserting (5.71) and (5.73) into (5.74), and dividing both sides by the common factor $-qL\, e^{ijka}e^{-i\omega t}$, we get the dispersion relation,

$$\omega^2 = -\frac{1}{LC}\left(1 - e^{ika}\right)\left(e^{-ika} - 1\right) = \frac{4}{LC}\sin^2\frac{ka}{2}\,. \tag{5.75}$$

This corresponds to (5.37) with $B = 1/LC$. This is just what we expect from (5.69). We will call (5.75) **the dispersion relation for coupled LC circuits.**

5.6.1 An Example of Coupled LC Circuits

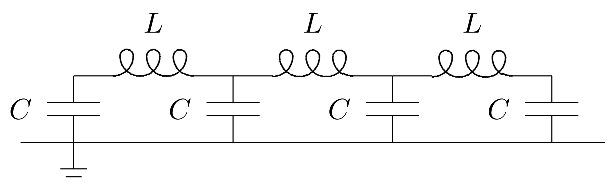


Figure 5.26: A circuit with three inductors.

Let us use the results of this section to study a finite example, with boundary conditions. Consider the circuit shown in figure 5.26. This circuit in figure 5.26 is analogous to the combination of springs and masses shown in figure 5.27.

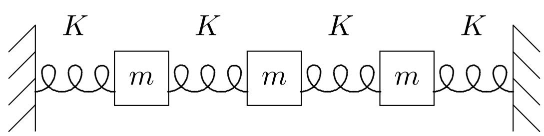


Figure 5.27: A mechanical system analogous to figure 5.26.

We already know that this is true for the middle. It remains only to understand the boundary conditions at the ends. If we label the inductors as shown in figure 5.28, then we can imagine that this system is part of the infinite system shown in figure 5.23, with the charges constrained to satisfy

$$Q_0 = Q_4 = 0\,. \tag{5.76}$$

This must be right. No charge can be displaced through inductors 0 and 4, because in figure 5.26, they do not exist. This is just what we expect from the analogy to the system in (5.27), where the displacement of the 0 and 4 blocks must vanish, because they are taking the place of the fixed walls.

Now we can immediately write down the solution for the normal modes, in analogy with (5.21) and (5.22),

$$Q_j \propto \sin \frac{jn}{4} \tag{5.77}$$

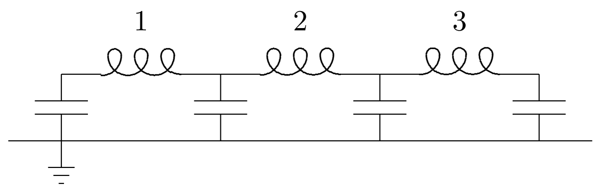


Figure 5.28: A labeling of the inductors in figure 5.26.

for $n = 1$ to 3.

5.6.2 A Forced Oscillation Problem for Coupled LC Circuits

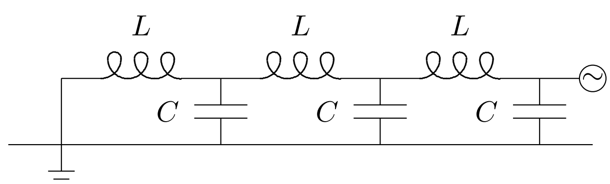


Figure 5.29: A forced oscillation with three inductors.

One more somewhat more practical example may be instructive. Consider the circuit shown in figure 5.29. The $\sim$ in figure 5.29 stands for a source of harmonically varying voltage. We will assume that the voltage at this point in the circuit is fixed by the source, $\sim$, to be

$$V \cos \omega t \,. \tag{5.78}$$

We would like to find the voltages at the other nodes of the system, as shown in figure 5.30, with

$$V_3 = V \cos \omega t \,. \tag{5.79}$$

We could solve this problem using the displaced charges, however, it is a little easier to use the fact that **all** the physical quantities in the infinite system in figure 5.23 are proportional to e^{ikx} in a mode with angular wave number k. Because this is a forced oscillation problem (and because, as usual, we are ignoring possible free oscillations of the system and looking for the steady state solution), k is determined from ω, by the dispersion relation for the infinite system of coupled LC circuits, (5.75).

The other thing we need is that

$$V_0 = 0 \,, \tag{5.80}$$

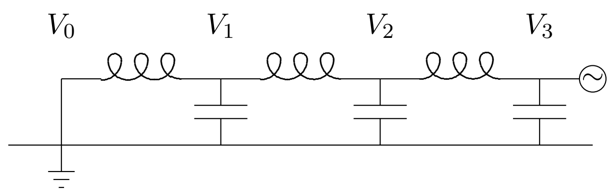


Figure 5.30: The voltages in the system of figure 5.29.

because the circuit is shorted out at the end. Thus we must combine the two modes of the infinite system, $e^{\pm ikx}$, into $\sin kx$, and the solution has the form

$$V_j \propto \sin jka\,. \tag{5.81}$$

We can satisfy the boundary condition at the other end by taking

$$V_j = \frac{V}{\sin 3ka} \sin jka \cos \omega t\,. \tag{5.82}$$

This is the solution.

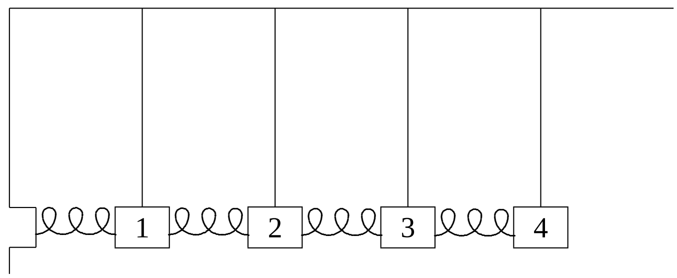
Chapter Checklist

You should now be able to:

1. Recognize a finite system as part of a space translation invariant infinite system;
2. Find the normal modes of the finite system as linear combinations of normal modes of the space translation invariant infinite system, consistent with the physics of the boundaries, by imposing boundary conditions;
3. Describe the normal modes of a space translation invariant system in terms of an angular wave number, k;
4. Find the dispersion relation that relates the angular frequency, ω, to the angular wave number, k;
5. Solve forced oscillation problems using boundary conditions;
6. Analyze space translation invariant systems of coupled LC circuits.

Problems

5.1. Consider the small longitudinal oscillations of the system shown below:



In the picture above, each bob has mass m, each pendulum has length ℓ, each spring has spring constant κ, and the equilibrium separation between bobs is a.

a. Find the $M^{-1}K$ matrix for this system in the basis in which the displacements of the blocks from equilibrium are all measured to the right and arranged into vector in the obvious way,

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{pmatrix} .$$

b. Classify as TRUE or FALSE each of the following questions about the normal modes of this system. If possible, explain your answers qualitatively, that is, in words, rather than by plugging into a formula, and discuss the generality of your results.

i. In the normal mode with the lowest frequency, all the blocks move in the same direction when they are moving at all.

ii. In the normal mode with the second lowest frequency, the 1st and 2nd blocks have the same displacement.

iii. In the normal mode with the highest frequency, neighboring blocks move in opposite directions when they are moving at all.

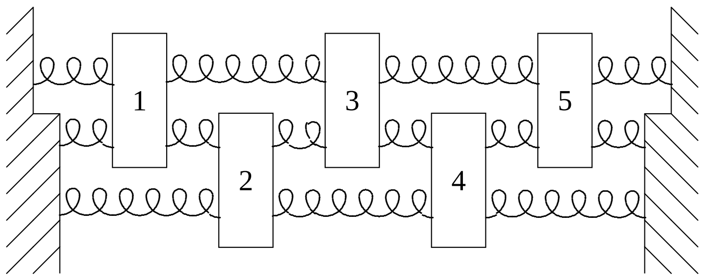
c. Find the angular frequencies of each of the normal modes. **Hint:** You may want to use the dispersion relation for coupled pendulums,

$$\omega^2 = 2B - 2C \cos ka$$

where

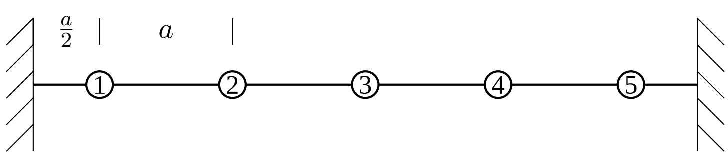
$$B = \frac{g}{2\ell} + \frac{\kappa}{m}, \quad C = \frac{\kappa}{m}.$$

5.2.



In the system shown above, all the blocks have mass m and they are constrained to move only horizontally. The long springs with six loops have spring constant K. The shorter springs, with three loops, have spring constant $2K$. The shortest springs, with two loops, have spring constant $3K$. As you will see in chapter 7, this is what we expect if the springs are all made out of the same material (see figure 7.1). Find the normal modes of the system and the corresponding frequencies. Make sure that you justify any assumptions you make about the normal modes. **Hint:** Try to find an infinite system with space translation invariance that contains this in such a way that you can put in the physics of the walls as a boundary condition. **Another Hint:** This works simply only if the three loop springs have exactly twice the spring constant of the long springs. Your answer should explain why.

5.3. In the beaded string shown below, the interval between neighboring beads is a, and the distance from the end beads to the walls is $a/2$. All the beads have mass m and are constrained to move only vertically, in the plane of the paper.

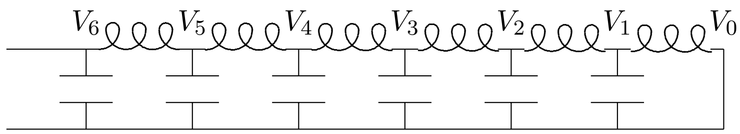


Show that the physics of the left-hand wall can be incorporated by going to an infinite system and requiring the boundary condition $A_0 = -A_1$.

a. **Easy.** Find the analogous boundary condition for the right-hand wall.

b. Find the normal modes and the corresponding frequencies.

5.4. Consider the following circuit:



All the capacitors have the same capacitance, $C \approx 0.00667\mu F$, and all the inductors have the same inductance, $L \approx 150\mu H$ and no resistance. The center wire is grounded. This circuit is an electrical analog of the space translation invariant systems of coupled mechanical oscillators that we have discussed in this chapter.

When you apply a harmonically oscillating signal from a signal generator through a coaxial cable to V_6, different oscillating voltages will be induced along the line. That is if

$$V_6(t) = V \cos \omega t \,,$$

then $V_j(t)$ has the form

$$V_j(t) = A_j \cos \omega t + B_j \sin \omega t \,.$$

Find A_j and B_j.

Chapter 6

Continuum Limit and Fourier Series

"Continuous" is in the eye of the beholder. Most systems that we think of as continuous are actually made up of discrete pieces. In this chapter, we show that a discrete system can look continuous at distance scales much larger than the separation between the parts. We will also explore the physics and mathematics of Fourier series.

Preview

In this chapter, we discuss the wave equation, the starting point for some other treatments of waves. We will get it as natural result of our general principles of space translation invariance and local interactions applied to continuous systems.

i. We will study the discrete space translation invariant systems discussed in the previous chapter in the limit that the separation between parts goes to zero. We will argue that the generic result is a continuous system obeying the wave equation.

ii. The continuum limit of the beaded string is a continuous string with transverse oscillations. We will discuss its normal modes for a variety of boundary conditions. We will see that the normal modes of a continuous space translation invariant system are the same as those of a finite system. The only difference is that there are an infinite number of them. The sum over the infinite number of normal modes required to solve the initial value problem for such a continuous system is called a Fourier series.

6.1 The Continuum Limit

Consider a discrete space translation invariant system in which the separation between neighboring masses is a. **If a is very small, the discrete system looks continuous.** To understand this statement, consider the action of the $M^{-1}K$ matrix, (5.8), in the notation of the last chapter in which the degrees of freedom are labeled by their equilibrium positions. The matrix $M^{-1}K$ acts on a vector to produce another vector. We have replaced our vectors by functions of x, so $M^{-1}K$ is something that acts on a function $A(x)$ to give another function. Let's call it $M^{-1}KA(x)$. It is easiest to see what is happening for the beaded string, for which $B = C = T/ma$. Then

$$M^{-1}KA(x) = \left(\frac{T}{ma}\right)(2A(x) - A(x+a) - A(x-a)) \,. \tag{6.1}$$

So far, (6.1) is correct for any a, large or small.

Whenever you say that a dimensional quantity, like the length a, is large or small, you must specify a quantity for comparison. You must say large or small compared to what?[1] In this case, the other dimensional quantity in the problem with the dimensions of length is the wavelength of the mode that we are interested in. Now here is where small a enters. If we are interested only in modes with a wavelength $\lambda = 2\pi/k$ that is very large compared to a, then ka is a very small dimensionless number and $A(x+a)$ is very close to $A(x)$. We can expand it in a Taylor series that is rapidly convergent. Expanding (6.1) in a Taylor series gives

$$M^{-1}K\,A(x) = -\frac{Ta}{m}\,\frac{\partial^2 A(x)}{\partial x^2} + \cdots \tag{6.2}$$

where the $\cdots$ represent higher derivative terms that are smaller by powers of the small number ka than the first term in (6.2). In the limit in which we take a to be really tiny (always compared to the wavelengths we want to study) we can replace m/a by the linear mass density ρ_L, or mass per unit length of the now almost continuous string and ignore the higher order terms. In this limit, we can replace the $M^{-1}K$ matrix by the combination of derivatives that appear in the first surviving term of the Taylor series, (6.2),

$$M^{-1}K \to -\frac{T}{\rho_L}\,\frac{\partial^2}{\partial x^2} \,. \tag{6.3}$$

Then the equation of motion for $\psi(x,t)$ becomes the wave equation:

$$\frac{\partial^2}{\partial t^2}\psi(x,t) = \frac{T}{\rho_L}\,\frac{\partial^2}{\partial x^2}\psi(x,t) \,. \tag{6.4}$$

The dispersion relation is

$$\omega^2 = \frac{T}{\rho_L}\,k^2 \,. \tag{6.5}$$

[1] A dimensionless quantity does not require this step. A dimensionless number is large if it is much greater than one and small if it is much smaller than one.

This can be seen directly by plugging the normal mode e^{ikx} into (6.4), or by taking the limit of (5.37)-(5.38) as $a \rightarrow 0$. **Equation (6.5) is the dispersion relation for the ideal continuous string.** The quantity, $\sqrt{T/\rho_L}$, has the dimensions of velocity. It is called the "phase velocity", v_φ. As we will discuss in much more detail in chapter 8 and following, this is the speed with which traveling waves move on the string.

We will call the approximation of replacing a discrete system with a continuous system that looks approximately the same for $k \gg 1/a$ the **continuum approximation**. Really, all of the mechanical systems that we will consider are discrete, at least on the atomic level. However, if we are concerned only about waves with macroscopic wavelengths, the continuum approximation is a very good one.

6.1.1 Philosophy and Speculation

Our treatment of the wave equation in (6.4) is a little unusual. In many treatments of wave phenomena, the wave equation is given a place of honor. In fact, the wave equation is only a restatement of the dispersion relation, (6.5), which is usually just an approximation to what is really going on. Almost all of the systems that we usually treat with the wave equation are actually discrete at very small distances. We cannot really get all the way to the continuum limit that gives (6.5). Light waves, which we will study in the chapters to come, for all we know, may be an exception to this rule, and be completely continuous. However, we don't really have the right to assume even that. It could be that at very short distances, far below anything we can look at today, the nature of light and even of space and time changes in some way so that space and time themselves have some tiny characteristic length scale a. **The analysis above shows that this doesn't matter!** As long as we can only look at space and time at distances much larger than a, they look continuous to us. Then because we are scientists, concerned about how the world looks in our experiments, and not how it behaves in some ideal regime far beyond what we can probe experimentally, we might as well treat them as continuous.

6.2 Fourier series

6.2.1 The String with Fixed Ends

6-1

If we stretch our continuous string between fixed walls so that $\psi(0) = \psi(\ell) = 0$, the modes are given by (5.33) and (5.34), just as for the discrete system. The only difference is that now n runs from 1 to ∞, or at least to such large n that the wavelength $2\pi/k = 2\ell/n$ is so small that the continuum approximation breaks down. This follows from (5.28), which because k is real here becomes

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a}\,. \tag{6.6}$$

As $a \to 0$ the allowed range of k increases to infinity.

These standing wave modes are animated in program 6-1 on the program disk, assuming the dispersion relation, (6.5).

We can now discuss the physical basis of the Fourier series. In (3.77) in chapter 3, we showed that the normal modes for a discrete system are linearly independent and complete. That means that any displacement of the discrete system can be written as a unique linear combination of the normal modes. Physically, this must be so to allow us to solve the initial value problem. Our picture of the continuous string is a limit of the beaded string in which the number of beads, N, goes to infinity and the beads get infinitely close together. For each N, the most general displacement of the system can be expanded as a linear combination of the N normal modes. If the limit $N \to \infty$ is reasonably well behaved, we might expect that the most general displacement of the limiting continuous string could be expanded in terms of the infinite number of normal modes of the continuous system. This expansion is a Fourier series. The displacement of the continuous system is described by a function of the position along the string. If the function is not too discontinuous, the expansion in normal modes works fine.

Consider the continuous string, stretched between fixed walls at $x = 0$ and $x = \ell$. The transverse displacement of this system at any time is described by a continuous function of x, $\psi(x)$ with

$$\psi(0) = \psi(\ell) = 0\,. \tag{6.7}$$

Thus we expect from the argument above that we can express any function that is not too discontinuous and satisfies (6.7) as a sum of the normal modes given by (5.33) and (5.34),

$$\psi(x) = \sum_{n=1}^{\infty} c_n \, \sin \frac{n\pi x}{\ell}\,. \tag{6.8}$$

The constants, c_n, are called the "Fourier coefficients." They can be found using the following identity:

$$\int_0^\ell dx \, \sin \frac{n\pi x}{\ell} \, \sin \frac{n'\pi x}{\ell} = \begin{cases} \ell/2 \text{ if } n = n' \\ 0 \text{ if } n \neq n' \end{cases} \tag{6.9}$$

so that

$$c_n = \frac{2}{\ell} \int_0^\ell dx \, \sin \frac{n\pi x}{\ell} \, \psi(x)\,. \tag{6.10}$$

This is just the method of normal coordinates adapted to the continuous situation.

6.2.2 Free Ends

6-2

Equation (6.8) is called the Fourier series for a function satisfying (6.7). Other boundary conditions yield different series. For example, consider a string with the $x = 0$ end fixed at

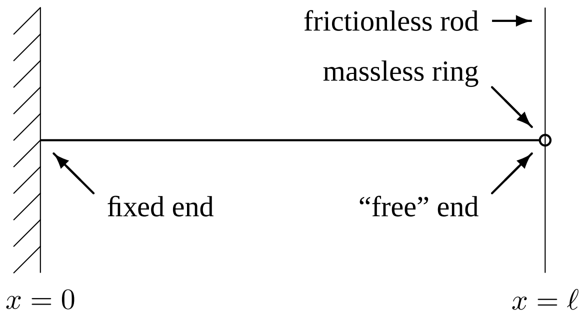


Figure 6.1: A continuous string with one end free to oscillate in the transverse direction.

$z = 0$. Suppose that the other end, at $x = \ell$ is attached to a massless ring that is free to slide along a frictionless rod in the z direction, as shown in figure 6.1. We say that this system has one "free end" because the end at $x = \ell$ is free to slide in the transverse direction, even though it is fixed in the x direction.

Because the rod is frictionless, the force on the ring due to the rod must have no component in the z direction. But because the ring is massless, the total force on the ring must vanish. Therefore, the force on the ring due to the string must have no component in the z direction. That implies that the string is horizontal at $x = \ell$. But the shape of the string at any given time is given by the graph of the transverse displacement, $\psi(x,t)$ versus x.[2] Thus the slope of $\psi(x,t)$ at $x = \ell$ must vanish. Therefore, the appropriate boundary conditions for the displacement is

$$\psi(0,t) = 0\,, \quad \frac{\partial}{\partial x}\,\psi(x,t)|_{x=\ell} = 0\,. \tag{6.11}$$

This implies that the normal modes also satisfy similar boundary conditions:

$$A_n(0) = 0\,, \quad A_n'(\ell) = 0\,. \tag{6.12}$$

The first condition implies that the solution must have the form

$$A_n(x) \propto \sin k_n x \tag{6.13}$$

for some k_n. The second condition determines the possible values of k_n. It implies that $\sin k_n x$ must have a maximum or minimum at $x = \ell$ which, in turn, implies that

$$k_n \ell = \frac{\pi}{2} + n\pi \tag{6.14}$$

[2]This is why transverse oscillations are easier to visualize than longitudinal oscillations — compare with (7.5).

where n is a nonnegative integer (nonnegative because we can choose all the $k_n > 0$ in (6.13) — negative values just change the sign of $A_n(x)$ and do not lead to new solutions). The solutions have the form

$$\sin\left(\frac{(2n+1)\pi x}{2\ell}\right) \qquad \text{for } n = 0 \text{ to } \infty. \tag{6.15}$$

These normal modes are animated in program 6-2. With these normal modes, we can describe an arbitrary function, $\psi(x)$, satisfying the boundary conditions for this system, (6.11).

$$\psi(0) = 0\,, \quad \psi'(\ell) = 0\,. \tag{6.16}$$

Thus for such a function, we can write

$$\psi(x) = \sum_{n=1}^{\infty} c_n \sin\left(\frac{(2n+1)\pi x}{2\ell}\right) \tag{6.17}$$

where

$$c_n = \frac{2}{\ell}\int_0^\ell dx\,\sin\left(\frac{(2n+1)\pi x}{2\ell}\right)\psi(x)\,. \tag{6.18}$$

6.2.3 Examples of Fourier Series

6-3

Let us find the Fourier coefficients for the following function, defined in the interval [0,1]:

$$\psi(x) = \begin{cases} x & \text{for} \quad x \le w\,, \\ \dfrac{w(1-x)}{1-w} & \text{for} \quad x > w\,. \end{cases} \tag{6.19}$$

For definiteness, we will take $w = 0.75$, so the function $\psi(x)$ has the form shown in figure 6.2.

We compute the Fourier coefficients using (6.10). Because $\ell = 1$, this has the following form (see problem (6.2)):

$$\begin{aligned} c_n &= \int_0^1 dx\,\sin n\pi x\,\psi(x) \\ &= \int_0^w dx\,x\sin n\pi x + \frac{w}{1-w}\int_w^1 dx\,(1-x)\sin n\pi x \\ &= \frac{\sin n\pi w}{(1-w)n^2\pi^2}\,. \end{aligned} \tag{6.20}$$

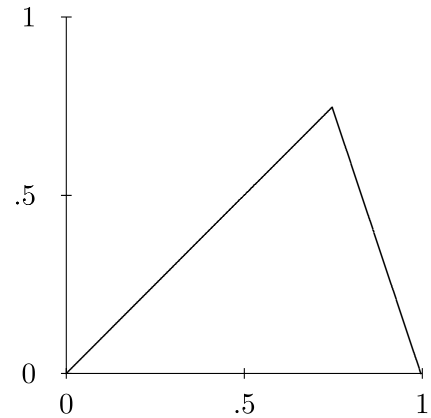


Figure 6.2: The function $\psi(x)$ for $w = 0.75$.

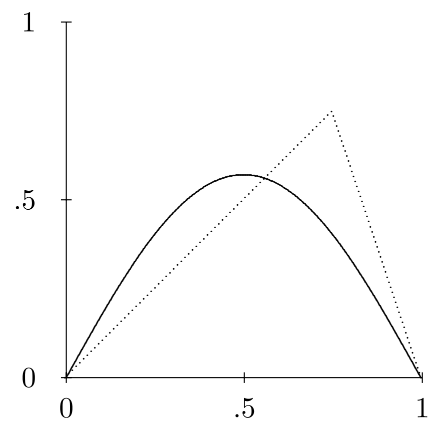


Figure 6.3: The first term in the Fourier series for $\psi(x)$. The dotted line is $\psi(x)$.

We can reconstruct the function, $\psi(x)$, as a sum over the normal modes of the string. Let us look at the first few terms in the series to get a feeling for how this works. The first term in the sum, for $w = 0.75$, is shown in figure 6.3. This is a lousy approximation, necessarily, because the function is not symmetrical about $x = 1/2$, while the first term in the sum is symmetrical. The first two terms are shown in figure 6.4. This looks much better.

The first six terms are shown in figure 6.5. This is now a pretty good approximation except where the function has a kink.

What is going on here is that if we include terms in the Fourier series only up to $n = N$,

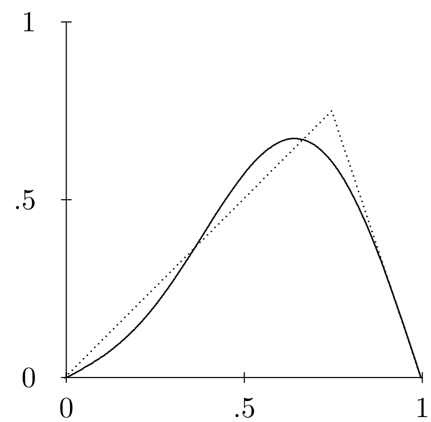


Figure 6.4: The sum of the first two terms in the Fourier series for $\psi(x)$. The dotted line is $\psi(x)$.

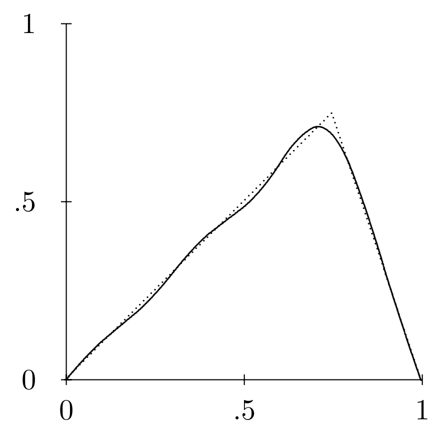


Figure 6.5: The sum of the first six terms in the Fourier series for $\psi(x)$. The dotted line is $\psi(x)$.

the truncated Fourier series

$$\psi(x) = \sum_{n=1}^{N} c_n \sin n\pi x \tag{6.21}$$

does not include any modes with very small wavelengths. The smallest wavelength that appears (for the highest angular wave number) is $2/N$ (no dimensions here because we took $a = 1$). Thus while the Fourier series can describe any features of the shape of the function that are larger than $2/N$, there is no way that it can pick up features that are much smaller. In this example, because the function has an infinitely sharp kink, the Fourier series never gets very good near $x = w$. However, eventually the discrepancy is squeezed into such a small region around the kink that the result will look OK to the naked eye.

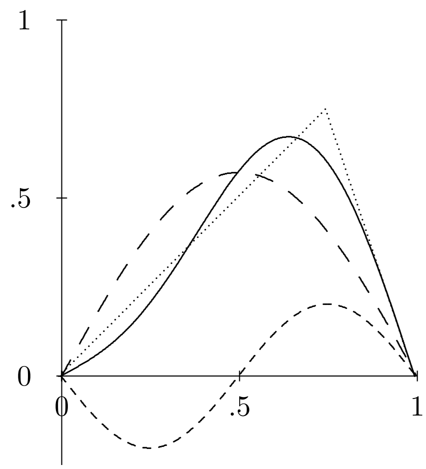


Figure 6.6: The first two terms in the Fourier series for $\psi(x)$ and their sum.

You can see how this works in more detail by studying figure 6.6. The curve of long dashes is the first term in the Fourier series. Evidently, it is less than the function, $\psi(x)$ (the dotted triangle), for large x and greater than $\psi(x)$ for small x. The sign and magnitude of the second term in the Fourier series, the curve of short dashes in figure 6.6, is chosen to make up for this discrepancy, so that the sum (the solid curve) is much closer to the actual function. The same process is repeated over and over again as you go to higher order in the truncated Fourier series.

You can play with the truncated Fourier series for the function $\psi(x)$ in program 6-3. This program allows you to vary the parameter w, and also the number of terms in the Fourier series. You should look at what happens near $w = 1$. You might think that this would cause problems for the Fourier series because the $(1-w)$ in the denominator of (6.20) goes to zero. However, the limit is actually well behaved because $\sin n\pi\omega$ also goes to zero as $w \to 0$. Nevertheless, the Fourier series has to work hard for $w = 1$ to reproduce a function that does not go to zero for $x = 1$ as a sum of sine functions, each of which do vanish at $x = 1$. This difficulty is reflected in the wiggles near $x = 1$ for any reasonable number of terms in the Fourier series.

6.2.4 Plucking a String

6-4

Let us now use this mathematics to solve a physics problem. We will solve the initial value problem for the string with fixed end for a particular initial shape. The initial value problem here is almost exactly like that discussed in chapter 3, (3.98)-(3.100), for a system with a finite number of degrees of freedom. The only difference is that now, because the number of degrees of freedom is infinite, the sum over modes runs to infinity. You shouldn't worry about the fact that the number of modes is infinite. What that "infinity" really means is "larger than any number we are going to care about." In practice, as we saw in the examples above, the higher modes eventually don't make much difference. They are associated with smaller and smaller features of the shape. When we say that the system is continuous and that it has an infinite number of degrees of freedom, we are actually assuming that the smallest features that we care about in the waves are still much larger than the distance between pieces of the system, so that we can truncate our Fourier series far below the limit and still have a good approximate description of the motion.

Suppose we pluck the string. Specifically, suppose that the string has linear mass density ρ_L, tension T, and fixed ends at $x = 0$ and ℓ. Suppose further that at time $t = 0$ the string is at rest, but pulled out of its equilibrium position into the shape, $\psi(x)$, given by (6.19). If the string is then released at $t = 0$, we can find the subsequent motion by summing over all the normal modes with fixed coefficients multiplied by $\cos\omega_n t$ and/or $\sin\omega_n t$, where ω_n is the frequency of the mode $\sin\frac{n\pi x}{\ell}$ with $k = \frac{n\pi}{\ell}$ (the frequency is given by (6.5))

$$\omega_n = \sqrt{\frac{T}{\rho_L}}\, k_n = \sqrt{\frac{T}{\rho_L}}\, \frac{n\pi}{\ell}\,. \tag{6.22}$$

In this case, only the $\cos\omega_n t$ terms appear, because the velocity is zero at $t = 0$. Thus we can write

$$\psi(x,t) = \sum_{n=1}^{\infty} c_n \,\sin\frac{n\pi x}{\ell}\, \cos\omega_n t\,. \tag{6.23}$$

This satisfies the boundary conditions at $t = 0$, by virtue of the Fourier series, (6.8). The disadvantage of (6.23) is that we are left with an infinite sum. For the simple dispersion relation, (6.5), there are other ways to solve this problem that we will discuss later when we learn about traveling waves. However, the advantage of the solution (6.23) is that it does not depend on the dispersion relation.

We can solve the problem approximately using (6.23) by adding up only the first few terms of the series. The computer can do this quickly. In program 6-4, the first twenty terms of the series are shown for $w = 1/2$ (and the dispersion relation still given by (6.5)). The result is amazingly simple. Check it out! Program 6-5 is the same idea, but allows you to vary w and the number of terms in the Fourier series. Try out $w = 0.75$ and compare with figures 6.3-6.5.

Chapter Checklist

You should now be able to:

i. Take the limit of a space translation invariant discrete system as the distance between the parts goes to zero, interpret the physics of the resulting continuous system, and find its dispersion relation;

ii. Use the Fourier series to set up and solve the initial value problem for a massive string with various boundary conditions.

Problems

6.1. Consider the continuous string of (6.7)-(6.10) as the continuum limit of a beaded string with W beads as $W \to \infty$. Write the analog of (6.8) and (6.10) for finite W. Show that the limit as $W \to \infty$ yields (6.10). **Hint:** This is an exercise in the definition of an integral as the limit of a sum. But to do the first part, you will either need to use normal coordinates, or prove the identity

$$\sum_{k=1}^{W} \sin \frac{nk\pi}{W+1} \sin \frac{n'k\pi}{W+1}$$

$$= \begin{cases} b & \text{if} \quad n = n' \neq 0 \\ 0 & \text{if} \quad n \neq n' \text{ and } n, n' > 0 \end{cases}$$

for a constant b and find b.

6.2. Do the integrals in (6.20). **Hint:** Use integration by parts and watch for miraculous cancellations.

6.3. Find the normal modes of the string with two free ends, shown in figure 6.7.

6.4. **Fun with Fourier Series and Fractals**

In this problem you will explore the Fourier series for an interesting set of functions. Consider a function of the following form, defined on the interval [0,1]:

$$f(t) = \sum_{j=0}^{\infty} h^j \, g(\mathrm{frac}(2^j t)) \, .$$

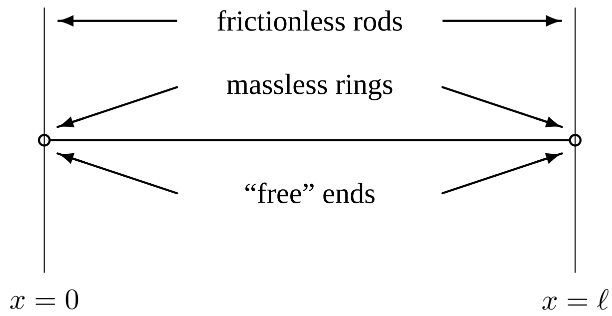


Figure 6.7: A continuous string with both ends free to oscillate in the transverse direction.

where

$$g(t) = \begin{cases} 1 \text{ for } 0 \leq t \leq w \\ 0 \text{ for } w < t < 1 - w \\ 1 \text{ for } 1 - w \leq t \leq 1 \end{cases}$$

and $\text{frac}(x)$ denotes the fractional part, *i.e.* $\text{frac}(4.39) = 0.39$. $f(t)$ thus depends on the two parameters h and w, where $0 < h < 1$ and $0 < w < 1/2$. For example, for $h = 1/2$ and $w = 1/4$, the h^0 term is shown in figure 6.8.

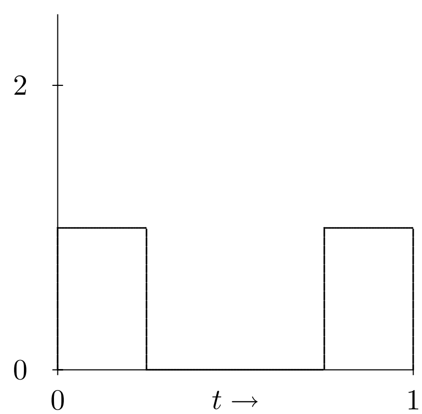


Figure 6.8: The h^0 term in $f(t)$ for $h = 1/2$ and $w = 1/4$.

If we add in the h^1 term we get the picture in figure 6.9.

Adding the h^2 term gives the picture in figure 6.10, and so on.

The final result is a very bumpy function, called a "fractal." You cannot compute this function exactly, but you can include enough terms to get to any desired accuracy. Because

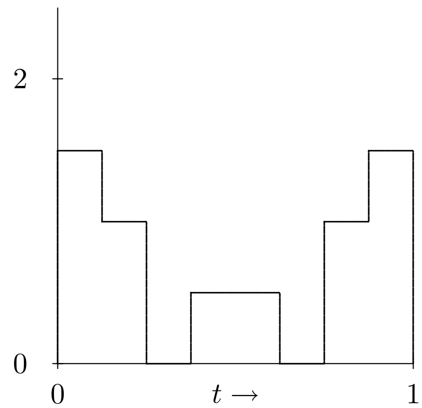


Figure 6.9: The first two terms in $f(t)$ for $h = 1/2$ and $w = 1/4$.

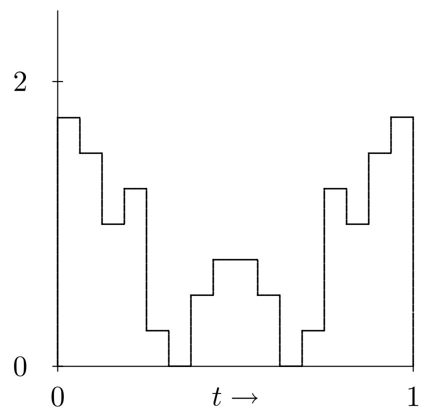


Figure 6.10: The first three terms in $f(t)$ for $h = 1/2$ and $w = 1/4$.

the function is symmetric about $t = 1/2$, it is really only necessary to plot it from 0 to $1/2$. Also because of the symmetry, it can be expressed in terms of a Fourier series of cosines,

$$f(t) = \sum_{k=0}^{\infty} b_k \cos 2\pi kt \,.$$

Show that the Fourier coefficients are given by

$$b_k = \frac{2}{\pi k} \sum_{j=0}^{\xi(k)} (2h)^j \sin(2\pi kw/2^j)$$

for $k \neq 0$, and

$$b_0 = \frac{2w}{1-h}$$

where the function, $\xi(k)$ is the number of times 2 appears as a factor of k. Thus $\xi(0) = \xi(1) = \xi(3) = 0$, $\xi(2) = 1$, $\xi(4) = 2$, *etc.*

Write a program to display and print the fractal for some set of parameters, h and w. Also, display the truncated Fourier series,

$$f_m(t) = \sum_{k=0}^{m-1} b_k \cos 2\pi k t$$

with m terms, for $m = 5$, 10, and 20 (or more if you have a fast computer).

Chapter 7

Longitudinal Oscillations and Sound

Transverse oscillations of a continuous system are easy to visualize because you can see directly the function that describes the displacement. The mathematics of longitudinal oscillations of a continuous linear space translation invariant system is the same. It must be, because it is completely determined by the space translation invariance. But the physics is different.

Preview

In this chapter, we introduce two physical systems with longitudinal oscillations: massive springs and organ pipes.

i. We describe the massive spring as the continuum limit of a system of masses connected by massless springs and study its normal modes for various boundary conditions.

ii. We discuss in some detail the system of a mass at the end of a massive spring. When the spring is "light," this is an important example of physics with two different "scales."

iii. We discuss the physics of sound waves in a tube, by analogy with the oscillations of the massive spring. We also introduce the "Helmholtz" approximation for the lowest mode of a bottle.

7.1 Longitudinal Modes in a Massive Spring

So far, in our extensive discussions of waves in systems of springs and blocks, we have assumed that the only degrees of freedom are those associated with the motion of the blocks. This is a reasonable assumption at low frequencies, when the blocks are very heavy compared to the springs, because the blocks move so slowly that the springs have time to readjust and are always nearly uniform.[1] In this case, the dispersion relation for the longitudinal oscillations of the blocks is just the dispersion relation for coupled pendulums, (5.35), in the limit in which we ignore gravity, and keep only the coupling between the masses produced by the spring constant, K. In other words, we take the limit of (5.35) as $g/\ell \to 0$. The result can be written as

$$\omega^2 = \frac{4K_a}{m} \sin^2 \frac{ka}{2} \tag{7.1}$$

where K_a is the spring constant of the springs, m the mass of the blocks, and a the equilibrium separation. We have put a subscript a on K_a because we will want to vary the spring constant as we vary the separation between the blocks in the discussion below.

Now what happens when the blocks are absent, but the spring is massive? We can find this out by considering the limit of (7.1) as $a \to 0$. In this limit, the massive blocks and the massless spring melt into one another, so that the result looks like a uniform, massive spring. In order to take the limit, however, we must understand what variables describe the massive spring, and have a finite limit as $a \to 0$. One such variable is the linear mass density,

$$\rho_L = \lim_{a \to 0} \frac{m}{a} \,. \tag{7.2}$$

We must take the masses of the blocks to zero as $a \to 0$ in order to keep ρ_L finite.

To understand what happens to K_a as $a \to 0$, consider what happens when you cut a spring in half. When a spring is stretched, each half contributes half the displacement. But the tension is uniform throughout the stretched spring. Thus the spring constant of half a spring is twice as great as that of the full spring, because half the displacement gives the same force. This relation is illustrated in figure 7.1. The spring in the center is unstretched. The spring on top is stretched by x to the right. The bottom shows the **same** stretched spring, still stretched by x, but now symmetrically. Comparing top and bottom, you can see that the return force from stretching the spring by x is the same as from stretching half the spring by $x/2$.

The diagram in figure 7.1 is an example of the following result. In general, the spring constant, K_a, depends not just on what the spring is made of, it depends on how long the spring is. But the quantity $K_a a$, where a is the length of the spring, is actually independent of a, for a spring made of uniform material. Thus we should take the limit $a \to 0$ holding $K_a a$ fixed.

This implies that the dispersion relation for the massive spring is

$$\omega^2 = \frac{K_a a}{\rho_L} \, k^2, \tag{7.3}$$

where we have used the Taylor series expansion of $\sin x$, (1.58), and kept only the first term.

[1]We will say this much more formally below.

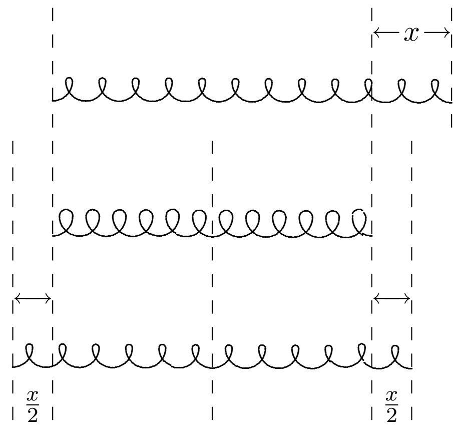


Figure 7.1: Half a spring has twice the spring constant.

According to the discussion above, we can rewrite this as

$$\omega^2 = \frac{K\ell}{\rho_L}\, k^2 \tag{7.4}$$

where ℓ is the length of the spring and K is the spring constant of the spring as a whole.

Note that in longitudinal oscillations in a continuous material in the x direction, the equilibrium position, x, doesn't actually describe the x position of the material. Because the displacement is longitudinal, the actual x position of the point on the spring with equilibrium position x is

$$x + \psi(x,t)\,, \tag{7.5}$$

where ψ is the displacement. You will need this to do problem (7.1).

7.1.1 Fixed Ends

7-1

Suppose that we have a massive spring with length ℓ and its ends fixed at $x = 0$ and $x = \ell$. Then the displacement, $\psi(x,t)$ must vanish at the ends,

$$\psi(0,t) = 0\,, \quad \psi(\ell,t) = 0\,. \tag{7.6}$$

The modes of the system are the same as for any other space translation invariant system. The linear combinations of the complex exponential modes of the infinite system that satisfy (7.6) are

$$A_n(x) = \sin\frac{n\pi x}{\ell}\,, \tag{7.7}$$

with angular wave number

$$k_n = \frac{n\pi}{\ell} \tag{7.8}$$

and frequency (from the dispersion relation, (7.4))

$$\omega_n = \sqrt{\frac{K\ell}{\rho_L}}\, k_n = \sqrt{\frac{K\ell}{\rho_L}}\, \frac{n\pi}{\ell}\,. \tag{7.9}$$

However, because the oscillations are longitudinal, the modes **look** very different from the transverse modes of the string that we studied in the previous chapter. The position of the point on the string whose equilibrium position is x, in the nth normal mode, has the general form (from (7.5))

$$x + \epsilon \sin \frac{n\pi x}{\ell} \cos(\omega_n t + \phi) \tag{7.10}$$

where ϵ and ϕ are the amplitude and phase of the oscillation.

The lowest 9 modes in (7.10) are animated in program 7-1. Compare these with the modes animated in program 6-1. The mathematics is the same, but the physics is very different because of (7.5). Stare at these two animations until you can visualize the relation between the two. Then you will have understood (7.5).

7.1.2 Free Ends

7-2

Now let us look at the situation in which the end of the spring at $x = 0$ is fixed, but the end at $x = \ell$ is free. The boundary conditions in this case are analogous to the normal modes of the string with one fixed end. The displacement at $x = 0$ must vanish because the end is fixed. Also, the derivative of the displacement at $x = \ell$ must vanish. You can see this by looking at the continuous spring as the limit of discrete masses coupled by springs. As we saw in (5.43), the last real mass must have the same displacement as the first "imaginary" mass,

$$\psi(\ell, t) = \psi(\ell + a, t)\,. \tag{7.11}$$

Therefore, for the finite system with a free end at ℓ, we have the relation

$$\frac{\psi(\ell, t) - \psi(\ell + a, t)}{a} = 0 \text{ for all } a. \tag{7.12}$$

In the limit that the distance between masses goes to zero, this becomes the condition that the derivative of the displacement, ψ, with respect to x vanishes at $x = \ell$,

$$\left.\frac{\partial}{\partial x}\, \psi(x, t)\right|_{x=\ell} = 0\,. \tag{7.13}$$

Thus the boundary conditions on the displacement are the same as in (6.11) for the transverse oscillation of a continuous string with $x = 0$ fixed and $x = \ell$ free,

$$\psi(0,t) = 0\,, \quad \frac{\partial}{\partial x}\,\psi(x,t)|_{x=\ell} = 0\,. \tag{7.14}$$

This, in turn, implies that the normal modes are the same as for the transversely oscillating string, (6.15),

$$A_n(x) = \sin\left(\frac{(2n+1)\pi x}{2\ell}\right) \quad \text{for } n = 0 \text{ to } \infty\,. \tag{7.15}$$

However, again because of (7.5), these modes look very different from those of the string. The first nine are animated in program 7-2 (compare with program 6-2).

7.2 A Mass on a Light Spring

Let us return to the system that we studied at the very beginning of the book, the harmonic oscillator constructed by putting a mass at the end of a light spring. We are now in a position to understand precisely what "light" means for this system, because we can now allow the spring to have a nonzero linear mass density, ρ_L, and find the normal modes of this system. We will then be able to see what happens as $\rho_L \to 0$.

To be specific, consider a spring with equilibrium length ℓ and spring constant K, fixed at $x = 0$ and constrained to oscillate only in the x direction (that is longitudinally). Now attach a mass, m, to the free end (with equilibrium position $x = \ell$). The spring, for $0 < x < \ell$, can be regarded as part of a space translation invariant system. To find the normal modes for this system, we look for linear combination of the modes of the infinite spring (for a given ω) that reproduces the physics at $x = 0$ and $x = \ell$. The fixed end at $x = 0$ is easy. This fixes the form of the modes to be proportional to

$$\sin k_n x \tag{7.16}$$

with frequency

$$\omega_n = \sqrt{\frac{K\ell}{\rho_L}}\,k_n\,. \tag{7.17}$$

As always, k_n and ω_n are related by the dispersion relation, (7.4). Now to determine the possible values of k_n, we require that $F = ma$ be satisfied for the mass. Suppose, for example, that the amplitude of the oscillation is A (a length). Then the displacement of the point on the spring with equilibrium position x is

$$\psi(x,t) = A \sin k_n x \cos \omega_n t\,, \tag{7.18}$$

and the displacement of the mass is determined by the displacement of the end of the spring,

$$x(t) \equiv \psi(\ell, t) = A \sin k_n \ell \cos \omega_n t \,. \tag{7.19}$$

The acceleration is

$$a(t) = \frac{\partial^2}{\partial t^2} \psi(\ell, t) = -\omega_n^2 A \sin k_n \ell \cos \omega_n t \tag{7.20}$$

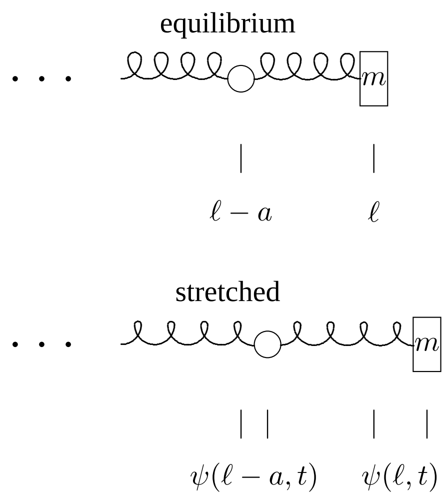


Figure 7.2: The stretching of the last spring is $\psi(\ell, t) - \psi(\ell - a, t)$.

To find the force on the mass, consider the massive spring as the continuum limit as $a \to 0$ of masses connected by massless springs of equilibrium length a, as at the beginning of the chapter. Then the force on the mass at the end is determined by the stretching of the last spring in the series. This, in turn, is the difference between the displacement of the system at $x = \ell$ and $x = \ell - a$, as illustrated in figure 7.2. Thus the force is

$$F = -K_a \left[\psi(\ell, t) - \psi(\ell - a, t) \right] \,. \tag{7.21}$$

In order to take the limit, $a \to 0$, rewrite this as

$$F = -K_a a \, \frac{\psi(\ell, t) - \psi(\ell - a, t)}{a} \,. \tag{7.22}$$

Now in the continuum limit, $K_a a$ is $K\ell$, and the last factor goes to a derivative, $\frac{\partial}{\partial x} \, \psi(x, t)|_{x=\ell}$. The final result for the force is therefore[2]

$$F = -K\ell \frac{\partial}{\partial x} \, \psi(x, t)|_{x=\ell} = -K\ell \, k_n \, A \cos k_n \ell \cos \omega_n t \,. \tag{7.23}$$

[2] Note that we can use this to give an alternate derivation of the boundary condition for a free end, (7.14).

Note that the units work. $K\ell$ is a force. $\frac{\partial}{\partial x}\psi$ is dimensionless.

Putting (7.20) and (7.23) into $F = ma$ and canceling a factor of $-A\cos\omega_n t$ on both sides gives,

$$K\ell k_n \cos k_n\ell = m\omega_n^2 \sin k_n\ell \,. \tag{7.24}$$

Using the dispersion relation to eliminate ω_n^2, we obtain

$$k_n\ell \tan k_n\ell = \frac{\rho_L \ell}{m} \,. \tag{7.25}$$

We have multiplied both sides of (7.25) by ℓ in order to deal with the dimensionless variables $k_n\ell$ (which is 2π times the number of wavelengths that fit onto the spring) and the dimensionless number

$$\epsilon \equiv \frac{\rho_L \ell}{m} \tag{7.26}$$

(which is the ratio of the mass of the spring, $\rho_L\ell$, to the mass, m). The spring is light if ϵ is much smaller than one.

The important point is that (7.25) has only one solution for $k_n\ell$ that goes to zero as $\epsilon \to 0$. Because $\tan k\ell \approx k\ell$ for small $k\ell$, it is

$$k_0\ell \approx \sqrt{\epsilon} \,. \tag{7.27}$$

For all the other solutions, the smallness of the left-hand side of (7.25) must come because $\tan k_n\ell$ is very small,

$$k_n\ell \approx n\pi \quad \text{for } n = 1 \text{ to } \infty \,. \tag{7.28}$$

But (7.28) implies

$$x(t) \equiv \psi(\ell, t) = A \sin k_n\ell \cos\omega_n t \approx 0 \quad \text{for } n = 1 \text{ to } \infty \,. \tag{7.29}$$

In other words, in all the solutions except k_0, the mass is hardly moving at all, and the spring is doing almost all the oscillating, looking very much like a system with two fixed ends. Furthermore, the frequencies of all the modes except the k_0 mode are large,

$$\omega_n \approx n\pi \sqrt{\frac{K}{\rho_L \ell}} \quad \text{for } n = 1 \text{ to } \infty \,, \tag{7.30}$$

while the frequency of the k_0 mode is

$$\omega_0 \approx \sqrt{\frac{K}{m}} \,. \tag{7.31}$$

For small ϵ (large mass), the k_0 mode is associated primarily with the oscillation of the mass, and has about the frequency we found for the case of the massless spring. The other modes are in an entirely different range of frequencies. They are associated with the oscillations of the spring. This is an important example of the way in which a single system can behave in very different ways in different regimes of frequency.

7.3 The Speed of Sound

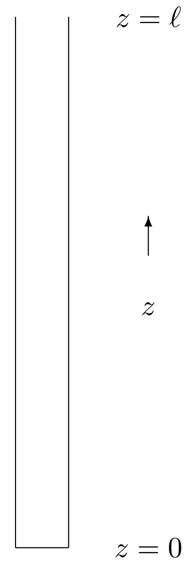


Figure 7.3: An organ pipe.

The physics of sound waves is obviously a three-dimensional problem. However, we can learn a lot about sound by considering motion of air in only one-dimension. Consider, for example, standing waves in the air in a long narrow tube like an organ pipe, shown in cartoon form in figure 7.3. Here, we will ignore the motion of the air perpendicular to the length of the pipe, and consider only the one-dimensional motion along the pipe. As we will see later, when we can deal with three-dimensional problems, this is a sensible thing to do for low frequencies, at which the transverse modes of oscillation cannot be excited. If we consider only one-dimensional motion, we can draw an analogy between the oscillations of the air in the pipe and the longitudinal waves in a massive spring.

It is clear what the analog of ρ_L is. The linear mass density of the air in the tube is

$$\rho_L = \rho A \tag{7.32}$$

where A is the cross-sectional area of the tube. The question then is what is $K\ell$ for a tube of air?

Consider putting a piston at the top of the tube, as shown in figure 7.4. With the piston at the top of the tube, there is no force on the piston, because the pressure of the air in the tube is the same as the pressure of the air in the room outside. However, if the piston is moved in a distance dz, as shown figure 7.5, the volume of the air in the tube is decreased by

$$-\,dV = A\,dz\,. \tag{7.33}$$

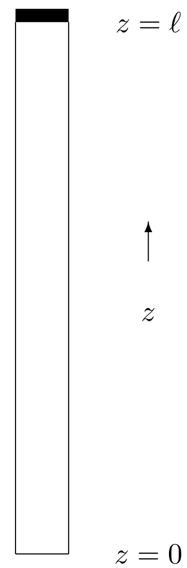


Figure 7.4: The organ pipe with a piston at the top. The air in the tube acts like a spring.

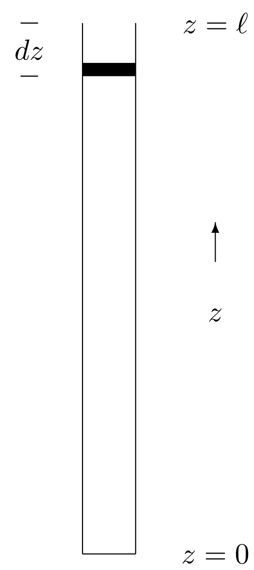


Figure 7.5: Pushing in the piston changes the volume of the air in the tube.

If the piston were moved in slowly enough for the temperature of the gas to stay constant, then the pressure would simply be inversely proportional to the volume. However, in a sound wave, the motion of the air is so rapid that almost no heat has a chance to flow in or out of the system. Such a change in the volume is called “adiabatic.” When the volume is decreased adiabatically, the temperature goes up (because the force on the piston is doing work) and the pressure increases faster than $1/V$, like

$$p \propto V^{-\gamma} \tag{7.34}$$

where γ is a positive constant that depends on the thermodynamic properties of the gas. More precisely, γ is the ratio of the specific heat at constant pressure to the specific heat at constant volume:[3]

$$C_P/C_V \tag{7.35}$$

In air, at standard temperature and pressure

$$\gamma_{\text{air}} \approx 1.40 \tag{7.36}$$

Now we can write from (7.34),

$$\frac{dp}{p} = -\gamma\,\frac{dV}{V} \tag{7.37}$$

or

$$dp = -\gamma p \frac{dV}{V} \approx \frac{\gamma A\,p_0}{V}\,dz = \frac{\gamma p_0}{\ell}\,dz \tag{7.38}$$

where p_0 is the equilibrium (room) pressure. Then the force on the piston is

$$dF = A\,dp = \frac{\gamma\,A^2\,p_0}{V}\,dz = \frac{\gamma\,A\,p_0}{\ell}\,dz \tag{7.39}$$

so that

$$K = \frac{dF}{dz} = \frac{\gamma\,A\,p_0}{\ell} \tag{7.40}$$

and $K\ell$ is

$$K\ell = \gamma\,A\,p_0\,. \tag{7.41}$$

Thus we expect the dispersion relation to be

$$\omega^2 = v_{\text{sound}}^2 k^2 = \frac{K\ell}{\rho_L}\,k^2 = \frac{\gamma\,p_0}{\rho}\,k^2 \tag{7.42}$$

where we have defined the “speed of sound”, v_{sound}, as

$$v_{\text{sound}}^2 = \frac{\gamma\,p_0}{\rho} \tag{7.43}$$

[3]See, for example, Halliday and Resnick.

For air at standard temperature and pressure,

$$v_{\text{sound}} \approx 332\,\frac{\text{m}}{\text{s}}\,. \tag{7.44}$$

As we will see in the next chapter, this is actually the speed at which sound waves travel. For now, it is just a parameter in our calculation of the normal modes.

In the pipe shown in (7.3), the displacement of the air, which we will call $\psi(z,t)$, must vanish at $z = 0$, because the bottom of the tube is closed and there is nowhere for the gas to go.

The z derivative of ψ must vanish at $z = \ell$, because the excess pressure is proportional to $-\frac{\partial}{\partial z}\psi$. The pressure is proportional to the force in our analogy with longitudinal waves in the massive spring. Using (7.41) and (7.23), we expect the longitudinal force to be

$$\pm\,\gamma A\,p_0\frac{\partial}{\partial z}\psi \tag{7.45}$$

or the excess pressure to be

$$p - p_0 = -\gamma\,p_0\frac{\partial}{\partial z}\psi\,. \tag{7.46}$$

We want the negative sign because for $\frac{\partial}{\partial z}\psi > 0$, the air is spreading out and has lower pressure.

Thus for a standing wave in the pipe, (7.3), we expect the boundary conditions

$$\psi(0,t) = 0\,, \quad \frac{\partial}{\partial z}\,\psi(z,t)\big|_{z=\ell} = 0\,, \tag{7.47}$$

for which the solution is

$$\psi(z,t) = \sin kz\cos\omega t \tag{7.48}$$

$$k = \frac{(n+1/2)\pi}{\ell}\,, \quad \omega = vk\,, \tag{7.49}$$

where $v = v_{\text{sound}}$, for nonnegative integer n. In particular, the lowest frequency mode of the tube corresponds to $n = 0$,

$$\omega = \frac{v\pi}{2\ell}\,, \quad \nu = \frac{\omega}{2\pi} = \frac{v}{4\ell}\,. \tag{7.50}$$

7.3.1 The Helmholtz Approximation

Let's consider a slightly different problem. What is the lowest frequency mode of a one-liter soda bottle, shown in figure 7.6? A typical set of parameters is given below:

$$\begin{array}{rcl} A \approx 2.85\,\text{cm}^2 & : & \text{area of neck} \\ \ell \approx 5.7\,\text{cm} & : & \text{length of neck} \\ L \approx 25\,\text{cm} & : & \text{length of bottle} \\ V_0 \approx 1000\,\text{cm} & : & \text{volume of body} \end{array} \tag{7.51}$$

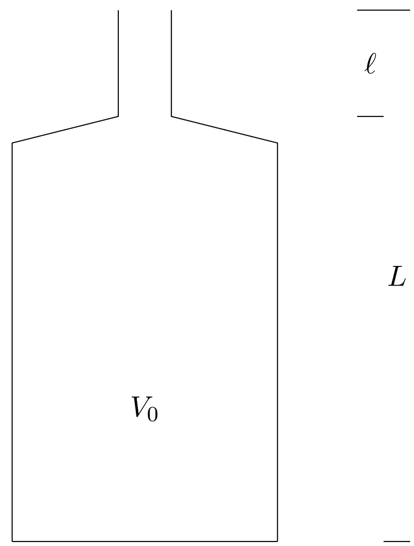


Figure 7.6: A one liter soda bottle.

Putting the length, L, of the bottle into (7.50) gives $\nu \approx 332\,\mathrm{hertz}$. In American standard pitch (see table 7.1), this is an E above middle C.

This is obviously wrong. If you have ever blown into your soda bottle, you know that the frequency of the lowest mode is much lower than that. The problem, of course, is that the soda bottle is not shaped anything like the tube. To determine the modes is a complicated three-dimensional problem. It turns out, however, that we can find the lowest mode to a decent approximation rather easily.

The idea is that in the lowest mode, the air in the neck of the bottle is moving rapidly, but in the body of the bottle, the air quickly spreads out so that it is not moving much at all. The idea of the Helmholtz approximation to try is to treat the air in the neck as a single chunk with mass

$$\rho A\ell\,, \tag{7.52}$$

and to treat the body as a spring, that contributes restoring force but no inertia (because the air is not moving much). Then all we must do is to compute the K of the "spring." That is easy, using (7.38). In this case,

$$dV = A\,dz\,, \tag{7.53}$$

so

$$dp = -\gamma p\,\frac{A\,dz}{V} \approx -\gamma p_0\,\frac{A\,dz}{V_0} \tag{7.54}$$

Table 7.1: American standard pitch (A440) — frequencies are in Hertz.

Equal-temperament Chromatic Scale					
note	ν	note	ν	note	ν
A	880	A	440	A	220
$G\sharp$	831	$G\sharp$	415	$G\sharp$	208
G	784	G	392	G	196
$F\sharp$	740	$F\sharp$	370	$F\sharp$	185
F	698	F	349	F	175
E	659	E	330	E	165
$E\flat$	622	$E\flat$	311	$E\flat$	156
D	587	D	294	D	147
$C\sharp$	554	$C\sharp$	277	$C\sharp$	139
C	523	C	262	C	131
B	494	B	247	B	123
$B\flat$	466	$B\flat$	233	$B\flat$	117

and

$$F \approx -\gamma p_0 \, \frac{A^2 \, dz}{V_0} \tag{7.55}$$

or

$$\text{“}K\text{”} = \gamma p_0 \, \frac{A^2}{V_0} \, . \tag{7.56}$$

Then using $\omega^2 = K/m$, we expect

$$\omega = \sqrt{\frac{\gamma A^2 \, p_0/V_0}{\rho A \ell}} = v \sqrt{\frac{A}{\ell V_0}} \, . \tag{7.57}$$

For the soda bottle, (7.6), this gives

$$\nu \approx 118 \text{hertz} \tag{7.58}$$

or roughly a $B\flat$ below low C. This is just about right (see problem 7.5).

7.3.2 Corrections to Helmholtz

There are many possible corrections to (7.57) that might be considered. One is to include the so-called “end effect.” The point is that the velocity of the air in the lowest mode does not drop to zero immediately when you go past the ends of the neck. Thus the actual mass is somewhat larger than $\rho A \ell$. The lore is that you can do better by replacing

$$\ell \to \ell + 0.6 \, r \tag{7.59}$$

where r is the radius of the neck.

Here we will discuss another correction that can be dealt with systematically using the methods of space translation invariance and local interactions. If the bottle has a long neck, it is probably not a good idea to treat the air in the neck as a solid mass. Furthermore, there is a simple alternative. A better analogy for the neck is a massive spring with $K\ell = \gamma A\, p_0$. Because the neck is a space translation invariant, essentially one-dimensional system, we expect a displacement of the form

$$y\, \cos \frac{\omega z}{v} \tag{7.60}$$

in the neck, where $z = 0$ is the open end and y is the displacement of the air at $z = 0$. Thus, where the neck attaches to the body, the displacement is

$$y\, \cos \frac{\omega \ell}{v}\,. \tag{7.61}$$

The force at this point from the compression of the air in the neck is (from (7.45))

$$F_{\text{neck}} = -\gamma A\, p_0 \frac{\partial \psi}{\partial z} = \frac{\gamma A\, p_0 \omega}{v}\, y \sin \frac{\omega \ell}{v}\,. \tag{7.62}$$

This must be the negative of the force from the air in the body, from (7.39),

$$-\, F_{\text{body}} = \frac{\gamma A^2 p_0}{V_0}\, y \cos \omega \ell / v\,, \tag{7.63}$$

or

$$\frac{\omega V_0}{A v}\, \tan \frac{\omega \ell}{v} = 1\,. \tag{7.64}$$

You will explore the consequences of this in problem 7.5.

This analysis does not distinguish between the area of the top and bottom of the neck. Perhaps the area at the bottom is more appropriate. What matters is the area at the bottom that determines the force per unit area where the wave in the neck matches onto the body.

Chapter Checklist

You should now be able to:

i. Find the motion of a point on a continuous spring oscillating longitudinally in one of its normal modes for various boundary conditions;

ii. Solve for the normal modes of a system of a mass attached to a massive spring;

iii. Be able to derive the dispersion relation for sound waves and find the normal modes for oscillations of air in a tube;

iv. Be able to use the Helmholtz approximation to estimate the frequency of the lowest mode of bottle.

Problems

7.1. Derive (7.45) directly by considering the volume of the chunk of air in the tube between z and $z + dz$, and using (7.38).

7.2. Use an analogy with (7.16)-(7.31) to find (approximately!) the normal modes and corresponding frequencies of the system shown in figure 6.1, but with a massive ring of mass m sliding on the frictionless rod.

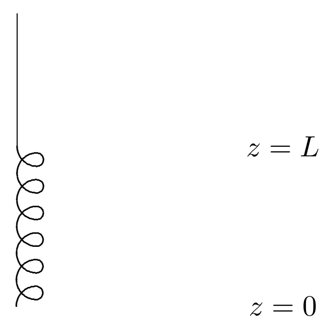


Figure 7.7: A hanging spring.

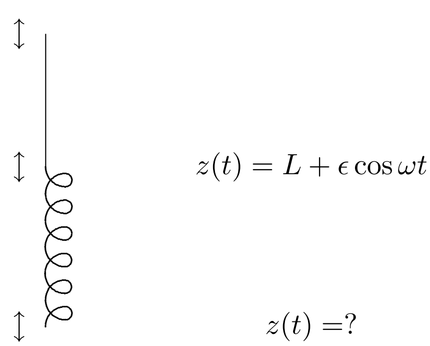


Figure 7.8: Problem 7.3.

7.3. A massive continuous spring with mass m, length L and spring constant K hanging vertically. The system is shown **at rest in its equilibrium configuration** in figure 7.7. The spring constant is large, satisfying $KL \gg mg$, so gravity plays no important role here except to keep the spring vertical. Now suppose that the supporting hanger is driven up and down so that the top of the spring moves vertically with displacement $\epsilon \cos \omega t$, as shown in figure 7.8. Find the z position of the bottom of the spring as a function of time. Ignore damping.

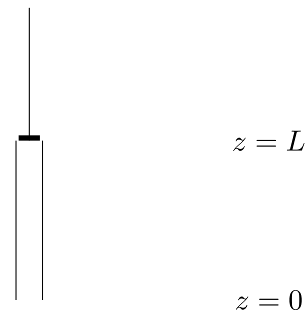


Figure 7.9: Problem 7.4.

7.4. A system analogous to that in problem 7.3 is a tube of air with a piston at the top and the bottom open, as shown in figure 7.9: If the cross sectional area of the tube is A, what is the analog in this system of the spring constant, K, in problem 7.3? Make sure that your answer has units of force per unit distance.

7.5. **PERSONAL EXPERIMENT —** Show that when $\omega\ell/v$ is small, (7.64) reduces to the Helmholtz approximation, (7.57), while for $V_0 \approx 0$, when the bottle is all neck, it reduces to the result for the modes of a uniform tube with one open and one closed end, (7.50).

Do the experiment! Find a selection of at least four bottles, at least one of which has a very long neck. Measure the frequency of the lowest mode of each, and describe how you did it. For each bottle, tabulate the following (in cgs units):

i. A description (ie. soda bottle, 1000 ml)
ii. A_t (the area of the top of the neck)
iii. A_b (the area of the bottom of the neck)
iv. r (the radius of the neck)
v. ℓ (the length of the neck)
vi. V_{body} (the volume of the body)
vii. ν (the frequency of the lowest mode)
viii. ω (the angular frequency of the lowest mode)
ix. $\omega^2 V_0 \ell / a v^2$ (=1 in the Helmholtz approximation)
x. $(\omega V_0 / A v)\ \tan(\omega \ell / v)$ (=1 in the approximation (7.64))

See whether you can see the end effect, (7.59), or distinguish the area of the top of the neck from the bottom — that is, see which works better in (7.57). Comment, as quantitatively as you can, on the errors in your experiment, and on the relative merits of the approximate expressions that you have tested.

Chapter 8

Traveling Waves

In this chapter, we show how the same physics that leads to standing wave oscillations also gives rise to waves that move in space as well as time. We then go on to introduce the important physical example of light waves.

Preview

In an infinite translation invariant system, traveling waves arise naturally from the complex exponential behavior of the solutions in space and time.

i. We begin by showing the connection between standing waves and traveling waves in infinite systems. A traveling wave in a linear system is a pair of standing waves put together with a special phase relation. We show how traveling waves can be produced in finite systems by appropriate forced oscillations.

ii. We then go on to discuss the force and power required to produce a traveling wave on a string, and introduce the useful idea of "impedance."

iii. We introduce and discuss the most important classical example of wave phenomena, electromagnetic waves and light.

iv. We reexamine the translation invariant systems of coupled LC circuits discussed in chapter 5 and show how they are related to electromagnetic waves.

v. We discuss the effects of damping in translation invariant systems, giving a simple physical interpretation of the effect of traveling waves.

vi. We discuss traveling waves in systems with damping and in systems with high and/or low frequency cut-offs.

8.1 Standing and Traveling Waves

8.1.1 What is It That is Moving?

8-1

We have seen that an infinite system with translation invariance has complex solutions of the form

$$e^{\pm ikx}e^{\pm i\omega t}, \tag{8.1}$$

where k and ω are related by the dispersion relation characteristic of the system. So far, we have considered standing wave solutions in which the space and time dependent factors are separately real, i.e.

$$\sin kx \cdot \cos \omega t \propto (e^{ikx} - e^{-ikx}) \cdot (e^{i\omega t} + e^{-i\omega t}) . \tag{8.2}$$

But we can put the same solutions together in a different way,

$$\psi(x,t) = \cos(kx - \omega t) \propto (e^{ikx}e^{-i\omega t} + e^{-ikx}e^{i\omega t}) . \tag{8.3}$$

This is called a **"traveling wave."** The underlying system that supports the wave is not actually traveling. Instead, what is moving is the wave itself. If we follow the point x for which $\psi(x,t)$ has some constant value, the point moves in the positive x direction at a constant velocity, called the **"phase velocity,"**

$$v_\phi = \omega(k)/k . \tag{8.4}$$

In (8.3), for example, $\psi(x,t)$ is equal to one for $x = t = 0$, because the argument of the cosine is zero (it is also equal to one for $x = 2n\pi/k$ for any integer n, but we will focus on just the single point, $x = 0$). As t increases, this point moves in the positive x direction because the argument of the cosine, $kx - \omega t$, vanishes for $x = \omega t/k = v_\phi t$. This is illustrated in program 8-1.

We will continue to define all the real modes to be real parts of complex modes proportional to $e^{-i\omega t}$. Thus (8.3) is

$$\cos(kx - \omega t) = \mathrm{Re}\left[e^{ikx}e^{-i\omega t}\right] . \tag{8.5}$$

In this notation a wave traveling to the left is

$$\cos(kx + \omega t) = \mathrm{Re}\left[e^{-ikx}e^{-i\omega t}\right] , \tag{8.6}$$

while a standing wave is

$$\begin{aligned} \cos kx \cos \omega t &= \frac{1}{2}\,\mathrm{Re}\left[e^{ikx}e^{-i\omega t} + e^{-ikx}e^{-i\omega t}\right] \\ &= \tfrac{1}{2}\left[\cos(kx - \omega t) + \cos(kx + \omega t)\right] . \end{aligned} \tag{8.7}$$

A standing wave is a combination of traveling waves going in opposite directions! Likewise, a traveling wave is a combination of standing waves. For example,

$$\cos(kx - \omega t) = \cos kx \cos \omega t + \sin kx \sin \omega t\,. \tag{8.8}$$

These relations are important because they show that **the relation between k and ω, the dispersion relation, is just the same for traveling waves as for standing waves!** A wave is a wave, whether traveling or standing. Indeed, we can go back and forth using (8.7) and (8.8). **The dispersion relation that relates k and ω is a property of the system in which the waves exist, not of the particular wave.**

The other side of this coin is that traveling waves exist for systems with any dispersion relation. Knowing the phase velocity, (8.4), for all k is equivalent to knowing the dispersion relation, because you must know $\omega(k)$. In particular, it is only for simple, continuous systems like the stretched string (see (6.5)) that $\omega(k)$ is proportional to k and the phase velocity is a constant, independent of k.

8.1.2 Boundary Conditions

8-2

Traveling waves can be produced in finite systems by forced oscillation with an appropriate phase for the oscillations at the two ends. A simple example involves a stretched string with tension T and linear mass density ρ. Given boundary conditions on the system so that

$$\psi(0,t) = A\cos\omega t\,, \quad \psi(L,t) = A\sin\omega t\,, \tag{8.9}$$

where L is the length of the string, the angular frequency ω is chosen so that

$$k = \frac{5\pi}{2L} = \omega\sqrt{\frac{\rho}{T}} = \frac{\omega}{v_\phi}\,. \tag{8.10}$$

As usual in a forced oscillation problem, we are interested in the steady state solution in which the system moves with the angular frequency, ω, of the forcing terms. We can solve this problem easily by breaking it up into two problems.

First consider the boundary condition:

$$\psi_1(0,t) = 0\,, \quad \psi_1(L,t) = A\sin\omega t\,. \tag{8.11}$$

This is easily solved by the methods of chapter 5. From the condition at $x = 0$, we know that the solution for $\psi_1(x,t)$ is proportional to $\sin kx$. Then the boundary condition at $x = L$ gives the standing wave solution:

$$\psi_1(x,t) = A\sin kx \sin\omega t\,. \tag{8.12}$$

Next consider the boundary condition

$$\psi_2(0,t) = A\cos\omega t\,, \quad \psi_2(L,t) = 0\,. \tag{8.13}$$

Analogous arguments (starting at $x = L$) show that the solution is the standing wave

$$\psi_2(x,t) = A\cos kx\cos\omega t\,. \tag{8.14}$$

Now we can obtain the solution for the boundary condition (8.9) simply by adding these:

$$\begin{aligned}\psi(x,t) &= \psi_1(x,t) + \psi_2(x,t) \\ &= A\cos kx\cos\omega t + A\sin kx\sin\omega t = A\cos(kx-\omega t)\,,\end{aligned} \tag{8.15}$$

which is a wave traveling from $x = 0$ to $x = L$. The crucial point is that the two standing waves out of which the traveling wave is built are $90°$ out of phase with one another both in time and in space. They get large at different points in space and also at different times and the interplay between the two produces the traveling wave. This is illustrated in figures 8.1-8.4 for $\omega t = 0$, $\pi/4$, $\pi/2$ and $3\pi/4$. In each of these figures, the top curve is the traveling wave. The middle curve is (8.14). The lower curve is (8.12).

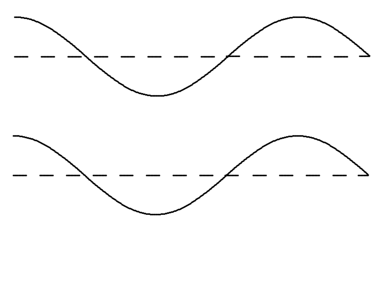


Figure 8.1: $t = 0$.

This system is animated in program 8-2. This animation is important. It is worth staring at it for a while to get a better feeling for how (8.15) works than you can from the still pictures in figures 8.1-8.4. If you concentrate on a particular point on the string, you will see that the traveling wave gets large either when one of the standing waves is a maximum with the other near zero, or (depending on where you looking) when both standing waves are positive.

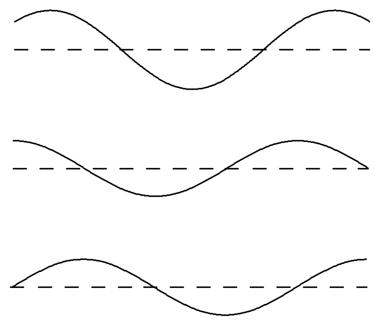


Figure 8.2: $t = \pi/4$.

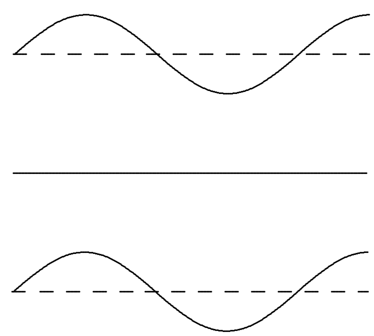


Figure 8.3: $t = \pi/2$.

8.2 Force, Power and Impedance

Whatever is enforcing the boundary conditions in the example of (8.9) must exert a force on the string. Of course, a horizontal force is required to keep the string stretched, but for small oscillations, this force is nearly constant and approximately equal to the string tension, T. Furthermore, there is no motion in the x direction so no work is done by this component of the force. The vertical component of the force is the negative of the force which the tension on the string produces. At $x = 0$, this is

$$F_0 = -T \frac{\partial}{\partial x} \psi(x,t)|_{x=0} \,. \tag{8.16}$$

This is illustrated in figure 8.5.

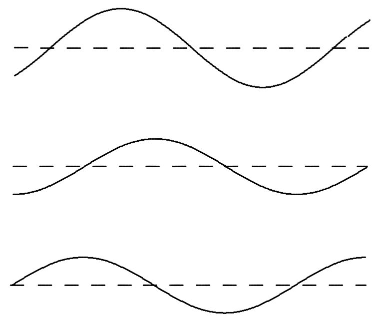


Figure 8.4: $t = 3\pi/4$.

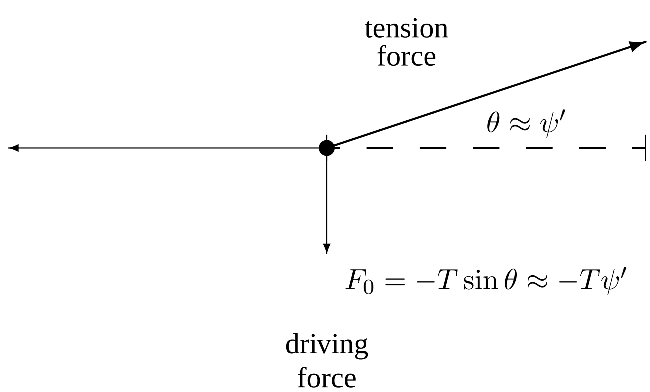


Figure 8.5: The force due to a string pulling in the $+x$ direction.

At $x = L$, because the string is coming in from the $-x$ direction, it is

$$F_L = T\frac{\partial}{\partial x}\psi(x,t)|_{x=L}\,, \tag{8.17}$$

as illustrated in the figure 8.6.

In the forced oscillation, the end of the string is moving only in the transverse direction. Thus the power supplied by the external force at $x = 0$, which is $\vec{F}\cdot\vec{v}$ is

$$P(t) = -T\frac{\partial}{\partial x}\,\psi(x,t)|_{x=0}\;\;\frac{\partial}{\partial t}\psi(0,t) \tag{8.18}$$

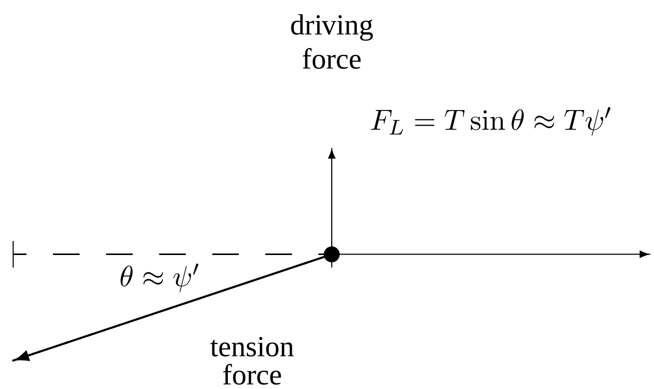


Figure 8.6: The force due to a string pulling in the $-x$ direction.

where as in (2.26), $\psi(x,t)$ is the **real** displacement from equilibrium for the piece of string at horizontal position x. We must take the real part first because the power is a **nonlinear** function of the displacement.

For a standing wave on the string (or any system with no frictional forces), the force and the velocity are 90° out of phase. For example, if the displacement is proportional to $\sin\omega t$, then the transverse force at each end is also proportional to $\sin\omega t$. The velocity, however, is proportional to $\cos\omega t$. Thus the power expended by the external force is

$$\propto \sin\omega t\cos\omega t = \frac{1}{2}\sin 2\omega t\,. \tag{8.19}$$

This averages to zero over a half-cycle. On the average, no power is required to keep the standing wave going (in the absence of damping).

In a traveling wave, on the other hand, the force and the velocity are proportional. From (8.15), you can see that

$$\frac{\partial}{\partial x}\psi(x,t) = -\frac{k}{\omega}\frac{\partial}{\partial t}\psi(x,t)\,. \tag{8.20}$$

Thus

$$F_0 = Z\frac{\partial}{\partial t}\psi(0,t)\,,\quad F_L = -Z\frac{\partial}{\partial t}\psi(L,t)\,, \tag{8.21}$$

where the constant Z,

$$Z = \frac{Tk}{\omega} = \sqrt{\rho T}\,, \tag{8.22}$$

is called the **“impedance”** of the string system. It measures the power required to produce the traveling wave. The power required at $x = 0$ is

$$P_0 = Z\,\left(\frac{\partial}{\partial t}\psi(0,t)\right)^2 = Z\,A^2\,\omega^2\,\sin^2\omega t\,. \tag{8.23}$$

The average power expended is thus

$$\langle P_0\rangle = Z\,A^2\omega^2/2\,. \tag{8.24}$$

The power expended at $x = 0$ to produce the traveling wave is given up by the string at $x = L$, because the power required at L is

$$P_L = -Z\,\left(\frac{\partial}{\partial t}\psi(L,t)\right)^2 = -Z\,A^2\omega^2\cos^2\omega t\,. \tag{8.25}$$

If the boundary conditions were such that the traveling waves were going in the opposite direction, the force in the above derivations would have the opposite sign from (8.20). Thus the positive power is always required to produce the wave and the negative power is required to absorb it. It may seem odd that the power fed into the wave in (8.23) and the power given up by the wave in (8.25) are not exactly equal and opposite. The sum vanishes on the average, but oscillates with time. The reason is that the length of the system is not an integral number of wavelengths. This allows the energy stored on the system, the sum of kinetic and potential, to oscillate as a function of time.

Note that the force required to absorb a traveling wave, in (8.21), is negative and proportional to the velocity. This is a typical frictional force. Thus a traveling wave can be absorbed completely by a frictional force (or a resistance) with exactly the right ratio of force to velocity. If the impedance of the **“dashpot”** (as such a resistance is called) is not exactly the same as that of the string, there will be some reflection. We will come back to this in the next chapter.

8.2.1 * Complex Impedance

For the stretched string, a system for which the dispersion relation is equivalent to the wave equation, (6.4), the force on the system and the displacement velocity, $\frac{\partial}{\partial t}\psi$, are proportional for any traveling wave.[1] In general, this is not true. For example, consider the beaded string of figure 5.4 stretched from $x = 0$ to some large x. Suppose further that there is a traveling wave in the system of the form,

$$\psi(x,t) = A\cos(kx - \omega t)\,, \tag{8.26}$$

illustrated in figure 8.7.[2] The dotted line is the equilibrium position of the string.

[1] We will see this in detail in chapter 10.

[2] For an animation of a traveling wave in a similar system, see program 8-6. The system shown in this program has the beads on springs, as well as on a string. However, the form of the traveling wave is the same. Only the

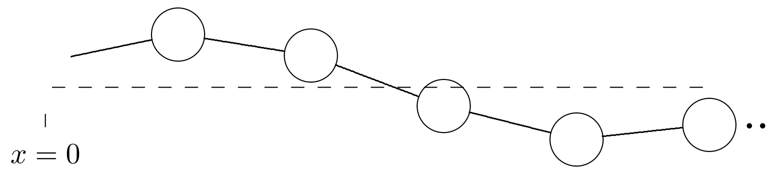


Figure 8.7: A snapshot of a traveling wave in a beaded string.

So long as k and ω are related by the dispersion relation, (5.39), then (8.26) is a solution to the equation of motion. The external transverse force at $x = 0$ required to produce the traveling wave is related to the difference between the displacement of the first block and the displacement of the end at $x = 0$ (see figure 5.5). It is

$$F_0 = \frac{TA}{a} \left(\cos(\omega t - ka) - \cos \omega t\right) . \tag{8.27}$$

This is approximately proportional to the velocity **only** if ka is very small, so that the right-hand side of (8.27) can be expanded in a Taylor series. Thus in this case, and in general for a discrete system, we cannot define the impedance simply as in (8.21).

However, suppose that instead of the real traveling waves, (8.26), we consider a complex harmonic traveling wave with **irreducible** time and space of the form

$$\psi(x,t) = A\, e^{-i(\omega t - kx)} . \tag{8.28}$$

Then because of the irreducible on t and x (that comes from translation invariance), we know immediately that the both the force and the t derivative of ψ are proportional to ψ. For an irreducible solution, everything is proportional to $e^{-i(\omega t - kx)}$. Thus they are also proportional to each other, and we can define the impedance,

$$F = -Z(k) \frac{\partial}{\partial t} \psi(x,t) = i\omega A\, Z(k)\, e^{-i(\omega t - kx)} . \tag{8.29}$$

For example, for the beaded string, if we replace the real solution, (8.26), with the irreducible complex solution, (8.28), the force becomes

$$F_0 = \frac{TA}{a} \left(e^{-i(\omega t - ka)} - e^{-i\omega t}\right) = \frac{TA}{a} \left(e^{ika} - 1\right) e^{-i\omega t} . \tag{8.30}$$

Thus from (8.29), the impedance, $Z(k)$, is

$$Z(k) = \frac{T}{\omega a} \frac{e^{ika} - 1}{i} = \frac{2T}{a}\, e^{ika/2}\, \frac{\sin \frac{ka}{2}}{\omega} . \tag{8.31}$$

dispersion relation is different.

Using the dispersion relation, (5.39), we can write this as

$$Z(k) = e^{ika/2}\sqrt{\frac{mT}{a}}\,. \tag{8.32}$$

The impedance, $Z(k)$, defined by (8.29) is, in general, complex, and k dependent. Nevertheless, we can find the average power required to produce the wave. Because the power is a **nonlinear** function of the displacement, we must first take the real parts of the complex velocity and complex force before computing the power, as in (2.26). For arbitrary complex $A = |A|e^{i\phi}$,

$$\begin{aligned} v &= \omega|A|\sin(\omega t - kx - \phi)\,, \\ F &= (\mathrm{Im}\,Z(k))\,\omega|A|\cos(\omega t - kx - \phi) + (\mathrm{Re}\,Z(k))\,\omega|A|\sin(\omega t - kx - \phi)\,, \end{aligned} \tag{8.33}$$

where we have put the phase of A into the $\cos$ and $\sin$ functions (see (1.96)-(1.98)) to make it clear that only the absolute value of A matters for the average power. Then, as in (2.26), only the $\sin^2$ term contributes to the time-averaged power, which is

$$\frac{1}{2}(\mathrm{Re}\,Z)\,\omega^2|A|^2\,. \tag{8.34}$$

8.3 Light

Light waves, like the sound waves that we discussed in the previous chapter, are inherently three-dimensional things. However, as with sound, we can say a lot about light that is more or less independent of the three-dimensional details.

8.3.1 Plane Waves

There is a simple way of concentrating on only one dimension. That is to look for solutions in which the other two dimensions do not enter at all. Consider Maxwell's equations in free space, in terms of the vector fields, $\vec{E}$ and $\vec{B}$ describing the electric and magnetic fields.

$$\begin{aligned} \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t} \\ \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \end{aligned} \tag{8.35}$$

$$\begin{aligned}\frac{\partial B_y}{\partial x}-\frac{\partial B_x}{\partial y}&=\mu_0\epsilon_0\frac{\partial E_z}{\partial t}\\ \frac{\partial B_z}{\partial y}-\frac{\partial B_y}{\partial z}&=\mu_0\epsilon_0\frac{\partial E_x}{\partial t}\\ \frac{\partial B_x}{\partial z}-\frac{\partial B_z}{\partial x}&=\mu_0\epsilon_0\frac{\partial E_y}{\partial t}\end{aligned} \tag{8.36}$$

$$\begin{aligned}\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}&=0\\ \frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}&=0\end{aligned} \tag{8.37}$$

where ϵ_0 and μ_0 are two constants called the permittivity and permeability of empty space.[3] Let us look for solutions to these partial differential equations that involve only functions of z and t. In this case, things simplify to:

$$0=-\frac{\partial B_z}{\partial t}\,,\quad -\frac{\partial E_y}{\partial z}=-\frac{\partial B_x}{\partial t}\,,\quad \frac{\partial E_x}{\partial z}=-\frac{\partial B_y}{\partial t}\,, \tag{8.38}$$

$$0=\mu_0\epsilon_0\frac{\partial E_z}{\partial t}\,,\quad -\frac{\partial B_y}{\partial z}=\mu_0\epsilon_0\frac{\partial E_x}{\partial t}\,,\quad \frac{\partial B_x}{\partial z}=\mu_0\epsilon_0\frac{\partial E_y}{\partial t}\,, \tag{8.39}$$

$$\frac{\partial E_z}{\partial z}=0\,,\quad \frac{\partial B_z}{\partial z}=0\,. \tag{8.40}$$

These equations imply that E_z and B_z are independent of z and t. Since we have already assumed that they depend only on z and t, this means that they are constants. We will ignore them because we are interested in the solutions with nontrivial z and t dependence. That leaves the x and y components, satisfying (8.38) and (8.39).

Then, because (8.38) and (8.39) are invariant under translations in z and t, we expect complex exponential solutions, in which all components are proportional to

$$e^{i(\pm kz-\omega t)}\,, \tag{8.41}$$

$$E_x(z,t)=\varepsilon_x^\pm e^{i(\pm kz-\omega t)}\,,\quad E_y(z,t)=\varepsilon_y^\pm e^{i(\pm kz-\omega t)}\,, \tag{8.42}$$

$$B_x(z,t)=\beta_x^\pm e^{i(\pm kz-\omega t)}\,,\quad B_y(z,t)=\beta_y^\pm e^{i(\pm kz-\omega t)}\,, \tag{8.43}$$

Direct substitution of (8.42) and (8.43) into (8.38) and (8.39) gives

$$\mp k\varepsilon_y^\pm=\omega\beta_x^\pm\,,\quad \pm k\varepsilon_x^\pm=\omega\beta_y^\pm\,, \tag{8.44}$$

$$\mp k\beta_y^\pm=-\mu_0\epsilon_0\omega\varepsilon_x^\pm\,,\quad \pm k\beta_x^\pm=-\mu_0\epsilon_0\omega\varepsilon_y^\pm\,. \tag{8.45}$$

As usual, we have written the wave with the irreducible time dependence, $e^{-i\omega t}$. To get the real electric and magnetic fields, we take the real part of (8.42) and (8.43). This works

[3] See, for example, Purcell, chapter 9.

because Maxwell's equations are linear in the electric and magnetic fields. The amplitudes, $\varepsilon_x^\pm$, etc, can be complex.

From (8.44) and (8.45), you see that $\varepsilon_y^\pm$ is related to $\beta_x^\pm$ and $\varepsilon_x^\pm$ is related to $\beta_y^\pm$. For each relation, there are two homogeneous simultaneous linear equations in the two unknowns. They are consistent only if the ratio of the coefficients is the same, which implies a relation between k and ω,

$$k^2 = \mu_0 \epsilon_0 \omega^2 \,. \tag{8.46}$$

This is a dispersion relation,

$$\omega^2 = c^2 k^2 = \frac{1}{\mu_0 \epsilon_0} k^2 \,. \tag{8.47}$$

The phase velocity, c, is the speed of light in vacuum (we will have more to say about this in chapters 10 and 11!).

Once (8.47) is satisfied, we can solve for the $\beta^\pm$ in terms of the $\varepsilon^\pm$:

$$\beta_y^\pm = \pm \frac{1}{c} \varepsilon_x^\pm \,, \quad \beta_x^\pm = \mp \frac{1}{c} \varepsilon_y^\pm \,. \tag{8.48}$$

These solutions to Maxwell's equations in free space are electromagnetic waves, or light waves. These simple solutions, depending only on z and t are an example of plane wave solutions. The name is appropriate because the electric and magnetic fields in the wave have the same value everywhere on each plane of constant z, for any fixed time, t. These planes propagate in the $\pm z$ direction at the phase velocity, c.

In general, electromagnetic waves can propagate in any direction in three-dimensional space. However, the electric and magnetic fields that make up the wave are always perpendicular to the direction in which the wave is traveling and perpendicular to each other.

The treatment of plane wave electromagnetic waves traveling in the z direction is analogous to our treatment of sound in chapter 7. There, also, the wave depended only on z. However, the electromagnetic waves are a little more complicated because the wave phenomenon depends on **both** the electric and magnetic fields. The reason that we have postponed until now the discussion of electromagnetic waves, even though they are one of the most important examples of wave phenomena, is that the relations, (8.48), between the electric and magnetic fields depend on the direction in which the wave is traveling (the $\pm$ sign!). It is much easier to write down the solutions for the traveling waves than for the standing waves. Even for the simple traveling plane waves we have described that depend only on z and t, this relation between $\vec{E}$ and $\vec{B}$ and the direction of the wave depends on the three-dimensional properties of Maxwell's equations. We will discuss these issues in much more detail in chapters 11 and 12.

8.3.2 Interferometers

One of the wonderful features of light waves is that it is relatively easy to split them up and reassemble them. This feature is used in many optical devices, one of the simplest of

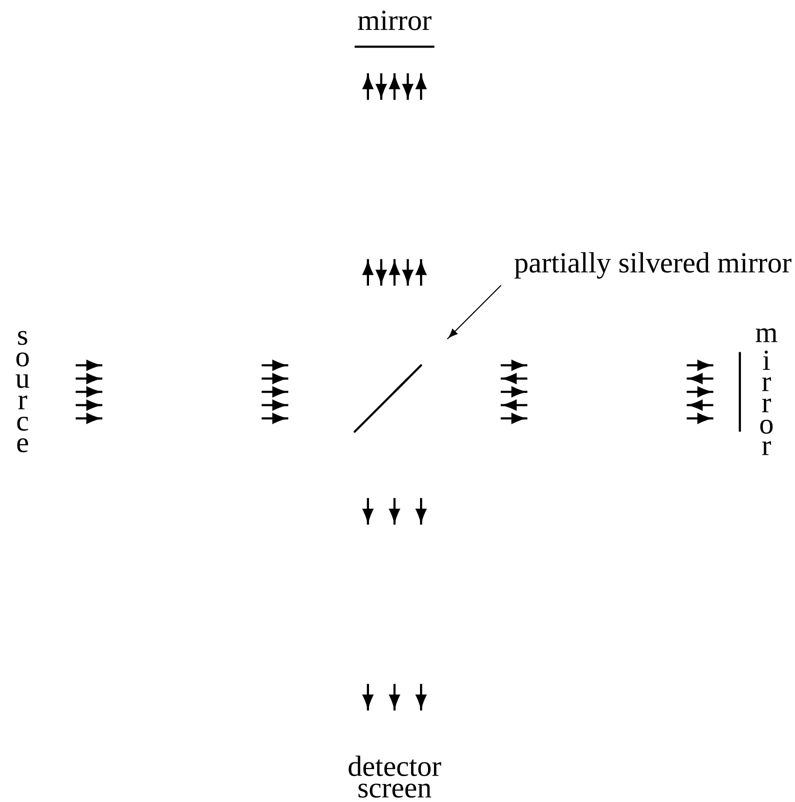


Figure 8.8: A schematic diagram of a Michelson interferometer.

which is an **“interferometer,”** one version of which (the Michelson interferometer) is shown in schematic form in figure 8.8. A source produces a plane wave (as we will discuss in chapter 13, it cannot be quite a plane wave, but never mind that for now). The partially silvered mirror serves as a “beam splitter” by allowing some of the light to pass through, while reflecting the rest. Then the mirrors at the top and the right reflect the beams back. Then the partially silvered mirror serves as a “beam reassembler,” combining the beams from the top and the right into a single beam that travels on to the detector screen where the beam intensity (proportional to the square of the electric field) is measured. The important thing is that the light wave reaching the detector screen is the sum of two components that are coherent and yet have traveled different paths. What “coherent” means in this context is not only that the frequency is the same, but that the phase of the waves is correlated. In this case, that happens simply because the two components reaching the screen arise from the same incoming plane wave.

Now the intensity of the light reaching the screen depends on the relative length of the two paths. Different path lengths will produce different phases. If the two components are in phase, the amplitudes will add and the screen will be bright. This is called “constructive interference”. If the two components are $180°$ out of phase, the amplitudes will subtract and the screen will be dark. There will be what is called “destructive interference.”

This sounds rather trivial, and indeed it is (at least for classical electromagnetic waves), but it is also extremely useful, because it provides a very sensitive measure of **changes** of the length of the paths. In particular, if one of the mirrors is moved a distance d (it might be part of an experimental setup designed to detect small motions, for example), the relative phase of the two components reaching the screen changes by $2kd$ where k is the angular wave number of the plane wave, because the path length of the reflected wave has changed by $2d$. Thus each time d changes by a quarter of the wavelength of the light, the screen goes from bright to dark, or vice versa.

This is a very useful way of measuring small distance changes. In practice, the incoming beam is not exactly a plane wave (that, as we will see in detail later, would require an infinite experiment!), so the intensity of the light is not uniform over the screen. Instead there are light areas and dark areas known as “fringes.” As the mirror is moved, the fringes move, and one can count the fringes that go past a given spot to keep track of the number of changes from bright to dark.

8.3.3 Quantum Interference

There is another wave of thinking about the interferometer that makes it seem much less trivial. As we will discuss several times in this book, and you will learn more about when you study quantum mechanics, light is not only a wave. It is **also** made up of individual particles of light called photons. You don’t notice this unless you turn the intensity of the light wave way down. But in fact, you can turn the intensity down so much that you can detect individual photons hitting the screen. Now it is not so clear what is happening. An individual photon cannot split into two parts at the beam splitter and beam reassembler. As we will see later, the energy of the photon is determined by the frequency of the light. It cannot be divided. You might think, therefore, that the individual photon would have to go one way or the other. But then how can one get an interference between the two paths? There is no answer to this question that makes “sense” in the classical physics of particles. Nevertheless, when the experiment is done, the number of photons reaching the screen depends on the difference in lengths between the two paths in just the way you expect from the wave description! The probability that a photon will hit a given spot on the screen is proportional to the intensity of the corresponding classical wave. If the path lengths produce destructive interference, no photons get through. Not only that, but similar experiments can be done with other particles, such as neutrons! Maybe interference is not so trivial after all.

8.4 Transmission Lines

We have seen that a translation invariant system of inductors and capacitors can carry waves. Let us ask what happens when we take the continuum limit of such a system. This will give an interesting insight into electromagnetic waves. The dispersion relation for the system of figure 5.23 is given by (5.75),

$$\omega^2 = \frac{4}{L_a C_a} \sin^2 \frac{ka}{2} \,. \tag{8.49}$$

where L_a and C_a are the inductance and capacitance of the inductors and capacitors for the system with separation a between neighboring parts. To take the continuum limit, we must replace the inductance and capacitance, L_a and C_a, by quantities that we expect to have finite limits as $a \to 0$. We expect from the analogy, (5.69), between LC circuits and systems of springs and masses, and the discussion at the beginning of chapter 7 about the continuum limit of the system of masses and springs that the relevant quantities will be:

$$\begin{aligned} \rho_L &\to \frac{L_a}{a} && \text{inductance per unit length} \\ K_a a &\to \frac{a}{C_a} && \frac{1}{\text{capacitance per unit length}} \end{aligned} \tag{8.50}$$

These two quantities can be computed directly from the inductance and capacitance of a finite length, ℓ, of the system that contains many individual units. The inductances are connected in series so the individual inductances add to give the total inductance. Thus if the length ℓ is na so that if the finite system contains n inductors, the total inductance is $L = nL_a$. Then

$$\frac{L}{\ell} = \frac{L_a}{a} \,. \tag{8.51}$$

The capacitances work the same way because they are connected in parallel, and parallel capacitances add. Thus

$$\frac{C}{\ell} = \frac{C_a}{a} \,. \tag{8.52}$$

Therefore, in taking the limit as $a \to 0$ of (8.49), we can write

$$L_a = a\frac{L}{\ell} \,, \quad C_a = a\frac{C}{\ell} \,. \tag{8.53}$$

This gives the following dispersion relation:

$$\omega^2 = \frac{\ell^2}{LC} \frac{4 \sin^2 \frac{ka}{2}}{a^2} \to \frac{\ell^2}{LC} k^2 \,. \tag{8.54}$$

A continuous system like this with fixed inductance and capacitance per unit length is called a transmission line. We will call (8.54) the dispersion relation for a resistanceless transmission line. A transmission line can be used to send electrical waves, just as a continuous string transmits mechanical waves. In the continuous system, the displacement variable, the displaced charge, becomes a function of position along the transmission line. If the transmission line is stretched in the z direction, we can describe the charges on the transmission line by a function $Q(z,t)$ that is the charge that has been displaced through the point z on the transmission line at time t. The time derivative of $Q(z,t)$ is the current at the point z and time t:

$$I(z,t) = \frac{\partial Q(z,t)}{\partial t}\,. \tag{8.55}$$

8.4.1 Parallel Plate Transmission Line

It is worth working out a particular example of a transmission line. The example we will use is of two long parallel conducting strips. Imagine an infinite system in which the strips are stretched parallel to one another in planes of constant y, going to infinity in the z direction. Suppose that the strips are sufficiently thin that we can neglect their thickness. Suppose further that the width of strips, w, is much larger than the separation, s. A cross section of this transmission line in the $x-y$ plane is shown in figure 8.9. In the figure, the z direction is out of the plane of the paper, toward you. We will keep track of the motion of the charges in the upper conductor and assume that the lower conductor is grounded (with voltage fixed at $V=0$).

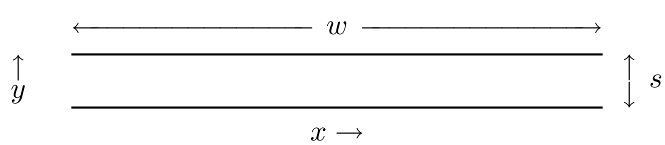


Figure 8.9: Cross section of a transmission line in the $x-y$ plane.

We will find the dispersion relation of the transmission line by computing the capacitance and inductance of a part of the line of length ℓ. It will be useful to do this using energy considerations. Suppose that there is a charge, Q, uniformly spread over the upper plate of the capacitor, and a current, I, flowing evenly out of the $x-y$ plane in the z direction along the upper conductor (and back into the plane along the lower conductor). The energy stored in the length, ℓ, of the transmission line is then

$$\frac{1}{2C}Q^2 + \frac{1}{2}LI^2\,, \tag{8.56}$$

where C and L are the capacitance and inductance.[4]

[4]See Haliday and Resnick, part 2.

The energy is actually stored in the electric and magnetic fields produced by the charge and current. In this configuration, the electric and magnetic fields are almost entirely between the two plates of the piece of the transmission line. If Q and I are positive, the electric and magnetic fields are as shown in figure 8.10 and figure 8.11. In figure 8.10, the dotted line is a cross section of a box-shaped region that can be used to compute the electric field, using Gauss's law. In figure 8.11, the dotted path can be used to compute the magnetic field, using Ampere's law. The electric and magnetic fields are approximately constant between the strips, but quickly fall off to near zero outside.

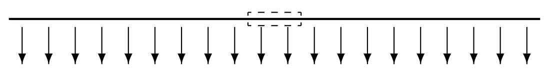


Figure 8.10: The electric field due to the charge on the transmission line.

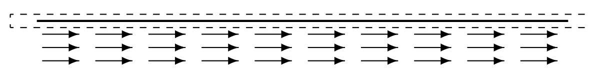


Figure 8.11: The magnetic field due to the current on the transmission line.

The charge density on the upper plate is approximately uniform and given by the total charge divided by the area, $w\ell$,

$$\sigma \approx \frac{Q}{w\ell}\,. \tag{8.57}$$

Then we can apply Gauss's law to a small box-shaped region, a cross section of which is shown in figure 8.10 and conclude that the electric field inside is given by

$$E_y \approx -\frac{Q}{\epsilon_0 w\ell} \tag{8.58}$$

The density of energy stored in the electric field between the plates is therefore

$$u_E = \frac{\epsilon_0}{2}E^2 \approx \frac{Q^2}{2\epsilon_0 w^2\ell^2}\,. \tag{8.59}$$

The total energy stored in the electric field is then obtained by multiplying u_E by the volume between the plates, yielding

$$\frac{1}{2}\frac{s}{\epsilon_0 w\ell}Q^2 \tag{8.60}$$

thus (comparing with (8.56))

$$C = \frac{\epsilon_0 w \ell}{s} \,. \tag{8.61}$$

We can calculate the inductance in a similar way. Ampere's law, applied to a path enclosing the upper conductor (as shown in figure 8.11) gives

$$B_x \approx \frac{\mu_0 I}{w} \,. \tag{8.62}$$

The density of energy stored in the magnetic field between the plates is therefore

$$u_B = \frac{1}{2\mu_0} B^2 \approx \frac{\mu_0 I^2}{2w^2} \,. \tag{8.63}$$

The total energy stored in the magnetic field is then obtained by multiplying u_B by the volume between the plates, yielding

$$\frac{1}{2} \frac{\mu_0 s \ell}{w} I^2 \tag{8.64}$$

thus (comparing with (8.56))

$$L = \frac{\mu_0 s \ell}{w} \,. \tag{8.65}$$

We can now put (8.61) and (8.65) into (8.54) to get the dispersion relation for this transmission line:

$$\omega^2 = \frac{1}{\mu_0 \epsilon_0} k^2 = c^2 k^2 \,, \tag{8.66}$$

where c is the speed of light!

8.4.2 Waves in the Transmission Line

The dispersion relation, (8.66), looks suspiciously like the dispersion relation for electromagnetic waves. In fact, the electric and magnetic fields **between** the strips of the transmission line have exactly the form of an electromagnetic wave. To see this explicitly, let us look at a traveling wave on the transmission line, and consider the charge, $Q(z,t)$, displaced through z, with the irreducible complex exponential z and t dependence,

$$Q(z,t) = q\, e^{i(kz-\omega t)} \,. \tag{8.67}$$

This wave is traveling in the positive z direction, out toward you in the diagram of figure 8.9.

At any fixed time, t and position, z, the electric and magnetic fields inside the transmission line look as shown in figure 8.10 and figure 8.11 (or both may point in the opposite direction). We can find the magnetic field just as we did above, because the current at any point along the line is given by (8.55), so

$$B_x(z,t) \approx \frac{\mu_0 I(z,t)}{w} = \frac{\mu_0}{w}\frac{\partial}{\partial t}Q(z,t) = -i\frac{\mu_0 \omega q}{w}\, e^{i(kz-\omega t)} \,. \tag{8.68}$$

To find the electric field as a function of z and t, we need the density of charge along the line. Once we have that, we can find the electric field using Gauss's law, as above. A nonzero charge density results if the amount of charge displaced **changes** as a function of z. It is easiest to find the charge density by returning to the discrete system discussed in chapter 5, and to (5.72). In the language in which we label the parts of the system by their positions, the charge, q_j, in the discrete system becomes $q(z,t)$ where $z = ja$. As $a \to 0$, this corresponds to a linear charge density along the transmission line of

$$\rho(z,t) = \frac{q(z,t)}{a} \,. \tag{8.69}$$

In this language, (5.72) becomes

$$q(z,t) = Q(z,t) - Q(z+a,t) \,, \tag{8.70}$$

where $Q(z,t)$ is the charge displaced through the inductor a position z at time t. Combining (8.69) and (8.70) gives

$$\rho(z,t) = \frac{Q(z,t) - Q(z+a,t)}{a} \,. \tag{8.71}$$

Taking the limit as $a \to 0$ gives

$$\rho(z,t) = -\frac{\partial}{\partial z}Q(z,t) = -ikq\, e^{i(kz-\omega t)} \,. \tag{8.72}$$

This linear charge density is spread out over the width of the upper strip in the transmission line, giving a surface charge density of

$$\sigma(z,t) = \frac{\rho(z,t)}{w} = -i\frac{kq}{w}\, e^{i(kz-\omega t)} \,. \tag{8.73}$$

Now the electric field from Gauss's law is

$$E_y = -\frac{\sigma(z,t)}{\epsilon_0} = i\frac{kq}{\epsilon_0 w}\, e^{i(kz-\omega t)} \,. \tag{8.74}$$

Comparing (8.68) with (8.74), you can see that (8.45) is satisfied, so that this pair of electric and magnetic fields form a part of a traveling electromagnetic plane wave.

What is happening here is that the role of the charges and currents in the strips of the transmission line is to **confine** the electromagnetic waves. Without the conductors it would impossible to produce a **piece** of a plane wave, as we will see in much more detail in chapter 13.

Meanwhile, note that the mode with $\omega = 0$ and $k = 0$ must be treated with care, as with the $\omega = k = 0$ mode of the beaded string discussed in chapter 5. The mode in which the displaced charge is proportional to z (see (5.41)) describes a situation in which the entire infinite transmission line is charged. This is not very interesting in the finite case. However, the mode that is independent of z, but increasing with time, proportional to t is important. This describes the situation in which a constant current is flowing through the conductors. Inside the transmission line, in this case, is a constant magnetic field.

8.5 Damping

It is instructive, at this point, to consider waves in systems with frictional forces. We have postponed this until now because it will be easier to understand what is happening in systems with damping now that we have discussed traveling waves.

The key observation is that in a translation invariant system, even in the presence of damping, the normal modes of the infinite system are exactly the same as they were without damping, because they are still determined by translation invariance. The normal modes are still of the form, $e^{\pm ikx}$, characterized by the angular wave number k. Only the dispersion relation is different. To see how this goes in detail, let us recapitulate the arguments of chapter 5.

The dispersion relation for a system without damping is determined by the solution to the eigenvalue equation

$$\left[-\omega^2 + M^{-1}K\right] A^k = 0\,, \tag{8.75}$$

where A^k is the normal mode with wave number k,

$$A_j^k \propto e^{ijka}\,, \tag{8.76}$$

with time dependence $e^{-i\omega t}$.[5] We already know that A^k is a normal mode, because of translation invariance. This implies that it is an eigenvector of $M^{-1}K$. The eigenvalue is some function of k. We will call it $\omega_0^2(k)$, so that

$$M^{-1}K\, A^k = \omega_0^2(k)\, A^k\,. \tag{8.77}$$

This function $\omega_0^2(k)$ **determines** the dispersion relation for the system without damping, because the eigenvalue equation, (8.75) now implies

$$\omega^2 = \omega_0^2(k)\,. \tag{8.78}$$

[5]In the presence of damping, the sign of i matters. The relations below would look different if we had used $e^{i\omega t}$, and we could not use $\cos\omega t$ or $\sin\omega t$.

We can now modify the discussion above to include damping in the infinite translation invariant system. In the presence of damping, the equation of motion looks like

$$M\frac{d^2}{dt^2}\psi(t) = -M\Gamma\frac{d}{dt}\psi(t) - K\psi(t)\,, \tag{8.79}$$

where $M\Gamma$ is the matrix that describes the velocity dependent damping. Then for a normal mode,

$$\psi(t) = A^k\, e^{-i\omega t}\,, \tag{8.80}$$

the eigenvalue equation now looks like

$$\left[-\omega^2 - i\Gamma\omega + M^{-1}K\right]\, A^k = 0\,. \tag{8.81}$$

Now, just as in (8.77) above, because of translation invariance, we know that A^k is an eigenvector of both $M^{-1}K$ and Γ,

$$M^{-1}K\, A^k = \omega_0^2(k)\, A^k\,, \quad \Gamma\, A^k = \gamma(k)\, A^k\,. \tag{8.82}$$

Then, as above, the eigenvalue equation becomes the dispersion relation

$$\omega^2 = \omega_0^2(k) - i\gamma(k)\omega\,. \tag{8.83}$$

For all k, $\gamma(k) \geq 0$, because as we will see in (8.84) below, the force is a frictional force. If $\gamma(k)$ were negative for any k, then the "frictional" force would be feeding energy into the system instead of damping it. Note also that if $\Gamma = \gamma I$, then $\gamma(k) = \gamma$, independent of k. However, in general, the damping will depend on k. Modes with different k may get damped differently.

In (8.83), we see the new feature of translation invariant systems with damping. **The only difference is that the dispersion relation becomes complex.** Both $\omega_0^2(k)$ and $\gamma(k)$ are real for real k. Because of the explicit i in (8.83), either ω or k (or both) must be complex to satisfy the equation of motion.

8.5.1 Free Oscillations

For free oscillations, the angular wave numbers, k, of the allowed modes are determined by the boundary conditions. Typically, the allowed k values are real and $\omega_0^2(k)$ is positive (corresponding to a stable equilibrium in the absence of damping). Then the modes of free oscillation are analogous to the free oscillations of a damped oscillator discussed in chapter 2. In fact, if we substitute $\alpha \to -i\omega$ and $\Gamma \to \gamma(k)$ in (2.5), we get precisely (8.83). Thus we can take over the solution from (2.6),

$$-\,i\omega = -\frac{\gamma(k)}{2} \pm \sqrt{\frac{\gamma(k)^2}{4} - \omega_0^2(k)}\,. \tag{8.84}$$

This describes a solution that dies out exponentially in time. Whether it oscillates or dies out smoothly depends on the ratio of $\gamma(k)$ to $\omega_0(k)$, as discussed in chapter 2.

8.5.2 Forced Oscillation

8-3 – 8-5

Now consider a forced oscillation, in which we drive one end of a translation invariant system with angular frequency ω. After the free oscillations have died away, we are left with oscillation at the single, real angular frequency ω. As always, in forced oscillation problems, we think of the real displacement of the end of the system as the real part of a complex displacement, proportional to $e^{-i\omega t}$. Then the dispersion relation, (8.83), applies. Now the dispersion relation determines k, and k must be complex.

You may have noticed that none of the dispersion relations that we have studied so far depend on the **sign** of k. This is not an accident. The reason is that all the systems that we have studied have the property of reflection symmetry. We could change $x \to -x$ without affecting the physics. In fact, a translation invariant system that did not have this symmetry would be a little peculiar. As long as the system is invariant under reflections, $x \to -x$, the dispersion relation cannot depend on the sign of k. The reason is that when $x \to -x$, the mode e^{ikx} goes to e^{-ikx}. If $x \to -x$ is a symmetry, these two modes with angular wave numbers k and $-k$ must be physically equivalent, and therefore must have the same frequency. Thus the two solutions for fixed ω must have the form:

$$k = \pm(k_r + ik_i) \tag{8.85}$$

Because of the $\pm$ sign, we can choose $k_r > 0$ in (8.85).

In systems with frictional forces, we always find

$$k_i \geq 0 \text{ for } k_r > 0\,. \tag{8.86}$$

The reason for this is easy to see if you consider the traveling waves, which have the form

$$e^{-i\omega t}\, e^{\pm i(k_r + ik_i)x} \tag{8.87}$$

or

$$e^{i(\pm k_r x - \omega t)}\, e^{\mp k_i x}\,. \tag{8.88}$$

From (8.88), it should be obvious what is going on. When the $\pm$ is +, the wave is going in the $+x$ direction, so the sign of the real exponential is such that the amplitude of the wave decreases as x increases. The wave peters out as it travels! This is what must happen with a frictional force. The other sign would require a source of energy in the medium, so that the wave amplitude would grow exponentially as the wave travels. A part of an infinite damped traveling wave is animated in program 8-3.

The form, (8.88) has some interesting consequences for forced oscillation problems in the presence of damping. In damped, **discrete** systems, even in a normal mode, the parts of the system do not all oscillate in phase. In damped, **continuous** systems, the distinction between traveling and standing waves gets blurred.

Consider a forced oscillation problem for the transverse oscillation of a string with one end, at $x = 0$ fixed, and the other end, $x = L$ driven at frequency ω. It will not matter until the end of our analysis whether the string is continuous, or has beads with separation a such that $na = L$ for integer n. The boundary conditions are

$$\psi(L,t) = A\cos\omega t\,, \quad \psi(0,t) = 0\,. \tag{8.89}$$

As usual, we regard $\psi(x,t)$ as the real part of a complex displacement, $\tilde{\psi}(x,t)$, satisfying

$$\tilde{\psi}(L,t) = Ae^{-i\omega t}\,, \quad \tilde{\psi}(0,t) = 0\,. \tag{8.90}$$

If k, for the given angular frequency ω, is given by (8.85), then the relevant modes of the infinite system are those in (8.87), and we must find a linear combination of these two that satisfies (8.89). The answer is

$$\tilde{\psi}(x,t) = A\,\left[\left(\frac{e^{i(k_r+ik_i)x} - e^{-i(k_r+ik_i)x}}{e^{i(k_r+ik_i)L} - e^{-i(k_r+ik_i)L}}\right)e^{-i\omega t}\right]. \tag{8.91}$$

The factor in parentheses is constructed to vanish at $x = 0$ and to equal 1 at $x = L$.

For a continuous string, the solution, (8.91), is animated in program 8-4. The interesting thing to notice about this is that near the $x = L$ end, the solution looks like a traveling wave. The reason is that here, the real exponential factors in (8.91) enhance the left-moving wave and suppress the right-moving wave, so that the solution is very nearly a traveling wave moving to the left. On the other hand, near $x = 0$, the real exponential factors are comparable, and the solution is very nearly a standing wave. We will discuss the more complicated behavior in the middle in the next chapter.

The same solution works for a beaded string (although the dispersion relation will be different). An example is shown in the animation in program 8-5. Here you can see very clearly that the parts of the system are not all in phase.

8.6 High and Low Frequency Cut-Offs

8.6.1 More on Coupled Pendulums

8-6

In the previous section, we saw how the angular wave number, k, can become complex in a system with friction. There is another important way in which k can become complex. Consider the dispersion relation for the system of coupled pendulums, (5.35), which we can rewrite as follows:

$$\omega^2 = \omega_\ell^2 + \omega_c^2 \sin^2\frac{ka}{2}\,. \tag{8.92}$$

Here a is interblock distance, ω_ℓ is the frequency of a single uncoupled pendulum, and ω_c^2 is a frequency associated with the coupling between neighboring blocks.

$$\omega_c^2 = \frac{4K}{m} \tag{8.93}$$

where m is the mass of a block and K is the spring constant of the coupling springs.

Traveling waves in a system with a dispersion relation like (8.92) are animated in program 8-6. To make the physics easier to see, this system is a beaded string with transverse oscillations. However, to produce the ω_ℓ^2 term in (8.92), we have also attached each bead by a spring to an equilibrium position along the dotted line. In this case, the coupling between beads comes from the string, so the analog of (8.93) is

$$\omega_c^2 = \frac{4T}{ma} \,. \tag{8.94}$$

The parameters in the system are chosen so that in terms of a reference frequency, ω_0,

$$\omega_\ell^2 = 25\omega_0^2 \,, \quad \omega_c^2 = 24\omega_0^2 \,. \tag{8.95}$$

The properties of waves in this system differ dramatically as a function of ω. One way to see this is to go backwards and note that for real k, because $\sin^2 \frac{ka}{2}$ must be between 0 and 1, ω is constrained,

$$\omega_\ell \leq \omega \leq \sqrt{\omega_\ell^2 + \omega_c^2} \equiv \omega_h \,. \tag{8.96}$$

For k in this "allowed" region,

$$\sin^2 \frac{ka}{2} = \frac{\omega^2 - \omega_\ell^2}{\omega_c^2} \tag{8.97}$$

is between 0 and 1, as is

$$\cos^2 \frac{ka}{2} = \frac{\omega_h^2 - \omega^2}{\omega_c^2} \,. \tag{8.98}$$

The two frequencies, ω_ℓ and ω_h, are called low and high frequency cut-offs. The system of coupled pendulums supports traveling waves only for frequency ω between the high and low frequency cut-offs. It is only in this region that the dispersion relation can be satisfied for real ω and k. For $\omega < \omega_\ell$ or $\omega > \omega_h$, the system oscillates, but there is nothing quite like a traveling wave. You can see this in program 8-6 by changing the frequency up and down with the arrow keys.

For any ω, we can always solve the dispersion relation. However, in some regions of frequency, the result will be complex, as in (8.85). We expect $k_i = 0$ in the allowed region (8.96). The solution of (8.92) for k_r and k_i as functions of ω are shown in the graphs in figure 8.12. Here, k_r and k_i are plotted against ω for the dispersion relation, (8.92), with $\omega_\ell = 5\omega_0$ and $\omega_h = 7\omega_0$. k_i is the dotted line. Note the very rapid dependence of k_i near the high and low frequency cut-offs.

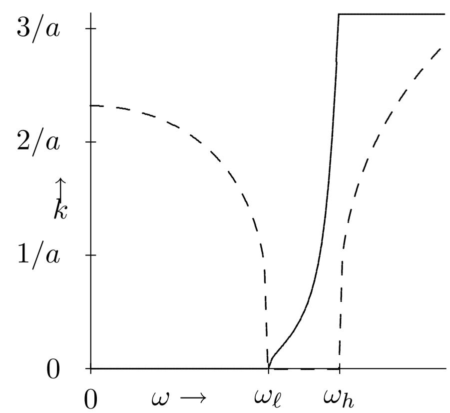


Figure 8.12: $k_r a$ and $k_i a$ versus ω.

As ω decreases, in the allowed region, (8.96), $\sin\frac{ka}{2}$ decreases. At the low frequency cut-off, $\omega = \omega_\ell$, $\sin\frac{ka}{2}$ and therefore k goes to zero. This means that as the frequency decreases, the wavelength of the traveling waves gets longer and longer, until at the cut-off frequency, it becomes infinite. At the low frequency cut-off, every pendulum in the infinite chain is oscillating in phase. The springs that couple them are then irrelevant because they always maintain their equilibrium lengths. This is possible precisely because ω_ℓ is the oscillation frequency of the uncoupled pendulum, so that no coupling is required for an individual pendulum to swing at frequency ω_ℓ.

If ω is below the low frequency cut-off, ω_ℓ, $\sin^2\frac{ka}{2}$ must become negative to satisfy the dispersion relation, (8.92). Therefore $\sin\frac{ka}{2}$ must be a pure imaginary number

$$k = \pm i k_i \,. \tag{8.99}$$

The general solution for the wave is then

$$\psi(x,t) = A\,e^{-k_i x}\,e^{-i\omega t} + B\,e^{k_i x}\,e^{-i\omega t}\,. \tag{8.100}$$

In a finite system of coupled pendulums, both terms may be present. In a semi-infinite system that is driven at $x = 0$ and extends to $x \to \infty$, the constant B must vanish to avoid exponential growth of the wave at infinity. Thus the wave falls off exponentially at large x. Furthermore, the solution is a product of a real function of x and a complex exponential function of t. This is a standing wave. There is no traveling wave. You can see this in program 8-6 at low frequencies.

The physics of this oscillation below the low frequency cut-off is particularly clear in the extreme limit, $\omega \to 0$. At zero frequency, there is no motion. The analog of a forced oscillation problem is just to displace one pendulum from equilibrium and look to see what happens to the rest. Clearly, what happens is that the displacement of the first pendulum causes a force on the next one because of the coupling spring that pulls it away from equilibrium, but not as far as the first. Its displacement is smaller than that of the first by some factor $\epsilon = e^{-k_i a}$. Then the second pendulum pulls the third, but again the displacement is smaller by the same factor. And so on! In an infinite system, this gives rise to the exponentially falling displacement in (8.100) for $B = 0$. As the frequency is increased, the effect of inertia (more precisely, the ma term in $F = ma$) increases the displacement of second (and each subsequent) block, until above the low frequency cut-off, the effect of inertia is large enough to compete on an equal footing with the effect of the restoring force, and a real traveling wave can be produced.

The low frequency cut-off is not peculiar to the discrete system. It occurs any time there is a restoring force for $k = 0$ in the infinite system. Later, in chapter 11, we will see that a similar phenomena can occur in two- and three-dimensional systems even when there is no restoring force at $k = 0$.

The high frequency cut-off, on the other hand, depends on the finite separation between blocks. As ω increases, in the allowed region, (8.96), $\sin\frac{ka}{2}$ increases, k increases, and therefore $\cos\frac{ka}{2}$ decreases. At the high frequency cut-off, $\omega = \omega_h$, $\sin\frac{ka}{2} = 1$ and $\cos\frac{ka}{2} = 0$. But

$$\sin\frac{ka}{2} = 1 \Rightarrow k = \frac{\pi}{a} \tag{8.101}$$

which, in turn means

$$e^{ika} = e^{-ika} = -1. \tag{8.102}$$

Thus the displacement of the blocks simply alternates, because

$$\psi_j = \psi(ja, t) \propto e^{ij\pi} = (-1)^j\,. \tag{8.103}$$

This is as wavy as the discrete system can get. In a discrete system with interblock separation, a, the maximum possible real part of k is $\frac{\pi}{a}$ (because k can be redefined by a multiple of $\frac{2\pi}{a}$ without changing the displacements of any of the blocks – see (5.28)). This bound is the origin of the high frequency cut-off.

You can see this in program 8-6. The frequency starts out at $6\omega_0$. At this point, $k_r a$ is quite small (and $k_i = 0$) and the wave looks smooth. As the frequency is increased toward ω_h, the wave gets more and more jagged looking, until at $\omega = \omega_h$, neighboring beads are moving in opposite directions.

For $\omega > \omega_h$, $\sin\frac{ka}{2}$ is greater than 1, and $\cos\frac{ka}{2}$ is negative. This implies that k has the form

$$k = \frac{\pi}{a} \pm i k_i\,. \tag{8.104}$$

Then the general solution for the displacement is

$$\psi(x,t) = A\, e^{-k_i x}\, e^{i\pi x/a}\, e^{-i\omega t} + B\, e^{k_i x}\, e^{i\pi x/a}\, e^{-i\omega t}\,. \tag{8.105}$$

As in (8.100), there is an exponentially falling term and an exponentially growing one. Here however, there is also a phase factor, $e^{i\pi x/a}$, that looks as if it might lead to a traveling wave. But in fact, this is not really a phase. It simply produces the alternation of the displacement from one block to the next. We see this if we look only at the displacements of the blocks (as in (8.103),

$$\psi_j = \psi(ja, t) = A\,(-1)^j\, e^{-k_i x}\, e^{-i\omega t} + B\,(-1)^j\, e^{k_i x}\, e^{-i\omega t}\,. \tag{8.106}$$

As for (8.100), in a semi-infinite system that extends to $x \to \infty$, we must have $B = 0$, and there is no travelling wave.

One of the striking things about program 8-6 is the very rapid switch from a traveling wave solution in the allowed region to a standing wave solution with a rapid exponential decay of the amplitude in the high and low frequencies regions. You see this also in figure 8.12 in the rapid change of k_i near the cut-offs. The reason for this is that k has a square-root dependence on the frequency near the cut-offs.

In the infinite system, the solution outside the allowed region is a pure standing wave. In the absence of damping, the work done by the force that produces the wave averages to zero over time. In a finite system, however, it is possible to transfer energy from one end of a system to the other, even if you are below the low frequency cut-off or above the high-frequency cutoff. The reason is that in a finite system, both the A and B terms in (8.100) (or (8.106)) can be nonzero. If A and B are both real (or relatively real — that is if they have the same phase), then there is no energy transfer. The solution is the product of a real function of x (or j) and an oscillating exponential function of t. Thus it looks like a standing wave. However if A and B have different phases, then the oscillation looks something like a traveling wave and energy can be transferred. This process becomes exponentially less efficient as the length of the system increases. We will discuss this in more detail in chapter 11.

Chapter Checklist

You should now be able to:

i. Construct traveling wave modes of an infinite system with translation invariance;

ii. Decompose a traveling wave into a pair of standing waves, and a standing wave into a pair of traveling waves “moving” in opposite directions;

iii. Solve forced oscillation problems with traveling wave solutions and compute the forces acting on the system.

iv. Compute the power and average power required to produce a wave, and define and calculate the impedance;

v. Analyze translation invariant systems with damping;

vi. Understand the physical origins of high and low frequency cut-offs and be able to analyze the behavior of systems driven above and below the cut-off frequencies.

Problems

8.1. An infinite string with tension T and linear mass density ρ is stretched along the x axis. A force is applied in the y direction at $x = 0$ so as to cause the string at $x = 0$ to oscillate in the y direction with displacement

$$A(t) = D \cos \omega t \, .$$

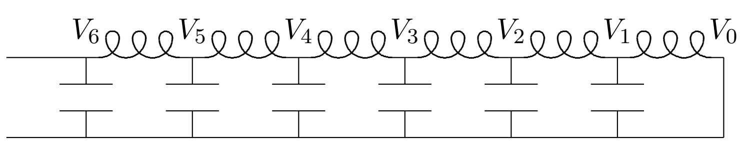
This produces two traveling waves moving away from $x = 0$ in the $\pm x$ directions.

a. Find the force applied at $x = 0$.

b. Find the average power supplied by the force.

8.2. For air at standard temperature and pressure, the pressure is 1.01×10^6 dyne/cm^2, the density is 1.29×10^{-3} gr/cm^3. Use these to find the displacement amplitude for sound waves with a frequency of 440 cycles/sec (Hertz) carrying a power per unit area of 10^{-3} watts/cm^2.

8.3. Consider the following circuit:



All the capacitors have the same capacitance, $C \approx 0.00667\mu F$, and all the inductors have the same inductance, $L \approx 150\mu H$ and the same resistance, $R \approx 15\Omega$ (this is the same problem as (5.4), but with nonzero resistance). The wire at the bottom is grounded so that $V_0 = 0$. This circuit is an electrical analog of the translation invariant systems of coupled mechanical oscillators that we have discussed in this chapter.

a. Show that the dispersion relation for this system is

$$\omega^2 + i\omega \frac{R}{L} = \frac{2}{LC} (1 - \cos ka) \, .$$

When you apply a harmonically oscillating signal from a signal generator through a coaxial cable to V_6, different oscillating voltages will be induced along the line. That is if

$$V_6(t) = V\cos\omega t\,,$$

then $V_j(t)$ has the form

$$V_j(t) = A_j\cos\omega t + B_j\sin\omega t\,.$$

b. Find A_1 and B_1 and $|A_1 + iB_1|$ and graph each of them versus ω from $\omega = 0$ to $2/\sqrt{LC}$. Never mind simplifying complicated expressions, so long as you can graph them. How many of the resonances can you identify in each of the graphs? **Hint:** Use the trigonometric identity of problem (1.2e),

$$\sin 6x = \sin x\,\left(32\cos^5 x - 32\cos^3 x + 6\cos x\right)$$

to express $A_1 + iB_1$ in terms of $\cos ka$. Note that this identity is true even if x is a complex number. Then use the dispersion relation to express $\cos ka$ in terms of ω. Find A_1 and B_1 by taking the real and imaginary parts of $A_1 + iB_1$. Finally, program a computer to construct the graphs.[6]

c. Find the positions of the resonances directly using the arguments of chapter 5, and show that they are where you expect them.

[6]This hint dates from the days before Mathematica was generally available. You may choose to to the problem differently, and that is OK as long as you explain clearly what you are doing and understand it!

Chapter 9

The Boundary at Infinity

Although the wave phenomena we can see in the laboratory live in finite regions of space, it is often convenient to analyze them as if the traveling waves come in from and go out to infinity. We have described traveling waves in infinite translation invariant systems. But traveling waves are more complicated and more interesting in systems in which there are boundaries that break the translation symmetry.

Preview

In this chapter, we introduce a new kind of "boundary condition" in systems that lack a boundary! It will enable us to discuss reflection and transmission, and in general, the phenomenon of scattering.

i. We discuss forced oscillation problems in semi-infinite systems, that extend to infinity in one direction. We show that we can impose a "boundary condition" even though there is no boundary, by specifying the amplitude of a wave traveling in one direction. We then discuss scattering problems in infinite systems, describing the amplitudes for transmission and reflection. We study the motion of a general wave with definite frequency.

ii. We discuss electromagnetic plane waves in a dielectric.

iii. We discuss reflection and transmission by a mass on a string and two masses on a string, showing how to use a "transfer matrix" to simplify the solution to the scattering problem. We analyze reflection from a boundary between regions with different wave number and show how to eliminate the reflection with a suitable "nonreflective coating."

9.1 Reflection and Transmission

9.1.1 Forced Oscillation

Consider the forced oscillation problem in a semiinfinite stretched string that runs from $x = 0$ to $x = \infty$. Suppose that

$$\psi(0,t) = A\cos\omega t\,. \tag{9.1}$$

Then what is $\psi(x,t)$? This is not a well-posed problem, because we only have a boundary condition on one side. Furthermore, $\psi(\infty,t)$ does not have a definite value. We can only talk about the value of a function at infinity if the function goes to a constant value. Here, we expect $\psi(x,t)$ to continue to oscillate as $x \to \infty$, so we cannot specify it. Instead, we can specify either the incoming (traveling toward the boundary at $x = 0$ in the $-x$ direction) or the outgoing (traveling away from $x = 0$ in the $+x$ direction) traveling waves in the system. This is called a **"boundary condition at ∞."**

For example, we could take our boundary condition at infinity to be that no incoming traveling waves appear on the string. Physically, this corresponds to the situtation in which the motion of the string at $x = 0$ is producing the waves. In general, we can write a solution with angular frequency ω as a sum of four real traveling waves

$$\begin{gathered}\psi(x,t) = a\cos(kx-\omega t) + b\sin(kx-\omega t)\\ +c\cos(kx+\omega t) + d\sin(kx+\omega t)\,.\end{gathered} \tag{9.2}$$

Then (9.1) implies

$$a + c = A\,, \quad b - d = 0\,, \tag{9.3}$$

and the boundary condition at ∞ implies

$$c = d = 0\,. \tag{9.4}$$

Thus

$$\psi(x,t) = A\cos(kx-\omega t)\,. \tag{9.5}$$

9.1.2 Infinite Systems

Now consider two semi-infinite strings with the same tension but different densities that are tied together at $x = 0$, as shown in figure 9.1. Suppose that in the $x \le 0$ region (Region I), there is an incoming traveling wave with amplitude A and angular frequency ω, and in the $x \ge 0$ region (Region II), there is no incoming traveling wave. This describes a physical situation in which the incoming wave in I is scattered by the boundary so that the other waves are a transmitted wave in II and a reflected wave in I, both outgoing.

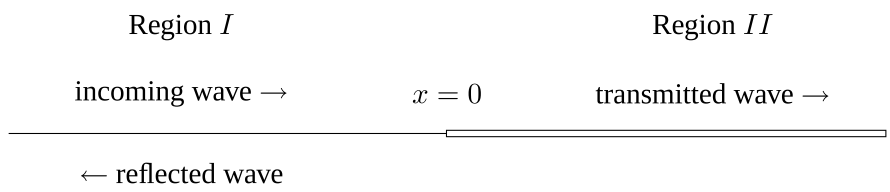


Figure 9.1: Two semi-infinite strings tied together at $x = 0$.

The key to this problem is to think of it as a forced oscillation problem. The incoming traveling wave in region I is what is "causing" all the oscillations. We have put the word in quotes, because the harmonic form, $e^{-i\omega t}$, for the oscillation implies that it has been going on forever, so that a philosopher might question this use of cause and effect. Nevertheless, it will help us to think of it this way. If the reflected and transmitted waves are produced by the incoming wave, their amplitudes will also be proportional to $e^{-i\omega t}$. As in a conventional forced oscillation problem, we could add on any free oscillations of the system. However, if there is any friction at all, these will die away with time, and we will be left only with the oscillation produced by the incoming traveling wave, proportional to $e^{-i\omega t}$. The important thing is that the frequency is the same in both regions, because as in a forced oscillation problem, the frequency is imposed on the system by an external agency, in this case, whatever produced the incoming traveling wave.

In our complex exponential notation in which everything has the irreducible time dependence, $e^{-i\omega t}$. Right moving waves are $\propto e^{ikx}\, e^{-i\omega t}$ and left moving waves are $\propto e^{-ikx}\, e^{-i\omega t}$. In this case, the boundary conditions at $\pm\infty$ require that

$$\psi(x,t) = e^{ikx}\, Ae^{-i\omega t} + R\, Ae^{-ikx}\, e^{-i\omega t} \tag{9.6}$$

for $x \leq 0$ in Region I, and

$$\psi(x,t) = \tau\, Ae^{ik'x}\, e^{-i\omega t} \tag{9.7}$$

for $x \geq 0$ in Region II. The k and k' are

$$k = \omega\, \sqrt{\rho_I/T}\,, \quad k' = \omega\, \sqrt{\rho_{II}/T}\,, \tag{9.8}$$

and R and τ are (in general) complex numbers that determine the reflected and transmitted waves. They are sometimes called the "reflection coefficient" and "transmission coefficient," or the "amplitudes" for transmission and reflection. Notice that we have defined the reflection and transmission coefficients by taking out a factor of the amplitude, A, of the incoming wave. The amplitude, A, then drops out of all the boundary conditions, and the dimensionless coefficients R and τ are independent of A. This must be so because of the linearity of the system. We know that once we have found the solution, $\psi(x,t)$, for an incoming amplitude, A, we can find the solution for an incoming amplitude, B, by multiplying our solution by B/A. We will keep the parameter, A, in our expressions for $\psi(x,t)$, mostly in order to keep the units right. A has units of length in this example, but in general, the amplitude of the incoming wave will have units of generalized displacement (as in (1.107) and (1.108)).

To determine R and τ, we need a boundary condition at $x = 0$ where (9.6) and (9.7) meet. Clearly $\psi(x,t)$ must be continuous at $x = 0$, thus

$$1 + R = \tau\,. \tag{9.9}$$

We have canceled the common factor of $Ae^{-i\omega t}$ from both sides. The x derivative must also be continuous (for a massless knot) because the vertical forces on the knot must balance, thus

$$ik(1 - R) = ik'\tau\,. \tag{9.10}$$

Solving for R and τ gives

$$\tau = \frac{2}{1 + k'/k}\,, \quad R = \frac{1 - k'/k}{1 + k'/k}\,. \tag{9.11}$$

9.1.3 Impedance Matching

Note that we could replace the string in Region II by a dashpot with the same impedance, Z_{II}. This must be true because of the local nature of the interactions. The only thing that the string for $x < 0$ knows about the string for $x > 0$ is that it exerts a force at $x = 0$ equal to

$$-\,Z_{II}\,\frac{\partial}{\partial t}\psi(0,t)\,. \tag{9.12}$$

Thus we also learned what happens when an incoming wave encounters a dashpot with the wrong impedance. The amplitude of the reflected wave is given by R in (9.11).

The reflected wave in (9.11) vanishes if $k = k'$. If $k = k'$, then $\rho_I = \rho_{II}$ (from (9.8)), and the impedance in region I is the same as the impedance in region II. This is a simple example of the important principle of “impedance matching.” There is no reflection if the impedance of the system in region II is the same as the impedance of the system in region I. The argument is the same as for the dashpot in the previous paragraph. What matters in the computation of the reflection coefficient are the forces that act on the string at $x = 0$. Those forces are determined by the impedances in the two regions. Nothing else matters. Consider, for example, the system shown in figure 9.2 of two semi-infinite strings connected at $x = 0$ to a massless ring which is free to slide in the vertical direction on a frictionless rod. The rod can exert a horizontal force on the ring, so the tensions in the two strings need not be the same. In such a system, we can change both the density and the tension in the string from region I to region II. There will be no reflection so long as the product of the linear mass density and the tension (and thus the impedance, from (8.22)) is the same in both regions,

$$Z_I = \sqrt{\rho_I T_I} = \sqrt{\rho_{II} T_{II}} = Z_{II}\,. \tag{9.13}$$

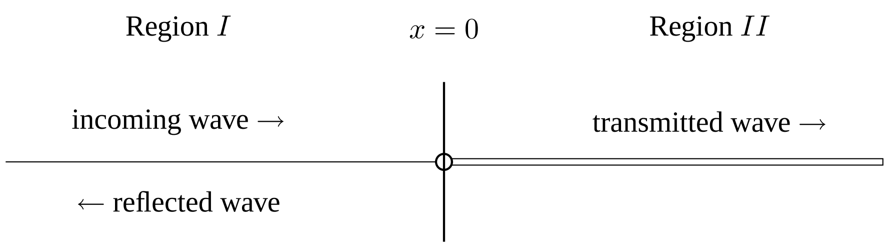


Figure 9.2: A system in which impedances can be matched.

It is instructive to solve the scattering problem completely for the more general case shown in figure 9.2. This will give us a feeling for the meaning of impedance. The form of the solution, (9.6) and (9.7) is unchanged, but now the angular wave numbers satisfy

$$k = \omega\sqrt{\rho_I/T_I}\,, \quad k' = \omega\sqrt{\rho_{II}/T_{II}}\,. \tag{9.14}$$

The boundary condition at $x = 0$ arising from the continuity of the string, (9.9), remains unchanged. However, (9.10) arose from the fact that the forces on the massless knot must sum to zero so the acceleration is not infinite. In this case, from (8.21), the contribution of each component of the wave to the total force is proportional to plus or minus the impedance in the relevant region depending on whether it is moving in the $+x$ or the $-x$ direction. Thus the boundary condition is

$$Z_I(1 - R) = Z_{II}\tau\,. \tag{9.15}$$

Then the reflection and transmission coefficients are

$$\tau = \frac{2Z_I}{Z_I + Z_{II}}\,, \quad R = \frac{Z_I - Z_{II}}{Z_I + Z_{II}}\,. \tag{9.16}$$

We have already discussed the case where the impedances match and the reflection coefficient vanishes. It is also interesting to look at the limits in which $R = \pm 1$. First consider the limit in which the impedance in region II goes to infinity,

$$\lim_{Z_{II}\to\infty} R = -1\,. \tag{9.17}$$

This is situation in which it takes an infinite force to produce a wave in region II. Thus the string in region II does not move at all, and in particular, the point $x = 0$ might as well be a fixed end. The solution, (9.17) ensures that the string does not move at $x = 0$, and therefore that the solution in region I is $\psi(x,t) \propto \sin kx$. This solution is an infinite standing wave with a fixed end boundary condition.

In the opposite limit, in which the impedance in region II is zero, we get

$$\lim_{Z_{II}\to 0} R = 1 \,. \tag{9.18}$$

This time, it takes no force at all to produce a wave in region II. Thus the end of region I at $x = 0$ feels no transverse force. It acts like a free end. The solution, (9.18) ensures that $\psi(x,t) \propto \cos kx$ in region I, so the slope of the string vanishes at $x = 0$. This solution is an infinite standing wave with a free end boundary condition.

9.1.4 Looking at Reflected Waves

9-1

In this section, we discuss what the displacement in Region I looks like. We will find a useful diagnostic for the presence of reflection. We will also conclude that standing waves are very special.

Look at a wave of the form

$$A\cos(kx-\omega t) + R\,A\cos(kx+\omega t)\,. \tag{9.19}$$

This describes an incoming traveling wave with some reflected wave of amplitude R (we could put in an arbitrary phase for the reflected wave but it would complicate the algebra without changing the physics).

For $R = \pm 1$, this is a standing wave. For $R = 0$, it is a traveling wave. To see how the system interpolates between these two extremes, consider the motion of the crest of the wave, a maximum of (9.19).

To find the maximum, we differentiate with respect to x and set the result to zero. Eliminating the irrelevant factor of A, we get

$$\sin(kx-\omega t) + R\sin(kx+\omega t) = 0 \;, \tag{9.20}$$

or

$$(1+R)\sin kx\cos\omega t = (1-R)\cos kx\sin\omega t \;, \tag{9.21}$$

or

$$\tan kx = \frac{1-R}{1+R}\,\tan\omega t \;. \tag{9.22}$$

(9.22) describes (implicitly — we could solve for x as a function of t if we felt like it) the motion of the maximum as a function of time. We can differentiate it to get the velocity:

$$k\,\left(1+\tan^2 kx\right)\,\frac{\partial x}{\partial t} = \frac{1-R}{1+R}\,\frac{\omega}{\cos^2\omega t}\;. \tag{9.23}$$

We have left $(1 + \tan^2 kx)$ in (9.23) so that we can eliminate it by using (9.22). Thus

$$\begin{aligned} \frac{\partial x}{\partial t} &= \frac{1-R}{1+R}\,\frac{\omega}{k}\,\frac{1}{(1+\tan^2 kx)\,\cos^2\omega t} \\ &= \frac{1-R}{1+R}\,\frac{\omega}{k}\,\frac{1}{\left(1+\left(\frac{1-R}{1+R}\right)^2\tan^2\omega t\right)\,\cos^2\omega t} \\ &= v\,\frac{(1+R)(1-R)}{(1+R)^2\cos^2\omega t+(1-R)^2\sin^2\omega t} \end{aligned} \tag{9.24}$$

where $v = \omega/k$ is the phase velocity. When $\sin\omega t$ vanishes, the speed of the maximum is smaller than the phase velocity by a factor of

$$\frac{1-R}{1+R}\,, \tag{9.25}$$

while when $\cos\omega t$ vanishes, the speed is larger than the v by the inverse factor,

$$\frac{1+R}{1-R}\,. \tag{9.26}$$

The wave thus appears to move in fits and starts. You can easily see this effect if you stare at a system with a lot of reflection. The effect is illustrated in program 9-1.

We can draw a more general moral from this discussion. The general case of wave motion is much more like a traveling wave than like a standing wave. Generically, except for $R = \pm 1$, the wave crests move with time. As we approach $R = \pm 1$, one of the two velocities in (9.25) and (9.26) goes to zero and the other goes to infinity. What happens when you are close to $R = \pm 1$ is then that the wave stays nearly still most of the time, and then moves very quickly to the next nearly stationary position. A standing wave is thus a degenerate special case of a traveling wave in which this motion is unobservable because, in a sense, it is infinitely fast.

9.1.5 Power and Reflection

It is instructive to consider the power required to produce a traveling wave that is partially reflected. That is, we consider the power required by a transverse force acting at $x = 0$ to produce a wave in the region $x > 0$ that is a linear combination of an outgoing wave moving in the $+x$ direction and an incoming wave moving in the $-x$ direction, such as might be produced by a reflection at some large value of x. Let us imagine the most general one-dimensional case, in a medium with impedance Z:

$$\begin{aligned} \psi(x,t) &= \mathrm{Re}\left(A_+ e^{i(kx-\omega t)} + A_- e^{i(-kx-\omega t)}\right) \\ &= R_+\cos(kx-\omega t+\phi_+) + R_-\cos(-kx-\omega t+\phi_-) \end{aligned} \tag{9.27}$$

where $R_\pm$ and $\phi_\pm$ are the absolute value and phase of the amplitude $A_\pm$. The velocity is

$$\frac{\partial}{\partial t}\psi(x,t) = \omega R_+ \sin(kx - \omega t + \phi_+) + \omega R_- \sin(-kx - \omega t + \phi_-)\,. \tag{9.28}$$

Now because (9.27) involves waves traveling both in the $+x$ and in the $-x$ direction, we cannot find the force required to produce the wave at the point x by simply multiplying (9.28) by the impedance, Z. However, we can use linearity. We can write $\psi(x,t) = \psi_+(x,t) + \psi_-(x,t)$, where $\psi_\pm(x,t)$ is the wave moving in the $\pm x$ direction. Then from (8.21), the force required to produce ψ_+ is

$$F_+(t) = Z\frac{\partial}{\partial t}\psi_+(0,t) \tag{9.29}$$

while the force required to produce ψ_- is

$$F_-(t) = -Z\frac{\partial}{\partial t}\psi_-(0,t)\,. \tag{9.30}$$

Then the total force required to produce ψ is

$$\begin{aligned} F(t) &= F_+(t) + F_-(t) \\ &= Z\omega R_+ \sin(-\omega t + \phi_+) - Z\omega R_- \sin(-\omega t + \phi_-)\,. \end{aligned} \tag{9.31}$$

Thus the power required is

$$\begin{aligned} P(t) &= F(t)\ \frac{\partial}{\partial t}\psi(x,t)\Big|_{x=0} \\ &= Z\omega^2 R_+^2 \sin^2(-\omega t + \phi_+) - Z\omega^2 R_-^2 \sin^2(-\omega t + \phi_-)\,. \end{aligned} \tag{9.32}$$

The average power is then given by

$$P_{\text{average}} = \frac{1}{2}Z\omega^2(R_+^2 - R_-^2) = \frac{1}{2}Z\omega^2\left(|A_+|^2 - |A_-|^2\right)\,. \tag{9.33}$$

The result, (9.32), has an obvious and important physical interpretation. Positive power is required to produce the outgoing traveling wave, while the incoming wave gives energy back to the system, and thus requires negative power. **The power required to produce a general traveling wave is thus proportional to the difference of the squares of the absolute values of the amplitudes of the outgoing and incoming waves.**

Note also that we can apply this discussion to the example of reflection at a boundary, discussed above. We can check that energy is conserved in this scattering. The average power required to produce the wave in region I is, from (9.33)

$$Z_I\omega^2 - Z_I\omega^2\, R^2\,. \tag{9.34}$$

The average power required to produce the wave in region II is,

$$Z_{II}\omega^2\tau^2\,. \tag{9.35}$$

Using (9.16), you can check that these are equal.

9.1.6 Mass on a String

9-2

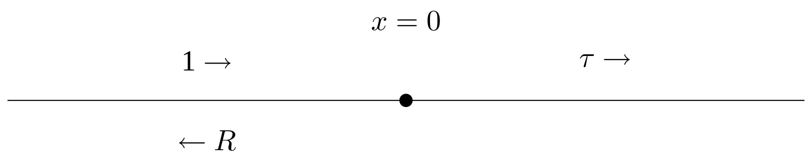


Figure 9.3: A mass on a string.

Consider the transmission and reflection of waves from a mass, m, at $x = 0$ on a string with linear mass density ρ and tension T, stretched from $x = -\infty$ to $x = \infty$, shown in figure 9.3. Before we calculate the coefficients for reflection and transmission, let us guess the result in two extreme limits.

m **small –** Here we expect that the reflection to be small and the transmission close to one, because in the limit

$$m \rightarrow 0 \Rightarrow \tau \rightarrow 1 \text{ and } R \rightarrow 0\,. \tag{9.36}$$

m **large –** Here we expect the transmission to be small and the reflection close to -1, because in the limit

$$m \rightarrow \infty \Rightarrow \tau \rightarrow 0 \text{ and } R \rightarrow -1\,. \tag{9.37}$$

"Large or small compared to what?" you ask! That we can answer by dimensional analysis. The relevant dimensional parameters are m, ω, k, ρ and T. However, one of these is not independent, because of the dispersion relation, (6.5). If we use (6.5) to eliminate T, then ω cannot be relevant to the question, because it is the only thing left that involves the unit of time. The only dimensionless quantity we can build is

$$\epsilon = \frac{m\,k}{\rho} = \frac{m\,\omega^2}{k\,T}\,. \tag{9.38}$$

Now that we have guessed, we can do the calculation. It follows from translation invariance and the boundary condition at $x = \infty$ that

$$\psi(x,t) = Ae^{ikx}\,e^{-i\omega t} + R\,Ae^{-ikx}\,e^{-i\omega t} \text{ for } x \leq 0 \tag{9.39}$$

$$\psi(x,t) = \tau\,Ae^{ikx}\,e^{-i\omega t} \text{ for } x \geq 0 \tag{9.40}$$

where, as usual, R and τ are "amplitudes" for the reflected and transmitted waves. The boundary conditions are

continuity – The fact that the string doesn't break implies that it is continuous, so that $\psi(0,t)$ can be computed with either (9.39) or (9.40). This implies

$$1+R=\tau\,. \tag{9.41}$$

$F=ma$ **–** The horizontal component of the tension in the string must be equal on the two sides. Both are about equal to T, for small displacements. However, if there is a kink in the string, the vertical components do not match, as shown in figure 9.4 (see also (8.16)-(8.17)). The force on the mass is then the tension times the slope for $x\geq 0$ minus the tension times the slope for $x\leq 0$, thus $F=ma$ becomes

$$\begin{aligned} T\left(\frac{\partial}{\partial x}\;\psi(x,t)|_{x=0^+}-\frac{\partial}{\partial x}\;\psi(x,t)|_{x=0^-}\right)\\ =m\,\frac{\partial^2}{\partial t^2}\,\psi(0,t)\end{aligned} \tag{9.42}$$

or

$$ik\,T\;(R-1+\tau)=-m\omega^2\,\tau\,. \tag{9.43}$$

Thus

$$1+R=\tau\,,\quad 1-R=(1-i\epsilon)\tau\,, \tag{9.44}$$

so that

$$\tau=\frac{2}{2-i\epsilon}\,,\quad R=\frac{i\epsilon}{2-i\epsilon}\,. \tag{9.45}$$

Clearly, this is in accord with our guess.

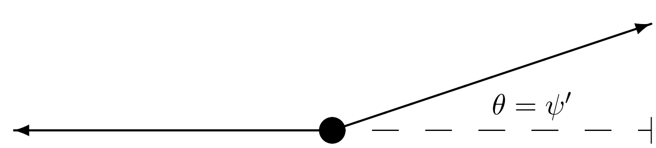


Figure 9.4: The force on the mass.

Note that these amplitudes, unlike those in (9.11), are complex numbers. The transmitted and reflected waves do not have the same phase as the incoming wave at the boundary. The phase difference between the transmitted (or reflected) wave is called a "phase shift." One interesting feature of the solution, (9.45), that we did not guess is that for large ϵ, the small transmitted wave is 90° out of phase with the incoming wave.

This scattering is animated in program 9-2. The solution is also decomposed into incoming, transmitted and reflected waves. Stare at the mass and see if you can understand how the kink in the string is related to its acceleration. You can also make the mass larger and smaller to approach the limits (9.36) and (9.37).

9.2 Index of Refraction

Matter is composed of electric charges. This is something of a miracle. We cannot understand it without quantum mechanics. In a purely classical world, there would be no stable atoms or molecules. Because of quantum mechanics, the world does not collapse and we can build stable chunks of matter composed of equal numbers of positive and negative charges. In a chunk of matter in equilibrium, the charge and current are very close to zero when averaged over any large smooth region. However, in the presence of external electric and magnetic fields, such as those produced by an electromagnetic wave, the charges out of which the matter is built can move. This gives rise to what are called "bound" charges and currents, distinguishable from the "free" charges that are not part of the matter itself. These bound charges and currents affect the relation between electric and magnetic fields. In a homogeneous and isotropic material, which is a fancy way of describing a material that does not have any preferred axis, the effects of the matter (averaged over large regions) can be incorporated by replacing the constants ϵ_0 and μ_0 by the permittivity and permeability, ϵ and μ. Then Maxwell's equations for electromagnetic waves, (8.35)-(8.37), are modified to[1]

$$\begin{gathered}\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\partial B_z}{\partial t}\,, \quad \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t}\,, \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t}\,,\end{gathered} \tag{9.46}$$

$$\begin{gathered}\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = \mu\epsilon\frac{\partial E_z}{\partial t}\,, \quad \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = \mu\epsilon\frac{\partial E_x}{\partial t}\,, \\ \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = \mu\epsilon\frac{\partial E_y}{\partial t}\,,\end{gathered} \tag{9.47}$$

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0\,, \quad \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0\,. \tag{9.48}$$

Now (8.41)-(8.47) are satisfied with the appropriate substitutions,

$$\epsilon_0 \to \epsilon\,, \quad \mu_0 \to \mu\,. \tag{9.49}$$

In particular, the dispersion relation, (8.47), becomes

$$\omega^2 = \frac{1}{\mu\epsilon}\,k^2\,, = \frac{\mu_0\epsilon_0}{\mu\epsilon}\,c^2\,k^2\,. \tag{9.50}$$

[1] See Purcell, chapter 10.

so electromagnetic waves propagate with velocity

$$v = \frac{\omega}{k} = c\sqrt{\frac{\mu_0\epsilon_0}{\mu\epsilon}}\,, \tag{9.51}$$

and (8.48) becomes

$$\beta_y^\pm = \pm\sqrt{\mu\epsilon}\,\varepsilon_x^\pm\,, \quad \beta_x^\pm = \mp\sqrt{\mu\epsilon}\,\varepsilon_y^\pm\,. \tag{9.52}$$

The factor

$$n = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}} \tag{9.53}$$

is called the index of refraction of the material. $1/n$ is the ratio of the speed of light in the material to the speed of light in vacuum. In terms of n, we can write (9.52) as

$$\beta_y^\pm = \pm\frac{n}{c}\varepsilon_x^\pm\,, \quad \beta_x^\pm = \mp\frac{n}{c}\varepsilon_y^\pm\,. \tag{9.54}$$

Note also that we can rewrite (9.50) in the following useful form:

$$k = n\frac{\omega}{c}\,. \tag{9.55}$$

For fixed frequency, the wave number is proportional to the index of refraction. For most transparent materials, μ is very close to 1, and can be ignored. But ϵ can be very different from 1, and is often quite important. For example, the index of refraction of ordinary glass is about 1.5 (it varies slightly with frequency, but we will discuss the interesting and familiar consequences of this later, when we treat waves in three dimensions).

9.2.1 Reflection from a Dielectric Boundary

Let us now consider a plane wave in the $+z$ direction in a universe that is filled with a dielectric material with index of refraction $n = \sqrt{\epsilon/\epsilon_0}$, for $z < 0$ and filled with another dielectric material with index of refraction $n' = \sqrt{\epsilon'/\epsilon_0}$, for $z > 0$. The boundary between the two dielectrics, the plane $z = 0$, is analogous to the boundary between two regions of the rope in figure 9.1. We would, therefore, expect some reflection from this surface.

Because the electric field in a plane electromagnetic wave is perpendicular to its direction of motion, we know that in this case that it is in the x-y plane. It doesn't matter in what direction the electric field of our incoming plane wave is pointing in the x-y plane. That is clear by symmetry. The system looks the same if we rotate it around the z axis, thus we can always rotate until our $\vec{e}_+$ vector is pointing in some convenient direction, say the x direction. It is then pretty obvious that the reflected and transmitted waves will also have their electric fields in the $\pm x$ direction. Actually, we can turn this into a symmetry argument too. If we reflect the system in the x-z plane, both the incoming wave and the dielectric are unchanged, but any y component of the transmitted or reflected waves would change sign. Thus these components must vanish, by symmetry. Magnetic fields work the other way, because of the cross product of vectors in their definition. Thus we can write

$$\left.\begin{aligned} E_x(z,t) &= Ae^{i(kz-\omega t)} + R\,Ae^{i(-kz-\omega t)} \\ B_y(z,t) &= \frac{n}{c}\,Ae^{i(kz-\omega t)} - \frac{n}{c}R\,Ae^{i(-kz-\omega t)} \end{aligned}\right. \quad \text{for } z < 0\,, \tag{9.56}$$

and

$$\left.\begin{aligned} E_x(z,t) &= \tau\,Ae^{i(kz-\omega t)} \\ B_y(z,t) &= \frac{n'}{c}\tau\,Ae^{i(kz-\omega t)} \end{aligned}\right. \quad \text{for } z > 0\,, \tag{9.57}$$

where we have continued our convention of calling the amplitude of the incoming wave A. Here, A has units of electric field. In (9.56) and (9.57), we have used (9.54) to get the B field from the E field.

To compute R and τ, we need the boundary conditions at $z = 0$. For this we go back to Maxwell. The only way to get a discontinuity in the electric field is to have a sheet of charge. In a dielectric, charge builds up on the boundary only if there is a polarization perpendicular to the boundary. In this case, the electric fields, and therefore the polarizations, are parallel to the boundary, and thus the E field is continuous at $z = 0$. The only way to get a discontinuity of the magnetic field, B, is to have a sheet of current. If μ were not equal to 1 in one of the materials, then we would have a nonzero magnetization, and we would have to worry about current sheets at the boundary. However, because these are only dielectrics, and $\mu = 1$ in both, there is no magnetization and the B field is continuous at $z = 0$ as well. Thus we can immediately read off the boundary conditions:

$$1 + R = \tau\,, \quad n(1-R) = n'\tau\,. \tag{9.58}$$

Because of (9.55), the boundary condition (9.58) is equivalent to

$$1 + R = \tau\,, \quad k(1-R) = k'\tau\,, \tag{9.59}$$

which looks exactly like (9.9) and (9.10). We can simply take over the results of (9.11),

$$\tau = \frac{2}{1 + k'/k}\,, \quad R = \frac{1 - k'/k}{1 + k'/k}\,. \tag{9.60}$$

9.3 * Transfer Matrices

9.3.1 Two Masses on a String

Next consider the reflection and transmission from two masses on a string, as in figure 9.5. Now translation invariance and the boundary condition at $x = \infty$ imply that

$$\psi(x,t) = Ae^{ikx}\,e^{-i\omega t} + R\,Ae^{-ikx}\,e^{-i\omega t} \text{ for } x \le 0\,, \tag{9.61}$$

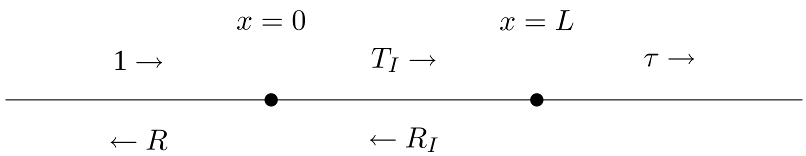


Figure 9.5: Two masses on a string.

$$\psi(x,t) = T_I\, Ae^{ikx}\, e^{-i\omega t} + R_I\, Ae^{-ikx}\, e^{-i\omega t} \text{ for } 0 \leq x \leq L\,, \tag{9.62}$$

$$\psi(x,t) = \tau\, Ae^{ikx}\, e^{-i\omega t} \text{ for } x \geq L\,. \tag{9.63}$$

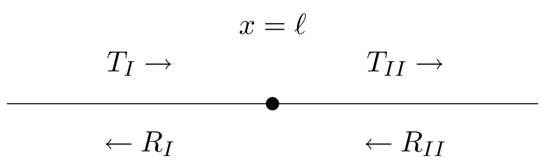


Figure 9.6: The general scattering problem from a mass on a string.

We could solve this problem in the same way, simply imposing boundary conditions twice, at $x = 0$ and at $x = L$, but there is a systematic way of doing this that is very useful. Consider first the general scattering problem from a single mass at $x = \ell$, with both incoming and outgoing waves on both sides, as shown in figure 9.6. This is the most general possible thing that can happen in scattering from a single mass, and we will be able to use the result to do much more complicated problems without any additional work. The general solution has the form

$$\psi(x,t) = T_I\, Ae^{ikx}\, e^{-i\omega t} + R_I\, Ae^{-ikx}\, e^{-i\omega t} \text{ for } x \leq \ell\,, \tag{9.64}$$

$$\psi(x,t) = T_{II}\, Ae^{ikx}\, e^{-i\omega t} + R_{II}\, Ae^{-ikx}\, e^{-i\omega t} \text{ for } x \geq \ell\,. \tag{9.65}$$

The boundary conditions are continuity —

$$T_I\, e^{ik\ell} + R_I\, e^{-ik\ell} = T_{II}\, e^{ik\ell} + R_{II}\, e^{-ik\ell} \tag{9.66}$$

and $F = ma$ —

$$\begin{gathered} T\left(\frac{\partial}{\partial x}\ \psi(x,t)|_{x=\ell^+} - \frac{\partial}{\partial x}\ \psi(x,t)|_{x=\ell^-}\right) \\ = m\,\frac{\partial^2}{\partial t^2}\,\psi(\ell,t) \end{gathered} \tag{9.67}$$

or

$$\begin{gathered} ik\,T\,\left((T_{II} - T_I)\, e^{ik\ell} + (R_I - R_{II})\, e^{-ik\ell}\right) \\ = -m\omega^2\,(T_{II}\, e^{ik\ell} + R_{II}\, e^{-ik\ell})\,. \end{gathered} \tag{9.68}$$

Solving for T_I and R_I gives

$$\begin{aligned} T_I &= \frac{1}{2}\left[(2-i\epsilon)\,T_{II} - i\epsilon\,R_{II}\,e^{-2ik\ell}\right] , \\ R_I &= \frac{1}{2}\left[(2+i\epsilon)\,R_{II} + i\epsilon\,T_{II}\,e^{2ik\ell}\right] . \end{aligned} \tag{9.69}$$

The important point is that because of linearity, the result (9.69) can be written in matrix form:

$$\begin{pmatrix} T_I \\ R_I \end{pmatrix} = d(\ell) \begin{pmatrix} T_{II} \\ R_{II} \end{pmatrix} \tag{9.70}$$

where the matrix $d(\ell)$

$$d(\ell) = \frac{1}{2}\begin{pmatrix} (2-i\epsilon) & -i\epsilon\,e^{-2ik\ell} \\ i\epsilon\,e^{2ik\ell} & (2+i\epsilon) \end{pmatrix} . \tag{9.71}$$

The matrix, $d(\ell)$, is a "transfer matrix." It allows us to get from the amplitudes in one region to those in the next by just doing a matrix multiplication. We can use this to solve the two mass problem without any further calculation except a matrix multiplication. Comparing the general result, (9.70), with the two mass problem, figure 9.5, we see immediately that

$$\begin{pmatrix} 1 \\ R \end{pmatrix} = d(0) \begin{pmatrix} T_I \\ R_I \end{pmatrix} , \tag{9.72}$$

and

$$\begin{pmatrix} T_I \\ R_I \end{pmatrix} = d(L) \begin{pmatrix} \tau \\ 0 \end{pmatrix} . \tag{9.73}$$

Thus

$$\begin{pmatrix} 1 \\ R \end{pmatrix} = d(0)\,d(L) \begin{pmatrix} \tau \\ 0 \end{pmatrix} . \tag{9.74}$$

Doing the matrix multiplication,

$$\begin{aligned} & d(0)\,d(L) = \frac{1}{4} \\ & \begin{pmatrix} (2-i\epsilon)^2 + \epsilon^2 e^{2ikL} & -i\epsilon\left((2-i\epsilon)e^{-2ikL} + (2+i\epsilon)\right) \\ i\epsilon\left((2-i\epsilon) + (2+i\epsilon)e^{2ikL}\right) & (2+i\epsilon)^2 + \epsilon^2 e^{-2ikL} \end{pmatrix} . \end{aligned} \tag{9.75}$$

So

$$\begin{aligned} \tau &= \frac{4}{(2-i\epsilon)^2 + \epsilon^2 e^{2ikL}} , \\ R &= i\epsilon\left((2-i\epsilon) + (2+i\epsilon)e^{2ikL}\right)\,\frac{\tau}{4} . \end{aligned} \tag{9.76}$$

Note that the reflection and transmission shows interesting resonance structure. For example, the reflection vanishes for

$$e^{2ikL} = -\frac{2 - i\epsilon}{2 + i\epsilon}\,. \tag{9.77}$$

In figure 9.7, $|\tau|$ and $|R|$ are plotted versus ϵ for $kL = 0.5$.

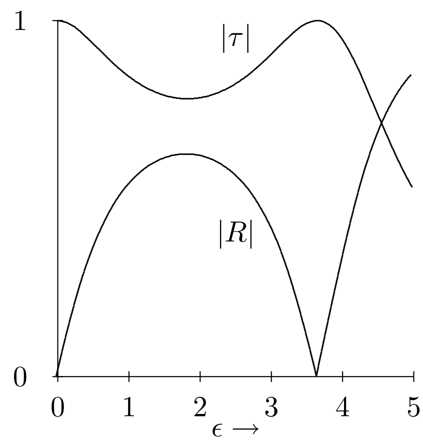


Figure 9.7: $|\tau|$ and $|R|$ plotted versus ϵ for two masses on a string.

9.3.2 *k* Changes

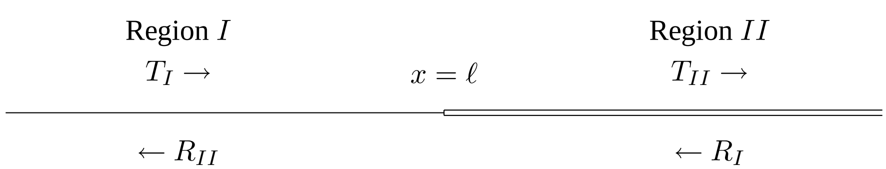


Figure 9.8: The general scattering problem for a change of k.

Let us return to the simple example at the beginning of the chapter of a boundary between two regions of string with different values of k. This is a very important example because its general features are characteristic of many important physical systems. For example, when a light-wave encounters a transparent medium, the k value changes. That situation is somewhat more complicated because of the three-dimensional nature of light waves and because of polarization. However the analogy between (9.59) and (9.9) and (9.10) means that we can take over the discussion of of the string directly to electromagnetic waves reflecting from a dielectric boundary perpendicular to the direction of the wave. In this section, we apply the general method of transfer matrices discussed in the previous section to this important example. Thus we consider the situation shown in figure 9.8. where the waves have the form

$$\psi(x,t) = Ae^{-i\omega t}\left(T_I\,e^{ik_1x} + R_I\,e^{-ik_1x}\right) \text{ in } I\,, \tag{9.78}$$

$$\psi(x,t) = Ae^{-i\omega t}\left(T_{II}\,e^{ik_2x} + R_{II}\,e^{-ik_2x}\right) \text{ in } II\,. \tag{9.79}$$

Then as in (9.9) and (9.10), the boundary conditions are that ψ is continuous at $x = \ell$, which implies

$$T_I\,e^{ik_1\ell} + R_I\,e^{-ik_1\ell} = T_{II}\,e^{ik_2\ell} + R_{II}\,e^{-ik_2\ell}\,, \tag{9.80}$$

and that the slope, $\partial\psi/\partial x$ is continuous at $x = \ell$, which implies

$$ik_1\,\left(T_I\,e^{ik_1\ell} - R_I\,e^{-ik_1\ell}\right) = ik_2\,\left(T_{II}\,e^{ik_2\ell} - R_{II}\,e^{-ik_2\ell}\right)\,. \tag{9.81}$$

Solving the simultaneous linear equations, (9.80) and (9.81), for T_I and R_I and expressing the result in matrix form, we find

$$\begin{pmatrix} T_I \\ R_I \end{pmatrix} = d(k_1,k_2,\ell)\,\begin{pmatrix} T_{II} \\ R_{II} \end{pmatrix}\,, \tag{9.82}$$

where

$$d(k_1,k_2,\ell) = \frac{1}{2}\begin{pmatrix} \left(1+\frac{k_2}{k_1}\right)e^{ik_2\ell-ik_1\ell} & \left(1-\frac{k_2}{k_1}\right)e^{-ik_2\ell-ik_1\ell} \\ \left(1-\frac{k_2}{k_1}\right)e^{ik_2\ell+ik_1\ell} & \left(1+\frac{k_2}{k_1}\right)e^{-ik_2\ell+ik_1\ell} \end{pmatrix}\,. \tag{9.83}$$

(9.82) is a very general result because k_1, k_2 and ℓ can be anything. Note that the relation is symmetrical:

$$\begin{pmatrix} T_{II} \\ R_{II} \end{pmatrix} = d(k_2,k_1,\ell)\,\begin{pmatrix} T_I \\ R_I \end{pmatrix}\,. \tag{9.84}$$

In matrix language, that implies that

$$d(k_2,k_1,\ell)\,d(k_1,k_2,\ell) = I\,. \tag{9.85}$$

It is also useful to use the properties of matrix multiplication to rewrite (9.83) in the following form:

$$d(k_1,k_2,\ell) = b(k_1,\ell)^{-1}\,\tau(k_1,k_2)\,b(k_2,\ell)\,, \tag{9.86}$$

where

$$b(k,\ell) = \begin{pmatrix} e^{ik\ell} & 0 \\ 0 & e^{-ik\ell} \end{pmatrix}\,, \tag{9.87}$$

and

$$\tau(k_1, k_2) = d(k_1, k_2, 0) = \frac{1}{2} \begin{pmatrix} \left(1 + \frac{k_2}{k_1}\right) & \left(1 - \frac{k_2}{k_1}\right) \\ \left(1 - \frac{k_2}{k_1}\right) & \left(1 + \frac{k_2}{k_1}\right) \end{pmatrix} . \tag{9.88}$$

You will see the utility of this in the computer problem, (9.6).

9.3.3 Reflection from a Thin Film

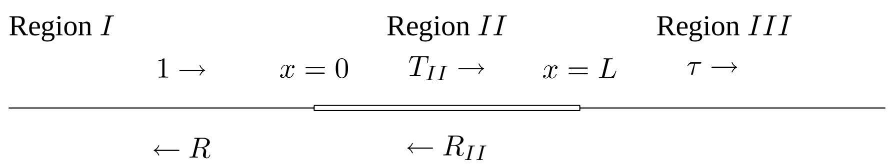


Figure 9.9: Reflection from a thin film.

Consider the situation shown in figure 9.9. where the wave numbers are k_1 for $x \le 0$, k_2 for $0 \le x \le L$ and k_3 for $x \ge L$. As usual, translation invariance plus the boundary condition at infinity (that the incoming wave in I has amplitude, A, and that there is only an outgoing wave in III) implies

$$\begin{aligned} \psi(x,t) &= Ae^{-i\omega t} \left(e^{ik_1 x} + R\, e^{-ik_1 x}\right) \quad \text{for } x \le 0\,, \\ \psi(x,t) &= Ae^{-i\omega t} \left(T_{II}\, e^{ik_2 x} + R_{II}\, e^{-ik_2 x}\right) \quad \text{for } 0 \le x \le L\,, \\ \psi(x,t) &= \tau\, Ae^{-i\omega t} e^{ik_3 x} \quad \text{for } L \le x\,. \end{aligned} \tag{9.89}$$

Then we know from the results of the previous section that

$$\begin{pmatrix} 1 \\ R \end{pmatrix} = d(k_1, k_2, 0) \begin{pmatrix} T_{II} \\ R_{II} \end{pmatrix} \tag{9.90}$$

and

$$\begin{pmatrix} T_{II} \\ R_{II} \end{pmatrix} = d(k_2, k_3, L) \begin{pmatrix} \tau \\ 0 \end{pmatrix} \tag{9.91}$$

and therefore

$$\begin{pmatrix} 1 \\ R \end{pmatrix} = d(k_1, k_2, 0)\, d(k_2, k_3, L) \begin{pmatrix} \tau \\ 0 \end{pmatrix} . \tag{9.92}$$

$$\begin{aligned} &d(k_1, k_2, 0)\, d(k_2, k_3, L) \\ &= b(k_1, 0)^{-1}\, \tau(k_1, k_2)\, b(k_2, 0)\, b(k_2, L)^{-1}\, \tau(k_2, k_3)\, b(k_3, L) \end{aligned} \tag{9.93}$$

Often we are interested in the situation $k_3 = k_1$, that describes a film (in one-dimension, a film is just a region in x) in an otherwise homogeneous medium. This is then a one-dimensional analog of the reflection of light from a soap bubble. Then the transfer matrix looks like

$$\frac{1}{4}\begin{pmatrix} \left(1+\frac{k_2}{k_1}\right) & \left(1-\frac{k_2}{k_1}\right) \\ \left(1-\frac{k_2}{k_1}\right) & \left(1+\frac{k_2}{k_1}\right) \end{pmatrix} \begin{pmatrix} e^{-ik_2L} & 0 \\ 0 & e^{ik_2L} \end{pmatrix} \begin{pmatrix} \left(1+\frac{k_1}{k_2}\right) & \left(1-\frac{k_1}{k_2}\right) \\ \left(1-\frac{k_1}{k_2}\right) & \left(1+\frac{k_1}{k_2}\right) \end{pmatrix} \begin{pmatrix} e^{ik_1L} & 0 \\ 0 & e^{-ik_1L} \end{pmatrix} \tag{9.94}$$

Thus

$$1 = \left(\cos k_2L - i\,\frac{k_1^2+k_2^2}{2k_1k_2}\sin k_2L\right)\, e^{ik_1L}\,\tau \tag{9.95}$$

and

$$R = -\left(i\,\frac{k_1^2-k_2^2}{2k_1k_2}\sin k_2L\right)\, e^{ik_1L}\,\tau \tag{9.96}$$

or

$$\tau = \left(\cos k_2L - i\,\frac{k_1^2+k_2^2}{2k_1k_2}\sin k_2L\right)^{-1} e^{-ik_1L} \tag{9.97}$$

and

$$R = -\left(i\,\frac{k_1^2-k_2^2}{2k_1k_2}\sin k_2L\right)\left(\cos k_2L - i\,\frac{k_1^2+k_2^2}{2k_1k_2}\sin k_2L\right)^{-1}. \tag{9.98}$$

Here we see the phenomenon of **resonant transmission**. The wave does not get reflected at all if the thickness of the film is an integral or half-integral number of wavelengths. Note, also, that when $k_2 \to k_1$, $\tau \to 1$ and $R \to 0$ as they should, because in this limit there is no boundary.

The reflection in (9.98) varies rapidly with k_2, as shown figure 9.10, where we plot the intensity of the reflected wave versus k_2 for fixed ratio $k_1/k_2 = 3$. It is this rapid variation of the intensity of reflected light as a function of wavelength that is responsible for the familiar color patterns on thin films like soap bubbles and oil slicks.

9.3.4 Nonreflective Coating

We will not work out the general case of $k_1 \neq k_3$, simply because the algebra is a mess. However, one important special case is worth noting. Suppose that you have a boundary between media in which the wave number of your traveling wave are k_1 and k_3. Normally, you find reflection at the boundary. The question is, can you add an intermediate film layer

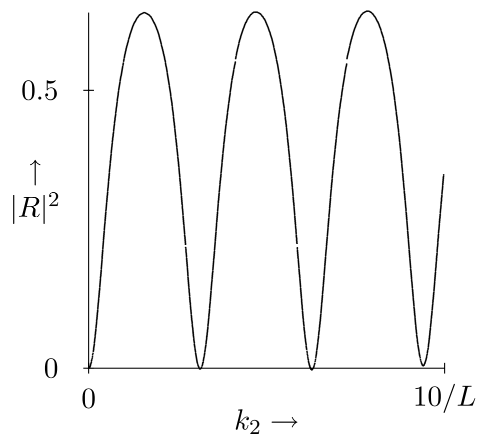


Figure 9.10: Graph of $|R|^2$ versus k_2 for $k_1/k_2 = 3$.

with wave number k_2, that eliminates all reflection? The answer is yes. First you must adjust the wave number in the film to be the geometric mean of k_1 and k_3, so that

$$\frac{k_2}{k_1} = \frac{k_3}{k_2} \,. \tag{9.99}$$

Then the transfer matrix becomes

$$\begin{aligned}\frac{1}{4}&\begin{pmatrix} \left(1+\frac{k_2}{k_1}\right) & \left(1-\frac{k_2}{k_1}\right) \\ \left(1-\frac{k_2}{k_1}\right) & \left(1+\frac{k_2}{k_1}\right) \end{pmatrix} \begin{pmatrix} e^{-ik_2L} & 0 \\ 0 & e^{ik_2L} \end{pmatrix} \\ &\begin{pmatrix} \left(1+\frac{k_2}{k_1}\right) & \left(1-\frac{k_2}{k_1}\right) \\ \left(1-\frac{k_2}{k_1}\right) & \left(1+\frac{k_2}{k_1}\right) \end{pmatrix} \begin{pmatrix} e^{ik_3L} & 0 \\ 0 & e^{-ik_3L} \end{pmatrix} \,.\end{aligned} \tag{9.100}$$

It is easy to check that the reflection vanishes when there are a half-odd-integral number of wavelengths in the intermediate region,

$$k_2L = (2n+1)\frac{\pi}{2} \,. \tag{9.101}$$

In qualitative terms, the reflection vanishes because of a destructive interference between the reflected waves from the two boundaries. This has practical applications to nonreflective coatings for optical components.

Chapter Checklist

You should now be able to:

i. Analyze scattering problems by imposing boundary conditions and computing reflection and transmission coefficients;

ii. Identify a wave with some reflection, and differentiate it from a pure traveling or standing wave;

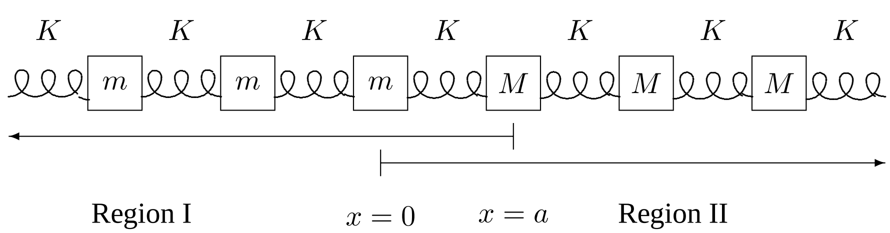
iii. Check energy conservation in scattering problems;

iv. Analyze electromagnetic plane waves in a dielectric, and the reflection from a dielectric boundary;

v. ***** Use transfer matrices to simplify the analysis of scattering from more than one boundary.

Problems

9.1.



Shown above is the boundary between two semi-infinite systems. To the left, there are identical blocks of mass m. To the right, there are identical blocks of mass M. They are connected as shown by identical massless springs with spring constant K, such that the equilibrium separation between neighboring blocks is a. Consider the reflection of a traveling longitudinal wave from the boundary between these two regions. That is, assume that in region I there is an incident wave of amplitude A traveling to the right and a reflected wave traveling to the left. In a complex notation, the displacement of the mass with equilibrium position x is

$$\psi(x,t) = Ae^{-i(\omega t - kx)} + R\,Ae^{-i(\omega t + kx)}$$

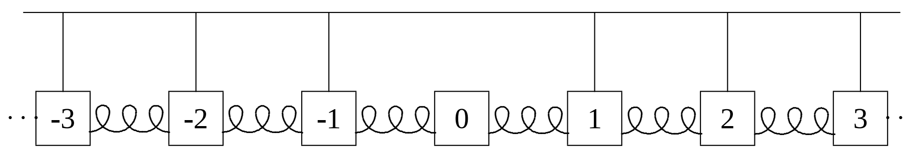
for $x \leq a$. What is the relation between ω and k?

In region II, there is only a transmitted wave:

$$\psi(x,t) = T\,Ae^{-i(\omega t - k'x)}$$

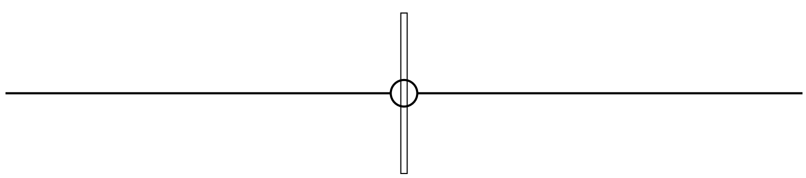
for $x \geq 0$. What is the relation between ω and k'? Find the appropriate boundary conditions that allow you to relate $\psi(x,t)$ in the two regions and solve for R (do not bother to simplify the complex number). Check your result by taking the limit of a, m and M going to zero with m/a and M/a fixed and comparing with an appropriate continuous system.

9.2. An infinite line of coupled pendulums supports traveling waves, but it has no standing wave normal modes in which the displacement of the pendulums goes to zero at infinity. Consider, however, the system shown below:



Here block 0 is free to slide longitudinally with no gravitational restoring force, only the coupling due to the springs. If the blocks have mass M, the springs' spring constant K, the separation between neighboring blocks is a, and the pendulums have length ℓ, find the frequency of the standing wave normal mode of the system in which the displacements are $A\,e^{-\kappa x}$ for $x \geq 0$ and $A\,e^{\kappa x}$ for $x \leq 0$. **Hint:** Consider the subsystem, $-a \leq x \leq a$, as part of an infinite system with appropriate boundary conditions. Then you can get the answer directly from the dispersion relation.

9.3.



Consider a string with linear mass density ρ, split into two pieces. The two halves are attached to a massless ring which slides vertically without friction on a rod at $x = 0$. One of the two halves is stretched in the negative x direction with tension T. The other is stretched in the positive x direction with tension T'. Note that the vertical rod is necessary to balance the horizontal forces on the massless ring from the two strings with different tensions.

Suppose that a traveling wave comes in from the negative x direction. Then the displacement of the strings in the two regions is

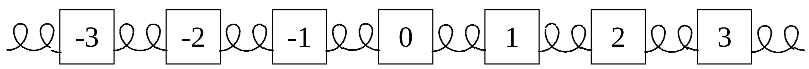
$$\psi(x,t) = Ae^{ikx}\,e^{-i\omega t} + R\,Ae^{-ik'x}\,e^{-i\omega' t} \text{ for } x \leq 0$$

$$\psi(x,t) = \tau\,Ae^{ik''x}\,e^{-i\omega'' t} \text{ for } x \geq 0\,.$$

a. Find k, k', ω', k'' and ω'' in terms of ω, T, T' and ρ. **Hint – this is easy!**

b. Write down the two boundary conditions at $x = 0$ and find R and τ.

9.4. Consider traveling waves in an infinite system, part of which is shown below, for longitudinal (horizontal) motion of the blocks.



All the blocks have mass m, except for block 0 **which is massless**. The springs are massless and have spring constant K. The separation between neighboring blocks is a. To the left of block 0, which we will take to be at $x = 0$, there is an incoming and a reflected wave, so that the longitudinal displacement of the blocks for $x \leq 0$ has the form

$$Ae^{ikx-i\omega t} + R\,Ae^{-ikx-i\omega t}\,.$$

To the right of the massless block, there is a transmitted wave, so that the longitudinal displacement of the blocks for $x \geq 0$ has the form

$$T\,Ae^{ikx-i\omega t}\,.$$

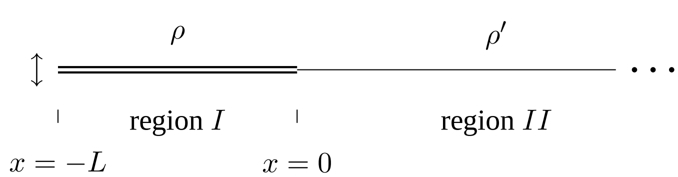
ω and k are related by the dispersion relation

$$\omega^2 = \frac{4K}{m}\sin^2\frac{ka}{2}\,.$$

a. Explain the physics of the boundary conditions at $x = 0$.

b. Find R and T.

9.5. Consider a semi-infinite system of two kinds of massive string with different densities, shown below:



The density of the string in region I is ρ and in region II is ρ'. The tension in both strings is T. Suppose that the end at $x = -L$ is oscillated in the transverse direction with displacement

$\chi \sin \omega t$. This produces an outgoing wave (moving to the right) in region II with no incoming wave. Suppose that $\omega = \frac{\pi}{2L}\sqrt{\frac{T}{\rho}}$. Find the displacement at the point $x = 0$ as a function of time.

9.6. If you are doing a reflection and transmission problem involving several different regions, and thus requiring several boundary conditions, the transfer matrix is very helpful. You saw this in the analysis of scattering from a thin film.

Your computer assignment is to extend this analysis to incorporate $2n$ such boundary conditions where n is some large integer. In particular, consider a continuous string with wave number k_2 for $L \leq x \leq 2L$, $3L \leq x \leq 4L$, $\cdots$, and $(2n-1)L \leq x \leq 2nL$, and k_1 elsewhere.

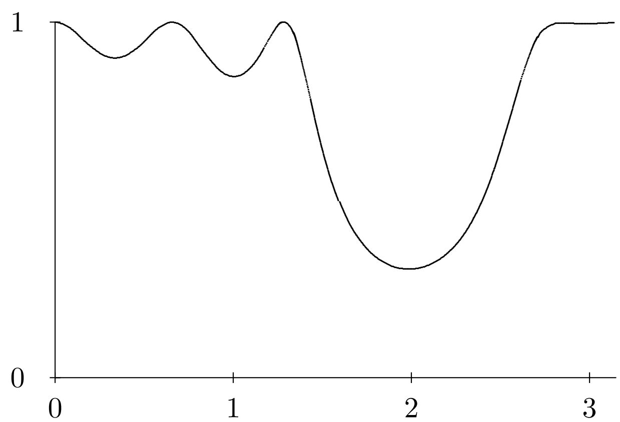


Figure 9.11: $n = 3$.

Take $k_1 = k$ and $k_2 = 2k$. Compute the amplitude for transmission of an incoming wave in this system as a function of L by doing the appropriate multiplication of $2n$ matrices. To do this, you must program your computer to multiply complex matrices. Organize your program in an iterative way, so that you can change n easily. This will allow you to start out with small n and go to larger n only when you are sure that the program is working.

If possible, you should present the results in the form of a graph of the absolute value of the transmission coefficient versus kL, for $0 \leq L \leq \pi/2k$. As you go to higher n, something interesting happens. The transmission coefficient drops nearly to zero in a region of L values. Even if you cannot produce a graph, you should be able to find the range of L for which the transmission goes to zero as n gets large.

Hint: For $n = 3$, the result should look like the graph in figure 9.11.

Chapter 10

Signals and Fourier Analysis

Traveling waves with a definite frequency carry energy but no information. They are just there, always have been and always will be. To send information, we must send a nonharmonic signal.

Preview

In this chapter, we will see how this works in the context of a forced oscillation problem. In the process, we will find a subtlety in the notion of the speed with which a traveling wave moves. The phase velocity may not be the same as the velocity of signal propagation.

i. We begin by studying the propagation of a transverse pulse on a stretched string. We solve the problem in two ways: with a trick that works in this special case; and with the more powerful technique of Fourier transformation. We introduce the concept of "group velocity," the speed at which signals can actually be sent in a real system.

ii. We discuss, by example and then in general, the counterpoint between a function and its Fourier transform. We make the connection to the physical concepts of bandwidth and fidelity in signal transmission and to Heisenberg's uncertainty relation in quantum mechanics.

iii. We work out in some detail an example of the scattering of a wave packet.

iv. We discuss the dispersion relation for electromagnetic waves in more detail and explore the question of whether light actually travels at the speed of light!

10.1 Signals in Forced Oscillation

10.1.1 A Pulse on a String

10-1

We begin with the following illustrative problem: the transverse oscillations of a semiinfinite string stretched from $x = 0$ to ∞, driven at $x = 0$ with some arbitrary transverse signal $f(t)$, and with a boundary condition at infinity that there are no incoming traveling waves. This simple system is shown in figure 10.1.

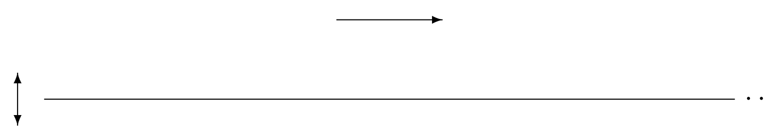


Figure 10.1: A semiinfinite string.

There is a slick way to get the answer to this problem that works **only** for a system with the simple dispersion relation,

$$\omega^2 = v^2\, k^2 \,. \tag{10.1}$$

The trick is to note that the dispersion relation, (10.1), implies that the system satisfies the wave equation, (6.4), or

$$\frac{\partial^2}{\partial t^2}\psi(x,t) = v^2\, \frac{\partial^2}{\partial x^2}\psi(x,t)\,. \tag{10.2}$$

It is a mathematical fact (we will discuss the physics of it below) that the general solution to the one-dimensional wave equation, (10.2), is a sum of right-moving and left-moving waves with arbitrary shapes,

$$\psi(x,t) = g(x - vt) + h(x + vt)\,, \tag{10.3}$$

where g and h are arbitrary functions. You can check, using the chain rule, that (10.3) satisfies (10.2),

$$\begin{aligned}\frac{\partial^2}{\partial t^2}(g(x-vt) + h(x+vt)) &= v^2\frac{\partial^2}{\partial x^2}(g(x-vt) + h(x+vt)) \\ &= v^2\,(g''(x-vt) + h''(x+vt))\,.\end{aligned} \tag{10.4}$$

Given this mathematical fact, we can find the functions g and h that solve our particular problem by imposing boundary conditions. The boundary condition at infinity implies

$$h = 0\,, \tag{10.5}$$

because the h function describes a wave moving in the $-x$ direction. The boundary condition at $x = 0$ implies

$$g(-vt) = f(t)\,, \tag{10.6}$$

which gives

$$\psi(x,t) = f(t - x/v)\,. \tag{10.7}$$

This describes the signal, $f(t)$, propagating down the string at the phase velocity v with no change in shape.

For the simple function

$$f(t) = \begin{cases} 1 - |t| & \text{for } |t| \le 1 \\ 0 & \text{for } |t| > 1 \end{cases} \tag{10.8}$$

the shape of the string at a sequence of times is shown in figure 10.2 and animated in program 10-1.

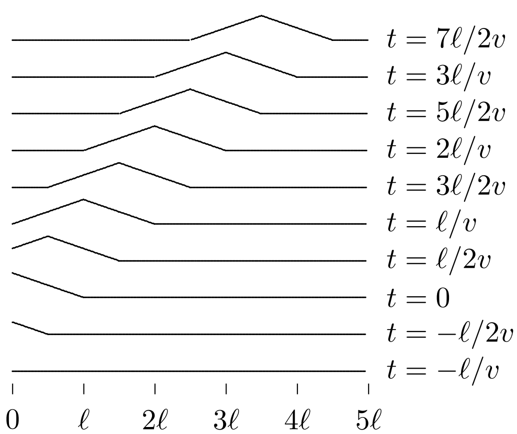


Figure 10.2: A triangular pulse propagating on a stretched string.

10.1.2 Fourier integrals

Let us think about this problem in a more physical way. In the process, we will understand the physics of the general solution, (10.3). This may seem like a strange thing to say in a section entitled, "Fourier integrals." Nevertheless, we will see that the mathematics of Fourier integrals has a direct and simple physical interpretation.

The idea is to use linearity in a clever way to solve this problem. We can take $f(t)$ apart into its component angular frequencies. We already know how to solve the forced oscillation problem for each angular frequency. We can then take the individual solutions and add them back up again to reconstruct the solution to the full problem. The advantage of this procedure is that it works for any dispersion relation, not just for (10.1).

Because there may be a continuous distribution of frequencies in an arbitrary signal, we cannot just write $f(t)$ as a sum over components, we need a Fourier integral,

$$f(t) = \int_{-\infty}^{\infty} d\omega \, C(\omega) \, e^{-i\omega t} \,. \tag{10.9}$$

The physics of (10.9) is just linearity and time translation invariance. We know that we can choose the normal modes of the free system to have irreducible exponential time dependence, because of time translation invariance. Since the normal modes describe all the possible motions of the system, we know that by taking a suitable linear combination of normal modes, we can find a solution in which the motion of the end of the system is described by the function, $f(t)$. The only subtlety in (10.9) is that we have assumed that the values of ω that appear in the integral are all real. This is appropriate because a nonzero imaginary part for ω in $e^{-i\omega t}$ describes a function that goes exponentially to infinity as $t \to \pm\infty$. Physically, we are never interested in such things. In fact, we are really interested in functions that go to zero as $t \to \pm\infty$. These are well-described by the integral over real ω, (10.9).

Note that if $f(t)$ is real in (10.9), then

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} d\omega \, C(\omega) \, e^{-i\omega t} \\ = f(t)^* &= \int_{-\infty}^{\infty} d\omega \, C(\omega)^* \, e^{i\omega t} = \int_{-\infty}^{\infty} d\omega \, C(-\omega)^* \, e^{-i\omega t} \end{aligned} \tag{10.10}$$

thus

$$C(-\omega)^* = C(\omega) \,. \tag{10.11}$$

It is actually easier to work with the complex Fourier integral, (10.9), with the irreducible complex exponential time dependence, than with real expansions in terms of $\cos\omega t$ and $\sin\omega t$. But you may also see the real forms in other books. You can always translate from (10.9) by using the Euler identity

$$e^{i\theta} = \cos\theta + i\sin\theta \,. \tag{10.12}$$

For each value of ω, we can write down the solution to the forced oscillation problem, incorporating the boundary condition at ∞. Each frequency component of the force produces a wave traveling in the $+x$ direction.

$$e^{-i\omega t} \to e^{-i\omega t + ikx} \,, \tag{10.13}$$

then we can use linearity to construct the solution by adding up the individual traveling waves from (10.13) with the coefficients $C(\omega)$ from (10.9). Thus

$$\psi(x,t) = \int_{-\infty}^{\infty} d\omega \, C(\omega) \, e^{-i\omega t + ikx} \,. \tag{10.14}$$

where ω and k are related by the dispersion relation.

Equation (10.14) is true quite generally for any one-dimensional system, **for any dispersion relation**, but the result is particularly simple for a nondispersive system such as the continuous string with a dispersion relation of the form (10.1). We can use (10.1) in (10.14) by replacing

$$k \to \omega/v\,. \tag{10.15}$$

Note that while k^2 is determined by the dispersion relation, the sign of k, for a given ω, is determined by the boundary condition at infinity. k and ω must have the same sign, as in (10.15), to describe a wave traveling in the $+x$ direction. Putting (10.15) into (10.14) gives

$$\psi(x,t) = \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i\omega t + i\omega x/v} = \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i\omega(t-x/v)}\,. \tag{10.16}$$

Comparing this with (10.9) gives (10.7).

Let us try to understand what is happening in words. The Fourier integral, (10.9), expresses the signal as a linear combination of harmonic traveling waves. The relation, (10.15), which follows from the dispersion relation, (10.1), and the boundary condition at ∞, implies that each of the infinite harmonic traveling waves moves at the same phase velocity. Therefore, the waves stay in exactly the same relationship to one another as they move, and the signal is never distorted. It just moves with the waves.

The nonharmonic signal is called a "wave packet." As we have seen, it can be taken apart into harmonic waves, by means of the Fourier integral, (10.9).

10.2 Dispersive Media and Group Velocity

For any other dispersion relation, the signal changes shape as it propagates, because the various harmonic components travel at different velocities. Eventually, the various pieces of the signal get out of phase and the signal is dispersed. That is why such a medium is called "dispersive." This is the origin of the name "dispersion relation."

10.2.1 Group Velocity

10-2

If you are clever, you can send signals in a dispersive medium. The trick is to send the signal not directly as the function, $f(t)$, but as a modulation of a harmonic signal, of the form

$$f(t) = f_s(t)\cos\omega_0 t\,, \tag{10.17}$$

where $f_s(t)$ is the signal. Very often, you want to do this anyway, because the important frequencies in your signal may not match the frequencies of the waves with which you want to send the signal. An example is AM radio transmission, in which the signal is derived from sound with a typical frequency of a few hundred cycles per second (Hz), but it is carried as a modulation of the amplitude of an electromagnetic radio wave, with a frequency of a few million cycles per second.[1]

You can get a sense of what is going to happen in this case by considering the sum of two traveling waves with different frequencies and wave numbers,

$$\cos(k_+ x - \omega_+ t) + \cos(k_- x - \omega_- t) \tag{10.18}$$

where

$$k_\pm = k_0 \pm k_s\,, \quad \omega_\pm = \omega_0 \pm \omega_s\,, \tag{10.19}$$

for

$$k_s \ll k_0\,, \quad \omega_s \ll \omega_0\,. \tag{10.20}$$

The sum can be written as a product of cosines, as

$$2\cos(k_s x - \omega_s t) \cdot \cos(k_0 x - \omega_0 t)\,. \tag{10.21}$$

Because of (10.20), the first factor varies slowly in x and t compared to the second. The result can be thought of as a harmonic wave with frequency ω_0 with a slowly varying amplitude proportional to the first factor. The space dependence of (10.21) is shown in figure 10.3.

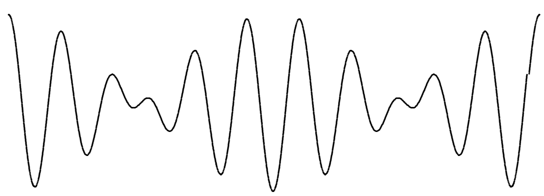


Figure 10.3: The function (10.21) for $t = 0$ and $k_0/ks = 10$.

You should think of the first factor in (10.21) as the signal. The second factor is called the "carrier wave." Then (10.21) describes a signal that moves with velocity

$$v_s = \frac{\omega_s}{k_s} = \frac{\omega_+ - \omega_-}{k_+ - k_-}\,, \tag{10.22}$$

while the smaller waves associated with the second factor move with velocity

$$v_0 = \frac{\omega_0}{k_0}\,. \tag{10.23}$$

[1] See (10.71), below.

These two velocities will not be the same, in general. If (10.20) is satisfied, then (as we will show in more detail below) v_0 will be roughly the phase velocity. In the limit, as $k_+ - k_- = 2k_s$ becomes very small, (10.22) becomes a derivative

$$v_s = \frac{\omega_+ - \omega_-}{k_+ - k_-} \to \left.\frac{\partial\omega}{\partial k}\right|_{k=k_0} . \tag{10.24}$$

This is called the "group velocity." It measures the speed at which the signal can actually be sent.

The time dependence of (10.21) is animated in program 10-2. Note the way that the carrier waves move through the signal. In this animation, the group velocity is smaller than the phase velocity, so the carrier waves appear at the back of each pulse of the signal and move through to the front.

Let us see how this works in general for interesting signals, $f(t)$. Suppose that for some range of frequencies near some frequency ω_0, the dispersion relation is slowly varying. Then we can take it to be approximately linear by expanding $\omega(k)$ in a Taylor series about k_0 and keeping only the first two terms. That is

$$\omega = \omega(k) = \omega_0 + (k - k_0) \left.\frac{\partial\omega}{\partial k}\right|_{k=k_0} + \cdots , \tag{10.25}$$

$$\omega_0 \equiv \omega(k_0) , \tag{10.26}$$

and the higher order terms are negligible for a range of frequencies

$$\omega_0 - \Delta\omega < \omega < \omega_0 + \Delta\omega . \tag{10.27}$$

where $\Delta\omega$ is a constant that depends on ω_0 and the details on the higher order terms. Then you can send a signal of the form

$$f(t) \cdot e^{-i\omega_0 t} \tag{10.28}$$

(a complex form of (10.17), above) where $f(t)$ satisfies (10.9) with

$$C(\omega) \approx 0 \;\text{ for }\; |\omega - \omega_0| > \Delta\omega . \tag{10.29}$$

This describes a signal that has a carrier wave with frequency ω_0, modulated by the interesting part of the signal, $f(t)$, that acts like a time-varying amplitude for the carrier wave, $e^{-i\omega_0 t}$. The strategy of sending a signal as a varying amplitude on a carrier wave is called amplitude modulation.

Usually, the higher order terms in (10.25) are negligible only if $\Delta\omega \ll \omega_0$. If we neglect them, we can write (10.25) as

$$\omega = vk + a , \quad k = \omega/v + b , \tag{10.30}$$

where a and b are constants we can determine from (10.25),

$$a = \omega_0 - v k_0 \,, \qquad b = k_0 - \omega_0 / v \tag{10.31}$$

and v is the group velocity

$$v = \left.\frac{\partial \omega}{\partial k}\right|_{k=k_0} \,. \tag{10.32}$$

For the signal (10.28)

$$\psi(0,t) = \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i(\omega+\omega_0)t} = \int_{-\infty}^{\infty} d\omega\, C(\omega-\omega_0)\, e^{-i\omega t} \,. \tag{10.33}$$

Thus (10.14) becomes

$$\psi(x,t) = \int_{-\infty}^{\infty} d\omega\, C(\omega-\omega_0)\, e^{-i\omega t}\, e^{ikx} \,, \tag{10.34}$$

but then (10.29) gives

$$\begin{aligned} \psi(x,t) &= \int_{-\infty}^{\infty} d\omega\, C(\omega-\omega_0)\, e^{-i\omega t + i(\omega/v+b)x} \\ &= \int_{-\infty}^{\infty} d\omega\, C(\omega-\omega_0)\, e^{-i\omega(t-x/v)+ibx} \\ &= \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i(\omega+\omega_0)(t-x/v)+ibx} \\ &= f(t-x/v)\, e^{-i\omega_0(t-x/v)+ibx} \,. \end{aligned} \tag{10.35}$$

The modulation $f(t)$ travels without change of shape at the group velocity v given by (10.32), as long as we can ignore the higher order term in the dispersion relation. The phase velocity

$$v_\phi = \frac{\omega}{k} \,, \tag{10.36}$$

has nothing to do with the transmission of information, but notice that because of the extra e^{ibx} in (10.35), the carrier wave travels at the phase velocity.

You can see the difference between phase velocity and group velocity in your pool or bathtub by making a wave packet consisting of several shorter waves.

10.3 Bandwidth, Fidelity, and Uncertainty

The relation (10.9) can be inverted to give $C(\omega)$ in terms of $f(t)$ as follows

$$C(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt\, f(t)\, e^{i\omega t} \,. \tag{10.37}$$

This is the "inverse Fourier transform." It is very important because it allows us to go back and forth between the signal and the distribution of frequencies that it contains. We will get this result in two ways: first, with a fancy argument that we will use again and explain in more detail in chapter 13; next, by going back to the Fourier series, discussed in chapter 6 for waves on a finite string, and taking the limit as the length of the string goes to infinity.

The fancy argument goes like this. It is very reasonable that the integral in (10.37) is proportional to $C(\omega)$ because if we insert (10.9) and rearrange the order of integration, we get

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega'\, C(\omega')\int_{-\infty}^{\infty} dt\, e^{i(\omega-\omega')t}\,. \tag{10.38}$$

The t integral averages to zero unless $\omega = \omega'$. Thus the ω' integral is simply proportional to $C(\omega)$ times a constant factor. The factor of $1/2\pi$ can be obtained by doing some integrals explicitly. For example, if

$$f(t) = e^{-\Gamma|t|}\,, \tag{10.39}$$

for $\Gamma > 0$ then, as we will show explicitly in (10.49)-(10.56), (10.37) yields

$$2\pi C(\omega) = 2\Gamma/(\Gamma^2+\omega^2)\,, \tag{10.40}$$

which can, in turn, be put back in (10.9) to give (10.39). For $t = 0$, the integral can be done by the trigonometric substitution $\omega \to \Gamma\tan\theta$:

$$\begin{aligned}1 = f(0) = e^{-\Gamma\cdot 0} &= \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i\omega\cdot 0}\\ = \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega \frac{\Gamma}{\Gamma^2+\omega^2} &\to \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} d\theta = 1\,.\end{aligned} \tag{10.41}$$

To get the inverse Fourier transform, (10.37), as the limit of a Fourier series, it is convenient to use a slightly different boundary condition from those we discussed in chapter 6, fixed ends and free ends. Instead, let us consider a string stretched from $x = -\pi\ell$ to $x = \pi\ell$, in which we assume that the displacement of the string from equilibrium at $x = \pi\ell$ is the same as the displacement at $x = -\pi\ell$,[2]

$$\psi(-\pi\ell, t) = \psi(\pi\ell, t)\,. \tag{10.42}$$

The requirement, (10.42), is called "periodic boundary conditions," because it implies that the function ψ that describes the displacement of the string is periodic in x with period $2\pi\ell$. The normal modes of the infinite system that satisfy (10.42) are

$$e^{inx/\ell}, \tag{10.43}$$

[2]A example of a physical system with this kind of boundary condition would be a string stretched around a frictionless cylinder with radius ℓ and (therefore) circumference $2\pi\ell$. Then (10.42) would be true because $x = -\pi\ell$ describes the same point on the string as $x = \pi\ell$.

for integer n, because changing x by $2\pi\ell$ in (10.43) just changes the phase of the exponential by 2π. Thus if $\psi(x)$ is an arbitrary function satisfying $\psi(-\pi\ell) = \psi(\pi\ell)$, we should be able to expand it in the normal modes of (10.43),

$$\psi(x) = \sum_{n=-\infty}^{\infty} c_n \, e^{-inx/\ell}. \tag{10.44}$$

Likewise, for a function $f(t)$, satisfying $f(-\pi T) = f(\pi T)$ for some large time T, we expect to be able to expand it as follows

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \, e^{-int/T} \,, \tag{10.45}$$

where we have changed the sign in the exponential to agree with (10.9). We will show that as $T \to \infty$, this becomes equivalent to (10.9).

Equation (10.44) is the analog of (6.8) for the boundary condition, (10.42). The sum runs from $-\infty$ to ∞ rather than 0 to ∞ because the modes in (10.43) are different for n and $-n$. For this Fourier series, the inverse is

$$c_m = \frac{1}{2\pi T} \int_{-\pi T}^{\pi T} dt \, e^{imt/T} \, f(t) \tag{10.46}$$

where we have used the identity

$$\frac{1}{2\pi T} \int_{-\pi T}^{\pi T} dt \, e^{imt/T} \, e^{-int/T} = \begin{cases} 1 \text{ for } m = n \,, \\ 0 \text{ for } m \neq n \,. \end{cases} \tag{10.47}$$

Now suppose that $f(t)$ goes to 0 for large $|t|$ (note that this is consistent with the periodic boundary condition (10.42)) fast enough so that the integral in (10.46) is well defined as $T \to \infty$ for all m. Then because of the factor of $1/T$ in (10.47), the c_n all go to zero like $1/T$. Thus we should multiply c_n by T to get something finite in the limit. Comparing (10.45) with (10.9), we see that we should take ω to be n/T.

Thus the relation, (10.45), is an analog of the Fourier integral, (10.9) where the correspondence is

$$\begin{aligned} T &\to \infty \\ \frac{n}{T} &\to \omega \\ c_n T &\to C(\omega) \,. \end{aligned} \tag{10.48}$$

In the limit, $T \to \infty$, the sum becomes an integral over ω.

Multiplying both sides of (10.46) by T, and making the substitution of (10.48) gives (10.37).

10.3.1 A Solvable Example

For practice in dealing with integration of complex functions, we will do the integration that leads to (10.40) in gory detail, with all the steps.

$$C(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt\, e^{-\Gamma|t|}\, e^{i\omega t}. \tag{10.49}$$

First we get rid of the absolute value —

$$= \frac{1}{2\pi} \int_{0}^{\infty} dt\, e^{-\Gamma t}\, e^{i\omega t} + \frac{1}{2\pi} \int_{-\infty}^{0} dt\, e^{\Gamma t}\, e^{i\omega t} \tag{10.50}$$

and write the second integral as an integral from 0 to ∞ —

$$= \frac{1}{2\pi} \int_{0}^{\infty} dt\, e^{-\Gamma t}\, e^{i\omega t} + \frac{1}{2\pi} \int_{0}^{\infty} dt\, e^{-\Gamma t}\, e^{-i\omega t} \tag{10.51}$$

$$= \frac{1}{2\pi} \int_{0}^{\infty} dt\, e^{-\Gamma t}\, e^{i\omega t} + \text{complex conjugate}, \tag{10.52}$$

but we know how to differentiate even complex exponentials (see the discussion of (3.108)), so we can write

$$\frac{\partial}{\partial t}\left(e^{-\Gamma t}\, e^{i\omega t}\right) = (-\Gamma + i\omega)\, e^{-\Gamma t}\, e^{i\omega t}. \tag{10.53}$$

Thus

$$\int_{0}^{\infty} dt\, e^{-\Gamma t}\, e^{i\omega t} = \frac{1}{-\Gamma + i\omega} \int_{0}^{\infty} dt\, \frac{\partial}{\partial t}\left(e^{-\Gamma t}\, e^{i\omega t}\right) \tag{10.54}$$

or, using the fundamental theorem of integral calculus,

$$= \frac{1}{-\Gamma + i\omega} \left.\left(e^{-\Gamma t}\, e^{i\omega t}\right)\right|_{t=0}^{\infty} = \frac{1}{\Gamma - i\omega}\,. \tag{10.55}$$

This function of ω is called a "pole." While the function is perfectly well behaved for real ω, it blows up for $\omega = -i\Gamma$, which is called the position of the pole in the complex plane. Now we just have to add the complex conjugate to get

$$\begin{aligned} C(\omega) &= \frac{1}{2\pi}\left(\frac{1}{\Gamma - i\omega} + \frac{1}{\Gamma + i\omega}\right) \\ &= \frac{1}{2\pi}\left(\frac{\Gamma + i\omega}{\Gamma^2 + \omega^2} + \frac{\Gamma - i\omega}{\Gamma^2 + \omega^2}\right) = \frac{1}{2\pi}\,\frac{2\Gamma}{\Gamma^2 + \omega^2} \end{aligned} \tag{10.56}$$

which is (10.40). We already checked, in (10.41), that the factor of $1/2\pi$ makes sense.

The pair (10.39)-(10.40) illustrates a very general fact about signals and their associated frequency spectra. In figure 10.4 we plot $f(t)$ for $\Gamma = 0.5$ and $\Gamma = 2$ and in figure 10.5, we plot $C(\omega)$ for the same values of Γ. Notice that as Γ increases, the signal becomes more sharply peaked near $t = 0$ but the frequency spectrum spreads out. And conversely if Γ is small so that $C(\omega)$ is sharply peaked near $\omega = 0$, then $f(t)$ is spread out in time. This complementary behavior is general. To resolve short times, you need a broad spectrum of frequencies.

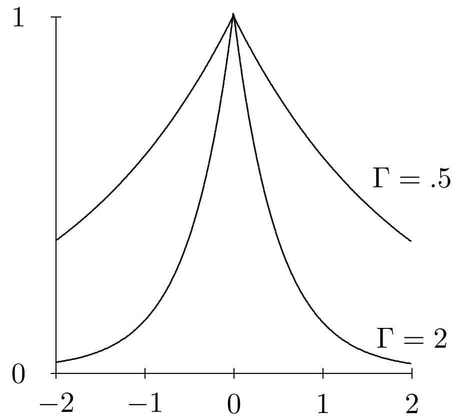


Figure 10.4: $f(t) = e^{-|\Gamma t|}$ for $\Gamma = 0.5$ and $\Gamma = 2$.

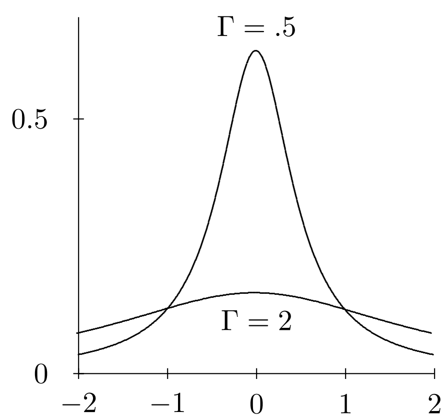


Figure 10.5: $C(\omega)$ for the same values of Γ.

10.3.2 Broad Generalities

We can state this fact very generally using a precise mathematical definition of the spread of the signal in time and the spread of the spectrum in frequency.

We will define the intensity of the signal to be proportional to $|f(t)|^2$. Then, we can define the average value of any function $g(t)$ weighted with the signal's intensity as follows

$$\langle g(t) \rangle = \frac{\int_{-\infty}^{\infty} dt\, g(t)\, |f(t)|^2}{\int_{-\infty}^{\infty} dt\, |f(t)|^2} \ . \tag{10.57}$$

This weights $g(t)$ most when the signal is most intense.

① Define that

For example, $\langle t \rangle$ is the average time, that is the time value around which the signal is most intense. Then

$$\left\langle [t - \langle t \rangle]^2 \right\rangle \equiv \Delta t^2 \tag{10.58}$$

measures the mean-square deviation from the average time, so it is a measure of the spread of the signal.

We can define the average value of a function of ω in an analogous way by integrating over the intensity of the frequency spectrum. But here is the trick. Because of (10.9) and (10.37), we can go back and forth between $f(t)$ and $C(\omega)$ at will. They carry the same information. We ought to be able to calculate averages of functions of ω by using an integral over t. And sure enough, we can. Consider the integral

$$\int_{-\infty}^{\infty} d\omega\, \omega\, C(\omega)\, e^{-i\omega t} = i\, \frac{\partial}{\partial t} \int_{-\infty}^{\infty} d\omega\, C(\omega)\, e^{-i\omega t} = i\, \frac{\partial}{\partial t}\, f(t)\,. \tag{10.59}$$

This shows that multiplying $C(\omega)$ by ω is equivalent to differentiating the corresponding $f(t)$ and multiplying by i.

Thus we can calculate $\langle \omega \rangle$ as

$$\langle \omega \rangle = \frac{\int_{-\infty}^{\infty} dt\, f(t)^*\, i \frac{\partial}{\partial t}\, f(t)}{\int_{-\infty}^{\infty} dt\, |f(t)|^2}\,, \tag{10.60}$$

and

$$\Delta\omega^2 \equiv \left\langle [\omega - \langle \omega \rangle]^2 \right\rangle = \frac{\int_{-\infty}^{\infty} dt\, \left| \left(i \frac{\partial}{\partial t} - \langle \omega \rangle \right)\, f(t) \right|^2}{\int_{-\infty}^{\infty} dt\, |f(t)|^2}\,. \tag{10.61}$$

$\Delta\omega$ is a measure of the spread of the frequency spectrum, or the "bandwidth."

Now we can state and prove the following result:

$$\Delta t \cdot \Delta\omega \geq \frac{1}{2}\,. \tag{10.62}$$

One important consequence of this theorem is that for a given bandwidth, $\Delta\omega$, the spread in time of the signal cannot be arbitrarily small, but is bounded by

$$\Delta t \geq \frac{1}{2\Delta\omega}\,. \tag{10.63}$$

The smaller the minimum possible value of Δt you can send, the higher the "fidelity" you can achieve. Smaller Δt means that you can send signals with sharper details. But (10.63) means that the smaller the bandwidth, the larger the minimum Δt, and the lower the fidelity.

To prove (10.62) consider the function[3]

$$\left([t - \langle t \rangle] - i\kappa \left[i \frac{\partial}{\partial t} - \langle \omega \rangle \right] \right)\, f(t) = r(t)\,, \tag{10.64}$$

[3]This is a trick borrowed from a similar analysis that leads to the Heisenberg uncertainty principle in quantum mechanics. Don't worry if it is not obvious to you where it comes from. The important thing is the result.

which depends on the entirely free parameter κ. Now look at the ratio

$$\frac{\int_{-\infty}^{\infty} dt\,|r(t)|^2}{\int_{-\infty}^{\infty} dt\,|f(t)|^2}\,. \tag{10.65}$$

This ratio is obviously positive, because the integrands of both the numerator and the denominator are positive. What we will do is choose κ cleverly, so that the fact that the ratio is positive tells us something interesting.

First, we will simplify (10.65). In the terms in (10.65) that involve derivatives of $f(t)^*$, we can integrate by parts (and throw away the boundary terms because we assume $f(t)$ goes to zero at infinity) so that the derivatives act on $f(t)$. Then (10.65) becomes

$$\Delta t^2 + \kappa^2\,\Delta\omega^2 + \kappa\,\frac{\int_{-\infty}^{\infty} dt\,f(t)^*\left(t\frac{\partial}{\partial t} - \frac{\partial}{\partial t}t\right)\,f(t)}{\int_{-\infty}^{\infty} dt\,|f(t)|^2}\,. \tag{10.66}$$

All other terms cancel. But

$$\frac{\partial}{\partial t}\left[t\,f(t)\right] = f(t) + t\,\frac{\partial}{\partial t}f(t)\,. \tag{10.67}$$

Thus the last term in (10.66) is just κ, and (10.65) becomes

$$\Delta t^2 + \kappa^2\,\Delta\omega^2 - \kappa\,. \tag{10.68}$$

(10.68) is clearly greater than or equal to zero for any value of κ, because it is a ratio of positive integrals. To get the most information from the fact that it is positive, we should choose κ so that (10.65) (=(10.68)) is as small as possible. In other words, we should find the value of κ that minimizes (10.68). If we differentiate (10.68) and set the result to zero, we find

$$\kappa_{\rm min} = \frac{1}{2\Delta\omega^2}\,. \tag{10.69}$$

We can now plug this back into (10.68) to find the minimum, which is still greater than or equal to zero. It is

$$\Delta t^2 - \frac{1}{4\Delta\omega^2} \geq 0 \tag{10.70}$$

which immediately yields (10.62).

Equation (10.62) appears in many places in physics. A simple example is bandwidth in AM radio transmissions. A typical commercial AM station broadcasts in a band of frequency about 5000 cycles/s (5 kc) on either side of the carrier wave frequency. Thus

$$\Delta\omega = 2\pi\Delta\nu \approx 3\times 10^4\,{\rm s}^{-1}\,, \tag{10.71}$$

and they cannot send signals that separate times less than a few $\times 10^{-5}$ seconds apart. This is good enough for talk and acceptable for some music.

A famous example of (10.62) comes from quantum mechanics. There is a completely analogous relation between the spatial spread of a wave packet, Δx, and the spread of k values required to produce it, Δk:

$$\Delta x \cdot \Delta k \geq \frac{1}{2}\,. \tag{10.72}$$

In quantum mechanics, the momentum of a particle is related to the k value of the wave that describes it by

$$p = \hbar k\,, \tag{10.73}$$

where $\hbar$ is Planck's constant h divided by 2π. Thus (10.72) implies

$$\Delta x \cdot \Delta p \geq \frac{\hbar}{2}\,. \tag{10.74}$$

This is the mathematical statement of the fact that the position and momentum of a particle cannot be specified simultaneously. This is Heisenberg's uncertainty relation.

10.4 Scattering of Wave Packets

In a real scattering experiment, we are interested not in an incoming harmonic wave that has always existed and will always exist. Instead we are interested in an incoming **wave packet** that is limited in time. In this section, we discuss two examples of scattering of wave packets.

10.4.1 Scattering from a Boundary

💻 10-3

We begin with the easier of the two examples. Consider the scattering of a wave packet from the boundary between two semi-infinite dispersionless strings both with tension T and different densities, ρ_I and ρ_{II}, as shown in figure 9.1. The dispersion relations are:

$$\omega^2 = \begin{cases} v_I^2 k^2 = \dfrac{T}{\rho_I}\, k^2 = & \text{in region } I \\ v_{II}^2 k^2 = \dfrac{T}{\rho_{II}}\, k^2 & \text{in region } II \end{cases} \tag{10.75}$$

where v_I and v_{II} are the phase velocities in the two regions.

Specifically, we assume that the boundary condition at $-\infty$ is that there is an incoming wave,

$$f(x - vt) \tag{10.76}$$

in region I, but no incoming wave in region II, and we wish to find the outgoing waves, the reflected wave in region I and the transmitted wave in region II.

We can solve this problem without decomposing the wave packet into its harmonic components with a trick that is analogous to that used at the beginning of this chapter to solve the forced oscillation problem, figure 10.1. The most general solution to the boundary conditions at $\pm\infty$ is

$$\psi(x,t) = \begin{cases} f(t - x/v_I) + g(t + x/v_I) & \text{in region } I \\ h(t - x/v_{II}) & \text{in region } II \end{cases} \tag{10.77}$$

where g and h are arbitrary functions. To actually determine the reflected and transmitted waves, we must impose the boundary conditions at $x = 0$, that the displacement is continuous (because the string doesn't break) and its x derivative is continuous (because the knot joining the two strings is massless):

$$f(t) + g(t) = h(t)\,, \tag{10.78}$$

and

$$\frac{\partial}{\partial x}\left[f(t - x/v_I) + g(t + x/v_I)\right]\Big|_{x=0} = \frac{\partial}{\partial x}\, h(t - x/v_{II})\Big|_{x=0}\,. \tag{10.79}$$

Using the chain rule in (10.79), we can relate the partial derivatives with respect to x to deriviatives of the functions,

$$\frac{1}{v_I}\left[-f'(t - x/v_I) + g'(t + x/v_I)\right]\Big|_{x=0} = -\frac{1}{v_{II}}\, h'(t - x/v_{II})\Big|_{x=0}\,, \tag{10.80}$$

or

$$-f'(t) + g'(t) = -\frac{v_I}{v_{II}}\, h'(t)\,. \tag{10.81}$$

Differentiating (10.78), we get

$$f'(t) + g'(t) = h'(t)\,, \tag{10.82}$$

Now for every value of t, (10.81) and (10.82) form a pair of simultaneous linear equations that can be solved for $g'(t)$ and $h'(t)$ in terms of $f'(t)$:

$$g'(t) = \frac{1 - v_I/v_{II}}{1 + v_I/v_{II}}\, f'(t)\,, \quad h'(t) = \frac{2}{1 + v_I/v_{II}}\, f'(t)\,. \tag{10.83}$$

Undoing the derivatives, we can write

$$g(t) = \frac{1 - v_I/v_{II}}{1 + v_I/v_{II}}\, f(t) + k_1\,, \quad h(t) = \frac{2}{1 + v_I/v_{II}}\, f(t) + k_2\,, \tag{10.84}$$

where k_1 and k_2 are constants, independent of t. In fact, though, we must have $k_1 = k_2$ to satisfy (10.78), and adding the same constant in both regions is irrelevant, because it just corresponds to our freedom to move the whole string up or down in the transverse direction. Thus we conclude that

$$g(t) = \frac{1 - v_I/v_{II}}{1 + v_I/v_{II}}\, f(t)\,, \quad h(t) = \frac{2}{1 + v_I/v_{II}}\, f(t)\,, \tag{10.85}$$

and the solution, (10.77), becomes

$$\psi(x,t) = \begin{cases} f(t - x/v_I) + \dfrac{1 - v_I/v_{II}}{1 + v_I/v_{II}}\, f(t + x/v_I) & \text{in region } I, \\ \dfrac{2}{1 + v_I/v_{II}}\, f(t - x/v_{II}) & \text{in region } II. \end{cases} \tag{10.86}$$

The same result emerges if we take the incoming wave packet apart into its harmonic components. For each harmonic component, the reflection and transmission components are the same (from (9.16)):

$$\begin{aligned} \tau &= \frac{2Z_I}{Z_I + Z_{II}} = \frac{2}{1 + v_I/v_{II}}\,, \\ R &= \frac{Z_I - Z_{II}}{Z_I + Z_{II}} = \frac{1 - v_I/v_{II}}{1 + v_I/v_{II}}\,. \end{aligned} \tag{10.87}$$

When we now put the harmonic components back together to get the scatter and transmitted wave packets, the coefficients, ρ and τ appear just as overall constants in front of the original pulse, as in (10.86).

This scattering process is animated in program 10-3. Here you can input different values of v_{II}/v_I to see how the reflection and transmission is affected. Notice that v_{II}/v_I very small corresponds to a large impedance ratio, Z_{II}/Z_I, which means that the string in region II does not move very much. Then we get a reflected pulse that is just the incoming pulse flipped over below the string. In the extreme limit, $v_{II}/v_I \to \infty$, the boundary at $x = 0$ acts like a fixed end. v_{II}/v_I very large corresponds to a small impedance ratio, Z_{II}/Z_I, in which case the string in region I hardly notices the string in region II. In the limit $v_{II}/v_I \to 0$, the boundary at $x = 0$ acts like a free end.

10.4.2 A Mass on a String

10-4

A more interesting example of the scattering of wave packets that can be worked out using the mathematics we have already done is the scattering of an incoming wave packet with the shape of (10.39) encountering a mass on a string. Here the dispersion relation is trivial, so the wave packet propagates without change of shape until it "hits" the mass. But then interesting things happen. This time, when we decompose the wave packet into its harmonic components, the reflection and transmission coefficients depend on ω. When we add them

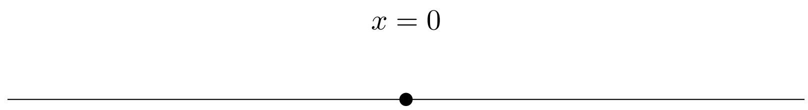


Figure 10.6: A mass on a string.

back up again to get the reflected and transmitted wave packets, we will find that the shape has changed. We will work this out in detail. The familiar setup is shown in figure 10.6.

For an incoming harmonic wave of amplitude A, the displacement looks like

$$\psi(x,t) = Ae^{ikx} \cdot e^{-i\omega t} + R\,Ae^{-ikx} \cdot e^{-i\omega t} \text{ for } x \leq 0 \tag{10.88}$$

$$\psi(x,t) = \tau\,Ae^{ikx} \cdot e^{-i\omega t} \text{ for } x \geq 0 \tag{10.89}$$

The solution for R and τ was worked out in the last chapter in (9.39)-(9.45). However, the parameter ϵ of (9.38) depends on ω. In order to disentangle the frequency dependence of the scattered wave packets, we write R and τ as

$$\tau = \frac{2\Omega}{2\Omega - i\omega}\,, \quad R = \frac{i\omega}{2\Omega - i\omega}\,, \tag{10.90}$$

where

$$\Omega \equiv \frac{T}{m\,v} = \frac{\sqrt{\rho\,T}}{m}\,, \tag{10.91}$$

is independent of ω — it depends just on the fixed parameters of the string and the mass. Note that in the notation of (9.38),

$$\Omega = \frac{\omega}{\epsilon}\,. \tag{10.92}$$

Suppose that we have not a harmonic incoming wave, but an incoming pulse:

$$\psi_{\text{in}}(x - vt) = Ae^{-\Gamma|t-x/v|}\,. \tag{10.93}$$

Now the situation is more interesting. We expect a solution of the form

$$\psi(x,t) = \psi_{\text{in}}(x - vt) + \psi_R(x + vt) \text{ for } x \leq 0 \tag{10.94}$$

$$\psi(x,t) = \psi_\tau(x - vt) \text{ for } x \geq 0 \tag{10.95}$$

where $\psi_\tau(x + vt)$ is the transmitted wave, traveling in the $+x$ direction, and $\psi_R(x + vt)$ is the reflected wave, traveling in the $-x$ direction. To get the reflected and transmitted waves, we will use superposition and take ψ_{in} apart into harmonic components. We can then use

(10.90) to determine the scattering of each of the components, and then can put the pieces back together to get the solution. Thus we start by Fourier transforming $\psi_{\rm in}$:

$$\psi_{\rm in}(x,t) = \int d\omega\, e^{-i\omega(t-x/v)}\, C_{\rm in}(\omega)\,. \tag{10.96}$$

We know from our discussion of signals that

$$\begin{gathered} C_{\rm in}(\omega) = \frac{1}{2\pi}\int dt\, e^{i\omega t}\,\psi_{\rm in}(0,t) \\ = \frac{1}{2\pi}\int_0^\infty dt\, Ae^{i\omega t}\, e^{-\Gamma t} + \text{h.c.} = \frac{1}{2\pi}\left(\frac{1}{\Gamma - i\omega} + \frac{1}{\Gamma + i\omega}\right)\,. \end{gathered} \tag{10.97}$$

Now to get the reflected and transmitted pulses, we multiply the components of $\psi_{\rm in}$ by the reflection and transmission amplitudes R and τ for unit $\psi_{\rm in}$

$$C_\tau(\omega) = A\,\frac{1}{2\pi}\left(\frac{1}{\Gamma - i\omega} + \frac{1}{\Gamma + i\omega}\right)\frac{2\Omega}{2\Omega - i\omega} \tag{10.98}$$

$$C_R(\omega) = A\,\frac{1}{2\pi}\left(\frac{1}{\Gamma - i\omega} + \frac{1}{\Gamma + i\omega}\right)\frac{i\omega}{2\Omega - i\omega} \tag{10.99}$$

Now we have to reverse the process and find the Fourier transforms of these to get the reflected and transmitted pulses. This is straightforward, because we can rewrite (10.98) and (10.99) in terms of single poles in ω:

$$\begin{gathered} C_\tau(\omega) = A\,\frac{1}{2\pi}\,\frac{2\Omega}{2\Omega - \Gamma}\cdot\left(\frac{1}{\Gamma - i\omega} - \frac{1}{2\Omega - i\omega}\right) \\ +\frac{1}{2\pi}\,\frac{2\Omega}{2\Omega + \Gamma}\cdot\left(\frac{1}{\Gamma + i\omega} + \frac{1}{2\Omega - i\omega}\right)\,; \end{gathered} \tag{10.100}$$

$$\begin{gathered} C_R(\omega) = A\,\frac{1}{2\pi}\,\frac{1}{2\Omega - \Gamma}\cdot\left(\frac{\Gamma}{\Gamma - i\omega} - \frac{2\Omega}{2\Omega - i\omega}\right) \\ +\frac{1}{2\pi}\,\frac{1}{2\Omega + \Gamma}\cdot\left(-\frac{\Gamma}{\Gamma + i\omega} + \frac{2\Omega}{2\Omega - i\omega}\right)\,. \end{gathered} \tag{10.101}$$

Now we can work backwards in (10.100) and (10.101) to get the Fourier transforms. We know from (10.55) that each term is the Fourier transform of an exponential. It is straightforward, but tedious, to put them back together. The result is reproduced below (note that we have combined the two terms in each expression proportional to $1/(2\Omega - i\omega)$).

$$\begin{gathered} \psi_\tau(x,t) = \frac{2\Omega}{2\Omega - \Gamma}\theta(t - x/v)\, Ae^{-\Gamma(t-x/v)} \\ -\frac{4\Omega\Gamma}{4\Omega^2 - \Gamma^2}\theta(t - x/v)\, Ae^{-2\Omega(t-x/v)} + \frac{2\Omega}{2\Omega + \Gamma}\theta(-t + x/v)\, Ae^{\Gamma(t-x/v)} \end{gathered} \tag{10.102}$$

and

$$\psi_r(x,t) = \frac{2\Gamma}{2\Omega-\Gamma}\theta(t+x/v)\,Ae^{-\Gamma(t+x/v)}$$
$$-\frac{4\Omega\Gamma}{4\Omega^2-\Gamma^2}\theta(t+x/v)\,Ae^{-2\Omega(t+x/v)} - \frac{2\Gamma}{2\Omega+\Gamma}\theta(-t-x/v)\,Ae^{\Gamma(t+x/v)} \tag{10.103}$$

where

$$\theta(t) = \begin{cases} 1 \text{ for } t \geq 0\,, \\ 0 \text{ for } t < 0\,. \end{cases} \tag{10.104}$$

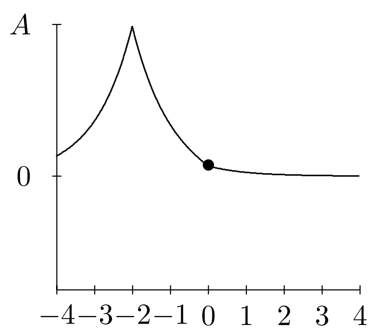


Figure 10.7: A wave packet on a stretched string, at $t = -2$.

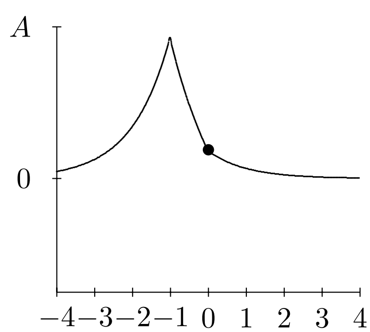


Figure 10.8: $t = -1$.

These formulas are not very transparent or informative, but we can put them into a computer and look at the result. We will plot the result in the limit $2\Omega \to \Gamma$. The results, (10.102) and (10.103) look singular in this limit, but actually, the limit exists and is perfectly smooth.[4] In figures 10.7-10.12, we show $\psi(x,t)$ for $\Gamma = v = 1$ in arbitrary units, for t values from -2

[4]The apparent singularity is similar to one that occurs in the approach to critical damping, discussed in (2.12).

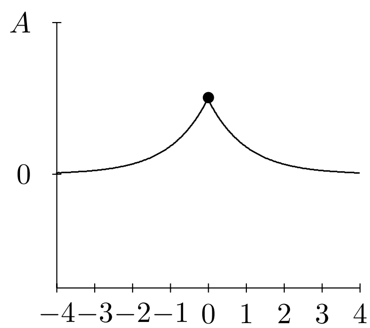


Figure 10.9: $t = 0$.

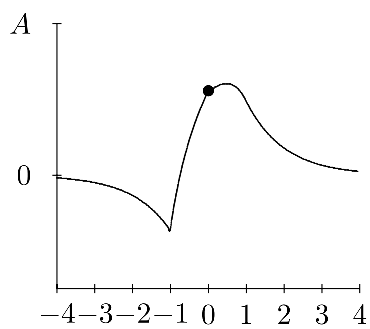


Figure 10.10: $t = 1$.

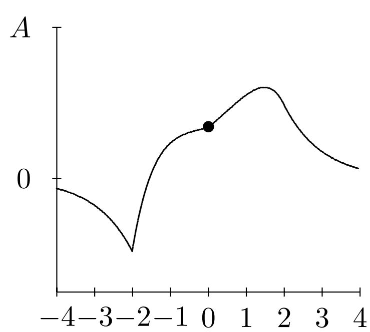


Figure 10.11: $t = 2$.

to 3. At $t = -2$, you see the pulse approaching the mass for negative t. At $t = -1$, you can begin to see the effect of the mass on the string. By $t = 0$, the string to the left of $x = 0$ is moving rapidly downwards. At $t = 1$, downward motion of the string for $x < 0$ has continued, and has begun to form the reflected pulse. For $t = 2$, you can see the transmitted and

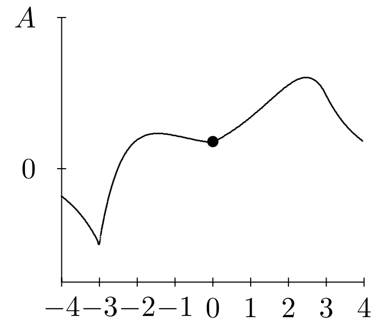


Figure 10.12: $t = 3$.

reflected waves beginning to separate. For $t = 3$, you can see the reflected and transmitted pulses have separated almost completely and the mass has returned nearly to its equilibrium position. For large positive t, the pulse is split into a reflected and transmitted wave.

The really interesting stuff is going on between $t = 0$ and $t = 1$, so we will look at this on a finer time scale in figures 10.13-10.16. To really appreciate this, you should see it in motion. It is animated in program 10-4.

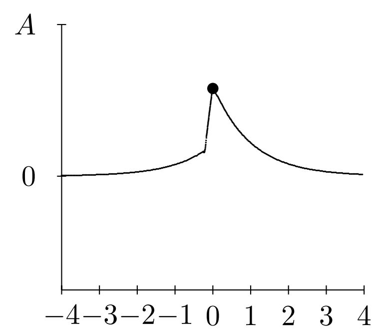


Figure 10.13: This is $t = .2$.

10.5 Is c the Speed of Light?

We have seen that an electromagnetic wave in the z direction satisfying Maxwell's equations in free space has the dispersion relation (8.47), so that light, at least in vacuum, travels at the speed of light. But is the theory right? How do we test the dispersion relation? In fact, the most sensitive tests of Maxwell's equations do not involve traveling waves. They come from observations of magnetic fields that extend over astrophysical distances (like the galaxy!). However, there is an interesting, if not very sensitive, way of looking for corrections to (8.47)

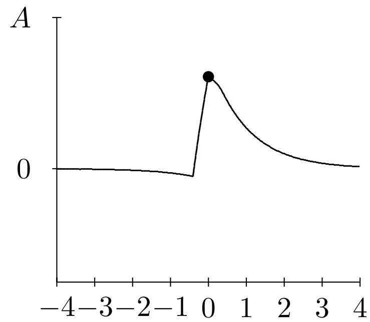


Figure 10.14: This is $t = .4$.

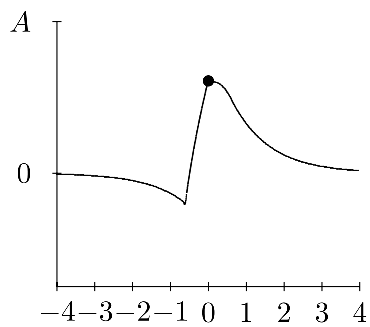


Figure 10.15: This is $t = .6$.

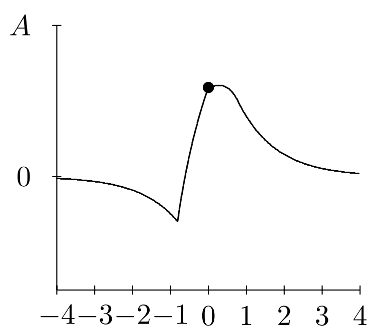


Figure 10.16: This is $t = .8$.

that involves the speed of light directly. Before discussing this, let us digress briefly to talk in more detail about photons, the particles of light that we described briefly in chapter 8.

Light is a wave phenomenon, as we have seen. Indeed, the wave properties of light are obvious in our everyday experience. It is less obvious from our everyday experience, but equally true, that light also consists of photons. This becomes obvious when you work with light at very low intensities and/or very high energies. That both of these statements can be true simultaneously is one of the (many) miracles of quantum mechanics.

Quantum mechanics tells us that all particles have wave properties. A particle with momentum p and energy E has an associated angular frequency and angular wave number related by

$$E = \hbar\,\omega\,, \quad p = \hbar\,k\,, \tag{10.105}$$

where $\hbar$ is Planck's constant divided by 2π. This combination appears so ubiquitously in quantum mechanics that it has its own symbol, and we physicists almost always use $\hbar$ rather than h. The reason is just that h is related to the frequency ν, rather than the angular frequency, ω, and we have seen that ω is the more convenient measure for most purposes. In addition, the energy and momentum of the particle are related as follows:

$$E^2 = p^2\,c^2 + m^2\,c^4\,, \quad v = c\,\frac{p\,c}{E} \tag{10.106}$$

where m is the rest mass and v is the classical velocity.

If we put (10.105) into (10.106), we get a dispersion relation for the quantum mechanical wave associated with the particle

$$\omega^2 = c^2 k^2 + \omega_0^2\,, \quad \omega_0 = \frac{mc^2}{\hbar}\,. \tag{10.107}$$

The classical velocity is the **group velocity** of the quantum mechanical wave!

$$v = \frac{\partial\omega}{\partial k} = c^2\,\frac{k}{\omega} = c\,\frac{p\,c}{E} \tag{10.108}$$

In fact, particles, in a quantum mechanical picture, correspond to wave packets that move with the group velocity.

The quantum mechanical dispersion relation, (10.107), agrees with (8.47) only if $m = 0$. Thus we can restate the question of whether (8.47) is correct by asking "Is the photon mass really zero?"

It would seem that we ought to be able to test this idea by looking at two photons with different frequencies emitted at the same time from a far away object and checking whether they arrive at the same time. There is an obvious flaw in this scheme. If the object is so far away that we cannot get there, how do we know that the two photons were emitted at the same time? In fact, astrophysics has provided us with a way around this difficulty. We can look at pulsars. Pulsars are (presumably) rotating neutron star remnants of supernova explosions that emit light toward the earth at regular intervals. For example, pulsar 1937+21 is so regular that the departure time of photons can be determined to within a few microseconds (μs).[5] It

[5] See G. Barbiellini and G. Cocconi, *Nature* 329 (1987) 21.

is also about 16,000 light years away, so the photons with the higher frequency (the faster ones) have plenty of time to get ahead. When this experiment is done, one finds a nonzero ω_0, of about $1.7 \times 10^4 s^{-1}$, corresponding to a mass of about 1.26×10^{-49}g. That seems like a rather small mass, but in fact, it is ridiculously large for a photon. From studies of the galactic magnetic field, we suspect that it is less than 4×10^{-65}g![6] Thus something else is going on.

The problem with this measurement as a test of the dispersion relation is that there are electrons lying around out there — free electrons in interstellar space (10^{-1} to 10^{-2}cm^{-3}). These electrons in space will wiggle in the E field — this will produce a current density that will affect Maxwell's equations, and that, in turn, will affect the dispersion relation. Let us analyze the effect of this dilute plasma assuming that the electron density is constant. Then (at least for the long wavelength radio waves of interest in these experiments) we can still use translation invariance to understand what is happening. Consider a plane wave in the z direction and suppose that the electric field of the plane wave is in the x direction. Then it is still true that at a given ω

$$E_x(\vec{r},t) = E_0\, e^{i(kz-\omega t)}\,, \quad B_y(\vec{r},t) = B_0\, e^{i(kz-\omega t)}\,, \tag{10.109}$$

for some k. To find k, we must look at the effect of the electric fields on the electrons, and then go back to Maxwell's equations. The fields are very small, and for small fields the induced electron velocities, v are small. Thus we can neglect B. Then the force on an electron at the point $(\vec{r},t)$ is

$$F_x(\vec{r},t) = e\, E_x(\vec{r},t) = e\, E_0\, e^{i(kz-\omega t)} = m\, a_x(\vec{r},t) \tag{10.110}$$

The displacement of the electron has the same form:

$$d_x(\vec{r},t) = d_0\, e^{i(kz-\omega t)} \tag{10.111}$$

which implies

$$a_x(\vec{r},t) = -\omega^2\, d_0\, e^{i(kz-\omega t)} \tag{10.112}$$

comparing (10.110) and (10.112) gives

$$d_0 = -\frac{e\, E_0}{m\, \omega^2}\,. \tag{10.113}$$

Thus the electrons are displaced 180° out of phase with the electric field and in the same direction. Then the electron velocity is

$$v_x = \frac{i\, e\, E_0}{m\, \omega}\, e^{i(kz-\omega t)}. \tag{10.114}$$

[6]Chibisov, Soviet Physics - *Uspekhi*, 19 (1986) 624.

The movement of the electrons gives rise to a current density: [7]

$$\mathcal{J}_x = \frac{i\,e^2 N\,E_0}{m\,\omega}\,e^{i(kz-\omega t)} \tag{10.115}$$

where N is the electron number density.

Putting this into the relevant Maxwell's equations, we find

$$k\,E_0 = \omega\,B_0\,, \quad -k\,B_0 = -\omega\mu_0\epsilon_0\,E_0 + \mu_0\,\frac{e^2 N\,E_0}{m\,\omega}\,, \tag{10.116}$$

or using $c = 1/\sqrt{\mu_0\epsilon_0}$, (8.47),

$$B_0 = \frac{k}{\omega}\,E_0\,, \quad -\frac{k^2}{\omega} = -\frac{\omega}{c^2} + \frac{e^2 N}{c^2\,m\,\epsilon_0\,\omega}\,, \tag{10.117}$$

or solving for ω^2

$$\omega^2 = c^2\,k^2 + \omega_0^2\,, \quad \text{with} \quad \omega_0^2 = \frac{e^2 N}{\epsilon_0 m}\,. \tag{10.118}$$

The constant ω_0 in (10.118) is called the **"plasma frequency."** The amazing thing is that it looks just like a photon mass. For $N \approx 10^{-2}\text{cm}^{-3}$, this is consistent with the observation from the pulsar.

Chapter Checklist

You should now be able to:

i. Solve a forced oscillation problem for a stretched string with arbitrary time dependent displacement at the end;

ii. Decompose an arbitrary signal into harmonic components using the Fourier transformation;

iii. Compute the group velocity of a dispersive system;

iv. Understand the relations between a function and its Fourier transform that lead to the relation between bandwidth and fidelity;

v. Be able to describe the scattering of a wave packet;

vi. Understand the effect of free charges on the dispersion relation of electromagnetic waves.

[7] Notice that the result is inversely proportional to the electron mass. This why we are concentrating on electrons rather than protons. The protons don't move as fast!

Problems

10.1. Is it possible for a medium that supports electromagnetic waves to have the dispersion relation $\omega^2 = c^2\, k^2 - \omega_0^2$ for real ω_0?

Why or why not?

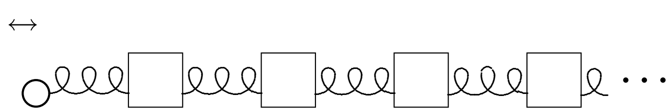
10.2. A beaded string has neighboring beads separated by a. If the maximum possible group velocity for waves on the string is v, find T/m.

10.3. In the next chapter, we will derive the dispersion relation for waves in water (or at least an idealized picture of water). If the water is deep, the dispersion relation is

$$\omega^2 = gk + \frac{Tk^3}{\rho}$$

where g is the acceleration of gravity, 980 in cgs units, T is the surface tension, 72, and ρ is density, 1.0. Find the group velocity and phase velocity as a function of wavelength. When are they equal?

10.4. Consider the longitudinal oscillations of the system of blocks and **massless** springs shown below:



Each block has mass m. Each spring has spring constant K. The equilibrium separation between the blocks is a. The ring on the left is moved back and forth with displacement $B\cos\omega t$. This produces a traveling wave in the system moving to the right for $\omega < 2\sqrt{K/m}$. There is no traveling wave moving to the left.

The dispersion relation for the system is

$$\omega^2 = \frac{4K}{m}\sin^2\frac{ka}{2}\,.$$

a. Suppose that $\omega = \sqrt{K/m}$. Find the phase velocity of traveling waves at this frequency.

b. For $\omega = \sqrt{K/m}$, find the displacement of the first block at time $t = \pi/2\omega$. Express the answer as B times a pure number.

c. Find the group velocity in the limit $\omega \to 2\sqrt{K/m}$.

d. Find the time average of the power supplied by the force on the ring in the limit $\omega \to 2\sqrt{K/m}$.

e. Explain the relation between the answers to parts **c** and **d**. You may be able to do this part even if you have gotten confused in the algebra. Think about the physics and try to understand what must be going on.

Chapter 11

Two and Three Dimensions

The concepts of space translation invariance and local interactions can be extended to systems with more than one space dimension in a straightforward way. But in two and three dimensions, these ideas alone are not enough to determine the normal modes of an arbitrary system. One needs extra tricks, or plain hard work.

Preview

Here, we will only be able to discuss the very simplest sort of tricks, but at least we will be able to understand why the problems are more difficult.

i. We begin by explaining why the angular wave number, k, becomes a vector in two or three dimensions. We find the normal modes of systems with simple boundary conditions.

ii. We then discuss scattering from planes in two- and three-dimensional space. We derive Snell's law of refraction and discuss total internal reflection and tunneling.

iii. We discuss the example of Chladni plates.

iv. We give a two-dimensional example of a waveguide, in which the waves are constrained to propagate only in one direction.

v. We study water waves (in a simplified version of water).

vi. We introduce the more advanced topic of spherical waves.

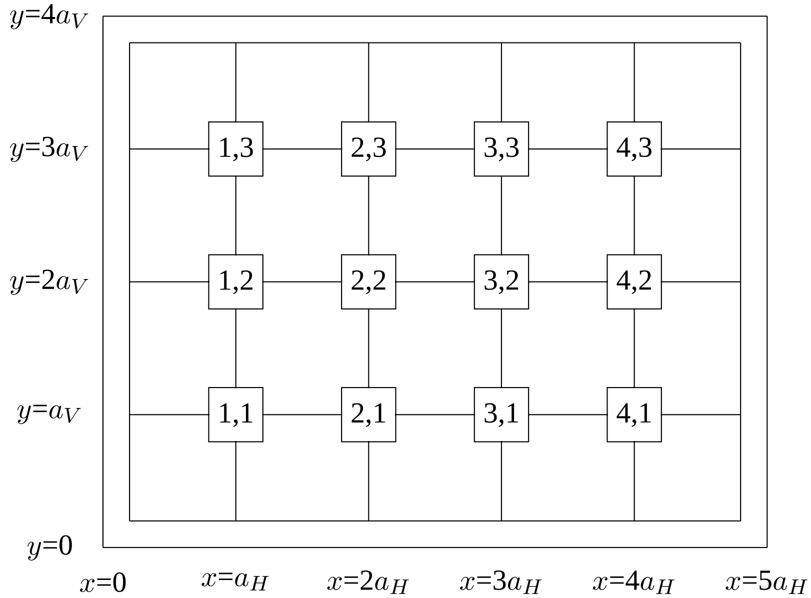


Figure 11.1: A two-dimensional beaded mesh.

11.1 The $\vec{k}$ Vector

Consider the two-dimensional beaded mesh, a two-dimensional analog of the beaded string, shown in figure 11.1. All the beads have mass m. The tension of the horizontal (vertical) strings is T_H (T_V) and the interbead distance is a_H (a_V). There is no damping. We can label the beads by a pair of integers (j, k) indicating their horizontal and vertical positions as shown. Alternatively, we can label the beads by their positions in the x, y plane according to

$$(x, y) = (ja_H, ka_V) \,. \tag{11.1}$$

Thus, we can describe their small transverse (out of the plane of the paper, in the z direction) oscillations either by a matrix $\psi_{jk}(t)$ or by a function

$$\psi(x, y, t) \,; \quad 0 \leq x \leq 5a_H \,, \; 0 \leq y \leq 4a_V \,. \tag{11.2}$$

We will use (11.2) because we can then extend the discussion to continuous systems more easily. We are interested only in the transverse oscillations of this system, in which the blocks move up and down out of the plane of the paper, because these oscillations do not stretch the strings very much (only to second order in the small displacements). The other oscillations of such a system have much higher frequencies and are strongly damped, so they are not very interesting.

As in the one-dimensional case, the first step is to remove the walls and consider the infinite system obtained by extending the interior in all directions. The oscillations of the resulting system can be described by a function $\psi(x, y, t)$, where x and y are not constrained.

This infinite system looks the same if it is translated by a_V vertically, or by a_H horizontally. We can write down solutions for the infinite system by using our discussion of the one-dimensional case twice. Because the system has translation invariance in the x direction, we expect that we can find eigenstates of the $M^{-1}K$ matrix proportional to

$$e^{ik_x x} \tag{11.3}$$

for any constant k_x. Because the system has translation invariance in the y direction, we expect that we can find eigenstates of the $M^{-1}K$ matrix proportional to

$$e^{ik_y y} \tag{11.4}$$

for any constant k_y. Putting (11.3) and (11.4) together, we expect that we can find eigenstates of the $M^{-1}K$ matrix that have the form

$$\psi(x, y) = A\, e^{ik_x x}\, e^{ik_y y} = A\, e^{i\vec{k}\cdot\vec{r}} \tag{11.5}$$

where $\vec{k} \cdot \vec{r}$ is the two-dimensional dot product

$$\vec{k} \cdot \vec{r} = k_x x + k_y y\,. \tag{11.6}$$

In other words, the wave number has become a vector.

As with the one-dimensional system, we can use (11.5) to determine the dispersion relation of the infinite system. Putting in the t dependence, we have a displacement of the form

$$\psi(x, y, t) = Ae^{i\vec{k}\cdot\vec{r}}\, e^{-i\omega t}. \tag{11.7}$$

The analysis is precisely analogous to that for the one-dimensional beaded string, with the result that ω^2 is simply a sum of vertical and horizontal contributions, each of which look like the dispersion relation for the one-dimensional case:

$$\omega^2 = \frac{4T_H}{ma_H} \sin^2 \frac{k_x a_H}{2} + \frac{4T_V}{ma_V} \sin^2 \frac{k_y a_V}{2}\ . \tag{11.8}$$

Equations (11.7) and (11.8) are the complete solution to the equations of motion for the infinite beaded mesh.

11.1.1 The Difference between One and Two Dimensions

11-1

So far, our analysis has been essentially the same in two dimensions as it was in one. The next step, though, is very different. In the one-dimensional case, where the normal modes are $e^{\pm ikx}$, there are only two modes with any given value of ω^2. Thus, no matter what the boundary conditions are, we only have to worry about superposing two modes at a time. But in the two-dimensional case, there are a continuously infinite number of solutions to (11.8) for any ω, because you can lower k_x and compensate by raising k_y. Thus a normal mode of the finite two-dimensional system with no damping (which is just some solution in which all the beads oscillate in phase with the same ω) may be a linear combination of an infinite number of the nice simple space translation invariant modes of the infinite system.

Sure enough, in general, the two-dimensional case is infinitely harder. If figure 11.1 were a system with a more complicated shape, we would not be able to find an analytic solution. But for the special case of a rectangular frame, aligned with the beads, the boundary conditions are not so bad, because both the modes, (11.5) and the boundary conditions can be simply expressed in terms of products of one-dimensional normal modes.

The boundary conditions for the system in figure 11.1 are;

$$\psi(0,y,t) = \psi(L_H,y,t) = \psi(x,0,t) = \psi(x,L_V,t) = 0\,, \tag{11.9}$$

where

$$L_H = 5a_H\,, \quad L_V = 4a_V\,. \tag{11.10}$$

In the corresponding infinite system, a piece of which is shown in figure 11.2, (11.9) implies that the beads along the dotted rectangle are all at rest. Comparing figure 11.1 and figure 11.2, you can see that this boundary condition captures the physics of the walls in figure 11.1.

Now to find the normal modes of the finite system in figure 11.1, we must find linear combinations of modes of the infinite system that satisfy the boundary conditions, (11.9). We can satisfy (11.9) by forming linear combinations of just four modes of the infinite system:[1]

$$Ae^{\pm ik_x x}e^{\pm ik_y y} \tag{11.11}$$

where

$$k_x = n\pi/L_H\,, \;\; k_y = n'\pi/L_V\,. \tag{11.12}$$

Then we can take the solutions to be a product of sines,

$$\begin{gathered}\psi(x,y) = A\sin(n\pi x/L_H)\,\sin(n'\pi y/L_V)\\ \text{for } n = 1 \text{ to } 4 \text{ and } n' = 1 \text{ to } 3\,.\end{gathered} \tag{11.13}$$

[1] There is a symmetry at work here! The modes in which the $\vec{k}$ vector is lined up along the x or y axes are those that behave simply under reflections through the center of the rectangle.

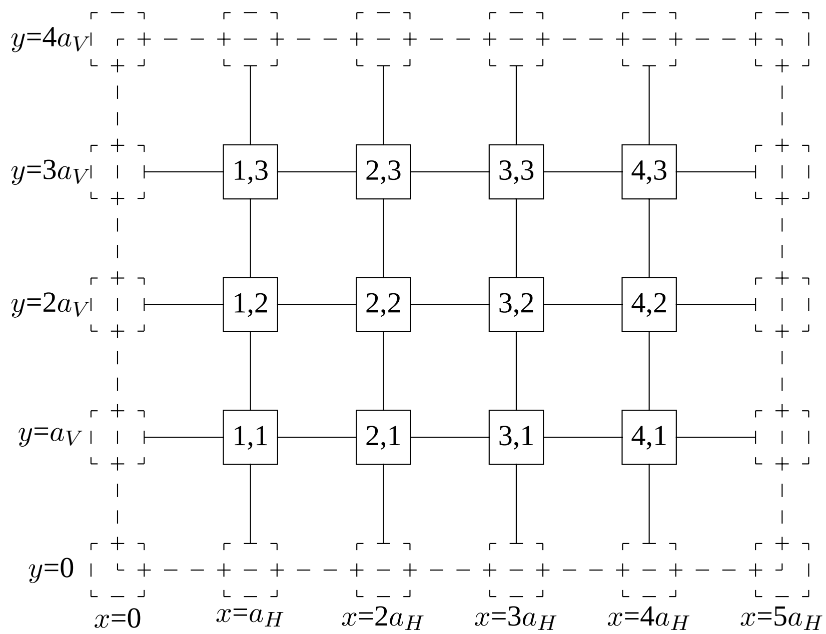


Figure 11.2: A piece of an infinite two-dimensional beaded mesh.

The frequency of each mode is given by the dispersion relation (11.8):

$$\omega^2 = \frac{4T_H}{ma_H} \sin^2 \frac{n\pi a_H}{2L_H} + \frac{4T_V}{ma_V} \sin^2 \frac{n'\pi a_V}{2L_V} \ . \tag{11.14}$$

These modes are animated in program 11-1.

The solution of this problem is an example of a technique called "separation of variables." In the right variables, in this case, x and y, the problem falls apart into one-dimensional problems. This trick works equally well in the continuous case, so long as the boundary surface is rectangular. If we take the limit in which a_V and a_H are very small compared to the wavelengths of interest, we can express (11.8) in terms of quantities that make sense in the continuum limit, just as in the analysis of the continuous one-dimensional string as the limit of the beaded string, in chapter 6. Assume, for simplicity, that

$$a_V = a_H = a \ \text{ and } \ T_V = T_H = T \tag{11.15}$$

(so that the x and y directions have the same properties). The quantities that characterize the surface in this case are the surface mass density,

$$\rho_s = \frac{m}{a^2} \,, \tag{11.16}$$

and the surface tension,

$$T_s = \frac{T}{a}\,. \tag{11.17}$$

The surface tension is the pull per unit transverse distance exerted by the membrane. When these quantities remain finite as the separation, a, goes to zero, (11.8) becomes

$$\omega^2 = \frac{T_s}{\rho_s}\left(k_x^2 + k_y^2\right) = \frac{T_s}{\rho_s}\left|\vec{k}\right|^2\,. \tag{11.18}$$

An argument that is precisely analogous to that for the one-dimensional case shows that in this limit, $\psi(x, y, t)$ satisfies the two-dimensional wave equation,

$$\frac{\partial^2}{\partial t^2}\psi(x,y,t) = v^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi(x,y,t) = v^2\,\vec{\nabla}^2\psi(x,y,t)\,. \tag{11.19}$$

Note that in this limit, the special properties of the x and y axes that were manifest in the finite system have completely disappeared from the equation of motion. The wave numbers k_x and k_y form a two-dimensional vector $\vec{k}$. The infinite number of solutions to the dispersion relation (11.18) are just those obtained by rotating $\vec{k}$ in all possible ways without changing its length. This makes it possible to solve for the normal modes in circular regions, for example. But we will not discuss these more complicated boundary conditions now. It is clear, however, that (11.13) is the solution for the rectangular region in the continuous case, and that the corresponding frequency is

$$\omega^2 = \frac{T_s}{\rho_s}\left[\left(\frac{n\pi}{L_H}\right)^2 + \left(\frac{n'\pi}{L_V}\right)^2\right]\,. \tag{11.20}$$

Now because the system is continuous, the integers n and n' run from zero to infinity (though $n = n'$ is not interesting), or until the continuum approximation breaks down.

11.1.2 Three Dimensions

The beaded mesh cannot be extended to three dimensions because there is no transverse direction. But a system of masses connected by elastic rods can be three-dimensional, and indeed, this sort of system is a good model of an elastic solid. This system is rather complicated because each mass can move in all three directions. A two-dimensional version of this is illustrated in figure 11.3. This system is the same as figure 11.1 except that the strings have been replaced by light, elastic rods, so that system is in equilibrium even without the frame. Now we are interested in the oscillations of this system **in the plane of the paper**. Compared to figure 11.1, this system has twice as many degrees of freedom, because each block can move in both the x and y direction, while in figure 11.1, the blocks moved only

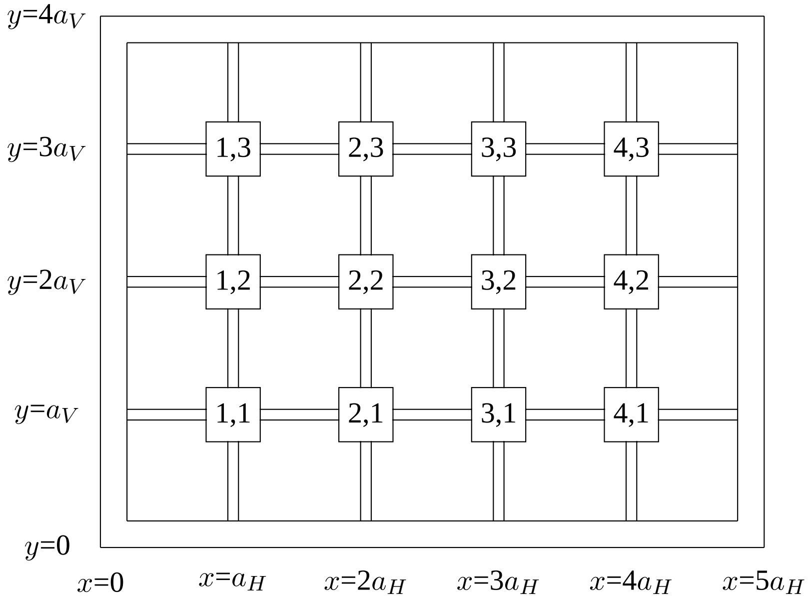


Figure 11.3: A two-dimensional solid, with masses connected by elastic rods.

in the z direction. This means that we cannot use space translation invariance alone, even to determine the modes of the infinite system.

For each value of $\vec{k}$, there will be four modes rather than the usual two. We would have to do some matrix analysis to see which combinations of x and y motion were actually the normal modes. We will not do this in general, but will discuss it briefly in the continuum limit, to remind you of some physics that is important for fields like geology.

Consider the continuous, infinite system obtained by taking the a's very small in figure 11.3, with other quantities scaling appropriately. Consider a wave with wave number $\vec{k}$. The normal modes will have the form

$$\vec{A}\, e^{\pm \vec{k}\cdot\vec{r}}\,, \tag{11.21}$$

for some vector $\vec{A}$ (in the three-dimensional case, $\vec{A}$ is a 3-vector, in our two-dimensional example, it is a 2-vector). If the system is rotation invariant, then there is no direction picked out by the physics except the direction of $\vec{k}$. Then the normal modes must be a longitudinal or "compressional" mode

$$\vec{A} \propto \vec{k}\,, \tag{11.22}$$

and a transverse or "shear" mode

$$\vec{A} \perp \vec{k}\,. \tag{11.23}$$

Each mode will have its own characteristic dispersion relation. In three dimensions, there will be two shear modes, because there are two perpendicular directions, and they will have the same dispersion relation, because one can be rotated into the other.

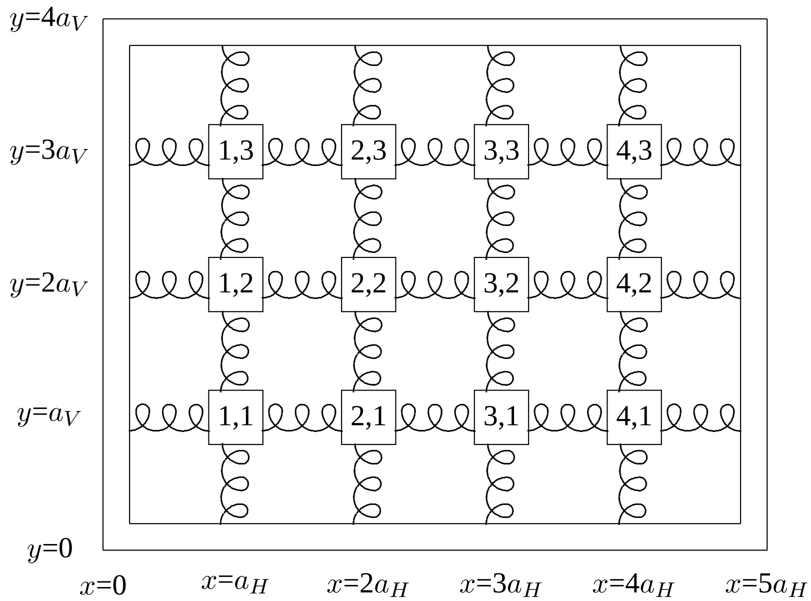


Figure 11.4: A two-dimensional system of beads and springs.

11.1.3 Sound Waves

In a liquid or a gas, there are no shear waves because there is no restoring force that keeps the system in a particular shape. The shear modes have zero frequency. If we replaced the rods in figure 11.3 with unstretched springs, we would get a system with the same property, shown in figure 11.4. Without the frame, this system would not be rigid. However, the compressional modes are still there. These are analogous to sound waves. For an approximately continuous system like air, we expect a dispersion relation of the form

$$\omega^2 = v^2 \, \vec{k}^2 \tag{11.24}$$

where v is constant unless k is too large. We have already calculated v, in (7.43), by considering one-dimensional oscillations. It is called the speed of sound because it is the speed of sound waves in an infinite or semi-infinite system.

We can describe the normal modes of a rectangular box full of air in terms of a function $P(x, y, z)$ that describes the gas pressure at the point (x, y, z). The pressure or density of the compressional wave is related to the displacement $\vec{\psi}(x, y, z)$:

$$\vec{\psi} \propto \vec{\nabla} P\,, \quad P \propto -\vec{\nabla} \cdot \vec{\psi}\,. \tag{11.25}$$

As in the two-dimensional system described above, we can use separation of variables and find a solution that is a product of functions of single variables. The only difference here is that the boundary conditions are different. Because of (11.25), which is the mathematical statement of the fact that gas is actually pushed from regions of high pressure to regions of low pressure, the pressure gradient perpendicular to the boundary must vanish. The gas at the boundary has nowhere to go. Thus the normal modes in a rectangular box, $0 \leq x \leq X$, $0 \leq y \leq Y$, $0 \leq z \leq Z$, have the form

$$P(x, y, z) = A \cos(n_x \pi x/X)\, \cos(n_y \pi y/Y)\, \cos(n_z \pi z/Z) \tag{11.26}$$

with frequency

$$\omega^2 = v^2 \left(\left(\frac{n_x \pi}{X}\right)^2 + \left(\frac{n_y \pi}{Y}\right)^2 + \left(\frac{n_z \pi}{Z}\right)^2 \right)\,. \tag{11.27}$$

The trivial solution $n_x = n_y = n_z = 0$ represents stationary air. If any of the n's is nonzero, the mode is nontrivial.

11.2 Plane Boundaries

The easiest traveling waves to discuss in two and three dimensions are "plane waves," solutions in the infinite system of the form

$$\psi(r, t) = A e^{i(\vec{k}\cdot\vec{r} - \omega t)}\,. \tag{11.28}$$

This describes a wave traveling the direction of the wave-number vector, $\vec{k}$, with the phase velocity in the medium. The displacement (or whatever) is constant on planes of constant $\vec{k} \cdot \vec{r}$, which are perpendicular to the direction of motion, $\vec{k}$. We will study more complicated traveling waves soon, when we discuss diffraction. Then we will learn how to describe "beams" of light or sound or other waves that are the traveling waves with which we usually work. We will see how to describe them as superpositions of plane waves. For now, you can think of a plane wave as being something like the traveling wave you would encounter inside a wide, coherent beam, or very far from a small source of nearly monochromatic light, light with a definite frequency. That should be enough to give you a physical picture of the phenomena we discuss in this section.

We are most interested in waves such as light and sound. However, it is much easier to discuss the transverse oscillations of a two-dimensional membrane, and many of our examples will be in that system. There are two reasons. One is that a two-dimensional membrane is easier to picture on two-dimensional paper. The other reason is that the physics is very simple, so we can concentrate on the wave properties. We will try to point out where things get more complicated for other sorts of wave phenomena.

Consider two two-dimensional membranes stretched in the $z = 0$ plane, as shown in figure 11.5. For $x < 0$, suppose that the surface mass density is ρ_s and surface tension T_s. For $x > 0$, suppose that the surface mass density is ρ'_s and surface tension T'_s. This is a two-dimensional analog of the string system that we discussed at length in chapter 9. The boundary between the two membranes must supply a force (in this case, a constant force per unit length) in the x direction to support the difference between the tensions, as in the system of figure 9.2. However, we will assume that whatever the mechanism is that supplies this force, it is massless, frictionless and infinitely flexible.

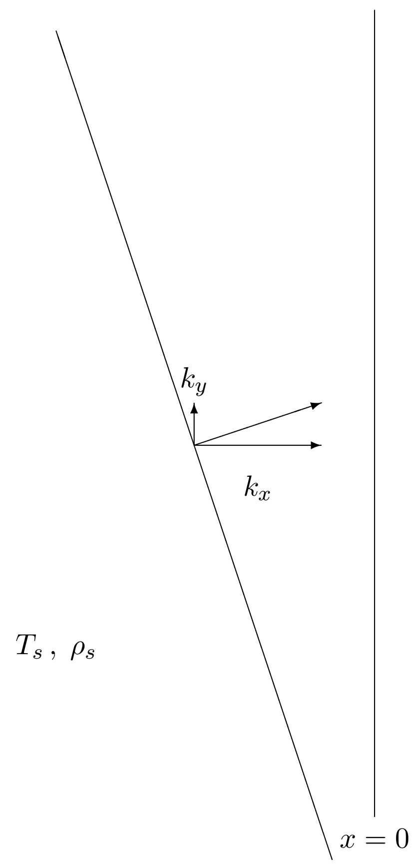


Figure 11.5: A line of constant phase in a plane wave approaching a boundary.

Now again, we can consider reflection of traveling waves. Thus, suppose that there is, in this membrane, a plane wave with amplitude A and wave number $\vec{k}$ for $x < 0$, traveling in toward the boundary at $x = 0$. The condition that the wave is traveling toward the boundary can be written in terms of the components of $\vec{k}$ as

$$k_x > 0 \,. \tag{11.29}$$

We would like to know what waves are produced by this incoming wave because of reflection and transmission at the boundary, $x = 0$. On general grounds of space translation invariance, we expect the solution to have the form

$$\begin{aligned} \psi(r,t) &= Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} + \sum_\alpha R_\alpha\, Ae^{i(\vec{k}_\alpha\cdot\vec{r}-\omega t)} && \text{for } x \leq 0 \\ \psi(r,t) &= \sum_\beta \tau_\beta\, Ae^{i(\vec{k}_\beta\cdot\vec{r}-\omega t)} && \text{for } x \geq 0 \end{aligned} \tag{11.30}$$

$$\vec{k}_\alpha^2 = \omega^2 \frac{\rho_s}{T_s} \;;\;\; \vec{k}_\beta^2 = \omega^2 \frac{{\rho'}_s}{T'_s} \,, \tag{11.31}$$

and

$$k_{\alpha x} < 0 \text{ and } k_{\beta_x} > 0 \text{ for all } \alpha \text{ and } \beta. \tag{11.32}$$

The α and β in (11.30) run over all the transmitted and reflected waves. We will show shortly that only one of each contributes for a plane boundary condition at $x = 0$, but (11.30) is completely general, following just from space translation invariance. Note that we have put in boundary conditions at $\pm\infty$ by requiring (11.29) and (11.32). Except for the incoming wave with amplitude A, all the other waves are moving away from the boundary. But we have not yet put in the boundary condition at $x = 0$.

11.2.1 Snell's Law — the Translation Invariant Boundary

11-2

As far as we know from considerations of the physics at $\pm\infty$, the reflected and transmitted waves could be a complicated superposition of an infinite number of plane waves going in various directions away from the boundary. In fact, if the boundary were irregularly shaped, that is exactly what we would expect. It is the fact that the boundary, $x = 0$, is itself invariant under space translations in the y directions that allows us to cut down the infinite number of parameters in (11.30) to only two. Because translations in the y direction leave the whole system invariant, **including the boundary**, we can find solutions in which all the components have the same irreducible y dependence. If the incoming wave is proportional to

$$e^{ik_y y} \,, \tag{11.33}$$

then all the components of (11.30) must also be proportional to $e^{ik_y y}$. Otherwise there is no way to satisfy the boundary condition at $x = 0$ **for all** y**.** That means that

$$k_{\alpha y} = k_y \,, \quad k_{\beta_y} = k_y \,. \tag{11.34}$$

But (11.34), together with (11.31) and (11.32), completely determines the wave vectors $\vec{k}_\alpha$ and $\vec{k}_\beta$. Then (11.30) becomes[2]

$$\begin{aligned} \psi(r,t) = Ae^{i\vec{k}\cdot\vec{r}-i\omega t} + R\,Ae^{i\vec{\tilde{k}}\cdot\vec{r}-i\omega t} &\equiv \psi_-(r,t) \quad \text{for } x \leq 0 \\ \psi(r,t) = \tau\,Ae^{i\vec{k}'\cdot\vec{r}-i\omega t} &\equiv \psi_+(r,t) \quad \text{for } x \geq 0 \end{aligned} \tag{11.35}$$

where

$$\tilde{k}_y = k_y \,, \quad k'_y = k_y \,, \tag{11.36}$$

and

$$\tilde{k}_x = -\sqrt{\omega^2/v^2 - k_y^2} = -k_x \,, \quad k'_x = \sqrt{\omega^2/v'^2 - k_y^2} \,, \tag{11.37}$$

with

$$v = \sqrt{\frac{T_s}{\rho_s}} \,, \quad v' = \sqrt{\frac{T'_s}{\rho'_s}} \,. \tag{11.38}$$

The entertaining thing about (11.35)-(11.37) is that we know everything about the directions of the reflected and transmitted waves, even though we have not even mentioned the details of the physics at the boundary. To get the directions, we needed only the invariance under translations in the y direction. The details of the physics of the boundary come in only when we want to calculate R and τ. The directions of the reflected and transmitted waves are the same for any system with a translation invariant boundary. Obviously, this argument works in three dimensions, as well. In fact, if we simply choose our coordinates so that the boundary is the $x = 0$ plane and the wave is traveling in the x-y plane, then nothing depends on the z coordinate and the analysis is exactly the same as above. For example, we can apply these arguments directly to electromagnetic waves. For electromagnetic waves in a transparent medium, because the phase velocity is $v_\phi = \omega/k$, the index of refraction, n, is proportional to k,

$$n = \frac{c}{v_\varphi} = k\frac{c}{\omega} \,. \tag{11.39}$$

(11.36)-(11.37) shows that the reflected wave comes off at the same angle as the incoming wave because the only difference between the k vectors of the incoming and reflected waves is a change of the sign of the x component. Thus the angle of incidence equals the angle of reflection. This is the rule of “specular reflection.” From (11.36), we can also derive Snell’s law of refraction for the angle of the refracted wave. If θ is the angle that the incident

[2]We have defined $\psi_\pm$ here to make it easier to discuss the boundary conditions, below.

wave makes with the perpendicular to the boundary, and θ' is the corresponding angle for the transmitted wave, then (11.36) implies

$$k \sin\theta = k' \sin\theta' . \tag{11.40}$$

For electromagnetic waves, we can rewrite this as

$$n \sin\theta = n' \sin\theta' . \tag{11.41}$$

For example, when an electromagnetic wave in air encounters a flat glass surface at an angle θ, $n' > n$ in (11.41). The wave is refracted toward the perpendicular to the surface. This is illustrated in figure 11.6 for $n' > n$.

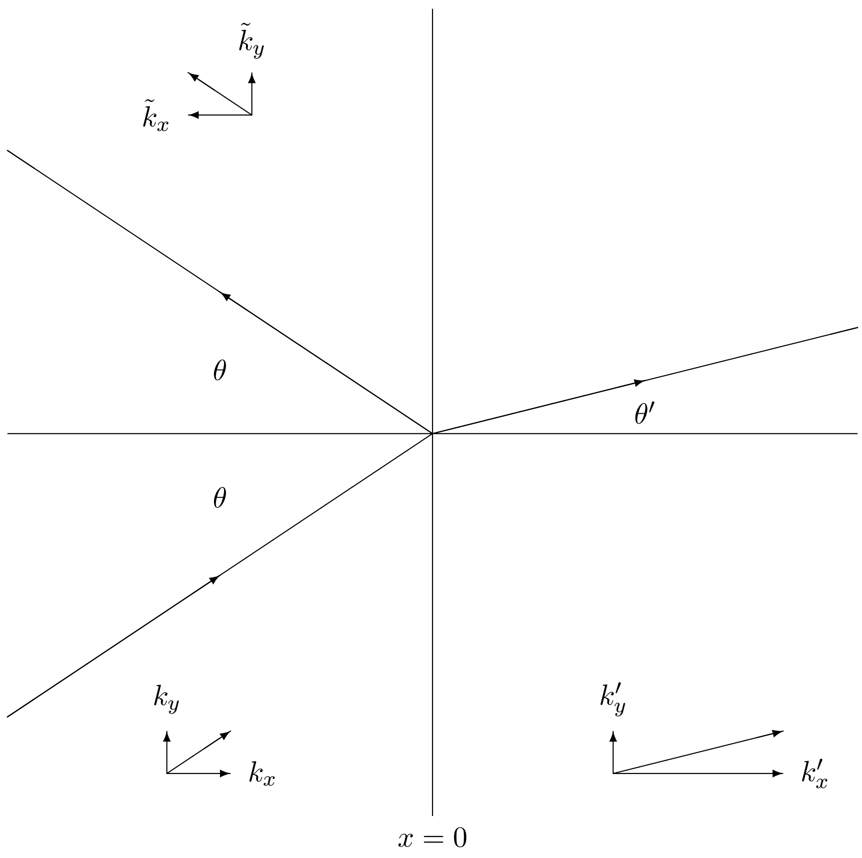


Figure 11.6: Reflection and transmission from a boundary.

Let us now finish the solution for the membrane problem by solving for R and τ in (11.35). To do this, we must finally discuss the boundary conditions in more detail. One is that the membrane is continuous, which from the form, (11.35), implies

$$\psi_-(r,t)\big|_{x=0} = \psi_+(r,t)\big|_{x=0} \;, \tag{11.42}$$

or

$$1 + R = \tau \,. \tag{11.43}$$

The other is that the vertical force on any small length of the membrane is zero. The force on a small length, $d\ell$, of the boundary at the point, $(0, y, 0)$, from the membrane for $x < 0$ is given by

$$-\,T_s d\ell \;\left.\frac{\partial\psi_-(r,t)}{\partial x}\right|_{x=0} \;. \tag{11.44}$$

This is analogous to the one-dimensional example illustrated in figure 8.6. The force of surface tension is perpendicular to the boundary, so for small displacements, only the slope of the displacement in the x direction matters. The slope in the y direction gives no contribution to the vertical force to first order in the displacement. Likewise, the force on a small length, $d\ell$, of the boundary at the point, $(0, y, 0)$, from the membrane for $x > 0$ is given by

$$T_s' d\ell \;\left.\frac{\partial\psi_+(r,t)}{\partial x}\right|_{x=0} \;. \tag{11.45}$$

Thus the other boundary condition is

$$T_s' d\ell \;\left.\frac{\partial\psi_+(r,t)}{\partial x}\right|_{x=0} = T_s d\ell \;\left.\frac{\partial\psi_-(r,t)}{\partial x}\right|_{x=0} \;, \tag{11.46}$$

or

$$T_s' k_x' \tau = T_s k_x (1 - R) \,. \tag{11.47}$$

Thus the solution is

$$\tau = \frac{2}{1+r} \,, \quad R = \frac{1-r}{1+r} \,, \tag{11.48}$$

where

$$r = \frac{T_s' k_x'}{T_s k_x} \,. \tag{11.49}$$

You can see from (11.48) and (11.49) that we can adjust the surface tension to make the reflected wave go away even when there is a change in the length of the $\vec{k}$ vector from one side of the boundary to the other. It is useful to think about refraction in this limit, because it will allow us to visualize it in a simple way. If $r = 1$ in (11.48), then $R = 0$ and $\tau = 1$. There is no reflected wave and the transmitted wave has the same amplitude as the incoming wave. Thus in each region, there is a single plane wave. Remember that a plane wave consists of infinite lines of constant phase perpendicular to the $\vec{k}$ vector, moving in the direction of the k vector with the phase velocity, $v_\varphi = \omega/|\vec{k}|$. In particular, suppose we look at lines on which the phase is zero, so that $\psi = A$. The perpendicular distance between two such lines is the wavelength, $2\pi/|\vec{k}|$, because the phase difference between neighboring lines is 2π. But here is the point. The lines in the two regions must meet at the boundary, $x = 0$, to satisfy the boundary condition, (11.43). If the incoming wave amplitude is 1 at $x = 0$, the outgoing wave amplitude is also 1. The lines where $\psi = A$ are continuous across the boundary, $x = 0$. This situation is illustrated in figure 11.7. The $\vec{k}$ vectors in the two regions are shown. Notice that the angle of the lines must change when the distance between them changes in order to maintain continuity at the boundary. In program 11-2, the same system is shown in motion.

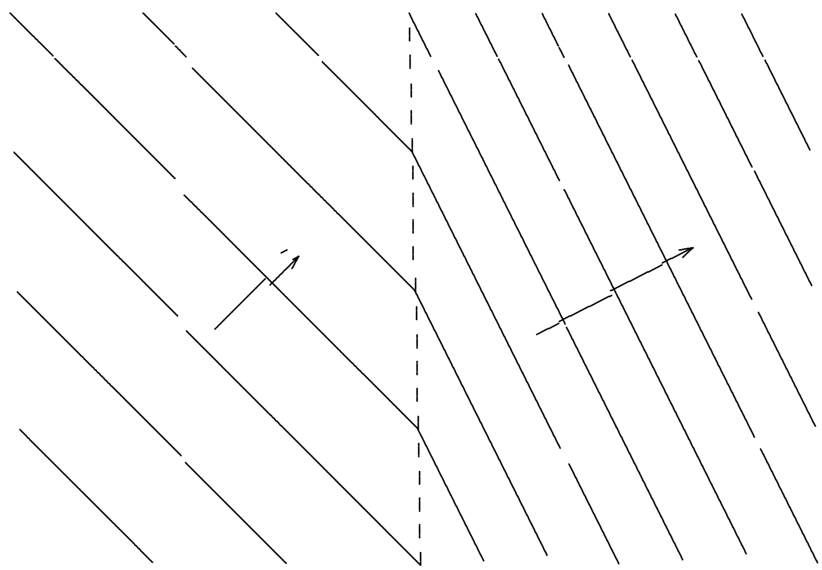


Figure 11.7: Lines of constant $\psi = 1$ for a system with refraction but no reflection.

11.2.2 Prisms

The nontrivial index of refraction of glass is the building block of many optical elements. Let us discuss the prism. In fact, to do the problem of the scattering of light waves by prisms entirely correctly would require much more sophisticated techniques than we have at our disposal at the moment. The reason is that the prism is not an infinite, flat surface with space translation invariance. In general, we would have to worry about the boundary. However, we can say interesting things even if we ignore this complication. The idea is to think not of an infinite plane wave, but of a wide beam of light incident on a face of the prism. A wide beam behaves very much like a plane wave, and we will ignore the difference in this chapter. We will see what the differences are in Chapter 13 when we discuss diffraction.

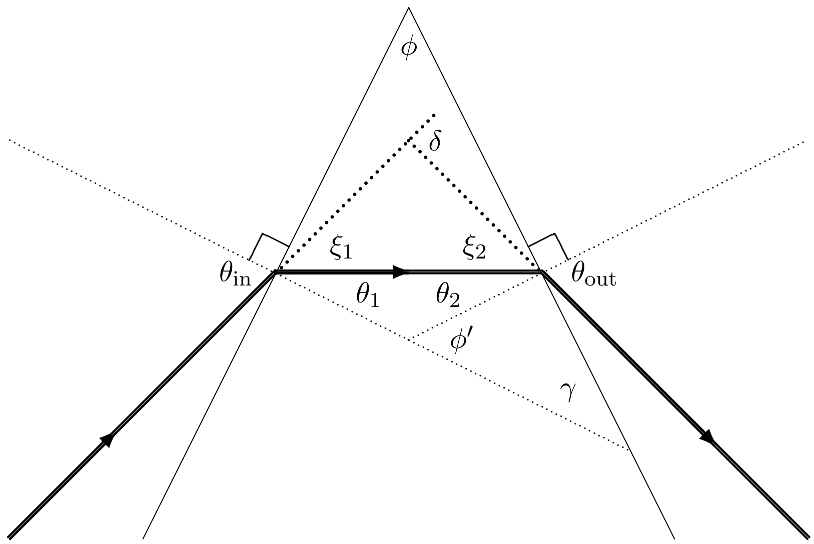


Figure 11.8: The geometry of a prism.

Thus we consider the following situation, in which a wide beam of light enters one face of a prism with index of refraction n and exits the other face. The geometry is shown in figure 11.8 (the directions of the beams are indicated by the thick lines). The interesting quantity is δ. This describes how much the direction of the outgoing beam has been deflected from the direction of the incoming beam by the prism. We can calculate it using simple geometry and Snell's law, (11.40). From Snell's law

$$\sin\theta_{\text{in}} = n\sin\theta_1 \tag{11.50}$$

and

$$\sin\theta_{\text{out}} = n\sin\theta_2\,. \tag{11.51}$$

Now for some geometry.

$$\theta_2 + \theta_1 = \phi' \tag{11.52}$$

— because the complement of ϕ', $\pi - \phi'$, along with θ_1 and θ_2 are the angles of a triangle, and thus add to π.

$$\phi = \phi' \tag{11.53}$$

— because ϕ and ϕ' are corresponding angles of the two similar right triangles with other acute angle γ. Thus

$$\delta = \xi_1 + \xi_2 = \theta_{\text{in}} + \theta_{\text{out}} - \theta_1 - \theta_2 = \theta_{\text{in}} + \theta_{\text{out}} - \phi \tag{11.54}$$

where we have used (11.52) and (11.53). But for small angles, from (11.50) and (11.51),

$$\theta_{\rm in} \approx n\,\theta_1\,, \quad \theta_{\rm out} \approx n\,\theta_2\,. \tag{11.55}$$

Thus

$$\delta \approx n\,(\theta_1+\theta_2) - \phi \approx (n-1)\,\phi\,. \tag{11.56}$$

The result, (11.56), is certainly reasonable. It must vanish when $n \to 1$, because there is no boundary for $n = 1$. If things are small and the answer is linear, it must be proportional to ϕ.

One of the most familiar characteristics of a prism results from the dependence of the index of refraction, n, on frequency. This causes a beam of white light to break up into colors. For most materials, the index of refraction increases with frequency, so that blue light is deflected more than red light by the prism. The physics of the frequency dependence of n is that of forced oscillation. The index of refraction of a material is related to the dielectric constant (see (9.53)), that in turn is related to the distortion of the electronic structure of the material caused by the electric field. For a varying field, this depends on the amplitude of the motion of bound charges within the material in an electric field. Because these charges are bound, they respond to the oscillating fields in an electromagnetic wave like a mass on a spring subject to an oscillating force. We know from our studies of forced oscillation that this amplitude has the form

$$\sum_{\substack{\rm resonances \\ \alpha}} \frac{C_\alpha}{\omega_\alpha^2 - \omega^2}\,, \tag{11.57}$$

where ω_α are the resonant frequencies of the system and the C_α are constants depending on the details of how the force acts on the degrees of freedom. We can estimate the order of magnitude of these resonant frequencies with dimensional analysis, if we remember that any material consists of electrons and nuclei held together by electrical forces (and quantum mechanics, of course, but $\hbar$ will not enter into our estimate except implicitly, in the typical atomic distance). The relevant quantities are[3]

The charge of the proton	$e \approx 1.6 \times 10^{-19}\,\rm C$	(11.58)
The mass of the electron	$m_e \approx 9.11 \times 10^{-31}\,\rm kg$	
Typical atomic distance	$a \approx 10^{-10}\,\rm m = 1\mathring{A}$	
The speed of light	$c = 299,792,458\,\rm m/s$	

In terms of these parameters, we would guess that the typical force inside the materials is of order $\frac{e^2}{4\pi\epsilon_0 a^2}$ (from Coulomb's law), and thus that the spring constant is of order $\frac{e^2}{4\pi\epsilon_0 a^3}$ (the typical force over the typical distance). Thus we expect

$$\omega_\alpha^2 \approx \sqrt{\frac{e^2}{4\pi\epsilon_0 a^3 m_e}} \tag{11.59}$$

[3]Note that it is the mass of the electron rather than the mass of the proton that is relevant, because the electrons move much more in electric fields.

and

$$\lambda_\alpha \approx \frac{2\pi c}{\omega_\alpha} \approx 2\pi c \sqrt{\frac{4\pi\epsilon_0 a^3 m_e}{e^2}} \approx 10^{-7}\,\mathrm{m} = 1000\,\text{Å}\,. \tag{11.60}$$

This is a wavelength in the ultraviolet region of the electromagnetic spectrum, shorter than that of visible light. That means that for visible light, $\omega < \omega_\alpha$, and thus the displacement, (11.57), increases as ω increases for visible light. The distortion of the electronic structure of the material caused by a varying electric field increases as the frequency increases in the visible spectrum. Thus the dielectric constant of the material increases with frequency. Thus blue light is deflected more.

Incidentally, this is the same reason that the sky is blue. Blue light is scattered more than red light because its frequency is closer to the important resonances of the air molecules.

11.2.3 Total Internal Reflection

The situation in which the wave comes from a region of large $|\vec{k}|$ into a region of smaller $|\vec{k}|$ has another feature that is surprising and very useful. This situation is depicted in figure 11.9 for a system with no reflection. For small, θ, as shown in figure 11.9, this looks rather

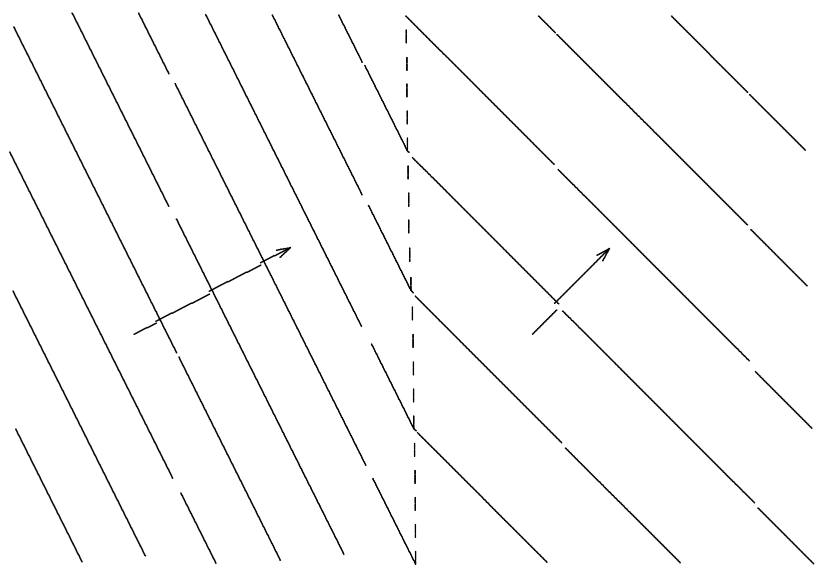


Figure 11.9: Lines of constant $\psi = 1$ for $n' < n$.

like figure 11.7, except that the wave is refracted away from the perpendicular to the surface instead of toward it. But suppose that the angle θ is large, satisfying

$$n\,\sin\theta/n' > 1\,. \tag{11.61}$$

Then there is no solution for real θ' in (11.41). Thus there can be no transmitted traveling wave. The incoming wave must be totally reflected by the boundary. This is total internal reflection. It happens when a plane wave tries to escape from a region of high $|\vec{k}|$ to a region of lower $|\vec{k}|$ at a grazing angle. It is extensively used in optical equipment and many other things. Let us investigate this peculiar phenomenon in more detail.

Suppose we start from $\theta = 0$ and increase θ. As θ increases, k_y increases and k_x decreases. This continues until we get to the boundary of total internal reflection, called the critical angle,

$$\sin\theta = \sin\theta_c \equiv \frac{n'}{n}\,. \tag{11.62}$$

The amplitudes for both the reflected and transmitted waves in (11.48) also increase. At the critical angle, k_x vanishes. The amplitude for the reflected wave is 1 and the amplitude for the transmitted wave is 2. However, even though the transmitted wave is nonzero, no energy is carried away from the boundary because the $\vec{k}$ vector points in the y direction.

As θ increases beyond the critical angle, k_y continues to increase. To satisfy the dispersion relation,

$$\omega^2 = v'^2\left(k_x^2 + k_y^2\right)\,, \tag{11.63}$$

k_x must be pure imaginary! The x dependence is then proportional to

$$e^{-\kappa x} \quad \text{where} \quad \kappa = \operatorname{Im} k_x\,. \tag{11.64}$$

Now the nature of the boundary condition at infinity changes. We can no longer require simply that $k_x > 0$. Instead, we must require

$$\operatorname{Im} k_x > 0\,. \tag{11.65}$$

The sign is important. If $\operatorname{Im} k_x$ were negative, the amplitude of the wave for $x > 0$ would increase with x, going exponentially to infinity as $x \to \infty$. This doesn't make much physical sense because it corresponds to a finite cause (the incoming wave for $x < 0$) producing an infinite effect. As we will see below, we can also come to this conclusion by going to this infinite system as a limit of a finite system.

We actually have three different boundary conditions at infinity for this situation:

$$\begin{aligned} \operatorname{Re} k_x > 0 \text{ for } \theta < \theta_c\,, \\ k_x = 0 \text{ for } \theta = \theta_c\,, \\ \operatorname{Im} k_x > 0 \text{ for } \theta > \theta_c\,. \end{aligned} \tag{11.66}$$

These three can be combined into a compound condition that is valid in all regions:

$$\operatorname{Re} k_x \geq 0\,, \quad \operatorname{Im} k_x \geq 0\,. \tag{11.67}$$

The condition, (11.67), is actually the most general statement of the outgoing traveling wave boundary condition at infinity. It is also correct in situations in which there is damping and both the real and imaginary parts of k_x are nonzero. This is the mathematical statement of the physical fact that the wave for $x > 0$, whatever its form, is produced at the boundary by the incoming wave.

From (11.48) and (11.49), you see that for $\theta > \theta_c$, the amplitude of the reflected wave becomes complex. However, its absolute value is still 1. All the energy of the incoming wave is reflected.

We have seen that in total internal reflections, the wave does penetrate into the forbidden region, but the x dependence is in the form of an exponential standing wave, not a traveling wave. The y dependence is that of a traveling wave. This is one of many situations in which the physics forces the nature of the two- or three-dimensional solution to have different properties in different directions.

It is easy to see total internal reflection in a fish-tank, a glass block, or some other rectangular transparent object with an index of refraction significantly greater than 1. You can look through one face of the rectangle and see the silvery reflection from an adjacent face, as illustrated in figure 11.10.

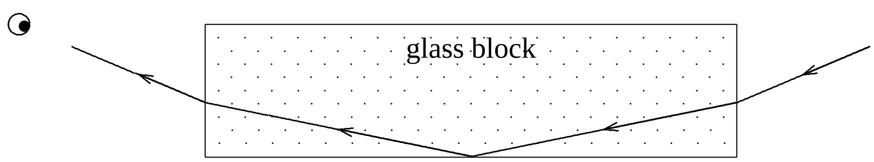


Figure 11.10: Total internal reflection in glass with index of refraction 2.

11.2.4 Tunneling

Consider the scattering of a plane wave in the system illustrated in figure 11.11. This is the same setup as in figure 11.10, except that another block of glass has been added a small distance, d, below the boundary from which there was total internal reflection. We have defined the positive x direction to be downwards for consistency with the discussion of Snell's law, above. Now does any of the light get through to the observer below, or is the light still totally reflected at the boundary, as in figure 11.10? The answer is that some light gets through. As we will see in detail in an example below, the presence of the other block of glass means that instead of a boundary condition at infinity, we have a boundary condition at the finite distance, d.

The details of this phenomenon for electromagnetic waves are somewhat complicated by polarization, which we will discuss in detail in the next chapter. However, there is a precisely analogous process in the transverse oscillation of membranes that we can analyze easily.

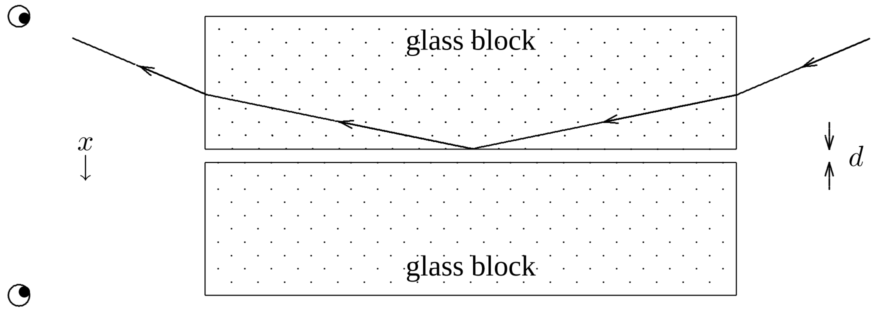


Figure 11.11: A simple experiment to demonstrate tunneling.

In fact, we will find that we have already analyzed it in chapter 9. Consider the scattering problem illustrated in figure 11.12. The unshaded region is a membrane with lower density. The arrows indicate the directions of the $\vec{k}$ vectors of the plane waves. The shaded regions

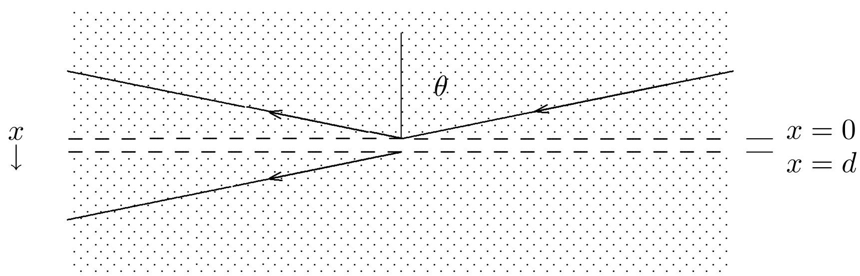


Figure 11.12: Tunneling in an infinite membrane.

have surface mass density ρ_s and surface tension T_s. The unshaded region, which extends from $x = 0$ to $x = d$, has the same surface tension but surface mass density $\rho_s/4$. Thus the ratio of phase velocities in the two regions is two, the same as the ratio from air to glass in figure 11.11. The dashed lines are massless boundaries between the different membranes.

We can now ask what are the coefficients, R and τ, for reflection and transmission. We have done this problem for a single boundary earlier in this chapter in (11.42)-(11.49). We could solve this one by putting two of these solutions together using the transfer matrix techniques of chapter 9. In fact, we do not even have to do that, because we can read off the result from (9.97) and (9.98) in the discussion of thin films in chapter 9. The point is that all the terms in our solution must have the same irreducible y dependence, $e^{ik_y y}$, because of the space translation invariance of the whole system including the boundary in the y direction. This common factor plays no role in the boundary conditions. If we factor it out, what is left looks like a one-dimensional scattering problem. Comparing (11.47) for $T_s = T_s'$ with

(9.10), you can see that the analyses become the same if we make the replacements

$$\begin{aligned} k_1 &\rightarrow k_x \\ k_2 &\rightarrow k_x' \\ L &\rightarrow d \end{aligned} \tag{11.68}$$

where k_x is the x component of the $\vec{k}$ vector of the incoming wave in the shaded region and k_x' is the x component of the $\vec{k}$ vector of the transmitted wave in the unshaded region. The result is

$$\tau = \left(\cos k_x' d - i\,\frac{k_x^2 + {k_x'}^2}{2k_x k_x'}\sin k_x' d\right)^{-1} e^{-ik_x d} \tag{11.69}$$

and

$$R = \left(i\,\frac{k_x^2 - {k_x'}^2}{2k_x k_x'}\sin k_x' d\right)\left(\cos k_x' d - i\,\frac{k_x^2 + {k_x'}^2}{2k_x k_x'}\sin k_x' d\right)^{-1}. \tag{11.70}$$

It may be a little easier to look at the intensity of the transmitted wave, which is proportional to

$$|\tau|^2 = \frac{2k_x^2 {k_x'}^2}{\left(k_x^4 + {k_x'}^4\right)\sin^2 k_x' d + 2k_x^2 {k_x'}^2}. \tag{11.71}$$

Note that we have not mentioned the critical angle or total internal reflection or anything like that. The reason is that our analysis in chapter 9 was perfectly general. It remains correct even if the angular wave number in the middle region becomes imaginary. All that happens for θ larger than the critical angle, θ_c, is that k_x' becomes imaginary. But this has a spectacular effect in (11.71). If $k_x' \rightarrow i\kappa$, where κ is real, then it follows from the Euler identity, (1.57) and (1.62), that

$$\sin k_x' d \rightarrow i \sinh \kappa d\,, \tag{11.72}$$

where $\sinh$ is the "hyperbolic sine", defined by

$$\sinh x \equiv \frac{e^x - e^{-x}}{2}. \tag{11.73}$$

Thus for angles above the critical angle, the denominator of (11.71) is an exponentially increasing function of d (the $e^{\kappa d}$ term in (11.73) dominates for large κd). The intensity of the transmitted wave therefore **decreases exponentially with** d. In the limit of large d, we quickly recover total internal reflection.

We can get some insight about what is happening by looking at the boundary conditions at $x = d$ for angles above the critical angle. For $x > d$, the wave has the form (suppressing the common factors of $e^{ik_y y}$ and $Ae^{-i\omega t}$)

$$\tau e^{ik_x x}. \tag{11.74}$$

For $0 \le x \le d$, the wave has the form

$$T_{II}e^{-\kappa x} + R_{II}e^{\kappa x}\,, \tag{11.75}$$

where I have called the coefficients T_{II} and R_{II} by analogy with transmitted and reflected waves, even though these are not traveling waves. The boundary conditions at $x = d$ are

$$\begin{aligned} \tau e^{ik_xd} &= T_{II}e^{-\kappa d} + R_{II}e^{\kappa d}\,, \\ ik_x\tau e^{ik_xd} &= \kappa\left(-T_{II}e^{-\kappa d} + R_{II}e^{\kappa d}\right)\,. \end{aligned} \tag{11.76}$$

This looks more complicated than it really is. If we solve for $T_{II}e^{-\kappa d}$ and $R_{II}e^{\kappa d}$ in terms of τe^{ik_xd}, the result is

$$T_{II}e^{-\kappa d} = \frac{2\kappa}{\kappa - ik_x}\tau e^{ik_xd}\,, \quad R_{II}e^{\kappa d} = \frac{2\kappa}{\kappa + ik_x}\tau e^{ik_xd}\,. \tag{11.77}$$

The important point is that the values of the two components of the wave, (11.75), at $x = d$, $T_{II}e^{-\kappa d}$ and $R_{II}e^{\kappa d}$, are more or less the same size. These two quantities do not have any exponential dependence on d. **This qualitative fact does not depend on the details of (11.76). It will be true for any reasonable boundary condition at $x = d$.**

Thus the coefficient, R_{II} of the "reflected" wave (in quotes because it is a real exponential wave, not a traveling wave) must be smaller than the "transmitted" wave by a factor of roughly $e^{2\kappa d}$. Notice that this justifies the statement, (11.67), of the boundary condition at infinity. As $d \to \infty$, for any reasonable physics at d, the wave becomes a pure negative exponential.

At $x = 0$, for large κd, the R_{II} term in wave will be completely negligible, and T_{II} term will be produced with some coefficient of order 1, just as in the limit of total internal reflection.

Thus what is happening in the boundary conditions for tunneling can be described qualitatively as follows. The incoming wave for $x < 0$ produces the $e^{-\kappa x}$ term in the region $0 \le x \le d$, with an exponentially small admixture of $e^{\kappa x}$. But at $x = d$, the two parts of the exponential wave are of the same size (both exponentially small), and they can produce the transmitted wave.

The rapid exponential dependence of the transmitted wave on d has some interesting consequences. It implies, for example, that the reflected wave is also very sensitive to the value of d, for small d (energy conservation implies $|R|^2 + |\tau|^2 = 1$). You can see this rapid dependence in the example of (11.10) by putting your finger on the bottom surface of the glass block or fish tank, where the wave is being reflected. You will see a ghostly fingerprint! The reason is that the tiny indentations on your finger are far enough away from the glass that κd is large and the wave is almost entirely reflected. But where the flesh is pressed tightly against the glass, the wave is absorbed. This is a simple version of a tunneling microscope.

Finally, before leaving the subject of tunneling, let us consider what happens when we turn down the intensity of the light wave in figure 11.11 so that we see the scattering of individual photons. The first thing to note is that each photon is either transmitted or reflected. The meaning of R and τ in this case is that the $|R|^2$ and $|\tau|^2$ are the **probabilities** of reflection and transmission. You cannot predict whether any particular photon will get through. In the quantum mechanical world, you can predict only the probabilities.

The second thing to note is that in the particle description, the whole phenomenon of tunneling is very peculiar. A classical photon, coming at the boundary of the glass plate at more than the critical angle could not enter at all into the air. It would be forbidden to do so by conservation of energy and conservation of the y component of momentum.[4] How can the particle get through to the $x > d$ side if it cannot exist for $0 < x < d$? Obviously, in classical physics, it cannot. Tunneling is, therefore, a truly quantum mechanical phenomenon. The wave manages to penetrate into the forbidden region, but only in the form of a real exponential wave, not a traveling wave. It is only for $x < 0$ and $x > d$, where the waves are traveling, that they can be interpreted as particles in anything like the classical sense.

11.3 Chladni Plates

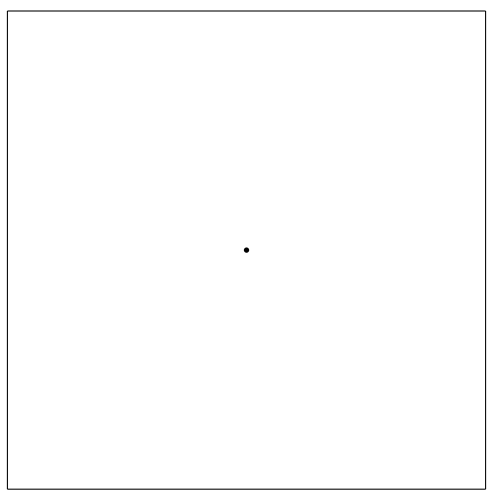


Figure 11.13: A Chladni plate.

[4]The boundary does not change p_y of the photon, because of the translation invariance in the y direction. However, there is no reason why the boundary cannot exert a force in the x direction and change p_x of the photon.

Chladni plates are a very pretty and instructive example of a two-dimensional oscillating system. A Chladni plate is simply a square metal plate that is driven transversely at its center. It is illustrated in figure 11.13. The dot in the center shows where the plate is driven in the transverse direction (out of the plane of the paper). The center, which we will take to have equilibrium position $\vec{r} = 0$, moves up and down out of the plane of the paper at a frequency ω. Let us assume that the square sits in the x-y plane and has side $2L$, and call the transverse displacement (in the z direction)

$$\psi(x, y, t) \quad \text{for} \quad |x|, |y| \leq L\,. \tag{11.78}$$

In principle this is a forced oscillation problem. We could take the boundary condition at the origin to be

$$\psi(0, 0, t) = A\cos\omega t \tag{11.79}$$

and try to find ψ everywhere else.

To find ψ, we must know the boundary condition at the edges of the plate. This depends on the details of the physics of the plate, because there are several ways that the plate can deform in response to the driving force. Just for simplicity, we will assume that the dominant deformation is shear, illustrated in figure 11.14. For this kind of displacement, to avoid an infinite acceleration, the slope of the plate must go to zero on the boundary in the direction perpendicular to the boundary, or in mathematics,

$$\hat{n} \cdot \vec{\nabla}\psi = 0 \tag{11.80}$$

on the edge, where $\hat{n}$ is a unit vector in the plane perpendicular to the edge. In this case,

$$\frac{\partial}{\partial x}\,\psi(x, y, t)|_{x=|L|} = \frac{\partial}{\partial y}\,\psi(x, y, t)|_{y=|L|} = 0\,. \tag{11.81}$$

While the general case is more complicated than this, we will use (11.81) for illustration. The instructive thing about Chladni plates, as we will see, is not what is happening at the edges, but what is happening in the middle!

The general solution to this forced oscillation problem is not easy to write down. However, we are primarily interested in the resonances. Those are the modes of free oscillation of

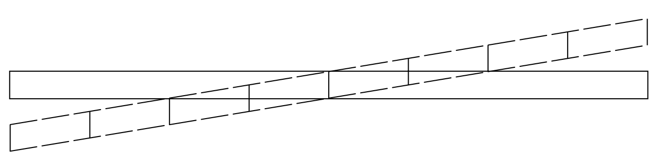


Figure 11.14: Shear.

the plate (subject to the boundary condition (11.81)) that can be excited by the driving force. These will be those modes that have nonzero values of the displacement at the origin.

The relevant free oscillation modes of the plate have the form[5]

$$\psi_{(n_x,n_y)}(x,y,t) = A\cos\frac{n_x\pi x}{L}\,\cos\frac{n_y\pi y}{L}\,\cos\omega t \tag{11.82}$$

with

$$\omega^2 = \omega_0^2(\vec{k}^2) \Rightarrow \omega^2 = f(n_x^2+n_y^2)\,. \tag{11.83}$$

If the frequencies of these modes were unique, (11.82) would be the whole story. But the interesting thing about this system is that the symmetry guarantees that there is **degeneracy** — that is that if $n_x \neq n_y$, there are two modes with the same frequency. We can get a physically equivalent mode by interchanging $n_x \leftrightarrow n_y$, because this just corresponds to a 90° rotation of the plate, which doesn't change the physics at all. When we have degenerate modes, then linear combinations of them are also modes, as shown in (3.117). Thus we have to ask **which linear combinations are excited by the driving force?** Another way of saying this is summarized in (11.83). Rotation invariance ensures that ω^2 depends only on $n_x^2+n_y^2$.

In particular, it is clear that the difference

$$\psi^-_{(n_x,n_y)}(x,y,t) = A\left(\cos\frac{n_x\pi x}{L}\,\cos\frac{n_y\pi y}{L} - \cos\frac{n_y\pi x}{L}\,\cos\frac{n_x\pi y}{L}\right)\cos\omega t \tag{11.84}$$

vanishes at the origin. **Only the sum couples to the driving force!**

$$\psi^+_{(n_x,n_y)}(x,y,t) = A\left(\cos\frac{n_x\pi x}{L}\,\cos\frac{n_y\pi y}{L} + \cos\frac{n_y\pi x}{L}\,\cos\frac{n_x\pi y}{L}\right)\cos\omega t \tag{11.85}$$

These are the resonant modes of a Chladni plate.

One reason that this is amusing is that it is easy to see. If you excite the plate, and sprinkle sand on it, the sand builds up in the regions where the plate is not moving — along the displacement nodes where $\psi = 0$. Thus we can get a visual picture of the zeros of ψ. Let's look at some of these modes (in order of increasing frequency) to see what to expect.

The mode $\psi^+_{(0,0)}$ is not interesting. It corresponds to the whole plate going up and down as a block. Obviously, the corresponding frequency is 0, because there is no restoring force. The first interesting mode is

$$\psi^+_{(1,0)}(x,y,t) = A\left(\cos\frac{\pi x}{L} + \cos\frac{\pi y}{L}\right)\cos\omega t\,. \tag{11.86}$$

This vanishes for

$$y = \pm L \pm x \tag{11.87}$$

so the Chladni sand pattern looks like the diagram in figure 11.15.

[5]There are also modes proportional to $\sin{(n_x+1/2)\pi x/L}$ and/or $\sin{(n_y+1/2)\pi y/L}$, but these vanish at the origin and are not excited by the driving force.

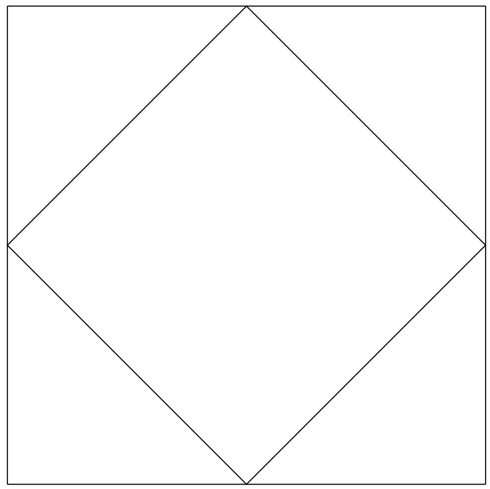


Figure 11.15: The Chladni pattern for the mode $(n_x, n_y) = (1, 0)$.

The next mode is

$$\psi^{+}_{(1,1)}(x, y, t) = 2A \cos\frac{\pi x}{L} \, \cos\frac{\pi y}{L} \, \cos\omega t\,. \tag{11.88}$$

Because this mode is not degenerate, it does not give rise to a very interesting pattern. It vanishes at

$$x = \pm\frac{L}{2} \quad \text{or} \quad y = \pm\frac{L}{2}\,, \tag{11.89}$$

which gives the pattern shown in figure 11.16. We won't consider any more of these boring modes with $n_x = n_y$.

The next mode is

$$\psi^{+}_{(2,0)}(x, y, t) = A \left(\cos\frac{2\pi x}{L} \, + \cos\frac{2\pi y}{L}\right) \cos\omega t\,, \tag{11.90}$$

which vanishes for

$$y = \pm\frac{L}{2} \pm x \quad \text{or} \quad y = \pm\frac{3L}{2} \pm x \tag{11.91}$$

so the pattern looks like figure 11.17.

Next comes

$$\psi^{+}_{(2,1)}(x, y, t) = A \left(\cos\frac{\pi x}{L} \, \cos\frac{2\pi y}{L} + \cos\frac{2\pi x}{L} \, \cos\frac{\pi y}{L}\right) \cos\omega t\,. \tag{11.92}$$

This vanishes for

$$\begin{aligned} c_x\,(2\,c_y^2 - 1) + c_y\,(2\,c_x^2 - 1) &= 0 \\ = (c_x + c_y)\,(2\,c_x\,c_y - 1) &= 0 \end{aligned} \tag{11.93}$$

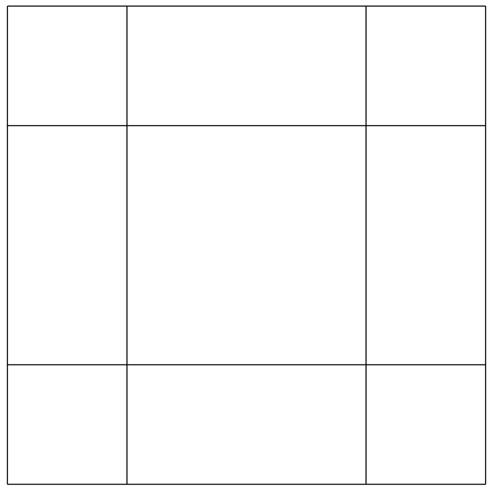


Figure 11.16: The Chladni pattern for the mode (1,1).

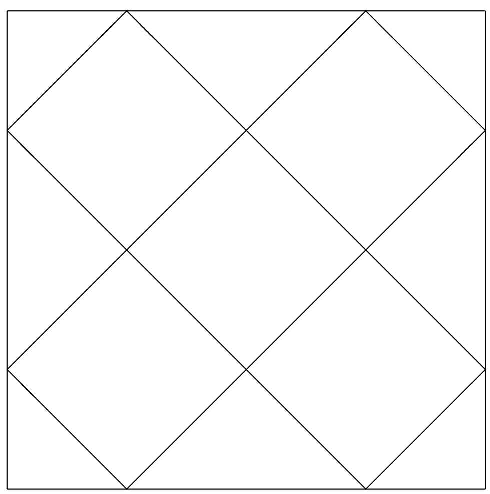


Figure 11.17: The Chladni pattern for the mode (2,0).

with $c_x \equiv \cos(\pi x/L)$ and $c_y \equiv \cos(\pi y/L)$. The pattern is shown in figure 11.18.

We could go on, but you should have the idea by now. Let us look at one last mode:

$$\psi^+_{(3,1)}(x,y,t) = A\left(\cos\frac{\pi x}{L}\,\cos\frac{3\pi y}{L} + \cos\frac{3\pi x}{L}\,\cos\frac{\pi y}{L}\right)\cos\omega t\,, \tag{11.94}$$

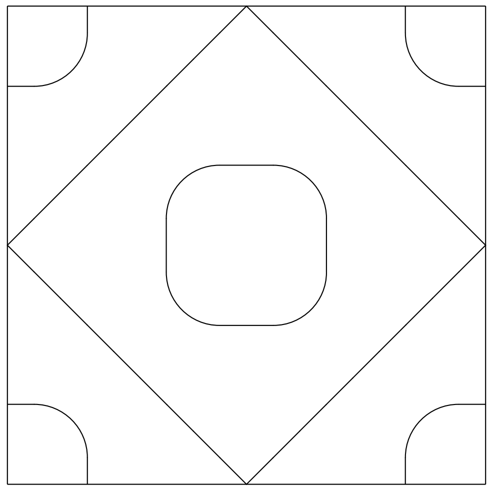


Figure 11.18: The Chladni pattern for the mode (2,1).

vanishing for

$$\begin{aligned} c_x\,(4\,c_y^3 - 3\,c_y) + c_y\,(4\,c_x^3 - c_x) = 0 \\ = c_x\,c_y\,(4\,c_x^2 + 4\,c_y^2 - 6) = 0 \end{aligned} \tag{11.95}$$

with pattern shown in figure 11.19.

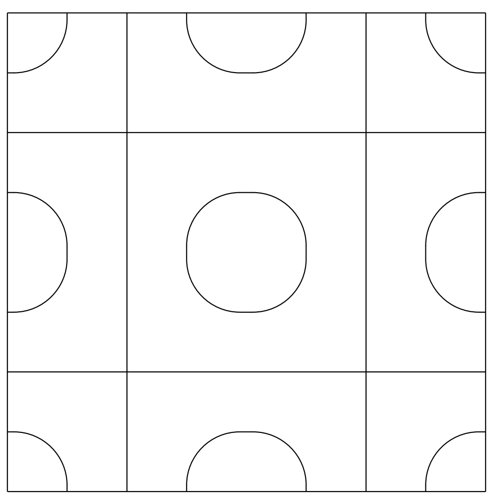


Figure 11.19: The Chladni pattern for the mode (3,1).

Moral: When there is more than one mode with the same frequency, look at linear combinations to determine which are excited!

11.4 Waveguides

Generically, a "waveguide" is a device that forces a traveling wave to propagate only where you want it to go. Typically, a waveguide is some kind of tube that allows the wave disturbance to propagate in one direction while confining it in the other directions. In this section, we will discuss the case of straight wave guides with simple uniform cross sections. The really interesting physics occurs when the width of the waveguide is not much larger than the wavelength of the wave. Then, as we will see, the physics of the waveguide has a dramatic effect on the propagation of the wave.

The simplest situation to discuss is the case of transverse oscillations of a membrane in the form of an infinite strip, as shown in figure 11.20. Consider a membrane with surface

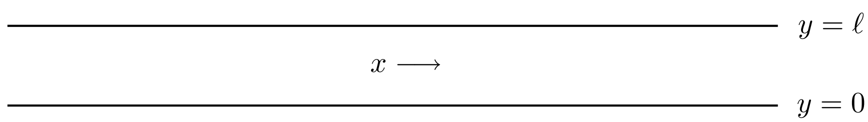


Figure 11.20: A section of an infinite strip of stretched membrane that acts as a waveguide.

mass density ρ_s and surface tension T_s, stretched in an infinite strip in the x-y plane between $y = 0$ and $y = \ell$ and from $x = -\infty$ to ∞. The edges, at $y = 0$ and $y = \ell$ are held fixed in the plane. We are interested in the oscillations of the interior of the strip up and down out of the plane.

This is a job for separation of variables. We can look for modes of this system which are products of a function of x and a function of y. In particular, we can satisfy the boundary conditions at $y = 0$ by combining two modes of the infinite system,

$$e^{ik_x x} e^{ik_y y} \text{ , and } e^{ik_x x} e^{-ik_y y} \tag{11.96}$$

into

$$\sin(k_y y)\, e^{ik_x x} \,. \tag{11.97}$$

Now this satisfies the boundary condition at $y = \ell$ if

$$k_y = \frac{n\pi}{\ell} \text{ for } n = 1 \text{ to } \infty \,. \tag{11.98}$$

Thus the modes look like this:

$$\psi_{n+}(x, y, t) = A \, \sin \frac{n\pi y}{\ell} \, e^{i(k_x x - \omega t)} \,, \tag{11.99}$$

and

$$\psi_{n-}(x,y,t) = A\,\sin\frac{n\pi y}{\ell}\,e^{i(-k_x x-\omega t)}\,. \tag{11.100}$$

For each value of n, these look like waves traveling in the $\pm x$ direction!

The dispersion relation for the membrane is given by (11.18). But the modes, $\psi_{n\pm}$, have $|k_y| = \frac{n\pi}{\ell}$. Thus the dispersion relation for the traveling waves, (11.99) and (11.100) is

$$\omega^2 = v^2 k_x^2 + \omega_n^2\,, \tag{11.101}$$

where

$$v = \sqrt{\frac{T_s}{\rho_s}} \tag{11.102}$$

and

$$\omega_n = \frac{n\pi v}{\ell}\,. \tag{11.103}$$

One interesting thing about (11.101) is that the dispersion relation has a low frequency cut-off that depends on n. For any given ω, the only modes that actually propagate are the finite number of modes with

$$n < \frac{\omega\ell}{\pi v}\,. \tag{11.104}$$

For example, for $\omega \le \pi v/\ell$, there are no traveling waves. For $\pi v/\ell < \omega \le 2\pi v/\ell$, there is only one, corresponding to $n = 1$, etc.

The modes satisfying (11.104) have a simple physical interpretation. They can be thought of as the plane waves, (11.96), of the infinite system, bouncing back and forth between the fixed edges, $y = 0$ and $y = \ell$. The requirement, (11.98), on the allowed values of k_y arises because for other values of k_y, the reflected waves get out of phase, giving destructive interference. You might expect a zig-zag wave of this kind to propagate in the x direction with a speed less than the phase velocity, v, of the waves in the infinite system by a factor of

$$\frac{k_x}{\sqrt{k_x^2+k_y^2}} = \frac{k_x}{\sqrt{k_x^2+(\omega_n/v)^2}}\,, \tag{11.105}$$

because it has to go that much farther as it bounces back and forth to move a given distance in x, as illustrated in figure 11.21. In fact, the phase velocity of the zig-zag waves for fixed

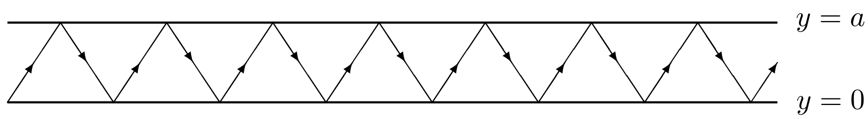


Figure 11.21: A zig-zag wave in the waveguide.

n, ω/k_x, is actually **larger** than v by the factor, (11.105), rather than smaller,

$$v_{n\phi} = \frac{\omega}{k_x} = v\frac{\sqrt{k_x^2 + (\omega_n/v)^2}}{k_x}\,. \tag{11.106}$$

However, the group velocity, $\partial\omega/\partial k_x$, of the zig-zag waves, the velocity with which you can actually send signals, is smaller by just the expected factor,

$$v_{gn} = \frac{\partial\omega}{\partial k_x} = v\frac{k_x}{\sqrt{k_x^2 + (\omega_n/v)^2}}\,. \tag{11.107}$$

For light waves, we can make a wave guide by making a tube of some conducting material, so that the electric field is nonzero only inside the tube. However, in this case, the details of the boundary conditions at the edges depend on the direction of the electic field. We will return to a related question in the next chapter.

11.5 Water

Water is pretty complicated stuff. It wets things. It has viscosity. It forms whirlpools and eddies and has nonlinear turbulent motions that we cannot hope to understand using the techniques that we have at our disposal. In this section, we consider a somewhat idealized fluid, that we will call "dry water" (after Feynman) that has none of this complicated structure. It has three features that we will keep in common with the real thing. It has mass density. It has surface tension, and it is nearly incompressible. Let's see how it waves.

Imagine an infinite universe full of an incompressible, frictionless liquid. This will allow us to see the consequences of the incompressibility in a simple, qualitative way. Consider the analog of a plane sound wave in such a system. That is, for example, a plane wave traveling in the x direction (with $k_y = k_z = 0$) with longitudinal displacements in the x direction. If the liquid is truly incompressible, the k_x must be zero for this wave, because any longitudinal displacement must be accompanied by compressions and rarefactions of the medium. Thus, for such a plane wave, $\vec{k} = 0$. **There are no nontrivial plane waves in the infinite system!** In general, we do not expect that all the components of the k vector must vanish, because even in an incompressible liquid, displacement in one direction is allowed if it is accompanied by appropriate motion in other directions. But what we have seen is that we cannot have a mode that has a real $\vec{k}$ vector. That would be a plane wave, which we have seen is not compatible with incompressibility. Instead, we expect that the constraint $k_x = 0$ will be replaced by a constraint on the rotation invariant length of the k vector, that $\vec{k}\cdot\vec{k} = 0$. If some of the components of the $\vec{k}$ vector are imaginary, this can be satisfied for nonzero $\vec{k}$.

Note that the condition $\vec{k}\cdot\vec{k} = 0$ is not exactly a dispersion relation, because it makes no reference to frequency. But it is the whole story for an infinite system of incompressible fluid. In fact, it is clear that there are no harmonic waves in the infinite system, because there is nothing to produce a restoring force. Even if there is a gravitational field, the pressure in the liquid just adjusts itself to cancel the effect of gravity. We can get a nontrivial dispersion relation only when there is a surface. The dispersion relation then depends on the physics of the surface. This would seem to violate our general principle that the dispersion relation is a property of the infinite system. What is happening is this. The relation, $\vec{k}\cdot\vec{k}=0$ is really the only dispersion relation that makes any sense for the three-dimensional infinite system. When we introduce a surface, we have **broken** the translation invariance in the direction normal to the surface. This allows us to get a nontrivial dispersion relation for the two-dimensional system parallel to the surface.

11.5.1 Mathematics of Water Waves

Now let us try to make these considerations quantitative. As usual, we will label our fluid in terms of the equilibrium positions of its parts. Then call the displacement from equilibrium of the fluid that is at the point $\vec{r}$ at equilibrium

$$\epsilon\,\vec{\psi}(\vec{r},t) \tag{11.108}$$

for some small ϵ. This means that the actual position of the water is[6]

$$\vec{R}(\vec{r},t)=\vec{r}+\epsilon\,\vec{\psi}(\vec{r},t)\,. \tag{11.109}$$

We can regard (11.109) as a kind of change of coordinates. It maps us from the equilibrium coordinates (a rather arbitrary label because the water is free to flow) to the physical coordinates that tell us where the water actually is. If the water is incompressible, which is a pretty good approximation, then a small volume element should have the same volume in equilibrium and in the physical coordinates.

$$dR_x\,dR_y\,dR_z=dx\,dy\,dz\,. \tag{11.110}$$

This will be the case if the determinant of the Jacobian matrix equals 1:

$$\det\begin{pmatrix}\frac{\partial R_x}{\partial x} & \frac{\partial R_x}{\partial y} & \frac{\partial R_x}{\partial z}\\ \frac{\partial R_y}{\partial x} & \frac{\partial R_y}{\partial y} & \frac{\partial R_y}{\partial z}\\ \frac{\partial R_z}{\partial x} & \frac{\partial R_z}{\partial y} & \frac{\partial R_z}{\partial z}\end{pmatrix}=1\,. \tag{11.111}$$

Because ϵ is small, we can expand (11.111) to lowest order in ϵ,

$$\begin{aligned}&=\det\begin{pmatrix}1+\epsilon\,\frac{\partial\,\psi_x}{\partial x} & \epsilon\,\frac{\partial\,\psi_x}{\partial y} & \epsilon\,\frac{\partial\,\psi_x}{\partial z}\\ \epsilon\,\frac{\partial\,\psi_y}{\partial x} & 1+\epsilon\,\frac{\partial\,\psi_y}{\partial y} & \epsilon\,\frac{\partial\,\psi_y}{\partial z}\\ \epsilon\,\frac{\partial\,\psi_z}{\partial x} & \epsilon\,\frac{\partial\,\psi_z}{\partial y} & 1+\epsilon\,\frac{\partial\,\psi_z}{\partial z}\end{pmatrix}\\ &=1+\epsilon\,\vec{\nabla}\cdot\vec{\psi}+\mathcal{O}(\epsilon^2)\,.\end{aligned} \tag{11.112}$$

[6]Here we can take ψ to be dimensionless and let the parameter, ϵ, be a small displacement.

Thus

$$\vec{\nabla} \cdot \vec{\psi} = 0\,. \tag{11.113}$$

(11.113) is very reasonable. It is the statement that the flux of displacement into or out of any region vanishes.[7] This is what we expected from our qualitative discussion.

To see what this means for waves, let us also assume that there are no eddies. The mathematical statement of this is

$$\vec{\nabla} \times \vec{\psi} = 0\,. \tag{11.114}$$

If we do not assume (11.114), angular momentum conservation becomes important and life becomes very complicated. You will have to wait for courses on fluid dynamics to learn more about it. With the simplifying assumption, (11.114), the displacement can be written as the gradient of a scalar function, χ,

$$\epsilon\vec{\psi} = \epsilon\nabla\chi\,. \tag{11.115}$$

This simplifies our life enormously, because we can now deal with the scalar quantity, χ. Space translation invariance tells us that we can find modes of the form

$$\chi = e^{i\vec{k}\cdot\vec{r}-i\omega t}, \tag{11.116}$$

which gives a displacement of the form

$$\epsilon\vec{\psi} = i\,\epsilon\,\vec{k}\,e^{i\vec{k}\cdot\vec{r}-i\omega t}. \tag{11.117}$$

The condition, (11.113) then becomes

$$\vec{k}\cdot\vec{k} = 0\,, \tag{11.118}$$

as anticipated in our qualitative discussion at the beginning of the section.

11.5.2 Depth

11-3

Let us now consider waves in an "ocean" of depth L, ignoring frictional forces, eddies and nonlinearities. We will further restrict our attention to a two-dimensional situation. Let y be the vertical direction, and consider water waves in the x direction. That is, we will take k_x real, because we are interested in wave propagation in the x direction, and k_y pure imaginary with the same magnitude, so that (11.118) is satisfied. Then we assume that nothing depends on the other coordinate, z. Having simplified things this far, we may as well assume that our ocean is a rectangular box. Then the modes of interest of the infinite system look like

$$\chi_\infty(x,y,t) = e^{\pm ikx\pm ky-i\omega t}. \tag{11.119}$$

[7] Note, however, that for large ϵ, incompressibility is the nonlinear constraint, (11.111).

If the ocean has a bottom at $y = 0$, then the vertical displacement must vanish at $y = 0$. Then (11.115) implies that we must combine modes of the infinite system to get a χ whose y derivative vanishes at $y = 0$, to get

$$\chi(x, y, t) \propto e^{\pm ikx - i\omega t} \cosh ky \,. \tag{11.120}$$

where $\cosh$ is the "hyperbolic cosine." defined by

$$\cosh x \equiv \frac{e^x + e^{-x}}{2} \,. \tag{11.121}$$

Then from (11.115), we get

$$\begin{aligned} \psi_x(x, y, t) &= \frac{\partial}{\partial x}\chi(x, y, t) = \pm i\, e^{\pm ikx - i\omega t} \cosh ky \,, \\ \psi_y(x, y, t) &= \frac{\partial}{\partial y}\chi(x, y, t) = e^{\pm ikx - i\omega t} \sinh ky \,. \end{aligned} \tag{11.122}$$

Before going further, note that we could extend these considerations by adding a z coordinate. Then (11.120) would become

$$\chi(x, y, t) \propto e^{(\pm ik_x x \pm ik_z z) - i\omega t} \cosh ky \tag{11.123}$$

where

$$k = \sqrt{k_x^2 + k_z^2} \,. \tag{11.124}$$

These are the two-dimensional wave modes of the infinite ocean of depth L. The y dependence is completely fixed by the boundary condition at the bottom and the condition $\vec{k} \cdot \vec{k} = 0$. The only interesting dependence, from the point of view of space translation invariance, is the dependence on x and z.

Now, let us return to the rectangular ocean, and the z-independent modes, (11.122). If our ocean has sides at $x = 0$ and $x = X$, we must choose linear combinations of the modes, (11.122), such that the x displacement vanishes at the sides. We can do this for $x = 0$ by forming the combinations

$$\begin{aligned} \psi_x(x, y, t) &= -\sin kx \cosh ky \, \cos \omega t \,, \\ \psi_y(x, y, t) &= \cos kx \sinh ky \, \cos \omega t \,. \end{aligned} \tag{11.125}$$

Then if

$$k = \frac{n\pi}{X} \,, \tag{11.126}$$

the boundary condition at $x = X$ is satisfied as well.

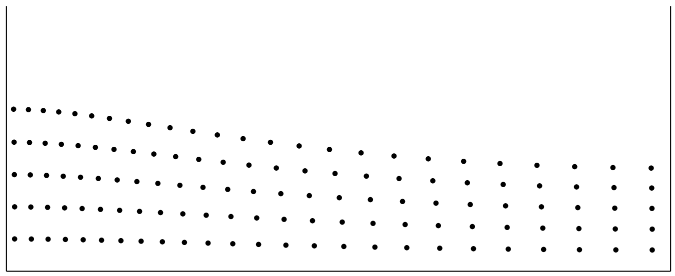


Figure 11.22: The motion of an incompressible fluid in a wave.

Now we know the mathematics of the displacement of the dry water. Before we go on to discuss the dispersion relation, let us pause to consider what this actually looks like. Imagine that we put a regular rectangular grid of points in the water in equilibrium. Then in figure 11.22, we show what the grid looks like in the mode, (11.125) with $n = 1$.

Each of the little rectangles in (11.22) was a square in equilibrium position (when $\psi = 0$). Note the way incompressibility works. When the water is squeezed in one direction, it is stretched in the other. You can see this in motion in program 11-3.

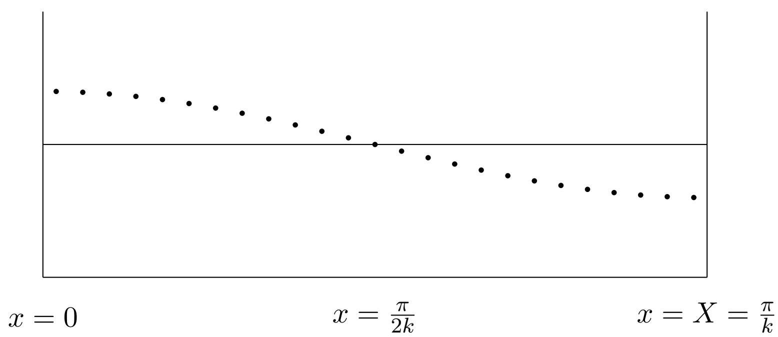


Figure 11.23: The surface of a water wave, with horizontal displacement suppressed.

Having stared at this, we can now forget about it for a while, and concentrate just on the surface. That is what matters for the dispersion relation. For ease of presentation in the diagrams below, we will exaggerate the displacement in the vertical y direction and forget about the displacement of the surface in the x direction (which won't matter anyway). Then the wave looks like the picture in figure 11.23. We will use energy arguments to get the dispersion relation. There are three contributions to the total energy of the standing wave, (11.125) — gravitational potential energy, energy stored in surface tension, and kinetic energy. Let us consider them in turn.

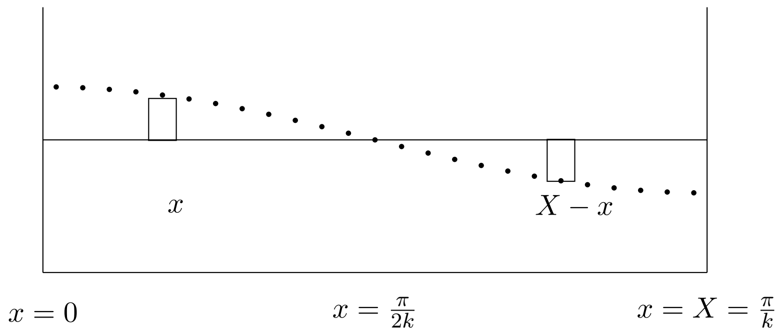


Figure 11.24: Water is removed from the rectangle in $X-x$ and raised to the rectangle at x.

Gravitational Potential

In the diagram in figure 11.24, you can see that the overall effect of the displacements in the mode (11.125) is to take a chunk of the water from $X-x$, raise it by $\epsilon\psi_y(x,L,t)$ (the vertical displacement of the surface), and move it over to x. The volume of this chunk is $W\,dx\,\epsilon\psi_y(x,L,t)$ where dx is the length of chunk and W is the width in the z direction (into the paper). Thus the total gravitational potential is

$$\begin{aligned} V_{\rm grav} &= \rho\, g \int dV\,\Delta h = \rho\, g\, W \int_0^{\frac{\pi}{2k}} dx\, |\epsilon\,\psi_y(x,L,t)|^2 + \mathcal{O}(\epsilon^3) \\ &= \rho\, g\, W \int_0^{\frac{\pi}{2k}} dx\, \epsilon^2\, \cos^2 kx \sinh^2 kL \cos^2 \omega t + \cdots \\ &= \frac{\pi}{4k}\,\rho\, g\, W\, \epsilon^2\, \sinh^2 kL \cos^2 \omega t + \cdots\,. \end{aligned} \tag{11.127}$$

Surface Tension

The energy stored in surface tension is W times the difference between the length of the surface and the equilibrium length (X). This requires that we be a little careful about the position of the surface, going back to (11.109). The position of the surface is

$$R_x(x,t) = x + \epsilon\psi_x(x,L,t)\,, \quad R_y(x,t) = \epsilon\psi_y(x,L,t)\,. \tag{11.128}$$

The length is then

$$\int_0^X dx\,\sqrt{\frac{\partial R_x}{\partial x}^2 + \frac{\partial R_y}{\partial x}^2}\,. \tag{11.129}$$

But

$$\frac{\partial R_x}{\partial x} = 1 + \epsilon \frac{\partial}{\partial x}\psi_x\,, \quad \frac{\partial R_y}{\partial x} = \epsilon \frac{\partial}{\partial x}\psi_y\,. \tag{11.130}$$

Thus

$$\begin{aligned} V_{\text{surface}} &= T \times (\text{Area} - \text{Area}_0) \\ &= T\,W \int_0^{\frac{\pi}{k}} dx \left(\sqrt{(1+\epsilon\partial\psi_x/\partial x)^2 + (\epsilon\partial\psi_y/\partial x)^2} - 1 \right) \\ &= T\,W \int_0^{\frac{\pi}{k}} dx \left(\epsilon\partial\psi_x/\partial x + \frac{1}{2}(\epsilon\partial\psi_y/\partial x)^2 + \mathcal{O}(\epsilon^3) \right)\,. \end{aligned} \tag{11.131}$$

The order ϵ term in (11.131) cancels when integrated of x, so

$$\begin{aligned} &= T\,W\,\epsilon^2 \int_0^{\frac{\pi}{k}} dx\,\frac{1}{2}\,k^2\,\sin^2 kx \sinh^2 kL \cos^2 \omega t + \cdots \\ &= \frac{\pi}{4k}\,T\,W\,\epsilon^2\,k^2\,\sinh^2 kL \cos^2 \omega t + \cdots\,. \end{aligned} \tag{11.132}$$

Kinetic Energy

The kinetic energy is obtained by integrating $\frac{1}{2}mv^2$ over the whole volume of the liquid:

$$\begin{aligned} KE &= \frac{1}{2}\,\rho \int dV\,\vec{v}^2 \\ &= \frac{1}{2}\,\rho\,W \int_0^{\frac{\pi}{k}} dx \int_0^L dy\,\left((\epsilon\partial\psi_x/\partial t)^2 + (\epsilon\partial\psi_y/\partial t)^2 \right) \end{aligned} \tag{11.133}$$

$$\begin{aligned} &= \frac{1}{2}\,\rho\,W\,\epsilon^2 \int_0^{\frac{\pi}{k}} dx \int_0^L dy\,\omega^2\,\sin^2 \omega t \\ &\cdot \left(\cos^2 kx \sinh^2 ky + \sin^2 kx \cosh^2 ky \right) \end{aligned} \tag{11.134}$$

$$\begin{aligned} &= \frac{\pi}{4k}\rho\,W\,\epsilon^2 \int_0^L dy\,\omega^2\,\sin^2 \omega t\,\left(\sinh^2 ky + \cosh^2 ky \right) \\ &= \frac{\pi}{4k}\rho\,W\,\epsilon^2 \int_0^L dy\,\omega^2\,\sin^2 \omega t\,\cosh 2ky \\ &= \frac{\pi}{8k^2}\rho\,W\,\epsilon^2\,\omega^2\,\sinh 2kL\,\sin^2 \omega t\,. \end{aligned} \tag{11.135}$$

Dispersion Relation

The total of (11.127)-(11.135) is

$$\begin{aligned} V_{\text{grav}} + V_{\text{surface}} + KE &= \frac{\pi}{4k}\,\rho\,g\,W\,\epsilon^2\,\sinh^2 kL \cos^2 \omega t \\ +\frac{\pi}{4k}\,T\,W\,\epsilon^2\,k^2\,\sinh^2 kL \cos^2 \omega t &+ \frac{\pi}{8k^2}\rho\,W\,\omega^2\,\epsilon^2\,\sinh 2kL\,\sin^2 \omega t + \cdots\,. \end{aligned} \tag{11.136}$$

This must be constant in time, which implies

$$\begin{aligned}\omega^2 &= \frac{2\sinh^2 kL\left(gk + \frac{T}{\rho}\,k^3\right)}{\sinh 2kL} \\ &= \left(gk + \frac{T}{\rho}\,k^3\right)\tanh kL\end{aligned} \tag{11.137}$$

where $\tanh$ is the "hyperbolic tangent," defined by

$$\tanh x \equiv \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}\,. \tag{11.138}$$

Note that in the twin limit of long wavelength and shallow water, the water waves become nondispersive — for $kL \ll 1$, and $\rho g k \gg T\,k^3$ — $\tanh kL \to kL$

$$\omega^2 \approx gL\,k^2\,. \tag{11.139}$$

Gravity versus Surface Tension

The dispersion relation, (11.139), involves a competition between gravity and surface tension. For long wavelengths gravity dominates and the gk term is most important. For short wavelengths, surface tension dominates and the $\frac{Tk^3}{\rho}$ term is more important. The cross-over occurs for wave numbers of order

$$k \approx k_0 = \sqrt{\frac{\rho\, g}{T}}\,. \tag{11.140}$$

The cross-over wavelength is actually a familiar distance. There is a much more familiar process that involves a similar competition between gravity and surface tension. Consider a water drop on a low friction surface, such as a teflon frying pan. A very tiny drop is nearly spherical. But as the size of the drop increases, it begins to flatten out. Then when the drop increases above a critical size, the height of the drop does not increase. It spreads out with a fixed height, h, as shown in cross-section in figure 11.25.

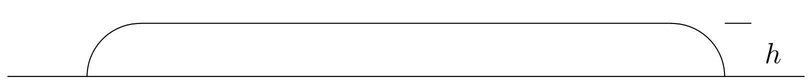


Figure 11.25: The cross-section of a water droplet on a frictionless surface.

As with the dispersion relation, we can understand what is going on by considering the energy. The total energy of the drop is a sum of the gravitational potential energy and the energy due to surface tension.

$$V_{\rm grav} \approx \frac{1}{2}\,\rho\, g\, h\, v\,, \tag{11.141}$$

where v is the volume of the drop and

$$V_{\text{surface}} \approx \frac{T\,v}{h}\,. \tag{11.142}$$

The volume is fixed, so the equilibrium value of h minimizes the sum

$$V_{\text{grav}} + V_{\text{surface}} \approx \frac{1}{2}\,\rho\,g\,h\,v + \frac{T\,v}{h}\,. \tag{11.143}$$

The minimum occurs for

$$T = \frac{1}{2}\,\rho\,g\,h^2\,. \tag{11.144}$$

The measured surface tension of water is $T \approx 72$ dynes/cm. This gives the familiar height of a water drop, $h \approx$ 0.4 cm. This height is related to k_0 by

$$k_0 = \sqrt{\frac{\rho\,g}{T}} \approx \frac{\sqrt{2}}{h}\,. \tag{11.145}$$

11.6 Lenses and Geometrical Optics

Geometrical Optics

The idea of geometrical optics is to understand the effects of refraction and reflection on beams of light, ignoring the effects of diffraction. This is really only Snell's law and geometry. One application of these ideas will be in the discussion of the rainbow in the next section. There we use what is called "ray tracing" which as the name suggests is simply keeping track of what each ray of light does as it passes through the drop. A spherical drop is a "thick lens." Obviously, there is no sense in which a sphere could be regarded as "thin." In this section we are going to see how to give a simpler approximate description of what a "thin lens" does. In fact, if we were designing a very precise optical instrument, we would still use ray tracing to get the fine details right. But the thin lens analysis is a good approximate starting point and will help us understand what is happening in some important situations.

Tecnically, what "thin" means in this context is that if a narrow beam of light approximately perpendicular to the plane of the lens comes into the lens at some point on one side, it comes out at about the same point on the other side. If we ignore the small change in position, this simplifies the analysis and gives us the thin lens formula.

Thin Spherical Lenses

In Chapter 11, we derive the formula for the angular change in a narrow (we are ignoring diffraction) beam of light due to a prism. The analysis is uses the geometrical construction

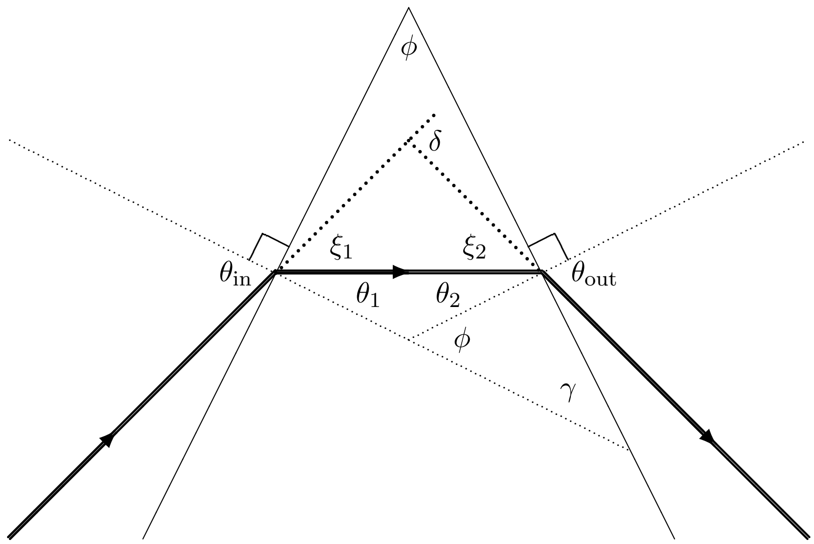


Figure 11.26:

shown in figure 11.26 and gives

$$\begin{aligned} \delta &= \theta_{\rm in} + \theta_{\rm out} - \theta_1 - \theta_2 \\ &\approx n\,(\theta_1 + \theta_2) - \phi \approx (n-1)\,\phi \end{aligned} \tag{11.146}$$

where the first is exact and the second follows in the limit in which the θ angles are small. In this limit, the angular deflection is independent of the incoming angle.

Thin lenses and small angles

We can use this result to understand how a lens focuses light. A lens is a device in which the angular change given to the beam is proportional to the distance from the axis for small angles and distances —

$$\delta \approx h/f \tag{11.147}$$

where f is length. This is approximately true for a piece of glass with surfaces that are parts of spheres. In figure 11.27 is a diagram showing how this works for a lens which is flat on one side and a partial sphere with radius r_1 on the other. In the diagram, θ_1 is the angle of the "effective prism" seen by the part of a beam at distance h from the axis. It should be clear from the figure that if θ_1 is small, it is proportional to h.

$$\theta_1 \approx \sin\theta_1 = \frac{h}{r_1} \tag{11.148}$$

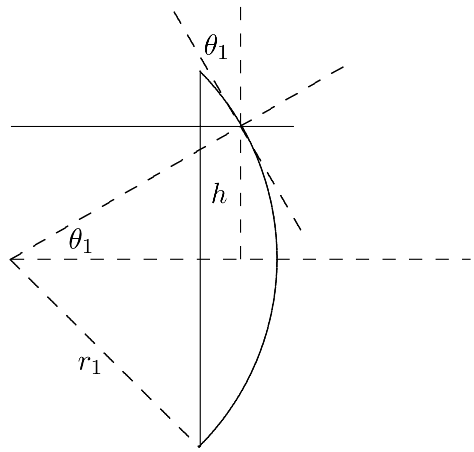


Figure 11.27:

More often, the lens is curved on both sides. If the radii are r_1 and r_2, the result looks like figure 11.28. Figure 11.28 shows the beam at the very tip of the lens for convenience, but as

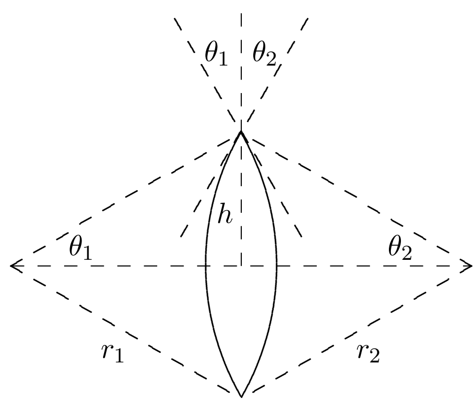


Figure 11.28:

the previous diagram should make clear, $\theta_1 + \theta_2$ is the "effective prism" angle for any h. The figure also exaggerates the curvature of the two sides, so that the lens pictured is not really "thin." A thin lens looks more like figure 11.29. This is important because if the lens is fat, the height h is not very well-defined because if the light inside the lens is not horizontal, we might have one h where the light enters the lens and a very different h where it come out. But if the lens is thin and if the light rays are not too far from the perpendicular, this ambiguity in

Figure 11.29:

h can be ignored just like other corrections to small angle relations (like $\sin\theta \approx \theta$).

Putting together the geometry from figure 11.28 with the formula for δ in a prism, we get the constant f for a thin spherical lens:

$$\begin{aligned} \delta &= (n-1)(\theta_1+\theta_2) \\ &\approx (n-1)\left(\frac{h}{r_1}+\frac{h}{r_2}\right) = \frac{h}{f} \end{aligned} \tag{11.149}$$

and thus

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right) \tag{11.150}$$

This is called the “lens-maker’s formula”

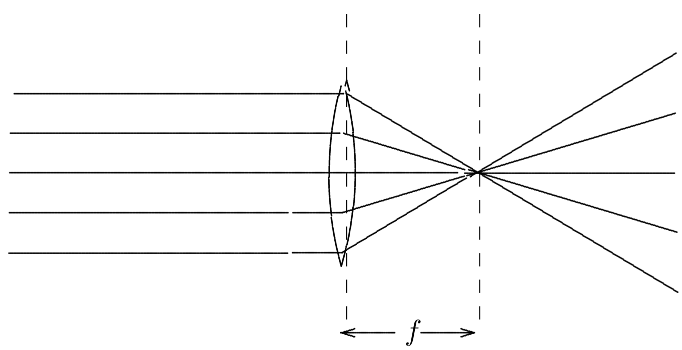


Figure 11.30:

A lens of this kind focuses parallel rays of light, as shown in figure 11.30. This works because $\delta \approx h/f$ as shown in figure 11.31. Parallel rays at any angle are focused onto a “focal plane” a distance f from the lens as shown in figure 11.32. The analytical way of explaining how this works is to note that the difference in the slopes of the rays on the two

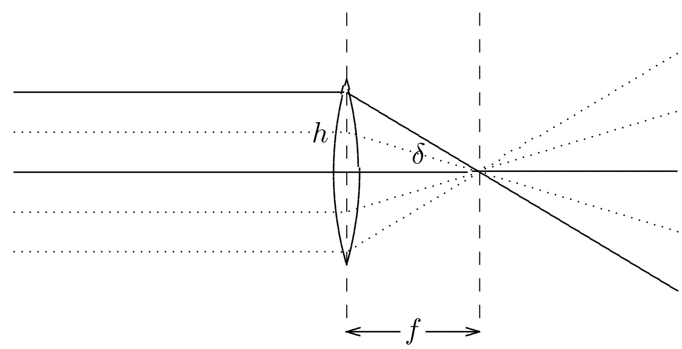


Figure 11.31:

sides of the lens is proportional to the height. Thus the in this case, because the slopes on one side are the same, the difference in slopes on the other side is proportional to the difference in height, and that means that they all come together at the same x.

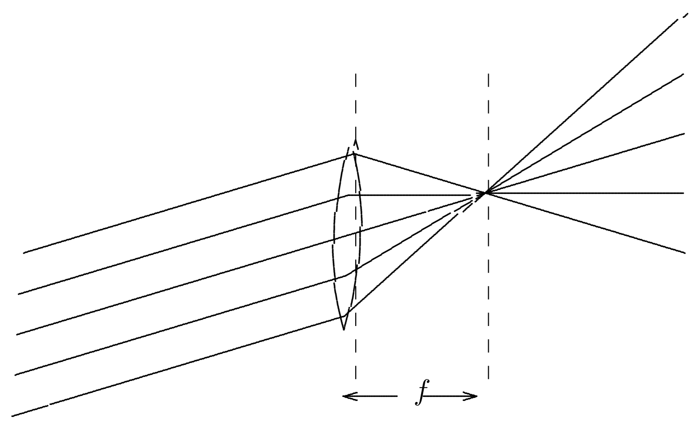


Figure 11.32:

Another way to see that this focusing must work is illustrated in figures 11.33 and 11.34. Note that if the parallel rays are coming in at an angle δ_i, the ray a distance $h_i = \delta f$ above the center of the lens is bent to the horizontal, as shown in figure 11.33 with the solid line. Then for the rays on either side of that ray (shown as dashed lines), because the dependence of the bending on the height in the lens is linear, the total angular bend, $\delta_i + \delta_o$ is f multiplied by the total distance from the center, $h_i + h_o$, but then $h_o = \delta_o f$, which is the condition for focusing. This is illustrated in figure 11.34.

For a bundle of parallel rays at any angle, you can determine where they hit the focal

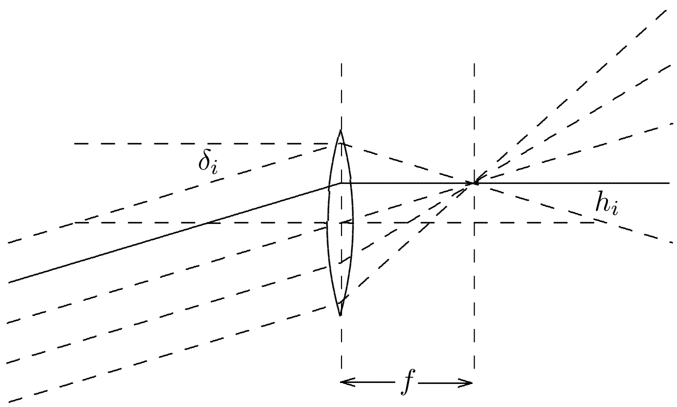


Figure 11.33:

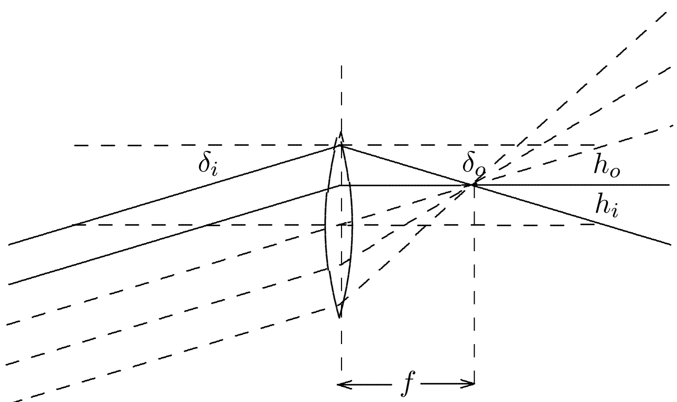


Figure 11.34:

plane by tracing any ray, the easiest being the one through the center of the lens, which is not bent at all, as shown in figure 11.35. The parallel rays (a part of a plane wave — we know this is impossible, but we are ignoring diffraction) can be thought of as coming from a point source at infinity. If there is a point source closer to the lens, it focuses farther away. Now play with the animation LENS.EXE.

To find the relation between d_1 and d_2, consider the diagram in figure 11.37 — the sum of the angles of deflection on the two sides equals δ:

$$\delta_1 + \delta_2 = \delta \tag{11.151}$$

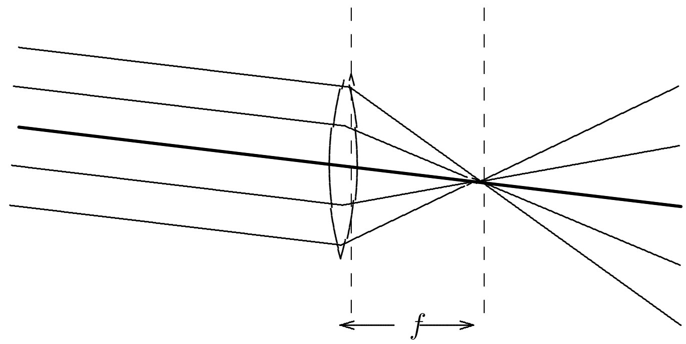


Figure 11.35:

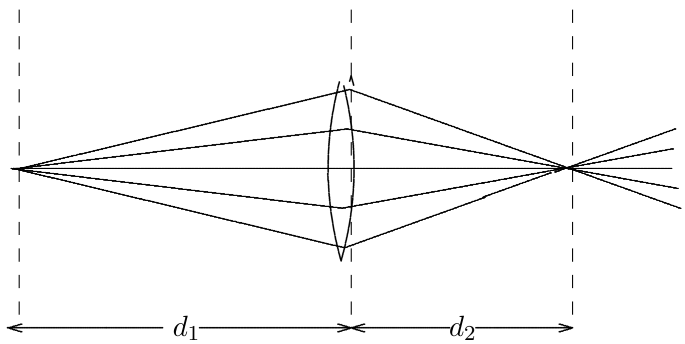


Figure 11.36:

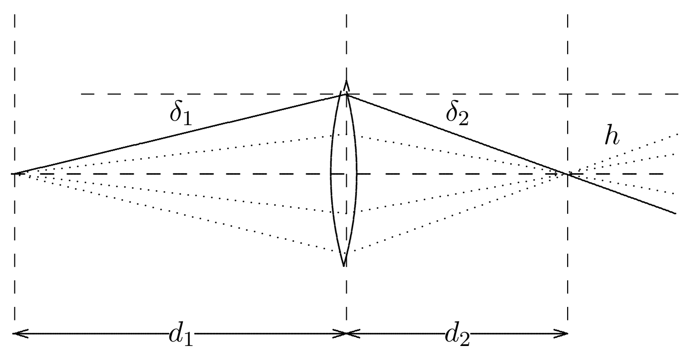


Figure 11.37:

which for small angles is equivalent to

$$\frac{h}{d_1} + \frac{h}{d_2} = \frac{h}{f} \tag{11.152}$$

or

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \tag{11.153}$$

This is called the "thin lens formula."

So far, we have discussed "converging" or "convex" lenses for which f is positive, but there are also "diverging" or "concave" lenses, for which f is negative. In this case, parallel rays are not focuses, but defocused, and appear to diverge from a plane a distance $-f$ (which is a positive number) beyond the lens, as shown in figure 11.38: The point from which the

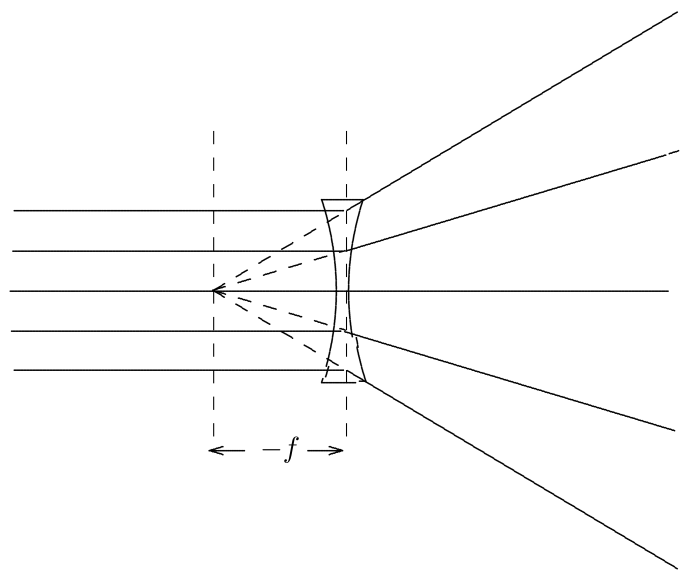


Figure 11.38:

outgoing rays diverge is called a "virtual image." In this case it is a virtual image of the point at infinity. Shown in figure 11.39 is the effect of a concave lens on a point source. Again there is a virtual image. Here the thin lens formula is still satisfied, but both f and d_2 are negative.

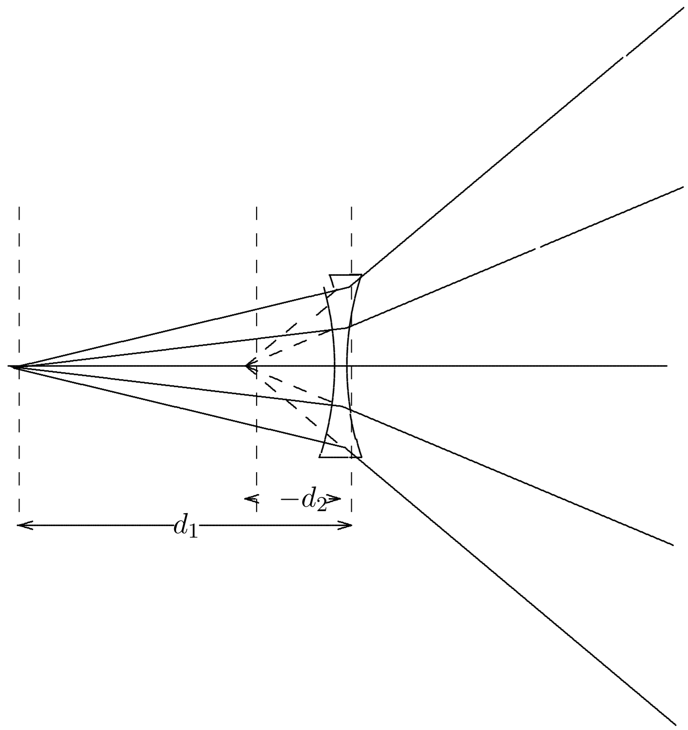


Figure 11.39:

Images

The focusing property of a lens can be used to project an image of an object on a surface, as shown in figure 11.40. What is happening is that light fanning out from each point on

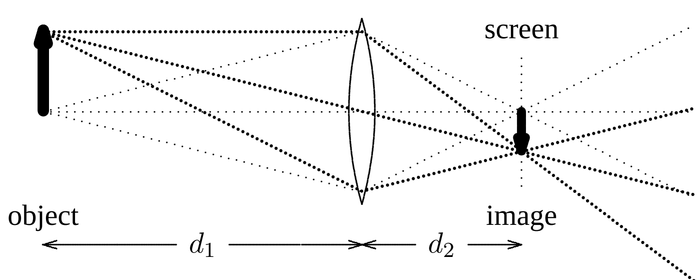


Figure 11.40:

the object is focused back to a single point on the screen. As in figures 11.36 and 11.37, the distances satisfy the thin lens formula,

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \tag{11.153}$$

This tell you where to put the screen. Note also that it is easy to see where on the screen the image of a particular point on the object appears because a ray of light that goes right through the center of the lens is not deflected at all (we also used this for parallel rays above figure 11.35). This plus simple geometry then implies that the ratio of the size of the image to the size of the object is d_2/d_1.

$$\frac{\text{size of image}}{\text{size of object}} = \frac{d_2}{d_1} \tag{11.154}$$

If the screen in figure 11.40 is removed, you can see that the light to the right of where the screen was is a copy of the light coming from the object, but upside down, and changed in size by d_2/d_1. If you have played with lenses, you know this.

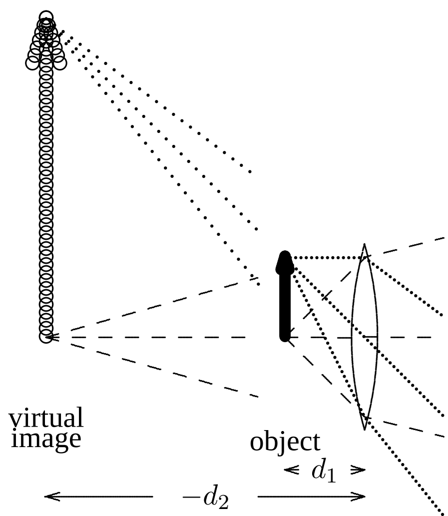


Figure 11.41:

Notice that (11.153) implies that neither d_1 nor d_2 can be less that f. If you bring the object too close to the lens, you do not get a real image on the other side. Instead, d_2 becomes negative and you get a "virtual image" on the same side of the lens as the object, and the light to the right of the lens is diverging as if it came from the virtual image. This situation is illustrated in figure 11.41. As we will discuss further below, this is how a magnifying glass works.

The image formation illustrated in figure 11.40 is what happens in a camera, and in your own eyeball. The lens focuses light from outside points onto points on the film, or your retina. Of course, the retina is not actually a plane. For the same reason, your eye lens is not a spherical lens, but some more complicated shape instead. The ray tracing has been done by evolution, however, so that objects in a plane get focused properly onto the retina.

Because the distance from your eye lens to your retina is fixed by the geometry of your eye, you must be able to adjust the shape of your lens. By doing so, you can change the focal length of your lens and thus change the distance at which points are perfectly in focus (this is called “accommodation”).

The formation of an image on your retina is illustrated in the diagram in figure 11.42. Again as in figure 11.40 the image is upside down. You cannot focus on objects that are too

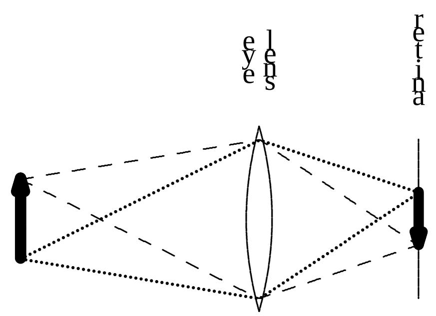


Figure 11.42:

close to your eye lens because the amount of accommodation you can do is limited. If you bring the object too closer than the smallest focal length your eye lens can produce, the real image is beyond your retina, the object will look fuzzy, as shown in figure 11.43.

A magnifying glass works by allowing you to produce a larger image of the object on your retina. It does this in two ways, both of which are illustrated in the diagram in figure 11.44 (with fewer light rays shown now because the diagrams are getting too busy).

Obviously, the image is larger. But note also that the magnifying glass changes the amount of accommodation required by your eye lens. Your eye is actually focusing on the virtual image which is much farther away, and that is easier. Thus when you look at an object in a magnifying glass, you can bring it much closer to your eye then you could without the glass. This further increases the magnifying effect, because closer objects look bigger. In this diagram you can also see a third salutary effect of the magnifying glass — more of the light from the object reaches your eye.

One of the magnifying effects of a lens can be obtained without a lens in a very simple way — with a pinhole. If you look at a nearby object through a pinhole, you can bring it

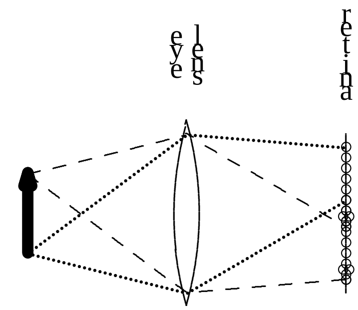


Figure 11.43:

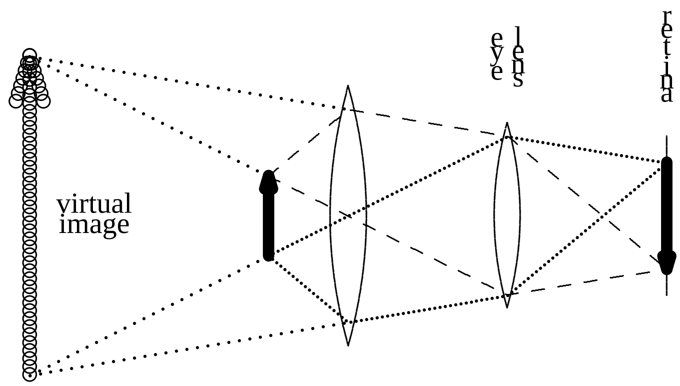


Figure 11.44:

much closer to your eye. The reason is that only a narrow beam of light get through the pinhole from each point on the object you are looking at, so not much focusing is required. The size of the image on your retina is not increased when you look at the object through a pinhole at the same same distance as without the pinhole, but with the pinhole, you can bring it much closer to your eye without fuzziness, and therefore you make it appear bigger.

You may also have played with pinhole cameras, in which you form an image on a screen in a dark box without a lens, as shown in figure 11.45.

One disadvantage to a pinhole camera is that you need a very bright object. You throw away most of the light coming from the object. You can get more light by making the pinhole larger, but that makes the image fuzzier. Actually, however, you cannot make the pinhole too small anyway. Ultimately, as we will see in chapter 13, diffraction limits the resolution of a pinhole camera. If you try to make the image very sharp by making the pinhole very tiny, the beam you get inside the camera will be spread by diffraction. The best you can do is choose

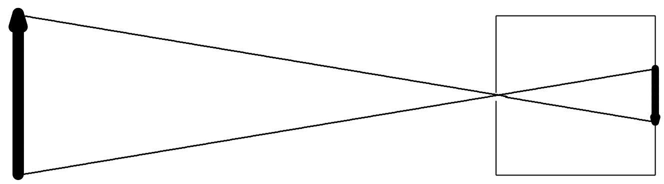


Figure 11.45:

the size of your pinhole so that the spreading at the screen due to diffraction just matches the size of the pinhole.

While we are on the subject, note that diffraction and the finite size of your pupil limits the angular resolution of your eye. As we will understand in detail in chapter 13, the finite size, s of your pupil introduces an angular spread of order d/λ for light of wavelength λ. Unless you have huge eyes, s is less than .25 cm, so for green light with wavelength 500 nanometers (550 is about the middle of the visible spectrum), the angular resolution is greater than about 2×10^{-4}. At a distance of 10 meters, for example, even if your eyes are perfect, you will not be able to resolve two objects less than a few millimeters apart.

You can use a pinhole to study your eyes in rather interesting ways. Put the pinhole close to your eye and look at a bright diffuse source of light. We will do this in lecture, but you can make your own pinhole by punching a small hole in a piece of aluminum foil with a pin and try this out. If you wear glasses, take them off. You won't need them. You should see a circular spot of light. This is the image of your pupil on your retina, as shown below:

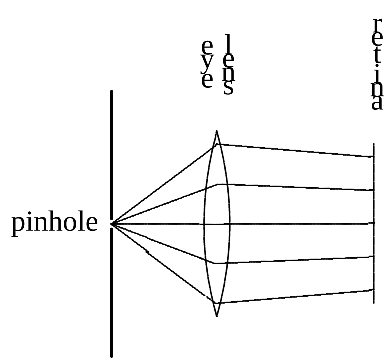


Figure 11.46:

You can watch the size of your pupil change with this arrangement. Just cover or close your other eye. Because you are now getting less light, both pupils will expand. Uncover the other eye and look at the bright light again and the pupils will contracts. Can you notice a short time-lag?

Now carefully bring a pen or pencil point up from below in between the pinhole and your eye, until it just begins to obscure your view. What do you see? This should convince you, if you were not sure before, that the image on your retina is upside down, as shown in figure 11.47. The bottom half of the image on your retina is missing. Your brain, being used

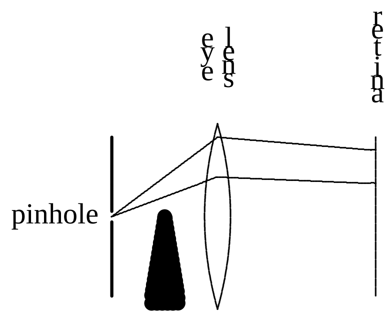


Figure 11.47:

to seeing images on the retina upside down, interprets this as an object coming down from above!

Magnification, telescopes, microscopes, and all that

By combining lenses in various ways, you can construct all sorts of interesting optical instruments. The simplest way to think about magnification is just to consider the angular size of the observed image, compared to the angular size you would see without the instrument.

A simple telescope is illustrated in figure 11.48. The distances are somwhat distorted. In a real telescope the object would be much farther way and the sizes of the lenses much smaller. When you look at a distant object (large L) with your telescope, the light arrives at the first ("objective") lens as a nearly parallel bundle of rays. We know from the thin lens formula

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \tag{11.153}$$

with $d_1 = L \gg f$ that a real image forms at a distance from the objective d_2 just slightly larger than its focal length f_1. The "eyepiece" is then placed a distance just beyond its focal length, f_2, from the real image, to make the light from the image into a nearly parallel bundle

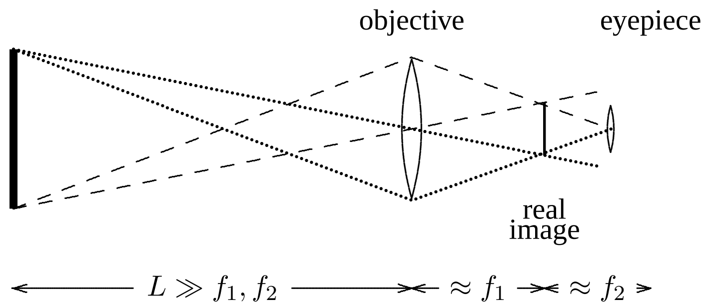


Figure 11.48:

again. Essentially what you are doing with the eyepiece is looking at the light from the real image with a magnifying glass.

We can understand how (and how much) a telescope magnifies distant objects by looking at the angles involved. If the object has size h_o, its angular size without the telescope is

$$\frac{h_o}{L+f_1+f_2} \approx \frac{h_o}{L} \tag{11.155}$$

By similar triangles, the size of the real image is

$$\frac{h_o}{L} \cdot f_1 \tag{11.156}$$

and thus the angular size of the real image at the eyepiece (and your eye) is

$$\frac{h_o}{L} \cdot \frac{f_1}{f_2} \tag{11.157}$$

Thus the magnification is approximately

$$\frac{f_1}{f_2} \tag{11.158}$$

Note that the telescope image appears upside down because what you are actually seeing is the real image.

A microscope looks something like what is shown in figure 11.49 (with even fewer light rays drawn because you should be getting used to them by this time.

The sample is placed just a little more than the focal length, f_1, away from the objective so that a real image forms that is much bigger than the sample. Then you look at the real image with the eyepiece as a magnifying glass, again positioned a little more than its focal

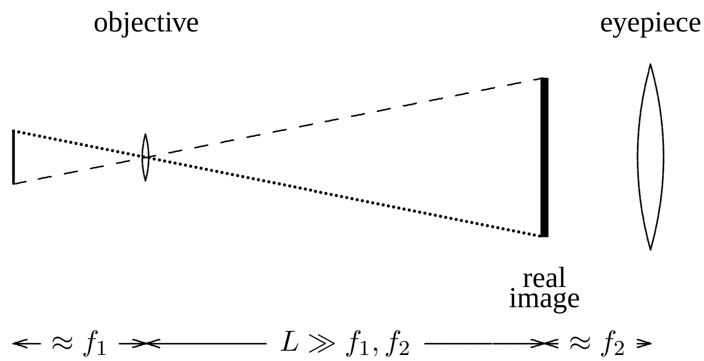


Figure 11.49:

length, f_2, away, to be able to view the image comfortably with your eyes relaxed. If the sample has size h_o, the size of the real image is

$$\frac{L}{f_1} \cdot h_o \tag{11.159}$$

and the angular size of the image at the eyepiece (and your eye) is

$$\frac{L\,h_o}{f_1 f_2} \tag{11.160}$$

This should be compared with the angular size of the object at some reference length, $L_0 \approx 25$ cm, at which you can view the object comfortably with your unaided eye, which is

$$\frac{h_o}{L_0} \tag{11.161}$$

Thus the magnification is

$$\frac{L\,L_0}{f_1 f_2} \tag{11.162}$$

11.7 Rainbows

Most elementary physics books either do not explain the rainbow at all, or explain it incorrectly (sometimes embarrassingly so). Obviously, it has something to do with the refraction of light by raindrops. We ought to be able to explain it just using Snell's law and geometrical optics — ray tracing. But it is a little subtle, as you will see.

To begin with, consider the refraction of a narrow ray of light from a spherical drop of water, illustrated in figure 11.50. The index of refraction of water, n, varies from about 1.332

for red light to about 1.343 for violet light. The ray enters somewhere on the drop, which can parameterize by the angle θ between the direction of the incoming light and the radius from the center of the drop to the point where the light enters. The angle θ is also the angle between the light ray and the perpendicular to the surface of the drop, so it is the appropriate to use in Snell's law. Thus the angle ϕ of the refracted ray inside the drop is given by

$$\sin\phi = \frac{1}{n}\sin\theta \tag{11.163}$$

or

$$\phi = \sin^{-1}\left(\frac{\sin\theta}{n}\right) \tag{11.164}$$

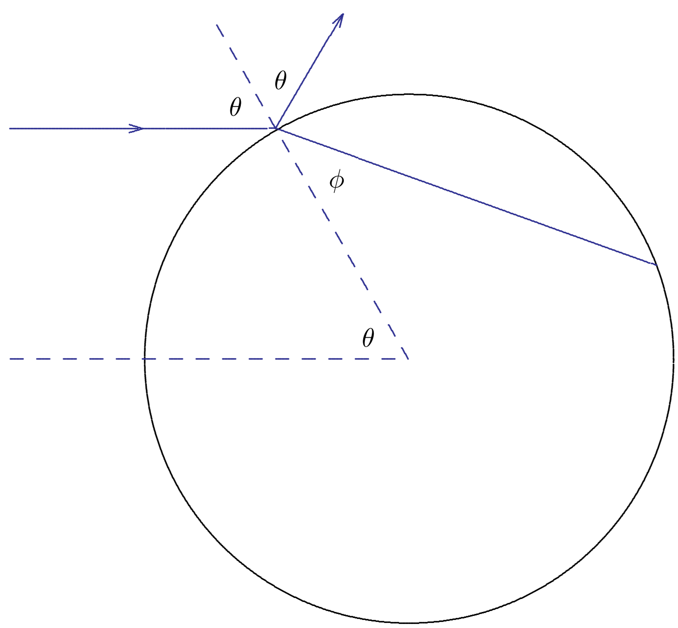


Figure 11.50:

Some of the light is also reflected from the drop. Note that the reflected light is reflected specularly. For $\theta = 0$, the light is reflected directly backwards. As θ increases from 0 the reflected ray is rotated counter-clockwise with respect to the incoming ray by an angle $\pi - 2\theta$ until at $\theta = \pi/2$ it just kisses the sphere and is not rotated at all.

The important geometrical fact that makes the problem fairly simple is that the angle between the ray and the perpendicular to the surface is the same when it comes out of the drop as when it comes in. Snell's law works in reverse, and the ray coming out of the drop makes an angle θ with the perpendicular. As you can see from figure 11.51, this means

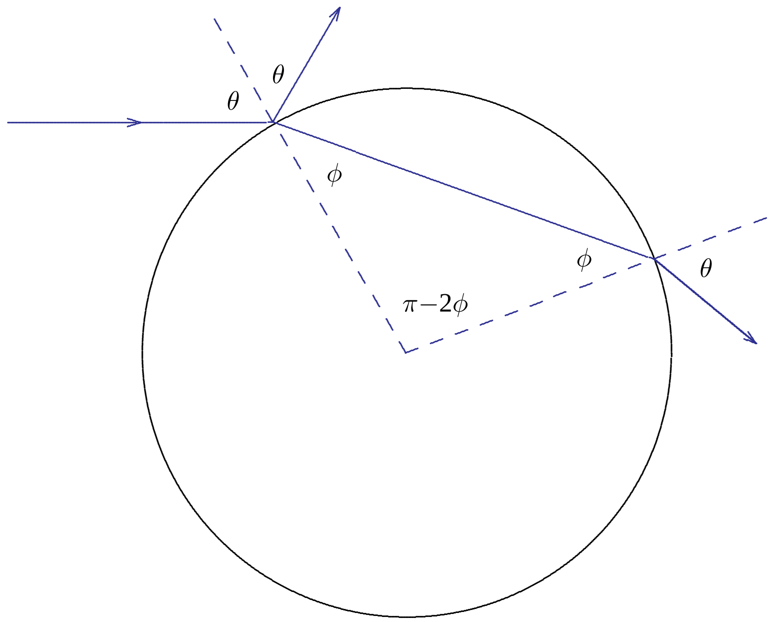


Figure 11.51:

that the refracted ray coming out of the drop This is just a version of the reflected ray in figure 11.50 rotated by $\pi - 2\phi$. This means that is it rotated by

$$\theta_1 = (\pi - 2\phi) - (\pi - 2\theta) = 2\theta - 2\phi \qquad (11.165)$$

from the original direction of the incoming light.

The trouble with this is that it has nothing to do with the rainbow. The problem is that the direction of the refracted ray is basically forward and it depends on θ, so that no particular value of θ is picked out. There are three mysterious things about the rainbow that this effect cannot explain.

i. The primary rainbow occurs at a definite angle, and

ii. the angle is in the **backwards** direction — at an angle of about 41° (about .7 radians) from the incoming light ray — that is rotated by about 2.4 radians from the original direction, and

iii. there is a second rainbow outside the first in which the colors go in the **opposite** direction!

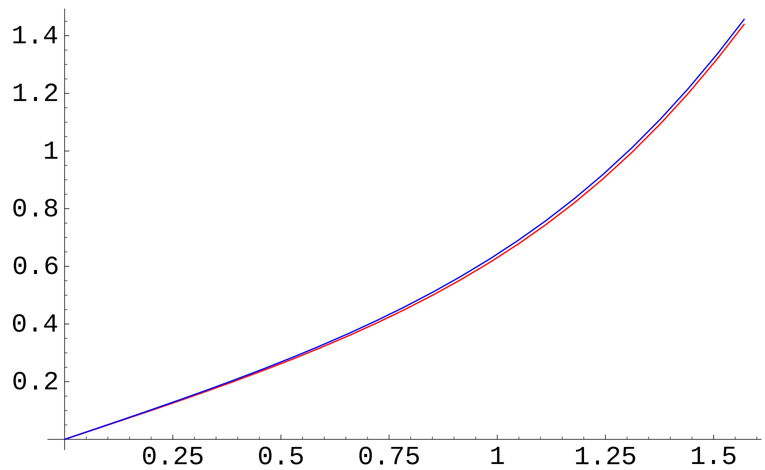


Figure 11.52: Plot of θ_1 versus θ for red light and blue light.

So what does this refraction do? The answer is almost nothing! The refracted ray is spread over a large range of angles, as shown in the graph in figure 11.52. At any particular outgoing angle, the light from this effect is very faint and hardly noticeable. Not only are the colors not separated very much, but all of them are spread more or less evenly over outgoing angle, so you don't see any rainbow from this refraction.

So where does the rainbow come from? The answer is that in addition to being refracted from the inside surface of the drop, the ray can also be reflected, and then come out at a still larger angle. The result looks like the picture in figure 11.53.

Comparing figure 11.51, figure 11.53 and equation (11.165), it is clear that for this path the light is rotated by

$$\theta_2 = 2(\pi - 2\phi) - (\pi - 2\theta) = 2\theta + \pi - 4\phi \tag{11.166}$$

And now here is the critical point. If we plot this θ_2 versus θ, the graph has a minimum! This is shown in figure 11.54.

Now the outgoing angle has a minimum for $\theta \approx 1.05$ (which is the value of θ illustrated in the diagrams). The outgoing angle $\theta_2 \equiv \theta_{\text{out}}$ corresponding to this θ gives the angular position of the rainbow. Here, because θ_2 does not change much for a small change in θ, you see the sum of the refracted light from a range of θs around the minimum. The angle is about what we expect, $\theta_{\text{out}} \approx \pi - .7$, where .7 radians$\approx 41^\circ$ is the angle between a vector from the water drop to the sun and the same drop to your eye, as shown in figure 11.55. The negative sign in $\pi - .7$ means that the light has not rotated by a full 180°, so the light reaching your eye entered the refracting water drop on the side farther away from you.

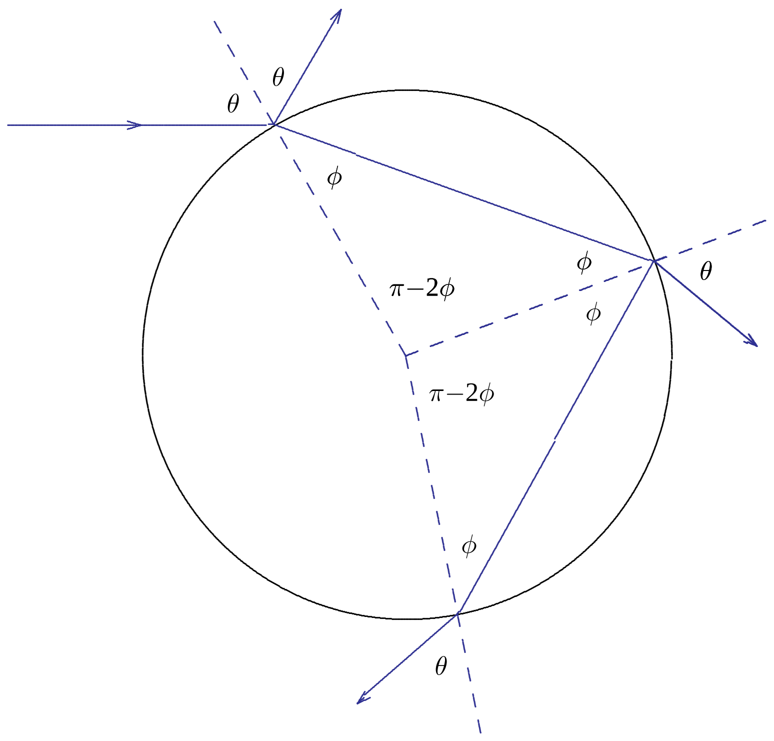


Figure 11.53:

You can also see from the graph in figure 11.54 that the colors are spread out. The red light is on the outside (farther away from 2π) and the blue light on the inside.

Mathematically, why does the light pile up at the edge? The energy from sunlight falling on a small part of the surface of the water drop between θ and $\theta + d\theta$ is proportional to $I\,d\theta$ (there are other factors, like $\cos\theta$, but they vary slowly, so let's forget them). The angle of the outgoing ray, θ_{out} is a function of θ, and the energy $\propto I_i\,d\theta$ is spread over an angular region between θ_{out} and $\theta_{\text{out}} + d\theta_{\text{out}}$. Thus the outgoing intensity is proportional to

$$\begin{array}{c}\text{incoming}\\ \text{energy between}\\ \theta \text{ and } \theta + d\theta\end{array} \propto I_i\,d\theta \tag{11.167}$$

$$= \begin{array}{c}\text{outgoing}\\ \text{energy between}\\ \theta_{\text{out}} \text{ and } \theta_{\text{out}} + d\theta_{\text{out}}\end{array} \propto I_o\,d\theta_{\text{out}} \tag{11.168}$$

$$I_o \propto \frac{I_i\,d\theta}{d\theta_{\text{out}}} = \frac{I_i}{\frac{d\theta_{\text{out}}}{d\theta}} \tag{11.169}$$

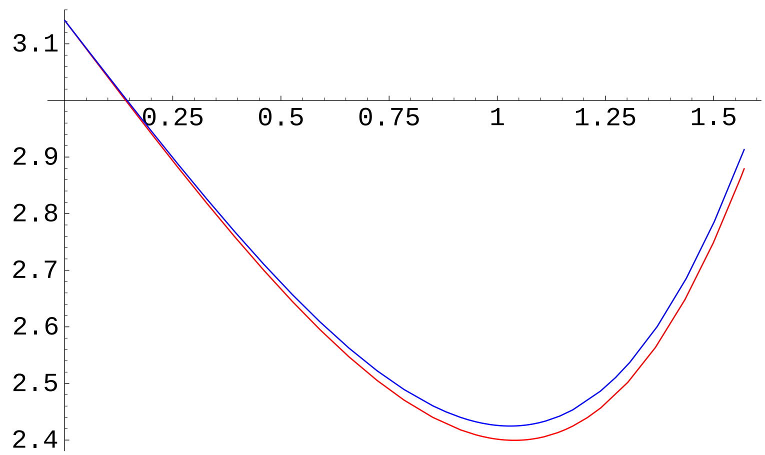


Figure 11.54: Plot of θ_2 versus θ for red light and blue light.

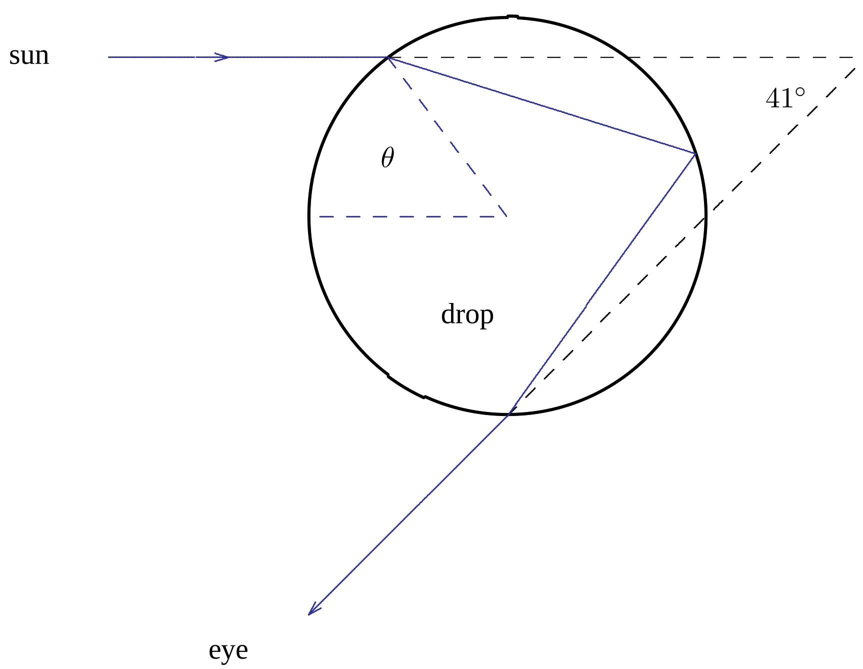


Figure 11.55:

When $d\theta_{\rm out}/d\theta = 0$, the intensity goes to infinity! The edge is infinitely more bright than the interior. That is why we see it!

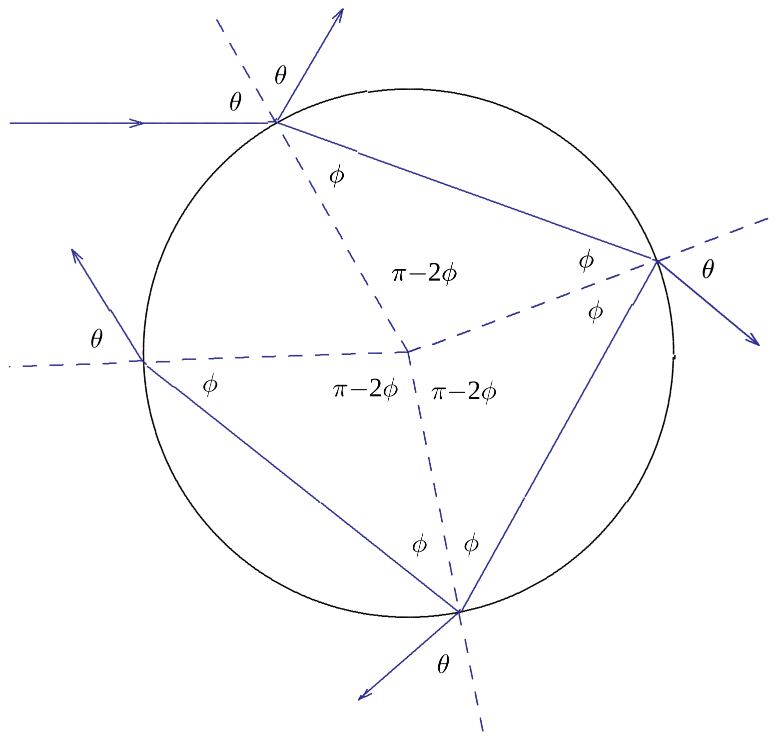


Figure 11.56:

We can now check this picture by seeing how it explains the second rainbow. As you might guess, this comes from yet another reflection, as shown in figure 11.56.

Now the light ray is rotated by

$$\theta_3 = 3(\pi - 2\phi) - (\pi - 2\theta) = 2\theta + 2\pi - 6\phi \tag{11.170}$$

This is shown, along with θ_2, in the plot in figure 11.57. The minimum of θ_3 is the position of the second rainbow. But now because the angle is greater than π, the light is reaching your eye from the side of the drop that is closer to you, and it is bending completely around.

This is why the colors are reversed. Again the blue is refracted more, but this time that means that the blue is on the outside, while the red in on the inside.

By accident, the minima for θ_2 and θ_3 are almost equally (within about .13 radians) displaced from π, though on opposite sides. This is why the two rainbows are fairly close

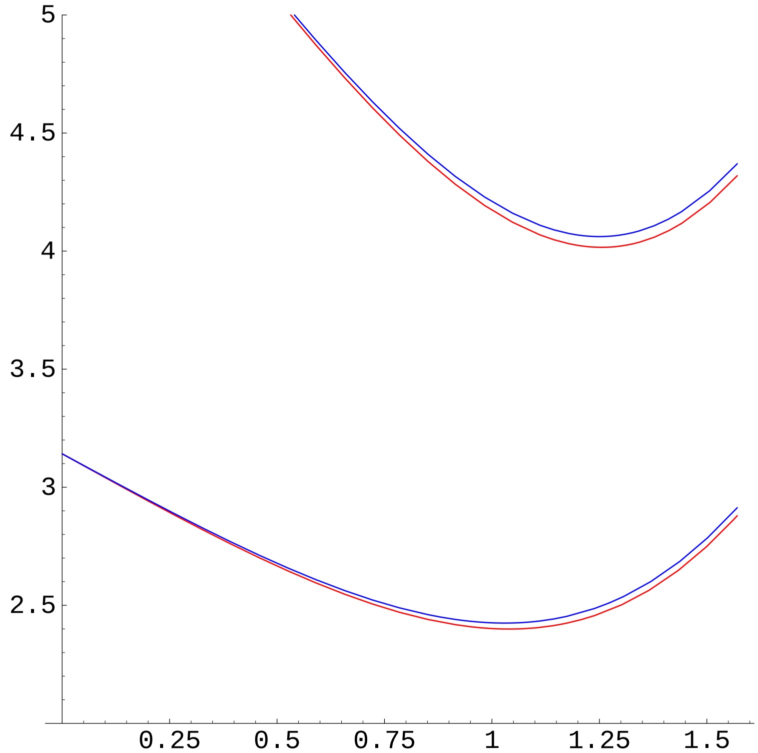


Figure 11.57: Plot of θ_2 and θ_3 versus θ for red light and blue light.

together in the sky.

Another prediction of this picture that can often be seen is "Alexander's dark band" that appears between the rainbows. The light that is not concentrated at the minimum value of θ is spread inside the first rainbow but outside the second rainbow, thus the region between the two rainbows (or outside the first if the second cannot be seen) is darker. If we plot the angular distance away from π as a function of the angle at the which the incoming sunlight enters the water drop, the first and second rainbows look like figure 11.58 (as usual, I have exagerated the difference in index of refraction between red and blue. Here you clearly see that the angle of first rainbow is smaller, and the dark band between the two.

11.8 Spherical Waves

Consider sound waves in a very large room with absorbing walls. In the middle of the room (we will take the middle of the room to be the origin of our coordinate system, $\vec{r} = 0$) is a

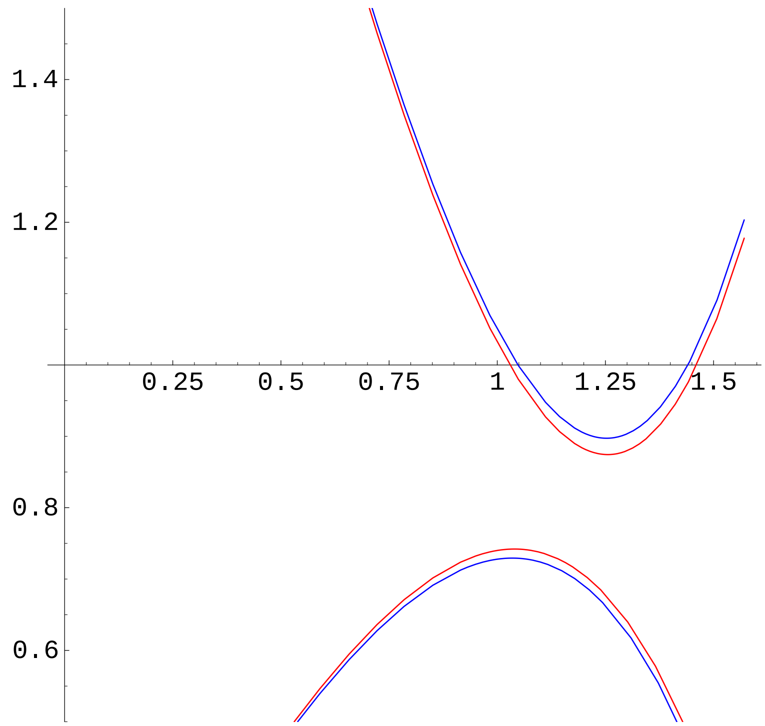


Figure 11.58: Both rainbows.

spherical loudspeaker, a sphere that produces an oscillating pressure at its surface (at radius R) of the form $p_0 \cos \omega t$. What sort of sound waves are produced? It seems rather silly to use our plane wave solutions with space translation invariance for this problem, because this system has a symmetry under rotations about the origin. Instead, let us look directly at the wave equation and make use of the spherical nature of the problem. That is, assume that the solution has the form $\psi(\vec{r}, t) = \chi(|\vec{r}|, t)$. Putting this into the wave equation gives (with $r \equiv |\vec{r}|$)[8]

$$\begin{aligned} \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \chi(r,t) = \vec{\nabla}^2 \chi(r,t) = \vec{\nabla} \cdot \vec{\nabla} \chi(r,t) \\ = \vec{\nabla} \cdot (\vec{\nabla} r) \frac{\partial}{\partial r} \chi(r,t) = \vec{\nabla} \cdot (\vec{r}/r) \frac{\partial}{\partial r} \chi(r,t) \end{aligned} \quad (11.171)$$

[8]If you have seen spherical coordinates, you may remember that you cannot compute the Laplacian, $\vec{\nabla} \cdot \vec{\nabla}$, simply as $\frac{\partial^2}{\partial r^2}$. You don't need to remember the details here because we compute it from scratch for the function, $\chi(|\vec{r}|, t)$.

$$
\begin{aligned}
&= (\vec{\nabla}\cdot\vec{r}/r)\,\frac{\partial}{\partial r}\chi(r,t) + (\vec{\nabla}r)\cdot(\vec{r}/r)\,\frac{\partial^2}{\partial r^2}\chi(r,t) \\
&= [(\vec{\nabla}\cdot\vec{r})/r + \vec{r}\cdot\vec{\nabla}(1/r)]\,\frac{\partial}{\partial r}\chi(r,t) + (\vec{r}/r)\cdot(\vec{r}/r)\,\frac{\partial^2}{\partial r^2}\chi(r,t) \\
&= \frac{2}{r}\,\frac{\partial}{\partial r}\chi(r,t) + \frac{\partial^2}{\partial r^2}\chi(r,t)\,.
\end{aligned}
\tag{11.172}
$$

We can rewrite this in the following useful form:

$$
\frac{1}{v^2}\,\frac{\partial^2}{\partial t^2}\,\chi(r,t) = \frac{1}{r}\,\frac{\partial^2}{\partial r^2}\,r\chi(r,t)\,. \tag{11.173}
$$

Thus $r\chi(r,t)$ satisfies the one-dimensional wave equation.

We can now solve the problem that we posed above. The solutions for $r\chi$ have the form $\sin(kr \pm \omega t)$ and $\cos(kr \pm \omega t)$, where $k = \omega/v$. Because the pressure at $r = R$ is $p_0 \cos\omega t$, we are interested in the combinations $\cos(kr - kR - \omega t)$ and $\cos(kr - kR + \omega t)$. These describe waves going outward from and inward toward the origin respectively. The appropriate boundary condition at infinity is to take the outgoing wave, so that the disturbance is produced entirely by the speaker. Thus

$$
\chi(r,t) = \frac{p_0 R}{r}\cos(kr - kR - \omega t)\,. \tag{11.174}
$$

The general features of the solution, (11.174), are easy to understand. The wave-fronts, along which the phase of oscillation is constant, are spheres centered about the origin, as they must be because of the rotational symmetry. The waves move out from the origin at speed v. As they move outward, their local intensity must decrease, because the same amount of energy is being spread over a larger area. This is the reason for the $1/r$ in (11.174). If the amplitude falls as $1/r$, the intensity of the wave falls as $1/r^2$, as it must. Though the physics is clear, the precise form of this solution is deceptively simple. In two dimensions, for example, it is not possible to find a solution to an analogous problem using the functions that you know from high school. In two dimensions, the amplitude of the wave must decrease roughly as $1/\sqrt{r}$. The solutions to the two-dimensional wave equation with this property are called Bessel functions. You will learn about them in more advanced courses.

11.9 Chapter Checklist

You should now be able to:

i. Interpret plane waves in two- and three-dimensional space in terms of a $\vec{k}$ vector, angular wave number;

ii. Analyze the scattering of a plane wave from a plane boundary between regions with different dispersion relations;

iii. Derive and use Snell's law;

iv. Understand the phenomenon of total internal reflection, along with the general statement of the boundary condition at infinity for complex $\vec{k}$;

v. Understand the physics and mathematics of tunneling phenomena;

vi. Understand how degeneracy of the frequencies of the normal modes affects the forced oscillation problem and find the sand patterns on square Chladni plates;

vii. Understand the propagation of waves in waveguides, using separation of variables to construct the modes and interpret the result in terms of zig-zag waves;

viii. Be able to analyze water waves, ignoring viscosity and angular momentum.

ix. Solve problems involving spherical waves where the displacement involves only r and t;

Problems

11.1. Consider the free transverse oscillations of the two-dimensional beaded string shown in figure 11.59. All the horizontal strings have tension T_h, all the vertical strings have tension T_v, all the solid circles are beads with mass m. The square frame is fixed in the $z = 0$ plane.

a. Find the normal modes and their corresponding frequencies.

b. Suppose that $T_v = 100T_h$. Draw nine diagrams, one for each normal mode, **in order of increasing frequency** indicating which beads are moving up (by a + sign), which are moving down (by a $-$ sign), and which are not moving (by a 0). You can interchange + and $-$ and still have the right answer by changing the setting of your clock, or multiplying your normal mode vector by -1. For example, the lowest frequency mode looks like

$$\begin{matrix} + & + & + \\ + & + & + \\ + & + & + \end{matrix}$$

while the mode with the fifth highest (and also the fifth lowest — in other words the one in the middle) looks like

$$\begin{matrix} - & 0 & + \\ 0 & 0 & 0 \\ + & 0 & - \end{matrix}$$

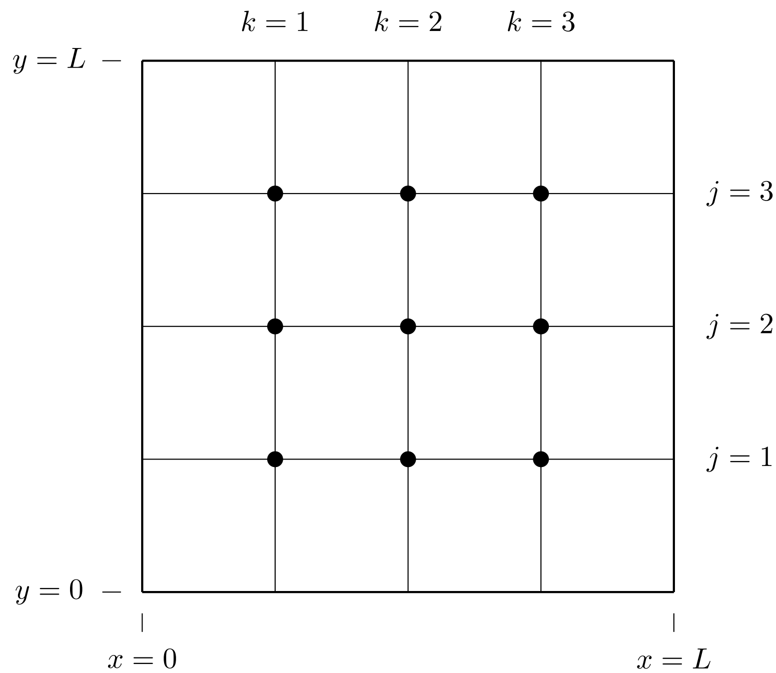


Figure 11.59: A two-dimensional beaded string.

Do the rest and **get the order right.** You should be able to do this even if you got confused by the details of part a.

11.2. Consider the forced transverse oscillations of the two-dimensional beaded string shown in figure 11.60. All the strings have tension T, all the solid circles are beads with mass m. The frame is held fixed in the $z = 0$ plane. The open circles are moved up and down out of the plane of the paper with the same transverse displacement,

$$z_1(t) = z_2(t) = z_3(t) = d\cos\omega t$$

where

$$\omega = 2\sqrt{\frac{T}{ma}}\,.$$

Find the displacement for each of the beads. You can do this by solving for the displacement, $z_{jk}(t)$, of the bead whose horizontal position is

$$x = \frac{kL}{4}\,, \quad y = \frac{jL}{4}$$

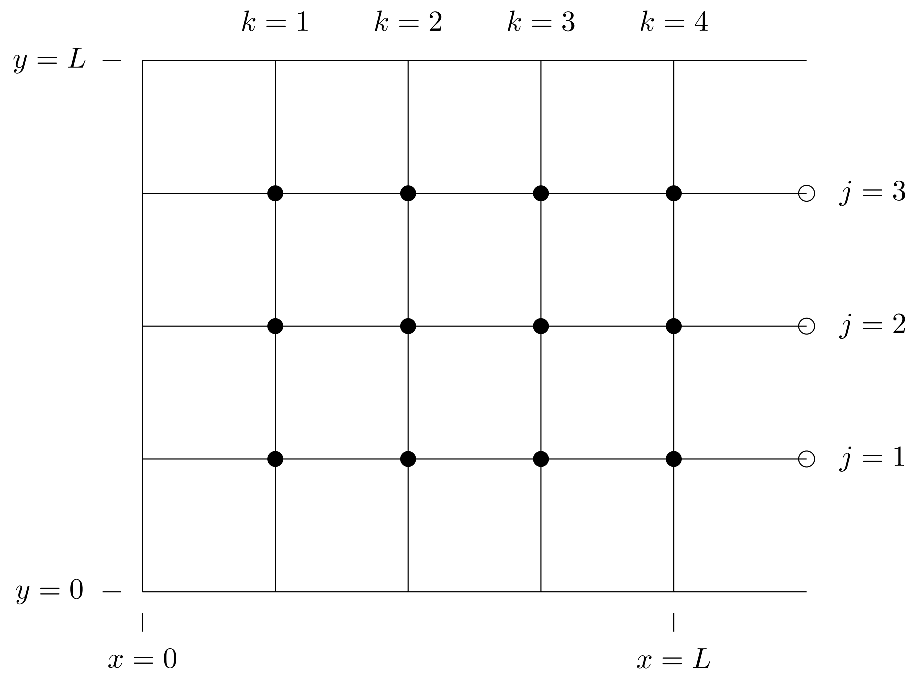


Figure 11.60: A two-dimensional beaded string.

for all relevant j and k. All displacements will be proportional to $d\cos\omega t$, so write your answer in the form of a table of the coefficients of $d\cos\omega t$ for each j and k:

$j \backslash k$	1	2	3	4
1	?	?	?	?
2	?	?	?	?
3	?	?	?	?

11.3. Consider the forced transverse oscillations of the semi-infinite two-dimensional beaded string shown in figure 11.61. All the strings have tension T, all the solid circles are beads with mass m. The equilibrium separations of the blocks are all a. The frame at $y=0$ and $y=4a$ is held fixed in the $z=0$ plane. The open circles at $x=0$ are moved up and

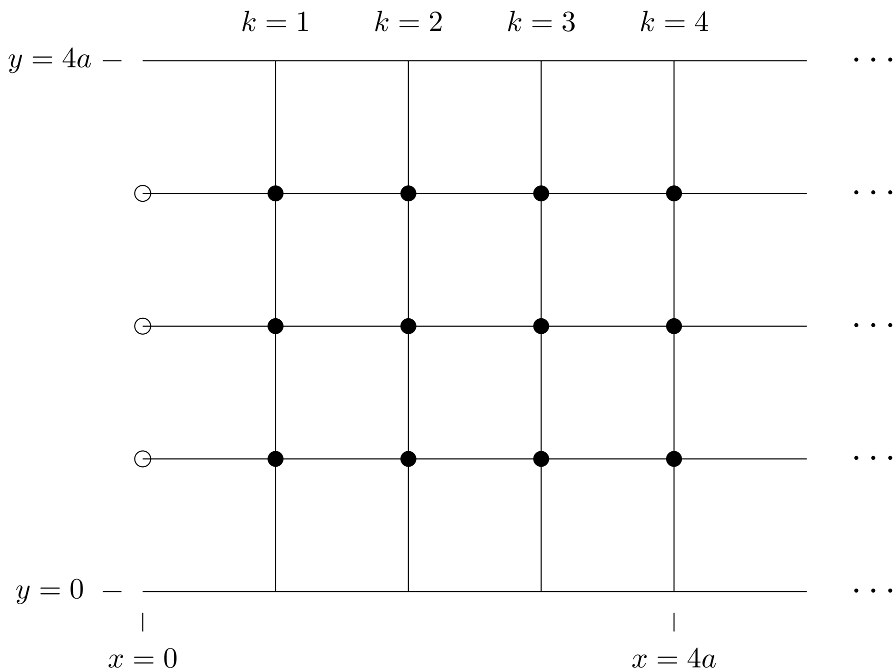


Figure 11.61: A semi-infinite two-dimensional beaded string.

down out of the plane of the paper with transverse displacement,

$$z_1(t) = z_3(t) = \frac{d}{\sqrt{2}}\cos\omega t\,, \quad z_2(t) = -d\cos\omega t\,,$$

for the values of ω given below. For each ω find the displacement for each of the beads as a function of its equilibrium position. That is, determine $\psi(x,y,t)$. Assume that the entire system is oscillating with frequency ω and that the displacement is well-behaved at $x=+\infty$.

a. Find $\psi(x,y,t)$ for

$$\omega^2 = \frac{T}{am}\,(2+\sqrt{2}-\epsilon^2)$$

In both **a** and **b**, assume that ϵ is a small real number, small enough so that you can approximate

$$\sinh\frac{\epsilon}{2} \approx \frac{\epsilon}{2}\,.$$

b. Find $\psi(x,y,t)$ for

$$\omega^2 = \frac{T}{am}\,(6+\sqrt{2}+\epsilon^2)\,.$$

11.4. A flexible membrane with surface tension τ_S and surface mass density ρ_S is stretched so that its equilibrium position is the $z = 0$ plane. Attached to the surface of the membrane at $x = 0$ is a string with tension τ_L and linear mass density ρ_L. Consider a traveling wave on the membrane with transverse displacement

$$\psi(x, y, t) = \psi_-(x, y, t) = Ae^{-i\omega t + ik_x x + ik_y y} + R\,Ae^{-i\omega t - ik_x x + ik_y y}$$

for $x \leq 0$, and

$$\psi(x, y, t) = \psi_+(x, y, t) = T\,Ae^{-i\omega t + ik_x x + ik_y y}$$

for $x \geq 0$.

In what direction is the reflected wave (for $x < 0$) traveling? **Easy!**

Newton's law for a small element of the string of length dy with equilibrium position $(0, y, 0)$ is

$$\tau_S\,dy\,\left[\frac{\partial}{\partial x}\psi_+(0, y, t) - \frac{\partial}{\partial x}\psi_-(0, y, t)\right] + \tau_L\,dy\,\frac{\partial^2}{\partial y^2}\psi_\pm(0, y, t)$$
$$= \rho_L\,dy\,\frac{\partial^2}{\partial t^2}\psi_\pm(0, y, t)\,.$$

Explain the physical significance of the term above, proportional to τ_S. What is pulling on what? Why does it have the form shown above?

11.5. Consider the transverse oscillations of an infinite flexible membrane stretched in the $z = 0$ plane with surface tension T_s and surface mass density D_s. Along the $z = 0$, $x = 0$ line, a string with linear mass density D_L but no tension of its own is attached to the membrane.

Consider a wave of the form:

$$Ae^{i(kx\cos\theta + ky\sin\theta - \omega t)} + R\,Ae^{i(-kx\cos\theta + ky\sin\theta - \omega t)} \quad \text{for } x < 0$$
$$T\,Ae^{i(k'x\cos\theta' + k'y\sin\theta' - \omega t)} \quad \text{for } x > 0$$

where $\cos\theta > 0$ and $\cos\theta' > 0$.

Find $\sin\theta'$ in terms of $\sin\theta$ (TRIVIAL!).

Find R and T.

Hint: Consider $F = ma$ for an infinitesimal piece of the weighted string, remembering that it has no tension of its own.

11.6. Two semi-infinite flexible membranes are stretched in the $z = 0$ plane. The first has surface tension 1 dyne/cm and mass density 169 gr/cm^2. It is fixed along the $z = 0$,

$y = 0$ axis and the $z = 0$, $y = a$ axis and extends from $x = 0$ to ∞ in the $+x$ direction. The second has the same surface tension but mass density 180 gr/cm^2. It is also fixed along the $z = 0$, $y = 0$ axis and the $z = 0$, $y = a$ axis and extends from $x = 0$ to $-\infty$ in the $-x$ direction. The two membranes are joined together with massless tape at $x = 0$. Consider the transverse oscillations of this system of the following form:

$$\psi(x,y,t) = A\sin(k_y y)(e^{-i(\omega t - k_x x)} + R\,e^{-i(\omega t + k_x x)}) \quad \text{for } x \leq 0;$$
$$\psi(x,y,t) = A\sin(k_y y)T\,e^{-i(\omega t - k'_x x)} \quad \text{for } x \geq 0$$

where $k_y = 12\pi$ cm^{-1} and $\omega = \pi$ s^{-1}.

Find k_x and k'_x.

Find R and T.

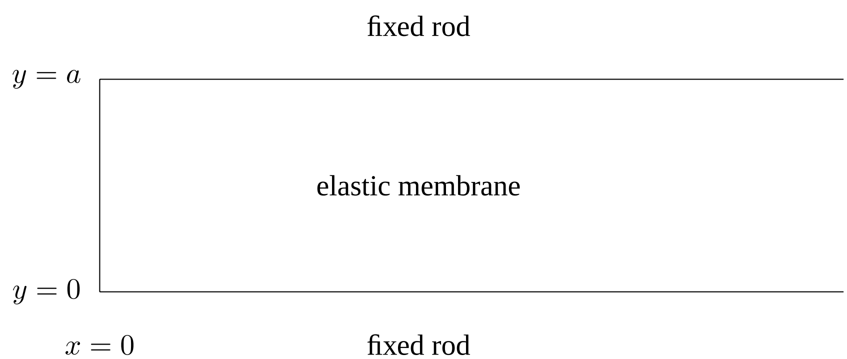


Figure 11.62: A forced oscillation problem in an elastic membrane.

11.7. A uniform membrane is stretched in the $z = 0$ plane, as shown in figure 11.62. It is attached to fixed rods along $y = 0$, $z = 0$ and $y = a$, $z = 0$ from $x = 0$ to ∞. $\psi(x,y,t)$ is the z displacement of the point on the membrane with equilibrium position $(x,y,0)$. For small oscillations, ψ satisfies the two-dimensional wave equation,

$$v^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi = \frac{\partial^2}{\partial t^2}\,\psi\,.$$

If this system is extended to an infinite system by continuing it to negative x, show that the normal modes of the infinite system take the form:

$$\psi(x,y) = A\sin(nk_0 y)\,e^{ikx}\,.$$

Find k_0. Suppose that the end of the membrane at $x = 0$ is driven as follows:

$$\psi(0, y, t) = \cos(5vk_0t)[B \sin(3k_0y) + C \sin(13k_0y)]$$

The boundary condition at ∞ is such that there is no wave traveling in the $-x$ direction along the membrane. Find $\psi(x, y, t)$.

Explain the following statement: For $\omega < 2vk_0$, the system acts like a one-dimensional wave carrier with the dispersion relation $\omega^2 = v^2k^2 + \omega_0^2$. What is ω_0?

11.8. Consider a rigid spherical shell of inner radius L filled with gas in which the speed of sound is v. In this sphere there are **standing wave** normal modes of many kinds. We will be interested in those in which the pressure depends only on the distance, r, from the center of the sphere. Suppose that $\psi(\vec{r}, t) = \chi(r, t)$ is the difference between the pressure of the gas in such a mode and the equilibrium pressure. We know from (11.173) that $\xi(r, t) \equiv r\,\chi(r, t)$ satisfies the one-dimensional wave equation:

$$\frac{\partial^2}{\partial t^2}\,\xi(r, t) = v^2\,\frac{\partial^2}{\partial r^2}\,\xi(r, t)\,.$$

Explain the physics of the boundary condition at $r = 0$.

In terms of an unknown wave number, k, find a form for $\chi(r, t)$ that satisfies the boundary condition at $r = 0$.

Explain the physics of the boundary condition at $r = L$.

Write down the mathematical statement of the boundary condition at $r = L$, the solutions of which give the allowed values of k for the normal modes.

Hints:. Remember that it is χ and not ξ that is the physical pressure difference. The lowest nontrivial mode has a k value which satisfies $k\,L \approx 4.4934$. The amplitude of the pressure oscillations in this mode as a function of r is shown in the graph in figure 11.63.

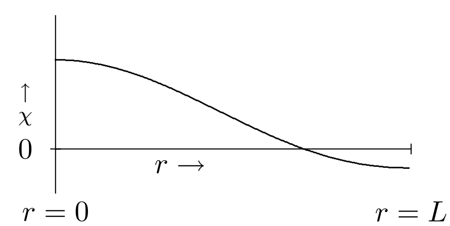


Figure 11.63: Amplitude of pressure oscillation versus r.

11.9. Consider a boundary between two semi-infinite membranes stretched in the x-y plane. The membrane for $x < 0$ has surface tension τ_s and surface mass density ρ_s. The membrane for $x > 0$ has the same surface tension τ_s but a different surface mass density ρ_s'. Along the boundary there is a device (I don't know exactly how it works) that produces a vertical frictional force, proportional to minus the vertical velocity of the membrane at the boundary. In other words, if $\psi(x, y, t)$ is the z displacement of the membrane as a function of (x, y), then the force (in the z direction) on a small chunk of the boundary stretching from the point $(0, y)$ to $(0, y + dy)$ is

$$dF = -dy\,\gamma\,\frac{\partial}{\partial t}\psi(0, y, t)\,.$$

On the membrane there is a plane wave of the form shown below, with displacement:

$$\psi(x, y, t) = A e^{i(kx\cos\theta + ky\sin\theta - \omega t)}$$

for $x < 0$, and

$$\psi(x, y, t) = A e^{i(k'x\cos\theta' + k'y\sin\theta' - \omega t)}$$

for $x > 0$. The setup is shown in figure 11.64. The dispersion relation for $x < 0$ is

$$\omega^2 = \frac{\tau_s}{\rho_s}\,\vec{k}^2\,.$$

Find k'.
Find θ'
Find γ. You should find $\gamma \to 0$ for $\rho_s \to \rho_s'$. Explain why.

11.10 💻 11-4. Instead of an open ocean, consider a system with a bottom at $y = 0$ and a fixed top at $y = 2L$, half full of water and half full of paint-thinner, another nearly incompressible fluid which is lighter than water and floats in the top half without mixing with the water.

Show that waves in this system have the form of (11.122) for $y \le L$ (in the water) and

$$\begin{aligned}\psi_x(x, y, t) &= \mp i e^{\pm ikx - i\omega t}\,\cosh[k(2L - y)]\,,\\ \psi_y(x, y, t) &= e^{\pm ikx - i\omega t}\,\sinh[k(2L - y)]\,,\end{aligned} \tag{11.175}$$

for $L \le y \le 2L$ (in the paint-thinner), by arguing that (11.175) and (11.122) satisfy the appropriate boundary conditions at $y = 0$ and $y = 2L$ and (for small displacements) at $y = L$, and show that (11.175), like (11.125), is consistent with incompressibility ($\vec{\nabla}\cdot\vec{\psi} = 0$).

Show that ψ_x is discontinuous at $y = L$ and explain physically what is happening at this boundary and why. When you have done this, take a look at program 11-4, in which this system is animated. If you look carefully, you will notice the effect of the breakdown of linearity for large displacements.

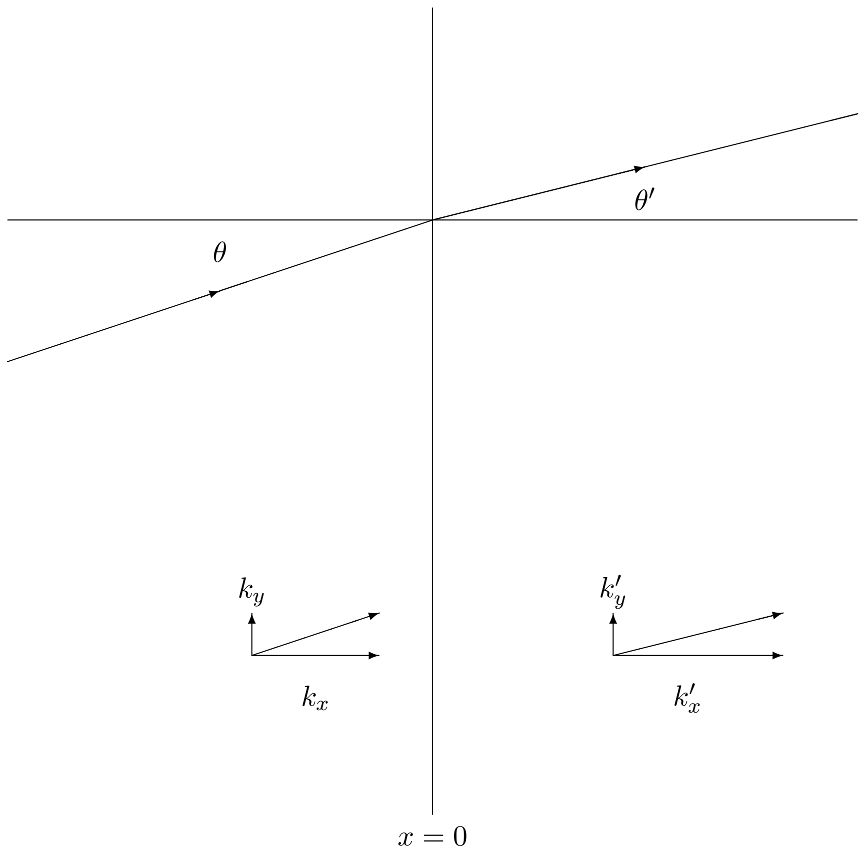


Figure 11.64: Scattering from a boundary in an elastic membrane.

Now suppose that the liquids are contained within vertical walls at $x = 0$ and $x = X$.

What boundary conditions are satisfied at the vertical boundaries, $x = 0$ and $x = X$?

Find the form of the displacements for the normal modes in this system. You may want to check that they satisfy $\vec{\nabla} \cdot \vec{\psi} = 0$.

Show that the dispersion relation for this system is

$$\omega^2 = \left[\frac{\rho_W - \rho_P}{\rho_W + \rho_P} \, gk + \frac{k^3 \tau_S}{\rho_W + \rho_P} \right] \tanh kL \,, \tag{11.176}$$

where ρ_P is the density of the paint-thinner, ρ_W is the density of the water, and τ_S is the surface tension of the boundary between the water and the paint-thinner. **Hint:** You use an energy argument analogous to (11.127)-(11.137), and just discuss how the various contributions change when you go from (11.137) to (11.176).

11.11. Consider the reflection of sound waves from a massless, infinitely flexible membrane that separates two gases with the same equilibrium pressure, p_0, but different densities. The membrane is in the $x = 0$ plane. The gas in region 1, for $x < 0$ has equilibrium density ρ_1, ratio of specific heat at constant pressure to specific heat at constant volume γ_1, and sound speed $\sqrt{\gamma_1 p_0/\rho_1}$ while the gas in region 2, for $x > 0$ has density ρ_2, specific heat ratio γ_2 and sound speed $\sqrt{\gamma_2 p_0/\rho_2}$. A pressure wave in the system has the following form:

$$P(r,t)/\delta p = A e^{i\vec{k}_1\cdot\vec{r}-i\omega t} + R\,A e^{i\vec{k}_R\cdot\vec{r}-i\omega t}$$

in region 1, for $x < 0$, and

$$P(r,t)/\delta p = T\,A e^{i\vec{k}_2\cdot\vec{r}-i\omega t}$$

in region 2, for $x > 0$, where $P(r,t) + p_0$ is the pressure of the gas whose equilibrium position is $\vec{r}$. The small pressure, δp, describes the amplitude of the pressure wave. R and T are the reflection and transmission coefficients.

The k vectors are

$$\vec{k}_1 = (k\,\cos\theta, k\,\sin\theta, 0)$$
$$\vec{k}_R = (-k_R\,\cos\theta_R, k_R\,\sin\theta_R, 0)$$
$$\vec{k}_2 = (k_2\,\cos\theta_2, k_2\,\sin\theta_2, 0)$$

where k, k_R, k_2, $\cos\theta$, $\cos\theta_R$, and $\cos\theta_2$ are all positive.

Find k_R and $\cos\theta_R$ in terms of k and θ.

Find k_2 and $\cos\theta_2$ in terms of k and θ.

Show that if $\rho_1/\gamma_1 > \rho_2/\gamma_2$, there is a critical value of θ above which the wave is totally reflected, and find the critical angle.

To find R and T, we need the boundary conditions at $x = 0$. One condition follows from the fact that the membrane is massless and infinitely flexible. That implies that there can be no force on it transverse to its surface.

Find this boundary condition. **Hint:** Where does the force transverse to the surface come from?

The other condition involves the transverse displacement of the membrane. The displacement can be obtained from the pressure:

$$\vec{\psi}(r,t) = \frac{1}{\rho_j\omega^2}\vec{\nabla}P(r,t)\,,$$

where $\vec{\psi}(r,t)$ is the displacement of the gas whose equilibrium position is $\vec{r}$ and j is the region label.

Thus

$$\vec{\psi}(r,t)/\delta p = \frac{iA}{\rho_1\omega^2}\left(\vec{k}_1\,e^{i\vec{k}_1\cdot\vec{r}-i\omega t} + R\,\vec{k}_R\,e^{i\vec{k}_R\cdot\vec{r}-i\omega t}\right)$$

in region 1, for $x < 0$, and

$$\vec{\psi}(r,t)/\delta p = \frac{iA}{\rho_2\omega^2}\, T\, \vec{k}_2\, e^{i\vec{k}_2\cdot\vec{r}-i\omega t}$$

in region 2, for $x > 0$.

Find the other boundary condition. **Hint:** Assume that the amplitude δp is small.

Find R and T.

11.12. Consider a universe filled with material that has a nonzero conductivity, σ. That is, in this material, there is a current proportional to the electric field (Ohm's law),

$$\vec{J}(\vec{r},t) = \sigma\, \vec{E}(\vec{r},t)\,. \tag{11.177}$$

We will assume that the material has no other electrical properties, in particular that there is no polarization or magnetization, and that no charge builds up anywhere, so that $\rho = 0$. Consider the propagation of a plane electromagnetic wave in this universe. Because this universe is perfectly space translation invariant and rotation invariant, and because (11.177) is linear, we would expect that there will be plane wave solutions in which the electric and magnetic fields are proportional to

$$e^{i(\vec{k}\cdot\vec{r}-\omega t)}$$

for $\vec{k}^2$ and ω related by some dispersion relation. In particular, consider propagation in the $+z$ direction with the electric field in the x direction and the magnetic field in the y directions:

$$E_x(\vec{r},t) = E\, e^{i(kz-\omega t)}\,, \quad E_y(\vec{r},t) = E_z(\vec{r},t) = 0$$

$$B_y(\vec{r},t) = B\, e^{i(kz-\omega t)}\,, \quad B_x(\vec{r},t) = B_z(\vec{r},t) = 0\,.$$

a. Show from the relevant Maxwell's equations,

$$\frac{\partial}{\partial z}E_x - \frac{\partial}{\partial x}E_z = -\frac{\partial B_y}{\partial t}$$

$$\frac{\partial}{\partial y}B_z - \frac{\partial}{\partial z}B_y = \mu_0\epsilon_0\frac{\partial E_x}{\partial t} + \mu_0 J_x$$

that such a plane wave can exist if

$$k^2 = \mu_0\epsilon_0\omega^2 + i\mu_0\sigma\omega\,.$$

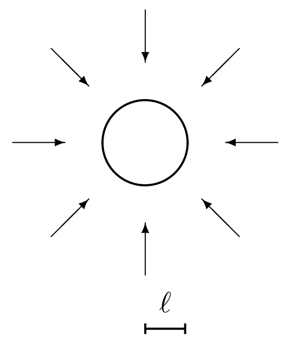


Figure 11.65: A spherical sound damper.

b. Assume that ω is real and positive and that the real part of k is positive. Find the sign of the imaginary part of k, and interpret your result physically. That is, explain why the sign had to come out the way it did.

11.13. Consider a spherical sound wave coming in from far away and being completely absorbed by a spherical sound damper at a radius $r = \ell$, as shown in figure 11.65. The pressure in is this system is described by the real part of the complex traveling wave below, depending only on the radius and time:

$$p(r,t) - p_0 = \frac{\epsilon}{r}\, e^{-i(kr+\omega t)}$$

where

$$\omega^2 = \frac{\gamma p_0}{\rho}\, \vec{k}^2$$

with p_0, the equilibrium pressure and ρ the equilibrium mass density of the gas. The typical displacement of the air from its equilibrium position in this wave is in the radial direction,

$$\psi_r(r,t) = \frac{1}{\rho\omega^2}\frac{\partial p}{\partial r}\,.$$

a. Find the time-averaged power absorbed by the spherical damper at $r = \ell$.

b. Explain (qualitatively) the factor of $1/r$ in the pressure.

Now suppose that there is a massless, flexible spherical boundary between two different gases at radius $r = r_b$, shown as the dashed circle in the diagram in figure 11.66. The equilibrium pressure, p_0, is the same on both sides of the boundary. Also, assume that γ is

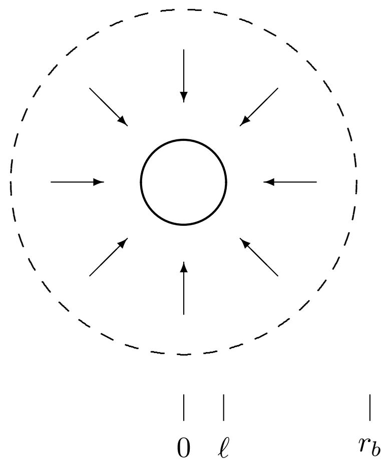


Figure 11.66: A spherical sound damper with a reflecting boundary.

the same for both gases and that the only difference is the densities. Inside the density is ρ, and outside the density is ρ'. Now for $\ell < r < r_b$, the pressure is still given as above, but in the region outside the dashed circle, there is a reflected wave as well as the incoming wave,

$$p(r,t) - p_0 = \frac{A}{r}\, e^{-i(k'r+\omega t)} + \frac{B}{r}\, e^{i(k'r-\omega t)}$$

where

$$\omega^2 = \frac{\gamma p_0}{\rho'}\, \vec{k'}^2 \, .$$

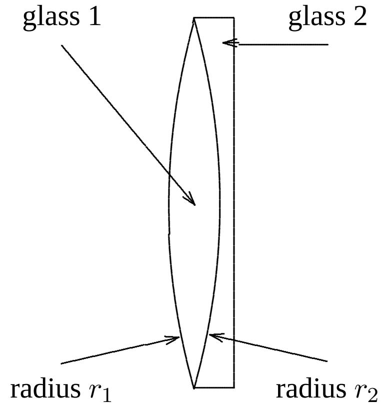
c. What are the boundary conditions at $r = r_b$ and why?

d. Find B/A and ϵ/A in the limit,

$$k, k' \gg \frac{1}{r_b}$$

in which you can drop terms proportional to $1/r_b$ compared to k or k'.

11.14. One of the problems with glass lenses is that the index of refraction of glass depends on frequency. Thus, according to the lens maker's formula, the focal length of a glass lens will depend of frequency, and that is not good, because if one color is focused sharply, the others will be fuzzy. This is called "chromatic aberration." Fortunately, different kinds of glass have different behavior in this respect, and this makes it possible to eliminate chromatic aberration. Suppose that you make a lens that looks like this by gluing together lenses made of two different types of glass.

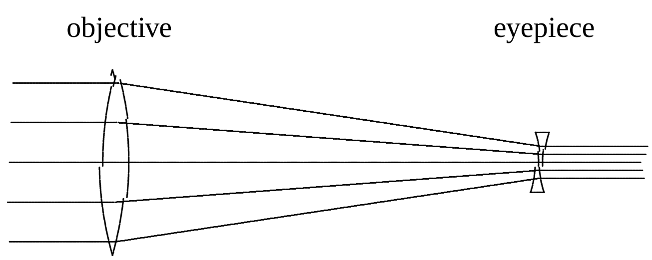


Suppose that the indices of refraction of the two glasses are

$$n_1(\lambda) = n_1^0 + \alpha_1\lambda\,, \qquad n_2(\lambda) = n_2^0 + \alpha_2\lambda\,. \tag{11.178}$$

What relation must be satisfied if the compound lens is to have a focal length that is independent of λ?

11.15. You can also make a telescope with one converging lens (the objective) and one diverging lens (the eyepiece).



The focal length of the convex lens is f_1 and the focal length of the concave lens is $-f_2$.

a. If the ray tracing works as shown, that is that parallel rays entering the objective are focused down to parallel rays leaving the eyepiece, find the distance, d, between the two lenses.

b. Compute the magnification by assuming that you are looking at a distance object which subtends an angular size θ. Then consider a ray at angle θ that passes through the center of the convex lens. By calculating where it passes through the concave lens, you should be able to determine its angle, θ_o, when it reaches the observers eye. The magnification is then θ_o/θ. What is it in terms of the focal lengths?

c. The image in this case is right-side-up. Draw a careful diagram to explain why.

11.16. The appearance of the rainbows depends dramatically on the index of refraction of water. Describe in detail what the rainbows look like if n were decreased by 0.03 for each frequency of light? Discuss the first and second rainbows and Alexander's dark band.

Chapter 12

Polarization

In this chapter, we return to (9.46)-(9.48) and examine the consequences of Maxwell's equations in a homogeneous material for a general traveling electromagnetic plane wave. The extra complication is polarization.

Preview

Polarization is a general feature of transverse waves in three dimensions. The general electromagnetic plane wave has two polarization states, corresponding to the two directions that the electric field can point transverse to the direction of the wave's motion. This gives rise to much interesting physics.

i. We introduce the idea of polarization in the transverse oscillations of a string.

ii. We discuss the general form of electromagnetic waves and describe the polarization state in terms of a complex, two-component vector, Z. We compute the energy and momentum density as a function of Z and discuss the Poynting vector. We describe the varieties of possible polarization states of a plane wave: linear, circular and elliptical.

iii. We describe "unpolarized light," and explain how to generate and manipulate polarized light with polarizers and wave plates. We discuss the rotation of the plane of linearly polarized light by optically active substances.

iv. We analyze the reflection and transmission of polarized light at an angle on a boundary between dielectrics.

12.1 The String in Three Dimensions

In most of our discussions of wave phenomena so far, we have assumed that the motion is taking place in a plane, so that we can draw pictures of the system on a sheet of paper. We have implicitly been restricting ourselves to two-dimensional waves. This is all right for longitudinal oscillations in three dimensions, because all the action is taking place along a single line. However, for transverse oscillations, going from two dimensions to three dimensions makes an enormous difference because there are two transverse directions in which the system can oscillate.

For example, consider a string in three dimensions, stretched in the z direction. Each point on the string can oscillate in both the x direction and the y direction. If the system were not approximately linear, this could be a horrendous problem. Linearity allows us to solve the problem of oscillation in the x-z plane separately from the problem of oscillation in the y-z plane. We have already solved these two-dimensional problems in chapter 5. Then we can simply put the results together to get the most general motion of the three-dimensional system. In other words, we can treat the x component of the transverse oscillation and the y component as completely independent.

Suppose that there is a harmonic traveling wave in the $+z$ direction in the string. The displacement of the string at z from its equilibrium position, $(0, 0, z)$, can be written as

$$\vec{\Psi}(z,t) = \mathrm{Re}\,\left[(\psi_1 \hat{x} + \psi_2 \hat{y})\; e^{i(kz-\omega t)}\right] \tag{12.1}$$

where $\hat{x}$ and $\hat{y}$ are unit vectors in the x and y direction and ψ_1 and ψ_2 are complex parameters describing the amplitude and phase of the oscillations in the x-z plane and the y-z plane,

$$\psi_j = A_j\, e^{i\phi_j} \quad \text{for } j = 1 \text{ to } 2\,. \tag{12.2}$$

It is convenient to arrange these parameters into a complex vector

$$Z = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} , \tag{12.3}$$

which gives a complete description of the motion of the string.

12.1.1 Polarization

12-1

"Polarization" refers to the nature of the motion of a point on the string (or other transverse oscillation). This motion is animated in program 12-1. You may want to read the discussion below with this program running.

If $\phi_1 = \phi_2$, or A_1 or A_2 is zero, then (12.3) represent a linearly polarized string. Linear polarization is easy to understand. It means that each point on the string is oscillating back and forth in a fixed plane. For example,

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{12.4}$$

$$u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{12.5}$$

represent strings oscillating in the x-z plane and the y-z plane respectively. A string oscillating in a plane an angle θ from the positive x axis (towards the positive y axis) is represented by

$$u_\theta = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} . \tag{12.6}$$

This is shown in the x-y plane in figure 12.1. The polarization vectors (12.4)-(12.6) can be multiplied by a phase factor, $e^{i\phi}$, without affecting the polarization state in any important way. This just corresponds to an overall resetting of the clock.

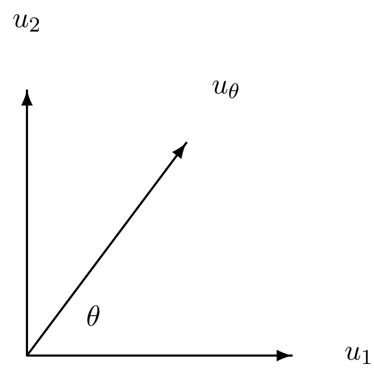


Figure 12.1: u_1, u_2 and u_θ.

More interesting is circular polarization. A circularly polarized wave in a string is represented by either

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \tag{12.7}$$

or

$$\begin{pmatrix} 1 \\ -i \end{pmatrix} . \tag{12.8}$$

In (12.7), the y component lags behind the x component by $\pi/2$ (=ϕ_2). Thus, at any fixed point in space, the field rotates from x to y, or in the counterclockwise direction viewed from the positive z axis (with the wave coming at you), as shown in figure 12.2. This is called "left-circular polarization" because the string resembles a left-handed screw. Likewise,

$$\begin{pmatrix} 1 \\ -i \end{pmatrix} . \tag{12.8}$$

represents clockwise rotation of the string. This is called "right-circular polarization."

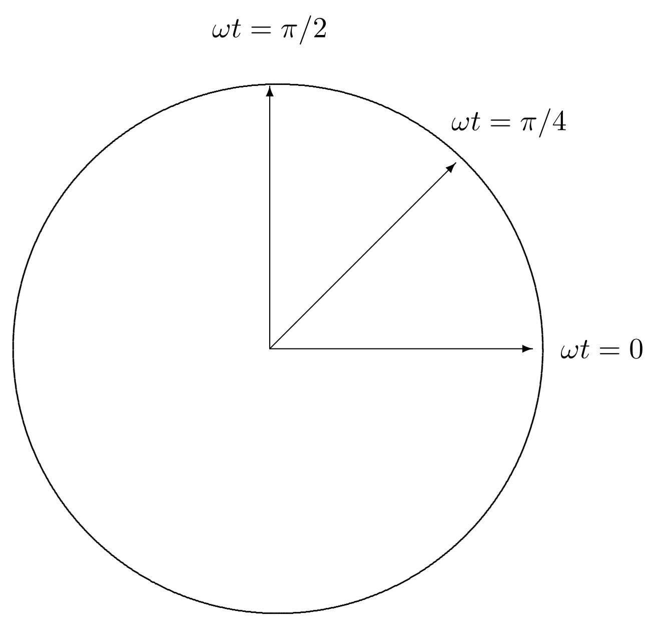


Figure 12.2: Circular polarization.

The vector

$$\begin{pmatrix} A \\ iB \end{pmatrix} \tag{12.9}$$

with $A > B > 0$ represents elliptical polarization. A point on the string traces out an ellipse with semi-major axis A along the 1 axis and semi-minor axis B along the 2 axis, with counterclockwise rotation, as shown in figure 12.3

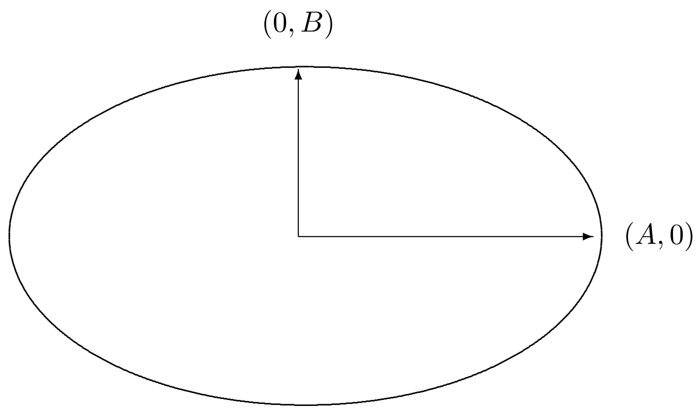


Figure 12.3: Elliptical polarization with long axis in the x direction.

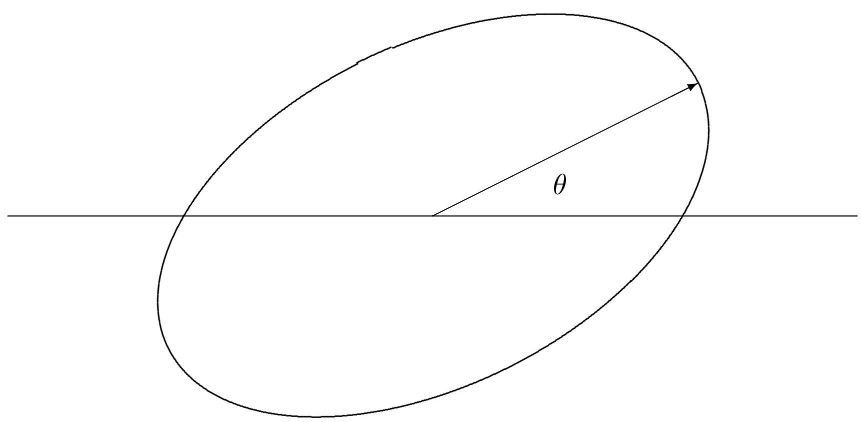


Figure 12.4: General elliptical polarization.

A completely general vector can be written in the following form:

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = e^{i\phi} \begin{pmatrix} A\cos\theta - iB\sin\theta \\ A\sin\theta + iB\cos\theta \end{pmatrix} \tag{12.10}$$

with $A \geq |B|$ and $0 \leq \theta < \pi$ and ϕ is real phase (which is not very relevant relevant to the physics but can be there to make the math look uglier). This represents elliptical polarization with semi-major axis A at an angle θ with the 1 axis, as in

$$u_\theta = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} . \tag{12.6}$$

and semi-minor axis B as shown in figure 12.4. If B is positive (negative), the rotation is counterclockwise (clockwise). The physically interesting parameters A, B and θ can be found from ψ_1 and ψ_2 as follows:

$$\begin{aligned} A^2 + B^2 &= |\psi_1|^2 + |\psi_2|^2 \,, \\ AB &= -\mathrm{Im}\,(\psi_1\,\psi_2^*) \,. \end{aligned} \tag{12.11}$$

Thus,

$$A \pm B = \sqrt{|\psi_1|^2 + |\psi_2|^2 \mp 2\mathrm{Im}\,(\psi_1\,\psi_2^*)} \,, \tag{12.12}$$

gives A and B. Then θ satisfies

$$\begin{aligned} (A^2 - B^2)\cos 2\theta &= |\psi_1|^2 - |\psi_2|^2 \,, \\ (A^2 - B^2)\sin 2\theta &= 2\mathrm{Re}\,(\psi_1\,\psi_2^*) \,. \end{aligned} \tag{12.13}$$

Notice that the overall phase factor $e^{i\phi}$ cancels out in (12.11)-(12.13).

12.2 Electromagnetic Waves

12.2.1 General Electromagnetic Plane Waves

12-1

We saw in chapters 8 and 9 that an electromagnetic plane wave traveling in the $+z$ direction looks like this,

$$E_x(z,t) = \varepsilon_x e^{i(kz-\omega t)} \,, \quad E_y(z,t) = \varepsilon_y e^{i(kz-\omega t)} \,, \tag{12.14}$$

$$B_x(z,t) = \beta_x e^{i(kz-\omega t)} \,, \quad B_y(z,t) = \beta_y e^{i(kz-\omega t)} \,, \tag{12.15}$$

$$E_z(z,t) = B_z(z,t) = 0 \,, \tag{12.16}$$

where the βs are determined by Maxwell's equations to be

$$\beta_y = \frac{n}{c}\varepsilon_x \,, \quad \beta_x = -\frac{n}{c}\varepsilon_y \,. \tag{12.17}$$

As usual, we have written the wave with the irreducible time dependence, $e^{-i\omega t}$. To get the real electric and magnetic fields, we take the real part of (12.14)-(12.15). Note, in particular, that the constants ε_j and β_j for $j = x$ and y may be complex.

The restriction to motion in the z direction is not important. Because the physics of Maxwell's equations is invariant under rotations in three-dimensional space, we can write down the form of a plane wave moving with an arbitrary $\vec{k}$ vector by extracting the features of (12.14)-(12.17) that do not depend on the direction. These are:

i. $\vec{k}$, $\vec{E}$ and $\vec{B}$ are mutually orthogonal vectors,

$$\vec{k}\cdot\vec{E} = \vec{k}\cdot\vec{B} = \vec{E}\cdot\vec{B} = 0\,; \tag{12.18}$$

ii. $\vec{B}$ is determined by the cross product

$$\vec{B} = \frac{n}{c}\hat{k}\times\vec{E} = \frac{1}{\omega}\vec{k}\times\vec{E}\,, \tag{12.19}$$

where $\hat{k}$ is a unit vector in the direction of the $\vec{k}$ vector, the direction of propagation of the wave.

These two conditions imply that a general real electromagnetic plane wave can be written as

$$\begin{aligned}\vec{E} &= \text{Re}\,\left(\vec{e}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right)\\ \vec{B} &= \text{Re}\,\left(\vec{b}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right)\end{aligned} \tag{12.20}$$

where the vectors, $\vec{e}$ and $\vec{b}$, are complex, in general, satisfying

$$\vec{b}(\vec{k}) = \frac{1}{\omega}\,\vec{k}\times\vec{e}(\vec{k}) = \frac{n}{c}\,\hat{k}\times\vec{e}(\vec{k}) \quad \text{and} \quad \hat{k}\cdot\vec{e}(\vec{k}) = 0\,. \tag{12.21}$$

There are two things to note about the relations, (12.21).

i. It is enough to specify the direction of the electric field, $\vec{e}(\vec{k})$. The magnetic field is then determined by (12.21). The vector, $\vec{e}$ is called the **"polarization"** of the electromagnetic wave.

ii. Because of (12.21), the polarization is perpendicular to $\vec{k}$, and thus lives in a two-dimensional vector space.

In the two-dimensional space perpendicular to $\vec{k}$, we can choose a basis of real vectors, $\hat{e}_1$ and $\hat{e}_2$, where

$$\hat{e}_1\cdot\hat{k} = \hat{e}_2\cdot\hat{k} = \hat{e}_1\cdot\hat{e}_2 = 0\,, \quad \hat{e}_1\times\hat{e}_2 = \hat{k}\,. \tag{12.22}$$

For example, for a plane wave traveling in the z direction, $\hat{k} = \hat{z}$, we could take $e_1 = \hat{x}$ and $e_2 = \hat{y}$. Then we can write

$$\vec{e}(\vec{k}) = \psi_1\,\hat{e}_1 + \psi_2\,\hat{e}_2\,. \tag{12.23}$$

The components, ψ_1 and ψ_2 go into the the two-dimensional vector, (12.3), that describes the polarization state of the electromagnetic wave, just as it describes the polarization state of the string.[1] We can always go back to the components of the electric field using (12.23) and (12.20) and then find the magnetic field using (12.21).

[1] This is sometimes called the Jones vector. See Hecht, page 323.

Now the entire discussion of transverse waves on a string from (12.4) to (12.13) can be taken over to describe polarized light. The direction of displacement of the string goes over directly into the direction of the electric field. Thus the animation in program 12-1 applies just as well to the electric field in a polarized wave as to polarization in a string.

12.2.2 Energy and Intensity

The energy density in an electromagnetic field is

$$\mathcal{E} = \frac{1}{2}\left(\epsilon\,\vec{E}^2 + \frac{1}{\mu}\,\vec{B}^2\right) . \tag{12.24}$$

Because the energy density is a nonlinear function of the field strengths, we must use **real** fields in (12.24). The momentum density is

$$\vec{\mathcal{P}} = \epsilon\,\vec{E}\times\vec{B}\,. \tag{12.25}$$

The Poynting vector, a measure of energy flow, is

$$\vec{S} = c^2\vec{\mathcal{P}}\,. \tag{12.26}$$

These quantities satisfy

$$\frac{\partial}{\partial t}\,\mathcal{E} + \vec{\nabla}\cdot\vec{S} = 0\,. \tag{12.27}$$

The Poynting vector is useful because it measures the intensity of the wave, the energy per unit time per unit area carried by the electromagnetic wave. The relation, (12.27), then expresses conservation of energy. The sum of the change in the energy density at any point plus the energy flowing away from it is zero.

To see what these quantities look like in terms of the vector, Z, let us compute the electric and magnetic fields explicitly using (12.20) and (12.21):

$$\begin{aligned}\vec{E} &= \mathrm{Re}\,\left(\vec{e}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right)\\ \vec{B} &= \mathrm{Re}\,\left(\vec{b}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right)\end{aligned} \tag{12.20}$$

$$\vec{b}(\vec{k}) = \frac{1}{\omega}\,\vec{k}\times\vec{e}(\vec{k}) = \frac{n}{c}\,\hat{k}\times\vec{e}(\vec{k}) \quad \text{and} \quad \hat{k}\cdot\vec{e}(\vec{k}) = 0\,. \tag{12.21}$$

The result is

$$\begin{aligned}\vec{E} &= A_1\,\hat{e}_1\cos(\vec{k}\cdot\vec{r}-\omega t+\phi_1) + A_2\hat{e}_2\cos(\vec{k}\cdot\vec{r}-\omega t+\phi_2)\,,\\ \vec{B} &= \sqrt{\mu\epsilon}\left(A_1\,\hat{e}_2\cos(\vec{k}\cdot\vec{r}-\omega t+\phi_1) - A_2\hat{e}_1\cos(\vec{k}\cdot\vec{r}-\omega t+\phi_2)\right)\,.\end{aligned} \tag{12.28}$$

Putting this into (12.24) and (12.26) gives

$$\mathcal{E} = \frac{\epsilon}{4\pi}\left(A_1^2\cos^2(\vec{k}\cdot\vec{r}-\omega t+\phi_1)+A_2^2\cos^2(\vec{k}\cdot\vec{r}-\omega t+\phi_2)\right) , \tag{12.29}$$

$$\vec{S} = \hat{k}\sqrt{\frac{\epsilon}{\mu}}\,\frac{c}{4\pi}\left(A_1^2\cos^2(\vec{k}\cdot\vec{r}-\omega t+\phi_1)+A_2^2\cos^2(\vec{k}\cdot\vec{r}-\omega t+\phi_2)\right) . \tag{12.30}$$

You can check explicitly that (12.27) is satisfied. Because ω is very large for interesting electromagnetic waves, we are almost always interested in only the time averaged values of $\mathcal{E}$ and $\vec{S}$. These are

$$\langle\mathcal{E}\rangle = \frac{\epsilon}{8\pi}\left(A_1^2+A_2^2\right) , \tag{12.31}$$

$$\left\langle\vec{S}\right\rangle = \hat{k}\sqrt{\frac{\epsilon}{\mu}}\,\frac{c}{8\pi}\left(A_1^2+A_2^2\right) . \tag{12.32}$$

Note that the time averaged values depend only on the quantity

$$|Z|^2 \equiv |\psi_1|^2+|\psi_2|^2 = A_1^2+A_2^2 . \tag{12.33}$$

The intensity of the light is proportional to $|Z|^2$.

12.2.3 Circular Polarization and Spin

Although linear polarization is more familiar and perhaps easier to understand, there is a sense in which circular polarization is the more fundamental. The plane electromagnetic wave in the $\hat{k}$ direction can be rotated around the $\hat{k}$ axis without changing anything but its polarization state. The rotation symmetry of the physics suggests that we ought to be able to find states that behave simply under such a rotation, and just get multiplied by a phase factor. These states are, in fact, the circular polarization states. The action of a rotation by an angle θ about the $\hat{k}$ axis on the polarization vector, Z, is represented by the matrix

$$R_\theta = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} . \tag{12.34}$$

For example, R_θ acting on u_1, (12.4), gives u_θ, (12.6):

$$R_\theta\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} . \tag{12.35}$$

But on the left- and right-circularly polarized states,

$$R_\theta\begin{pmatrix}1\\i\end{pmatrix} = e^{-i\theta}\begin{pmatrix}1\\i\end{pmatrix} , \quad R_\theta\begin{pmatrix}1\\-i\end{pmatrix} = e^{i\theta}\begin{pmatrix}1\\-i\end{pmatrix} . \tag{12.36}$$

This is related to the fact that the circularly polarized states carry the maximum angular momentum possible, which in turn is related to the quantum mechanical property of the spin of the photon.

12.3 Wave Plates and Polarizers

One reason that polarization is important is that the polarization state of an electromagnetic wave can be easily manipulated. Two of the most important devices for such manipulation are polarizers and wave plates.

12.3.1 Unpolarized Light

12-2

In any beam of light, at any given point and time, the electric field points in a particular direction. Likewise, because any plane electromagnetic wave with a definite angular frequency can be described by (12.20) and (12.21),

$$\begin{aligned} \vec{E} &= \mathrm{Re}\,\left(\vec{e}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right) \\ \vec{B} &= \mathrm{Re}\,\left(\vec{b}(\vec{k})\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\right) \end{aligned} \tag{12.20}$$

$$\vec{b}(\vec{k}) = \frac{1}{\omega}\,\vec{k}\times\vec{e}(\vec{k}) = \frac{n}{c}\,\hat{k}\times\vec{e}(\vec{k}) \quad \text{and} \quad \hat{k}\cdot\vec{e}(\vec{k}) = 0\,. \tag{12.21}$$

every plane wave is polarized. However, in an “unpolarized” beam, the light wave consists of a range of angular frequencies with different polarizations. As a result of the interference of the different harmonic components of the wave, the polarization wanders more or less randomly as a function of time and space, and on the average, no particular polarization is picked out. A simple example of what this looks like is animated in program 12-2, where we plot an electric field of the form

$$\begin{aligned} E_x(t) &= \cos(\omega_1 t + \phi_1) + \cos(\omega_2 t + \phi_2)\,, \\ E_y(t) &= \cos(\omega_3 t + \phi_3) + \cos(\omega_4 t + \phi_4)\,, \end{aligned} \tag{12.37}$$

where the phases are random and the frequencies are chosen at random in a small range around a central frequency. You can watch the $\vec{E}$ field wandering in the x-y plane, eventually filling it up. The narrower the range of frequencies in the wave, the more slowly the polarization wanders. In the example in program 12-2, the range of frequencies is of the order of 10% of the central frequency, so the polarization wanders rapidly. But for a beam with a fairly well-defined frequency, the polarization will be nearly constant over many cycles of the wave. The time over which the polarization is approximately constant is called the coherence time of the wave. For a plane wave of definite frequency, the coherence time is infinite.

12.3.2 Polarizers

A **"polarizer"** is a device that allows light polarized in a particular direction (the "easy transmission axis" of the polarizer) to pass through with very little absorption, but absorbs most of the light polarized in the perpendicular direction. Thus an unpolarized light beam, passing though the polarizer, emerges polarized along the easy axis.

For the transverse oscillations of a string, a polarizer is simply a slit that allows the string to oscillate in one transverse direction but not in the perpendicular direction.

For electromagnetic waves, the most familiar example of a polarizer, Polaroid, was invented by Edwin Land over 50 years ago, partly in experiments done in the attic of the Jefferson Physical Laboratory, where he worked as an undergraduate at Harvard. The idea of polaroid is to make material that conducts electricity (poorly) in one direction, but not in the other. Then the electric field in the conducting direction will be absorbed (the energy going to resistive loss), while the electric field in the nonconductive direction will be unaffected. One way of doing this is to make sheets of polymer (polyvinyl alcohol) stretched (to align the polymer molecules along a preferred axis) and doped with iodine (to allow conduction along the polymer molecules).[2]

12.3.3 Wave Plates

"Wave plates" are optical elements that change the relative phase of the two components of Z. Wave plates are possible because there are materials in which the index of refraction depends on the polarization. This property is called "birefringence." It can happen in various ways.

For example, the transparent polymer material cellophane is made into thin sheets by stretching. Because of the stretching, the polymer strands tend to be oriented along the stretch direction. The dielectric constant in this material depends on the direction of the electric field. It is easier for charges to move along the polymer strands than across them. Thus the dielectric constant is larger for electric fields in the stretch direction.

The same effect may arise because of the inherent structure of a transparent crystal. An example is the naturally occurring mineral, calcite, a crystalline form of calcium carbonate, $CaCO_3$. Crystals of calcite have the fascinating property of splitting a beam of unpolarized light into its two polarization states. Birefringence can even be produced mechanically, by stressing a transparent material, squeezing the electronic structure in one direction.

However the birefringence is produced, we can make a wave plate by orienting the material so that the x and y directions correspond to different indices of refraction, n_x and n_y, and then making a slice of the material in the form of a plate in the x-y plane, with some thickness ℓ in the z direction. Now an electromagnetic wave traveling in the z direction through the

[2]See Sears, Zemansky and Young, page 813.

plate has different k values depending on its polarization:

$$k = \begin{cases} \dfrac{n_x}{c}\,\omega \text{ for polarization in the } x \text{ direction} \\ \dfrac{n_y}{c}\,\omega \text{ for polarization in the } y \text{ direction} \end{cases} \tag{12.38}$$

In particular, the phase **difference**, between x and y polarized light in going through the plate is

$$\Delta\phi = \frac{n_x - n_y}{c}\,\omega\,\ell\,. \tag{12.39}$$

Note that in general the phase difference, $\Delta\phi$, depends on the frequency of the light. Even if n_x and n_y depend on frequency, it would be a bizarre accident if that dependence canceled the ω dependence from the explicit factor of ω in (12.39).

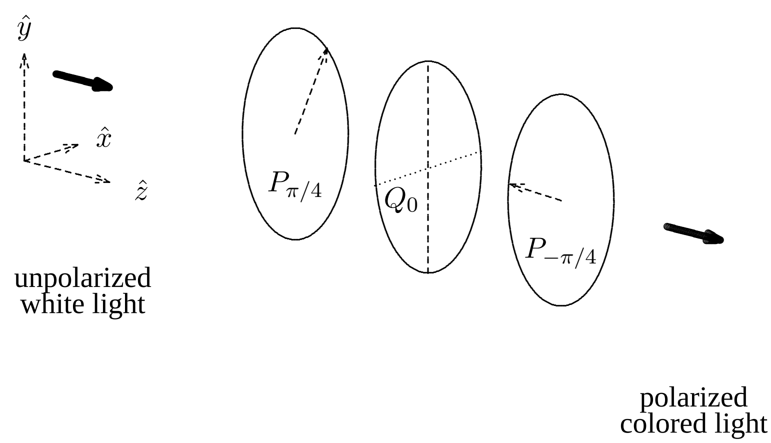


Figure 12.5: Initially unpolarized light passing through a pair of crossed polarizers with a wave plate in between.

Consider, now, putting such a wave plate between two crossed polarizers, oriented at $\pm 45°$, as shown in figure 12.5. Without the wave plate, no light would get through because the first polarizer transmits only light polarized at $45°$, described by the Z vector

$$Z = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \tag{12.40}$$

and the second polarizer absorbs it.

Coming out of the first polarizer, the vector, Z, looks like (12.40) for all the frequency components in the white light. But when the wave plate is inserted in between, a frequency dependent phase difference is added, so that the Z vector coming out of the wave plate (up to an irrelevant overall phase) looks like

$$Z = \begin{pmatrix} 1/\sqrt{2} \\ e^{-i\Delta\phi}/\sqrt{2} \end{pmatrix} . \tag{12.41}$$

For frequencies such that $e^{-i\Delta\phi}$ is -1, the light is polarized in the $-45°$ direction, and gets through the second polarizer without further attenuation. But for frequencies such that $e^{-i\Delta\phi}$ is 1, the light is still absorbed by the second polarizer. Intermediate frequencies are partially absorbed.

It is this frequency dependence that produces the interesting patterns of color that you see when you put cellophane or a stressed piece of plastic between polarizers.

12.3.4 Matrices

The effects of wave plates and polarizers and the like can be summarized by multiplication of the Z vector by 2×2 matrices. For example, a perfect polarizer with an axis at an angle θ from the 1 axis can be represented by

$$P_\theta = \begin{pmatrix} \cos^2\theta & \cos\theta\,\sin\theta \\ \cos\theta\,\sin\theta & \sin^2\theta \end{pmatrix} . \tag{12.42}$$

The object P_θ is called a **"projection operator,"** because it projects the vector onto the direction parallel to u_θ. It satisfies

$$P_\theta P_\theta = P_\theta \,, \tag{12.43}$$

as it must, since the first polarizer produces polarized light and the second one transmits it perfectly. P_θ acting on a vector transmits the component in the θ direction. This is easiest to visualize if $\theta = 0$ or $\pi/2$. The matrices

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} , \quad P_{\pi/2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} , \tag{12.44}$$

represent polarizers along the 1 and 2 axes respectively.

A wave plate in which the phase difference is $\pi/2$ is called a "quarter wave plate." For a wave plate in which the phase difference is between 0 and π, it is conventional to call the axis with the smaller phase the "fast axis." A quarter wave plate with fast axis along the 1 axis is represented by

$$Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} . \tag{12.45}$$

Notice that we can write

$$Q_0 = P_0 + iP_{\pi/2} \,. \tag{12.46}$$

This should convince you that in general if the fast axis is in the θ direction, the quarter wave plate looks like

$$Q_\theta = P_\theta + iP_{\theta+\pi/2} \,. \tag{12.47}$$

The discussion of (12.39) shows that in general, a wave plate will only be a quarter wave plate for light of a definite frequency.

A wave plate in which the phase difference is π is called a "half wave plate." A half wave plate is obtained by replacing the i in (12.45)-(12.47) by -1. Thus,

$$H_\theta = P_\theta - P_{\theta+\pi/2} \,. \tag{12.48}$$

Notice that

$$H_\theta = Q_\theta Q_\theta \,; \tag{12.49}$$

two quarter wave plates make a half wave plate.

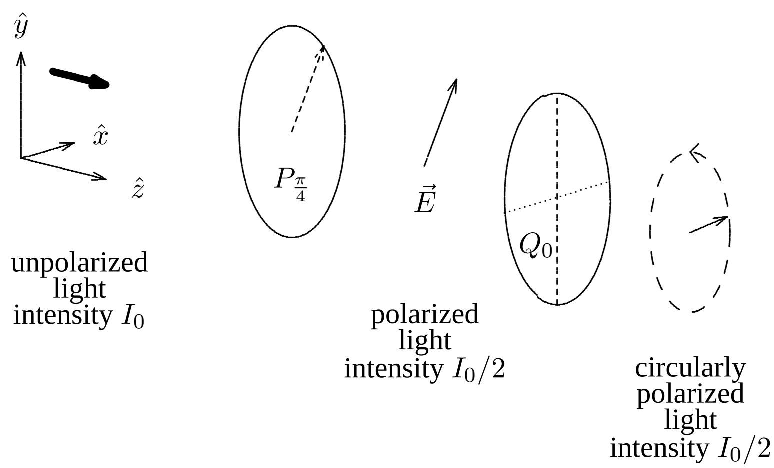


Figure 12.6: Producing circularly polarized light.

Here are two amusing devices that you can make with these optical elements (or matrices). Consider the combination of first a polarizer at $45°$ and then a quarter wave plate, as shown in figure 12.6. By forming the matrix product, $Q_0\, P_{\pi/4}$, you can see that this produces counterclockwise circularly polarized light from anything with a component of polarization in the $\pi/4$ direction. The argument goes like this. The product is

$$Q_0\, P_{\pi/4} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ i/2 & i/2 \end{pmatrix} \,. \tag{12.50}$$

When this acts on an arbitrary vector you get circularly polarization unless the vector is annihilated by $P_{\pi/4}$.

$$Q_0\, P_{\pi/4} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \frac{\psi_1 + \psi_2}{2} \begin{pmatrix} 1 \\ i \end{pmatrix} \,. \tag{12.51}$$

In the opposite order, $P_{\pi/4}\,Q_0$ is an analyzer for circularly polarized light. It annihilates counterclockwise light and converts clockwise polarized light to light linearly polarized in the $\pi/4$ direction.

12.3.5 Optical Activity

"Optical activity" is a property of many organic and some inorganic compounds. An optically active material rotates the polarization of light without absorbing either component of the polarization. A familiar example of such a material is corn syrup, a thick aqueous solution of sugar that you probably have in your kitchen. If you put a rectangular container of corn syrup between polarizers, as shown in figure 12.7, and rotate the second polarizer until the intensity of the light getting through is a maximum, you will find that direction of the second polarizer is not the same as that of the first. The plane of the polarization has been rotated by some angle θ. The rotation angle, θ, is proportional to the thickness of the container, the length of the region of syrup that the light goes through.

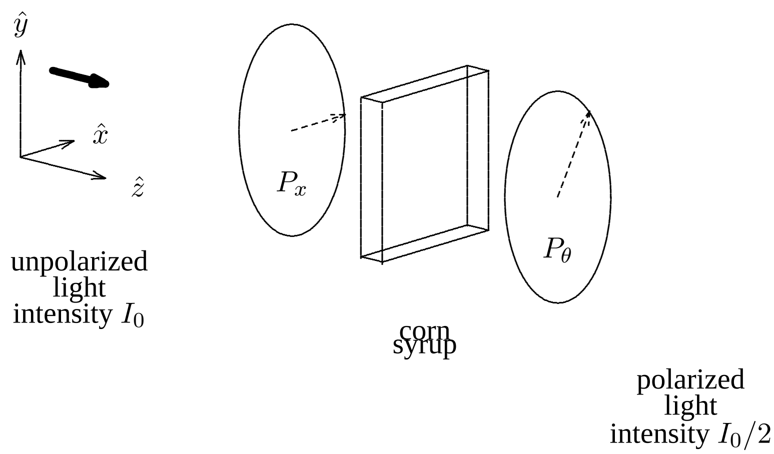


Figure 12.7: A rectangular container of corn syrup between polarizers.

Clearly, the optical activity of corn syrup cannot depend on crystal structure, because the stuff is a perfectly uniform liquid, completely invariant under rotations in three-dimensional space. It can have no special axes, or any such thing. Optical activity must work very differently from birefringence.

You can find a clue to the nature of optical activity by considering what it looks like if you look at it in a mirror. If you reflect the system illustrated in figure 12.7 in the x-z plane, by changing the sign of all the y coordinates, the angle θ changes to $-\theta$. Thus the corn syrup that you see in a mirror must be fundamentally different from the corn syrup in your kitchen. This is not so strange. After all, your right hand looks like a left hand when you look at it in a mirror. The corn syrup must have the same property and have a definite "handedness." In fact, because of the tetrahedral bonding of the carbon atoms of which they are built, the sugar molecules in the corn syrup can and do have such a handedness.

Because of the handedness of the sugar molecules, the index of refraction of the corn syrup actually depends on the handedness of the light. It is slightly different for left- and right-circularly polarized light. This happens because the $\vec{E}$ field of a circularly polarized beam twists slightly as it traverses each sugar molecule and sees a slightly different electronic structure depending on the direction of the twist. Then, because the indices of refraction are slightly different, the left- and right-circularly polarized components of the light get different phase factors ($k\ell$) in passing through a thickness, ℓ, of the syrup.

We can now use our matrix language to see how this leads to optical activity. Up to an irrelevant overall phase, we can choose the phase produced on the left-circularly polarized light to be $-\theta$ and that on the right-circularly polarized light to be θ. Then we can represent the action of the syrup on an arbitrary wave by the matrix

$$e^{-i\theta}\,P_+ + e^{i\theta}\,P_- \,, \tag{12.52}$$

where $P_\pm$ are matrices that pick out the left- and right-circularly polarized components, respectively. They satisfy

$$P_\pm \begin{pmatrix} 1 \\ \pm i \end{pmatrix} = \begin{pmatrix} 1 \\ \pm i \end{pmatrix} \,, \quad P_\pm \begin{pmatrix} 1 \\ \mp i \end{pmatrix} = 0 \,. \tag{12.53}$$

You can check that the matrices are

$$P_\pm = \frac{1}{2} \begin{pmatrix} 1 & \mp i \\ \pm i & 1 \end{pmatrix} \,. \tag{12.54}$$

Then (12.52) becomes

$$e^{-i\theta}\,\frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} + e^{i\theta}\,\frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \,. \tag{12.55}$$

This is just the rotation matrix R_θ, of (12.34)! R_θ rotates both components of any light by an angle θ.

One might wonder about the reason for the handedness of the sugar molecules. In fact, there are physical processes, the weak interactions that give rise to β-radioactivity, that look different when reflected in a mirror[3] and thus in principle could distinguish between left-handed and right-handed molecules. However, these interactions are most likely irrelevant to the handedness of corn syrup. Probably, the reason is biology rather than physics. Long ago, when the beginnings of life emerged from the primordial soup, **purely by accident**, the right-handed sugars were used. From then on, the handedness was maintained by the processes of reproduction.

[3]They violate what is called "parity" symmetry.

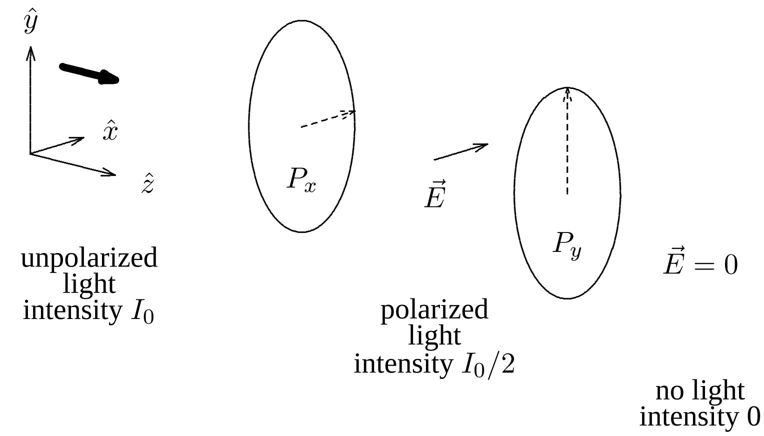


Figure 12.8: Initially unpolarized light passing through a pair of crossed polarizers.

12.3.6 Crossed Polarizers and Quantum Mechanics

Polarization offers many opportunities to get confused when you think of the light wave in terms of photons. Let us imagine turning down the intensity of the light to the point where one photon at a time is going through the polarizers and consider first the deceptively simple situation of light moving in the z direction through crossed polarizers in the x-y plane. Suppose that the first polarizer transmits light polarized in the x direction, and the second transmits light polarized in the y direction. This is deceptively simple because it seems that we can interpret what is going on simply in terms of photons. The situation is depicted in figure 12.8. This seems simple enough to interpret in terms of photons. The unpolarized light in region I is composed equally of photons polarized in the x direction and in the y direction (goes the wrong "classical" argument). Those polarized in the x direction get through the first polarizer, so half the photons are still around in region II, where the intensity is reduced by half. Then none of these get through the second polarizer, so that the intensity in region III is zero.

But compare this with the apparently similar situation in which the second polarizer transmits light polarized at $45°$ in the x-y plane, as shown in figure 12.9. Now the wave description tells us that the intensity in region III is reduced by another factor of 2 from that in region II. This is impossible to interpret in terms of classical particles. To see this, it is only necessary to turn down the intensity so that only one photon comes through at a time. Then the first polarizer is OK. As before, if the photon is polarized in the x direction, it get through. But now what happens at the second polarizer. The photon cannot split up. Either it gets through or it doesn't. To be consistent with the wave description, in which the intensity is reduced by another factor of two, the transmission at the second polarizer must be a probabilistic event. Half the time the photon gets through. Half the time it is absorbed. There is no way for the

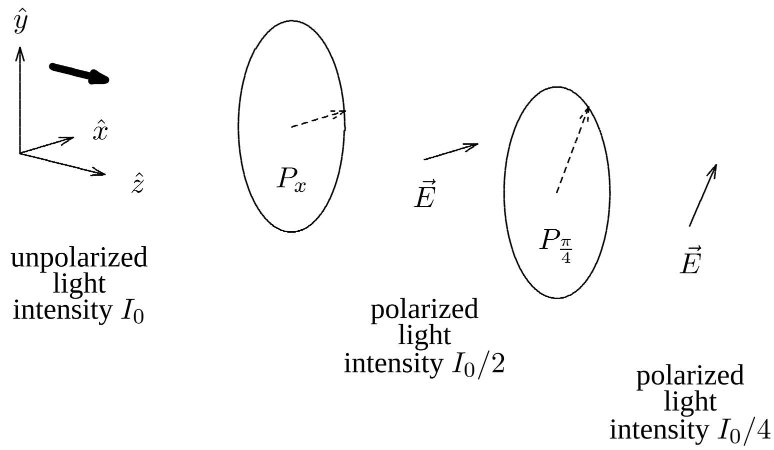


Figure 12.9: Initially unpolarized light passing through a pair of polarizers at with axes at $45°$.

photon in region II to tell whether it is going to make it! It is random. God plays dice.

12.4 Boundary between Dielectrics

Let us return to the infinite plane boundary between two dielectrics that we discussed in chapter 9, but now consider an electromagnetic wave coming in at an arbitrary angle. As in chapter 5, we will assume that the boundary is the $z = 0$ plane, and that for $z < 0$ we have dielectric constant ϵ, while for $z > 0$, dielectric constant ϵ'. We assume $\mu = 1$ everywhere.

On the general grounds of translation invariance and local interactions discussed in the previous chapter, all the components of the electric and magnetic fields will have the general form

$$\begin{aligned} \psi(r,t) &\propto e^{i\vec{k}\cdot\vec{r}} + R\, e^{i\vec{\tilde{k}}\cdot\vec{r}} \quad \text{for } z \le 0 \\ \psi(r,t) &\propto \tau\, e^{i\vec{k}'\cdot\vec{r}} \qquad\qquad \text{for } z \ge 0 \end{aligned} \tag{12.56}$$

where

$$\tilde{k}_x = k_x\,, \quad k'_x = k_x\,, \tag{12.57}$$

and

$$\begin{aligned} \tilde{k}_z &= -\sqrt{\omega^2/v^2 - k_x^2} = -k_z \\ k'_z &= \sqrt{\omega^2/v'^2 - k_x^2}\,. \end{aligned} \tag{12.58}$$

Thus Snell's law is satisfied, with θ and θ' defined as shown in figure 12.10.

$$\begin{aligned} k \cdot \sin\theta &= k' \, \sin\theta' \, . \\ |\vec{k}| &= \sqrt{\mu\epsilon}\, \frac{\omega}{c} = n\, \frac{\omega}{c} \\ n\, \sin\theta &= n' \, \sin\theta' \, . \end{aligned} \tag{12.59}$$

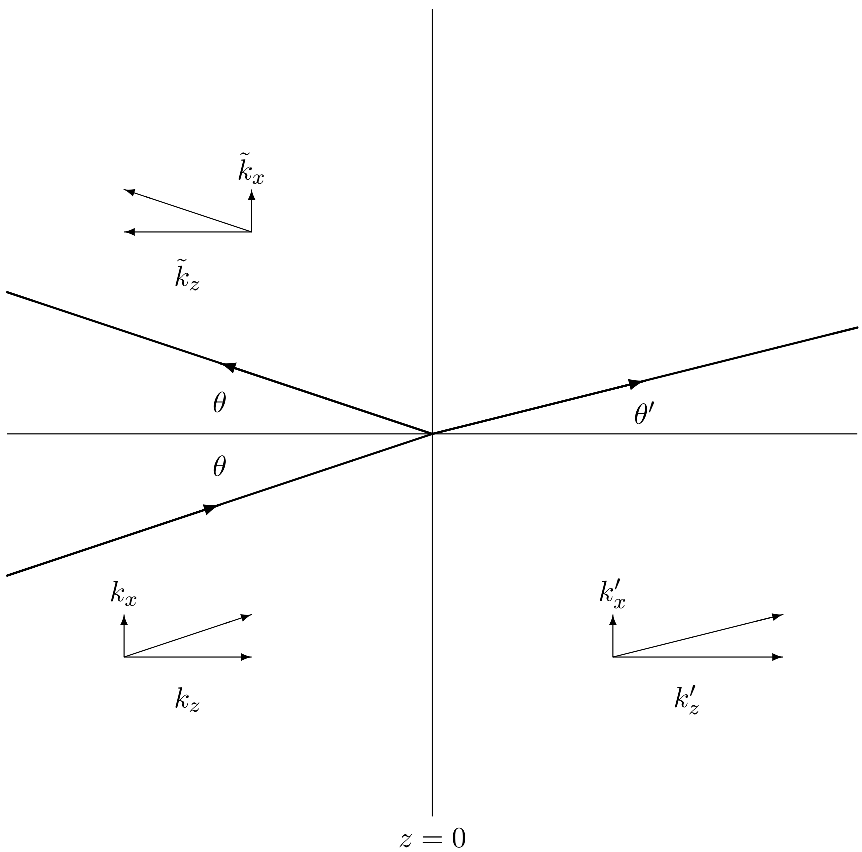


Figure 12.10: Scattering of plane waves from a plane boundary.

The details of the scattering will depend on the polarization. It is clear (by symmetry as usual) that the two cases will be polarization in the x-z plane and polarization perpendicular to the x-z plane. Of course, we lose nothing by considering these two separately, because of linearity. Any polarization for the incoming wave can be dealt with by forming a linear combination of the parallel and perpendicular solutions.

12.4.1 Polarization Perpendicular to the Scattering Plane

Let us first consider perpendicular polarization. This means that the electric field is in the y direction (out of the plane of the paper), while the magnetic field is the x-z plane:[4]

$$\begin{aligned} E_y(r,t) &= Ae^{i(\vec{k}\,\vec{r}-\omega t)} + R_\perp\, Ae^{i(\vec{\tilde{k}}\cdot\vec{r}-\omega t)} && \text{for } z \le 0 \\ E_y(r,t) &= \tau_\perp\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} && \text{for } z \ge 0 \\ E_z &= E_x = 0 \end{aligned} \tag{12.60}$$

Using (12.19))

$$\vec{B} = \frac{n}{c}\hat{k}\times\vec{E} = \frac{1}{\omega}\vec{k}\times\vec{E}\,, \tag{12.19}$$

we can write

$$\begin{aligned} B_x(r,t) &= -\frac{n}{c}\,A\cos\theta\, e^{i(\vec{k}\cdot\vec{r}-\omega t)} + \frac{n}{c}\,\cos\theta\, R_\perp\, Ae^{i(\vec{\tilde{k}}\cdot\vec{r}-\omega t)} && \text{for } z \le 0 \\ B_x(r,t) &= -\frac{n'}{c}\,\cos\theta'\,\tau_\perp\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} && \text{for } z \ge 0 \\ B_z(r,t) &= \frac{n}{c}\,\sin\theta\, Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} + \frac{n}{c}\,\sin\theta\, R_\perp\, Ae^{i(\vec{\tilde{k}}\cdot\vec{r}-\omega t)} && \text{for } z \le 0 \\ B_z(r,t) &= \frac{n'}{c}\,\sin\theta'\,\tau_\perp\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} && \text{for } z \ge 0 \\ B_y &= 0\,. \end{aligned} \tag{12.61}$$

The system is shown in figure 12.11. This figure shows the directions of the magnetic fields of the incoming ($\vec{B}_i$), reflected ($\vec{B}_r$), and transmitted ($\vec{B}_t$) component waves in scattering of an electromagnetic plane wave polarized parallel to a plane dielectric boundary. The $\vec{k}$ vectors are shown directly beneath the magnetic fields. The nontrivial boundary conditions are that E_y and B_x are continuous (the latter because we have assumed $\mu = 1$ so there is no sheet of bound current on the boundary). B_z is also continuous, but that provides no new information. Thus

$$1 + R_\perp = \tau_\perp \tag{12.62}$$

[4]The quantities, $R_\perp$ and $\tau_\perp$ in this section and $R_\|$ and $\tau_\|$ in the next are conventionally called “Fresnel coefficients.” See Hecht, page 97.

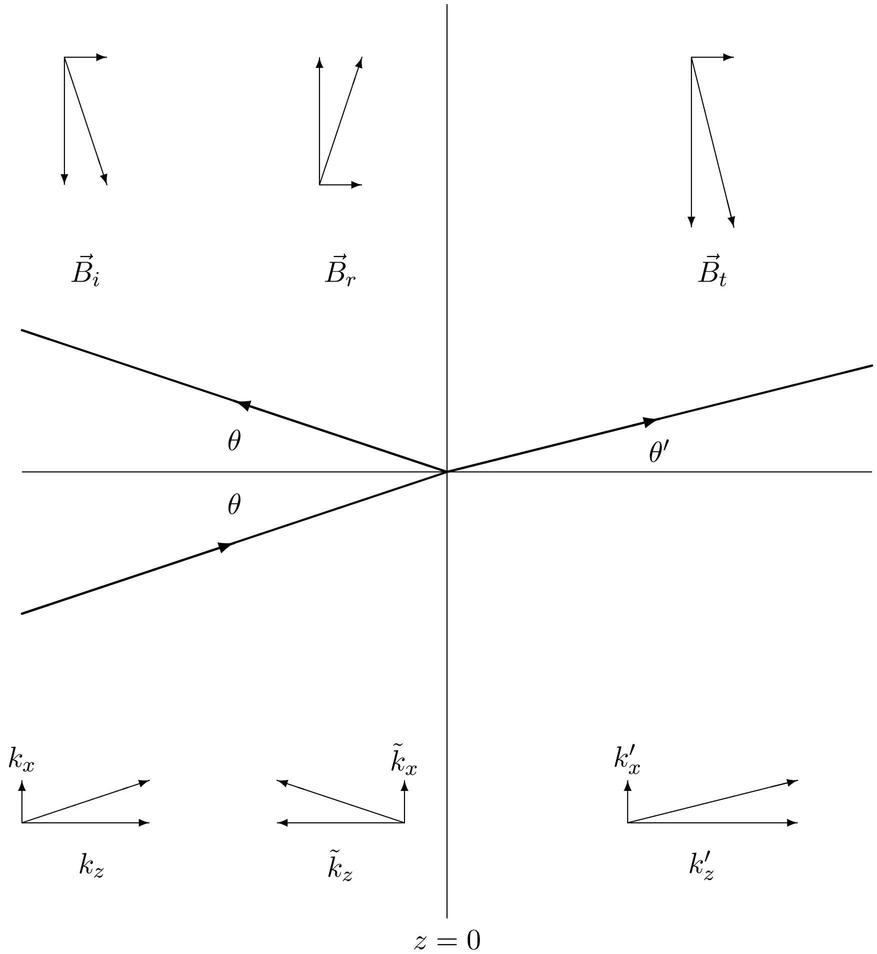


Figure 12.11: Scattering of an electromagnetic plane wave polarized parallel to a dielectric boundary.

$$n \cos\theta \, (1 - R_\perp) = n' \cos\theta' \, \tau_\perp \tag{12.63}$$

or because $n \propto |\vec{k}|$

$$k_z \, (1 - R_\perp) = k'_z \, \tau_\perp \, . \tag{12.64}$$

Thus

$$\tau_\perp = \frac{2}{1 + \xi_\perp} \tag{12.65}$$

$$R_\perp = \frac{1 - \xi_\perp}{1 + \xi_\perp} \tag{12.66}$$

where

$$\xi_\perp = \frac{k'_z}{k_z}\,. \tag{12.67}$$

12.4.2 Polarization in the Scattering Plane

Polarization in the x-z plane looks like

$$\begin{aligned} B_y(r,t) &= Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} + R_\parallel\, Ae^{i(\tilde{\vec{k}}\cdot\vec{r}-\omega t)} \quad \text{for } z \le 0 \\ B_y(r,t) &= \tau_\parallel\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} \quad \text{for } z \ge 0 \end{aligned} \tag{12.68}$$

$$B_z = B_x = 0\,,$$

where, for convenience, we have defined the reflection and transmission coefficients in terms of the magnetic fields, and

$$\begin{aligned} E_x(r,t) &= \frac{c}{n}\cos\theta\, Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} - \frac{c}{n}\cos\theta\, R_\parallel\, Ae^{i(-\tilde{\vec{k}}\cdot\vec{r}-\omega t)} \quad \text{for } z \le 0 \\ E_x(r,t) &= \frac{c}{n'}\cos\theta'\,\tau_\parallel\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} \quad \text{for } z \ge 0 \\ E_z(r,t) &= -\frac{c}{n}\sin\theta\, Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} - \frac{c}{n}\sin\theta\, R_\parallel\, Ae^{i(-\tilde{\vec{k}}\cdot\vec{r}-\omega t)} \quad \text{for } z \le 0 \\ E_z(r,t) &= -\frac{c}{n'}\sin\theta'\,\tau_\parallel\, Ae^{i(\vec{k}'\cdot\vec{r}-\omega t)} \quad \text{for } z \ge 0 \\ E_z &= 0\,. \end{aligned} \tag{12.69}$$

Now the nontrivial boundary conditions are the continuity of B_y and E_x. E_z is not continuous because a surface bound charge density builds up on the dielectric boundary. The boundary conditions yield

$$1 + R_\parallel = \tau_\parallel \tag{12.70}$$

$$\frac{\cos\theta}{n}\,(1 - R_\parallel) = \frac{\cos\theta'}{n'}\,\tau_\parallel \tag{12.71}$$

or

$$\tau_\parallel = \frac{2}{1+\xi_\parallel} \tag{12.72}$$

$$R_\parallel = \frac{1-\xi_\parallel}{1+\xi_\parallel} \tag{12.73}$$

where

$$\xi_{\parallel} = \frac{\cos\theta'/n'}{\cos\theta/n} = \frac{n^2\,k'_z}{n'^2\,k_z}\,. \tag{12.74}$$

One of the interesting things about (12.74) is that when

$$\frac{n^2\,k'_z}{n'^2\,k_z} = 1 \tag{12.75}$$

there is no reflection. This condition is satisfied for a special angle of incidence called Brewster's angle. We can understand the significance of Brewster's angle as follows:

$$\text{from Snell's law,} \quad \frac{n^2}{n'^2} = \frac{\sin^2\theta'}{\sin^2\theta} \tag{12.76}$$

$$\frac{k'_z}{k_z} = \frac{k_x/k_z}{k'_x/k'_z} = \frac{\tan\theta}{\tan\theta'} \tag{12.77}$$

$$\frac{n^2\,k'_z}{n'^2\,k_z} = \frac{\sin\theta'\cos\theta'}{\sin\theta\cos\theta} = 1\,. \tag{12.78}$$

Thus $\sin 2\theta = \sin 2\theta'$. Because $\theta \neq \theta'$ (that would be the trivial situation with no boundary), this means that

$$\theta = \pi/2 - \theta'\,. \tag{12.79}$$

In other words, Brewster's angle is defined by the condition that the reflected and transmitted plane waves are perpendicular, as shown in the diagram in figure 12.12. The relevance of this condition is that the reflected wave can be thought of as being produced by the motion of the charges on the boundary. But if these are moving in a direction perpendicular to the electric field in the would-be reflected wave, then the wave cannot be produced.

12.5 Radiation

In this section, we write down the electric and magnetic fields associated with changing charge and current densities.

12.5.1 Fields of moving charges

Because Maxwell's equations are partial differential equations, lots of initial conditions or boundary conditions must be specified to determine the solutions. For example, a constant electric field everywhere is a solution to the free-space Maxwell's equations, and therefore you can add a constant field to any solution and it will still be a solution. Such things must be determined by physical initial conditions or boundary conditions. One set of conditions that

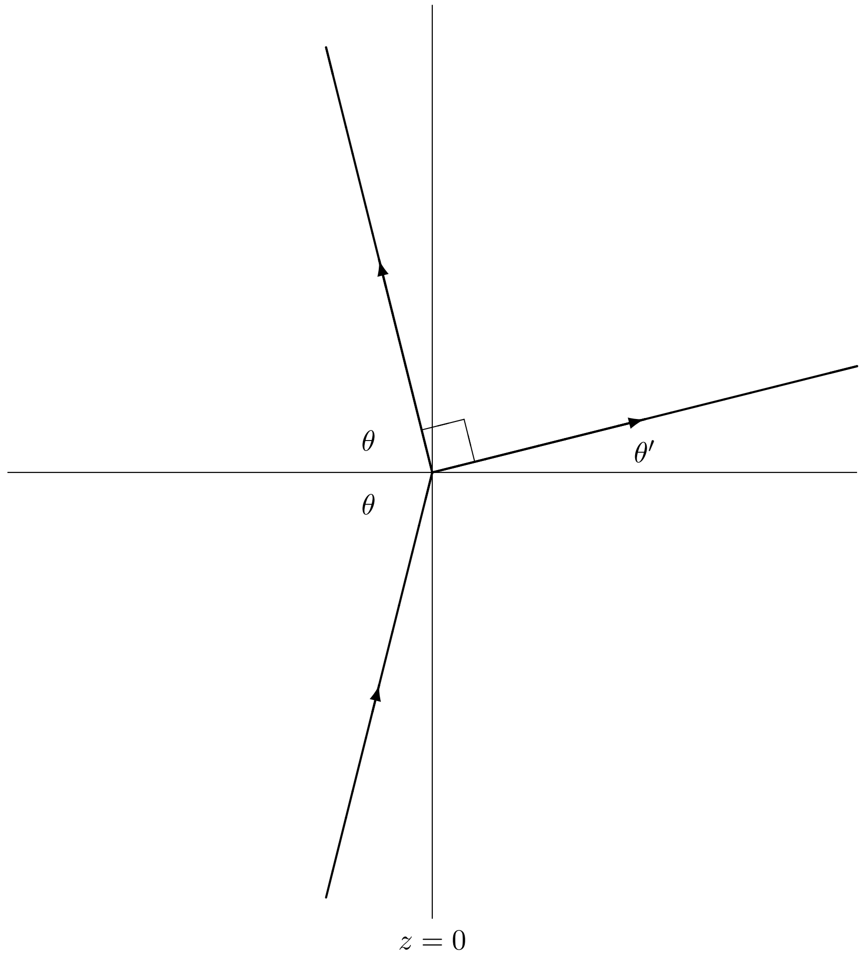


Figure 12.12: Brewster's angle.

is frequently interesting is an analog of the boundary condition at infinity that we discussed for one-dimensional waves. Suppose that you have a universe which is initially stationary, with no electric currents, no magnetic fields, and only electric fields due to stationary charges (which you know how to compute from Physics 15b). At some time, you begin to move charges around in some finite region of space. What are the electric and magnetic fields produced in this way? This question has a relatively simple answer that is a nice intuitive generalization of the relations you learned in 15b for the electric and vector potentials from stationary charge and current distributions. These relations were

$$\phi(\vec{r}) = \int d^3r' \, \frac{\rho(\vec{r'})}{|\vec{r} - \vec{r'}|} \tag{12.80}$$

$$\vec{A}(\vec{r}) = \frac{1}{c} \int d^3r' \, \frac{\vec{\mathcal{J}}(\vec{r'})}{|\vec{r} - \vec{r'}|} \tag{12.81}$$

The generalizations are

$$\phi(\vec{r}, t) = \int d^3r' \, \frac{\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)}{|\vec{r} - \vec{r'}|} \tag{12.82}$$

$$\vec{A}(\vec{r}, t) = \frac{1}{c} \int d^3r' \, \frac{\vec{\mathcal{J}}(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)}{|\vec{r} - \vec{r'}|} \tag{12.83}$$

It is a straightforward, but tedious, exercise in vector calculus to show these satisfy Maxwell's equations. I am not going to talk about this (I'll write down the derivation in an appendix for those of you who are interested), but it is worth trying to understand what these relations mean physically. The important physical point that these relations imply is that if the charge and current distributions depend on time, and if they are producing the fields, then what determines what the field is at some point $\vec{r}$ is the values of the charge and current distributions at earlier times. The farther away the charge is, the earlier the time has to be. That is what the factor of $t - |\vec{r} - \vec{r'}|/c$ is telling us. The appearance of this factor is a kind of boundary condition at infinity. It is consistent with the relativistic version of the principle of causality. Because information cannot be transferred faster than light, a charge distribution at a space-time point $(\vec{r'}, t')$ can effect the fields at the space-time point $(\vec{r}, t)$, only if $t \geq t'$ and

$$\frac{|\vec{r} - \vec{r'}|}{t - t'} \leq c \tag{12.84}$$

In these relations, (12.82) and (12.83), however, the condition is even stronger — a charge distribution at a space time point $(\vec{r'}, t')$ can effect the fields at the space time point $(\vec{r}, t)$ only if light can travel directly from $(\vec{r'}, t')$ to $(\vec{r}, t)$ — that is if $t \geq t'$ and

$$\frac{|\vec{r} - \vec{r'}|}{t - t'} = c \tag{12.85}$$

or

$$t - t' = |\vec{r} - \vec{r'}|/c \tag{12.86}$$

or

$$t' = t - |\vec{r} - \vec{r'}|/c \tag{12.87}$$

These are just words. We have not derived this! The real justification of this discussion comes when you check that the relations actually satisfy Maxwell's equations. That can wait for Physics 153 or 232 (or the appendix if you are in a hurry). However, I hope that this discussion at least makes the result reasonable. In fact you have already seen the result in action in 15b in Purcell's discussion of the electric field from a charge that starts and stops. Look at the ANIMATIONS - PURCELL - the field from a charge that suddenly accelerates. This is an animation of a famous figure in Purcell's book. The interesting thing about the animation is the kink in the electric field that propagates out from the acceleration event at the velocity of light — because it is light. Inside the kink, the fields are those of the moving charge. Outside the kind, the fields are those of the stationary charge. The kink — the electromagnetic way — is what connects the two asymptotic regions together. It is also fun to compare with PURCELL2 which illustrates what happens if an initially moving charge stops suddenly.

Now let's see at what the electric and magnetic fields look like in an important limit. The connection between the potentials and the fields is the following:

$$\vec{E} = -\vec{\nabla}\phi - \frac{1}{c}\frac{\partial}{\partial t}\vec{A} \tag{12.88}$$

$$\vec{B} = \vec{\nabla} \times \vec{A} \tag{12.89}$$

These relations are completely general. The special limit that I want to consider is one in which the charges and currents are confined to a small region around $\vec{r} = 0$. Then we will look at the electric and magnetic fields produced by the moving charges far away, for large $|\vec{r}|$. It is actually easiest to look at the magnetic field:

$$\vec{B} = \vec{\nabla} \times \vec{A} = \vec{\nabla} \times \frac{1}{c}\int d^3r'\, \frac{\vec{\mathcal{J}}(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)}{|\vec{r} - \vec{r'}|} \tag{12.90}$$

The point is that the curl ($\vec{\nabla}\times$) can operate in two different places, either on the $1/|\vec{r} - \vec{r'}|$ or on the $-|\vec{r} - \vec{r'}|/c$ in the time dependence of $\mathcal{J}$. The first gives a contribution that drops of like $1/r^2$ for large r, just like the magnetic field from a time-independent distribution of currents. But the second gives a contribution that only falls off like $1/r$. Thus this contribution dominates for large r. Explicitly (using the chain rule), it is

$$\vec{B} = -\frac{1}{c^2}\int d^3r'\, \frac{\vec{r} - \vec{r'}}{|\vec{r} - \vec{r'}|^2} \times \frac{d}{dt}\vec{\mathcal{J}}(\vec{r'}, t - |\vec{r} - \vec{r'}|/c) \tag{12.91}$$

$$\to -\frac{1}{c^2}\frac{1}{r}\int d^3r'\, \hat{r} \times \frac{d}{dt}\vec{\mathcal{J}}(\vec{r'}, t - |\vec{r} - \vec{r'}|/c) \tag{12.92}$$

where in (12.92) we have dropped a $\vec{r'}$ in the numerator because this term falls like $1/r^2$ for large r.

This is the magnetic field of an electromagnetic wave. Notice that it is perpendicular to the direction of motion ($\hat{r}$). The $1/r$ fall-off is what we expect for an electromagnetic wave, because the energy density goes like the square of the field, and falls off like $1/r^2$ as the wave spreads.

The electric field can be computed in a similar way, although you also need to use the conservation of electric charge.

$$\frac{\partial}{\partial t}\rho + \vec{\nabla}\cdot\vec{\mathcal{J}} = 0 \tag{12.93}$$

As you would expect, the result is that the electric field has the same magnitude as the magnetic field and is perpendicular to both direction of motion and to the magnetic field. The piece that corresponds to a traveling electromagnetic wave can be written as

$$\vec{E} \to -\frac{1}{c^2}\int d^3r'\,\frac{\vec{r}-\vec{r'}}{|\vec{r}-\vec{r'}|}\times\left(\frac{\vec{r}-\vec{r'}}{|\vec{r}-\vec{r'}|^2}\times\frac{d}{dt}\vec{\mathcal{J}}(\vec{r'},t-|\vec{r}-\vec{r'}|/c)\right) \tag{12.94}$$

$$\to \frac{1}{c^2}\,\frac{1}{r}\int d^3r'\,\hat{r}\times\left(\hat{r}\times\frac{d}{dt}\vec{\mathcal{J}}(\vec{r'},t-|\vec{r}-\vec{r'}|/c)\right) \tag{12.95}$$

To lowest order in $1/r$ for charges moving with velocities much smaller than c, we can simplify the electric field in (12.95) by substituting

$$|\vec{r}-\vec{r'}| \to r \tag{12.96}$$

and write the result as

$$\vec{E}(\vec{r},t) \approx \frac{1}{c^2}\,\frac{1}{r}\int d^3r'\,\hat{r}\times\left(\hat{r}\times\frac{d}{dt}\vec{\mathcal{J}}(\vec{r'},t-r/c)\right) \tag{12.97}$$

The reason for the restriction to nonrelativistic motion of charges is that if a charged particle is moving at a speed close to the velocity of light, then we cannot neglect its position, $\vec{r'}$, when it is moving towards $\vec{r}$. To see this, consider the impossible limit in which the charge is moving towards the point $\vec{r}$ at the speed of light. Then if the charge contributes to the electric field at $\vec{r}$ at one time, then it also contributes at later times because the particle keeps up with the moving light wave. While $v = c$ is impossible, for $v \approx c$, the $\vec{r'}$ dependence cannot be ignored because it leads to very rapid time dependence of the potentials, and hence to large fields. What happens is that the contribution of charges moving relativistically to the electric fields in front of them get enhanced by factors of $\frac{c}{c-v}$. This effect is widely used today to produce intense "light" from particle accelerators — so called synchrotron radiation. You can see this effect in the ANIMATIONS if you make v close to 1.

A particularly important and instructive case of (12.97) is the nonrelativistic motion of a single charge, Q, moving along a trajectory $\vec{R}(t)$. For this system,[5]

$$\vec{\mathcal{J}}(\vec{r},t) = Q\,\vec{v}(t)\,\delta^3(\vec{r}-\vec{R}(t)) = Q\frac{d\vec{R}(t)}{dt}\,\delta^3(\vec{r}-\vec{R}(t)) \tag{12.98}$$

Then the integration over d^3r' in (a) eliminates the δ-function, and the electric field of the outgoing electromagnetic wave is proportional to the acceleration,

$$\vec{E}(\vec{r},t) \approx \frac{1}{c^2}\,\frac{1}{r}\,Q\,\hat{r}\times(\hat{r}\times\vec{a}(t-r/c)) \tag{12.99}$$

where

$$\vec{a}(t) = \frac{d^2\vec{R}(t)}{dt^2} \tag{12.100}$$

All that the cross products with $\hat{r}$ do is to pick out minus the component of $\vec{a}(t-r/c)$ that is perpendicular to $\vec{r}$. It follows from the famous "bac-cab" identity,

$$\vec{a}\times\left(\vec{b}\times\vec{c}\right) = \vec{b}\left(\vec{a}\cdot\vec{c}\right) - \vec{c}\left(\vec{a}\cdot\vec{b}\right), \tag{12.101}$$

that

$$\vec{E}(\vec{r},t) \approx -\frac{1}{c^2}\,\frac{1}{r}\,Q\left(\vec{a}(t-r/c) - \hat{r}\left(\hat{r}\cdot\vec{a}(t-r/c)\right)\right). \tag{12.102}$$

This had to happen because the electric field of an electromagnetic wave is perpendicular to its direction of motion. In this case, for large r, the wave is nearly a plane wave moving in direction $\vec{r}$.

12.5.2 The Antenna Pattern

Let us do an even more explicit example by considering a charge that oscillates harmonically along the z axis,

$$\vec{R}(t) = \ell\,\hat{z}\cos\omega t\,. \tag{12.103}$$

so that

$$\vec{a}(t) = -\ell\omega^2\,\hat{z}\cos\omega t\,. \tag{12.104}$$

---

[5]This equation makes use of δ-function notation. To a physicist, a δ-function $\delta(x)$ is just a function that has area 1 and is so sharply peaked around $x = 0$ that we don't care exactly what it looks like. All that matters is the area and where the peak is. The $\delta^3\left(\vec{r}-\vec{R}(t)\right)$ in the equation is actually the product of three delta functions, for the x, y and z components, and just tells you that $\vec{r} = (x,y,z) = \vec{R}(t) = \left(X(t),Y(t),Z(t)\right)$ — that is that the particle is moving along the trajectory $\vec{R}(t)$. For a mathematical discussion of the δ-function you can look at http://mathworld.wolfram.com/DeltaFunction.html. But don't be frightened. It is just a simple device for ignoring small details that we don't care about. If you translate the integral into words or pictures, it may help.

$$\vec{E}(\vec{r},t) \approx \frac{\ell\omega^2}{c^2}\,\frac{1}{r}\,Q\left(\hat{z}-\hat{r}(\hat{r}\cdot\hat{z})\right)\cos[\omega(t-r/c)]\,. \tag{12.105}$$

The vector $\hat{z}-\hat{r}(\hat{r}\cdot\hat{z})$ is the component of $\hat{z}$ perpendicular to $\vec{r}$, as illustrated in figure 12.13.

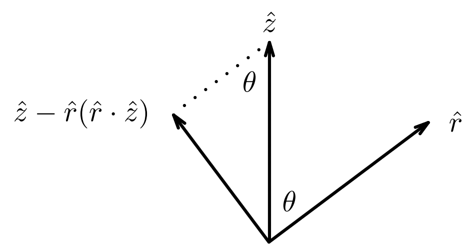


Figure 12.13:

Evidently, the magnitude of $\hat{z}-\hat{r}(\hat{r}\cdot\hat{z})$ is $\sin\theta$. This means that the intensity of the electromagnetic wave at an angle θ from the z axis is proportional to $\sin^2\theta$. The intensity pattern can be conveniently represented in polar coordinates, where we plot the intensity as a function of θ. The result is shown below. This is the "antenna pattern" for the oscillating

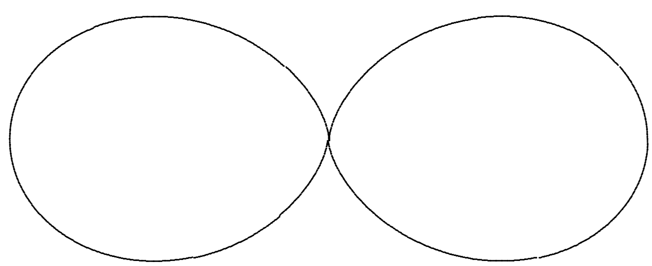


Figure 12.14:

dipole in the z direction. It is shown in figure 12.14. The two lobes of the pattern arise because the field is highest in the x-y plane, for $\theta = \pi/2$, and drops to zero as we approach the z axis, $\theta = 0$ or $\theta = \pi$.

12.5.3 * Checking Maxwell's equations

These things are called retarded potentials. This is a confusing name, since there is really nothing special about the potentials themselves. What is special is the assumption of a particular relation between the potentials and the charges and currents — that the fields are being produced entirely by the charges and currents. Here I show that they satisfy Maxwell's equations. I call this an appendix because you are NOT responsible for knowing the details. I include it for your general education.

Some mathematical things to notice about the solution:

$$\frac{\partial}{\partial t}\rho + \vec{\nabla}\cdot\vec{\mathcal{J}} = 0 \tag{12.106}$$

implies

$$\frac{\partial}{\partial t}\phi + c\vec{\nabla}\cdot\vec{A} = 0 \tag{12.107}$$

This is called the Lorentz gauge condition.

$$\vec{\nabla}\cdot\vec{E} = -\nabla^2\phi - \frac{1}{c}\frac{\partial}{\partial t}\vec{\nabla}\cdot\vec{A} \tag{12.108}$$

$$= \left(-\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi \tag{12.109}$$

$$= \int d^3r' \left(\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)\left(-\nabla^2\right)\frac{1}{|\vec{r} - \vec{r'}|}\right. \tag{12.110}$$

$$+ \frac{1}{|\vec{r} - \vec{r'}|}\left(-\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c) \tag{12.111}$$

$$\left. - 2\left(\vec{\nabla}\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)\right)\cdot\left(\vec{\nabla}\frac{1}{|\vec{r} - \vec{r'}|}\right)\right) \tag{12.112}$$

The first term is the one we want. It is

$$= \int d^3r'\, \rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)\, 4\pi\, \delta^3\left(\vec{r} - \vec{r'}\right) \tag{12.113}$$

$$= 4\pi\, \rho(\vec{r}, t) \tag{12.114}$$

The other two terms cancel because of the special form of the variable $t - |\vec{r} - \vec{r'}|/c$.

$$\frac{1}{|\vec{r} - \vec{r'}|}\left(-\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c) \tag{12.115}$$

$$- 2\left(\vec{\nabla}\rho(\vec{r'}, t - |\vec{r} - \vec{r'}|/c)\right)\cdot\left(\vec{\nabla}\frac{1}{|\vec{r} - \vec{r'}|}\right) \tag{12.116}$$

$$= \frac{1}{|\vec{r} - \vec{r'}|}\,\frac{1}{c^2}\ddot{\rho}(\vec{r'}, t - |\vec{r} - \vec{r'}|/c) \tag{12.117}$$

$$+\frac{1}{|\vec{r}-\vec{r'}|}\frac{1}{c}\vec{\nabla}\cdot\frac{\vec{r}-\vec{r'}}{|\vec{r}-\vec{r'}|}\dot{\rho}(\vec{r'},t-|\vec{r}-\vec{r'}|/c) \tag{12.118}$$

$$-2\frac{1}{c}\frac{1}{|\vec{r}-\vec{r'}|}\dot{\rho}(\vec{r'},t-|\vec{r}-\vec{r'}|/c) \tag{12.119}$$

where $\dot{}$ means differentiation with respect to the time variable:

$$\dot{\rho}(\vec{r'},t-|\vec{r}-\vec{r'}|/c) \equiv \frac{\partial}{\partial t'}\rho(\vec{r'},t')\Big|_{t'=t-|\vec{r}-\vec{r'}|/c} \tag{12.120}$$

i and ii come from a (from the $\frac{\partial^2}{\partial t^2}$ and $-\nabla^2$ terms respectively) and iii comes from b. Now the $\vec{\nabla}$ in ii gives two terms — acting on $\dot{\rho}$ cancels i and acting on $\frac{\vec{r}-\vec{r'}}{|\vec{r}-\vec{r'}|}$ cancels iii. Thus

$$\vec{\nabla}\cdot\vec{E} = 4\pi\rho\,. \tag{12.121}$$

Likewise,

$$\vec{\nabla}\times\vec{B}-\frac{1}{c}\frac{\partial}{\partial t}\vec{E} \tag{12.122}$$

$$=\vec{\nabla}\times\left(\vec{\nabla}\times\vec{A}\right)-\frac{1}{c}\frac{\partial}{\partial t}\left(-\vec{\nabla}\phi-\frac{1}{c}\frac{\partial}{\partial t}\vec{A}\right) \tag{12.123}$$

$$=\vec{\nabla}\left(\vec{\nabla}\cdot\vec{A}\right)-\nabla^2\vec{A}-\vec{\nabla}\left(\frac{1}{c}\frac{\partial}{\partial t}\phi\right)+\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\vec{A} \tag{12.124}$$

or using the Lorentz gauge condition,

$$=\left(-\nabla^2+\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{A} \tag{12.125}$$

Form here on, the derivation is the same as for $\vec{\nabla}\cdot\vec{E}$, and we find

$$\vec{\nabla}\times\vec{B}-\frac{1}{c}\frac{\partial}{\partial t}\vec{E}=\frac{4\pi}{c}\vec{\mathcal{J}} \tag{12.126}$$

QED.

Chapter Checklist

You should now be able to:

i. Describe polarization on a beaded or continuous string;

ii. Write down the general form of an electromagnetic plane wave and relate it to the two-dimensional vector, Z;

iii. Find the energy and momentum density of a plane electromagnetic wave;

iv. Understand the possible polarization states of a plane wave;

v. Analyze systems of polarizers and wave plates using matrix multiplication;

vi. Understand the connection between optical activity and handedness;

vii. Calculate the reflection and transmission of a plane electromagnetic wave from a plane boundary between dielectric for any angle and find and explain Brewster's angle.

Problems

12.1* . A pane of glass with index of refraction $n = 2$ sits in the x-y plane, from $z = 0$ to $z = \ell$. A plane wave with wave number k (outside the glass) comes at the pane at an angle θ from the perpendicular in the x-y plane, with $k_z = k\cos\theta$ and $k_x = k\sin\theta$.

For each of the two polarization states (in the y direction, and in the x-z plane), some fraction of the intensity is reflected as a function of θ and k. In this problem, we will use the method of transfer matrices, discussed in Chapter 9 to find it. We will work out the case of polarization perpendicular to the x-z scattering plane in detail. Then your job will be to repeat the calculation for polarization in the x-z plane. To do it, we must generalize the analysis of (12.62)-(12.63) and (12.70)-(12.71) to a situation with arbitrary incoming and outgoing waves on both sides and to a boundary at arbitrary z (rather than y for this problem). For the perpendicular polarization state, the boundary conditions look like:

$$e^{ik_z z}T_\perp^1 + e^{-ik_z z}R_\perp^1 = e^{ik'_z z}T_\perp^2 + e^{-ik'_z z}R_\perp^2$$

$$n\cos\theta\left(e^{ik_z z}T_\perp^1 - e^{-ik_z z}R_\perp^1\right) = n'\cos\theta'\left(e^{ik'_z z}T_\perp^2 + e^{-ik'_z z}R_\perp^2\right)$$

which gives

$$\begin{pmatrix} T_\perp^1 \\ R_\perp^1 \end{pmatrix} = d(z)\begin{pmatrix} T_\perp^2 \\ R_\perp^2 \end{pmatrix}$$

where the transfer matrix, $d(z)$ is

$$\frac{1}{2}\begin{pmatrix} e^{-ik_z z} & 0 \\ 0 & e^{ik_z z} \end{pmatrix}\begin{pmatrix} 1+h_\perp & 1-h_\perp \\ 1-h_\perp & 1+h_\perp \end{pmatrix}\begin{pmatrix} e^{ik'_z z} & 0 \\ 0 & e^{-ik'_z z} \end{pmatrix}$$

with

$$h_\perp = \frac{n\cos\theta}{n'\cos\theta'}\,.$$

Going from index n' to index n at z gives a transfer matrix that is the inverse of $d(z)$. Applying this to the present problem, if $R_\perp$ and $\tau_\perp$ are the reflection and transmission coefficients from the pane of glass, we have

$$\begin{pmatrix} 1 \\ R_\perp \end{pmatrix} = d(0)\, d(\ell)^{-1} \begin{pmatrix} 0 \\ \tau_\perp \end{pmatrix}$$

which implies

$$\tau_\perp = \frac{2h_\perp e^{ik_z\ell}}{2h_\perp \cos k'_z\ell - i(1+h_\perp^2)\sin k'_z\ell}$$

$$R_\perp = \frac{-i(1-h_\perp^2)\sin k'_z\ell}{2h_\perp \cos k'_z\ell - i(1+h_\perp^2)\sin k'_z\ell}$$

The fraction of the reflected intensity is

$$|R_\perp|^2 = \frac{(1-h_\perp^2)^2 \sin^2 k'_z\ell}{4h_\perp^2 \cos^2 k'_z\ell + (1+h_\perp^2)^2 \sin^2 k'_z\ell}$$

Now, do the same analysis for the polarization in the x-z plane. Find $\left|R_\parallel\right|^2$. What happens at Brewster's angle?

12.2. Consider a boundary at $x = 0$ between two regions of empty space. On the boundary surface at $x = 0$, there is a thin layer of stuff with surface conductivity σ. That means that an electric field, $\vec{E}$, with a component parallel to the surface (in the y-z plane) produces a surface current density in the boundary layer:

$$\vec{\mathcal{J}}(y,z) = (0, \sigma\, E_y(0,y,z), \sigma\, E_z(0,y,z))\,.$$

In this system, there is an electric field of the form shown below:

$$E_z(x,y,t) = Ae^{i(kx\cos\theta + ky\sin\theta - \omega t)} + R\,Ae^{i(-k'x\cos\theta' + k'y\sin\theta' - \omega t)}$$

for $x < 0$, and

$$E_z(x,y,t) = T\,Ae^{i(k''x\cos\theta'' + k''y\sin\theta'' - \omega t)}$$

for $x > 0$. E_x and E_y vanish everywhere.

Find k', k'', θ' and θ''. Find T in terms of R. Find the current density on the boundary, $\vec{\mathcal{J}}(y,z)$. Find the magnetic field everywhere. Find R.

Check your result for R by explaining the limit $\sigma \to \infty$, a superconducting surface. What happens to R in this limit and why?

Hint: Use Maxwell's equations to find $\vec{B}$ and then look at the discontinuity of the magnetic field across the surface current.

12.3. Suppose that on the planes $z = 0$ and $z = a$ for $x \geq 0$, there are two flat semi-infinite conducting planes. Suppose, further, that the oscillation of the system is forced by some device that produces an electric field in the $x = 0$ plane for $0 \leq z \leq a$ with the following properties: $\vec{E}$ points in the y direction but its y-component is independent of y and equal to $E_0 \sin(3\pi z/a)\, \cos(\omega t)$, where $\omega > 3\pi c/a$ and c is the speed of light in vacuum. If this produces a traveling wave in the $+x$ direction, find the form of the electric field everywhere between the plates. If this traveling wave is used as a carrier wave for amplitude modulated signals, with what speed does the signal travel?

12.4. Consider the standing electromagnetic waves in a cubical evacuated box with **perfectly conducting** sides at $x = 0$, $x = L$, $y = 0$, $y = L$, $z = 0$ and $z = L$. There exist modes in which the electric and magnetic fields vanish outside the box, and inside take the following form:

$$E_z(x, y, z, t) = A\,\omega\, \sin k_x x \sin k_y y \cos \omega t$$

$$B_x(x, y, z, t) = -A\, k_y \sin k_x x \cos k_y y \sin \omega t$$

$$B_y(x, y, z, t) = A\, k_x \cos k_x x \sin k_y y \sin \omega t$$

$$E_x = E_y = B_z = 0\,.$$

You can check that inside the box and for properly chosen ω, these satisfy Maxwell's equations,

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

$$\vec{\nabla} \times \vec{B} = \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} + \mu_0 \vec{J}\,,$$

$$\vec{\nabla} \cdot \vec{E} = \rho\,, \quad \vec{\nabla} \cdot \vec{B} = 0\,.$$

Find ω as a function of k_x and k_y.

There are no charges or currents inside the box, but there will be charges and currents built up on the boundary to confine the electric and magnetic fields inside the box. For example, a nonzero surface charge density appears on the top ($z = L$) and bottom ($z = 0$). The charges oscillate back and forth from top to bottom while nonzero surface current densities appear on all sides. The form above is constructed to satisfy appropriate boundary conditions on the four sides $x = 0$, $y = 0$, $z = 0$ and $z = L$.

Explain the physics of the boundary conditions for the $\vec{E}$ field on the sides $x = L$ and $y = L$ and find the allowed values of k_x and k_y. Then explain the physics of the boundary conditions for the $\vec{E}$ field on the sides $x = L$ and $y = L$ adn draw a diagram to explain what is going on for the lowest possible values of k_x and k_y. **Hint:** Remember that the magnetic field vanishes outside the box.

12.5. A plane wave of light traveling in the $+z$ direction is polarized at an angle θ from the x axis in the $x - y$ plane. When it encounters a sheet of polaroid in the $z = L$ plane that transmits only light polarized and an angle $\theta + \frac{\pi}{2}$, the wave is completely absorbed. However, if the plane wave first passes through a sheet of cellophane in the $z = 0$ plane with the "fast axis" along x axis, some of the light gets through. Suppose that the cellophane introduces a phase difference of ϕ between the component of the light wave polarized along the fast (x) axis and the component polarized along the slow (y) axis. Find the ratio of the intensity of the transmitted wave beyond the polaroid to the incoming wave intensity as a function of θ and ϕ. **Hint:** Does your answer go to zero as $\phi \to 0$? What happens as $\theta \to 0$?

12.6. A plane wave of light traveling in the $+z$ direction is polarized in the x direction. When it encounters a sheet of polaroid in the $z = L$ plane that transmits only y polarized light, the wave is completely absorbed. However, if the plane wave first passes through a sheet of cellophane in the $z = 0$ plane with the "fast axis" at an angle θ with the x axis, some of the light get through. Suppose that the cellophane introduces a phase difference of ϕ between a wave polarized along the fast axis and one polarized along the slow axis. Find the ratio of the intensity of the transmitted wave beyond the polaroid to the incoming wave intensity as a function of θ and ϕ.

Compare the result with the previous problem and explain what is going on.

12.7. Suppose that a charge Q is stationary at the origin until $t = 0$. From time $t = 0$ to $t = \Delta t$, the charge experiences uniform acceleration $a\,\hat{x}$.

a. Use (12.102) to find an approximate expression for the electric field at a large distance $r \gg a\Delta t^2$ from the origin.

b. How does this compare with what you see in the animation PURCELL?

Chapter 13

Interference and Diffraction

A "beam" of light is very familiar. A laser pointer, for example, produces a pattern of light that is almost like a transverse section of a plane wave. But not quite. The laser beam spreads as it travels. You might think that this is simply due to the imperfections in the laser. But, in fact, no matter how hard you try to perfect your laser, you cannot avoid some spreading. The problem is **"diffraction."**

Interference is a crucial part of the physics of diffraction. We have seen it already in one-dimensional situations such as interferometers and reflection from thin films. Here we begin to see what amazing things it does in more than one dimension.

Preview

In this chapter, we show how the phenomena of interference and diffraction arise from the physics of the forced oscillation problem and the mathematics of Fourier transformation.

i. We begin by discussing interference from a double slit. This is the classic example of interference. We give a heuristic discussion of the physics, and generalize it to get the fundamental result of Fourier optics.

ii. We then continue our quantitative analysis of interference and diffraction by discussing the general problem again as a forced oscillation problem. We show the connection with making a beam. We find the relevant boundary condition at infinity and express the solution in the form of an integral.

iii. We show how the integral simplifies in two extreme regions — very close to the source of the beam, where it really looks like a beam — and very far away, where diffraction takes over and the intensity of the wave is related to a Fourier transform of the wave pattern at the source, the same result that we found in our heuristic discussion of interference.

iv. We apply these techniques to examples involving beams made with one or more slits and rectangular regions.

v. We prove a useful result, the convolution theorem, for combining Fourier transforms.

vi. We show how periodic patterns lead to sharp diffraction patterns, and discuss the example of the diffraction grating in detail.

vii. We apply the same ideas to the three-dimensional example of x-ray diffraction from crystals.

viii. We describe a hologram as a rather complicated diffraction pattern.

ix. We discuss interference fringes and zone plates.

13.1 Interference

13.1.1 The Double Slit

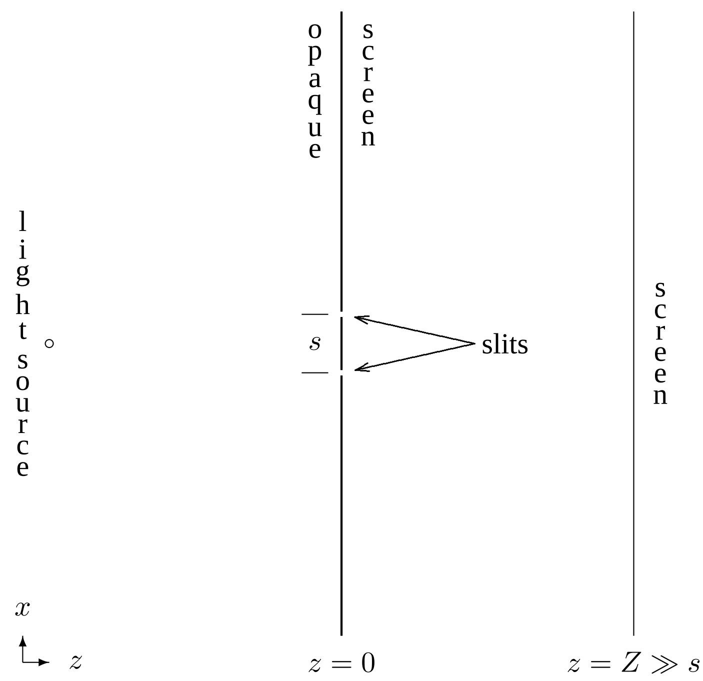


Figure 13.1: The double slit experiment.

The classic arrangement of the double slit experiment is illustrated in figure 13.1. There is an opaque screen with two narrow slits in it in the $z = 0$ plane (shown in cross section in the x-z plane — the slits come out of the paper in the y direction) a small distance s apart. The opaque screen is illuminated by a "point" source of light. For example, this could be a light with a clear glass bulb and a colored filter to pick out a narrow frequency range, far away in the $-z$ direction. A laser beam spread out with a lens would serve just as well. The important thing is to produce illumination at the opaque screen in which the frequency is in a narrow range and the phase of the light reaching the two slits is correlated. This will certainly be true if the illumination for $z < 0$ is nearly a plane wave.

Now an interesting thing happens at the second screen, at $z = Z$. This "screen" could be a photographic plate, a translucent screen, or even your retina. What appears on this screen is a series of parallel lines of brightness in the y direction (parallel to the slits). If one of the slits is covered up, the lines disappear.

What is going on is interference between the two possible straight-line paths by which the light can reach the screen. We will give a heuristic, physical discussion of the interference in this section. Then in the next section, we will derive the same result using the kind of forced oscillation and boundary condition arguments that you know from our study of one-dimensional waves.

The physical picture is this. The electric field at $z = Z$ is a sum of the fields that come from the two slits. At $x = 0$, in the symmetrical arrangement shown in figure 13.1, the two possible paths for the light have the same length. Therefore, the two components of the field have the same phase. Therefore they interfere "constructively" and there is a bright line at $x = 0$. As x changes, at $z = Z$, the relative length of the two paths changes. We will then get alternating positions of constructive and destructive interference. This gives rise to the bright lines.

We can understand the effect quantitatively by computing the path length explicitly. Consider a point on the screen at $x = X$. This is shown in figure 13.2.

The length of the dotted line in figure 13.2 is

$$\sqrt{X^2 + Z^2}\,. \tag{13.1}$$

For the upper and lower slits, the path lengths are slightly shorter and longer respectively. The total difference in path length is

$$\Delta\ell = \sqrt{(X + s/2)^2 + Z^2} - \sqrt{(X - s/2)^2 + Z^2}\,. \tag{13.2}$$

For $Z \gg s$, we can expand $\Delta\ell$ in (13.2) in a Taylor series,

$$\Delta\ell \approx \frac{sX}{\sqrt{X^2 + Z^2}}\,. \tag{13.3}$$

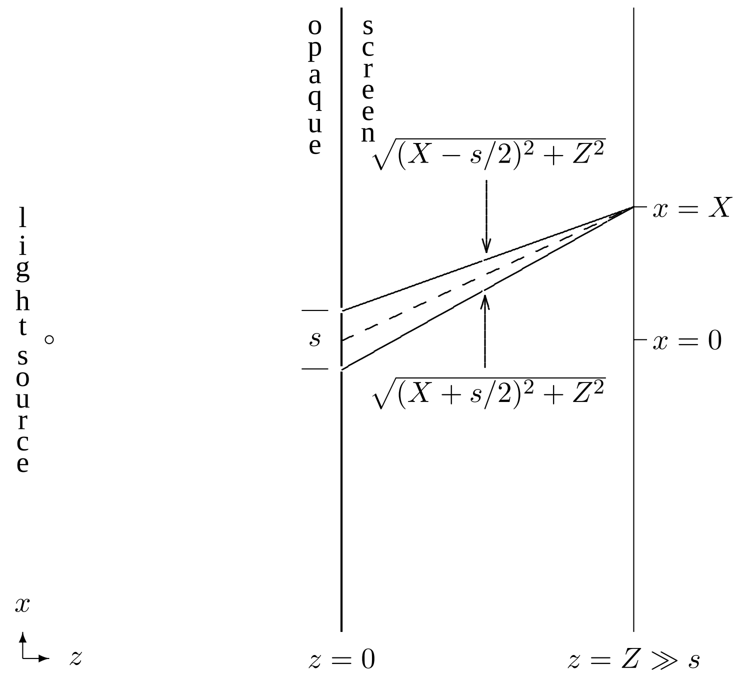


Figure 13.2: Path lengths.

Therefore if the angular wave number of the light is k, the **phase difference** between the two paths is

$$\frac{ksX}{\sqrt{X^2+Z^2}} . \tag{13.4}$$

We get an intensity maximum every time the phase is a multiple of 2π, when

$$\frac{ksX}{\sqrt{X^2+Z^2}} = 2n\pi . \tag{13.5}$$

In terms of the wavelength, $\lambda = 2\pi/k$, this is

$$\frac{X}{\sqrt{X^2+Z^2}} = n\frac{\lambda}{s} . \tag{13.6}$$

13.1.2 Fourier Optics

Suppose that instead of a simple pattern of two slits, there is some more complicated pattern on the opaque screen. In general, we can describe the wave disturbance in the $z = 0$ plane by some function of x and y,[1]

$$f(x,y)\,. \tag{13.7}$$

Our strategy will be to think of the wave produced for $z > 0$ by this general function as a sum of the effects of tiny holes at all the values of x and y for which $f(x,y)$ is nonzero. For each little piece of the function, we can compute the path length to some point on the screen at $z = Z$. Then we can add up all the pieces.

Suppose, for simplicity, that $f(x,y)$ is only nonzero in some small region around the origin, so that x and y will be small

$$x, y \ll Z\,, \tag{13.8}$$

for all relevant values of x and y. Now the path length from the point $(x,y,0)$ on the screen at $z = 0$ to the point (X,Y,Z) on the screen at $z = Z$ is

$$\sqrt{(X-x)^2+(Y-y)^2+Z^2}\,. \tag{13.9}$$

Using (13.8), we can expand this as follows:

$$R + \Delta\ell(x,y) + \cdots\,, \tag{13.10}$$

where

$$R = \sqrt{X^2+Y^2+Z^2} \tag{13.11}$$

and

$$\Delta\ell(x,y) = -\frac{xX+yY}{R}\,. \tag{13.12}$$

Thus the wave on the path from $(x,y,0)$ to (X,Y,Z) gets a phase of approximately

$$e^{ik(R+\Delta\ell)}\,. \tag{13.13}$$

Now we can put the pieces of the wave back together to see how the interference works at the point (X,Y,Z). We just sum over all values of x and y, with a factor of the phase and the function, $f(x,y)$. Because x and y are continuous variables, the sum is actually an integral,

$$\int dx \int dy\, f(x,y)\, e^{ik(R+\Delta\ell)} = e^{ikR} \int dx \int dy\, f(x,y)\, e^{-i(xX+yY)k/R}\,. \tag{13.14}$$

As we will see in more detail below, this is a two-dimensional Fourier transform of the function, $f(x,y)$.

The equation, (13.14), is the fundamental result of Fourier optics. It contains much of the physics of diffraction. We have made a number of assumptions in deriving it that need further discussion. In the next section, we will derive it in a different way, treating the wave for $z > 0$ as the result of a forced oscillation, produced by the wave in the $z = 0$ plane. This will give us an alternative physical description of diffraction. But it will be useful to keep the simple picture of adding up all the possible paths in mind as we get deeper into the phenomena of interference and diffraction.

[1]We are ignoring polarization.

13.2 Beams

13.2.1 Making a Beam

Consider a system with an opaque barrier in the $z = 0$ plane. If it is illuminated by a plane wave traveling in the $+z$ direction, the barrier absorbs the wave completely. Now cut a hole in the barrier. You might think that this would produce a beam of light traveling in the direction of the initial plane wave. But it is not that simple. This is actually the same problem that we considered in the previous section, (13.7)-(13.14), with the function, $f(x, y)$, given by

$$f(x,y)\,e^{-i\omega t} \tag{13.15}$$

where

$$f(x,y) = \begin{cases} 1 & \text{inside the opening} \\ 0 & \text{outside the opening.} \end{cases} \tag{13.16}$$

In fact, it will be useful to think about the more general problem, because the the function, (13.16), is discontinuous. As we will see later, this leads to more complicated diffraction phenomena than we see with a smooth function. In particular, we will assume that $f(x, y)$ is signifigantly different from zero only for small x and y and goes to zero for large x and y. Then we can talk about the position of the "opening" that produces the beam, near $x = y = 0$.

We can think of this problem as a forced oscillation problem. It is much easier to analyze the physics if we ignore polarization, so we will discuss scalar waves. For example, we could consider the transverse waves on a flexible membrane or pressure waves in a gas. Equivalently, we could consider light waves that depend only on two dimensions, x and z, and polarized in the y direction. We will not worry about these niceties too much, because as usual, the basic properties of the wave phenomena will be determined by translation invariance properties that are independent of what it is that is waving!

13.2.2 Caveats

It is worth noting that there are other approaches to the diffraction problem besides the ones we discuss here. The physical setup we are considering is slightly different from the standard setup of Huygens-Fresnel-Kirchhoff diffraction, because we are studying a different problem. In Huygens-Fresnel-Kirchhoff diffraction,[2] you consider the diffraction of a plane wave from a finite object, whereas, our opaque screen is infinite in the x-y plane. In the Huygens-Fresnel case, the appropriate boundary condition is that there are no incoming **spherical waves** coming back in from infinity toward the object that is doing the diffracting. The diffraction produces outgoing spherical waves only. We will not discuss this alternative physical setup

[2] For example, see Hecht, chapter 10.

in detail because it leads deeper into Bessel functions[3] than we (and probably the reader as well) are eager to go. The advantage of our formulation is that we can set it up entirely with the plane wave solutions that we have already discussed. We will simply indicate the differences between our treatment and Huygens-Fresnel diffraction. For diffraction in the forward region, at large z and not very far from the z axis, the diffraction is the same in the two cases.

The reader should also notice that we have not explained exactly how the oscillation, (13.15),

$$f(x,y)\,e^{-i\omega t} \tag{13.15}$$

in the $z=0$ plane is produced. This is by no means a trivial problem, but we will not discuss it in detail. We are concentrating on the physics for $z>0$. This will be quite interesting enough.

13.2.3 The Boundary at ∞

To determine the form of the waves in the region $z>0$ (beyond the barrier), we need boundary conditions both at $z=0$ and at $z=\infty$. At $z=0$, there is an oscillating amplitude given by (13.15).[4] At $z=\infty$, we must impose the condition that there are no waves traveling in the $-z$ direction (back toward the barrier) and that the solutions are well behaved at ∞.

The normal modes have the form

$$e^{i\vec{k}\cdot\vec{r}-i\omega t} \tag{13.17}$$

where $\vec{k}$ satisfies the dispersion relation

$$\omega^2 = v^2\,\vec{k}^2\,. \tag{13.18}$$

Thus given two components of $\vec{k}$, we can find the third using (13.18). So we can write the solution as

$$\psi(\vec{r},t) = \int dk_x dk_y\, C(k_x,k_y)\, e^{i\vec{k}\cdot\vec{r}-i\omega t} \text{ for } z>0 \tag{13.19}$$

where

$$k_z = \sqrt{\omega^2/v^2 - k_x^2 - k_y^2}\,. \tag{13.20}$$

Note that (13.20) does not determine the sign of k_z. But the boundary condition at ∞ does. If k_z is real, it must be positive in order to describe a wave traveling to the right, away from the barrier. If k_z is complex, its imaginary part must be positive, otherwise $e^{i\vec{k}\cdot\vec{r}}$ would blow up as z goes to ∞. Thus,

$$\text{if } \operatorname{Im} k_z = 0, \text{ then } \operatorname{Re} k_z > 0; \text{ otherwise } \operatorname{Im} k_z > 0. \tag{13.21}$$

---

[3]See the discussion starting on page 314.

[4]Note that in a real physical situation, the boundary conditions are often much more complicated than (13.16), because the physics of the boundary matters. However, this often means that diffraction in a real situation is even larger.

We discussed the physical signifigance of the boundary condition, (13.21), in our discussion of tunneling starting on page 274. There is real physics in the boundary condition at infinity. For example, consider the relation between this analysis and the discussion of path lengths in the previous section. In the language of the last chapter, we cannot describe the effects of the waves with imaginary k_z. However, the boundary condition, (13.21), ensures that these components of the wave will go to zero rapidly for large z.

13.2.4 The Boundary at z=0

All we need to do to determine the form of the wave for $z > 0$ is to find $C(k_x, k_y)$. To do that, we implement the boundary condition at $z = 0$ by using (13.19)

$$\psi(\vec{r}, t) = \int dk_x dk_y \, C(k_x, k_y) \, e^{i\vec{k}\cdot\vec{r} - i\omega t} \text{ for } z > 0 \tag{13.19}$$

and setting

$$\psi(\vec{r}, t)|_{z=0} = f(x, y) \, e^{-i\omega t} \tag{13.22}$$

to get (13.15). Taking out the common factor of $e^{-i\omega t}$, this condition is

$$f(x, y) = \int dk_x dk_y \, C(k_x, k_y) \, e^{i(k_x x + k_y y)} \, . \tag{13.23}$$

If $f(x, y)$ is well behaved at infinity (as it certainly is if, as we have assumed, it goes to zero for large x and y), then only real k_x and k_y can contribute in (13.23). A complex k_x would produce a contribution that blows up either for $x \to +\infty$ or $x \to -\infty$. Thus the integrals in (13.23) run over real k from $-\infty$ to ∞.

(13.23) is just a two-dimensional Fourier transform. Using arguments analogous to those we used in our discussion of signals, we can invert it to find C.

$$C(k_x, k_y) = \frac{1}{4\pi^2} \int dx dy \, f(x, y) e^{-i(k_x x + k_y y)} \, . \tag{13.24}$$

Inserting (13.24) into (13.19) with (13.20) and (13.21)

$$k_z = \sqrt{\omega^2/v^2 - k_x^2 - k_y^2} \tag{13.20}$$

$$\text{if } \operatorname{Im} k_z = 0, \text{ then } \operatorname{Re} k_z > 0; \text{ otherwise } \operatorname{Im} k_z > 0 \tag{13.21}$$

gives the result for the wave, $\psi(\vec{r}, t)$, for $z > 0$. This result is really very general. It holds for any reasonable $f(x, y)$.

13.3 Small and Large z

But what do we do with it? The integral in (13.19) is too complicated to do analytically. Below, we will give some examples of how it works by doing the integral numerically. However, for small z and for large z, the integral simplifies in different ways.

13.3.1 Small z

For sufficiently small z, we would expect on physical grounds that we really have produced a beam and projected an image of the function, $f(x,y)$. To see this explicitly, we will use the fact that for a particular (well behaved) $f(x,y)$, the Fourier transform $C(k_x,k_y)$ is a function that goes to zero for

$$k \equiv \sqrt{k_x^2+k_y^2} \gg 1/L \tag{13.25}$$

for some L much larger than the wavelength. The distance L is determined by the smoothness of $f(x,y)$. Typically, L is the size of the smallest important feature in $f(x,y)$, the smallest distance over which $f(x,y)$ changes appreciably. We saw this in our discussion of Fourier transforms in connection with signals in Chapter 10. We will see more examples below. We can expand $k_z z$ in the exponential in a Taylor expansion,

$$\begin{aligned} k_z\, z &= z\,\sqrt{\omega^2/v^2-k_x^2-k_y^2} \\ &= \frac{z\,\omega}{v}\sqrt{1-\frac{v^2\,(k_x^2+k_y^2)}{\omega^2}} \\ &\approx \frac{z\,\omega}{v}-\frac{zv\,(k_x^2+k_y^2)}{2\omega}\,. \end{aligned} \tag{13.26}$$

Because of (13.25), the largest value of $\sqrt{k_x^2+k_y^2}$ that we need in the integral, (13.19)

$$\psi(\vec{r},t) = \int dk_x dk_y\, C(k_x,k_y)\, e^{i\vec{k}\cdot\vec{r}-i\omega t} \text{ for } z>0 \tag{13.19}$$

is of order $1/L$. For much larger values, the integrand is zero. Thus the largest possible value of the second term in the expansion (13.26) that matters in the integral, (13.19) is of the order of

$$\frac{zv}{2\omega L^2}\,. \tag{13.27}$$

Therefore, if L is finite and z is small ($\ll \omega L^2/v$), the second term is small and we can keep only the first term, $z\omega/v$. Then putting this back into the integral, (13.19), we have

$$\begin{aligned} \psi(\vec{r},t) &= \int dk_x dk_y\, C(k_x,k_y)\, e^{i\vec{k}\cdot\vec{r}-i\omega t} \\ &\approx \int dk_x dk_y\, C(k_x,k_y)\, e^{i(k_x x+k_y y+z\,\omega/v-\omega t)} \\ &\approx \int dk_x dk_y\, C(k_x,k_y)\, e^{i(k_x x+k_y y)}\, e^{i(z\,\omega/v-\omega t)} \approx f(x,y)\, e^{i\omega(z-vt)/v}\,. \end{aligned} \tag{13.28}$$

This is just what we expect — a beam with the shape of the original function, propagating in the z direction with velocity v.

The result (13.28) begins to break down when the next term in the Taylor series, (13.26), becomes important. That is when

$$\frac{z\,v\,(k_x^2+k_y^2)}{\omega}\approx 1\,. \tag{13.29}$$

Thus

$$z\approx\frac{\omega\,L^2}{v}=\frac{2\pi\,L^2}{\lambda} \tag{13.30}$$

marks the transition from a simple beam to the onset of important diffraction effects.

If $L=0$, which is the situtation in the example of a single slit of width $2a$, that we will analyze in detail later, important diffraction effects start immediately because the slit has sharp edges. However, the beam maintains some semblance of its original size until $z\approx a^2/\lambda$.

For z larger than $\omega L^2/v$, the k_x and k_y dependence from the $e^{ik_z z}$ factor cannot be ignored. In general, the evaluation of the integral, (13.19), is very hard. However, for very large z, $z\gg L$, we can use a physical argument to find the result of the integral, (13.19).

13.3.2 Large z

Suppose that you are very far away, at a point $\vec{R}=(X,Y,Z)$,

$$(x,y,z)=(X,Y,Z)\text{ for }Z\gg\omega L^2/v\,. \tag{13.31}$$

Then you cannot see the details of the shape of the opening or other details of $f(x,y)$, only its position. The wave you detect at some far-away point must have come from the opening and if you are far enough away, it is almost a plane wave. This is called "Fraunhofer" or "far-field" diffraction. If this condition is not satisfied, the problem is called "Fresnel" or "near-field" diffraction. For the light to actually reach your eye in the far-field situation, the propagation vector must point from the opening to you. The situation is depicted in the diagram in figure 13.3. In the near-field region, the spreading due to diffraction is of the same order as the size of the opening. For much larger Z, in the far-field region, the $\vec{k}$ vector must point back to the opening.

Thus the only contribution to the integral, (13.19),

$$\psi(\vec{r},t)=\int dk_x dk_y\,C(k_x,k_y)\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\text{ for }z>0 \tag{13.19}$$

that counts is that proportional to $e^{i\vec{k}\cdot\vec{R}}$ where $\vec{k}$ points from the opening to your eye. **Because the integrand in (13.19) has a factor of $C(k_x,k_y)$, the amplitude of the wave is proportional to $C(k_x,k_y)$ where**

$$(k_x,k_y,k_z)=\left(k_x,k_y,\sqrt{\omega^2/v^2-k^2}\right)\propto(X,Y,Z)\,. \tag{13.32}$$

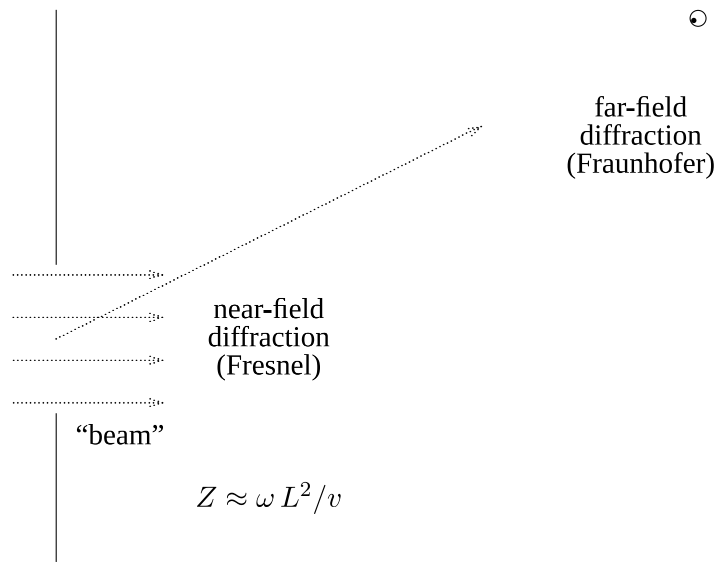


Figure 13.3: The basic diffraction problem — making a beam.

The amplitude is also inversely proportional to

$$R = \sqrt{X^2 + Y^2 + Z^2}\,, \tag{13.33}$$

because the intensity must fall off as R^{-2}, as in a spherical wave, by energy conservation.

There are other factors that contribute to the variation of the amplitude besides $C(k_x, k_y)$ (we will see one below). However, typically, all the other factors are very slowly varying and can be ignored. Thus we expect that the intensity for large Z is approximately

$$\frac{|C(k_x, k_y)|^2}{R^2}\,, \tag{13.34}$$

where $\vec{k}$ and $\vec{R}$ are related by (13.32).

$$(k_x, k_y, k_z) = \left(k_x, k_y, \sqrt{\omega^2/v^2 - k^2}\right) \propto (X, Y, Z) \tag{13.32}$$

which implies

$$\frac{k_x}{X} = \frac{k_y}{Y} = \frac{k_z}{Z} = \frac{k}{R} = \frac{\omega/v}{R}\,, \tag{13.35}$$

or

$$k_x = \frac{kX}{R}\,, \quad k_y = \frac{kY}{R}\,. \tag{13.36}$$

Now here is the point! Inserting (13.36) into (13.24)

$$C(k_x,k_y)=\frac{1}{4\pi^2}\int dxdy\,f(x,y)e^{-i(k_xx+k_yy)} \tag{13.24}$$

gives the integral in (13.14) that came from our physical argument about interference!

$$\int dx\int dy\,f(x,y)\,e^{ik(R+\Delta\ell)}=e^{ikR}\int dx\int dy\,f(x,y)\,e^{-i(xX+yY)k/R} \tag{13.14}$$

Thus our description of the wave for $z>0$ as a forced oscillation problem contains the same factor that describes the interference of all the paths that the wave can take from the opening to $\vec{R}$. The advantage of our present approach is that it is a real derivation.

We can also write this result in terms of angles:

$$\sin\theta_x=\frac{X}{R}=\frac{k_x\,v}{\omega}\,,\quad \sin\theta_y=\frac{Y}{R}=\frac{k_y\,v}{\omega}\,. \tag{13.37}$$

where θ_x and θ_y are the angles of the vector $\vec{r}$ from the $X=y=0$ line in the x and y directions. Or equivalently,

$$X=\frac{Z\,k_x}{\sqrt{\omega^2/v^2-k_x^2-k_y^2}}\,,\quad y=\frac{Z\,k_y}{\sqrt{\omega^2/v^2-k_x^2-k_y^2}}\,. \tag{13.38}$$

This is illustrated in the diagram in figure 13.4.

13.3.3 * Stationary Phase

Mathematically, (13.32)

$$(k_x,k_y,k_z)=\left(k_x,k_y,\sqrt{\omega^2/v^2-k^2}\right)\propto(X,Y,Z) \tag{13.32}$$

arises for large Z because the phase of the exponential in (13.19)

$$\psi(\vec{r},t)=\int dk_xdk_y\,C(k_x,k_y)\,e^{i\vec{k}\cdot\vec{r}-i\omega t}\text{ for }z>0 \tag{13.19}$$

is very rapidly varying as a function of k_x and k_y **except for special values of k_x and k_y where the derivatives of the phase with respect to k_x and k_y vanish.** If the function is centered at $x=y=0$ and is smooth,[5] the k derivatives of $C(k_x,k_y)$ are of order L and are

[5] See, however, the discussion on page 383.

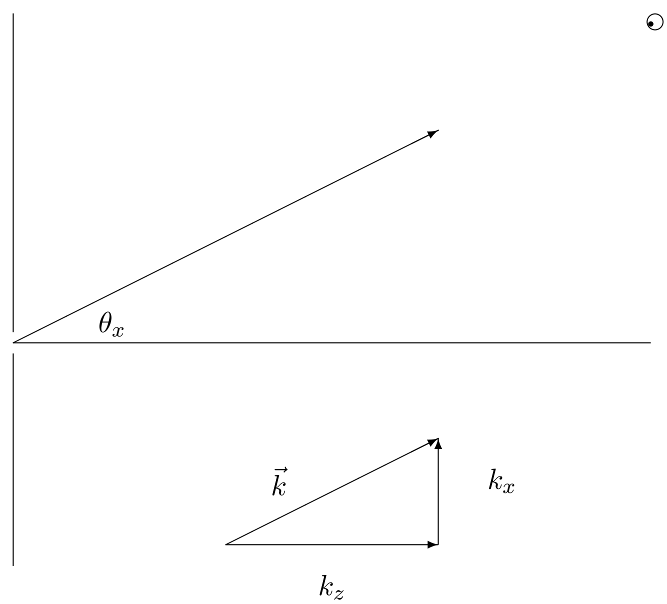


Figure 13.4:

irrelevant. Thus the contribution comes from k_x, k_y such that

$$\frac{\partial}{\partial k_x}\left(X\,k_x + Y\,k_y + Z\sqrt{\omega^2/v^2 - k_x^2 - k_y^2}\right) = X - \frac{Z\,k_x}{\sqrt{\omega^2/v^2 - k_x^2 - k_y^2}} = 0\,,$$
$$\frac{\partial}{\partial k_y}\left(X\,k_x + Y\,k_y + Z\sqrt{\omega^2/v^2 - k_x^2 - k_y^2}\right) = Y - \frac{Z\,k_y}{\sqrt{\omega^2/v^2 - k_x^2 - k_y^2}} = 0\,,\tag{13.39}$$

which is equivalent to (13.38). A careful evaluation of the integral, taking account of the k_x and k_y dependence in the neighborhood of the critical value determined by (13.38) yields an additional factor in the amplitude of the wave of

$$\frac{Z}{r^2} = \frac{\cos\theta}{r}\,,\tag{13.40}$$

where θ is the angle of the vector $\vec{r}$ to the z axis. We expected the $1/r$ factor because of the spreading of the diffracted wave with distance. The factor of $\cos\theta$ is actually the only place where the details of the boundary condition at infinity, (13.21), enter into our expression for the diffracted wave. This factor guarantees that the diffracted wave vanishes as we go to the surface of the opaque screen far from the opening. This is analogous to the "obliquity" factor $(1+\cos\theta)/2$, in the Fresnel-Kirchhoff diffraction theory. The difference between the two is due to the different boundary conditions (our infinite flat barrier versus the lack of incoming spherical waves). We will usually ignore this factor, and indeed it generally does not make much difference where diffraction is important in the forward direction. The important thing is that everything else about the diffraction in the far-field region is determined just by linearity, translation invariance and local interactions.

13.3.4 Spot Size

A useful way to think about the transition from near-field (Fresnel) to far-field (Fraunhofer) diffraction is to consider the size of the spot formed by the beam of figure 13.3 as a function of z. This is a competition between two effects. Increasing the size of the opening makes the spot size larger at small z. However, decreasing the size of the opening increases the spread in k_x, thus increasing the diffraction, and making the spot size larger at large z. For a given z, the best you can do is to choose the size of your opening so that these two effects are of the same order of magnitude. Suppose that the size of your opening is ℓ. Then the spread in k_x is of order $2\pi/\ell$. At large z, the beam spreads into a cone with an opening angle of order

$$\theta \approx \frac{\lambda}{\ell}\,. \tag{13.41}$$

Thus when

$$\frac{\lambda}{\ell} \approx \frac{\ell}{z}\,, \tag{13.42}$$

the spreading of the spot due to diffraction is of the same order of magnitude as the size of the opening. We conclude that to minimize the spot size for a given z, you should choose an opening of size

$$\ell \approx \sqrt{\lambda z}\,. \tag{13.43}$$

The relation, (13.41), up to factors of π, is what defines the region of Fresnel diffraction in figure 13.3. Another way of summarizing the result of this discussion is that for

$$z \gg \frac{\ell^2}{\lambda}\,, \tag{13.44}$$

the spreading due to diffraction is much larger than the spreading due to the size of the opening. This defines the region of far-field, or Fraunhofer diffraction.

13.3.5 Angles

What happens if the plane wave in (13.15) is coming in toward the opaque barrier at an angle, rather than head on? To be specific, suppose that the $\vec{k}$ vector of the wave makes an angle θ with the perpendicular in the x-z plane, so that

$$k_z = k\cos\theta\,, \quad k_x = k\sin\theta\,. \tag{13.45}$$

Then it is reasonable to assume that the analog of (13.15), the amplitude of the wave in the $z = 0$ plane, is[6]

$$f_\theta(x,y) = f(x,y)\, e^{ixk\sin\theta} \tag{13.46}$$

where the additional x dependence has simply been inherited from the x dependence of the incoming wave. We can write the Fourier transform of f_θ in terms of that of f as follows:

$$\begin{aligned} f_\theta(x,y) &= \int dk_x\, dk_y\, C(k_x,k_y)\, e^{i(k_x x+k_y y)}\, e^{ixk\sin\theta} \\ &= \int dk_x\, dk_y\, C(k_x - k\sin\theta, k_y)\, e^{i(k_x x+k_y y)}\,, \end{aligned} \tag{13.47}$$

which implies

$$C_\theta(k_x,k_y) = C(k_x - k\sin\theta, k_y)\,. \tag{13.48}$$

This is entirely reasonable. If the maximum of $C(k_x,k_y)$ occurs at $k_x \approx 0$, the maximum of $C_\theta(k_x,k_y)$ occurs at $k_x = k\sin\theta$. Thus the diffraction pattern appears where a line through the opening in the direction of the incoming plane wave crosses the screen, just as we would expect from a skew beam.

13.4 Examples

13.4.1 The Single Slit

Suppose

$$f(x,y) = \begin{cases} 1 \ \text{ for } \ -a \le x \le a \\ 0 \ \text{ for } |x| > a \end{cases} \tag{13.49}$$

independent of y. This is really a two-dimensional problem, because we can keep $k_y = 0$ and ignore it (except for a factor of 2π, that we won't worry about) by dropping the k_y integral from (13.19). (13.24)

$$C(k_x,k_y) = \frac{1}{4\pi^2}\int dxdy\, f(x,y) e^{-i(k_x x+k_y y)} \tag{13.24}$$

[6] Again, this is simplistic, ignoring complications from the boundaries in the same way as (13.15).

becomes (with the 2π corrected to make it one-dimensional)[7]

$$\begin{aligned} C(k_x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\, f(x)\, e^{-ik_x x} \\ = \frac{1}{2\pi}\int_{-a}^{a} dx\, e^{-ik_x x} &= \frac{1}{-2i\pi k_x}\; e^{-ik_x x}\Big|_{-a}^{a} = \frac{\sin k_x a}{\pi\, k_x}\,. \end{aligned} \tag{13.50}$$

Thus we expect that the intensity of the wave at large z is proportional to $|C(k_x)|^2$,

$$I(x,y) \propto \frac{\sin^2(k_x a)}{k_x^2} \tag{13.51}$$

where

$$\frac{x}{r} = \frac{k_x}{k} = \frac{k_x}{\omega/v} \tag{13.52}$$

or

$$k_x = \frac{\omega}{v}\,\frac{x}{r}\,. \tag{13.53}$$

Thus if we measure the intensity of the diffracted beam, a distance r from the opening, the intensity goes as follows:[8]

$$I(x,y) \propto \frac{\sin^2(2\pi a x/r\lambda)}{x^2} \tag{13.54}$$

where λ is the wavelength of the light. A plot of I as a function of x is shown in figure 13.5. This is called a diffraction pattern. In the important case of light passing through a small aperture, the diffraction pattern can be easily observed by projecting the diffracted beam onto a screen. The features of this pattern worth noting are the large maximum at $x = 0$, with twice the width of all the other maxima, and the periodic zeros for $x = nr\lambda/2a$. Note also that as the width, a of the slit decreases, the size of the diffraction pattern increases.

Moral: This inverse relation between the size of the slit and the size of the diffraction pattern is another illustration of the general feature of Fourier transforms discussed in Chapter 10.

13.4.2 Near-field Diffraction

We will pause here to discuss the region for intermediate z, Fresnel diffraction, where the diffraction problem is complicated. All we can do is to evaluate the integral, (13.19), numerically, by computer, and find the intensity approximately at various values of z. For example, suppose that we take

$$\frac{\omega}{c} = \frac{2\pi}{\lambda} = \frac{100}{a}\,, \tag{13.55}$$

[7] Note that $\sin ka/k$ is well-defined ($= a$) at $k = 0$.

[8] Here we are assuming small angles, so that $\sin\theta \approx \tan\theta$. In our discussion of diffraction gratings below, we will see what happens when the difference in important.

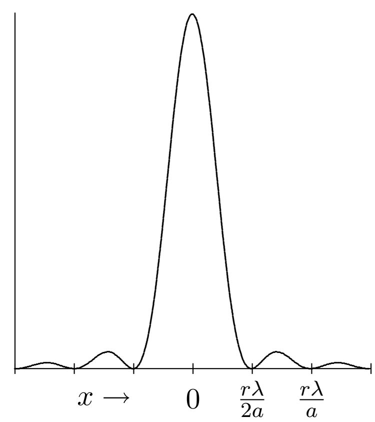


Figure 13.5: The intensity of the diffraction pattern as a function of x.

corresponding to a rather small slit, with a width of only $100/\pi \approx 32$ times the wavelength of the wave. We will then use (13.19) to calculate the intensity of the wave at various values of z, in units of a. For small z, the result is shown in figure 13.6. You can see that the basic beam shape is maintained for a while, as we expected from (13.28). However, wiggles develop immediately. The rather large wiggly diffraction is due to the sharp edges. Below, we will give another example in which the diffraction is much gentler. For intermediate z, shown in figure 13.7, the wiggles begin to coalesce and dramatically change the overall shape of the beam. At the same time, the beam begins to spread out.

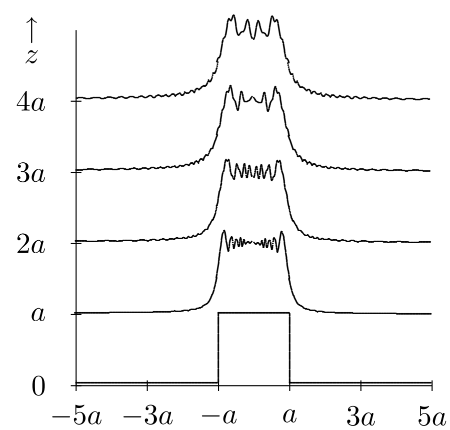


Figure 13.6: The intensity of a wave passing through a slit, for small z.

Finally, in figure 13.8, we show the approach to the large z regions, where diffraction takes over completely and the far field diffraction pattern, (13.54), appears.

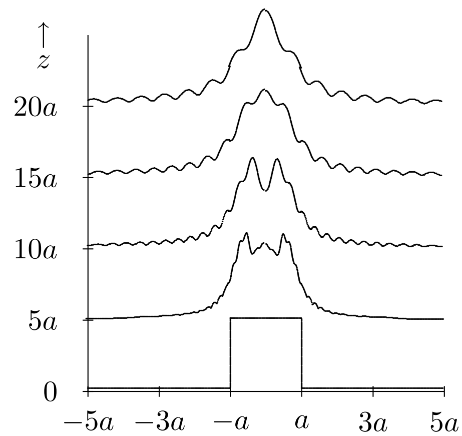


Figure 13.7: The intensity of a wave passing through a slit, for intermediate z.

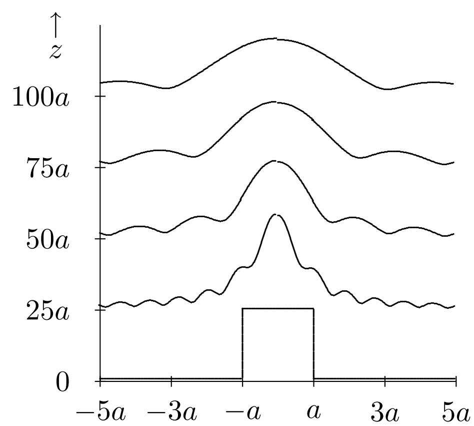


Figure 13.8: The intensity of a wave passing through a slit, as z gets large.

One more example may be interesting. Suppose that instead of being a simple hole in the opaque screen, the opening is shaded in such a way that the wave disturbance at $z = 0$ has the form

$$f(x, y) = e^{-|x|/a} \,. \tag{13.56}$$

The Fourier transform here was done in Chapter 10 in (10.49)-(10.56). Substituting $\omega \to k_x$ and $\Gamma \to 1/a$ in (10.56) gives

$$C(k_x) = \frac{1}{\pi} \frac{a}{1 + a^2 k_x^2} \,. \tag{13.57}$$

This determines the intensity distribution at large z. However, unlike the previous example, this pattern gives very gentle diffraction. For small z, the intensity pattern is shown in figure 13.9. The sharp point in (13.56) disappears, but otherwise the change is very gradual because the initial pattern is very smooth except at $x = 0$. For intermediate and large z, the intensity patterns are shown in figure 13.10 and figure 13.11.

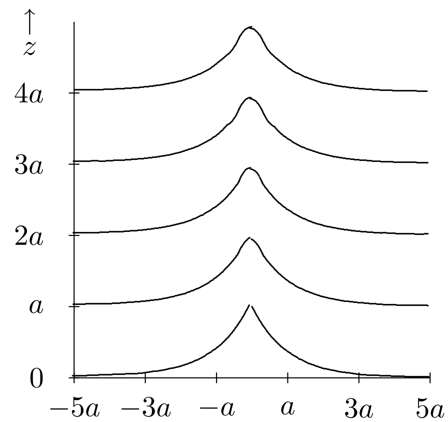


Figure 13.9: The intensity distribution from (13.56) for small z.

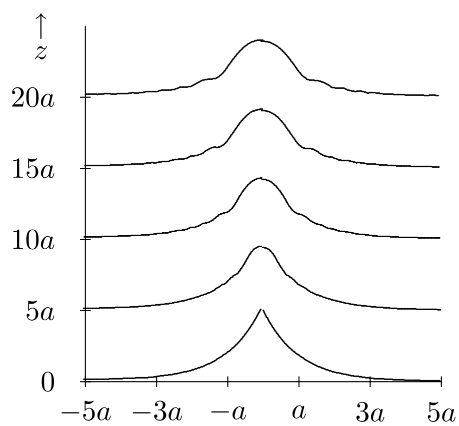


Figure 13.10: The intensity distribution from (13.56) for intermediate z.

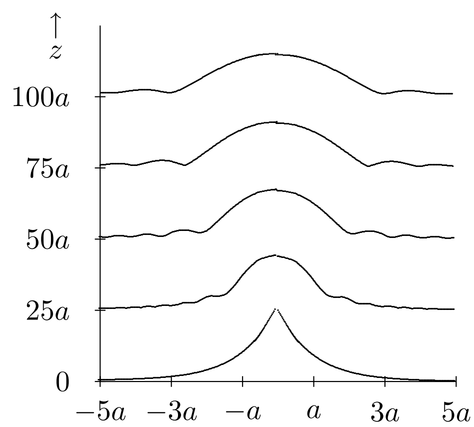


Figure 13.11: The intensity distribution from (13.56) for large z.

13.4.3 The Rectangle

Suppose

$$f(x,y) = \begin{cases} 1 \text{ for } -a_x \le x \le a_x \text{ and } -a_y \le y \le a_y\,, \\ 0 \text{ otherwise.} \end{cases} \tag{13.58}$$

This is the product of a single slit pattern in x with a single slit pattern in y. The Fourier transform is the product of the one-dimensional Fourier transforms

$$\begin{aligned} C(k_x,k_y) &= \frac{1}{4\pi^2}\int_{-a_x}^{a_x} dx\, e^{-ik_x x}\int_{-a_y}^{a_y} dy\, e^{-ik_y y} \\ &= \frac{\sin(k_x a_x)}{\pi\, k_x}\,\frac{\sin(k_y a_y)}{\pi\, k_y}\,. \end{aligned} \tag{13.59}$$

Thus the intensity looks approximately like

$$I(x,y) \propto \frac{\sin^2(2\pi a_x x/r\lambda)}{x^2}\,\frac{\sin^2(2\pi a_y y/r\lambda)}{y^2}\,. \tag{13.60}$$

Of course, once again, because of the general properties of the Fourier transform, if the rectangle is narrow in x, the diffraction pattern is spread out in k_x, and similarly for y.

13.4.4 δ "Functions"

As the slit in (13.49) gets narrower, the diffraction pattern spreads out. Of course, the intensity also decreases. The intensity at $k_x = 0$ is related to the Fourier transform of f at zero, which is just the integral of f over all x. As the slit gets narrower, this integral decreases. But suppose that we increase the intensity of the incoming beam, as a decreases, to keep the intensity of the maximum of the diffraction pattern fixed. Ignoring the y dependence, we require

$$f_a(x) = \begin{cases} \dfrac{1}{2a} \text{ for } -a \le x \le a\,, \\ 0 \text{ for } |x| > a\,. \end{cases} \tag{13.61}$$

The limit of f_a as $a \to 0$ doesn't really exist as a function. It is zero everywhere except $x = 0$. But it goes to ∞ very fast at $x = 0$, so that

$$\lim_{a\to 0} \int dx\, f_a(x) = 1\,. \tag{13.62}$$

It is extraordinarily convenient to invent an object with these properties, called a "δ-function". That is, $\delta(x)$ has the property that it is zero except at $x = 0$, and that

$$\int dx\, \delta(x) = 1\,. \tag{13.63}$$

In fact, this object makes a kind of mathematical sense, so long as you do **not** square it. δ-functions can be manipulated like ordinary functions, added together, multiplied by constants or smooth functions — δ-functions of different variables can even be multiplied — just don't square them! For example, a delta function can be multiplied by an ordinary continuous function:

$$f(x)\,\delta(x) = f(0)\,\delta(x) \tag{13.64}$$

where the equality follows because the delta function vanishes except at $x = 0$, so that only the value of f at 0 matters.

Now it should be clear from (13.63) and (13.64) that the Fourier transform of $\delta(x)$ is just a constant:

$$C(k) = \frac{1}{2\pi} \int dx\, e^{-ikx}\, \delta(x) = \frac{1}{2\pi}\,. \tag{13.65}$$

The diffraction pattern for this thing is thus very boring. There is uniform illumination at all angles.

Of course, in physics, we can't make δ-functions. However, if a, in (13.61) is much smaller than the wavelength of the wave, then it might as well be a δ-function, because it only matters what $C(k_x)$ is for $k_x < k = 2\pi/\lambda$. Larger k_x correspond to exponential waves that die off rapidly with z. But for such k_x, the product $k_x a$ is very small, thus

$$C(k_x) = \frac{1}{2\pi}\, \frac{\sin k_x a}{k_x a} \to \frac{1}{2\pi} \left(1 - \frac{(k_x a)^2}{6} + \cdots\right) \approx \frac{1}{2\pi} \tag{13.66}$$

and we still get uniform diffraction over all angles.

Moral: δ-functions are simply a convenience. When physicists talk about a δ-function, they mean (or at least they should mean) a function like $f_a(x)$, where a is smaller than any physical distance that is important in the problem. Once a gets that small, it is often easier to keep track of the math when you go all the way to the unphysical limit, $a = 0$.

13.4.5 Some Properties of δ-Functions

The Fourier transform of a δ-function is a complex exponential:

$$\text{if } f(x) = \delta(x-a) \text{ then } C(k) = \tfrac{1}{2\pi} e^{-ika}. \tag{13.67}$$

The Fourier transform of a complex exponential is a δ-function:

$$\text{if } f(x) = e^{-i\ell x} \text{ then } C(k) = \delta(k-\ell). \tag{13.68}$$

A δ-function can be reached as a limit in a variety of different ways. For example, from (13.68), we would expect that as $a \to \infty$, the Fourier transform of (13.49) should approach a δ-function:

$$\lim_{a\to\infty} \frac{\sin k_x a}{k_x} = \delta(k_x)\,. \tag{13.69}$$

13.4.6 One Dimension from Two

Using δ-functions, we can say more elegantly what is meant by the statement we made above that if $f(x,y)$ does not depend on y, the problem is one-dimensional. If we look at the limit of (13.58) as $a_y \to \infty$, it goes over into (13.49). In other words, when a rectangle is infinitely long, it is a slit. In this limit, the Fourier transform, (13.59) goes into

$$\frac{\sin(k_x a_x)}{\pi\, k_x}\,\delta(k_y)\,. \tag{13.70}$$

This is the real meaning of (13.50). It is one-dimensional in the sense that k_y is stuck at 0. There is no diffraction in the y direction.

13.4.7 Many Narrow Slits

An interesting application of δ-functions is to the diffraction pattern for several narrow slits. We will use this later in various ways. Consider a function, $f(x,y)$ of the form

$$\sum_{j=0}^{n-1} \delta(x - jb) \tag{13.71}$$

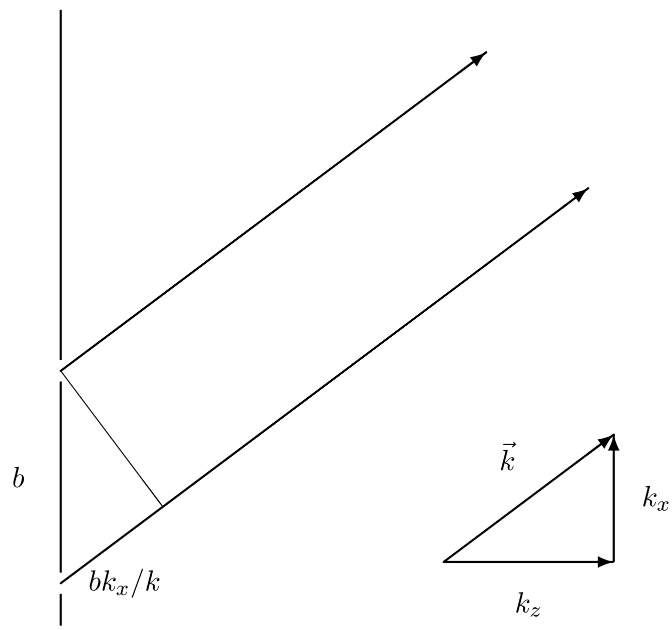


Figure 13.12: If $bk_x/k = n\lambda$, the interference is constructive.

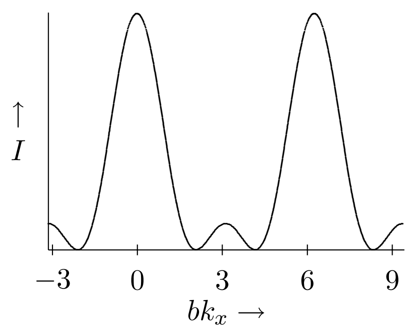


Figure 13.13: The diffraction pattern for three narrow slits.

This describes a series of n narrow slits[9] at $x = 0$, $x = b$, $x = 2b$, *etc.*, up to $x = (n-1)b$. The Fourier transform of (13.71) is a sum of contributions from the individual δ-functions,

[9]"Narrow" here means narrow compared to the wavelength of the light — see the moral above.

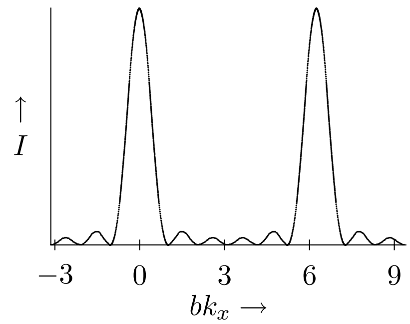


Figure 13.14: The diffraction pattern for 6 narrow slits.

from (13.67) and (13.68)

$$C(k_x, k_y) = \delta(k_y)\,\frac{1}{2\pi}\sum_{j=0}^{n-1} e^{-ijbk_x}\,. \tag{13.72}$$

But the sum is a geometric series that can be done explicitly:

$$\begin{gathered}\sum_{j=0}^{n-1} e^{-ijbk_x} = \frac{1-e^{-inbk_x}}{1-e^{-ibk_x}} \\ = \frac{e^{-inbk_x/2}\left(e^{inbk_x/2}-e^{-inbk_x/2}\right)}{e^{-ibk_x/2}\left(e^{ibk_x/2}-e^{-ibk_x/2}\right)} = e^{-i(n-1)bk_x/2}\,\frac{\sin nbk_x/2}{\sin bk_x/2}\,.\end{gathered} \tag{13.73}$$

Thus the diffraction pattern intensity is proportional to

$$\frac{\sin^2 nbk_x/2}{\sin^2 bk_x/2}\,. \tag{13.74}$$

For $n = 2$, (13.74) is just

$$4\cos^2\frac{bk_x}{2} = 2(1+\cos bk_x)\,. \tag{13.75}$$

This is the problem with which we started the chapter. When $bk_x = 2m\pi$ for integer m, then the wave from one slit travels farther than the wave from the other by $m\lambda$, where $\lambda = 2\pi/k$ is the wavelength. Thus for $bk_x = 2m\pi$ the interference is constructive, as illustrated in figure 13.12.

For larger n, we still get constructive interference for $bk_x = 2m\pi$, but the maxima are sharper, because with more slits, there are more possibilities for destructive interference at other angles. In figure 13.13 and figure 13.14, we plot (13.74) versus bk_x from ($-\pi$ to 3π so that you can see two full periods) for $n = 3$ and 6. Notice the appearance of $n-2$ secondary maxima between the primary maxima of the intensity. We will return to these relations when we discuss diffraction gratings.

13.5 Convolution

There is a rather simple theorem, know as the convolution theorem, that is extremely useful in dealing with Fourier transforms. Suppose that we have two functions, $f_1(x)$ and $f_2(x)$. Define the function $f_1 \circ f_2$ as follows:

$$f_1 \circ f_2(x) = \int_{-\infty}^{\infty} dy\, f_1(x-y)\, f_2(y)\,. \tag{13.76}$$

This integral will be well defined if $f_1(x)$ and $f_2(x)$ fall off fast enough at infinity (and certainly if they are nonzero only in a finite region of x). Note that $f_1 \circ f_2$ is a function of a single variable. It is also symmetric under the exchange of the two functions, because by a simple change of variables ($y \to x - y$)

$$f_1 \circ f_2(x) = \int_{-\infty}^{\infty} dy\, f_1(x-y)\, f_2(y) = \int_{-\infty}^{\infty} dy\, f_1(y)\, f_2(x-y) = f_2 \circ f_1(x)\,. \tag{13.77}$$

Now the theorem is that the Fourier transform of the convolution is 2π times the product of the Fourier transforms of the two functions. The proof is immediate (all integrals run from $-\infty$ to ∞):

$$\begin{aligned} C_{f_1\circ f_2}(k) &= \frac{1}{2\pi}\int dx\, e^{ikx}\, f_1 \circ f_2(x) \\ &= \frac{1}{2\pi}\int dx\, e^{ikx} \int dy\, f_1(x-y)\, f_2(y)\,. \end{aligned} \tag{13.78}$$

Now we substitute $x \to y + z$ and write the integral over y and z,

$$\begin{aligned} &= \frac{1}{2\pi}\int dz\, e^{ik(y+z)} \int dy\, f_1(x-y)\, f_2(y) \\ &= \frac{1}{2\pi}\int dz\, e^{ikz}\, f_1(z) \int dy\, e^{iky}\, f_2(y) = 2\pi\, C_{f_1}(k)\, C_{f_2}(k)\,. \end{aligned} \tag{13.79}$$

The two-dimensional analog of (13.79) is a straightforward extension. The two-dimensional convolution is

$$f_1 \circ f_2(x,y) = \int dx'\, dy'\, f_1(x-x', y-y')\, f_2(x', y') \tag{13.80}$$

$$C_{f_1\circ f_2}(k_x, k_y) = 4\pi^2\, C_{f_1}(k_x, k_y)\, C_{f_2}(k_x, k_y)\,. \tag{13.81}$$

13.5.1 Repeated Patterns

The convolution theorem can be used to understand many interesting situations. Consider the following very instructive pattern of two wide slits:

$$f(x,y) = \begin{cases} 1 & \text{for } -a \le x \le a \\ 1 & \text{for } -a \le x - b \le a \\ 0 & \text{otherwise} \end{cases} \tag{13.82}$$

for $b > 2a$. A piece of the pattern is shown in figure 13.15 for $b = 3.5a$.

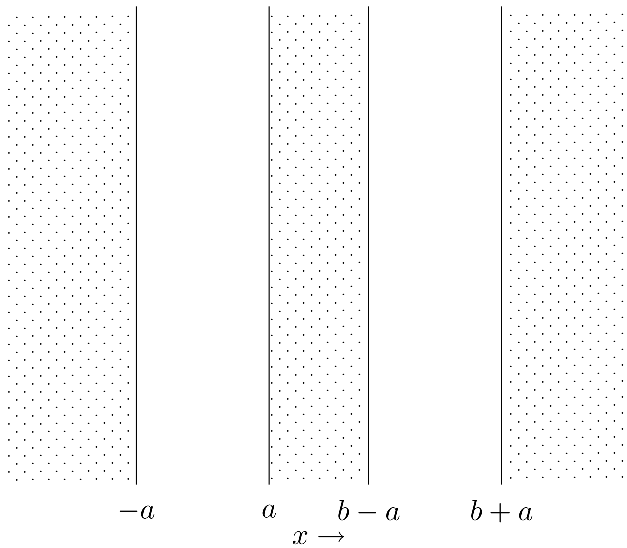


Figure 13.15: A piece of the opaque barrier with two wide slits.

This can be regarded as the convolution of two functions:

$$f = f_1 \circ f_2 \tag{13.83}$$

where

$$f_1(x,y) = \begin{cases} 1 & \text{for } -a \le x \le a \\ 0 & \text{otherwise} \end{cases} \tag{13.84}$$

and

$$f_2(x,y) = \delta(x)\,\delta(y) + \delta(x-b)\,\delta(y)\,. \tag{13.85}$$

The corresponding Fourier transforms are, from (13.70)

$$C_{f_1}(k_x,k_y) = \frac{\sin(k_x a)}{\pi\, k_x}\,\delta(k_y) \tag{13.86}$$

and from (13.73)

$$C_{f_2}(k_x,k_y) = \frac{1}{4\pi^2}\cos\frac{bk_x}{2}\,e^{-ibk_x/2}\,. \tag{13.87}$$

Now applying the convolution theorem gives

$$C_{f_1\circ f_2}(k_x,k_y) = \cos\frac{bk_x}{2}\,e^{-ibk_x/2}\,\frac{\sin(k_x a_x)}{\pi\, k_x}\,\delta(k_y)\,. \tag{13.88}$$

Because $b > 2a$, this describes a pattern that oscillates rapidly on the scale set by $1/b$, with an amplitude that varies with the single slit diffraction pattern characterized by size $1/a$. The intensity pattern on a distant screen is shown in figure 13.16, for $b = 3.5a$ The dotted line is the pattern for a single wide slit (compare (13.5)).

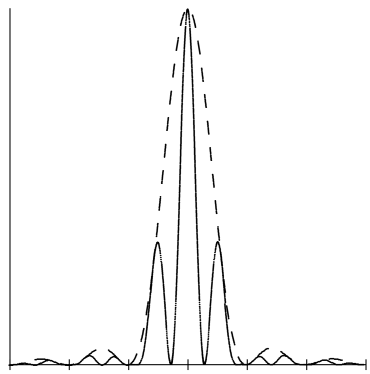


Figure 13.16: The diffraction pattern for two wide slits.

13.6 Periodic $f(x, y)$

Suppose $f(x, y)$ is periodic in x with period a. That is

$$f(x + a, y) = f(x, y)\,. \tag{13.89}$$

Then $C(k_x, k_y)$ can only be nonzero if

$$k_x = \frac{2\pi\, n}{a}\,. \tag{13.90}$$

To see this, insert (13.89) into (13.24),

$$C(k_x, k_y) = \frac{1}{4\pi^2} \int dx\, dy\, f(x + a, y)\, e^{i(k_x x + k_y y)}\,. \tag{13.91}$$

If we change variables from $x \to x - a$, (13.91) is

$$C(k_x, k_y) = \frac{1}{4\pi^2} \int dx\, dy\, f(x, y)\, e^{i(k_x x - k_x a + k_y y)} = e^{-ik_x a}\, C(k_x, k_y) \tag{13.92}$$

because the constant phase factor can be taken outside the integral. (13.90) follows because (13.92) implies that either $C(k_x, k_y) = 0$ or $e^{-ik_x a} = 1$.

An example of this general principle is (13.74). In the limit that $n \to \infty$, (13.74) goes to 0 except for $k_x = 2\pi m/b$ for integer m (where it is infinite). This example is simple because the slits are narrow, so the intensity is independent of m. However, with repeated wide slits, or some more complicated pattern, we could use the convolution theorem and (13.74) to see that (13.90) emerges as $n \to \infty$. The details of the pattern of each slit will then determine the relative intensity of the diffraction pattern at different m.

Thus any infinite regular pattern produces a discrete sequence of k's. For example, a transmission diffraction grating, that consists of lots of equally spaced lines in the y direction with x separation a on a transparent substrate, produces a $C(k_x, k_y)$ that is nonzero only for $k_y = 0$ (because there is no y dependence at all) and $k_x = 2n\pi/a$. Then (13.19) becomes

$$\sum_n C_n \, e^{i(2n\pi x/a + z\sqrt{\omega^2/v^2 - (2n\pi/a)^2} - \omega t)}. \tag{13.93}$$

This describes a linear superposition of plane waves fanning out at angles in the x direction given by

$$\sin\theta_n = \frac{2\pi\, n\, v}{a\,\omega} = \frac{n\lambda}{a} \tag{13.94}$$

as shown in figure 13.17.

Typically, for a transmission grating, most of the light goes into the central line, which is to say that you can see right through the grating. Note that the even spacing in $\sin\theta_n$ in (13.94) corresponds to an increasing spacing of the lines projected onto a screen at fixed large z (for example, a screen like your retina!) because the distance along the screen is determined by

$$\tan\theta_n = \frac{n\lambda}{\sqrt{a^2 - n^2\lambda^2}}\,. \tag{13.95}$$

There is a maximum value of n, above which no propagating wave is produced (because it corresponds to $\sin\theta > 1$ and thus imaginary k_z).

Note also the dependence of (13.94) on wavelength. The larger the wavelength of the light, the larger the angles in the pattern from the diffraction grating. This, of course, is why the diffraction grating is useful. It can separate light of different frequencies. The different colors of the rainbow are spread out along a line, for each value of n. This is illustrated in the figure 13.18, for three frequencies, blue light with wavelength 4300 Å, green light with wavelength 5200 Å and red light with wavelength 6300 Å, incident on a diffraction grating with 10,000 lines per inch. We have shown (13.95) for $n = -3$ to 3 and labeled the colors for the $n = 1$ secondary maximum. As you see, in a realistic grating, the angles of diffraction can be large, and it is a very bad idea to use a small angle approximation.

13.6.1 Twisting the Grating

Some interesting examples of the effects discussed in (13.48) occur when the incoming light wave comes at the grating at an angle with respect to the perpendicular. Starting with the

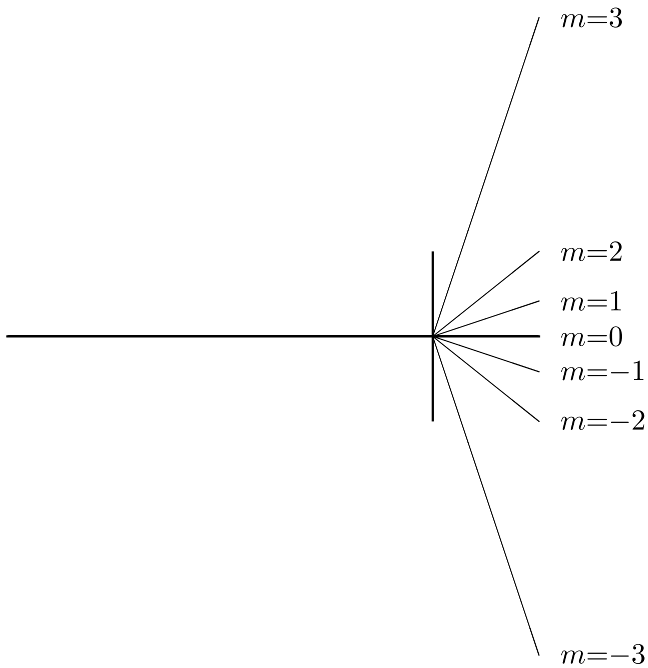


Figure 13.17: A transmission diffraction grating splits a beam of a single frequency.

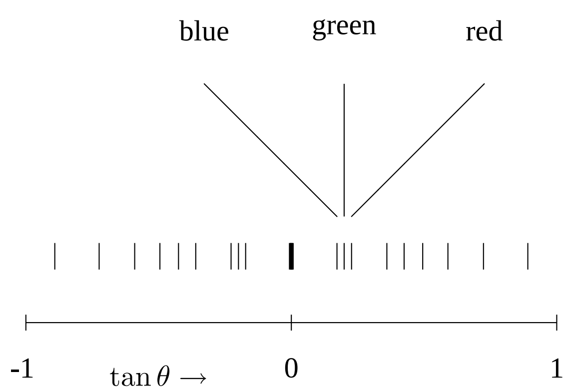


Figure 13.18: The pattern of three frequencies of light from a grating.

grating lines in the y direction and the grating in the x-y plane, there are two different effects.

1: Twisting Around the y Axis

Suppose that the light comes in at an angle $\theta_{\rm in}$ from the perpendicular in the x-z plane. Then from (13.48),

$$C_{\theta_{\rm in}}(k_x,k_y) = C(k_x - k\sin\theta_{\rm in}, k_y) \tag{13.96}$$

where C is Fourier transform for the perpendicular grating,

$$C(k_x,k_y) \neq 0 \quad \text{for} \quad k_y = 0,\ k_x = \frac{2\pi n}{a}\,. \tag{13.97}$$

Thus

$$\begin{gathered} C_{\theta_{\rm in}}(k_x,k_y) \neq 0 \quad \text{for} \\ k_y = 0,\ k_x = k\sin\theta_{\rm in} + \frac{2\pi n}{a} \end{gathered} \tag{13.98}$$

or

$$\sin\theta = \frac{k_x}{k} = \sin\theta_{\rm in} + \frac{n\lambda}{a}\,. \tag{13.99}$$

In other words, $\sin\theta$ is simply displaced by $\sin\theta_{\rm in}$. For example, this means that if $\theta = \pi/a$, the pattern is exactly the same, but the central maximum has moved over, as shown in figure 13.19.

2: Twisting Around the x Axis

Suppose that the light comes in at an angle θ from the perpendicular in the y-z plane. Then from (13.48).

$$C_{\theta_{\rm in}}(k_x,k_y) = C(k_x, k_y - k\sin\theta_{\rm in}). \tag{13.100}$$

Now instead of being 0, k_y is fixed at $k\sin\theta_{\rm in}$

$$k_y = k\sin\theta_{\rm in}\,, \quad k_x = \frac{2\pi n}{a}\,. \tag{13.101}$$

Now the diffracted waves make nontrivial angles from the perpendicular both in x and in y

$$\sin\theta_y = \frac{k_y}{\sqrt{k_y^2+k_z^2}} = \frac{k_y}{\sqrt{k^2-k_x^2}} = \frac{\sin\theta_{\rm in}}{\sqrt{1-n^2\lambda^2/a^2}} \tag{13.102}$$

and

$$\sin\theta_x = \frac{k_x}{\sqrt{k_x^2+k_z^2}} = \frac{k_x}{\sqrt{k^2-k_y^2}} = \frac{n\lambda}{a\cos\theta_{\rm in}}\,. \tag{13.103}$$

Again, as in (13.95), what we see if we project the pattern onto a perpendicular screen at fixed z are the tangents,

$$(x,y)_{\rm screen} = z\,(\tan\theta_x, \tan\theta_y)\,, \tag{13.104}$$

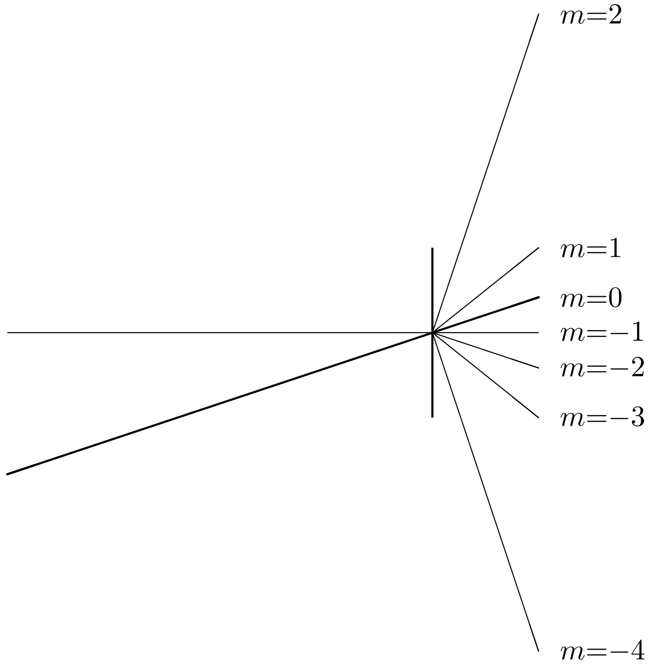


Figure 13.19: The pattern for a beam at an angle, $\theta_{\rm in} = \arcsin \lambda/a$.

where

$$\tan\theta_x = \frac{k_x}{k_z}\,, \quad \tan\theta_y = \frac{k_y}{k_z}\,. \tag{13.105}$$

Thus the diffraction pattern appears curved. What one sees on a screen or a retina is the colors of the rainbow spread out along a curved line. This is shown in figure 13.20, where we plot $\tan\theta_x$ versus $\tan\theta_y$ for a light source and grating as in (13.18), above, but with $\sin\theta_{\rm in} = 0.5$. Note that the pattern has not only curved, it has spread out, compared to (13.18). Here you really see the three-dimensional $\vec{k}$ vector in action. As $\tan\theta_y$ increases, for fixed k_x, $\tan\theta_x$ increases as well, because k_z decreases.

13.6.2 Resolving Power

The discussion so far has assumed that the diffraction grating is truely periodic. But this is only possible if the grating is infinite! In a finite grating, only the middle is periodic. The edges break the periodicity. In a grating consisting of only a finite number of grooves, n, the diffraction peaks are not infinitely sharp. They are not delta functions. However, as discussed at the beginning of this section, we actually already know what they look like in the finite

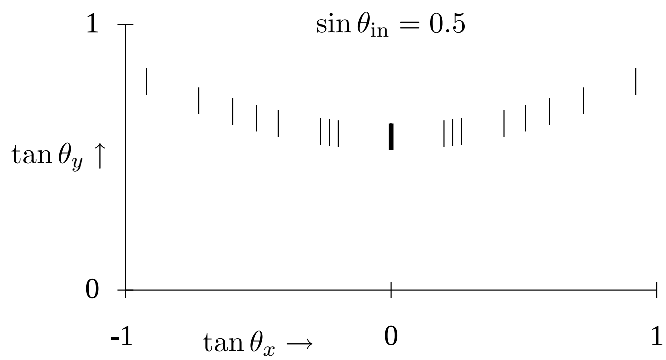


Figure 13.20: The diffraction pattern from a twisted grating.

case because we have solved the problem of diffraction from n evenly spaced narrow slits, in (13.74). In the general situation for n identical grooves, the intensity looks like (13.74) multiplied by some slowly varying function that depends on the shape of the grooves (by the convolution theorem, (13.79)). The important consequence of this is that the shape of a diffraction peak for an n-slit grating is roughly given by (13.74).

The shape of the diffraction peak is important for the following practical question. Suppose that you have a beam of light that consists of a superposition of light of two different frequencies. How close together do the frequencies have to be before their nontrivial diffraction peaks melt together, so that you cannot use your diffraction grating to distinguish them? The larger the number of grooves in the grating, the sharper the diffraction peaks and the easier it is to distinguish different frequencies.

Rayleigh's criterion is an historically important way of answering this question. Rayleigh assumed that it would be possible to distinguish the diffraction maxima from equally intense waves of slightly different wavelengths if the maximum of one frequency coincides with the first minimum of the other. For a grating of 6 lines, this criterion is illustrated in figure 13.21. The solid line is the total intensity of a wave consisting of two slightly different frequencies. The contributions from the separate frequency components are indicated by the dotted and dashed lines.

Any such fixed criterion for resolving power should be regarded not as a fact about nature, but as a conventional definition that facilitates communication between experimenters. It is always possible to do better than any given definition by accumulating accurate data on the line shape and modeling the details.

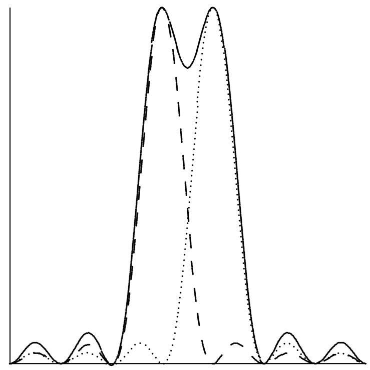


Figure 13.21: Rayleigh's criterion for a grating with 6 lines.

13.6.3 Blazed Gratings

As a spectroscope, the transmission diffraction grating has a disadvantage compared to a prism. The difficulty is that, as we noted above, most of the light impinging on the grating goes right through and is not split into its component frequencies. This is a very serious problem in devices in which the total amount of light is limited. It is often important to have the bulk of the light going into a single **nonzero** value of n in (13.94). Then nearly all of the photons can be used for the measurement, rather than being wasted in the $n = 0$ maximum (which carries no information about the frequency). As we argued above, there is no theoretical reason why such a thing cannot be done. The general principles of translation invariance and local interactions determine the possible angles of diffraction, but not how much light goes to which angle.

In fact, there is a practical and widely used method in reflection gratings. A reflecting surface with a series of evenly spaced parallel lines scored into it acts as a reflection grating, as illustrated in figure 13.22. This shows a reflection grating in which the predominant reflection of a beam coming in perpendicular to the plane of the grating is also perpendicular. What we want instead is shown in figure 13.23. To construct such a grating, you can shape the grooves in the grating so that the specular reflection from the individual grooves directs the beam into the nontrivial diffraction maximum, as shown in figure 13.24.

To do this, you can choose the angle of the blaze to be half the angle of the first maximum, $\theta_1 = 2\pi v/a\omega$, in (13.94), as shown in the blow-up of a groove in figure 13.25.

13.7 * X-ray Diffraction

A beautiful three-dimensional example of diffraction from a periodic function is x-ray diffraction from crystals. A crystal is a regular array of atoms whose positions can be described by

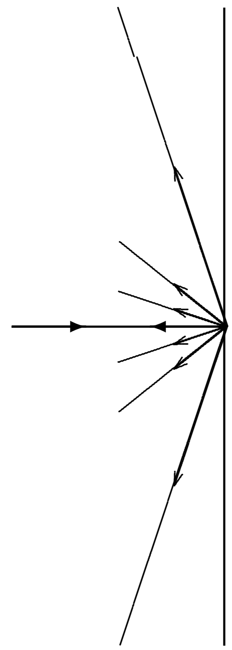


Figure 13.22: A reflection diffraction grating splits a beam of a single frequency.

a periodic function

$$f(\vec{r}) = f(\vec{r} + \vec{a}) \tag{13.106}$$

where $\vec{a}$ is any vector from one point on the lattice to another. Mathematically, we can define the lattice as the set of all such vectors. Note that the lattice always includes the zero vector, the point at the origin. The three-dimensional Fourier transform of $f(\vec{r})$ is nonzero **only** for wave number vectors of the form

$$2\pi \sum_{j=1}^{3} n_j \vec{\ell}_j \tag{13.107}$$

where $\vec{\ell}_j$ are the basis vectors for the **“dual”** or **“reciprocal”** lattice that satisfies

$$\vec{a} \cdot \vec{\ell}_j = \text{integer, for all } \vec{a}\,. \tag{13.108}$$

The idea here is the same as the one-dimensional discussion of the diffraction grating, that $k_x = 2\pi n/a$, (13.90). The derivation of (13.107) is precisely analogous to that of (13.90).

We can visualize the relation between the lattice and the dual lattice more easily for two-dimensional “crystals.” For example, consider a lattice of the form

$$\vec{a} = n_x a_x \hat{x} + n_y a_y \hat{y} \tag{13.109}$$

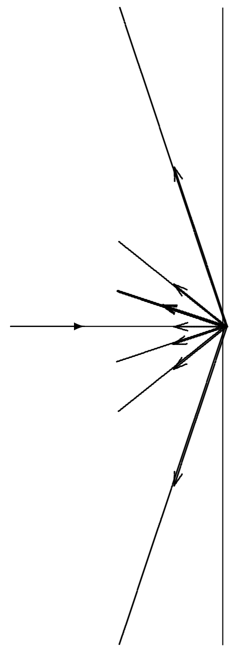


Figure 13.23: A blazed grating directs the beam into a nontrivial diffraction maximum.

shown in figure 13.26 (for $a_x = 2a_y$).

It is clear that vectors of the form

$$\vec{\ell}_1 = \frac{1}{a_x}\hat{x}\,, \quad \ell_2 = \frac{1}{a_y}\hat{y}\,, \tag{13.110}$$

satisfy (13.108). Furthermore, a little thought will convince you that these are the shortest pair of linearly independent vectors with this property. Thus we can take (13.110) to be the basis vectors for the dual lattice, so that the dual lattice looks like

$$\vec{d}_m = \left(\frac{m_x}{a_x}\hat{x} + \frac{m_y}{a_y}\hat{y}\right) \tag{13.111}$$

as shown in figure 13.27. Note that the long and short axes are interchanged, as usual in a diffraction process.

Now suppose that there is a plane wave passing through the infinite lattice,

$$e^{i\vec{k}\cdot\vec{r}-i\omega t}\,. \tag{13.112}$$

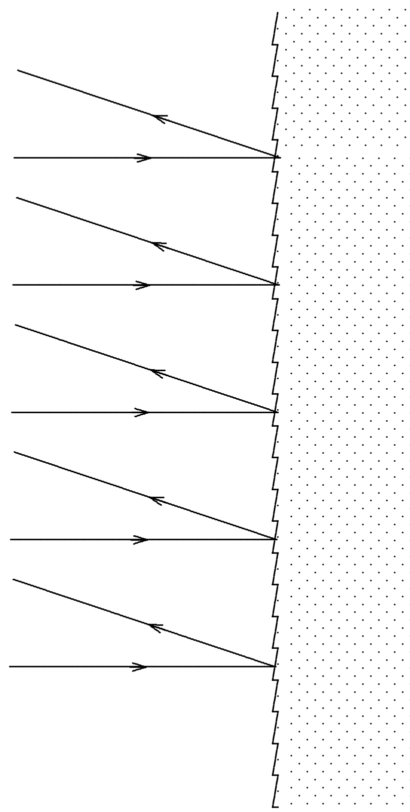


Figure 13.24: The grooves of a blazed grating.

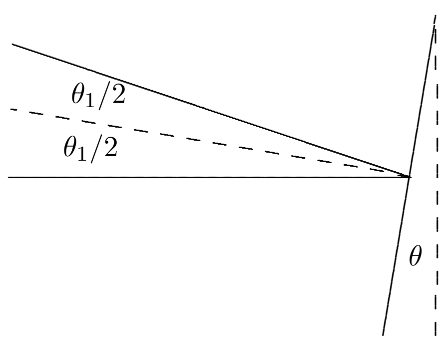


Figure 13.25: $\theta \approx \theta_1/2$.

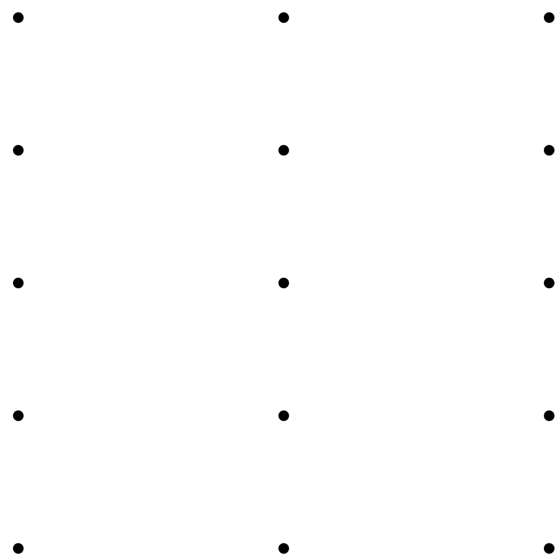


Figure 13.26: A crystal lattice.

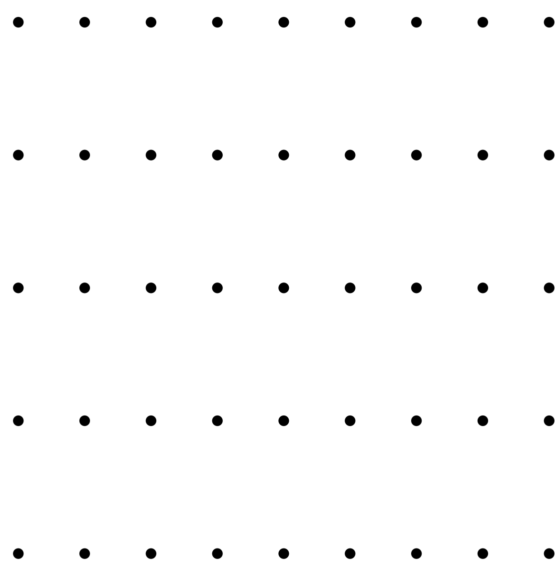


Figure 13.27: The dual lattice.

The wave that results from the interaction of the plane wave with the lattice then has the form

$$e^{i\vec{k}\cdot\vec{r}-i\omega t}\,g(\vec{r})\,, \tag{13.113}$$

where $g(\vec{r})$ is a periodic function, like $f(\vec{r})$ in (13.106). To find the possible refracted waves, we must write this in the form:

$$e^{i\vec{k}\cdot\vec{r}-i\omega t}\,g(\vec{r}) = \sum_{\substack{\text{diffracted}\\ \text{waves}\,,\ \alpha}} C_\alpha\, e^{i\vec{k}_\alpha\cdot\vec{r}-i\omega t}\,. \tag{13.114}$$

But we also know from the discussion above that the Fourier transform of g is nonzero only for values of $\vec{k}$ of the form (13.107). Thus (13.114) takes the form

$$e^{i\vec{k}\cdot\vec{r}-i\omega t}\int d^3k'\, e^{i\vec{k}'\cdot\vec{r}}\, C_g(\vec{k}') = e^{i\vec{k}\cdot\vec{r}-i\omega t}\sum_{n_j} C_{n_j} e^{2\pi i\sum_j n_j\vec{\ell}_j\cdot\vec{r}}\,. \tag{13.115}$$

Therefore, the $\vec{k}_\alpha$ in (13.114) must have the form

$$\vec{k}_\alpha = \vec{k} + 2\pi\sum_j n_j\vec{\ell}_j\,. \tag{13.116}$$

But this is only possible if $\vec{k}_\alpha$ satisfies the dispersion relation in the material, which means, if the material is rotation invariant so that ω^2 depends only on $|\vec{k}|^2$, that

$$|\vec{k}_\alpha|^2 = |\vec{k}|^2\,. \tag{13.117}$$

Thus we get a diffracted wave only for n_j such that (13.117) is satisfied. X-ray diffraction from a crystal, therefore, can provide direct information about the dual lattice and thus about the crystal lattice itself.

There is a more physical way of thinking about the dual lattice. Consider any vector in the **dual** lattice that is not a multiple of another,

$$\vec{d} \equiv \sum_j n_j\vec{\ell}_j\,. \tag{13.118}$$

Now look at the subset of vectors on the **lattice** that satisfy

$$\vec{d}\cdot\vec{a} = 0\,. \tag{13.119}$$

This subset is the set of lattice points that lie in the plane, $\vec{d}\cdot\vec{r} = 0$, that is the plane perpendicular to $\vec{d}$ passing through the origin. Now consider the subset

$$\vec{d}\cdot\vec{a} = 1\,. \tag{13.120}$$

This subset is the set of lattice points that lie in the plane, $\vec{d}\cdot\vec{r} = 1$, that is parallel to the plane $\vec{d}\cdot\vec{r} = 0$, in the lattice. This plane is also perpendicular to $\vec{d}$ and passes through the point (which may not be a lattice point)

$$r_1 = \frac{\vec{d}}{|\vec{d}|^2}\,. \tag{13.121}$$

Therefore, the perpendicular distance (that is in the $\vec{d}$ direction) between the two planes is

$$\hat{d} \cdot \vec{r}_1 = \frac{1}{|\vec{d}|} \,. \tag{13.122}$$

We can continue this discussion to conclude that the subset of lattice points satisfying

$$\vec{d} \cdot \vec{a} = m \text{ for integer } m = -\infty \text{ to } \infty \tag{13.123}$$

is the set of lattice points lying on parallel planes perpendicular to $\vec{d}$, with adjacent planes separated by $1/|\vec{d}|$. **But this set must be all the lattice points!** This is true because $\vec{d} \cdot \vec{a}$ is an integer for all lattice points by the definition of the dual lattice. Thus all lattice points lie in one of the planes in (13.123).

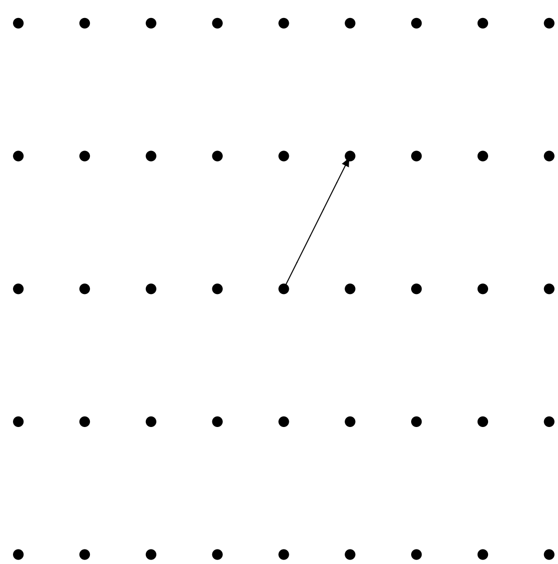


Figure 13.28: A vector in the dual lattice.

These considerations are illustrated in the two-dimensional crystal in the pictures below. If the vector $\vec{d}$ in the dual lattice is as shown in figure 13.28, then the perpendicular planes in the lattice are shown in figure 13.29.

Now suppose that $\vec{d}$ is one of the special points in the dual lattice that gives rise to a refracted wave, so that

$$|\vec{k} + 2\pi\vec{d}\,|^2 = |\vec{k}\,|^2 \Rightarrow \vec{d} \cdot \left(\vec{k} + \pi\vec{d}\right) = 0 \,. \tag{13.124}$$

This relation is shown in figure 13.30. This shows that the k vector of the refracted wave, $\vec{k} + 2\pi\vec{d}$, is just $\vec{k}$ reflected in a plane perpendicular to $\vec{d}$. We have seen that there are an

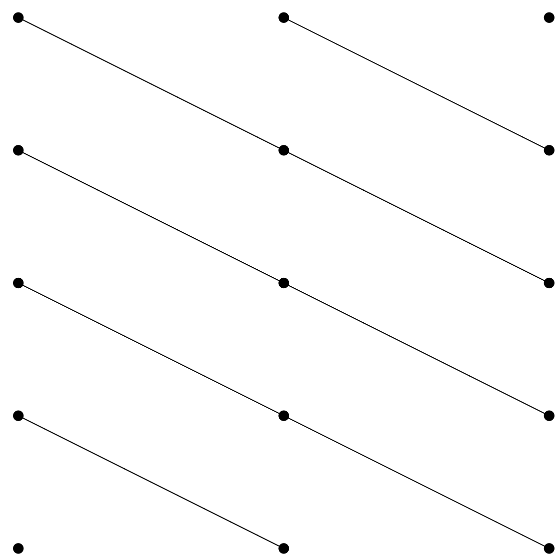


Figure 13.29: The corresponding planes in the lattice.

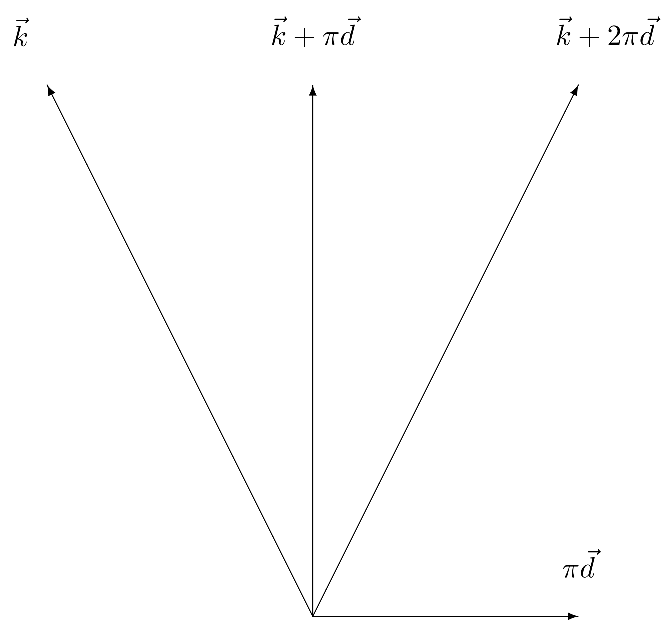


Figure 13.30: The Bragg scattering condition.

infinite number of such planes in the lattice, separated by $1/|\vec{d}|$. The contribution to the scattered wave from each of these planes adds **constructively** to the refracted wave. To see this, consider the phase difference between the incoming wave, $e^{i\vec{k}\cdot\vec{r}-i\omega t}$ and the diffracted wave $e^{i\vec{k}_\alpha\cdot\vec{r}-i\omega t}$ for $\vec{k}_\alpha = \vec{k} + 2\pi\vec{d}$. Evidently, the phase difference at any point $\vec{r}$ is

$$2\pi\vec{d}\cdot\vec{r}\,. \tag{13.125}$$

This phase difference is an integral multiple of 2π on all the planes

$$\vec{d}\cdot\vec{r} = m \text{ for integer } m = -\infty \text{ to } \infty\,. \tag{13.126}$$

Thus the contribution to scattering from all of the planes of lattice points adds constructively, because the phase relation between the incoming and diffracted wave is the same on all of them. Conversely, if $\vec{k}_\alpha \neq \vec{k} + 2\pi\vec{d}$, then the contribution from different planes will interfere destructively, and no diffracted wave will result.

This physical interpretation goes with the name “Bragg scattering.” The planes, (13.123) (or (13.126)) are the Bragg planes of the crystal. Note that as the vector $\vec{d}$ in the dual lattice gets longer, the corresponding Bragg planes get closer together, but they are also less dense, containing fewer scattering centers per unit area. Generally the scattering is weaker for large $|\vec{d}|$.

13.8 Holography

Nothing prevents us from doing the analysis of a diffraction pattern from a more complicated function, $f(x,y)$, than that discussed in (13.16). A hologram is just such a diffraction pattern. One of the simplest versions of a hologram is one in which an object is illuminated by a laser, that provides essentially a plane wave. The reflected light, and a part of the laser beam (extracted by some beam splitting technique) are incident on a photographic plate at slightly different angles, as shown schematically in figure 13.31. The wave incident on the photographic plate has the form

$$e^{-i\omega t}\left(e^{ikz} + \int dk_x\,dk_y\,C(k_x,k_y)\,e^{i\vec{k}\cdot\vec{r}}\right) \tag{13.127}$$

where

$$k = |\vec{k}| = \omega/v\,. \tag{13.128}$$

(13.127) describes the two coherent parts of the light wave incident on the photographic plate. For simplicity, we will assume that the signal in which we are actually interested, the reflected wave with Fourier transform $C(k_x,k_y)$, is small compared to the reference wave e^{ikz}. This signal is what we would see if the photographic plate were removed and we placed

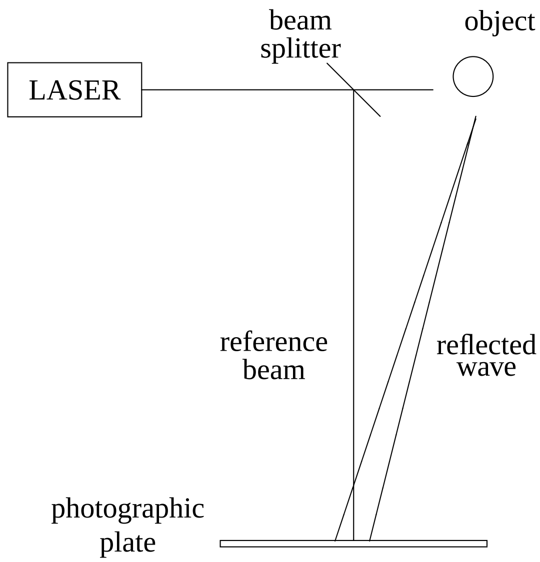


Figure 13.31: Making a hologram.

our eyes in the path of the reflected wave, but out of the path of the laser beam, as shown in figure 13.32.

The photographic plate (we'll assume it's at $z = 0$) records only the intensity of the total wave, proportional to

$$1 + 2\text{Re} \int dk_x \, dk_y \, C(k_x, k_y) e^{i(k_x x + k_y y)} + \mathcal{O}(C^2) \, . \tag{13.129}$$

We will drop the terms of order C^2, assuming that C is small, although we will be able to see later that they will not actually not make any difference even if C is large. If we now make a positive slide from the plate and shine through it a laser beam with the same frequency, ω, the wave "gets through" where the light intensity on the plate was large and is absorbed where the intensity was small. Thus we have a forced oscillation problem of exactly the sort that we discussed above, with (13.129) playing the role of $f(x, y)$. The solution for $z > 0$ (from (13.19)-(13.24)) is

$$e^{-i\omega t} \left(e^{ikz} + \int dk_x \, dk_y \, C(k_x, k_y) \, e^{i\vec{k}\cdot\vec{r}} + \text{c.c.} \right) \tag{13.130}$$

where c.c. is the complex conjugate wave obtained by taking the complex conjugate of the signal and changing the sign of the z dependence to get a wave traveling in the $+z$ direction. **The important thing to note about the complex conjugate wave is that it represents a**

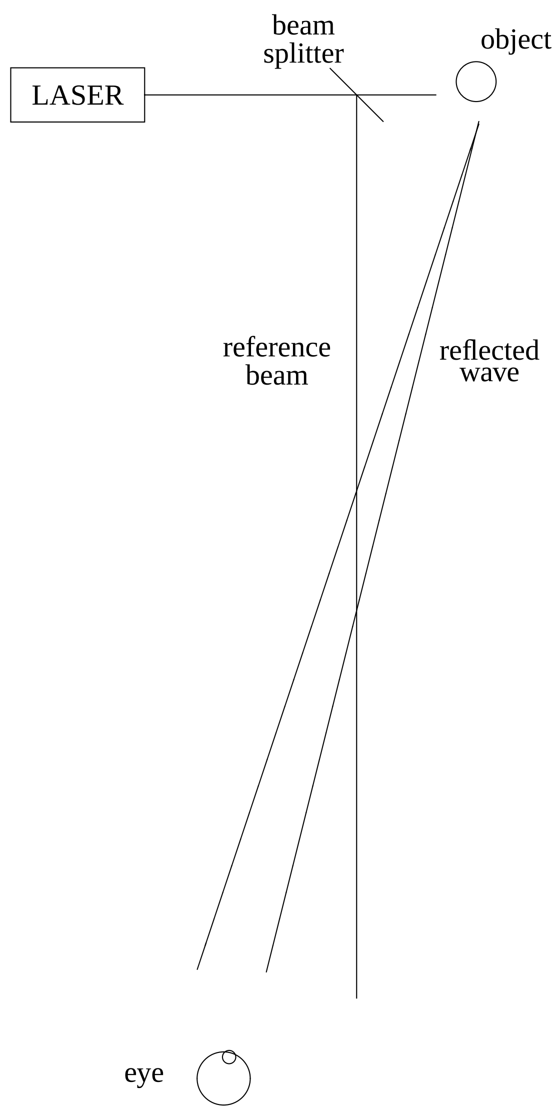


Figure 13.32: Viewing the object.

beam traveling in a different direction from either the signal or the reference beam, because the complex conjugation has changed the sign of k_x and k_y.

The resulting system is shown schematically in figure 13.33. Your eye sees a reconstructed version of the reflected wave that you would have seen without the photographic plate, as in (13.32). Note that neither the reference beam nor the complex conjugate beam get

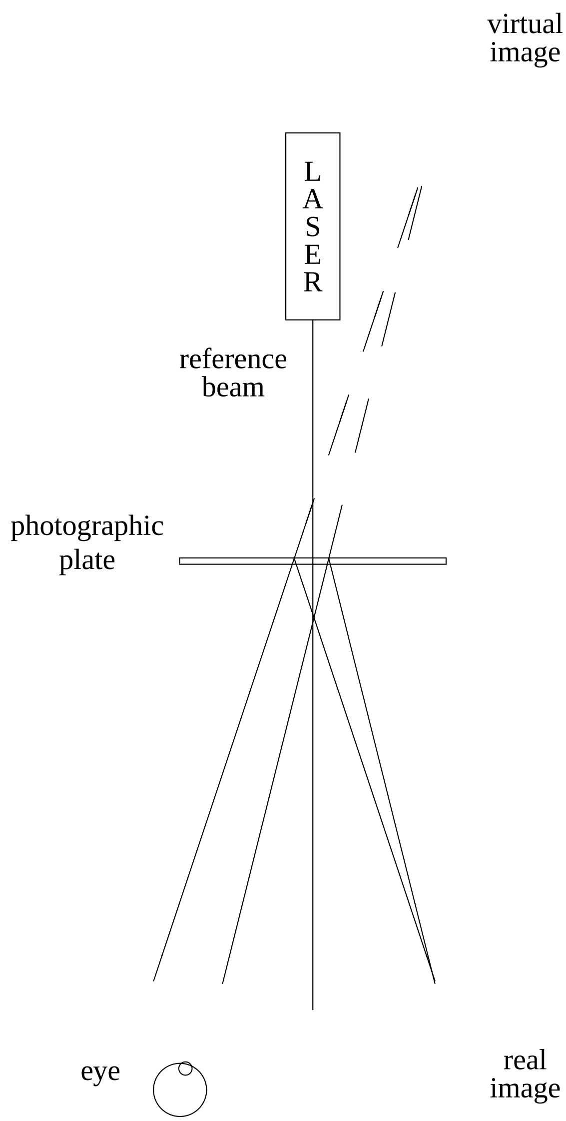


Figure 13.33: Viewing the holographic image.

in the way of your viewing, because they go off at slightly different angles. This is a hologram. Because it is not a picture but a reconstruction of the actual wave that you would have seen in (13.32), it has the surprising property of three-dimensionality that makes a hologram striking.

One might wonder why we choose the angle between the reference beam and the signal to be small. A large angle would have the advantage of getting the reference beam farther out of the way, but it would have an important disadvantage. Consider the intensity pattern on the photographic plate that records the hologram. It is an oscillating pattern with a typical wave number given by the typical value of k_x or k_y. These are of order $k\,\sin\theta$, where θ is the angle between the reference beam and the signal. But the distance between neighboring maxima on the photographic plate is therefore of order

$$\frac{2\pi}{k\,\sin\theta} = \frac{\lambda}{\sin\theta} \tag{13.131}$$

where λ is the wavelength of the light. Since λ is a very small distance, it pays to pick θ small to spread out the pattern on the photographic plate.

Note, also, that the order C^2 terms that we dropped really don't do any harm even if C is not small. Because their x and y dependence is proportional to that of the signal times its complex conjugate, the typical k_x and k_y for these terms is zero and they travel roughly in the direction of the reference beam. They don't reach your eye in (13.33).

13.9 Fringes and Zone Plates

13.9.1 The Holographic Image of a Point

One of the simplest of holographic images is the image of a single point. If a plane wave encounters a very small object in its path, the object will produce a spherical wave. If the plane wave and the spherical wave then are absorbed by a photographic plate, as shown in figure 13.34, an interference pattern is produced in the form of concentric circles, or fringes.

Specifically, suppose that the plane wave is propagating in the z direction, the photographic plate is in the x-y plane at $z = z_0$ and we put the origin of our coordinate system at the position of the source of the spherical wave, as shown in figure 13.34. Then the linear combination of plane wave plus spherical wave has the form (ignoring polarization)

$$Ae^{ikz} + \frac{B}{r}e^{ikr}\,, \tag{13.132}$$

where $r = \sqrt{x^2+y^2+z^2}$. We will assume, for simplicity, that A and B are real which means that the two waves are in phase at the object. The intensity of the wave at $z = z_0$, on the photographic plate is therefore

$$A^2 + \frac{B^2}{r_0^2} + \frac{2AB}{r_0}\cos[k(r_0 - z_0)] \tag{13.133}$$

where r_0 is the distance from the object for a point in the $z = z_0$ plane,

$$r_0 = \sqrt{z_0^2 + R^2} \tag{13.134}$$

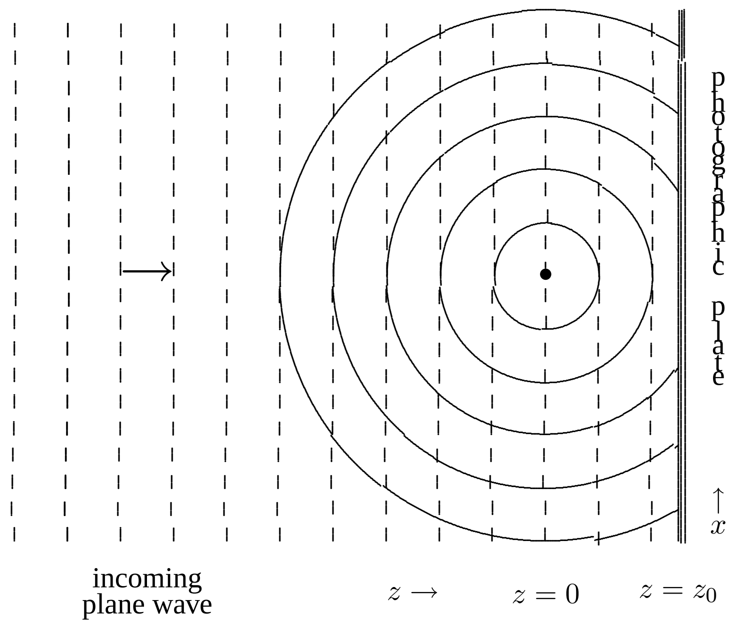


Figure 13.34: Fringes.

and

$$R = \sqrt{x^2 + y^2} \tag{13.135}$$

is the distance from the z axis in the x-y plane. The intensity depends only on R, as it must because of the symmetry of the system under rotations around the z axis.

Usually, we are interested in the region, $z_0 \gg R$, because, as we will see, the intensity pattern is most interesting for small R. In this region, the distance, r_0 is very nearly equal to z_0. We can ignore the variation of r_0 in the amplitude, B/r_0. However, there is interesting dependence in the cosine term in (13.133). In this term, we can expand r_0 in a Taylor series around $R = z_0$,

$$r_0 = z_0\sqrt{1 + R^2/z_0^2} = z_0 + \frac{1}{2}\frac{R^2}{z_0} + \cdots \,. \tag{13.136}$$

Putting all this together, the intensity is given approximately for $z_0 \gg R$ by

$$A^2 + \frac{B^2}{z_0^2} + \frac{2AB}{z_0}\cos\frac{kR^2}{2z_0} \,. \tag{13.137}$$

The intensity pattern, (13.137), describes concentric circular "zones" of intensity variation. The zones can be labeled by the maxima and minima of the cosine, at

$$\frac{kR^2}{2z_0} = n\pi \tag{13.138}$$

or

$$R^2 = n\lambda z_0 \tag{13.139}$$

where λ is the wavelength of the wave. For n even, the cosine has a maximum and for n odd, a minimum. The intensity variation is greatest if the plane wave and the spherical wave have approximately the same amplitude at the plate,

$$\frac{B}{z_0} = A\,. \tag{13.140}$$

Then the amplitude actually goes to zero at the minima. The intensity distribution as a function of R is shown in figure 13.35. The positions of the maxima and minima, or "zones," are shown on the R axis. On the photographic plate, this intensity distribution gives rise to circular fringes.

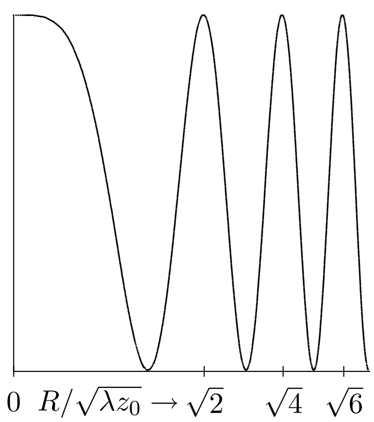


Figure 13.35: The intensity distribution.

If the plate is developed and illuminated by a plane wave, the original spherical wave is reproduced along with another spherical wave moving inward toward a point on the z axis a distance z_0 beyond the plate, as shown in figure 13.36. This wave is the real image of figure 13.33. When a plane wave (dotted lines) illuminates the photographic plate produced in figure 13.34, diverging (dotted lines) and converging (solid lines) spherical waves are produced.

13.9.2 Zone Plates

The hologram of figure 13.34 can be used to bring part of plane wave to a focus. The converging spherical wave shown in figure 13.36 is much stronger than the rest of the wave disturbance at the focus, $z = 2z_0$, $x = y = 0$, because the amplitude of this part of the wave

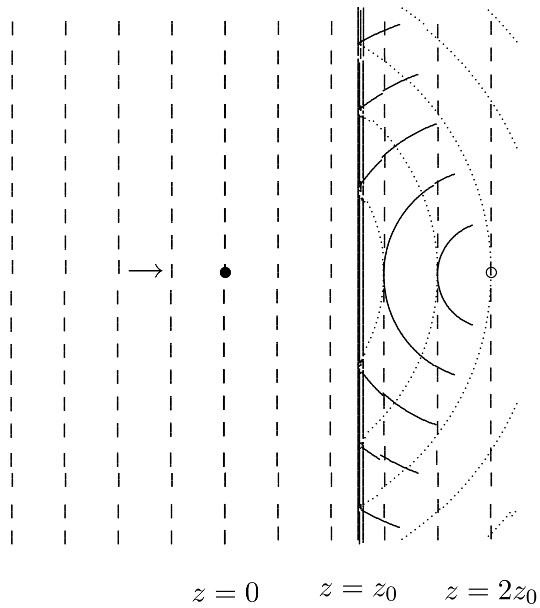


Figure 13.36: A plane wave illuminating the photographic plate.

increases as it approaches the focus. It has the form

$$\frac{1}{r'}e^{ikr'} \tag{13.141}$$

where

$$r' = \sqrt{(z - 2z_0)^2 + x^2 + y^2}\,. \tag{13.142}$$

The same effect can be produced with a cartoon version of the photographic plate made by taking a transparent plate and blacking out the zones for negative n in (13.138) where the intensity distribution is less than half the maximum. For example, the first negative zone is the region $\lambda z_0/2 < R^2 < 3\lambda z_0/2$. The second is the region $5\lambda z_0/2 < R^2 < 7\lambda z_0/2$, etc. The result is a "zone plate." An example, produced by blacking out the first 4 negative zones is shown in figure 13.37. These things are quite useful, because they can be easily produced and tailored to any wavelength.

Chapter Checklist

You should now be able to:

i. Be able to set up a diffraction problem as a forced oscillation problem and write the diffracted wave as a Fourier integral;

Figure 13.37: A zone plate.

ii. Interpret the Fourier integral in the far-field region and find the diffraction pattern;

iii. Analyze the diffraction patterns in beams made with one or more slits and rectangles;

iv. Use the convolution theorem to simplify the calculation of Fourier transforms;

v. Analyze the scattering from a diffraction grating and x-ray diffraction from crystals;

vi. Interpret a hologram as a diffraction pattern;

vii. Understand how a zone plate can focus a plane wave.

Problems

13.1. Consider the transverse oscillations of a semi-infinite, flexible membrane with surface tension T_S and surface mass density ρ_S. The membrane is stretched in the $z = 0$ plane from $y = -\infty$ to ∞ and from $x = 0$ to ∞. The membrane is held fixed along the half lines, $x = z = 0,\ a \le y \le \infty$ and $x = z = 0,\ -\infty \le y \le -a$. For y between a and $-a$, the membrane is driven with frequency ω so that the end at $x = 0$ moves with transverse displacement

$$\psi(0, y, t) = f(y)\, e^{-i\omega t}$$

where

$$f(y) = \begin{cases} b\left(1 - \dfrac{y}{a}\right) & \text{for } 0 \le y \le a \\ b\left(1 + \dfrac{y}{a}\right) & \text{for } -a \le y \le 0 \\ 0 & \text{for } |y| \ge a\,. \end{cases}$$

The transverse displacement is given by

$$\psi(x,y,t) = \int_{-\infty}^{\infty} dk_y\, C(k_y)\, e^{i(yk_y + xk(k_y) - \omega t)}$$

where $k(k_y)$ is some function of k_y and

$$C(k_y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dy\, f(y)\, e^{-ik_y y} = \frac{b}{\pi k_y^2 a}\,(1 - \cos k_y a)\,.$$

Find the function $k(k_y)$.

If the intensity of the wave at $x = L$, $y = 0$ for large L is I_0, find the intensity for $x = L$ and any value of y. **Hint:** Assume that you are in the far field region, and account for all the relevant factors contributing to the ratio of the intensity to I_0.

13.2. Consider an opaque barrier in the x-y plane at $z = 0$, with a single slit along the x axis of width $2a$, but with regions on either side of the slit each with width $2a$ which are partially transparent, designed to reduce the intensity by a factor of 2. When this barrier is illuminated by a plane wave in the z direction, the amplitude of the oscillating field at $z = 0$ is

$$f(x,y)\, e^{-i\omega t}$$

for

$$f(x,y) = \begin{cases} 1 & \text{for} \quad |y| < a \\ 1/\sqrt{2} & \text{for} \quad a < |y| < 3a \\ 0 & \text{for} \quad 3a < |y|\,. \end{cases}$$

Near the slit, this just produces a beam which is less intense by a factor of two on the edges. Far away, however, the diffraction pattern is quite different from that of the single slit. At a fixed large distance $R = \sqrt{y^2 + z^2}$ away from the slit, the intensity as a function of

$$\xi = k_y a = \frac{\omega\, y\, a}{c\, R}$$

is shown in the graph in figure 13.38 for positive ξ. The value of the peak at $\xi = 0$ is normalized to 1, but has been suppressed in the graph to show the details of the secondary maxima.

Find the smallest positive value of ξ for which the intensity vanishes.

Find the ratio of the intensity at $\xi = \pi/2$ to that at $\xi = 0$.

So far we have not mentioned the polarization of the light, assuming that it is irrelevant. In fact, we get the pattern shown above for any polarization, so long as the shading doesn't effect the polarization (and ξ is small). However, if the light is initially polarized in the direction

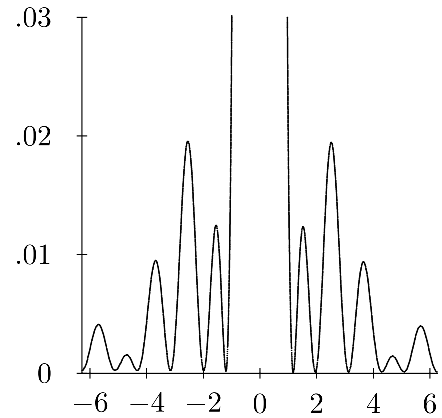


Figure 13.38: Problem 13.2.

45° from the x axis, we could reduce the intensity by two by passing it through a perfect polarizer aligned with the y axis. Suppose that our slit between $-a$ and a is completely empty, but between $-3a$ and $-a$ and between a and $3a$, we put such a polarizer. Now, as before, the beam close to the slit just has the intensity on the edges reduced by a factor of 2. Now, however, the diffraction pattern is quite different. As a function of ξ, the intensity at large fixed R is

$$\propto \frac{1}{10}\left[\left(\frac{\sin 3\xi}{\xi}\right)^2 + \left(\frac{\sin \xi}{\xi}\right)^2\right] .$$

which looks nothing like the pattern above. Explain the difference.

13.3. Consider an opaque barrier in the x-y plane at $z = 0$, with identical holes centered at $(x, y) = (n_x a, n_y a)$ for all integers n_x and n_y. Suppose that the barrier is illuminated from $z < 0$ by a plane wave traveling in the z direction with wavelength $\lambda = a\sqrt{3}/2$.

For $z > 0$, the wave has the form

$$\sum_{m_x,m_y} C_{m_x,m_y}\, e^{i(m_x\rho\, x + m_y\rho\, y + k_z(m_x,m_y)z - \omega t)}$$

where m_x and m_y run over all integers.

Find ρ.

For large z, only a finite number of terms in the sum are important. How many and how do you know?

Now suppose that instead of coming in the z direction, a plane wave with the same wavelength is moving for $z < 0$ at 45° to the z axis both in the x-z plane, and in the y-z plane. That is

$$\frac{k_x}{k_z} = \frac{k_y}{k_z} = \tan 45^\circ = 1 .$$

Now for $z > 0$, the wave has the form

$$\sum_{m_x,m_y} C_{m_x,m_y}\, e^{i[(m_x\rho+\xi_x)\,x+(m_y\rho+\xi_y)\,y+k_z(m_x,m_y)z-\omega t]}$$

where m_x and m_y run over all integers.

Find ξ_x and ξ_y.

Again for large z, only a finite number of terms in the sum are important. Which ones — that is, what values of m_x and m_y?

13.4. Describe the diffraction pattern that results when a transmission diffraction grating with line separation distance S is illuminated by a plane wave of monochromatic light with wavelength L that is traveling in a direction perpendicular to the grating lines and at an angle θ to the perpendicular from the surface of the grating.

13.5. An opaque screen with four narrow slits at x=± 0.6 mm and x=± 0.4 mm is blocking a beam of coherent light with wavelength 4×10^{-5} cm. Describe the diffraction pattern that appears on a screen 5 meters away.

13.6. A semi-infinite flexible membrane is stretched in the $z = 0$ plane for $x \geq 0$ with surface tension T_s and surface mass density ρ_s. The membrane is clamped down at $z = 0$ along the two semi-infinite lines, $z = 0$, $x = 0$, $y \geq a$ and $z = 0$, $x = 0$, $y \leq -a$. For $-a \leq y \leq a$ and $x = 0$, the membrane is forced to oscillate with an amplitude of the form

$$z = B\, e^{i\omega t} \cos \frac{\pi y}{2a}\,.$$

Draw a diagram of the $z = 0$ half plane for $x \geq 0$ and indicate where the average of the absolute value square of the transverse displacement of the membrane is large (i.e. not much smaller than $B^2 a/r$, where r is the distance from the origin). For your diagram, assume that the distance a is about 5 times the wavelength of the waves.

Find the intensity of the disturbance on the membrane produced by this forced oscillation as a function of $\theta = \tan^{-1}(y/x)$ on a large semicircle, $x^2 + y^2 = R^2$, for $R^2 >> a^4\, \omega^2\, \rho_s/T_s$.

Hint: This is similar to a single slit diffraction problem. Note that even though the disturbance is a cosine, you will have to do a Fourier integral (although not a difficult one) to do part b, because the disturbance is confined to $-a \leq y \leq a$ at $x = 0$.

13.7. Suppose that a diffraction grating with line separation d is etched onto the top of a thick piece of glass with index of refraction n. If light of frequency ω is incident on the top, coming in at an angle θ from the perpendicular to the face and perpendicular to the grating lines, find the angles of the components of the wave in the glass.

13.8. Shown in figure 13.39 are 4 diffraction patterns such as might be produced by shining laser light (nearly a plane wave) through a slit or slits, and projecting the pattern onto a photographic plate far away. The patterns are each produced by about 500 individual photons striking the plate with a probability density proportional to the intensity of the diffracted wave.

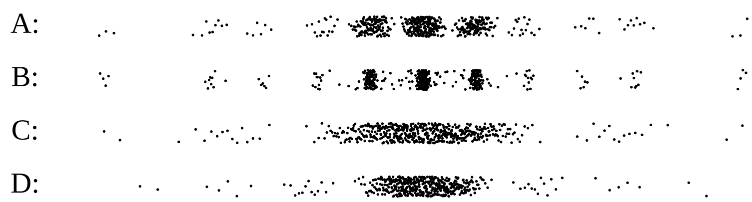


Figure 13.39: Four diffraction patterns.

The four objects that produced these patterns were, in a random order,

i. A single slit, 1 mm wide;

ii. A single slit, 0.6 mm wide;

iii. Two slits, each 0.6 mm wide, with centers 1.5 mm apart;

iv. Six slits, each 0.6 mm wide, with adjacent centers 1.5 mm apart.

a. Which is which?

b. How do you know?

Chapter 14

Shocks and Wakes

In this chapter, we apply the tools of the previous chapter to analyze some beautiful and interesting phenomena — shock waves and the Kelvin boat wake.

Preview

The

i. ****

14.1 * Boat Wakes

Combining our analysis of the dispersion relation of water and the discussion of section 11.5 and of group velocity in section 10.2.1 will allow us to give a simple interpretation of one of the most beautiful and subtle of all wave phenomena — the Kelvin wake.

14.1.1 Wakes

The general subject of wakes is very complicated. However, in the simplest case of motion with constant velocity, the symmetry of the system makes it possible to do a linear analysis rather simply. This will allow us to understand some of the most obvious features of the wake in a straightforward way, including the angle and the regular waves that appear along the wake, showing up at sunset like pearls on a string, as in figure 14.2.[2] While I do not think that there is much that is original in the treatment here, the approach is a little unusual,

[2]This is one of many beautiful photographs by Ian Alexander - http://easyweb.easynet.co.uk/ iany/patterns/wake.htm.

Figure 14.1: Wakes at sea.[1]

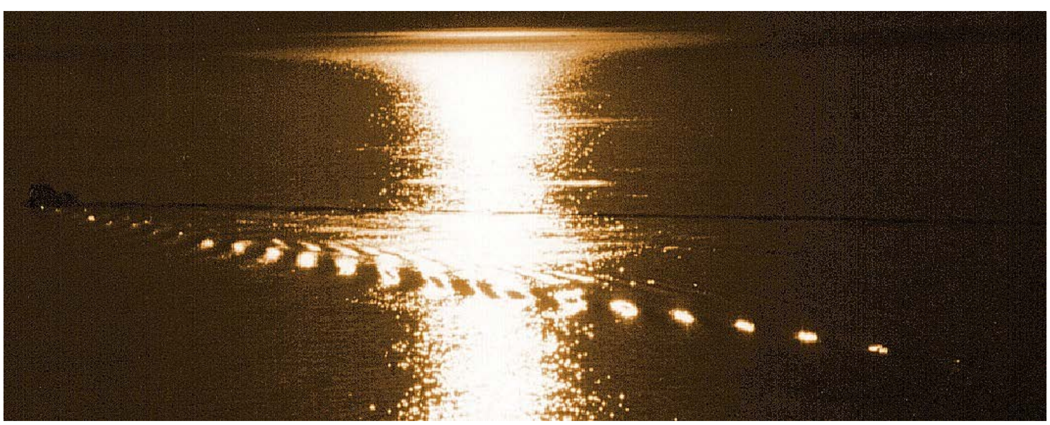


Figure 14.2: A wake at sunset.

like the approach to diffraction in chapter 13), and I think it gives a useful slant not just on the wake, but also on the crucial concept of group velocity.

14.1.2 Linear analysis of the Kelvin wake

Consider an infinite ocean in the x-y plane with a boat (duck, whatever) moving with constant velocity $v > 0$ along the x axis. The path of the boat divides the surface of the ocean into two regions, related to one another by reflection in the path. We can therefore without loss of generality focus on the half-plane $y > 0$. We will not try to describe in detail what goes on near the $y = 0$ line. In many situations, this involves turbulance, and is well beyond the scope of a beginning waves course. But away from $y = 0$, it is possible to apply a linear analysis, and think of the waves for $y > 0$ as linear combinations of plane waves with appropriate boundary conditions along the $y = 0$ line.[3] We will assume that whatever happens near $y = 0$ produces a localized disurbance on the y axis that moves along with the boat. This is necessarily a wave packet involving a range of frequencies. The integration over all these frequencies gives rise to the wake. That is the plan. We will do this for some simple illustrative boundary conditions, and we will argue that much can be understood about the system that is independent of the details of the boundary condition. The idea is to make use of the fact that the disturbance is a wave packet, and understand the appropriate analog in this two dimensional situation of the group velocity by which wave packets move.

This system is invariant under simultaneous translations in space and time.

$$t \to t + \tau \qquad x \to x + v\,\tau \tag{14.1}$$

We are looking for a steady-state solution in which the only things going on are the waves induced by the motion of the boat. This solution should depend only on the combination

$$x - v\,t \tag{14.2}$$

and be invariant under (14.1). This means that the wake wave is stationary in a frame of reference moving along with the boat.

An equivalent (and perhaps more physical) way to say this is that the solution that describes only the waves produced by the moving boat is a linear combination of plane waves which are moving along with the boat in the x direction — that is they have $k_x = \omega/v$.

Either way, the general solution looks like

$$\int d\omega\, f(\omega)\, e^{-i[\omega(t-x/v)-k_y y]} \tag{14.3}$$

[3] One could, if necessary, look only at $y > a > 0$ for some fixed a. This would not change the analysis in any essential way.

The function $f(\omega)$ describes the wave packet in frequency, and it is determined by the displacement of the wave packet at $y = 0$ and $t = 0$, through the Fourier transform

$$\psi(x) = \int d\omega\, f(\omega)\, e^{i\omega x/v} \Rightarrow f(\omega) = \frac{1}{2\pi} \int d\tau\, \psi(v\tau) e^{-i\omega\tau} \tag{14.4}$$

The relation between ω and

$$k = \sqrt{k_x^2 + k_y^2} \tag{14.5}$$

is given by the dispersion relation for water waves. It will turn out that it is usually a good approximation to ignore surface tension, so will use a simple approximation in which the dispersion relation depends only on gravity. We will also assume that the water is deep. Then the dispersion relation is simply

$$\omega^2 = g\, k \tag{14.6}$$

The magnitude of the phase velocity is thus

$$v_\phi = \frac{\omega}{k} = \frac{g}{\omega} \tag{14.7}$$

Note that coefficient of the right hand side in the dispersion depends on the details of the physics, but the overall structure of the relation follows simply from dimensional analysis. The only combination of g and k with units of ω^2 is $g\, k$.

For (14.3), we have

$$k = \left(k_x^2 + k_y^2\right)^{1/2} \tag{14.8}$$

with

$$k_x = \omega/v \tag{14.9}$$

Thus

$$\omega^4 = g^2(\omega^2/v^2 + k_y^2) \tag{14.10}$$

$$k_y = \left(\omega^4/g^2 - \omega^2/v^2\right)^{1/2} \tag{14.11}$$

where the sign of the square-root is determined by the boundary condition at $y = +\infty$. If k_y is real, k_y/ω is a positive number so that the phase waves in (14.3) propagate out from the $y = 0$ line. If k_y is imaginary, it is i times a positive number, so that that the amplitude vanishes as $y \to +\infty$ for $y > 0$. These signs are opposite for $y < 0$, but nothing else in the analysis changes, so the solution is symmetrical about $y = 0$, as it must be.

Thus for $y > 0$, we have

$$\vec{k} = (k_x, k_y) = \left(\omega/v, \left(\omega^4/g^2 - \omega^2/v^2\right)^{1/2}\right) \tag{14.12}$$

and

$$\frac{\vec{k}}{\omega} = \frac{\hat{v}_\phi}{v_\phi} = \left(1/v, \left(\omega^2/g^2 - 1/v^2\right)^{1/2}\right) \tag{14.13}$$

This is the key to the Kelvin wake. The different frequency components of the wave packet at $y = 0$ travel in different directions as they move away from the boat for $y > 0$.

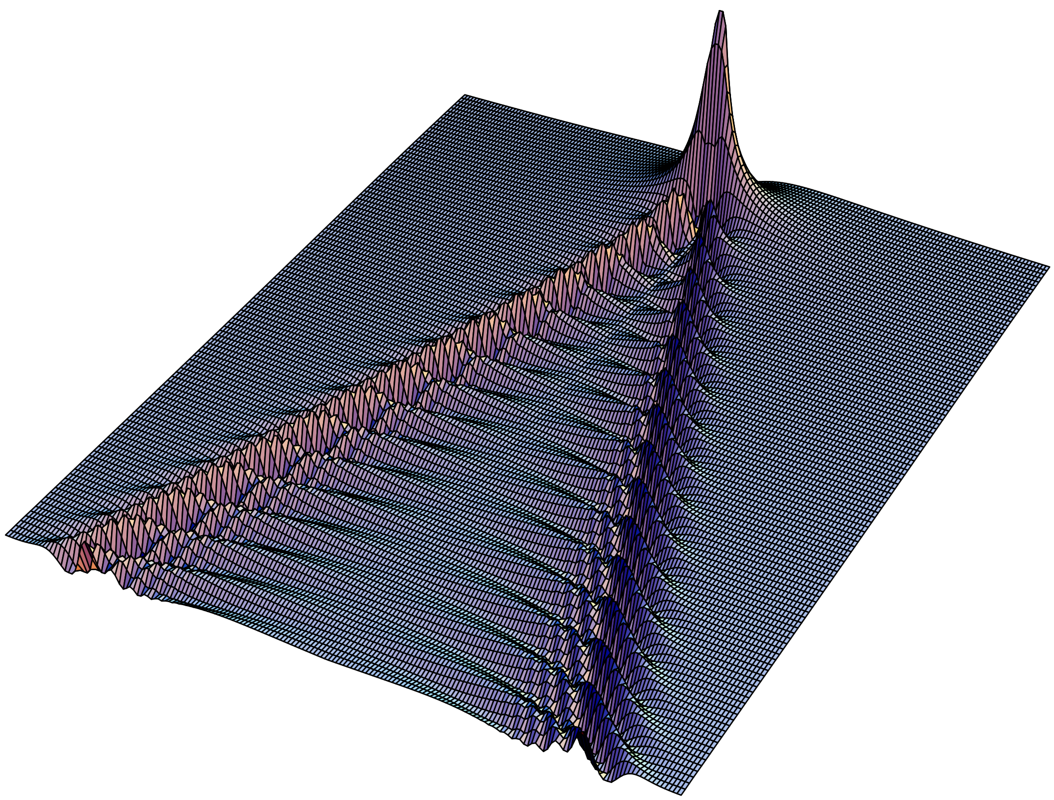


Figure 14.3: A wake constructed numerically from (14.3) for $f(\omega) = e^{-\omega^2 v/g}$.

We note in passing that the vector quantity $\vec{k}/\omega$ is a very important one, and it should probably have some important sounding name. It is sometimes called the “slowness vector” in the literature. This is the object that most naturally appears in the description of a plane

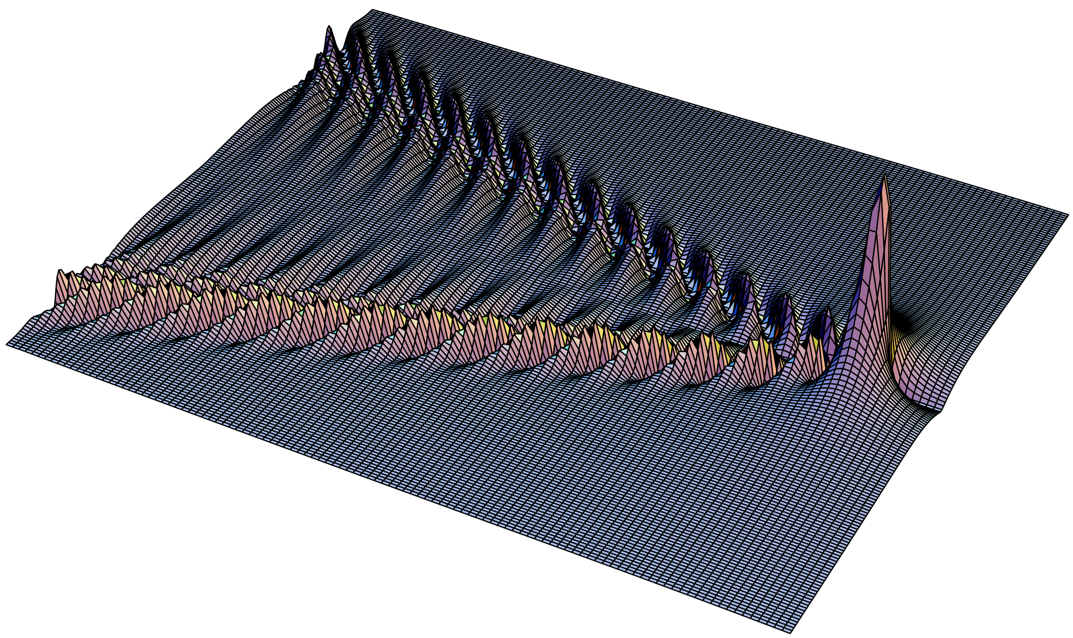


Figure 14.4: Another view of the wake in figure 14.3.

wave with frequency ω:

$$e^{-i\omega(t-\vec{r}\cdot\vec{k}/\omega)} \tag{14.14}$$

"Slowness vector" does not seem to me to capture adequately the relation of this with the phase velocity, so I plan to start a campaign to call $\vec{k}/\omega$ the "phase segnocity"

$$\vec{s}_\phi = \frac{\vec{k}}{\omega} \tag{14.15}$$

from the Latin *segnis* meaning slow, because the segnocity increases in magnitude as the points of constant phase on the wave move more slowly.[4] The phase velocity itself can be easily constructed from the phase segnocity, (14.13), but it is really (14.13) that appears naturally in the equations.

We can calculate $f(\omega)$ given a boundary condition at $t = y = 0$, and then construct the wake numerically using (14.3). For illustration, consider a Gaussian $f(\omega)$,

$$f(\omega) = e^{-\omega^2 v/g} \tag{14.16}$$

[4]The inspiration for this comes from paleontologists, who have fossil evidence for the *segnosaurus* - "slow lizard" and also the *velocisaurus* - "fast lizard."

which corresponds to a gaussian disturbance along $y = 0$. The resulting wake pattern is shown in figures 14.3 and 14.4. This looks like a wake! Thus we seem to have captured some of the essential physics of the wake in (14.3).

But what we are really interested in are the more general properties that will allow us to understand the key features of the wake without any numerical integration. We will look at how the various plane-wave components propagate. (14.11) implies that the wake waves propagate away from $y = 0$ only for

$$\omega^2 > g^2/v^2 \tag{14.17}$$

We now need to think carefully about the physics of (14.3) and (14.12)-(14.13). In any small range of ω, the integration over ω will produce some kind of wave packet that moves along with the boat. Different ω ranges are associated with wave packets moving out at different angles from the x axis. **The envelope of a wave packet in a narrow range of frequency centered on ω will move at the effective group velocity, v_g, constructed as follows**

$$\vec{s}_g = \frac{\partial \vec{k}}{\partial \omega} = \frac{\hat{v}_g}{|\vec{v}_g|} = \left(\frac{1}{v}, \frac{2\omega^3/g^2 - \omega/v^2}{(\omega^4/g^2 - \omega^2/v^2)^{1/2}} \right) \tag{14.18}$$

and this determines the angle of the wave packet as a function of ω. Again, as with (14.13), the object in (14.18) is **very important** and deserves a fancy name. I am going to call it the "group segnocity," $\vec{s}_g$, because the envelope of the wave group moves more slowly as $\vec{s}_g$ increases. Suggestions for better names for $\vec{s}_\phi$ and $\vec{s}_g$ are welcome.

At this point, the reader may (justifiably) wonder why the group velocity is not the conventional group velocity for water waves,

$$V_g = \frac{\partial \omega}{\partial k} \tag{14.19}$$

where the relation between ω and k is given by (14.6) so that

$$V_g = \frac{g}{2\omega} = \frac{v_\phi}{2} \tag{14.20}$$

This gives a group velocity in the same direction as the phase velocity and just half the magnitude. To understand the difference, we must generalize the the formula for group velocity in section 10.2.1. There we saw that simplest way to understand group velocity is to think about the superposition of two plane waves that are close together in both ω and $\vec{k}$

$$\begin{aligned} &\cos(\omega_1 t - \vec{k}_1 \cdot \vec{r}) + \cos(\omega_2 t - \vec{k}_2 \cdot \vec{r}) \\ &= 2\cos\left(\frac{\omega_1 - \omega_2}{2} t - \frac{\vec{k}_1 - \vec{k}_2}{2} \cdot \vec{r}\right) \cos\left(\frac{\omega_1 + \omega_2}{2} t - \frac{\vec{k}_1 + \vec{k}_2}{2} \cdot \vec{r}\right) \end{aligned} \tag{14.21}$$

If the ωs and $\vec{k}$s are close together, the first factor is a slowly oscillating envelope, like the envelope of the wave packet, with segnocity vector

$$\vec{s}_- = \frac{\vec{k}_1 - \vec{k}_2}{\omega_1 - \omega_2} \tag{14.22}$$

The second factor describes the more rapidly oscillating carrier wave, with segnocity vector

$$\vec{s}_+ = \frac{\vec{k}_1 + \vec{k}_2}{\omega_1 + \omega_2} \tag{14.23}$$

If $\vec{k}_1$ and $\vec{k}_2$ arise from a well defined smooth function $\vec{k}(\omega)$,

$$\vec{k}_1 = \vec{k}(\omega_1) \quad \text{and} \quad \vec{k}_2 = \vec{k}(\omega_2) \tag{14.24}$$

then

$$\lim_{\omega_1,\,\omega_2 \to \omega} \vec{s}_- = \vec{s}_g \quad \text{and} \quad \lim_{\omega_1,\,\omega_2 \to \omega} \vec{s}_+ = \vec{s}_\phi \tag{14.25}$$

as expected from (14.18). But this depends on what is being held fixed between $\vec{k}_1$ and $\vec{k}_2$ as ω changes. The point is the usual one that in one dimension, the dispersion relation determines $k(\omega)$ up to a sign. But in more than one dimension, infinitely many $\vec{k}$s satisfy the dispersion relation for a given ω. We must specify exactly how the function $\vec{k}(\omega)$ is determined before the limit in (14.25) is well defined. Thus (14.18) is the general formula for the group velocity. But what the derivative means (in more than one dimension) depends on the situation. The conventional physics assumes that all the frequency components of the wave are in the same direction. Then the direction of the $\vec{k}$ vector does not change with ω, and (14.18) reduces to (14.19).[5] But for a wave driven by the moving boat, what is held fixed is not the direction of the waves, but rather the x component of the segnocity vector, k_x/ω, because all the components of the wave packet move along with the boat that produces them. Then the condition (14.9) and the dispersion relation (14.6) taken together force the direction of the $\vec{k}$ vector to change as a function of ω. Thus we must use the general form, (14.18). As we will see, this implies that the effective group velocity not only has a very different magnitude from the conventional (14.20), but also that the effective group velocity and the phase velocity are not even in the same direction.

The red lines in figure 14.5 show the position of the wave packet corresponding to $\omega = 1.05\,g/v$. The cyan vectors show the group velocities for $y > 0$ and $y < 0$. These are perpendicular to the lines of the wave packets. If we take the origin to be the position of the boat (either instantaneously or by going to a moving coordinate system in which the boat is stationary at the origin), The points $\vec{r}$ on the wave packet satisfy

$$\vec{r} \cdot \hat{v}_g = 0 \tag{14.26}$$

[5]It looks different because in (14.18) we are looking at the group segnocity rather than the group velocity, but it is an elementary exercise in calculus to get from one to the other.

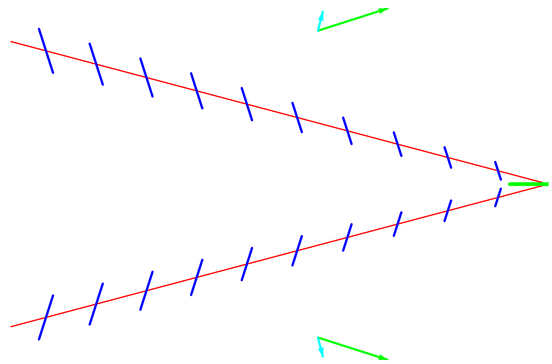


Figure 14.5: Wave packets (red) and phase waves (blue) for $\omega \approx 1.05\, g/v$.

Note also that because of (14.18),

$$\vec{r} \cdot \frac{\partial \vec{k}}{\partial \omega} = 0 \tag{14.27}$$

But there is additional structure within the wave packet because of the phase waves. The phase velocities are shown by the green vectors, and are not in the same direction as the group velocity. Thus there are oscillations along the wave packet corresponding to maxima and minima of the phase waves. Assuming that the maximum occurs at the position of the boat (that we are taking to be the origin), the other maxima occur at

$$\vec{r}_j \cdot \vec{k} = -2\pi j \quad \text{for } j = 1 \text{ to } \infty \tag{14.28}$$

These maxima represent the interaction of the plane wave with frequency ω (and moving along with the boat) and the group wave packet, so we indicate them in the figure by blue lines of constant phase perpendicular to $\vec{k}$.

Now we can build up a picture of the wake by putting these together for the range of important ω. As ω changes, the direction of the red wave packet lines and the blue phase waves will change according to (14.27) and (14.12). For $\omega \approx g/v$, the wave packets are nearly horizontal and the phase waves are moving nearly horizontally. This is shown in figure 14.6 for $\omega = 1.001\, g/v$: As ω increases, the angle of the wave packets increases for a while, giving a situation like that shown in figure 14.5. But the interesting thing is that there is a critical value of ω that gives the maximum angle for the group velocity.

It is particularly simple to analyze (14.18) because $\partial k_x/\partial\omega$ is constant. The ω dependence of $\partial k_y/\partial\omega$ is shown in figure 14.7 Because it goes to infinity as $\omega \to g/v$ and as $\omega \to \infty$, in these two limits the group velocity goes to zero, and its direction goes to $\hat{y}$. The minimum of $\partial k_y/\partial\omega$ corresponds to the maximum of $|\vec{v}_g|$, and also to the maximum angle of the wake wave propagation from the y direction. This maximum angle is particularly important for two reasons. Not only does the maximum angle correspond to the edge of the

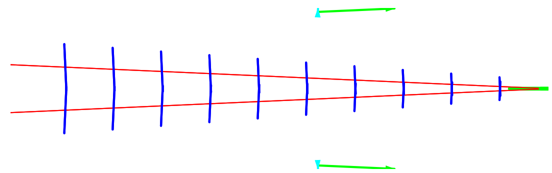


Figure 14.6: Wave packets (red) and phase waves (blue) for $\omega \approx 1.001\, g/v$.

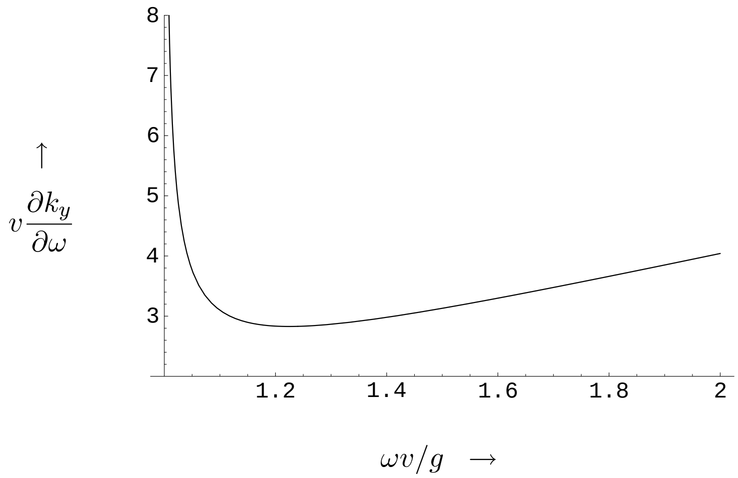


Figure 14.7:

disturbance produced by the boat, but more importantly, at all smaller angles, the energy is spread over neighboring angles. At the maximum, because the angle is stationary with respect ω, neighboring ω contribute constructively and the wave at the edge is generically much more intense than for smaller angles.[6]

To determine the maximum angle explicitly, we differentiate $\partial k_y/\partial\omega$ again, which gives

$$\frac{\partial^2 k_y}{\partial \omega^2} = \frac{(\omega^2/g^2)(2\omega^4/g^2 - 3\omega^2/v^2)}{(\omega^4/g^2 - \omega^2/v^2)^{3/2}} \tag{14.29}$$

so the minimum occurs at

$$\omega^2 = \frac{3g^2}{2v^2} \tag{14.30}$$

corresponding to wave number

$$k = \frac{3g}{2v^2} \tag{14.31}$$

[6]This pile-up at a stationary point is the same phenomenon that picks out the angle at which we see the rainbow.

and magnitude of phase velocity

$$v_\phi = \sqrt{\frac{2}{3}}\, v \tag{14.32}$$

The direction of the phase velocity is is

$$\hat{v}_\phi = \left(\sqrt{\frac{2}{3}}, \sqrt{\frac{1}{3}}\right) \tag{14.33}$$

at an angle from the x axis of

$$\theta_{max} = \arcsin\left(1/\sqrt{3}\right) = 35.26^\circ \tag{14.34}$$

The wavelength is

$$\lambda = \frac{4\pi v^2}{3g} \tag{14.35}$$

Even at low speeds like a meter per second (about 2 knots), this wavelength is large compared to the scale (a few centimeters) at which surface tension becomes important in the dispersion relation, so (14.6) is usually a good approximation.

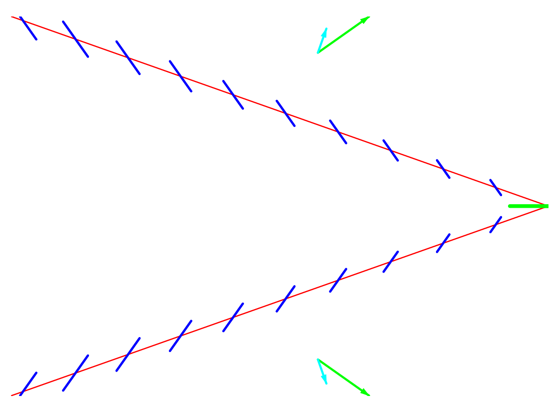


Figure 14.8: Wave packets (red) and phase waves (blue) for the critical value $\omega \approx \sqrt{3/2}\, g/v$.

At the critical angle, our wave packets and phase waves are shown in figure 14.8. Often, as in the wakes in figure 14.1, the phase waves along the maximum angle are all you see.

For ω larger than $\sqrt{3/2}g/v$, the wave packet angles starts to decrease again but the phase waves continue to get closer together, as shown in figure 14.9 for $\omega = 2\, g/v$.

We can now put these together for a range of ω from g/v up past the critical value. The result is shown in figure 14.10. It may be easier to see what is happening if we do not draw the wave packets, but just the phase waves. These, after all, show the places where the oscillation is a maximum. The result is shown in figure 14.11. This gives nearly the same

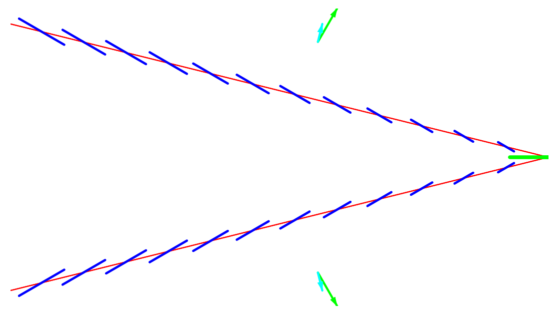


Figure 14.9: Wave packets (red) and phase waves (blue) for $\omega \approx 2\,g/v$.

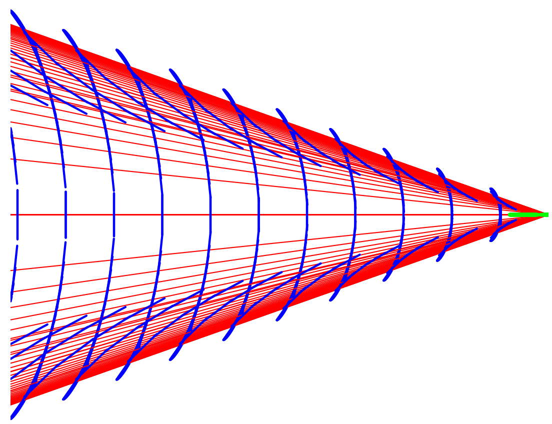


Figure 14.10: Combined wave packets and phase waves for a range of frequencies.

picture as a parametric plot of the points on the waves packets where the phase waves have there maxima, which is shown in figure 14.12. You should be able to recognize the basic features of figure 14.11 both in our numerical construction, figure 14.3 and in the beautiful picture of a real wake from Wikipedia shown in figure 14.13 on page 436. We can make this even more obvious by rotating figure 14.12 in three dimensions, as shown in figure 14.14. In the wake in figure 14.13, the phase waves are clearly visible both for small ω (the forward-moving waves in the center) and for large ω (the outward-moving and closely spaced waves just inside the maximum angle). From figure 14.11 and 14.13 you should be able to see why the phase waves are sometimes called "featherlet waves." The phase waves for $\omega > \sqrt{3/2}g/v$ look like delicate feathers on the wing of the wake.

It is worth making one more comment about figure 14.11. You have probably already noticed that the phase waves fit together into continuous curves. This is not an accident. Physically, of course, the position at which the oscillation is maximum must certainly vary

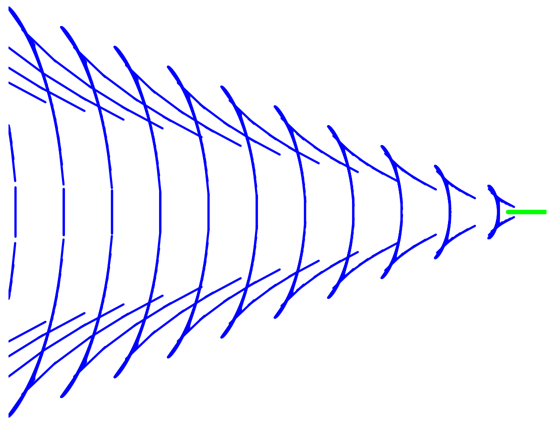


Figure 14.11: Combined phase waves for a range of frequencies.

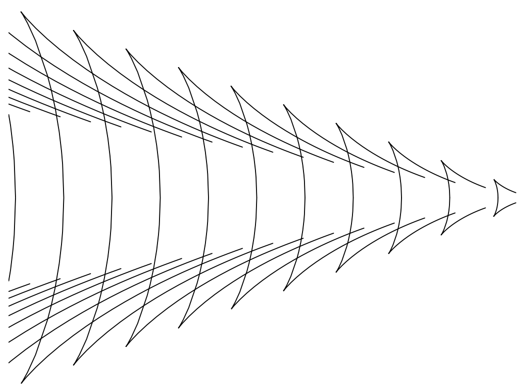


Figure 14.12: Parametric plot of the phase maxima.

continuously as a function of ω. But also, we can see that the tangent to the curve that describes the maxima is perpendicular to $\vec{k}$, and thus our phase waves in the figures are just these tangents. To see this, note that from (14.27) and (14.28), $\vec{r}_j$ satisfies

$$\vec{r}_j \cdot \vec{k} = -2\pi j \quad \text{and} \quad \vec{r}_j \cdot \frac{\partial \vec{k}}{\partial \omega} \tag{14.36}$$

Differentiating the first with respect to ω and using the second gives

$$\frac{\partial \vec{r}_j}{\partial \omega} \cdot \vec{k} = 0 \tag{14.37}$$

Figure 14.13: Compare figures 14.3 , 14.11, 14.12 and 14.14 with a real wake.

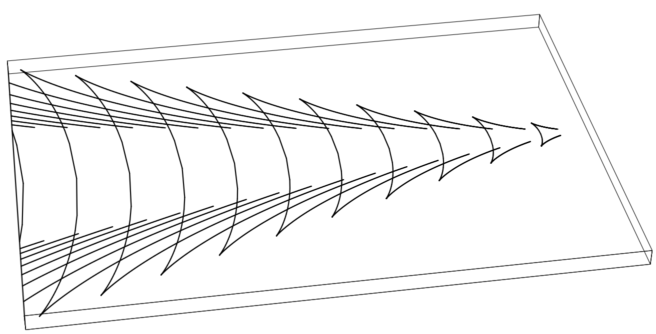


Figure 14.14: 3D Parametric plot of the phase maxima.

which says that the tangent to the curve described parametrically by $\vec{r}_j(\omega)$ is perpendicular to $\vec{k}$ and thus parallel to the phase waves.[7]

14.1.3 Shocks versus Wakes

It is interesting to compare the above analysis of the Kelvin wake with the creation of a shock wave by motion faster than the propagation speed in a dispersionless medium. For simplicity of comparison, consider a 2-dimensional system like a flexible membrane. Again, we will put the system in the x-y plane and consider the effect of an object moving with constant velocity v along the x axis. The 3-dimensional extension is straightforward. Once again, the waves associated with the object have $k_x = \omega/v$ and are described by (14.3). This time, however,

$$k^2 = k_x^2 + k_y^2 = \omega^2/v_0^2 \tag{14.38}$$

where v_0 is the speed of transverse waves in the membrane. Therefore k_y is also proportional to ω and the only issue is whether $v < v_0$, in which case k_y is imaginary and all propagation is in the x direction, or $v > 0$, in which case k_y is real and there is a propagating shockwave. In either case, the group velocity and the phase velocity are equal because the medium is dispersionless.

Let's see what (14.3) looks like in this case for $v > v_0$. Then

$$k_y = \pm\omega\sqrt{1/v_0^2 - 1/v^2} \tag{14.39}$$

The segnocity vector, $\vec{k}/\omega$ has the form

$$\vec{k}/\omega = \hat{v}_\phi/v_0 = \begin{cases} \left(1/v, \sqrt{1/v_0^2 - 1/v^2}\right) & \text{for } y > 0 \\ \left(1/v, -\sqrt{1/v_0^2 - 1/v^2}\right) & \text{for } y < 0 \end{cases} \tag{14.40}$$

For $v > v_0$ we do not even need to do a Fourier transform. We can just write down the answer and check that it works. If the displacement on the x axis is

$$f(t - x/v) \tag{14.41}$$

then the displacement in general is

$$f(t - \vec{r}\cdot\hat{v}_\phi/v_0) \tag{14.42}$$

Thus in contrast to a wake, a shock wave is completely boring. The pattern on the $y = 0$ axis just gets progagated without change in shape. It looks like the figure in figure 14.15 for a Gaussian envelope and $v = 5v_0/3$.

[7]There is one caveat worth noting here. In (14.28), we assumed that the phases of the different components of the wave relative to the boat do not vary with ω. This is just an approximation, and it might fail seriously for long (compared to v^2/g) boats.

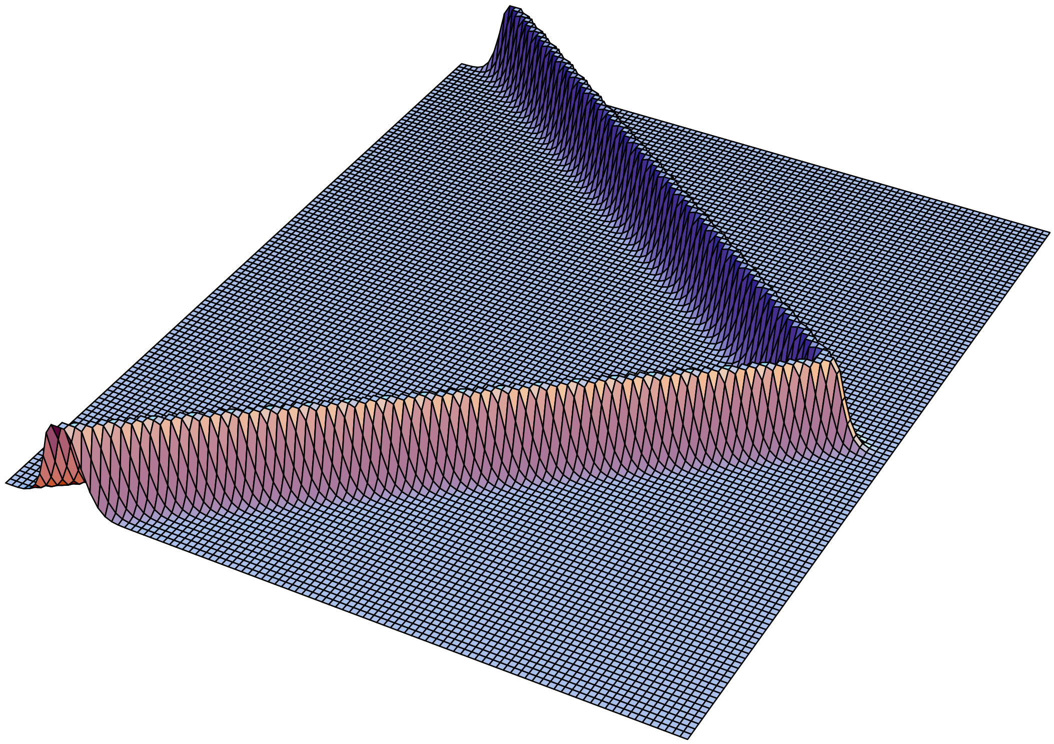


Figure 14.15: A shock wave.

14.2 Chapter Checklist

You should now be able to:

i. ****

Problems

Bibliography

[1] F. S. Crawford, Jr., **Waves**, Berkeley Physics Course, Volume 3 (McGraw-Hill: New York, 1968).

[2] E. Hecht, **Optics**, Second Edition (Addison-Wesley: Reading, Mass, 1987).

[3] E. M. Purcell, **Electricity and Magnetism** (McGraw-Hill: New York, 1985).

[4] R. Resnick and D. Halliday, **Physics**, Third Edition (John Wiley & Sons: New York, 1977).

[5] F. W. Sears, M. W. Zemansky and H. D. Young, **University Physics**, Sixth Edition (Addison-Wesley: Reading, Mass, 1981).

Appendix A

The Programs

The programs are simple moving illustrations, except for a few that are interactive and allow you to play with the concepts discussed in the text. They are available, along with many other animations, on my web page, but they need some updating. The best version is ALLSLOW.EXE - which is linked from Physics of Waves page.

Running ALLSLOW.EXE in a DOS window gives the master menu. You can access any of the programs from the master menu by typing the program number. To return to the menu from any program, touch the function key F10.

The PC programs are written in POWER BASIC by Spectra Publishing. A program works by drawing pictures on an invisible screen while you look at the visible screen. Then the program switches the two screens and repeats the process. This allows animation of nonharmonic motion in a simple way.

Description of programs

1-1 — pages 9-11 — The connection between harmonic motion and uniform circular motion.

1-2 — pages 14-17 — Multiplication in the complex plane. Move the complex number z around in the complex plane with the arrow keys.

2-1 — pages 48-49 — A damped forced harmonic oscillator with one degree of freedom. The driving frequency can be adjusted up or down (with the up or down arrow keys).

3-1 — pages 57-77 — Two coupled pendulums.

4-1 — pages 98-99 — Beats in two coupled pendulums.

4-2 — pages 99-103 — The modes of the hacksaw oscillator. This programs shows a complicated motion of the oscillator, but the individual modes can be seen.

You see the modes by pressing the function keys, as follows:

F1 the mode $A^1 + A^5$;

F2 the mode $A^2 + A^4$;

F3 the mode A^3;

F4 the mode $A^2 - A^4$;

F5 the mode $A^1 - A^5$;

F6 the mode A^0;

F9 all modes — full motion;

F10 quits the program.

5-1 — pages 113-114 — Standing waves in a system of coupled pendulums with fixed ends.

5-2 — pages 119-121 — Standing waves on a beaded string with fixed ends.

5-3 — pages 122-124 — Standing waves on a beaded string with free ends.

6-1 — pages 141-142 — Normal modes of the continuous string with fixed ends, with $k = n\pi/L$ for $n = 1$ to ∞. The up and down arrow keys increase n.

6-2 — pages 142-144 — Normal modes of the continuous string with one fixed end and one free end, with $k = n\pi/L - \pi/2L$ for $n = 1$ to ∞. The up and down arrow keys increase n.

6-3 — pages 144-147 — The Fourier series for the function of (6.19),

$$\psi(x) = \begin{cases} x & x \le w \\ \dfrac{w(1-x)}{1-w} & x > w\,. \end{cases} \tag{A.1}$$

6-4 — pages 148-148 — Plucking an ideal string.

6-5 — pages 148-148 — Same program as 6-4, but with variable inputs.

7-1 — pages 155-156 — Longitudinal modes of a continuous spring with fixed ends. F1-F9 give modes 1-9. F10 quits.

7-2 — pages 156-157 — Longitudinal modes of a continuous spring with one fixed end and one free end. F1-F9 give modes 1-9. F10 quits.

8-1 — pages 172-172 — A traveling wave with a circle moving along the maximum of the wave at the phase velocity.

8-2 — pages 173-174 — A traveling wave built out of two standing waves.

8-3 — pages 192-192 — A traveling wave with damping. It peters out as it travels.

8-4 — pages 192-193 — A forced oscillation problem for a continuous string with damping and one end fixed.

8-5 — pages 192-193 — A forced oscillation problem for a beaded string with damping and one end fixed.

8-6 — pages 193-197 — High- and low-frequency cut-offs in a forced oscillation problem. The up and down arrow keys increase and decrease the frequency. The left and right arrow keys decrease and increase the increment in ω.

9-1 — pages 206-207 — Looking at reflected waves. You can see the uneven motion of a traveling wave with a small reflected amplitude.

9-2 — pages 209-211 — Reflection and transmission from a mass on a string.

10-1 — pages 226-227 — A triangular pulse propagating on a stretched string.

10-2 — pages 229-231 — Group velocity (sum of two cosines).

10-3 — pages 239-241 — Scattering of a pulse by a boundary between regions of different k.

10-4 — pages 241-246 — Scattering of a pulse by a mass on a string.

11-1 — pages 256-257 — The modes of a two-dimensional beaded string. On the PC, you can change n between 1 and 4 with the left and right arrows. You can change n' between 1 and 3 with the up and down arrows.

11-2 — pages 263-267 — Snell's law with no reflection.

11-3 — pages 286-288 — Water sloshing in a rectangular container.

11-4 — pages 324-324 — Two immiscible liquids sloshing. Note the mismatch between the upper and lower liquids in the middle. This is the result of the nonlinearity of the constraint of incompressibility.

12-1 — pages 334-340 — Polarization in the two-dimensional harmonic oscillator, or in an electromagnetic wave. This shows the position of a string stretched in the z direction. The transverse position is shown in the x-y plane along with the x and y components. Alternatively, this can represent E_x and E_y in the electromagnetic wave propagating in the z direction and the total $\vec{E}$ field. In the upper left-hand corner is the complex two dimensional vector,

$$Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

that describes the polarization.

You can change u_1 between 1 and 0 with the left and right arrows. You can change $|u_2|$ between 1 and 0 with the up and down arrows. F1 and F2 decrease and increase the phase of u_2 between π and $-\pi$.

12-2 — pages 342-342 — The wandering of the electric field in unpolarized light. The electric field direction in the x-y plane is indicated by the trace. The color of the line changes occasionally to make it visible.

Appendix B

Solitons

Consider a system of nonlinear coupled pendulums. In particular, suppose that there is a rigid rod along the z axis. At regular intervals, a, along the rod, there are masses attached to light rods of length, ℓ, connected to the central rod by a frictionless sleeve, so that each one is free to rotate around the axis of the central rod in a plane of fixed z. Each such pendulum looks, in its x-y plane, like the one shown in figure 3.5. Now finally, assume that each pendulum bob is connected to those of its two nearest neighbors by springs with spring constant K, and a small equilibrium length (much less that a) which we will ignore.[1]

The lowest energy state of this system of oscillators is one in which all the pendulum bobs are hanging straight down, as shown in figure B.1.

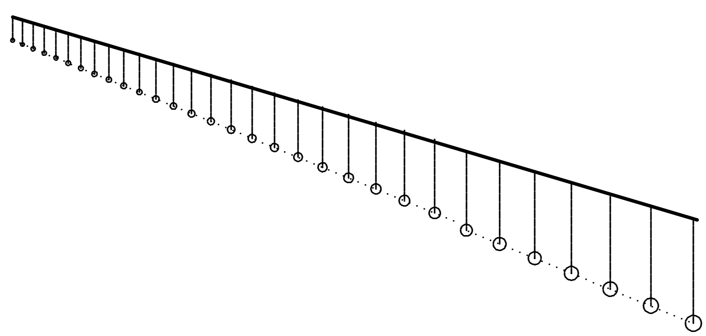


Figure B.1: A finite portion of the infinite system of coupled pendulums in its lowest energy state. The coupling springs are shown as dotted lines.

[1]In this approximation of ignoring the small equilibrium length of the springs, the system of coupled oscillators gets a linear restoring force from the springs even for rather large displacements. This is what Crawford call the “slinky approximation.”

For small oscillations about this equilibrium position, this system is approximately linear and has the dispersion relation:

$$\omega^2 = \frac{g}{\ell} + \frac{4K}{m}\sin^2\frac{ka}{2}\,. \tag{B.1}$$

Interesting things happen when we push this system into the nonlinear regime. There are other stable equilibrium configurations of the infinite system besides that shown in figure B.1. As in figure B.1, the pendulums must hang straight down as $z \to \pm\infty$. However, in between $-\infty$ and $+\infty$, the system may have **twists**. In particular, suppose that as we go from $-\infty$ to $+\infty$, there is a right-handed twist in the system, as shown in figure B.2.

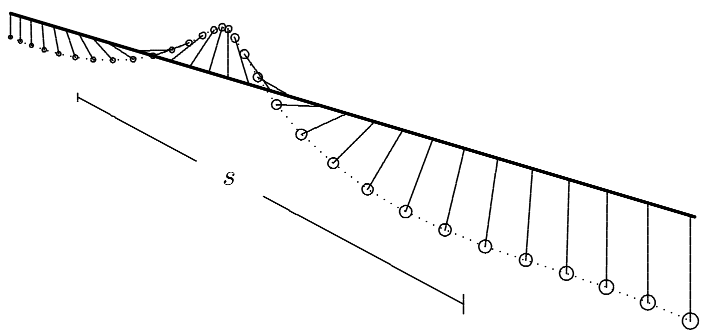


Figure B.2: A twist, or soliton, in the system of coupled pendulums. The "size," s, of the soliton is shown.

This object is completely stable. The twist cannot be removed without breaking one on the springs joining the masses. This twist is called a "soliton," short for solitary wave. Far from the center of the twist, the angular displacement of the pendulums from the downward direction falls off exponentially. The soliton has a size, s, indicated in figure B.2, which is the distance in the z direction over which the angular displacement is of order 1. The size, s, is determined by a competition between two physical effects. Gravity would like to have as many of the pendulums as possible hanging nearly straight down. This tends to make the soliton shrink to a smaller length in the z direction. However, as the soliton gets smaller, the difference in angular displacement between the neighboring pendulums in the soliton get larger, and thus the coupling springs get more stretched. Thus the effect of the coupling springs is to make the soliton spread out.

We can make an order of magnitude estimate of s by making more explicit estimates of the energy stored in gravitational potential energy and in the potential energy of the springs. To do this exactly would require knowing the precise solution for the displacements of the pendulums in the soliton, which we will not try to work out. But we can estimate the order of magnitude rather easily. To get a general idea, we will assume that the angular displacements from one pendulum to the next are roughly equal inside the soliton where the angular displacements are large.

First, consider the gravitational potential energy. The number of masses that are significantly displaced in the twist is of the order

$$n \approx \frac{s}{a}\,. \tag{B.2}$$

The masses are raised above their lowest position by between 0 and 2ℓ, depending on where in the soliton they are. On the average, they are raised by about ℓ. Thus the total gravitational potential energy stored in the soliton is of order

$$PE_G \approx nmg\ell \approx \frac{mg\ell s}{a}\,. \tag{B.3}$$

As we anticipated, the gravitational potential energy grows as s increases.

Now, let us consider the potential energy stored in the springs. The angles of the pendulums change by a total of 2π as we go through the soliton. Thus from (B.2) and the assumption that the displacements are approximately uniform, we find that the angular displacement from one rod to the next is of the order of

$$\Delta\theta \approx \frac{2\pi}{n} \approx \frac{2\pi a}{s}\,. \tag{B.4}$$

The length of the stretched spring is thus

$$\approx \sqrt{a^2 + \Delta\theta^2\ell^2} \approx \sqrt{a^2 + 4\pi^2 a^2\ell^2/s^2}\,. \tag{B.5}$$

The change in the potential energy from the downward position of a single spring is then

$$\approx \frac{1}{2}K\left(a^2 + 4\pi^2 a^2\ell^2/s^2\right) - \frac{1}{2}Ka^2 \approx 2K\pi^2 a^2\ell^2/s^2\,. \tag{B.6}$$

The total change in the potential energy in the n stretched springs is (from (B.2) and (B.6))

$$PE_K \approx 2K\pi^2 a\ell^2/s\,. \tag{B.7}$$

Again, we anticipated the result. The potential energy stored in the springs decreases as s increases.

The total energy of the soliton, from (B.3) and (B.7), is

$$PE \approx \frac{mg\ell s}{a} + 2K\pi^2 a\ell^2/s\,. \tag{B.8}$$

The soliton will adjust itself to minimize this total potential energy. We can find the minimum by differentiating with respect to s and setting the result to zero:

$$\frac{dPE}{ds} \approx \frac{mg\ell}{a} - 2K\pi^2 a\ell^2/s^2\,, \tag{B.9}$$

which implies

$$s \approx \pi a\sqrt{\frac{K\ell}{mg}}\,. \tag{B.10}$$

The soliton is particularly interesting in the continuum limit. The limit of this system as $a \to 0$ is a light, elastic ribbon, attached by a flexible sleeve to the rod on the z axis and weighted along the other side of the ribbon. The soliton exists for the continuous ribbon. The size can obtained by taking the continuum limit of (B.10), in the sense of chapter 6. The result is

$$s \approx \pi\sqrt{\frac{T\ell}{\rho g}}\,, \tag{B.11}$$

where $T = Ka$ is the tension in the elastic ribbon, and ρ is the linear mass density of the weighted ribbon.

There is an interesting difference between properties of the soliton in the continuous and the discrete cases that points up the differences between nonlinear and linear systems. We have seen in our study of linear systems that the discrete space translation invariance of a discrete system and the continuous space translation invariance of a continuous system have very similar implications. Both lead, through the magic of linearity, to modes of the form $e^{\pm ikx}$. However, for the nonlinear system, there is an important difference between the discrete and continuous case. In the discrete case, equilibrium soliton solutions exist only for discrete positions of the center of the soliton. There are an infinite number of possible positions, separated by multiples of a. But in the continuous case, the space translation invariance of the continuous system ensures that the center of the soliton can sit anywhere and be equally happy. This implies, in turn, that the soliton can move at constant velocity along the rod in the z direction.

The soliton behaves, in many ways, like a particle. We have seen that this is also true of a wave packet in a linear system. However, the important difference between a soliton and a wave packet is that the soliton never dissipates. It is held together indefinitely by the nonlinear interactions.

Appendix C

Goldstone Bosons

Consider an infinitely long rope, stretched along the x axis from $x = -\infty$ to ∞. We have seen that this system has space translation invariance for translations in the x direction. Because the system is linear for small oscillation, this implies, by the arguments chapters 4 and 5, that the normal modes are exponential waves of the form

$$e^{\pm i(kx \mp \omega t)} . \tag{C.1}$$

Spontaneous Symmetry Breaking

But the rope system has a much more subtle and even more interesting symmetry. The rope lives in three-dimensional space, and the laws that govern its motion are invariant under translations in the y and z direction, as well as the x direction. However, the rope, stretched in the x directions, sits at some definite value of y and z. Whereas, for an infinite, featureless rope, you cannot tell when it moves in x direction, you see it moving immediately if it is moved in the transverse directions, y and z. In a situation like this, the y and z translation invariance is said to be spontaneously broken.

A symmetry is said to be spontaneously broken if the underlying laws of nature are symmetrical, but the lowest energy state of the system is not. A simple example is a pencil balanced on its tip. This is a system in unstable equilibrium. The slightest nudge in any direction will cause the pencil to fall. But which way does it fall? The pencil is perfectly symmetrical under rotation about an axis through the center of its lead. Thus all directions are equally good. Nevertheless, the pencil does fall. Some random direction is picked out by the small fluctuations that cause the pencil to fall. Any direction is a good as any other. But once the direction is picked out, the pencil falls that way and the rotation symmetry of the unstable state, and of the underlying physical laws is spontaneously broken when the pencil is lying on its side in a lowest energy state. In the same way, the stretched rope must pick out some definite value of y and z. Any values will do.

You might think that we could not get any useful information out of a symmetry, like the y and z translation symmetry of the rope that is spontaneously broken. But despite the fact that the position of the rope breaks the symmetry, the invariance of the underlying laws of motion under y and z translations has an important consequence. It implies that the dispersion relation of the rope has the property that $\omega \to 0$ as $k \to 0$. The point is that we can move the entire, infinite rope in the y or z direction, with no restoring force, because the symmetry guarantees that the rope is equally happy to sit at any value of y and z. Therefore, as we make k smaller and smaller in the wave mode $e^{\pm i(kx \mp \omega t)}$, so that the wavelength of the waves gets larger and larger, the restoring force, and therefore the angular frequency ω, gets smaller and smaller, going to zero in the limit $k \to 0$.

In other words, the spontaneously broken symmetry actually gives you a different kind of information. It tells you something important about the dispersion relation:

$$\omega \to 0 \quad \text{as} \quad k \to 0\,. \tag{C.2}$$

The rope can carry traveling waves of arbitrarily low frequency. There is no low frequency cut-off. This is the dynamical consequence of the spontaneously broken translation symmetry.

In quantum mechanics, these waves correspond to particles, with $p = \hbar k$ and $E = \hbar\omega$. And waves for which $\omega \to 0$ as $k \to 0$ correspond to massless particles which travel at the speed of light in a relativistic world.

The fact that these massless particles are an inevitable consequence of a spontaneously broken symmetry in a relativistic theory is called the "Goldstone Theorem," in honor of one of the first particle physicists to state it clearly. The massless particles themselves are called Goldstone bosons.