IISER Pune

PHY 102

The World of Physics: Waves and Matter

Spring 2016

Course Instructors: Bhas Bapat and Sunil Mukhi

Version dated April 13, 2016

Contents

1 Oscillations **4**

1.1 General features, definitions 4

1.2 Harmonic oscillations in one dimension 6

1.2.1 General solution 6

1.2.2 Energy conservation 8

1.2.3 Superposition of different frequencies 9

1.3 Harmonic oscillator in two and three dimensions 14

1.3.1 Isotropic case 14

1.3.2 Anisotropic case 16

1.4 Damped harmonic oscillator 18

1.5 Forced oscillator and resonance 23

1.6 Forced harmonic oscillator with damping 25

1.7 Power absorbed by a forced oscillator 29

1.8 Electrical circuit as an oscillator 30

2 Coupled Oscillations **31**

2.1 Two identical coupled oscillators 31

2.2 N identical coupled oscillators 34

2.3 Large number of identical coupled oscillators 35

2.4 Transverse Displacements 36

3 A stretched vibrating string **38**

3.1 Generalised displacement of a string and harmonic analysis 39

3.2 Fourier Decomposition 40

4 Travelling Waves **41**

4.1 Speed of a wave 42

4.2 Dispersion of waves 43

4.3 Transmission across a boundary 43

4.4 Sound Waves 44

5 Wave Pulses **46**

5.1 A pulse in time 47

5.2 Group of frequencies 48

5.3 Dispersion of waves 50

6 Electromagnetic waves **51**

6.1 Plane electromagnetic waves 52

6.2 Standing electromagnetic waves 55

6.3 Energy of an electromagnetic wave 56

6.4 Polarisation 57

6.5 Interference 58

6.6 Coherence and bandwidth 59

7 Elastic Properties of Matter **61**

7.1 Stress and Strain 61

7.2 Uniform Strain 63

7.3 Shear 64

7.4 Torsion 66

8 Bending of Beams **67**

Guide to these notes

These notes are accompanied by exercises. They are of three kinds:

Exercise (level A): An exercise to fill in missing steps in a derivation, or to generalise a formula. These exercises will be very simple and straightforward. Everyone should attempt them in order to get a clearer understanding of the notes.

Exercise (level B): An exercise that is somewhat challenging and requires thought and/or calculation. These are supposed to prepare the student for the kind of questions that could be asked in the exams or quizzes. In fact, some of the Exercises of level B may actually appear in an exam or quiz of this course. Everyone is strongly advised to solve these and understand the method properly.

Exercise (level C): These are more mathematical. They are intended for students who like mathematics and want to understand the course material in greater depth. However, everyone need not do them and questions of this level will not appear in any of the exams/quizzes.

Exercise (level X): This is not really an exercise. It is usually just a comment that challenges you to think.

Evaluation methods

This course will be evaluated by two quizzes, a mid-semester exam, two more quizzes and a final exam. The portion for each evaluation consists of *all the material* that has been taught up to that time.

The mid-sem and final will count for 30 marks each and will have "subjective" type questions where the answer has to be derived and written out. Each quiz will count for 10 marks and will be of "objective" type with multiple-choice answers. The exams and quizzes are intended to be done by each student individually with no help from anyone else.

There will also be Exercises (for more details, see below) given in the notes or assigned in class. These can be treated as homework assignments. You are welcome to do them together with others, though it will be more helpful if you think about them by yourself first. These exercises will not be collected or graded, but if you have any difficulties or just want your answers checked, you are welcome to come to our offices in person.

The exams and quizzes are supposed to provide a just and fair ranking of each student's performance. If any student uses unfair methods or cheating of any kind, this purpose cannot be fulfilled. Therefore there will be absolutely no tolerance of cheating in this course and penalties will be imposed on anyone who is caught. Please note that copying an answer from another person, as well as supplying an answer to another person, are both forms of cheating and both are equally punishable. refer to the Section "Evaluation Methods" at the end of these notes, for useful information about quizzes, exams etc.

One comment about the course

If you rely on memorisation, you will not do well in this course! Our goal is to help you think for yourself and understand.

1 Oscillations

1.1 General features, definitions

Oscillations are among the most familiar phenomena seen in daily life:

- Our heartbeat
- Our breathing
- Our voice
- A bouncing ball
- Waves on the seashore
- The pendulum of a clock
- Expansion and contraction of a spring
- The string of a musical instrument
- Vibration of air in a flute
- The buzzing of an insect
- The rotation of the earth on its axis
- The revolution of the earth around the sun

The key property is that these are *periodic* motions. Such motions have a typical *period* after which the system returns to its initial state. The period varies widely across physical systems. For example our heart is a pump that contracts and expands around 80 times a minute. Therefore its period (the amount of time for one oscillation) is $\frac{1}{80}$ minutes, or $\frac{60}{80} \sim 0.75$ seconds. Our lungs expand and contract more slowly, around $12 - 20$ times a minute. So their period is about $3 - 5$ seconds.

The number of oscillations per unit time is called the *frequency*. It is the inverse of the period. In physics we typically denote the period by T or τ (indicating "time") and the frequency by ν. Thus:

$$\nu = \frac{1}{T}$$

A musical string playing the note called "middle C" vibrates around 256 times a second. The highest frequency that humans can hear is about 22,000 vibrations per second.

The unit of vibration/oscillation per second is the Hertz (Hz) named after the physics Heinrich Hertz. It also used to be called "cps" for "cycles per second". One also uses:

$$\begin{aligned} \text{KHz (kilo Hertz):} &\quad 1\,\text{KHz} = 10^3\,\text{Hz} \\ \text{MHz (mega Hertz):} &\quad 1\,\text{MHz} = 10^6\,\text{Hz} \\ \text{GHz (giga Hertz):} &\quad 1\,\text{GHz} = 10^9\,\text{Hz} \end{aligned}$$

There is another common unit which is closely related. Suppose we have an object moving in a circle. This may not look like a "vibration" or "oscillation" but in fact it is. The motion is periodic, since the object keeps returning to the same place. For such objects we use the "angular frequency" ω, defined as the *angle in radians* swept out per second. Clearly when the object undergoes one full

revolution, it sweeps out an angle of 2π. So if it undergoes ν revolutions per second, its angular frequency will be $2\pi\nu$. Thus:

$$\omega = 2\pi\nu = \frac{2\pi}{T}$$

One interesting feature of oscillations is that one can *superpose* them and the result is rather complicated. For example, the earth rotates around its axis *and* revolves around the sun. As a result, any fixed point of earth executes a rather complicated motion. Superposed oscillations are the basis of the "spirograph" toy. A more complicated example is a bouncing ball on a merry-go-round. As viewed from the sun, its motion is a superposition of (i) bouncing off the floor, (ii) rotation of the merry-go-round, (iii) rotation of the earth on its axis, and (iv) revolution of the earth. The resulting motion is very complicated and may not appear to be periodic.

Another nice example of superposed oscillations is the sound of a musical instrument. If we play the note "middle C" on a sitar, a sarod, a veena, a shehnai, a flute, a piano or a guitar, it sounds very different. The reason is that although the main vibration involved is 256 Hz, there are additional vibration frequencies that are multiples of this one (e.g. 512 Hz, 1024 Hz). Moreover, vibrations are induced in the body of the instrument. These depend on the material and its shape and therefore tend to be very complex. Also, a sitar has "sympathetic" strings that vibrate on their own when the main string is plucked. All these effects combine to produce the characteristic sound of the instrument.

A second interesting feature about oscillations is that they typically die down due to an effect called *damping*. Every one of the effects described above (including the earth's rotation around its axis!) will eventually come to an end unless some energy is supplied to keep it going. If energy is supplied then of course the vibration/oscillation can be maintained, and even increased.

A third interesting feature about oscillations is the concept of *amplitude* which tells us *how much* oscillation is happening. If you gently hit a note on the piano or bang it hard, you get the same frequency but in one case the sound is soft and in the other, loud. This is true of any kind of sound. One may think the rotation of the earth around its own axis has no concept like a period, but imagine two planets, one much bigger than the other, each rotating around its own axis at the same rate (e.g. one full turn in 24 hours). We can think of the differing sizes as the analogue of vibration amplitude.

Often the amplitude seems unrelated to the time period. For example, the pendulum of a clock may swing back and forth in one second, independent of whether it is swinging through a large or small angle. However, in general the amplitude and period for a pendulum are actually linked. The linkage is not visible for small amplitudes but becomes significant for large amplitudes. Only very special, idealised systems have completely de-linking between the amplitude and the period.

A final observation is that if we have a system that is not oscillating, and we couple it to a system that is oscillating, then the latter system can "induce" or "force" oscillations in the former. Such forced oscillations will not have the frequency typical of the original system but rather, those of the forcing system. Something special happens when the two frequencies match: a phenomen called *resonance* takes place and the original system will oscillate with increasing amplitude.

This has been a collection of qualitative facts about vibrations without using any mathematics. From now on our goal will be a precise, quantitative analysis of vibrations. These will justify many of the above statements that were based purely on observation and common sense. A quantitative understanding of physical phenomena is the primary goal of physics.

1.2 Harmonic oscillations in one dimension

1.2.1 General solution

The space in which we live has three dimensions. The possible motions of a point object in three dimensions are very complicated – it can trace any path through space. But if the object is confined to move in just one dimension then the motion is simpler. Suppose the single direction is denoted as the x-axis with $-\infty < x < \infty$. Oscillatory motion then means that for some time the object is moving towards larger x, then it reverses and moves towards smaller x, then reverses again. Each time the direction reverses, the object has to stop for an instant.

Imagine an idealised point object whose position evolves as a function of time. This is denoted by $x(t)$. We use the notational conventions:

$$\dot{x} = \frac{dx}{dt}, \quad \ddot{x} = \frac{d^2x}{dt^2}$$

To find a simple equation we can assume the object experiences a *linear restoring force* that always pulls it towards the origin. Thus:

$$F = m\ddot{x} = -kx \tag{1.1}$$

An object obeying this law is called a *simple harmonic oscillator*. Here k is the *spring constant* of the oscillator.

Some simple properties can be deduced by looking at the equation. When x is positive the force is along the negative direction, and when x is negative the force is along the positive direction. When $x = 0$ there is no force. The above equation can be easily rewritten:

$$\ddot{x} = -\omega^2 x \tag{1.2}$$

where $\omega = \sqrt{\frac{k}{m}}$. We have deliberately used the symbol for angular frequency here. Soon we will see that this is the correct meaning of ω.

Note that the above is a *linear* equation. If we rescale x by any factor, say $x \to \lambda x$ where λ is a nonzero constant, the equation remains the same. This is special to the harmonic oscillator. More general oscillators will be *nonlinear*.

Now we must decide what is the *initial condition*. A very simple condition is that at $t = 0$ the object is at rest at $x = 0$ (this is written as $x(0) = 0, \dot{x}(0) = 0$). Then at this instant, the above equation reduces to:

$$F(t = 0) = 0 \tag{1.3}$$

Since the force is zero, the object will not move. Therefore at any later time the object will simply remain at $x = 0$. Clearly this is a very special initial condition and the resulting motion is not interesting. So let us look at more general initial conditions.

One possibility is that at $t = 0$ the object is at rest at some point A. Thus $x(0) = x_0, \dot{x}(0) = 0$. Another possibility is that at $t = 0$ the object is at the origin but moving with some velocity v_0. Thus $x(0) = 0, \dot{x}(0) = v_0$. The most general initial condition is that at $t = 0$ the position and velocity are *both* arbitrary:

$$x(0) = x_0, \quad \dot{x}(0) = v_0$$

where x_0, v_0 are arbitrary numbers that can be independently positive, negative or zero. Now let us consider the following function:

$$x(t) = A\cos\omega t + B\sin\omega t \tag{1.4}$$

By differentiating repeatedly we find:

$$\begin{aligned} \dot{x}(t) &= -A\omega\sin\omega t + B\omega\cos\omega t \\ \ddot{x}(t) &= -A\omega^2\cos\omega t - B\omega^2\sin\omega t \end{aligned} \tag{1.5}$$

Comparing Eqs.(1.4) and the second equation in (1.5), we see that $\ddot{x} = -\omega^2 x$ which is the equation we were trying to solve. Also by putting $t = 0$ in the previous equations we find:

$$x(0) = x_0 = A, \quad \dot{x}(0) = v_0 = \omega B \tag{1.6}$$

Thus the proposed function indeed solves our force equation as well as the given initial conditions! In fact this is the most general solution of the above problem. We will not derive this fact, but you can assume it to be true.

What are the period and frequency of this oscillator? We know that the sin and cos functions have the following *periodicity* property:

$$\sin(\theta + 2\pi) = \sin\theta, \quad \cos(\theta + 2\pi) = \cos\theta$$

The period of the oscillator is defined by saying that $x(t+T) = x(t)$. This will be true if $\omega(t+T) = \omega t + 2\pi$. It follows that $\omega = \frac{2\pi}{T}$. This proves that we have given ω the correct interpretation of angular frequency.

Some books write the motion of the simple harmonic oscillator in one dimension as:

$$x(t) = C\sin(\omega t + \alpha) \tag{1.7}$$

This looks quite different from Eq.(1.4)! But in fact it is the same, in a different notation. Let us use the identity:

$$\sin(X + Y) = \sin X \cos Y + \cos X \sin Y \tag{1.8}$$

Then we can write the above solution as:

$$x(t) = (C\cos\alpha)\sin\omega t + (C\sin\alpha)\cos\omega t$$

This is the same as Eq.(1.4) if we identify $A = C\cos\alpha$ and $B = C\sin\alpha$. So the two ways of writing the solution are simply related to each other. The advantage of the second way can be seen if we plot the function. We see that C is just the maximum amplitude while α, known as the "phase", determines where on the graph the particle starts out at $t = 0$.

Notice that C is completely arbitrary and has no relation to ω. This is the statement that the *amplitude* of oscillations has nothing to do with the *frequency* of oscillations. This property is special to the harmonic oscillator with $F = -kx$. If we had started with a different restoring force, we would have the anharmonic oscillator and such a rule would not be true. We will soon explain this in more detail.

Exercise (level A): (i) Solve for C and α in terms of A and B. (ii) Sketch the graph of Eq.(1.7) for yourself by plotting a number of points on it.

Exercise (level C): For those who know complex numbers. Show that

$$x(t) = De^{i\omega t} + D^* e^{-i\omega t}$$

is another way to write the solution of the simple harmonic oscillator, where D is a complex constant. Find the relation between D and A, B in Eq.(1.4).

Exercise (level X): Do you believe the identity Eq.(1.8). If yes, why? Do you really know it is true? Can you find ways to make it more believable to yourself?

Since the equation $F = -kx$ is linear, we can *superpose* two different harmonic motions for the same particle of mass m and spring constant k. Since these constants determine the angular frequency

ω, this amounts to superposing two different expressions of the form Eq.(1.4). Thus the superposed motion is given by:

$$\begin{aligned} x(t) &= A\cos\omega t + B\sin\omega t + A'\cos\omega t + B'\sin\omega t \\ &= (A+A')\cos\omega t + (B+B')\sin\omega t \end{aligned} \tag{1.9}$$

It is obvious that the superposition is also a solution of the same form Eq.(1.4) with the constant coefficients just added. In particular this means physically that the original position of the particle becomes the sum of the original positions, and the original velocity also becomes the sum of the original velocities. Things are not this simple if we superpose two motions with *different* angular frequencies. We will come back to this case somewhat later.

Exercise (Level B): Consider the above superposition in the amplitude and phase representation Eq.(1.7). Find the amplitude and phase of the superposed motion as a function of the original amplitudes and phases.

Exercise (Level B): A particle moving in one dimension is subjected to three harmonic motions of amplitudes 0.25 mm, 0.20 mm and 0.15 mm respectively. The phase difference between the second and first motion is 45° and between the second and the third, 30°. Find the amplitude of the resulting motion, as well as its phase (relative to the first oscillator).

1.2.2 Energy conservation

It is useful to calculate the total energy of the simple harmonic oscillator and explicitly derive the fact of energy conservation. There is a useful trick to achieve this, which works for very general systems in one dimension but we will only apply it to the simple harmonic oscillator. Multiply both sides of the original equation Eq.(1.2) by $\dot{x}$ to get:

$$\dot{x}\ddot{x} = -\omega^2 x\dot{x} \tag{1.10}$$

Now, notice that:

$$\frac{d}{dt}(x^2) = 2x\dot{x}, \quad \frac{d}{dt}(\dot{x}^2) = 2\dot{x}\ddot{x} \tag{1.11}$$

Thus our equation becomes:

$$\frac{d}{dt}(\dot{x}^2) = -\omega^2\frac{d}{dt}(x^2) \tag{1.12}$$

It follows that:

$$\dot{x}^2 = -\omega^2 x^2 + C \tag{1.13}$$

where C is an arbitrary constant.

There is a nice physical interpretation for the above equation. By multiplying both sides by $\frac{1}{2}m$ and rearranging, we find:

$$\frac{1}{2}mC = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2 \tag{1.14}$$

Now notice that the first term on the RHS is just $\frac{1}{2}mv^2$, known as the *kinetic energy* of the particle. The equation is telling us that the kinetic energy plus another term $\frac{1}{2}m\omega^2x^2$ is a constant. We interpret this second term as the *potential energy* and the constant sum as the *total energy*. Thus we can write:

$$E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2 \tag{1.15}$$

We see that as the kinetic energy increases, the potential energy decreases and vice-versa. The fact that E is constant is the law of conservation of energy.

Exercise (level B): A 500 gm cube connected to a light spring for which the force constant is 20 N/m oscillates on a horizontal, frictionless track. (a) Calculate the total energy of the system and the maximum speed of the cube if the amplitude of the motion is 3 cm. (b) What is the velocity of the cube when the displacement is 2 cm? (c) Compute the kinetic and potential energies of the system when the displacement is 2 cm. (d) At what value of x is the speed of the cube equal to 10 cm/sec?

Exercise (level B): Find the energy in terms of the coefficients A and B in Eq.(1.4) and also in terms of C and α in Eq.(1.7). Does the energy depend on the phase α? What is the physical reason?

Exercise (level C): Rewrite the above equation as:

$$\int \frac{dx}{\sqrt{\frac{2E}{m} - \omega^2 x^2}} = \int dt \tag{1.16}$$

Impose suitable limits of integration and integrate to find $x(t)$. Verify that your solution is of the form in Eq.(1.4).

1.2.3 Superposition of different frequencies

In a previous section we considered superposing two harmonic oscillators with the same frequencies but possibly different amplitudes and phases. The result was again a harmonic oscillator of the same frequency, with its amplitude and phase determined by those of its components.

Now we will consider the superposition of two harmonic oscillators with different frequencies. Such superpositions occur in many different contexts in nature. As a physical example we can imagine two different musical strings tuned to different frequencies.

Let us start by considering two oscillators with unit amplitude and zero phase, but distinct frequencies:

$$x_1(t) = \sin \omega_1 t, \quad x_2(t) = \sin \omega_2 t \tag{1.17}$$

Obviously the superposition is:

$$x(t) = x_1(t) + x_2(t) = \sin \omega_1 t + \sin \omega_2 t \tag{1.18}$$

We can use the trigonometric formula[1]:

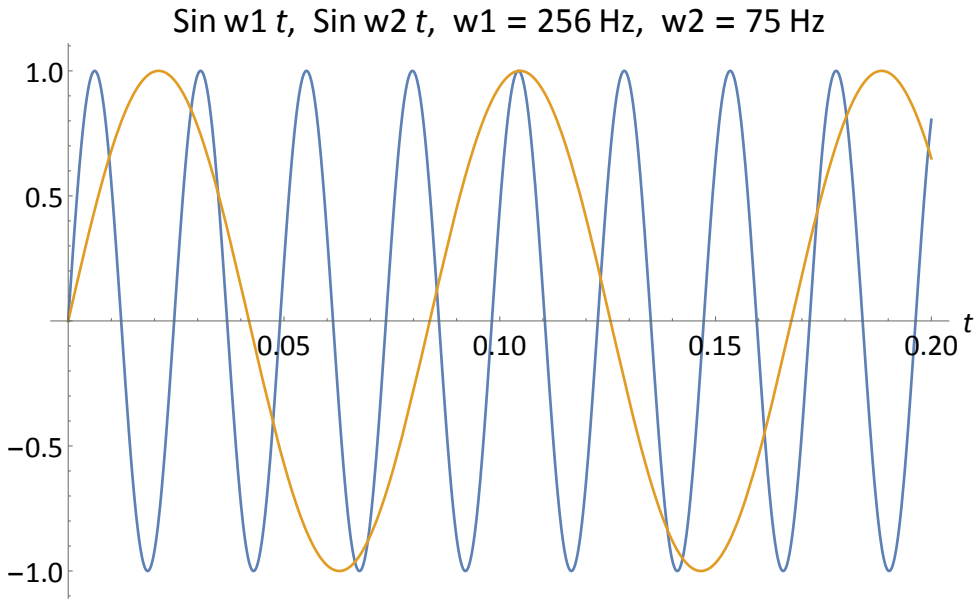
$$\sin a + \sin b = 2 \cos \frac{a-b}{2} \sin \frac{a+b}{2} \tag{1.19}$$

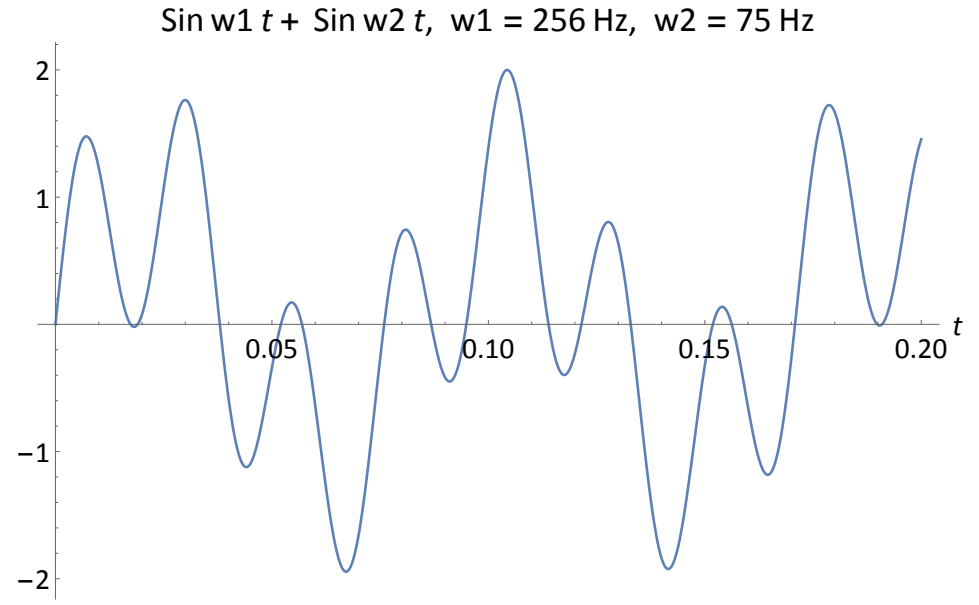
to rewrite the superposed oscillation as:

$$x(t) = 2 \cos \frac{(\omega_1 - \omega_2)t}{2} \sin \frac{(\omega_1 + \omega_2)t}{2} \tag{1.20}$$

It is very illuminating to plot such a superposition. First we plot the case of $\omega_1 = 256$ Hz, $\omega_2 = 75$ Hz:

[1] You can easily derive this formula using the representation of sine and cosine in terms of exponentials.





Clearly the sum is *not* a simple harmonic motion! In fact it shows features of two types of oscillations, one faster and another slower. From the figure, it is not even obvious that the resulting motion is periodic. We must think clearly what "periodic" means – it means that there is a finite portion of the graph (not a sinusoidal curve, but some more complicated one) which repeats indefinitely. The alternative is that the motion might *never* repeat.

We can derive the mathematical condition for the motion to repeat. This simply says that:

$$\sin\omega_1(t+T)+\sin\omega_2(t+T)=\sin\omega_1 t+\sin\omega_2 t$$

Since we have two different frequencies ω_1 and ω_2, it is no longer true that $T=\frac{2\pi}{\omega}$. So what is T? Let us denote by T_1, T_2 the periods of the individual oscillators. The full motion will be periodic only if *both* oscillators return to the same state at the same time. The first one returns after every lapse of T_1 seconds, so in general:

$$\sin\omega_1(t+n_1T_1)=\sin\omega_1 t$$

for any integer n_1. Similarly the second returns after a lapse of n_2T_2 seconds for any integer n_2. For both to return together, we must have:

$$n_1T_1=n_2T_2$$

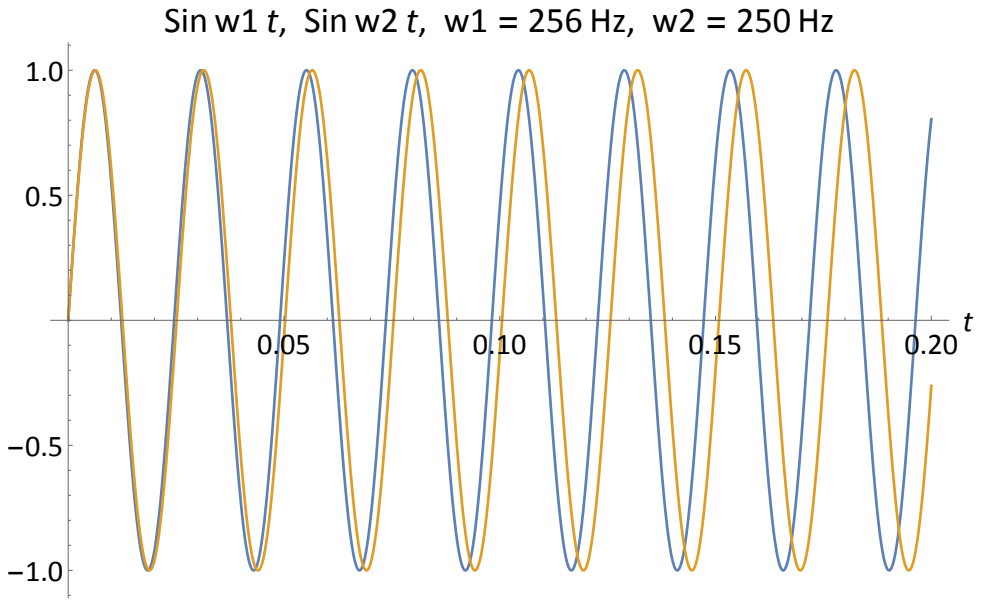
for some integers n_1, n_2. This is not always possible. For example if $T_1 = \sqrt{2}$ and $T_2 = \sqrt{3}$ seconds, then there are *no* integers n_1, n_2 such that the above equation is satisfied. The exact condition is found by rewriting the above equation as:

$$\frac{n_1}{n_2} = \frac{T_2}{T_1} = \frac{\omega_1}{\omega_2}$$

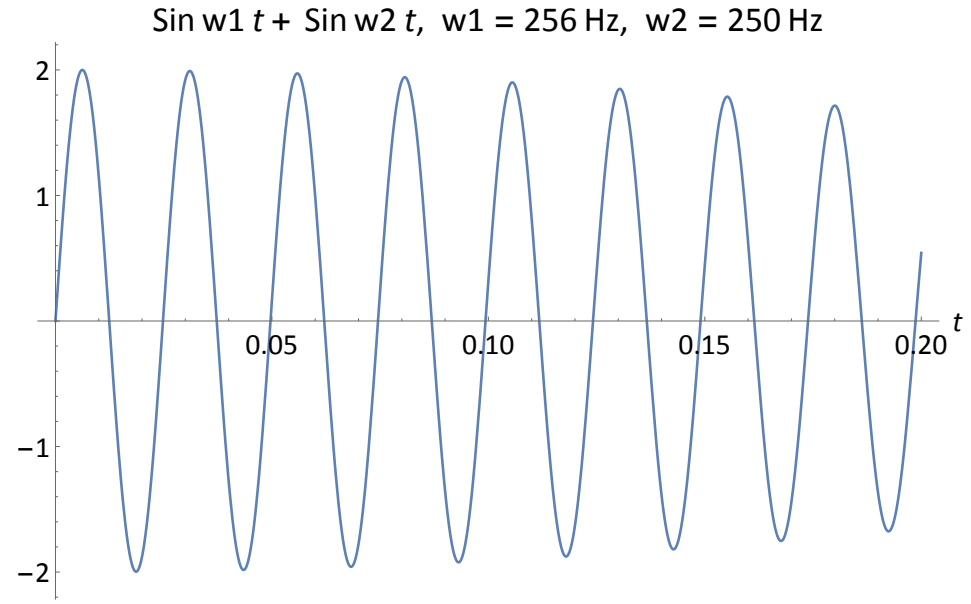
Since we know the frequencies of the two oscillators, the RHS is given to us. If it is a rational number then integers n_1, n_2 can be found to satisfy the equation. In this case the two periods are said to be *commensurable*. Otherwise such integers cannot be found and then the oscillators are *incommensurable*.

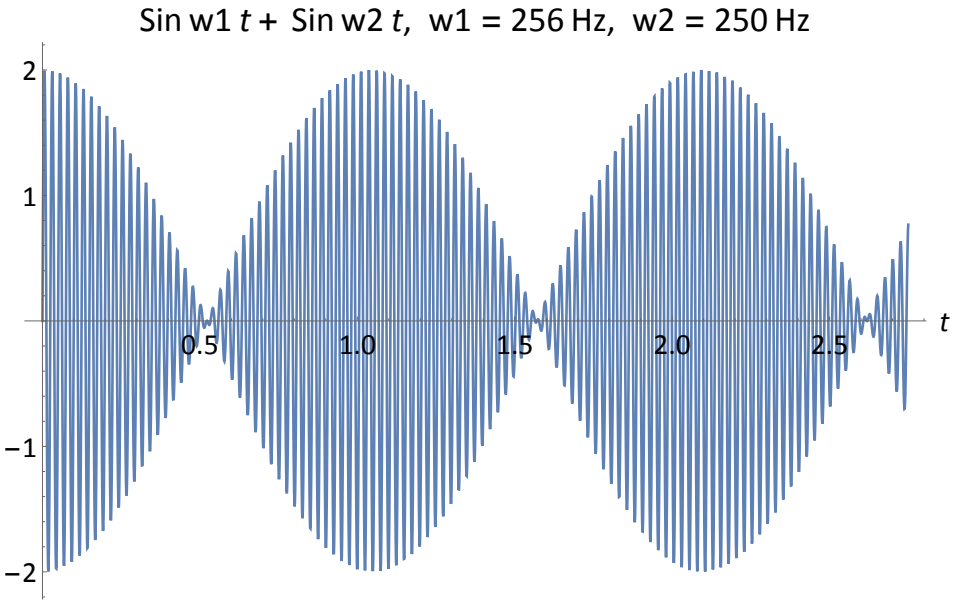
An interesting phenomenon emerges if we consider a closer pair of frequencies, $\omega_1 = 256$ Hz, $\omega_2 = 250$ Hz. For this particular example, Eq.(1.20) is:

$$x(t) = 2(\cos 3\,t)(\sin 253\,t) \tag{1.21}$$



First we look at the plot of the sum. This appears to show a slightly decreasing amplitude as t varies from 0 to 0.2. We can see this more clearly if we expand the range of t to cover 0 to 2.75.





Now we see that there is a "fast" oscillation of angular frequency 253 Hz multiplied by a "slow" one of angular frequency 3 Hz. The fast frequency is the average of the two original frequencies. It appears to have a varying amplitude, which is the slowly varying "envelope" of the shown curve. Inside the envelope, rapid oscillations take place.

When the oscillation corresponds to a musical sound, the human ear hears a note of the fast frequency that seems to swell and disappear periodically. These are called "beats". Notice that even though the envelope has a frequency $\frac{\omega_1-\omega_2}{2}$, the recurrence rate of beats is twice that, namely $\omega_1 - \omega_2$. This is clear from the figure. By the time the envelope completes a full sine wave, there are two beats contained in it.

We can see this in the above example. The the beat frequency is $\frac{(\omega_1-\omega_2)}{2} = 3$ Hz, but from what we said above, the recurrence rate should corresponds to double this frequency, namely 6 Hz. This frequency corresponds to a period of $T = 2\pi/6$ which is roughly 1 second. And indeed, as we see in the figure, the duration of a beat is roughly 1 second.

Another interesting point to note is that:

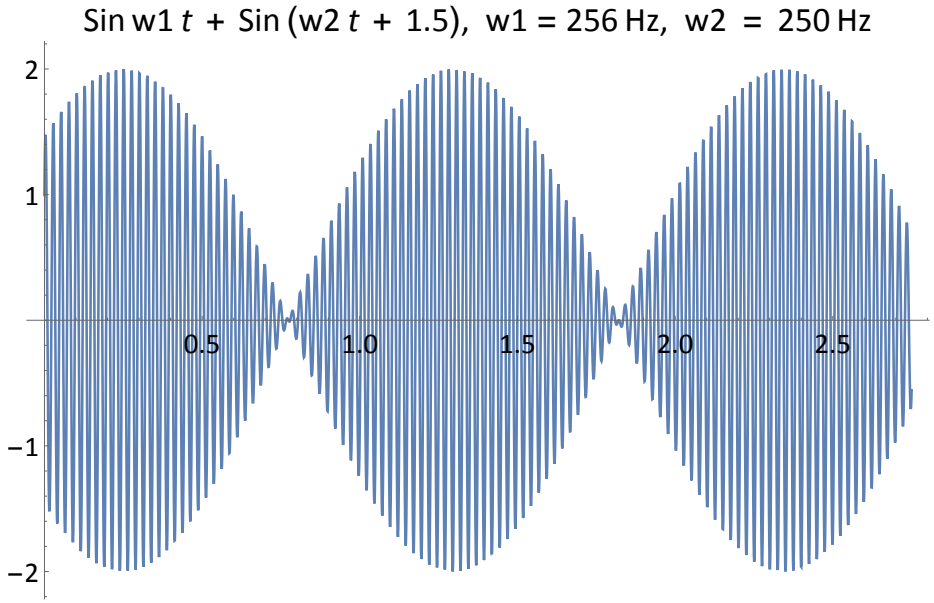
$$\frac{(\omega_1+\omega_2)/2}{(\omega_1-\omega_2)/2} = \frac{253}{3} \sim 84$$

Thus during one cycle of the envelope, there should be 84 fast oscillations. Since one cycle of the envelope is 2 beats, there must be roughly 42 oscillations inside a beat. One can count this number on the figure.

Above, we looked at a simple example where two harmonic oscillations were superposed with the same amplitude and phase. Let us try to see what happens if we vary one of these. First, we vary the phase. Thus, we consider the sum of two oscillations:

$$x(t) = \sin(256\,t) + \sin(250t + 1.5)$$

In this case, the sum of the two oscillations is given by the following figure:

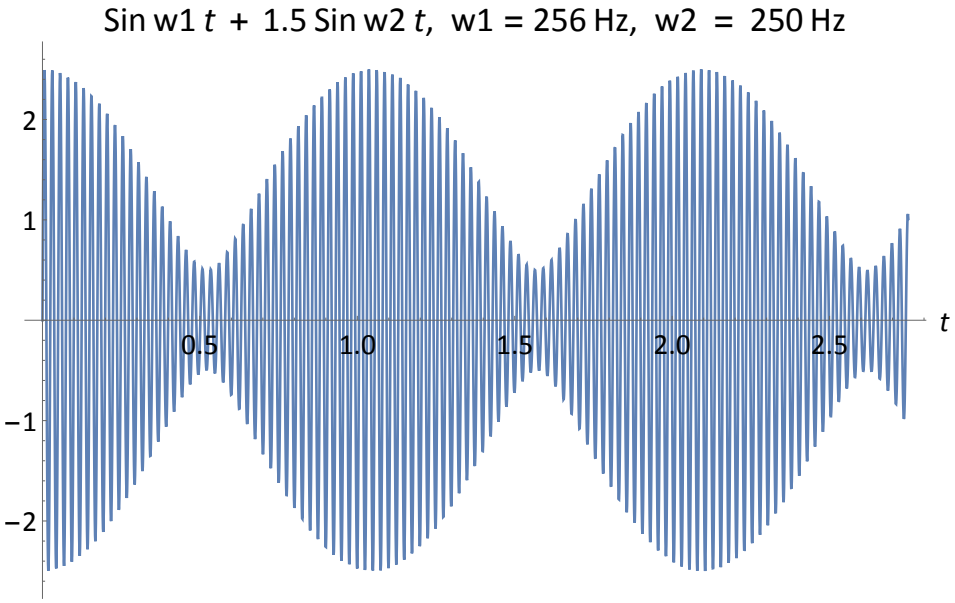


Comparing to the previous one, we see that the phase shift has only shifted the entire figure by the corresponding amount. But the structure of beats remains the same.

Next we may try to vary the amplitude. So consider the sum of two oscillations:

$$x(t) = \sin(256\,t) + 1.5\sin(250t)$$

The second oscillation has a different amplitude from the first. This time the graph looks as follows:



Although there are still beats, the amplitude never dies down completely to 0. This is obvious since a larger amplitude can never be completely cancelled by a smaller one.

Exercise (level B): Consider the superposition of two one-dimensional oscillators of frequencies $\omega_1 =$ 200 Hz and $\omega_2 = 75$ Hz and unit amplitude. What is the period of the combined motion? Suppose the frequency of the second oscillator is changed to 74 Hz, what is the period now?

1.3 Harmonic oscillator in two and three dimensions

1.3.1 Isotropic case

A two-dimensional *isotropic* harmonic oscillator satisfies the equation:

$$\vec{F} = -k\vec{x}$$

where both sides are two-component vectors. "Isotropic" means "same in all directions". This is the case because the spring constant is the same for motion in the x and y directions. In fact the above equation can be written, in components, as:

$$F_x = -kx, \quad F_y = -ky$$

The general solution is clearly:

$$x(t) = C\sin(\omega t + \alpha), \qquad y(t) = D\sin(\omega t + \beta),$$

It is easy, but not necessarily useful, to write down a general solution for the trajectory. We have:

$$\omega t = \sin^{-1}\left(\frac{x}{C}\right) - \alpha, \quad \omega t = \sin^{-1}\left(\frac{y}{D}\right) - \beta$$

Equating the two gives:

$$\sin^{-1}\left(\frac{x}{C}\right) - \sin^{-1}\left(\frac{y}{D}\right) - \alpha + \beta = 0$$

In principle this provides the orbit followed by the oscillator. However it is not very illuminating to express it in this form. Instead, we first look at some simple cases.

If $C = D = 1$ and $\alpha = \beta = 0$ then we have:

$$x(t) = \sin\omega t, \quad y(t) = \sin\omega t$$

It follows that $x(t) = y(t)$. Thus the path traced by the oscillator is a straight line at 45° to the x and y axes. Even though it is allowed to move in two dimensions, in this special case it just moves back and forth in one (slanted) direction.

Another special case is:

$$x(t) = \sin\omega t, \quad y(t) = \cos\omega t$$

It is easy to see that $x^2 + y^2 = 1$. Thus the path of the oscillator is a *unit circle* in two dimensions. We can easily check what is the direction in which it traces this circle. At $t = 0$ the object is at $x = 0, y = 1$ which is the top of the circle. After a little time, one sees from the functional form that x has increased while y has decreased. Thus the particle follows a *clockwise* trajectory.

To get additional insight into the general structure, let us generalise a little more. Suppose we repeat the previous cases but with different amplitudes:

$$x(t) = \sin\omega t, \quad y(t) = C\sin\omega t$$

This time the relation is $y(t) = Cx(t)$. This is still a straight line, but with a slope C instead of 1. Thus if $C > 1$ it will be steeper than 45° while if $C < 1$ then it will have a smaller angle.

Similarly, consider

$$x(t) = \sin\omega t, \quad y(t) = C\cos\omega t$$

The relation between x and y is now:

$$x^2 + \frac{y^2}{C^2} = 1$$

This is the equation of an ellipse.

Let us now leave the amplitudes equal but consider several different phases. Thus:

$$x(t) = \sin\omega t, \quad y(t) = \sin(\omega t + \alpha) \tag{1.22}$$

Above we have discussed the cases $\alpha = 0$ and $\alpha = \frac{\pi}{2}$. Suppose now that $\alpha = \frac{\pi}{4}$. Then:

$$y(t) = \sin(\omega t + \frac{\pi}{4}) = \frac{1}{\sqrt{2}}\big(\sin\omega t + \cos\omega t\big)$$

Eliminating $\sin\omega t$ between these equations, we get:

$$\cos\omega t = \sqrt{2}y - x$$

Since we also have $\sin\omega t = x$, we can square both equations and sum them, using $\sin^2\omega t + \cos^2\omega t = 1$, to get:

$$x^2 + (\sqrt{2}y - x)^2 = 1$$

which simplifies to:

$$2x^2 + 2y^2 - 2\sqrt{2}xy = 1$$

Cancelling a factor of 2, we have:

$$x^2 + y^2 - \sqrt{2}xy = \frac{1}{2}$$

We can simplify this by defining new coordinates:

$$x' = \frac{1}{\sqrt{2}}(x + y), \quad y' = \frac{1}{\sqrt{2}}(x - y)$$

Then it is easy to show that the above equation reduces to:

$$\left(1 - \frac{1}{\sqrt{2}}\right)x'^2 + \left(1 + \frac{1}{\sqrt{2}}\right)y'^2 = 1$$

This is an ellipse whose axes are oriented along the diagonal to the original x, y coordinate system.

We see that upon varying the relative phase between y and x, the combination describes: (i) a straight line (for $\alpha = 0$), (ii) an ellipse (for $\alpha = \frac{\pi}{4}$, (iii) a circle (for $\alpha = \frac{\pi}{2}$). Continuing in this way, one can get a picture of the trajectory as the phase varies from 0 to 2π.

Exercise (level A): Find the trajectory for phase differences $\frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}$ and 2π. It is not necessary to repeat the calculation in each case! Simple trigonometric identities will reduce all of these to the cases we have already considered, upto some signs. Consider also the phase $\frac{\pi}{3}$, and see if this helps you understand the overall picture better.

Exercise (level A): For each of the above trajectories, mark an arrow to show the direction (clockwise or counter-clockwise) of the particle's motion.

Exercise (level B): Find the equation satisfied by x and y (after eliminating t) for an arbitrary phase α in Eq.(1.22). Do not use inverse trigonometric functions. You should be able to find a quadratic equation in x and y.

Exercise (level C): Find a rotation that takes you from the equation in the previous exercise to the standard form of an ellipse.

Now let us briefly discuss the isotropic oscillator in 3 dimensions. We can imagine a point mass, free of gravitational force, attached to a central point by a spring. To specify its motion we need to choose six parameters – these can be the initial position (three components) and the initial velocity (three

more components), or equivalently the amplitudes and phases of the three independent oscillators. Thus:

$$\begin{aligned} x(t) &= C\sin(\omega t + \alpha) \\ y(t) &= D\sin(\omega t + \beta) \\ z(t) &= E\sin(\omega t + \gamma) \end{aligned} \tag{1.23}$$

From this formula it seems that harmonic motion in three dimensions can be extremely complicated. However there is an important simplifying feature. It is a theorem that any motion under a central force takes place in a plane. To prove this, assume a general central force:

$$\vec{F} = f(\vec{x})\frac{\vec{x}}{|\vec{x}|} \tag{1.24}$$

In our case, $f(\vec{x}) = -k|\vec{x}|$. But we will prove the theorem for any central force. We start by constructing the angular momentum vector:

$$\vec{L} = m\,\vec{x} \times \dot{\vec{x}} \tag{1.25}$$

The right hand side seems to depend on time, since both $\vec{x}$ and $\dot{\vec{x}}$ are functions of time. However, we can easily show that the angular momentum is *conserved* (independent of time). First, observe that:

$$\frac{d\vec{L}}{dt} = m\frac{d}{dt}\left(\vec{x} \times \dot{\vec{x}}\right) = m(\dot{\vec{x}} \times \dot{\vec{x}} + \vec{x} \times \ddot{\vec{x}}) \tag{1.26}$$

The first term on the RHS is clearly zero. For the second term, we use:

$$\ddot{\vec{x}} = \frac{\vec{F}}{m} = -\frac{f(\vec{x})}{m|\vec{x}\,|}\vec{x}$$

This shows that $\ddot{\vec{x}}$ is parallel to $\vec{x}$, therefore $\vec{x} \times \ddot{\vec{x}} = 0$. Thus we have proved that $\frac{d\vec{L}}{dt} = 0$.

Since $\vec{L}$ is conserved, it remains fixed in magnitude and direction for all time. And from its definition, we easily see that $\vec{L} \cdot \vec{x} = 0$. Thus, as the particle moves, its position vector remains permanently perpendicular to the fixed vector $\vec{L}$. This is the statement that it moves in a plane.

Knowing that the particle will move in a plane, we can reduce the three-dimensional harmonic oscillator to a two-dimensional one (lying in that plane). Therefore simple harmonic motion in three dimensions is no more complicated than in two dimensions.

1.3.2 Anisotropic case

We now consider the anisotropic harmonic oscillator in two dimensions. This object has the force law:

$$F_x = -k_1 x, \quad F_y = -k_2 y \tag{1.27}$$

where the spring constants are different: $k_1 \neq k_2$. Defining

$$\omega_1 = \sqrt{\frac{k_1}{m}}, \quad \omega_2 = \sqrt{\frac{k_2}{m}}$$

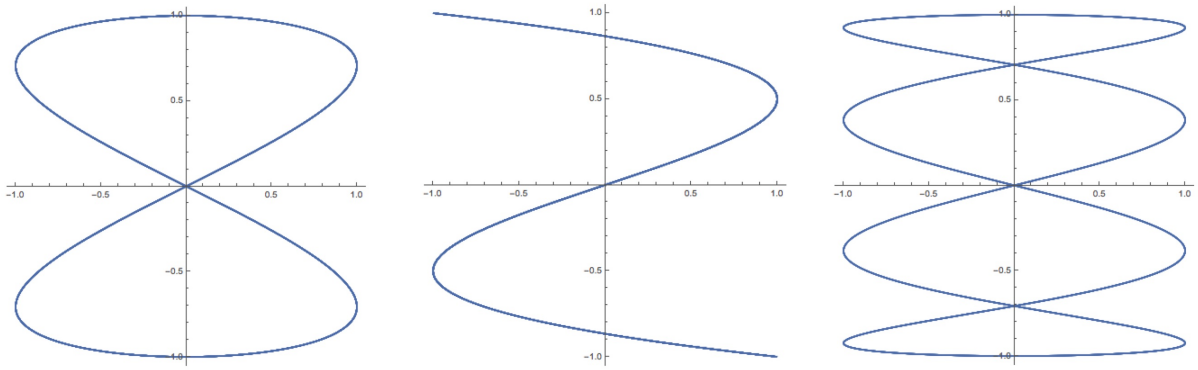
we find the solution to be:

$$x(t) = \sin\omega_1 t, \quad y(t) = C\sin(\omega_2 t + \alpha) \tag{1.28}$$

Here we have scaled both the axis so that the amplitude along x is unity. Also we have chosen the initial time $t = 0$ such that the phase in the x-oscillator is zero. Thus, the solution depends on two frequencies ω_1, ω_2 and two arbitrary constants C, α.

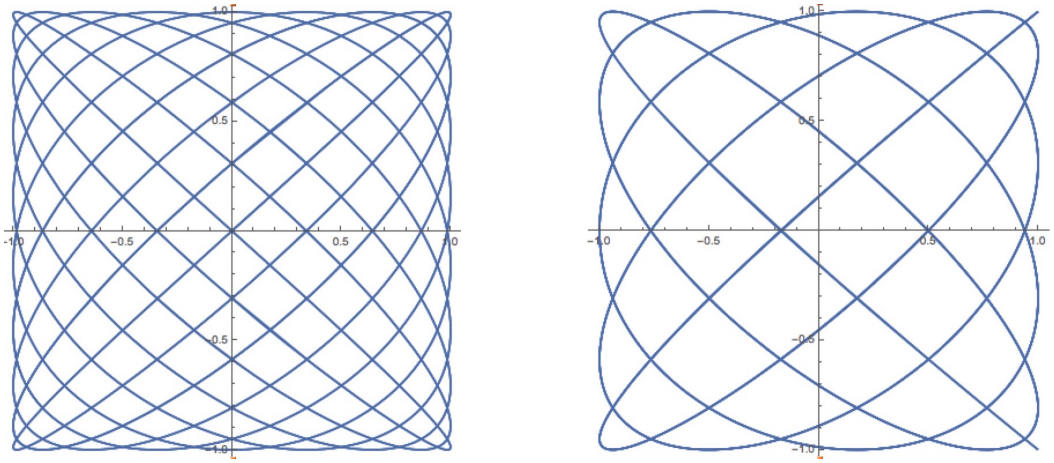
Perhaps surprisingly, the resulting motion is extremely complicated and varies significantly as we vary the frequencies, relative amplitude and relative phase. The resulting trajectories in two dimensions are called "Lissajous figures". One can view them on http://lissajousfigure.netne.net. or http://demonstrations.wolfram.com/LissajousFigures.

Some features of this combined motion are as follows. The shape of the figure depends on the ratio $\frac{\omega_2}{\omega_1}$ as well as the phase difference. Changing the relative amplitude only distorts the figure but does not change its basic form. First consider $\alpha = 0$ and $\frac{\omega_2}{\omega_1} = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$. The resulting figures are as follows:



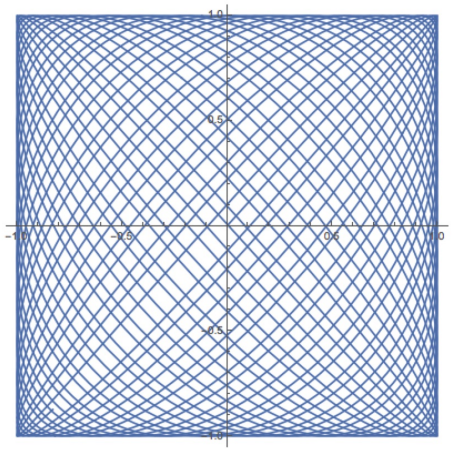
Exercise (level A): Try to understand each of the above trajectories by imagining a particle with the given frequency ratio and tracing out its path by hand. Can you see why the case of $\frac{1}{3}$ looks like an open orbit? What will happen to that case if we change the phase slightly? What if we change it all the way to $\alpha = \frac{\pi}{2}$?

Lissajous figures change dramatically even for slightly varying frequency ratios. For example if $\frac{\omega_2}{\omega_1} = 1$ then we get a straight line/ellipse/circle (depending on the phase difference). However if this ratio is $\frac{9}{10}$ then we get the following figures:



The first one is at $\alpha = 0$ and the second, at $\alpha = \frac{\pi}{2}$. What has happened in the second picture is that the paths are superimposed over themselves so we see fewer lines. This can happen upon varying the phase, as we already saw in the Exercise above.

Another interesting example is when the ratio $\frac{\omega_2}{\omega_1} = 0.906175$ (a randomly chosen decimal with many terms). This is extremely close to 0.9, yet the result at $\alpha = 0$ is very different from the first diagram above, and looks like:



We see that it is the *commensurability* of frequencies that is important in determining how much of the square gets filled by the particle's trajectory.

Recall that in our first example above, we had $\frac{\omega_2}{\omega_1} = \frac{9}{10}$. Inspecting the first figure, we see that there are 9 turning points along the x-axis at $y = \pm 1$, and 10 turning points along the y axis at $x = \pm 1$ (counting the corners in both cases). Let's try to derive this result in a more general situation. We write:

$$\frac{\omega_2}{\omega_1} = \frac{n_2}{n_1}$$

for some integers n_1, n_2 with no common factors. If the periods are T_1, T_2 then let

$$T = n_1 T_1 = n_2 T_2$$

Clearly this is the period of the 2d oscillator (just as it was when we considered superpositions in 1d), because after this time one of the oscillators has completed n_1 full oscillations and the other has completed n_2 oscillations, so both are back at their starting point.

Let us work at phase $\alpha = 0$. During the time T, the x oscillator has reversed itself n_1 times at each end $x = \pm 1$ and the y oscillator has reversed itself n_2 times at each end $y = \pm 1$. Thus we expect to see n_1 turning points on the boundaries at $x = \pm 1$ and n_2 turning points on the boundaries at $y = \pm 1$. Looking back at our example, we had $n_1 = 10$ and $n_2 = 9$, which agrees with our prediction.

1.4 Damped harmonic oscillator

Let us return to one dimension. What is the force law when a *damping force* acts on an object? We expect it to be proportional to the *velocity* since this will have the effect of slowing down a moving particle. Thus, we write:

$$F = -kx - r\dot{x} \tag{1.29}$$

where $r > 0$ is the damping constant. Thus we want to solve the equation:

$$m\ddot{x} + r\dot{x} + kx = 0$$

Let us rewrite it as:

$$\ddot{x} + 2\lambda\dot{x} + \omega^2 x \tag{1.30}$$

where we have defined $\lambda = \frac{r}{2m}$ and, as usual, $\omega = \sqrt{\frac{k}{m}}$.

In Physics, it is useful to understand the *dimensions* of each constant. In the equation $\ddot{x} = -\omega^2 x$, we know that x has dimensions of length and d/dt has dimensions of 1/time. Thus $\ddot{x}$ has dimensions of length/(time)^2. It follows that ω has dimensions of 1/time, which is what we expect for a frequency. Next examine the term proportional to $\lambda\dot{x}$. This too must have the same dimension, length/(time)^2.

For this, λ must have dimensions of 1/time. Thus λ and ω have the same dimensions, so they can be compared with each other. We will see shortly that the behaviour of a damped harmonic oscillator depends on whether $\lambda > \omega, \lambda = \omega$ or $\lambda < \omega$.

First recall that the undamped harmonic oscillator is just the same as the equation above, but without the damping term – i.e., with $\lambda = 0$. The resulting equation was solved by functions $\sin \omega t$ and $\cos \omega t$. Let us now try to re-obtain that solution in a useful way. This will help us eventually solve the damped equation.

It is well-known that sin and cos functions can be expressed in terms of exponentials:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}, \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Also we know how to differentiate an exponential:

$$\frac{d}{d\theta}\left(e^{a\theta}\right) = a\left(e^{a\theta}\right)$$

Let us now solve the undamped oscillator equation:

$$\ddot{x} + \omega^2 x = 0$$

We insert a trial solution Ce^{at}, where a is a constant to be determined while C is an arbitrary constant. Then we find:

$$a^2 + \omega^2 = 0$$

Since ω is a given real number, the second term is positive. If a is real then the first term will also be positive. In that case, the sum can never vanish. So instead, a has to be an imaginary number:

$$a = \pm i\omega$$

Since $a^2 = -\omega^2$, the above equation is satisfied. In this way we find the general solution:

$$x(t) = Ce^{i\omega t} + C^* e^{-i\omega t} \tag{1.31}$$

and it is easy to check that this is the same as an arbitrary linear combination of $\sin \omega t$ and $\cos \omega t$ with real coefficients.

Exercise (level A): Decompose the constant C as well as the complex exponentials in Eq.(1.31) into real and imaginary parts, and show that Eq.(1.31) is the same as the original solution Eq.(1.4) of the simple harmonic oscillator.

Now we want to apply the same method to the damped equation. So we insert $x = Ce^{at}$ into Eq.(1.30). This leads to:

$$a^2 + 2\lambda a + \omega^2 = 0$$

This is a quadratic equation whose solution is:

$$a = -\lambda \pm \sqrt{\lambda^2 - \omega^2} \tag{1.32}$$

As promised earlier, we see that there are three types of solution:

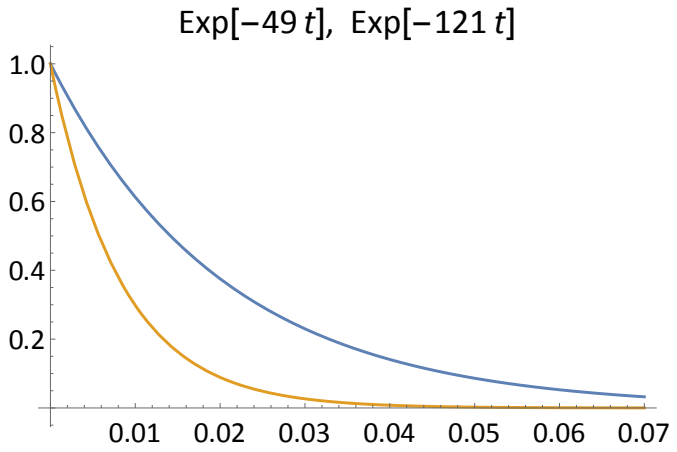
(i) If $\lambda > \omega$ then there are two distinct real values for a and both are negative,
(ii) If $\lambda = \omega$ there is a single, real, negative value for a,
(iii) If $\lambda < \omega$, there are two distinct complex values for a which are complex-conjugate of each other.

We now examine each of these cases in turn.

(i) This case is called *overdamped*. If the two solutions in Eq.(1.32) are denoted $-a_1, -a_2$ (where both a_1 and a_2 are positive real numbers and $a_1 > a_2$) then the motion is given by:

$$x(t) = Ae^{-a_1 t} + Be^{-a_2 t} \tag{1.33}$$

In this case, the particle rapidly slows down and comes to rest. Let us plot this for a concrete case. Suppose the frequency $\omega = 77$ Hz and the damping frequency is $\lambda = 85$ Hz. Then $a_1 = 121$ Hz and $a_2 = 49$ Hz. We see that there are two possible damped motions, one faster and the other slower. This is shown in the following plot:



As usual, the initial conditions will determine which one (or both) of the terms plays a role. We have:

$$x(0) = x_0 = A + B, \qquad \dot{x}(0) = v_0 = -(a_1 A + a_2 B)$$

By a suitable choice of x_0 and v_0, one can select A and B. We are free to choose the initial conditions such that $B = 0$ in Eq(1.33), then only the faster damping will take place. Or we can choose them such that $A = 0$, then only the slower damping will occur. If both terms are present, the slower damping will dominate. We can see this by writing:

$$x(t) = e^{-a_2 t}\left(Ae^{-(a_1 - a_2)t} + B\right)$$

In a very short time, the A term reduces to zero relative to the B term.

Exercise (level A): Consider the overdamped oscillator with $A = B = 1$. Suppose $a_1 = 400$ Hz and $a_2 = 300$ Hz. In how much time will the faster damping fall to $\frac{1}{e}$ of the total? How much time will it take to go below 1% of the total?

(ii) This is called *critically damped*. In this case the solution reduces to:

$$x(t) = Ae^{-\omega t}$$

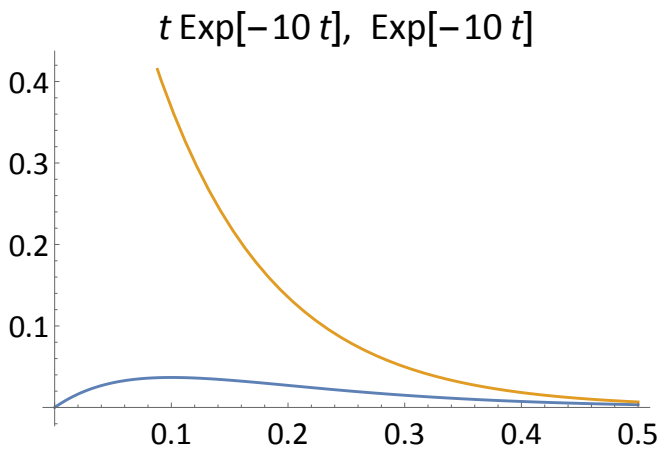
This is puzzling, since a second-order differential equation is supposed to have two solutions. The answer is that there is a second solution which goes like $te^{-\omega t}$. Let us verify it. The function and its first and second derivatives are:

$$\begin{aligned} x(t) &= te^{-\omega t} \\ \dot{x}(t) &= (1 - \omega t)e^{-\omega t} \\ x(t) &= (\omega^2 t - 2\omega)e^{-\omega t} \end{aligned} \tag{1.34}$$

Next, set $\lambda = \omega$ in Eq.(1.30) and insert the above results into it. We find that the equation is satisfied. Thus the general solution in the critically damped case is:

$$x(t) = (A + Bt)e^{-\omega t} \tag{1.35}$$

We can see the two terms in the following plot:



Exercise (level A): Decompose the constant C as well as the complex exponentials in Eq.(1.31) into real and imaginary parts, and show that Eq.(1.31) is the same as the original solution Eq.(1.4) of the simple harmonic oscillator.

Exercise (level B): It is easy to see that if a critically damped oscillator starts at $x_0 = 0$, its motion is of the form $x(t) = te^{-\omega t}$ without the other term. In how much time does this system reach its maximum displacement?

(iii) Finally we turn to the case $\lambda < \omega$ which is called *underdamped*. In this case, $\sqrt{\lambda^2 - \omega^2}$ is imaginary. Let us denote this by $i\omega'$. Then the motion is given by:

$$\begin{aligned} x(t) &= Ae^{(-\lambda + i\omega')t} + Be^{(-\lambda - i\omega')t} \\ &= e^{-\lambda t}(Ae^{i\omega' t} + Be^{-i\omega' t}) \end{aligned} \tag{1.36}$$

Both A and B can be complex numbers, but we must have $B = A^*$ for the above expression to be real.

The expression in brackets is one way of writing the motion of the undamped simple harmonic oscillator. To see this, note that:

$$e^{i\omega t} = \cos\omega t + i \sin\omega t$$

It follows that:

$$\begin{aligned} Ae^{i\omega' t} + Be^{-i\omega' t} &= A(\cos\omega' t + i\sin\omega' t) + B(\cos\omega' t - i\sin\omega' t) \\ &= (A + B)\cos\omega' t + i(A - B)\sin\omega' t \\ &= A'\cos\omega' t + B'\sin\omega' t \end{aligned} \tag{1.37}$$

where $A' = A + B$,

The two key physical features of the underdamped oscillator are: (i) it oscillates with a modified frequency $\omega' = \sqrt{\omega^2 - \lambda^2}$, (ii) its amplitude is modulated by a decreasing exponential. The period of oscillation is given as usual by:

$$T' = \frac{2\pi}{\omega'}$$

but since $\omega' < \omega$ the period is larger than that of the undamped oscillator: $T' > T$.

The exponential decay of amplitude also has a rate, governed by the constant λ, which as we have seen has dimensions of 1/(time). The time period:

$$\tau = \frac{1}{\lambda}$$

is called the *relaxation time*. If we divide the amplitude after a lapse of this time by the amplitude at the beginning, we find:

$$\frac{e^{-\lambda(t+\tau)}}{e^{-\lambda t}} = e^{-\lambda\tau} = \frac{1}{e}$$

Thus the relaxation time is the time taken for the amplitude to fall to $\frac{1}{e}$ of its original value.

Since damped systems eventually come to rest, their energy is not conserved but rather, decreases with time. To find the rate of decrease of energy, note that the energy at any given time is proportional to the *square* of the overall amplitude. Therefore the time taken for the energy to decay to $1/e$ of its original value is $\frac{1}{2\lambda} = \frac{\tau}{2}$. In fact we can write:

$$E(t) = E_0\, e^{-2\lambda t}$$

Now during the time $\frac{\tau}{2}$, the system undergoes $\frac{\tau}{2T'}$ oscillations. Using the relations above, this number can be written as:

$$\frac{\tau}{2T'} = \frac{\omega'}{4\pi\lambda}$$

So this number represents the number of oscillations undergone by the system while its energy decays to $\frac{1}{e}$ of its original value. Engineers like to instead talk about the number of *radians* swept out by the system during this time. Since each oscillation corresponds to 2π radians, the corresponding number is $\frac{\omega'}{2\lambda}$.

For a damped oscillator, we define the "Q-factor" or "quality factor" by:

$$Q = \frac{\omega}{2\lambda} = \frac{\sqrt{km}}{r} \tag{1.38}$$

where in the second expression we have used the constants m, k, r that appear in the original force equation – these are the mass, spring constant and damping constant.

The Q-factor is essentially the ratio of the natural frequency to the constant λ associated to damping. Thus an underdamped oscillator has $Q > \frac{1}{2}$, a critically damped one has $Q = \frac{1}{2}$ and an overdamped one has $Q < \frac{1}{2}$. Note that for very light damping, we can also write $Q \sim \frac{\omega'}{2\lambda}$ in terms of the frequency of the damped oscillator. This is because with light damping, $\omega' \sim \omega$.

Exercise (level B): For a normal (undamped) harmonic oscillator whose amplitude is C, we know that the total energy is $E_0 = \frac{1}{2}m\omega^2C^2$ (see the Exercise after Eq.(1.15)). Now consider the underdamped oscillator, for which we have seen that the energy at any instant is $E = E_0e^{-2\lambda t}$. Show that the energy lost per cycle of the system is $2\lambda ET'$. Show that for light damping, this can also be written in terms of the Q-factor defined above, as $\sim \frac{2\pi E}{Q}$. It follows that:

$$\frac{\text{Energy stored in the system}}{\text{Energy lost per cycle}} \sim \frac{Q}{2\pi}$$

Finally let us describe a nice physical example of a damped oscillator. Suppose an room is equipped with a *door closer*. This is a spring which pulls the door so that it shuts by itself. However if this spring is undamped, the door will simply slam into the wall at full speed (an oscillator has maximum speed as it passes through the origin). To avoid this result we introduce a damping force into the spring by, for example, immersing it in a viscous liquid. If the spring is underdamped the door will still slam into the wall (perhaps a bit more slowly than before). As we increase the damping, we will reach a value at which the door just closes smoothly. That means it attains zero velocity by the time it reaches the origin. This is a critically damped door closer. If we increase the damping further the door will still close smoothly. However it will take longer and longer to do close, which is not convenient. Therefore a critically damped door closer is ideal. Note that we can achieve critical damping in two different ways: by varying the spring constant, or by varying the damping constant. Only the ratio of the two appears in the Q-factor.

1.5 Forced oscillator and resonance

In this section we discuss the harmonic oscillator when perturbed by an external force that is also of sinusoidal form. The external force is taken to have an amplitude F_0 and an angular frequency ω_E. Then we have:

$$F = -kx + F_0 \cos \omega_E t \tag{1.39}$$

After rearranging and dividing by the mass m as usual, we find:

$$\ddot{x} + \omega^2 x - \frac{F_0}{m} \cos \omega_E t = 0 \tag{1.40}$$

Before trying to solve this, let us describe our qualitative expectations for small and large values of the driving frequency ω_E. We should restrict these predictions to very early times, because with an external force the amplitude is able to increase without limit. So we will consider situations where this has not yet happened.

If the frequency is very low, thus $\omega_E \ll \omega$, and if the strength of the driving force is also large, then we expect the system to oscillate slowly at the driving frequency. In this situation the acceleration would be very small so we can neglect the first term of the above equation and find:

$$x(t) \sim \frac{F_0}{m\omega^2} \cos \omega_E t \tag{1.41}$$

We can estimate the error we made by neglecting the acceleration term $\ddot{x}$ in the differential equation. For the above solution, this term is equal in magnitude to $\frac{F_0 \omega_E^2}{m\omega^2} \cos \omega_E t$. Compare this to the other two terms, which are both of the order of $\frac{F_0}{m} \cos \omega_E t$. We see that the neglected term is $\frac{\omega_E^2}{\omega^2}$ times the terms we are keeping, and this is a small quantity in this approximation. In this approximation where the acceleration is neglected, we say the reponse is controlled by the stiffness of the spring.

Suppose now that the driving frequency is very high, $\omega_E \gg \omega$. Then we can neglect the second term $\omega^2 x$ in the equation which is the spring term. To see how this works, first note that upon neglecting the spring term one has the solution:

$$x(t) \sim -\frac{F_0}{m\omega_E^2} \cos \omega_E t \tag{1.42}$$

Then the first and third terms are of order $\frac{F_0}{m} \cos \omega_E t$, while the middle term is of order $\frac{F_0 \omega^2}{m\omega_E^2} \cos \omega_E t$. This is negligible as long as $\frac{\omega}{\omega_E} \ll 1$. In this case, where the stiffness is neglected but acceleration is important, we say the response is controlled by inertia. Notice something that we did not point out earlier: in a harmonic oscillator, the acceleration is always *out of phase* with the motion. This is simply the minus sign in $\ddot{x} = -\omega^2 x$. So in the rapidly driven case, the particle moves at the driving frequency but out of phase with it.

Exercise (level A): Carefully verify the statements in the last few paragraphs showing that the neglected term in the given limits is genuinely small. This justifies the approximations made.

In the above discussion we found that a driven oscillator moves at the frequency ω_E of the driving force both for small ω_E and for large ω_E. Could it be true that it does this for all values of ω_E? To find out, let us insert a trial solution:

$$x(t) = C_E \cos \omega_E t \tag{1.43}$$

into the full equation Eq.(1.40). We find:

$$-C_E \omega_E^2 \cos \omega_E t + C_E \omega^2 \cos \omega_E t - \frac{F_0}{m} \cos \omega_E t = 0 \tag{1.44}$$

from which we have:

$$C_E = \frac{F_0/m}{\omega^2 - \omega_E^2} \tag{1.45}$$

In the limits $\omega_E \ll \omega$, $\omega_E \gg \omega$ we see that this perfectly reproduces Eqs.(1.41) and (1.42) respectively. In the latter case we even get the negative sign (out of phase behaviour) that we already discussed!

If the driving frequency ω_E is exactly equal to the natural frequency ω then the amplitude appears to diverge. What does this mean? Another interesting feature is that as ω_E crosses from being greater to being smaller than ω, the minus sign appears.

Before addressing both these questions, let us highlight a puzzle. The solution Eq.(1.43), together with the constant specified in Eq.(1.45), seems to be completely determined – it has no free parameters. This should not be the case. We have mentioned a few times that the solution of a second-order differential equation should have two free parameters, but so far we have not even found one free parameter! Another way of seeing this problem is that our trial solution does not satisfy arbitrary initial conditions. Specifically at $t = 0$ we find $x(0) = C_E$ with C_E given by Eq.(1.45). Also the initial velocity is $\dot{x}(0) = 0$. This is incompatible with, for example, the initial condition $x(0) = 0$ and non-zero initial velocity. That seems very strange.

The resolution is simple. Suppose we add to our trial solution Eq.(1.43) an *arbitrary* solution of the *unforced system*. Then the total is:

$$x(t) = x_1(t) + x_2(t) = C_E \cos\omega_E t + C\cos(\omega t + \alpha) \tag{1.46}$$

The first term x_1 has a fixed amplitude given by Eq.(1.45) and the fixed frequency of the driving force. It has no arbitrary constants. However the second term x_2 has an arbitrary amplitude C and an arbitrary phase α, and it carries the natural frequency of the oscillator. We have chosen x_1 so that it satisfies the inhomogeneous (forced) equation, while x_2 is the general solution of the homogeneous (unforced) equation. Inserting this sum into the equation we have:

$$(\ddot{x}_1 + \ddot{x}_2) + \omega^2(x_1 + x_2) - \frac{F_0}{m}\cos\omega_E t = \left(\ddot{x}_1 + \omega^2 x_1 - \frac{F_0}{m}\cos\omega_E t\right) + (\ddot{x}_2 + \omega^2 x_2) = 0 \tag{1.47}$$

The first bracket vanishes because x_1 is a solution of the full (inhomogeneous) equation, while the second term vanishes because x_2 satisfies the corresponding homogeneous equation. Thus the equation is satisfied. The full solution has two arbitrary constants, so we now have the most general solution of the forced oscillator.

With the general solution there is no longer a problem in satisfying any arbitrary initial conditions. Suppose $x(0) = x_0$ and $\dot{x}_0(0) = v_0$. From the general solution we find:

$$C_E + C\cos\alpha = x_0, \quad -\omega C\sin\alpha = v_0 \tag{1.48}$$

We can solve these two equations to determine C, α in terms of x_0, v_0:

$$C = \sqrt{(x_0 - C_E)^2 + \left(\frac{v_0}{\omega}\right)^2}, \quad \alpha = \tan^{-1}\left(\frac{v_0}{\omega(C_E - x_0)}\right) \tag{1.49}$$

Exercise (Level A): Verify the above equation. Also check that if we send the driving force to zero by taking $F_0 = 0$, the above equations are precisely the ones that determine the amplitude and phase of a *simple* harmonic oscillator in terms of the initial position and velocity.

Exercise (level A): Suppose we want to perform an experiment that displays only the $C_E\cos\omega_E t$ behaviour of the forced oscillator. What initial conditions on position and velocity should we choose so that the other term ($C\cos\omega t$) vanishes?

Exercise (level A): Suppose we take an oscillator at rest at the equilibrium position, and apply a sinusoidal force to it. Find the solution with these initial conditions. Expand it around $t = 0$ upto quadratic order in t and show that the particle just behaves as if it has experienced a constant force in this period.

Exercise (Level B): Consider the general solution Eq.(1.46) of the forced harmonic oscillator and find its behaviour in the limits $\omega_E \ll \omega$ and $\omega \ll \omega_E$.

Exercise (Level C): Look up the mathematical theorem that for a linear inhomogeneous differential equation, the most general solution is the sum of (i) *any* particular solution of the equation, and (ii) the general solution of the corresponding *homogeneous* equation. Verify that the solution Eq.(1.46) for the forced oscillator is an example of this theorem.

We still need to understand the behaviour when ω_E is close to ω. In this limit C_E diverges. This means that the component of the oscillator's motion which has the driving frequency develops a huge amplitude. Such behaviour suggests we have neglected some physical aspect of the system, and indeed it is due to our neglect of damping. Hence we re-introduce this in the next subsection.

Exercise (level B): A mass of 4 kg is suspended from a spring that has a force constant of 200 N/m. The system is undamped and is subjected to a harmonic force with a frequency of 10 Hz, which results in a forced-motion amplitude of 2 cm. Determine the maximum value of the force.

1.6 Forced harmonic oscillator with damping

Now we consider the equation:

$$\ddot{x} + 2\lambda\dot{x} + \omega^2 x - \frac{F_0}{m}\cos\omega_E t = 0 \tag{1.50}$$

As before, we only need to find *one* particular solution for the forced equation. Then we can add a general solution to the un-forced equation to get the complete solution.The latter is something we have already done when we studied the damped oscillator. So let us focus on finding a particular solution when there is both forcing and damping.

If we insert $x(t) = C_E \cos\omega_E t$ as we did in the undamped case, this time we don't get a solution. This is because the forcing term is proportional to $\dot{x}$ which behaves as $\sin\omega_E t$. It turns out that if we generalise the particular solution to $x(t) = C_E \cos(\omega_E t + \alpha_E)$ then we will find a solution to the equation that determines both C_E and α_E. However the mathematics becomes very tedious and is left as an exercise. Instead, we resort to the exponential method, which involves complex numbers. Let us think of the above equation as:

$$\ddot{x} + 2\lambda\dot{x} + \omega^2 x = \frac{F_0}{m}\,\mathrm{Re}\left(e^{i\omega_E t}\right) \tag{1.51}$$

Here we used the fact that

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \mathrm{Re}\left(e^{i\theta}\right)$$

and Re is the instruction to take the real part. Now we similarly insert:

$$x(t) = C_E \mathrm{Re}\left(e^{i(\omega_E t + \alpha_E)}\right)$$

Since both sides of the above equation have the Re instruction, we can temporarily drop it and impose it at the very end. Then we have:

$$\left[C_E(\omega^2 - \omega_E^2) + 2i\lambda C_E \omega_E\right] e^{i\omega_E t} e^{i\alpha_E} = \frac{F_0}{m} e^{i\omega_E t} \tag{1.52}$$

Now we can cancel out the factor $e^{i\omega_E t}$ from both sides, and also divide both sides by $e^{i\alpha_E}$, to get:

$$\Big[C_E(\omega^2-\omega_E^2)+2i\lambda C_E\omega_E\Big]=\frac{F_0}{m}e^{-i\alpha_E}=\frac{F_0}{m}\Big(\cos\alpha_E-i\sin\alpha_E\Big) \tag{1.53}$$

Equating the real and imaginary parts on both sides, we have two equations:

$$\begin{aligned} C_E(\omega^2-\omega_E^2)&=\frac{F_0}{m}\cos\alpha_E \\ 2\lambda C_E\omega_E&=-\frac{F_0}{m}\sin\alpha_E \end{aligned} \tag{1.54}$$

These two equations determine C_E and α_E. First, square both equations and add them to get:

$$C_E^2\Big[(\omega^2-\omega_E^2)^2+(2\lambda\omega_E)^2\Big]=\left(\frac{F_0}{m}\right)^2 \tag{1.55}$$

Hence:

$$C_E=\frac{F_0/m}{\sqrt{(\omega^2-\omega_E^2)^2+(2\lambda\omega_E)^2}} \tag{1.56}$$

Comparing this with Eq.(1.45), we see that the above equation reduces to it when $\lambda=0$. However when $\lambda\neq 0$, i.e. in the presence of damping, there is no longer a divergence when $\omega_E=\omega$. Instead, at least for small damping, there will be a smooth peak when $\omega=\omega_E$. Finally, we have:

$$\alpha_E=-\tan^{-1}\frac{2\lambda\omega_E}{\omega^2-\omega_E^2} \tag{1.57}$$

We see that the phase is fixed by a combination of the damping frequency, the driving frequence and the natural frequency. It is nonzero as long as there is damping. Note that this phase is not a free parameter (nor is the amplitude). This is because we are looking at a particular solution in the presence of forcing. The general solution will have additional terms (which we have already studied, and whose nature depends on whether the oscillator is underdamped, critically damped or overdamped). These terms, corresponding to the general solution of the homogeneous equation, have an arbitrary amplitude and phase as always. These provide the two parameters to be fixed by initial conditions.

To conclude, we have shown that a particular solution for a damped, forced oscillator is given by:

$$x(t)=C_E\cos(\omega_E t+\alpha_E) \tag{1.58}$$

with C_E,α_E given above.

A comment on the phase: only for those interested. This is not mandatory for everyone. Note that α_E is equivalent to $\alpha_E+2\pi$, since $\cos(\theta+2\pi)=\cos\theta$. Hence we need to fix a convention for α_E. A standard choice is $-\pi<0\leq\pi$. Now there is an apparent subtlety in calculating α_E for the present problem. We determined it via the equation:

$$\tan\alpha_E=-\frac{2\lambda\omega_E}{\omega^2-\omega_E^2}$$

But $\tan(\alpha_E+\pi)=\tan\alpha_E$, so this equation can only determine α_E upto the addition of a multiple of π, rather than 2π. However we can get more information directly from Eq.(1.54). Once we choose $\lambda,\omega,\omega_E,C_E$ to be all positive numbers, then we see that $\sin\alpha_E$ is always negative. That fixes $-\pi<\alpha_E\leq 0$. Next we check the sign of $\cos\alpha_E$ and narrow down the range of α_E, always remaining within the above region. For $\omega_E<\omega$, $\cos\alpha_E$ is positive which tells us that $-\frac{\pi}{2}<\alpha_E\leq 0$. On the other hand for $\omega_E>\omega$, $\cos\alpha_E$ is negative and this means that $-\pi<\alpha_E\leq-\frac{\pi}{2}$. It follows that as ω_E varies from 0 to ∞, the phase decreases monotonically from 0 to $-\pi$. This is shown in the figure.

At resonance, the phase is $-\frac{\pi}{2}$. Note that once we chose $-\pi < \alpha_E \leq \pi$ and also choose fixed signs for all the parameters, everything is precisely determined with no further assumptions.

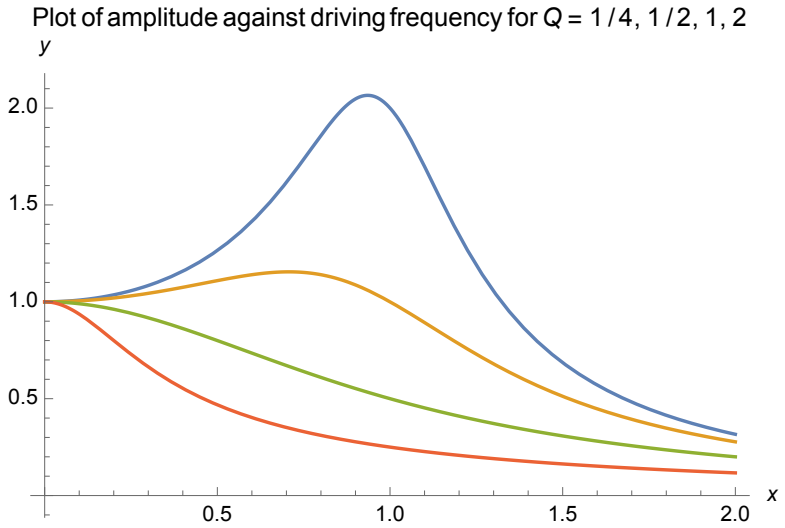
The fact that $\sin\alpha_E$ is negative also resolves the question of the sign of power input to the system, as we will see in a later subsection.

For the general solution we simply have to add the two-parameter solution of the unforced, damped oscillator to this. However, now there is a new feature. We know that an unforced, damped oscillator decays after a long enough time and comes to rest (regardless of whether it is underdamped or overdamped). In contrast, the particular solution above persists for all time, as one can see from the function. It follows that the late-time behaviour of the forced, damped oscillator is given *completely* by the above expression, without any other term. The terms we would have added, which exist only for a short period of time, are called *transients*. So the general behaviour of a forced, damped system is that initially there are two types of oscillations (one at the forcing frequency and one at the natural frequency) but after some time the latter die out and the system oscillates at precisely the forcing frequency as in Eq.(1.58). Thus, this solution is called the *steady state* solution.

Let us try to sketch the amplitude C_E as a function of ω_E in Eq.(1.56). Physically, this means we are varying the driving frequency for fixed natural frequency and damping. To make this easier, let us re-define $y = \frac{m\omega^2 C_E}{F_0}$. Next, define $\frac{\omega_E}{\omega} = x$ and recall that $\frac{\omega}{2\lambda} = Q$, the Q-factor. Then the equation becomes:

$$y = \frac{1}{\sqrt{(1-x^2)^2 + \frac{x^2}{Q^2}}}$$

This can be sketched on the $x-y$ plane, varying Q across some different values. Recall that $Q < \frac{1}{2}$ is the overdamped case and $Q > \frac{1}{2}$ is overdamped. So it is reasonable to choose, say, $Q = \frac{1}{4}, \frac{1}{2}, 1, 2$ to get a sample of the possible behaviours. The answer looks like this:

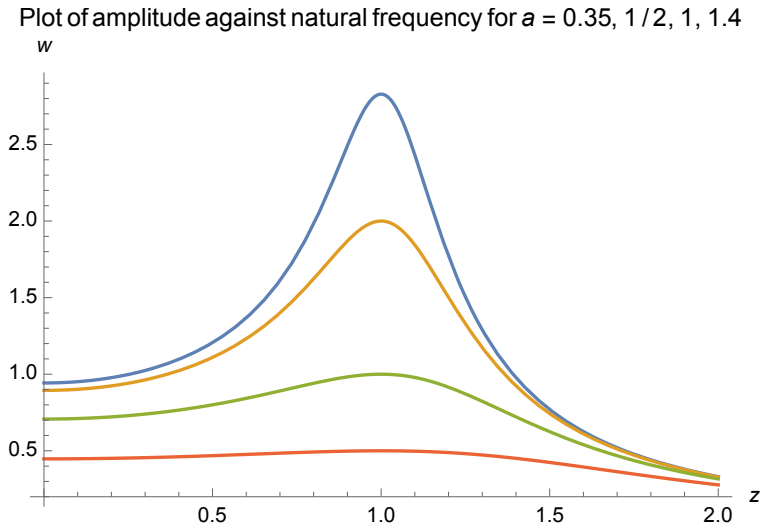


Notice that for the underdamped case the figure has a turning point, at least for large Q. This is a sign of resonance.

Alternatively, we can rewrite the above equation as a function of $z = \frac{\omega}{\omega_E}$. In this case we are varying the natural frequency, for fixed damping and fixed driving frequency. This time we will define $w = \frac{m\omega_E^2 C_E}{F_0}$ and $a = \frac{2\lambda}{\omega_E}$. The function is then:

$$w = \frac{1}{\sqrt{(1-z^2)^2 + a^2}}$$

This function always has a peak at $z = 1$, i.e. when the natural frequency equals the driving frequency.

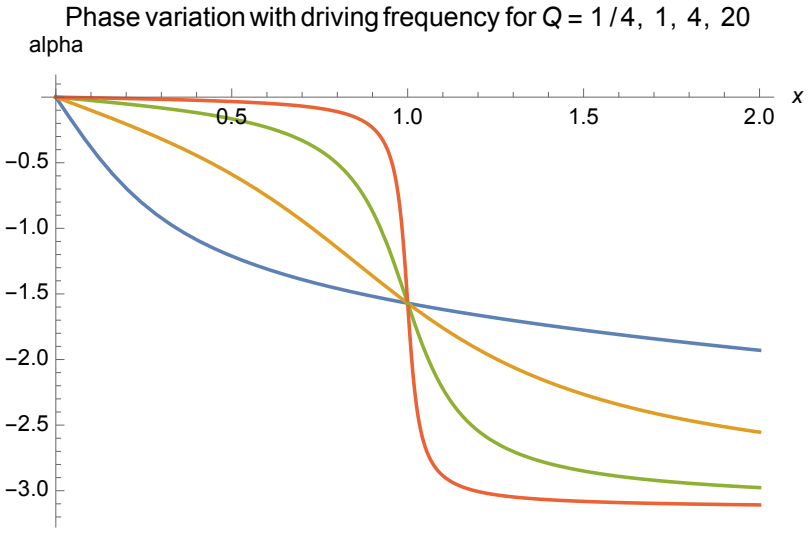


exercise teaches us the lesson that physical information depends very much on what we keep fixed and what we vary! But in any case, for very large Q-factor (very low damping) there is not much difference between the two cases, and we find an extremely sharp resonance at $\omega = \omega_E$.

So far we were only discussing the amplitude. Now we return to the phase. The goal is clear. If we use the x variable with ω fixed, then we have:

$$\alpha = -\tan^{-1}\frac{1}{Q}\frac{x}{1-x^2}$$

We have already determined the phase conventions as x varies from 0 to ∞. Let us see how this looks for different values of Q.



We see that as Q becomes large (less damping), the jump from $\alpha_E = 0$ to $\alpha_E = -\pi$ becomes more sudden at resonance. Damping smoothens this transition.

A simple example of a forced damped oscillator is a simple pendulum whose point of support executes simple harmonic motion. The motion of the point of support provides the driving force. To create such an object experimentally, we need a motor that vibrates a block in harmonic motion, then we suspend the pendulum from the block. But there is a simpler way. Suppose we take an extremely heavy pendulum bob. This will oscillate in simple harmonic motion like any pendulum. But now we suspend a much lighter bob from the heavy one. For the light bob, the point of suspension is the heavy bob. The only concern could be that the motion of the light bob influences that of the heavier one, but precisely because it is light, we don't expect this. So such an apparatus can experimentally realise a forced oscillator.

Exercise (level A): In the three plots above, make sure you understand which colour corresponds to which value of Q or a.

Exercise (level B): Instead of using complex exponentials, try to solve Eq.(1.50) by inserting $\cos(\omega_E t + \alpha_E)$ as a trial solution. Show that this leads to:

$$C_E(\omega^2 - \omega_E^2)\cos(\omega_E t + \alpha_E) - 2\lambda\omega_E C_E \sin(\omega_E t + \alpha_E) - \frac{F_0}{m}\cos\omega_E t = 0$$

Expanding the cos and sin functions, convert this to:

$$\left[C_E(\omega^2 - \omega_E^2)\cos\alpha_E - 2\lambda\omega_E C_E \sin\alpha_E - \frac{F_0}{m}\right]\cos\omega_E t$$
$$= \left[C_E(\omega^2 - \omega_E^2)\sin\alpha_E + 2\lambda\omega_E C_E \cos\alpha_E\right]\sin\omega_E t \quad (1.59)$$

Since sin and cos cannot be equal for all values of the argument, the cofficients on both sides must vanish. Use this to solve for C_E, α_E and check that you get the same answers as we got using complex exponentials.

1.7 Power absorbed by a forced oscillator

Consider first the forced, undamped oscillator moving according to:

$$x(t) = C_E \cos\omega_E t$$

The instantaneous power input is given by the force times the velocity:

$$P = Fv$$

Now $F = F_0\cos\omega_E t$ and $v = \dot{x} = -\omega_E C_E \sin\omega_E t$. Hence,

$$P = -\omega_E C_E F_0 \sin\omega_E t\cos\omega_E t = -\frac{1}{2}\omega_E C_E F_0 \sin 2\omega_E t$$

If we look at the function $\sin 2\omega_E t$, it is positive for $0 < t \leq \frac{\pi}{2\omega_E}$ and then turns negative. But this is just a *quarter-cycle*. Thus power if fed into the system for a quarter-cycle, then taken out for a quarter-cycle and so on. The average power input over a cycle is zero (in fact it is zero even over a half-cycle).

With damping, things are different. Now we have $x = C_E\cos(\omega_E t + \alpha_E)$ and the phase shift α_E will make all the difference. One has:

$$P = -\omega_E C_E F_0 \cos\omega_E t \sin(\omega_E t + \alpha_E) \quad (1.60)$$

But $\sin(\omega_E t + \alpha_E) = \sin\omega_E t\cos\alpha_E + \cos\omega_E t\sin\alpha_E$. Inserting this into the above, we have to average two terms: one proportional to $\sin\omega_E t\cos\omega_E t$ which, as we have seen, is zero, and the other proportional to $\cos^2\omega_E t$. Now this function is always positive (because it is a square) so its average can never be zero! Indeed, over a full cycle the average of $\cos^2\omega_E t$ is $\frac{1}{2}$ (see the Exercise below). Thus the average power input during a cycle is:

$$\bar{P} = -\frac{1}{2}\omega_E C_E F_0 \sin\alpha_E \quad (1.61)$$

From Eq.(1.54) one sees that $\sin\alpha_E$ is negative, therefore the power input is positive as it should be. Since the amplitude C_E becomes maximum at resonance (assuming the underdamped case), the power input is also maximal at resonance.

1.8 Electrical circuit as an oscillator

One of the most fascinating aspects of nature is that the concepts of harmonic oscillation, damping, forcing and resonance are *universal* – they occur in widely different contexts but follow the same basic equations and principles. Here we will discuss how these concepts apply to an electrical circuit.

Let us start with the simplest case: an LC circuit. This consists of an inductance coil and a capacitor connected in series. We know that if I is the current in the circuit and V_L is the voltage across the coil, then

$$V_L = L\frac{dI}{dt}$$

Recall that current is the rate of flow of charge, which we can write as $I = \frac{dq}{dt}$. Thus we can write the above equation as:

$$V_L = L\frac{d^2q}{dt^2} = L\ddot{q}$$

We also know that if a voltage is applied across a capacitor of capacitance C, then:

$$V_C = \frac{q}{C}$$

where q is the charge stored in the capacitor.

We can relate these two by the fact that if no external voltage is applied, the total voltage drop across the entire circuit is zero: $V_L + V_C = 0$. It follows that:

$$L\ddot{q} + \frac{1}{C}q = 0$$

This is precisely the equation of a simple harmonic oscillator! Indeed we can rewrite it as:

$$\ddot{q} + \omega^2 q = 0$$

where

$$\omega = \sqrt{\frac{1}{LC}}$$

Physically it is not hard to understand the oscillations. Suppose we start with a charged capacitor. As it tries to discharge, it sends a current through the inductance. This resists the buildup of current, by Lenz's law. After passing through the inductance the charges pile up on the other plate of the capacitor, creating a voltage in the opposite direction. Then the charges are pushed back through the circuit. In this way an LC circuit will, in principle, oscillate forever. We see that capacitance acts as a "stiffness" while inductance acts as "inertia".

We do not need to do any work to understand this system further. All the mathematics we have developed simply applies. The charge in the circuit will oscillate as:

$$q(t) = C_E \cos(\omega t + \alpha_E)$$

The only thing we need to do is understand the different possible initial conditions.

Exercise (level B): Try to understand the conditions under which any desired value of initial charge and initial current can be achieved.

Next, suppose we introduce a resistance into the circuit. The voltage drop across a resistor is $V_R = IR = R\dot{q}$. Adding this into the equation we have:

$$L\ddot{q} + R\dot{q} + \frac{1}{C}q = 0$$

This is a damped harmonic oscillator! The resistor provides the damping. We easily see that the damping constant λ is $\frac{R}{2L}$. The system is now an RLC circuit. Its Q-factor is:

$$Q = \frac{\omega}{\lambda} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

So if $Q > \frac{1}{2}$ or $Q < \frac{1}{2}$ then we have an underdamped or overdamped circuit. In the first case the charges will oscillate for a while and die down, while in the second case they will die down without oscillating.

Finally, we can apply an AC current of a frequence ω_E to the circuit. Thus we had a term $V_0 \cos \omega_E t$ to the original equation. So we have a forced oscillator and V_0 plays the role of the amplitude of the driving force, F_0, in our previous analysis. Our previous analysis, including the possibility of resonance, applies in a straightforward way. Moreover it is easy to tune either ω_E or ω, the first one by changing the AC frequency and the second by varying L or C.

2 Coupled Oscillations

In this section we discuss coupling between harmonic oscillators. This is not the same as superposition, as we will see in a moment. The problem of coupled oscillators is of fundamental importance in physics. This is the concept underlying wave behaviour. It is relevant in many-body condensed matter physics, quantum field theory and string theory, however in all these cases one also has to introduce quantum mechanics. Here we will only study what happens in the context of classical mechanics.

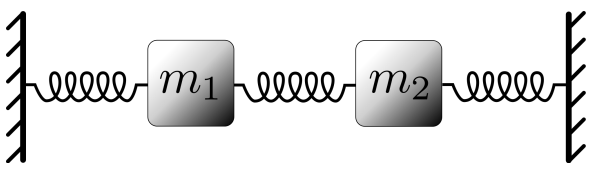
2.1 Two identical coupled oscillators

Coupling of two oscillators means that while each one is connected to an independent centre by a spring, the two are also connected to each other by a spring. This makes three springs in all! The problem can be realised using two simple pendulums with separate points of suspension and then joining them with a spring.

Let us start with two independent oscillators of the same mass m. The first one has coordinate x_1, the second one has coordinate x_2 (do not confuse this with the x_1, x_2 that we used earlier to specify the particular and general solutions for one oscillator). Each one is connected by a spring of the same stiffness k to their equilibrium positions, $x_1 = 0, x_2 = 0$. Thus we have:

$$\begin{aligned} m\ddot{x}_1 &= -kx_1 \\ m\ddot{x}_2 &= -kx_2 \end{aligned} \tag{2.1}$$

Each oscillator has its own solution and there is nothing very interesting about this system! It becomes interesting when we connect these two masses by a spring which is at equilibrium precisely when both masses are at rest at their respective origins. This is illustrated in the figure.



If this new spring has a stiffness k_{12}, then the force due to it on each of the masses is proportional to $k_{12}(x_1 - x_2)$. What is the sign? Suppose $x_1 > x_2$. This means the distance between the masses is less than the equilibrium length of the connecting spring. So this spring is compressed. To restore

itself to equilibrium, it forces particle 1 to the left and particle 2 to the right. It follows that the new equations are:

$$\begin{aligned} m\ddot{x}_1 &= -kx_1 - k_{12}(x_1 - x_2) \\ m\ddot{x}_2 &= -kx_2 + k_{12}(x_1 - x_2) \end{aligned} \tag{2.2}$$

This is the coupled simple harmonic oscillator. To find the motion of the system we just need to solve this pair of coupled differential equations. In the present example this is quite easy as we will see. Just add the two equations, to get:

$$m(\ddot{x}_1 + \ddot{x}_2) = -k(x_1 + x_2) \tag{2.3}$$

Next we take the difference of the two equations, to get:

$$m(\ddot{x}_1 - \ddot{x}_2) = -k(x_1 - x_2) - 2k_{12}(x_1 - x_2) = -(k + 2k_{12})(x_1 - x_2) \tag{2.4}$$

Notice that the first equation depends only on $x_1 + x_2$ while the second only depends on $x_1 - x_2$. So we can define:

$$X_1 = x_1 + x_2, \quad X_2 = x_2 - x_2$$

It is clear that X_1 is proportional to the centre-of-mass position $\frac{x_1+x_2}{2}$ while X_2 is the position of one mass relative to the other. Now the equations read:

$$\ddot{X}_1 = -\omega_1^2 X_1, \qquad \ddot{X}_2 = -\omega_2^2 X_2 \tag{2.5}$$

where

$$\omega_1^2 = \frac{k}{m}, \quad \omega_2^2 = \frac{k + 2k_{12}}{m}$$

Notice that just be changing variables, we have found two *decoupled* equations! The price we pay is that neither X_1 nor X_2 are coordinates of any physical object. Instead they are called *normal modes* of the system. Normal modes are coordinates in terms of which a coupled system appears decoupled. Once we find normal modes, the mathematics becomes easy. However the physical interpretation can be subtle.

In the present example, the solutions are:

$$X_1(t) = C_1 \cos(\omega_1 t + \alpha_1), \qquad X_2(t) = C_2 \cos(\omega_2 t + \alpha_2) \tag{2.6}$$

The angular frequencies ω_1, ω_2 have been determined above and $C_1, \alpha_1, C_2, \alpha_2$ are arbitrary constants. It follows that we can excite each of the normal modes independently of the other, and it only remains to interpret these normal modes. For this, we need to understand what happens if $X_1(t)$ oscillates while X_2 remains at rest, and the other way around.

The first one is simple. If $X_2(t) = 0$, it means the two particles maintain a constant distance from each other. Moreover, this is the distance at which the spring k_{12} is in equilibrium. This means that both masses move together *as if* they were a rigid body. This is called "centre-of-mass motion". During this motion the middle spring always stays in equilibrium, so its stiffness does not contribute to the frequency. This is why the frequency of this motion is ω.

The second independent motion arises when $X_1(t) = 0$ while $X_2(t)$ oscillates with angular frequency ω_2. In this motion, the centre-of-mass is fixed so the two masses simultaneously move towards each other, and then away from each other. When moving towards each other, the outer springs get stretched while the middle spring is compressed. The stiffness of all three springs reacts to push the masses apart. When they move apart, the middle spring is stretched and the outer ones compressed. Again their combined stiffness comes into play. This explains why the frequency ω_2 depends on both k and k_{12}. The general motion of the system is a combination of both normal modes, with a free choice of two amplitudes and two phases. In general it may look complicated but since it is just a decoupled combination, the mathematics is very simple and completely predicts the motion.

The energy in each normal mode must follow the standard formula:

$$\begin{aligned} E_1 &= \frac{1}{2}M_1\dot{X}_1^2 + \frac{1}{2}M_1\omega_1^2 X_1^2 \\ E_2 &= \frac{1}{2}M_2\dot{X}_2^2 + \frac{1}{2}M_2\omega_2^2 X_2^2 \end{aligned} \tag{2.7}$$

except that we have not yet determined the effective masses M_1, M_2 of the normal modes. The total energy is $E_1 + E_2$. To find the masses, rewrite the above kinetic terms as:

$$\frac{1}{2}M_1(\dot{x}_1 + \dot{x}_2)^2 + \frac{1}{2}M_2(\dot{x}_1 - \dot{x}_2)^2$$

The coupling of oscillators is only through stiffness, not through inertia. So the kinetic energy of the coupled system must be the same as that of the uncoupled one:

$$\frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2)$$

Comparing the two expressions above, we see that $M_1 = M_2 = \frac{1}{2}m$. It follows that the total energy is:

$$E_1 + E_2 = \frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2) + \frac{1}{2}\frac{m(\omega_1^2 + \omega_2^2)}{2}(x_1^2 + x_2^2) + m(\omega_1^2 - \omega_2^2)x_1x_2 \tag{2.8}$$

Now in the limit $k_{12} \to 0$, we have $\omega_1 = \omega_2 = \omega$. In this limit the last term vanishes and the potential energy becomes the standard one, $\frac{1}{2}m\omega^2x^2$ for each oscillator. We see that the last term represents the coupling of energy between the two oscillators. Due to this coupling, the energy in the system can get transferred back and forth between the two masses as time progresses. This is similar to beats, and can be seen by solving the following exercise.

Exercise (level B): Consider the above coupled oscillator with initial conditions $x_1(0) = 2a, x_2(0) = 0, \dot{x}_1(0) = \dot{x}_2(0) = 0$. Find the total energy of the system. Solve for the motion with these initial conditions. Show that the amplitude of oscillation of the first mass builds up to a maximum while the second mass has a small amplitude, and then after some time the first mass has a small amplitude while that of the second mass builds up. In this way energy gets "exchanged" between the two oscillators.

A nice physical realisation of the above problem is to take two identical simple pendulums of the same length, and join the bobs by a spring. The centre-of-mass mode will have both bobs moving in phase. The other mode has them alternately coming together and flying apart.

The above example was particularly simple because the two separate oscillators were identical. What if they are not? Let us keep both masses equal to m but take two independent stiffnesses k_1, k_2. It is easy to see that the coupled equations are now:

$$\begin{aligned} m_1\ddot{x}_1 &= -k_1x_1 - k_{12}(x_1 - x_2) \\ m_2\ddot{x}_2 &= -k_2x_2 + k_{12}(x_1 - x_2) \end{aligned} \tag{2.9}$$

This time, taking the sum and difference does not serve to decouple them. We have to *diagonalise* the system to find the normal modes.As in the previous case, after diagonalisation we will have two decoupled modes. It is just a slightly tedious exercise to find them. Hence this is left as an optional exercise for interested students.

Exercise (level C): Diagonalise the above general system of two coupled oscillators and find the normal modes. When $m_1 = m_2$ you only need to diagonalise the RHS and this if fairly easy. However if $m_1 \neq m_2$ then the situation is more complicated.

2.2 N identical coupled oscillators

The above problem has an easy generalisation to n coupled oscillators, where each one is coupled to the next one on its left and right by a spring. We take all masses to be equal to m and all springs to be identical, with stiffness k. The positions of the masses are $x_1(t), x_2(t), \cdots x_n(t)$ and the equilibrium position is $x_i = 0$ for all i. There are $N+1$ springs in all. The ith mass experiences a force of magnitude $k(x_{i-1} - x_i)$ from the spring on its left, and $k(x_i - x_{i+1})$ from the spring on its right. It is easy to see that the sign is positive for the former and negative for the latter. Thus:

$$\begin{aligned} m\ddot{x}_1 &= -kx_1 - k(x_1 - x_2) \\ m\ddot{x}_2 &= k(x_1 - x_2) - k(x_2 - x_3) \\ m\ddot{x}_3 &= k(x_2 - x_3) - k(x_3 - x_4) \\ &\cdots \\ m\ddot{x}_{n-1} &= k(x_{n-2} - x_{n-1}) - k(x_{n-1} - x_n) \\ m\ddot{x}_n &= k(x_{n-1} - x_n) - kx_n \end{aligned} \tag{2.10}$$

Notice that if we set $x_3, x_4, \cdots x_N = 0$ then we recover the pair of coupled oscillators that we already studied in the previous subsection (in the special case $k_{12} = k$).

We can rewrite the above as a matrix equation:

$$m \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ \cdots \\ x_{n-1} \\ x_n \end{pmatrix} = -k \begin{pmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & 1 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ \cdots \\ x_{n-1} \\ x_n \end{pmatrix} \tag{2.11}$$

The matrix on the RHS is a very famous one in mathematics: it is called the "Cartan matrix of the Lie algebra $\mathcal{A}_N$". The name is not relevant for us, however it will help us find a useful result. What we would now like to do is take suitable linear combinations of $x_1, x_2, \cdots, x_N$ such that we find N decoupled normal modes. This is the same as finding the *eigenfunctions and eigenvalues* of the stiffness matrix, which is just ω^2 times the Cartan matrix, where $\omega = \sqrt{\frac{k}{m}}$. Finding these by hand is difficult, but we can look up the result. It is known that the eigenvalues of this Cartan matrix are:

$$\lambda_n = 4\sin^2 \frac{\pi n}{2(N+1)}, \quad n = 1, 2, \cdots N \tag{2.12}$$

Thus, once we diagonalise the matrix through appropriate linear combinations, we will have N decoupled normal modes $X_1, X_2, \cdots X_N$ satisfying:

$$\ddot{X}_n = -\omega_n^2 X_n = -\omega^2 \lambda_n X_n$$

Thus the normal mode frequencies are:

$$\omega_n = \omega\sqrt{\lambda_n} = 2\omega \sin \frac{\pi n}{2(N+1)}, \quad \omega = \sqrt{\frac{k}{m}} \tag{2.13}$$

We can verify this for the case of two coupled oscillators, which we have already solved. Putting $N = 2$, the frequencies ω_1, ω_2 are $(2\omega \sin \frac{\pi}{6}, 2\omega \sin \frac{\pi}{3})$ which is the same as $(\omega, \sqrt{3}\omega)$. Previously we had found $\omega_1 = \omega$ and $\omega_2 = \sqrt{\frac{k+2k_{12}}{m}}$. With $k_{12} = k$, we have $\omega_2 = \sqrt{\frac{3k}{m}} = \sqrt{3}\omega$. So the two calculations agree.

2.3 Large number of identical coupled oscillators

Let us now consider a slightly different problem, which will eventually help us in analysing the vibrations of a stretched string. We deviate a bit from the previous case in that we consider chain of identical oscillators numbered $0, 1, \ldots (N+1)$, each having mass m and connected to the left and right neighbours by identical springs of stiffness k, and later consider the limit $N \to \infty$ as an approximation to a string under tension. We use the variable y to indicate the displacement, rather than x, keeping in mind that we will need x to show the position of a particle constituting the string. The end-point masses are taken to be fixed (stationary). The equation of motion for the p^{th} oscillator is given by

$$m\ddot{y}_p = k(y_{p-1} - y_p) - k(y_p - y_{p+1})$$

We attempt to solve this set of equations by assuming a harmonic solution in time:

$$y_p = A_p \cos(\omega t)$$

where we have taken ω to represent a normal mode frequency. We expect to find as many solutions (values) of ω as the number of oscillators. We set $k/m = \omega_0^2$ as usual. Substituting the trial solution into the equation of motion we find the following relationship between the coefficients A_p:

$$\left[-\omega^2 + 2\omega_0^2\right] A_p - \omega_0^2 \left[A_{p+1} + A_{p-1}\right] = 0$$

The equation can be rewritten as

$$\frac{A_{p-1} + A_{p+1}}{A_p} = \frac{-\omega^2 + 2\omega_0^2}{\omega_0^2},$$

which is a recursive relation for A_p. Once again, we assume a harmonic trial solution, but now for the the amplitude:

$$A_p = C \sin(p\theta)$$

in which C is a constant.

Substituting the trial solution in the recursive relation for A_p, and noting that $\sin\alpha + \sin\beta = 2\sin[(\alpha+\beta)/2]\,\cos[(\beta-\alpha)/2]$ we get

$$\frac{A_{p-1} + A_{p+1}}{A_p} = 2\cos\theta = \frac{-\omega^2 + 2\omega_0^2}{\omega_0^2}$$

Two boundary conditions are to be satisfied, corresponding to the fixed masses at the end of the chain of oscillators: $A_0 = 0, A_{N+1} = 0$. The first condition is satisfied by the trial solution for $p = 0$ automatically, while the second condition enforces

$$C \sin[(N+1)\theta] = 0$$

This implies

$$\theta = \frac{n\pi}{N+1}, \qquad (n = 0, 1, 2, \ldots \infty).$$

Thus, the permitted amplitudes are

$$A_p = C \sin\left(p\,\frac{n\pi}{N+1}\right)$$

Substituting this form of A_p in the recursion relation we get

$$2\cos\left(\frac{n\pi}{N+1}\right) = \frac{-\omega^2 + 2\omega_0^2}{\omega_0^2}$$

Using the trigonometric relationship $\cos\alpha = 1 - 2\sin^2\alpha/2$, we get

$$\omega^2 = 4\,\omega_0^2 \sin^2\left[\frac{n\pi}{2(N+1)}\right],$$

each value of n representing a different mode Thus the normal modes of this system of coupled oscillators have frequencies

$$\omega_n = 2\,\omega_0 \sin\left[\frac{n\pi}{2(N+1)}\right].$$

How many distinct values can ω_n take? The sin function is periodic, spans its range in the domain $[0, \pi/2]$, and only positive values of ω_n are meaningful, so, there will be N allowed values.

Exercise (level A): Find out, for what values of n are the amplitudes of the corresponding mode zero.

Exercise (level B): Show that if we let n take values beyond $N+1$, the resulting values of ω get repeated following the pattern $\omega_{N+1-p} = \omega_{N+1+p}$.

The displacement of the p^{th} oscillator in the n^{th} normal mode is

$$y_{p,n} = A_{pn}\cos(\omega_n t)$$

which we will write once in the long form:

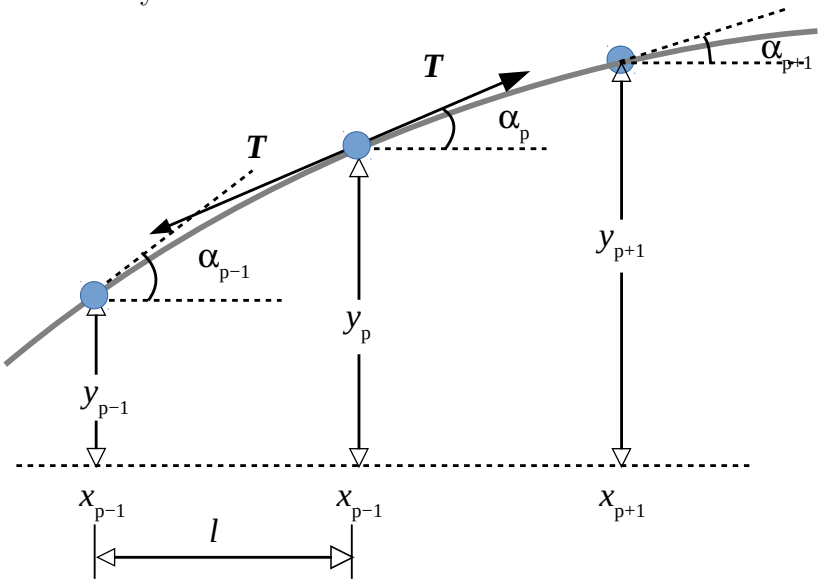
$$y_{p,n} = C_n \sin\left(\frac{pn\pi}{N+1}\right) \cdot \cos\left[2\,\omega_0 \sin\left(\frac{n\pi}{2(N+1)}\right) t\right]$$

in which C_n is an arbitrary constant, representing the amplitude of the n^{th} mode. This number is essentially determined by the initial impulse that sets the system into oscillation and it need not be the same for all modes. Note, that we have assumed at $t = 0$ all particles are at rest. If the initial conditions are different from this, the new conditions can be easily accommodated by allowing a *phase* in the time part of the solutions:

$$y_{p,n} = A_{pn}\cos(\omega_n t + \delta_n)$$

2.4 Transverse Displacements

The motion of the oscillators we have considered so far is longitudinal. What happens if the oscillations are transverse? Consider a string of identical masses m connected by elastic links of length l. Let T be the tension offered by the links.



Let these masses be in transverse oscillations (along y), with the end masses fixed as usual. The force on the p^{th} oscillator can be written component-wise as

$$\begin{aligned} F_x &= T\cos(\alpha_p) - T\cos(\alpha_{p-1}) \\ F_y &= T\sin(\alpha_p) - T\sin(\alpha_{p-1}) \end{aligned}$$

Since the string is displaced only slightly from its mean position, we can take α_p to be small for p. Hence we only need to retain terms of order α in simplifying the force equation for small oscillations. In this limit $\cos\alpha \approx 1$ and $\sin\alpha \approx \alpha$. Hence we have

$$\begin{aligned} F_x &= 0 \\ F_y &= T(\alpha_p) - \alpha_{p-1}) \end{aligned}$$

From the accompanying diagram we can see that $\alpha_p = (y_{p+1} - y_p)/l$ and $\alpha_{p-1} = (y_p - y_{p-1})/l$. The force on the p^{th} oscillator is mass times its acceleration, so the equation of motion for the y component becomes

$$m\frac{d^2y_p}{dt^2} = \frac{T}{l}(y_{p+1} - y_p)$$

If we set $\omega_0 = (T/ml)^2$, we find that the equation of motion becomes

$$\frac{d^2y}{dt^2} + 2\omega_0^2 y_p - \omega_0^2 y_{p-1} - \omega_0^2 y_{p+1} = 0$$

which is indeed identical to the equation we obtained for coupled longitudinal oscillations. So this model can be extended to the transverse oscillations of a string under tension. What we need to do is consider the limit when $N \to \infty$.

If l is the equilibrium separation between the oscillators, then the length of the equivalent string is $L = l(N+1)$ and the tension in the string is $T = kl$. In the limiting case, we will have have $m, l \to 0$, while the mass per unit length, $\mu = m/l$ remains constant. Thus

$$\omega_0^2 = \frac{k}{m} = \frac{T/l}{m} = \frac{T}{\mu l^2}$$

From the equation for ω_n we see that the highest mode, $n = (N+1)$ has a frequency $2T/\mu l^2$, while for the low modes ($n \ll N$), we get approximately

$$\begin{aligned} \omega_n &= 2\left(\frac{T}{\mu l^2}\right)^{1/2}\left(\frac{n\pi}{2(N+1)}\right) \\ &= \left(\frac{T}{\mu}\right)^{1/2}\left(\frac{n\pi}{L}\right) \end{aligned}$$

We have thus found the frequencies of the lowest normal modes. In these modes, the particle displacements are given by A_{pn}

$$A_{pn} = C\sin\left(p\,\frac{n\pi}{N+1}\right)$$

In the continuum limit $N \to \infty$, we can locate the p^{th} particle by its position coordinate $x_p = pl = pL/(N+1)$. Hence

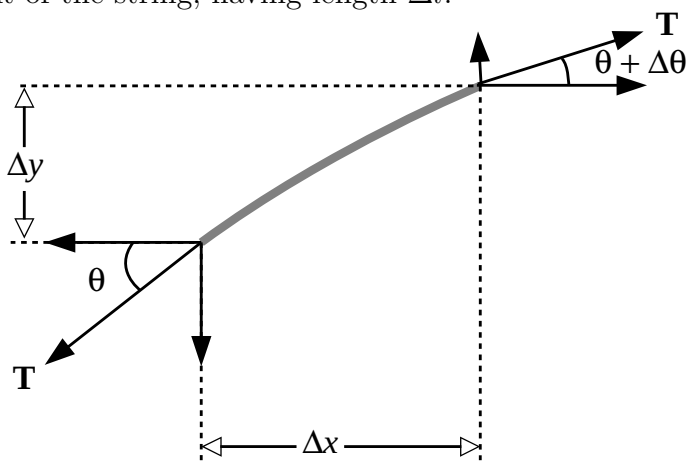
$$A_p = C \sin\left(\frac{n\pi x_p}{L}\right)$$

and the displacement of an arbitrary particle on the the string in the the low modes is given by

$$y_n = C \sin\left(\frac{n\pi x}{L}\right) \cos\left[\left(\frac{T}{\mu}\right)^{1/2} \left(\frac{n\pi}{L}\right) t\right]$$

3 A stretched vibrating string

Let us now analyse the vibrations of string by a different method that does not invoke the idea of an array of coupled oscillators. Let the string be under tension T and have mass per unit length μ. Consider a small segment of the string, having length Δl.



Referring to the diagram the forces on the string along the y and x directions are

$$\begin{aligned} F_x &= T\cos\theta - T\cos(\theta + \Delta\theta) \\ F_y &= T\sin\theta - T\sin(\theta + \Delta\theta) \end{aligned}$$

Assuming that the distortion of the string from a straight segment shape is minor, we expand the trigonometric functions and ignore terms of order $(\Delta\theta)^2$. In that approximation

$$F_x = 0, \qquad F_y = T\Delta\theta,$$

and $\Delta l \approx \Delta x$. Hence the equation for the transverse displacement of the string is

$$F_y = (\mu\Delta x)\frac{d^2y}{dt^2} = T\Delta\theta$$

We can write $\Delta\theta$ in terms of x, y by recognising that

$$\tan\theta = \frac{dy}{dx}; \qquad \sec^2\theta = \frac{d^2y}{dx^2}\frac{dx}{d\theta}$$

Once again, ignoring terms of order $(\Delta\theta)^2$, we have $\sec\theta \approx 1$ and we can write

$$\Delta\theta = \frac{d^2y}{dx^2}\Delta x$$

The equation of motion thus becomes

$$(\mu\Delta x)\frac{d^2y}{dt^2} = T\frac{d^2y}{dx^2}\Delta x$$

which we rearrange and write as

$$\frac{d^2y}{dx^2} = \frac{\mu}{T}\frac{d^2y}{dt^2}$$

Comment: The quantity $(\mu/T)^{-1/2}$ has dimensions of speed. We will see later that this is indeed the speed u at which the disturbance propagates along the string. As usual, we search for harmonic solutions to this equation

$$y(x,t) = f(x)\cos(\omega t)$$

so that

$$\frac{d^2y}{dx^2} = \frac{d^2f}{dx^2}\cos(\omega t); \qquad \frac{d^2y}{dt^2} = -\omega^2 f(x)\cos(\omega t)$$

Hence

$$\frac{d^2f}{dx^2}\cos(\omega t) = \frac{\mu}{T}\left[-\omega^2 f(x)\cos(\omega t)\right]$$

From the above equation it is clearly seen that $f(x)$ must be a harmonic function with a solution of the form

$$f(x) = A\sin(\omega x/u); \qquad u = (\mu/T)^{-1/2}$$

which must meet the boundary conditions, $f(0) = f(L) = 0$. These conditions give us the normal mode frequencies

$$\omega_n = \frac{n\pi u}{L}$$

and the complete solution for the displacement of an arbitrary particle:

$$y_n(x,t) = A_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{n\pi u t}{L}\right)$$

If we compare this solution with the solution obtained from the analysis of an infinite string of coupled oscillators, we find that the solutions are essentially the same.

3.1 Generalised displacement of a string and harmonic analysis

Consider the most general form of the displacement of a point on a stretched string, as a function of time, which is a linear combination of all harmonics:

$$y(x,t) = \sum_0^\infty C_n y_n(x,t) = \sum_0^\infty A_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{n\pi u t}{L}\right)$$

Let us now consider the displacement of a point located at x, *at a particular instant of time, t*. Then the displacement at that instant will be given by the expression

$$y(x,t) = \sum_0^\infty B_n \sin\left(\frac{n\pi x}{L}\right)$$

where B_n are new constants. Depending on which time we have chosen to view the string at, these constants will in general, be different. But the main point is that now the displacement of a point at an instant is a sum of several harmonic displacements, each of which has the form $\sin(n\pi x/L)$. What is not known are the values of B_n. But the values of B_n can be found by exploiting some properties of the sine and cosine functions. Let us consider the following equality based on the previous equation.

$$\int_0^L y(x,t)\sin\left(\frac{m\pi x}{L}\right)dx = \int_0^L \sum_0^\infty B_n \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx; \quad m \geq 1$$

Now the RHS integral, I_R, can be written as a sum of cosine functions by using the identity

$$\sin A \sin B = [\cos(A-B) - \cos(A+B)]/2$$

So

$$I_R = \int_0^L \sum_0^\infty B_n \left[\cos\left(\frac{(m-n)\pi x}{L}\right) - \cos\left(\frac{(m+n)\pi x}{L}\right)\right] dx$$
$$= \sum_0^\infty B_n \left[\frac{\sin\left(\frac{(m-n)\pi x}{L}\right)}{(m-n)\pi} - \frac{\sin\left(\frac{(m+n)\pi x}{L}\right)}{(m+n)\pi}\right]$$

The second term of I_R is zero always, but the first term has a $0/0$ form when $m = n$. So this is the only term of the series that is non-zero. In other words, of all terms contributing to I_R, we need consider only the m^{th} term. Thus,

$$I_R = \int_0^L B_m \sin^2\left(\frac{m\pi x}{L}\right) dx$$

which can be easily integrated as

$$I_R = \frac{1}{2}\int_0^L B_m \left[1 - \cos\left(\frac{2m\pi x}{L}\right)\right] dx$$
$$= \frac{1}{2} B_m \left[L - 0\right]$$
$$= B_m L/2$$

We thus obtain the value of B_m

$$B_m = \frac{2}{L}\int_0^L y(x,t) \sin\left(\frac{m\pi x}{L}\right) dx; \quad m \geq 1$$

Once the values of B_m are determined by this prescription, we have completed our task of breaking down the general displacement of a vibrating string (at a particular instant of time) into harmonics. Note, that these are harmonics in real space, not harmonics in time.

3.2 Fourier Decomposition

The analysis pertaining to the string obeys the boundary condition that the end-points have zero displacement (i.e. the end points are nodes) . This enforces the use of sine functions in the sum. If the boundary conditions were different, say for example, that the end points are anti-nodes, then the solutions would be in terms of cosine functions, and the harmonic analysis would involve cosine functions alone. So in the most general case we will need to allow for both types of harmonics. In such a general case we will have

$$y(x) = \frac{a_0}{2} + \sum_{n=1}^\infty \left[a_n \cos\left(\frac{2n\pi x}{L}\right) + b_n \sin\left(\frac{2n\pi x}{L}\right)\right]$$

This decomposition of a periodic function into harmonics is called Fourier analysis and the series is called the Fourier series. Let us consider some simple periodic functions, such as a sawtooth or a rectangular function.[2] A sawtooth wave is one in which the amplitude goes from zero to some finite non-zero value in some time T and then drops rapidly to zero.

$$y = y_0(x/L) \qquad\qquad 0 < x \leq L$$

[2]Typically such functions are useful in electronics and acoustics, where these can be realised as waveforms that are periodic in *time*, not in space, which is what have been discussing. Fourier analysis works out exactly the same way in time as it did in real space. That analysis is called time domain fourier analysis. It is important to remember this distinction.

A rectangular wave is one which has a constant (non-zero) amplitude over some distance l_1, followed by a certain distance l_2 during which the amplitude is zero. The cycle then repeats. If $l_1 = l_2$ ($\equiv L/2$), and the amplitude swings from $-A$ to $+A$ then it is called a square wave.

For ease of representation, we can set $A = 1$ and scale the x coordinate so that the period of the function is 2π with a range $[-\pi, \pi]$.

$$\begin{aligned} y &= -1 & -\pi < x < 0 \\ &= +1 & 0 < x < \pi \end{aligned}$$

Exercise (level B): Show that for the "standard" square wave shown above, the Fourier series is

$$y(x) = \sum_{m=1}^{\infty} b_m \sin(mx); \quad b_m = 4/m\pi \quad \text{for even } m; \quad b_m = 0 \quad \text{for odd } m$$

and all cosine terms are zero.

Exercise (level B): Show that a sawtooth wave is represented by the Fourier series

$$y(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{2}{n\pi} \sin\left(\frac{n\pi x}{L}\right)$$

---

4 Travelling Waves

Consider a finite string, open at one end and rigidly held at the other end. If it is given an impulse at the open end, a ripple travels towards the fixed end. The ripple (or a train of ripples) travels back after reflection, and if conditions are right a standing wave is formed. The shape of the string, or this standing wave, is very much like the normal mode oscillations of a fixed string that we have seen before. So, the travelling wave should be described by the normal mode solutions. Let us see how the travelling wave case is contained in the normal mode solutions.

The normal mode solution is

$$y_n(x,t) = A_n \sin\left(\frac{n\pi x}{L}\right) \cos(\omega_n t)$$

where $\omega_n = n\pi v/L$ is the n^{th} mode frequency and v is the wave speed. By using the trigonometric identities for addition of sines we can re-write the above equation as

$$y_n(x,t) = \frac{1}{2} A_n \left[\sin\left(\frac{n\pi x}{L} + (\omega_n t)\right) + \sin\left(\frac{n\pi x}{L} - (\omega_n t)\right)\right]$$

The first term represents a wave travelling in the $-x$ direction, while the second term represents the one travelling to the $+x$ direction. This can be seen as follows.

Take a specific displacement y_0^+ given by the second term corresponding to a certain combination of x, t. Since y_0^+ is a harmonic function, the same value of y_0^+ will reappear at some other value of its arguments, $x + \Delta x, t + \Delta t$. Hence we must have

$$\sin\left(\frac{n\pi x}{L} - \omega_n t\right) = \sin\left(\frac{n\pi(x + \Delta x)}{L} - \omega_n(t + \Delta t)\right)$$

This implies

$$n\pi \Delta x/L = \omega_n \Delta t$$

or, since $\omega_n = n\pi v/L$,

$$\Delta x/\Delta t = v$$

which tells us that the wave is propagating in the $+x$ direction with speed v. Similarly the other term in the standing wave solution represents a wave travelling in the $-x$ direction.

There is one subtlety which we have overlooked. The normal mode solutions which we used above, were obtained in the first place with the constraints that the end points of the string are fixed. But the travelling wave is obtained for one open end and a fixed other end. How are we justified in extending the first solution to the second case? The justification is based on two conditions: first, that setting up the travelling wave is a transient phenomenon (the impulse acts for a short time, compared to the period of the wave) and eventually (after the wave has travelled to the fixed end and back), a steady state is attained. Even if the wave train is finite (and much shorter than the length of the string, it is a periodic phenomenon, although the value of y is zero for most of the string coordinates. And we have seen that *any* periodic function can always be written as a (infinite) sum of sines and cosines (Fourier decomposition). The normal modes are just the components whose sum would represent any arbitrary pulse or wave train.

In short, the normal mode solutions cover the case of travelling waves, not only for waves that span entire strings, but also for short wave trains.

4.1 Speed of a wave

We have analysed the vibrations of a string in two ways. First, as a system of N coupled oscillators in the limit $N \to \infty$ as then a continuous medium which exhibits transverse oscillations subject to the constraint that its end-point are fixed. For the ideal string, having length L, mass per unit length μ and under tension T, the relationship between the wavelength, frequency of oscillation and the velocity of the wave was obtained as

$$v = (T/\mu)^{1/2}, \qquad \lambda_n = 2L/n, \qquad \omega_n = \pi n v/L$$

with no upper limit on n, at least in principle.

However, in the limiting case of N coupled oscillators, the normal mode frequencies were given by

$$\omega_n = 2\,\omega_0 \sin\left[\frac{n\pi}{2(N+1)}\right].$$

These are N different frequencies, the successive normal modes are *not* integer multiples of a basic frequency $\omega_0 = (T/\mu)^{1/2}/L$. Instead, the values are closer together and the highest possible value is twice the basic frequency. However, when n/N is small, i.e. for the low modes the two solutions agree. Experience tells us that the second analysis is closer to reality than the first one. One might argue that any finite length of string contains an enormous number of atoms, so the string is always in the limit $(n/N) \to 0$, even for relatively large values of n, so there is really no difference between the two cases. This argument is of limited validity, because every real string has a finite thickness, and for any non-zero thickness of the string there will always be transverse stresses or shearing forces, which will invalidate our assumption of the string being composed of perfect one-dimensional simple harmonic oscillators.

An important consequence of the unequal spacing of the normal mode frequencies implies that the product $\omega_n\lambda$

$$\omega_n\lambda = \left(\frac{(T/\mu)^{1/2}}{L}\right)\sin\left[\frac{n\pi}{2(N+1)}\right]\cdot\frac{2L}{n},$$

is not constant. This implies that the velocity of the wave is mode-dependent; higher modes have a lower velocity.

4.2 Dispersion of waves

This observation that the velocity is dependent on the mode has an interesting and important consequence. If a bunch of waves with slightly different characteristics are travelling in a medium, the waves will separate out as time progresses. This phenomenon is called dispersion.

Consider two waves,

$$y_{1,2} = A\sin(k_{1,2} - \omega_{1,2}t)$$

in which the difference between the two k values and the ω values is small. If these waves are superposed, i.e. they travel through along the same region of a medium, then the displacement due to the two together will take the form

$$\begin{aligned} y &= y_1 + y_2 \\ &= 2A\left[\sin\left(\frac{k_1+k_2}{2}x - \frac{\omega_1+\omega_2}{2}t\right) + \cos\left(\frac{k_1-k_2}{2}x - \frac{\omega_1-\omega_2}{2}t\right)\right] \\ &= 2A\left[\sin\left(kx-\omega t\right) + \cos\left(\frac{\Delta k}{2}x - \frac{\Delta\omega}{2}t\right)\right] \end{aligned}$$

in which we have defined k to be the average of k_1, k_2 and Δk to be the difference $k_2 - k_1$, and likewise for ω and $\Delta\omega$. This expression is a product of two travelling waves solutions, both travelling in the $+x$ directions. The first travelling wave is the one represented by the sine function, with wave speed ω/k. The other wave is the one represented by the cosine function, with wave speed $\Delta\omega/\Delta k$. What are these waves? A comparison with the phenomenon of beats will help. Beats occur when oscillations with two slightly differing frequency are superposed. That superposition has a very similar functional form to the above equation. There the cosine function was an envelope to the time-varying oscillation pattern, while the sine function was the oscillation itself. Likewise, here, the cosine function is an envelope to the crests of the travelling wave. It envelopes a group of waves. The envelope moves with a velocity equal to $\Delta\omega/\Delta k$, and this is called the group velocity. It is the speed corresponding to the average wave vector, and also the speed at which energy is transmitted. The speed corresponding to the sine function is the usual wave velocity, also called the phase velocity, since it (also) the speed at which the phase of any point on the wave changes.

Dispersion is seen in all kinds of waves, a well-known example is the dispersion of white light into rays of different colours by a prism. Red wavelengths have a lower speed than violet wavelengths in glass which causes dispersion – separation according to wavelength – when they travel unequal distances. Dispersion does not occur in passage through air. A corollary is that the refractive index of glass is wavelength dependent.

4.3 Transmission across a boundary

A standing wave is bounded by fixed end points, while a travelling wave is not. What happens when the travelling wave meets a boundary? To answer this let us first generalise what we mean by a boundary. A boundary, in general, is a location where the wave meets a resistance or opposition to motion. If the resistance is infinite, it is the same having a fixed point, that is the point does not yield to the disturbance. (The end points of a finite stretched strings are boundaries of this kind.) From such a boundary the wave will get reflected with complete reversal of the displacement, because the rigid boundary will exert an equal and opposite force on the string and hence invert the displacement. In general, however, when the boundary is not perfectly rigid, there is partial transmission and partial reflection across the boundary. Sticking to the protoptypical string, let us construct a simple boundary problem, and understand what happens at the boundary.

A boundary can be created by merely considering the junction of two straight strings, a light one (say, on the left) and a heavy one (on the right). Let the junction correspond to $x = 0$ and the mass

per unit length of the two strings be be μ_L and μ_R. If a wave is incident from the left, in the $+x$ direction, we will have the incident and reflected wave in the left portion and a transmitted wave in the right portion. The displacements of the strings in the two portions will be

$$y_L = A\sin(k_L x - \omega t) + B\sin(-k_L x - \omega t)$$
$$y_R = C\sin(k_R x - \omega t)$$

Note, that k_R and k_L are, in general, different, since $\mu_L \neq \mu_R$, and the velocities of the waves in the two media are not the same (recall that $k = \omega/v$). The *frequencies*, however are the same, for if they were not, then it might happen that the point at the boundary will be forced to oscillate at different frequencies at the same time, which is an inconsistency.

The boundary conditions to be satisfied are

$$\text{At} \quad x = 0 \quad y_L = y_R \quad \text{and} \quad \frac{\partial y_L}{\partial x} = \frac{\partial y_R}{\partial x}$$

These boundary conditions merely imply continuity of the string, and the absence of shearing of the boundary. Furthermore, the continuity of the first derivative is essential for defining the second derivative, which characterises the wave.

Substituting the two expressions for displacements in these boundary conditions, we get (after setting $x = 0$)

$$A\sin(-\omega t) + B\sin(-\omega t) = C\sin(-\omega t)$$
$$k_L A\cos(-\omega t) - k_L B\cos(-\omega t) = k_R C\cos(-\omega t)$$

or

$$A + B = C$$
$$A - B = \frac{k_R}{k_L} C$$

From these we find

$$\frac{C}{A} = \frac{2}{1 + k_R/k_L}$$
$$\frac{B}{A} = \frac{1 - k_R/k_L}{1 + k_R/k_L}$$

The ratios of the amplitudes of the incident, reflected and transmitted waves are thus related to the wave vectors in the two media, or in turn, the velocity of the waves in the two media:

$$\frac{C}{A} = \frac{2}{1 + v_L/v_R}$$
$$\frac{B}{A} = \frac{1 - v_L/v_R}{1 + v_L/v_R}$$

4.4 Sound Waves

So far, we have been mostly discussing waves on a string. When a string is plucked we see a standing wave pattern, and we hear a sound. The sound is heard not merely because the string vibrates,

but also because these vibrations are transmitted through air to our eardrum. This suggests that the column of air between the source and the ear responds to the vibrations in a harmonic manner, thereby agitating the eardrum. Let us see how the air column (or any gas, for that matter) responds to vibrations.

A gas is characterised – leaving aside for the time being the chemical structure of its constituents – by density and pressure. The pressure is a function of the density: $P = f(\rho)$. A tiny impulse, such as that created by a vibrating string or a vibrating membrane, will create a local pressure disturbance. Small changes in pressure from the quiet condition can be written as

$$P = P_0 + \frac{\partial P}{\partial \rho}\Delta\rho$$

where $\Delta\rho = \rho - \rho_0$ is the change in the density due to the impulse. Note, that the pressure and density changes are localised, because the disturbance is an impulse (sudden and short-lived application of a force). The effect of the impulse propagates with a finite speed through the medium as we will shortly see. This is in contrast to the case when we gradually and continuously press a piston in a cylinder–piston arrangement which is commonly referred to in the study of thermodynamics of gases.

Since the disturbance is an impulse, it is useful to investigate what happens to a parcel of the gas, whose thickness along the direction of the impulse is negligible. We need not concern ourselves with what happens to each molecule (which has a random motion, and no single molecule travels along the applied impulse), as it is sufficient to investigate what happens to them on the average. However, we must make allowance for the fact that a gas is easily compressed and is not a rigid body. So, if we have of two parcels of the gas with some separation between them, there is no guarantee that under the application of a force that displaces these parcels, the separation between them will remain unchanged. In general if x_1 and x_2 are the coordinates of the parcels, then their displacement under an external force will lead to new coordinates x_1' and x_2' given by

$$x_1' = x_1 + u(x_1) \quad \text{and} \quad x_2' = x_2 + u(x_2)$$

where $u(x)$ is a function that specifies the displacement at various locations, and is not the same at all locations, in general. To be more precise, u is also a function of time, for if it were not, we would not be able to account for the fact that the density of a gas tends to become homogeneous with time.

Let a certain parcel of the gas be specified by two boundaries, perpendicular to the direction of the impulse, located at the coordinates x and $x + \Delta x$. Thus, in the quiet condition, the parcel has thickness Δx along the direction of the impulse and a cross section area A and density ρ_0. As a result of the impulse the pressure and the density within this parcel changes, as does the thickness of the parcel, but the mass of the parcel remains unchanged. This gives us the relationship

$$\rho_0(A\,\Delta x) = \rho'(A\,\Delta x')$$

As a result of the impulse the boundary at x is displaced to $x + u(x)$ while the boundary at $x + \Delta x$ is displaced to $(x + \Delta x) + u(x + \Delta x)$. The new thickness of the parcel is given by

$$\begin{aligned}\Delta x' &= [(x + \Delta x) + u(x + \Delta x)] - [x + u(x)] \\ &= \Delta x + u(x + \Delta x) - u(x) \\ &= \Delta x + \frac{\partial u}{\partial x}\Delta x\end{aligned}$$

Using this in the mass conservation equation then gives us

$$\rho_0 = \rho' \left[1 + \frac{\partial u}{\partial x}\right]$$

which can be re-cast as

$$\rho - \rho_0 = -\rho_0 \frac{\partial u}{\partial x}$$

Now let us recall that the displacement function $u(x)$ is function of time too, which means that the displaced parcel is accelerating. The acceleration is related to the net force on the parcel, or in other words to the difference in the pressure on the two boundaries of the parcel. So

$$F_x - F_{x+\Delta x} = (\rho_0 A \Delta x) \frac{\partial^2 u}{\partial t^2}$$

Since pressure is force per unit area, we divide both sides by A to get

$$\begin{aligned} P_x - P_{x+\Delta x} &= \rho_0 \Delta x \frac{\partial^2 u}{\partial t^2} \\ \therefore -\frac{\partial P}{\partial x} \Delta x &= \rho_0 \Delta x \frac{\partial^2 u}{\partial t^2} \\ \therefore -\frac{\partial P}{\partial x} &= \rho_0 \frac{\partial^2 u}{\partial t^2} \end{aligned}$$

Using $P = P_0 + \frac{\partial P}{\partial \rho}(\rho - \rho_0)$, we have

$$-\frac{\partial}{\partial x}\left[P_0 + (\rho - \rho_0)\frac{\partial P}{\partial \rho}\right] = \rho_0 \frac{\partial^2 u}{\partial t^2}$$

butP_0 is a constant and $\rho - \rho_0 = -\rho_0 \frac{\partial u}{\partial x}$, and $(\partial P/\partial \rho) = K/\rho$, where K is the adiabatic bulk modulus of the gas. Hence,

$$\begin{aligned} -\frac{K}{\rho_0}\frac{\partial}{\partial x}\left[-\rho_0 \frac{\partial u}{\partial x}\right] &= \rho_0 \frac{\partial^2 u}{\partial t^2} \\ \frac{\partial^2 u}{\partial x^2} &= \frac{\rho_0}{K}\frac{\partial^2 u}{\partial t^2} \end{aligned}$$

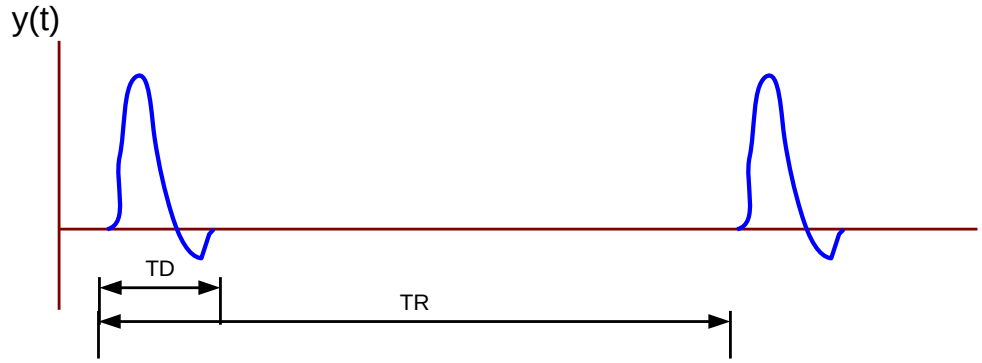
This is exactly the wave equation we have been dealing with. Here it represents a wave comprised of the longitudinal displacements of a slice of the gas when subject to an impulse. In other words, this is the equation for a sound wave. The speed of sound in a gas is thus $\sqrt{K/\rho}$. The compressibility of the gas permits the setting up of a longitudinal sound wave in it, and this is how the effect of vibrations of a string reaches our eardrum.

5 Wave Pulses

When we shake or twirl a rope, we create a pulse that travels along the rope. Typically this pulse will propagate at a speed equal to $\sqrt{T/\mu}$. This is an elementary form of sending a *signal*; another example of an elementary signal is when you clap to draw someone's attention. The wave pattern for neither of these signals are anything like the harmonic solutions for waves or oscillations that we have been analysing so far. The main differentiating feature is that these signals persist only for a short duration T_D, whereas the harmonic solutions apply at times, and also over the full range of the spatial coordinates (since the form sin or cos $(\omega t - kx)$ has finite values for all t and x). So does it mean that the harmonic solutions developed so far will be useless to describe these displacements (of the rope, or the air parcel) that last for only a finite time?

5.1 A pulse in time

If we concentrate at any point on the rope (i.e. we keep x fixed) and observe how this point moves in time we might get a displacement versus time graph as shown in the figure below, shown as repeating after some time.



Without bothering about the details of the shape of the pulse, we can say that pulse is a disturbance, that results in a displacement that has a general form

$$\begin{aligned} [x \text{ fixed}] : y(t) &\neq 0 & 0 < t < T_D \\ &= 0 & T_D < t < T_R \end{aligned}$$

where T_R is a arbitrary time, usually much larger that T_D, and can in principle be ∞. T_R would usually be the time after which the signal is repeated. Repeated signals are used, for example, in real life communications including speech, the latter being far more complicated that communication by electronic signals.

As we have just seen, a pure sinusoid exists for infinite time, so the $y(t)$ above cannot be such function. Some common signal shapes might be a single sawtooth, a single rectangular pulse, or just any lump. We know from our earlier discussion of Fourier analysis, that each such disturbance in time can be written as an infinite sum of harmonics of a base frequency ω_0

$$y(t) = \sum_{n=1}^{\infty} C_n \cos(\omega_n t + \delta_n); \qquad \omega_n = n\omega_0$$

Since such decomposition can always be done, and each of the components is a harmonic function, we can apply all our analysis of travelling waves and oscillations, to understand the propagation of any arbitrary pulses.

Let us consider now a special (if rather artificial) case, of a signal pulse which is a train of m cycles of the p^{th} harmonic of the base frequency and that the signal is repeated after a time that is equal to the period corresponding to the base frequency.

Such a signal would be described by

$$\begin{aligned} y(t) &= A\sin(p\omega_0 t) & 0 < t < T_D \\ &= 0 & T_D < t < T_R \end{aligned}$$

where $T_D = 2m\pi/p\omega_0$ is the duration of m cycles of the p^{th} harmonic and $T_R = 2\pi/\omega_0$. This function is neither symmetric nor antisymmetric as it stands, but we can make it appear antisymmetric (since it is a sin function) by merely shifting the time reference.

$$\begin{aligned} y(t) &= 0 & -\pi/\omega_0 < t < -m\pi/p\omega_0 \\ &= A\sin(p\omega_0 t) & -m\pi/p\omega_0 < t < +m\pi/p\omega_0 \\ &= 0 & +m\pi/p\omega_0 < t < +\pi/\omega_0 \end{aligned}$$

We find the C_n as follows

$$C_n = \frac{2}{T_R}\int_{-T_R/2}^{+T_R/2} dt\, y(t)\sin(n\omega_0 t)$$

We have to change the limits of integration now to account for the fact that the signal is non-zero only for a small duration, $(-m\pi/p\omega_0, m\pi/p\omega_0)$, which is less than the repetition period. So

$$\begin{aligned}
C_n &= \frac{\omega_0 A_0}{\pi}\int_{-m\pi/p\omega_0}^{+m\pi/p\omega_0} dt\ \sin(p\omega t)\sin(n\omega t) \\
&= \frac{\omega_0 A_0}{2\pi}\int_{-m\pi/p\omega_0}^{+m\pi/p\omega_0} dt\ [\cos(p-n)\omega_0 t - \cos(p+n)\omega_0 t] \\
&= \frac{\omega_0 A_0}{2\pi}\left[\frac{\sin(p-n)\omega_0 t}{(p-n)\omega_0} - \frac{\sin(p+n)\omega_0 t}{(p+n)\omega_0}\right]_{-m\pi/p\omega_0}^{+m\pi/p\omega_0} \\
&= \frac{\omega_0 A_0}{\pi}\left[\frac{\sin(p-n)\frac{m}{p}\pi}{(p-n)\omega_0} - \frac{\sin(p+n)\frac{m}{p}\pi}{(p+n)\omega_0}\right]
\end{aligned}$$

For $n \approx p$ the first term is very large compared to all other terms. In this approximation we have

$$\begin{aligned}
C_n &= \frac{A_0}{\pi}\left[\frac{\sin[(p-n)\frac{m}{p}\pi]}{(p-n)}\right] \\
C_n &= \frac{mA_0}{\pi}\left[\frac{\sin\theta_n}{\theta_n}\right]
\end{aligned}$$

where $\theta_n = \frac{m}{p}\pi(p-n)$. The values of C_n are large only when $\theta_n \approx 0$, i.e. $p \approx n$ and fall off rapidly as θ approaches π and then they oscillate. This means that the Fourier component corresponding the harmonic which the signal is made of is the dominant component, as expected. Now consider the case when $m = p$, i.e. the the number of cycles of the p^{th} harmonic that constitutes the signal is p itself. Then there is only one term in the expansion, C_p, since the signal is a complete wave train over the entire repetition period, not restricted to a few cycles.

If the signal pulse is only 1 cycle of the p^{th} harmonic, then $C_n = A_0/p$ for $n = p$ and $C_n = \frac{A_0}{\pi}[\frac{\sin\theta_n}{\theta_n}]$ for $n \neq p$. and the amplitudes for all $n \approx p$ are equal.

5.2 Group of frequencies

In the previous section we looked at a pulse in time, that propagates on a string or through a medium, and showed that it could be treated as a linear combination of multiple harmonics. The narrower the pulse is in time, the greater is the number of harmonics which contribute significantly to the linear combination describing the pulse.

Let us know do the reverse exercise. We will start with a group of frequencies which are associated with the oscillation of a point, and determine how the displacement appears as a function of time. Let this group of frequencies be made up of n equally-spaced values in the range ω_0 to $\omega_0 + \Delta\omega$, with each frequency having the same amplitude a. Let the spacing between adjacent frequencies be $\delta\omega$. The average frequency of this group, $\bar{\omega} = \omega_0 + \delta\omega(n-1)/2$. The resultant displacement of the particle (at a fixed x coordinate) at any time t will be given by

$$y(t) = \sum_0^{n-1} a\cos(\omega_0 + i\delta\omega t)$$

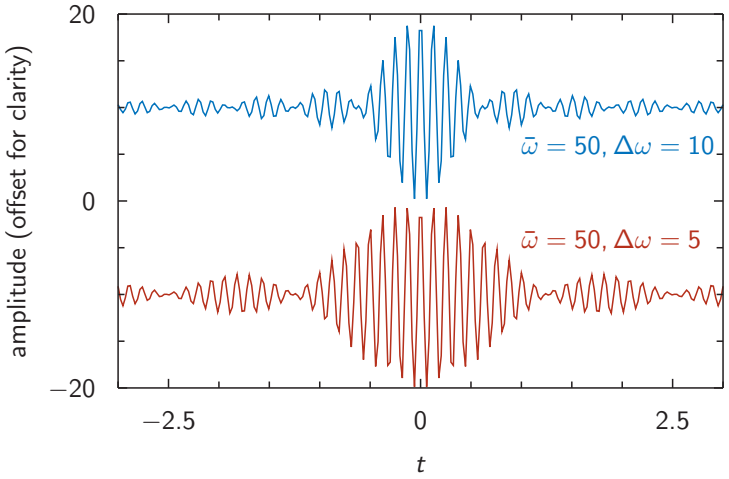
The sum can be effected by a geometric method, or by converting the cosines to imaginary exponentials (see The Physics of Vibrations and Waves, Chapter 1, H. J. Pain). We get

$$\begin{aligned} y(t) &= a\frac{\sin[n(\delta\omega)t/2]}{\sin[(\delta\omega)t/2]}\cos[(n-1)(\delta\omega t)/2] \\ &= a\frac{\sin[(\Delta\omega)t/2]}{\sin[(\Delta\omega)t/2n]}\cos[\bar{\omega}t] \end{aligned}$$

For t not much different from 0, and for a large number of frequencies with a narrow spread compared to their mean value, the denominator can be simply written as $(\Delta\omega)t/2n$. Then we have

$$y(t) = na\cos[\bar{\omega}t]\frac{\sin[(\Delta\omega)t/2]}{[(\Delta\omega)t/2]}$$

The cosine term varies rapidly compared to the sine term, since $\bar{\omega}$ is greater than $\Delta\omega$, and the graph of $y(t)$ appears as shown below.



The envelope function is the sine function. Thus we have an oscillation at frequency $\bar{\omega}$, whose amplitude is *modulated* by a $\sin((\Delta\omega)t/2)/((\Delta\omega)t/2)$ function. The amplitude is maximum $(= na)$ at $t = 0$, since all oscillations are initially in phase. As time progresses, they go out of phase and when $(\Delta\omega)t/2$ becomes π the amplitude of the envelope becomes zero. Thus the time Δt in which the amplitude of the envelope falls from its maximum value na to zero is $2\pi/(\Delta\omega)$. We also find an important relationship between $\Delta\omega$ and Δt

$$\Delta\omega\Delta t = 2\pi.$$

This is just a more concrete statement of what we had found in the previous section: the narrower the spread in time of a pulse (small Δt), the larger the number of frequencies (large $\Delta\omega$) needed to represent the pulse and vice-versa, keeping the product of their spreads fixed. (This statement is called the bandwidth condition, for reasons we will shortly see.)

We can look at the final expression for $y(t)$ obtained above as a rapidly varying harmonic function whose amplitude is modulated gradually in time. The function

$$y(t) = [a + b\cos(\omega' t)]\cos(\omega t)$$

is also a similar function overall, but with a different modulation. Suppose now, that this modulated ocsillation travels. That is,

$$y(x,t) = [a + b\cos(\omega' t)]\cos(\omega t - kx)$$

If $\omega' \ll \omega$, then we can think of this as a wave at frequency ω carrying a *signal* of a lower frequency, the signal being understood as the modulation of its amplitude. This form of sending a signal by

amplitude modulation is commonly adopted in radio transmission; the frequency ω' is the frequency of the speech or sound that is being transmitted (note that this frequency is not single valued, it is typically any value in the range of audible frequencies). If this travelling wave is interrupted at some position x_R, the oscillation at that point will have the form

$$y(x,t) = [a + b\cos(\omega' t)]\cos(\omega t - \delta)$$

or

$$y(t) = a\cos(\omega t - \delta) + \frac{b}{2}\cos[(\omega' + \omega)t - \delta] + \frac{b}{2}\cos[(\omega t + \omega')t + \delta]$$

The first term is simply oscillation at the carrier wave frequency, while the other terms are frequencies on either side of the main frequency, also side-bands. To retrieve the signal at the frequency ω', we need to decouple it from the carrier frequency, which an electronic implementation of the orthogonality property of cosines, applied at different frequencies in the side bands. Thus, when we tune in to a radio station, the receiver has to be sensitive to not just to the carrier frequency, but also to the side bands. This is a reason why radio senders operate at frequencies that are generously separated from each other.

5.3 Dispersion of waves

In the discussion about a carrier wave and amplitude modulation for transmitting a signal, we assumed that all frequencies would travel at the same speed v. However this need not always be the case. In an earlier section on dispersion we looked at two waves,

$$y_{1,2} = A\sin(k_{1,2} - \omega_{1,2}t)$$

in which the difference between the two k values and the ω values is small are superposed. As they travel along the same region of a medium, the displacement due to the two together is

$$y = 2A\left[\sin\left(\frac{k_1 + k_2}{2}x - \frac{\omega_1 + \omega_2}{2}t\right) + \cos\left(\frac{k_1 - k_2}{2}x - \frac{\omega_1 - \omega_2}{2}t\right)\right]$$

The cosine component represents an envelope of the more rapidly oscillating sine wave. The envelope over the group of two waves has a velocity v_g, given by

$$v_g = \frac{\Delta\omega}{\Delta k} = \frac{\omega_1 - \omega_2}{k_1 - k_2}.$$

If we divide the numerator and the denominator by $k_1 k_2$, and note, that $v_1 = \omega_1/k_1$ (and similarly for 2), we get

$$\frac{\Delta\omega}{\Delta k} = \frac{v_1/k_2 - v_2/k_1}{1/k_2 - 1/k_1}.$$

If v_1 and v_2 are the same, that is the medium is non-dispersing, then each of them will be equal to v_p, the rate of change of the phase of the oscillation, or the phase velocity. In that case we can write the following condition for a non-dispersive medium

$$v_g = \frac{\Delta\omega}{\Delta k} = v_p$$

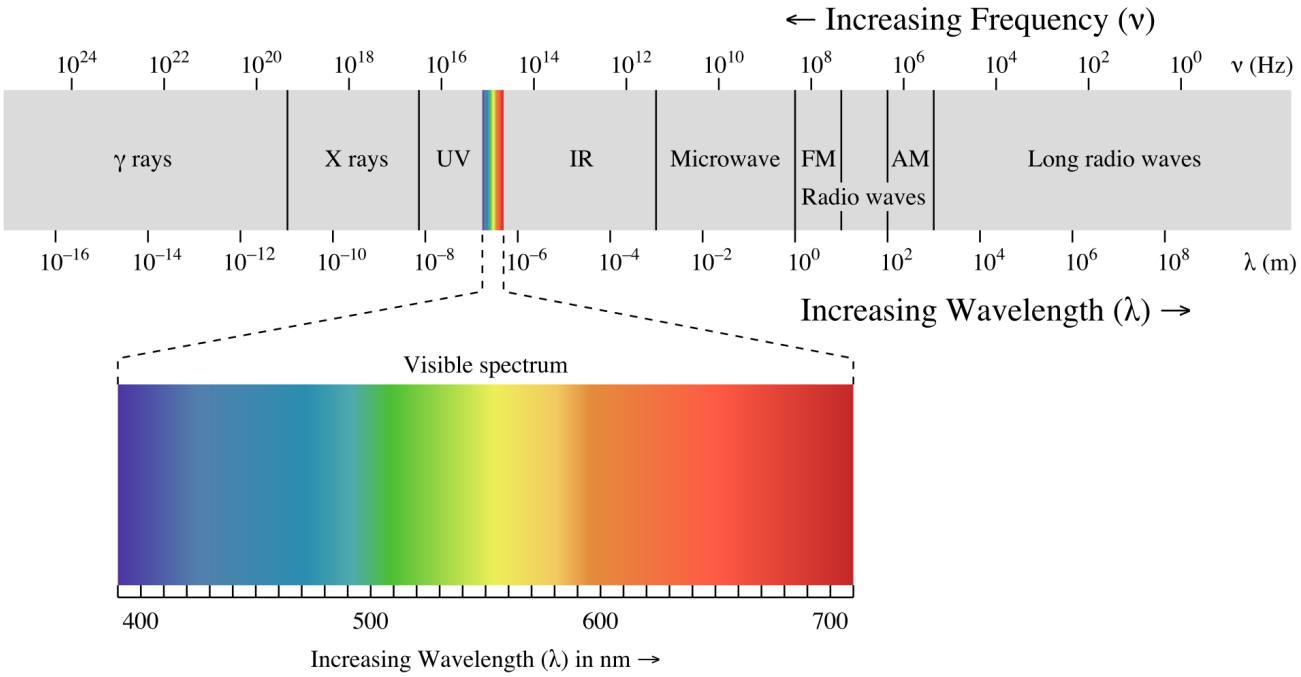
Radio wave broadcast, by time-dependent modulation of the amplitude, which results in the creation of side-bands, can be effective only when this condition is satisfied, otherwise the receiver will not receive all side-bands simultaneously. Dispersion is actually the norm rather than the exception. Light passing through any transparent solid exhibits dispersion, as seen most dramatically in a prism; seismic waves are dispersed in the earth's crust, and such dispersion becomes a tool for estimating the composition of the earth's interior and also for testing of materials for fractures and inhomogeneities and in medical diagnostics using ultrasonic waves.

6 Electromagnetic waves

In previous sections we have considered stationary and travelling waves on a stretched string, or collection of coupled oscillators. We also discussed sound waves in a medium. Now we will turn our attention to electromagnetic waves. They have some similarities with the above systems, but also some very important differences.

Electromagnetic waves are also called "electromagnetic radiation". The most familiar form of this radiation is light, which is usually studied under the heading of "optics" where we talk about focusing of rays, lenses, interference, diffraction etc. However, we now understand that visible light occupies a very small segment within the range of electromagnetic radiation. The rest of the range is occupied by radiation having names like infrared, ultraviolet, X-rays, gamma-rays, radio waves etc.

In terms of fundamental properties, all types of radiation are absolutely the same. They can all exhibit the basic properties of light, such as interference and diffraction. They can all be polarised or unpolarised. They only differ from each other in their frequency/wavelength. Of course in terms of interaction with materials and human beings, the various types of radiation behave extremely differently from each other. For example, unlike light some of them can cause damage, some of them can pass through solid matter, some can communicate radio signals effectively, some can transmit heat. Lenses may or may not bend all types of radiation. But ultimately all these different behaviours are only due to the difference in frequency!



Thus it makes sense to consider the basic properties of electromagnetic waves and this is what we will do here.

The first question is, what are the waves made up of? By analogy with a sound wave, we may imagine an electromagnetic wave is an oscillation in some medium. However this contradicts some simple experiments. First place a music player in a glass jar and evacuate the jar with a vacuum pump. You can no longer hear any music. However you can still see the music player! Apparently light has travelled in the vacuum. You may argue that the vacuum was not perfect enough. But consider that sunlight reaches us across space. So either it travels through a vacuum, or there is a medium that we do not know about, filling all of space, and light is an oscillation of that medium.

Such a medium was hypothesised under the name of "ether". If it exists, then motion through the ether should change the speed of light (as motion through air changes the speed of a sound wave).

However the very sophisticated Michelson-Morley experiment tested for such a change in the speed of light and found no such effect. This experiment led to the modern view that there is no ether, and electromagnetic waves indeed propagate in a vacuum.

Of course, light also propagates in materials. In this case it is found to propagate more slowly. But this does not mean that the light wave is an oscillation of that material. In fact, the light wave keeps striking atoms of the material and undergoing reflections, and this effectively slows down the wave. But microscopically (i.e. in between collisions), it is always moving at the same speed – the speed of light c, which is roughly 3×10^8 m/sec. Einstein understood that the speed of light is a fundamental constant of the universe and it plays an essential role in the special theory of relativity.

But we are no closer to understanding electromagnetic radiation. If it is not the vibration of a medium, what is it the vibration of? The answer lies in the concept of a *field*, a basic entity in modern physics. The electromagnetic field consists of the two vectors $\vec{E}(\vec{x}\,,t)$ and $\vec{B}(\vec{x}\,,t)$, known as the electric and magnetic field, which pervade all of space. These fields carry energy and exert force. For example a charged particle moving in an electromagnetic field experiences the famous Lorentz force:

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$$

The important point is that a field is a quantity defined at all points of space and time. It satisfies certain equations called "field equations". For the electromagnetic field, these are called "Maxwell's equations". These were abstracted from a number of experiments performed by Coulomb, Oersted, Biot, Savart, Ampère, and Faraday among others. We are not going to write down all these equations here, but let us mention that they are of the following general form. They involve first derivatives of the fields $\vec{E}, \vec{B}$ with respect to both space and time, and they also involve some given distribution of charges $\rho(\vec{x}\,,t)$ and currents $\vec{J}\,(\vec{x}\,,t)$. They take the form:

$$\Big[\text{Some linear function of } \tfrac{\partial \vec{E}}{\partial t}, \tfrac{\partial \vec{E}}{\partial x}, \tfrac{\partial \vec{B}}{\partial t}, \tfrac{\partial \vec{B}}{\partial x}\Big] = \Big[\text{Some linear function of } \rho, \vec{J}\Big]$$

Here we used $\frac{\partial \vec{E}}{\partial x}$ as shorthand for the three quantities $\frac{\partial \vec{E}}{\partial x}, \frac{\partial \vec{E}}{\partial y}, \frac{\partial \vec{E}}{\partial z}$. Once we specify the sources, the above equations will determine the field – of course subject to imposing suitable boundary conditions.

To be explicit, one of Maxwell's equations is:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0}$$

where ϵ_0 is a constant called the "permittivity of the vacuum". Since $\vec{\nabla} \cdot \vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}$, we see that it is of the general form given above. The other Maxwell equations are more technical and each of them embodies some property of electromagnetism.

The important thing is that *every possible electromagnetic field* has to be a solution of Maxwell equations. That includes the field of a point charge, of a magnet, of a conducting wire, of a solenoid, of a conductor of any shape etc, But most of these solutions are not of interest to us here. We will restrict ourselves to the *special class* of electromagnetic field configurations that correspond to electromagnetic radiation.

6.1 Plane electromagnetic waves

Let us consider electromagnetism in the absence of sources: $\rho = \vec{J} = 0$. In that case, the Maxwell equations take the source-free form:

$$\Big[\text{Some linear function of } \tfrac{\partial \vec{E}}{\partial t}, \tfrac{\partial \vec{E}}{\partial x}, \tfrac{\partial \vec{B}}{\partial t}, \tfrac{\partial \vec{B}}{\partial x}\Big] = 0$$

From these equations one can derive the following simple results:

$$\begin{aligned}\vec{\nabla}^2\vec{E} - \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = 0 \\ \vec{\nabla}^2\vec{B} - \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = 0\end{aligned} \tag{6.1}$$

This is the electromagnetic wave equation. Mathematically it is identical to equations that we have already seen. It is basically the same as Eq.(**??**) describing the propagation of a disturbance along a stretched string (that equation was written assuming the disturbance is only along a fixed direction, say x. We can do the same here if we restrict our attention to fields that vary only along x and are independent of y and z. In that case $\vec{\nabla}^2 \to \frac{\partial^2}{\partial x^2}$ and we get precisely the same equation).

Recall that in the equation for a vibrating string, the velocity v of the disturbance appeared in the same place as c in the electromagnetic equation. But there is an important difference in the two quantities. For the string, we have $v = \sqrt{\frac{T}{\mu}}$ which depends on physical properties of the string, namely its tension T and mass per unit length μ. However in the electromagnetic case the velocity of the wave is a fixed quantity c – a constant of nature independent of any material. We will take this electromagnetic wave equation in vacuum as our starting point.

One may wonder how there can be electromagnetic waves in the absence of sources! What provides the energy for these waves? The answer goes like this: we can imagine that some disturbance created the wave in some region of space and time. The wave will spread and we observe it far away, in a region of space that is completely empty. The above equation tells us what we can expect to see in this region far away from sources. If we want to understand what we would see near the source, that would require us to solve the wave equation with a source. That is a more difficult job and we avoid it for now.

One enormous convenience is that we do not need to learn new mathematics to solve the electromagnetic wave equation. The general solution of the electromagnetic wave equation is:

$$\begin{aligned}\vec{E}(\vec{x}\,,t) = \vec{E}_0 e^{i(\vec{k}\cdot\vec{x}-\omega t)} \\ \vec{B}(\vec{x}\,,t) = \vec{B}_0 e^{i(\vec{k}\cdot\vec{x}-\omega t)}\end{aligned} \tag{6.2}$$

where $\omega = c|\vec{k}\,|$ and $\vec{k}$ is called the wave vector. Of course, linear combinations of the above solutions are also allowed. As we have seen previously, the principle of superposition holds and we can superpose any of the above solutions to get a general one. Note that we have written the general solution as a complex one, but the physical electric and magnetic fields $\vec{E}, \vec{B}$ are supposed to be real. So we have to make sure our linear combination is real. One way is to write "Re( )" around every expression, which just means to take the real part. To avoid complicating our formulae we will not write this "Re" every time, but it should be assumed to be present.

Now let us consider waves propagating along the z-direction. In that case, $\vec{k} = (0, 0, k)$ where the number k can be either positive or negative, and $\omega = |k|c = \pm kc$. Thus we can have two terms:

$$u_k(z,t) = Ae^{i|k|(z-ct)} + Be^{-i|k|(z+ct)} \tag{6.3}$$

The first term corresponds to waves propagating to the right, while the second describes waves propagating to the left.

Given the basis above, we can take superpositions of the u_k and build up any function of $(z - ct)$ and any other function of $(z + ct)$ by Fourier analysis. Thus the most general electromagnetic wave in the z-direction will have components:

$$u(z,t) = f(z-ct) + g(z+ct) \tag{6.4}$$

where f, g are independent functions of one variable.

Now recall that the second-order wave equation was obtained by differentiating the first-order Maxwell equations. Therefore a solution of Maxwell equations will satisfy the wave equation, but a solution of the wave equation need not satisfy the Maxwell equations! So out of the general class of solutions of the above form where u stands for any of the six quantities $E_x, E_y, E_z, B_x, B_y, B_z$, we should not expect that all of them are genuine electromagnetic waves. So let us go back to the general form of the wave and write it as follows:

$$\begin{aligned} \vec{E} &= \vec{\epsilon}_E E_0\, e^{i(\vec{k}\cdot\vec{x}-\omega t)} \\ \vec{B} &= \vec{\epsilon}_B B_0\, e^{i(\vec{k}\cdot\vec{x}-\omega t)} \end{aligned} \tag{6.5}$$

So far we have done nothing, because the exponential is the general one and the coefficient has been parametrised as by a magnitude (E_0 and B_0) and unit vectors $\vec{\epsilon}$ that are called "polarisation vectors". Note that the polarisation vectors do not depend on $\vec{x}\,, t$ but they can depend on $\vec{k}\,$.

However, now we must impose Maxwell's equations. There are two relatively simple equations, which in free space are $\vec{\nabla}\cdot\vec{E}=0, \vec{\nabla}\cdot\vec{B}=0$. Imposing these on the above solutions, we find:

$$\vec{k}\cdot\vec{\epsilon}_E = 0 = \vec{k}\cdot\vec{\epsilon}_B \tag{6.6}$$

Since $\vec{k}$ is the direction of motion of the wave, this tells us that the polarisation of the wave (i.e. the direction of the vectors $\vec{E}, \vec{B}$) is *perpendicular* or *transverse* to the direction of motion. We see that electromagnetic waves are transverse waves.

The remaining Maxwell equations lead to the following additional condition:

$$\vec{B} = \frac{1}{c}\hat{k}\times\vec{E}$$

From this it follows that:

$$B_0 = \frac{1}{c}E_0, \quad \epsilon_B = \hat{k}\times\epsilon_E \tag{6.7}$$

where $\hat{k}$ is the unit vector $\frac{\vec{k}}{|\vec{k}|}$ along the direction of motion.

To summarise, the most general electromagnetic waves in free space are given by:

$$\begin{aligned} \vec{E} &= \mathrm{Re}\left(E_0\,\vec{\epsilon}_E e^{i(\vec{k}\cdot\vec{x}-\omega t)}\right) \\ \vec{B} &= \mathrm{Re}\left(\frac{E_0}{c}(\hat{k}\times\vec{\epsilon}_E)\, e^{i(\vec{k}\cdot\vec{x}-\omega t)}\right) \end{aligned} \tag{6.8}$$

and linear combinations. Notice that the three vectors $\vec{\epsilon}_E, \vec{\epsilon}_B, \hat{k}$ are all unit vectors and pairwise orthogonal. Thus they form an orthonormal set. One can make this more explicit by going into a special basis. Suppose, as before, we take the wave to be propagating along the z-direction. Then $\hat{k} = (0,0,1)$. Now the transversality condition says that the polarisation vector is orthogonal to this. By rotating our $x-y$ plane, we can align the electric polarisation along the x axis. Thus $\vec{\epsilon}_E = (1,0,0)$. Finally $\vec{\epsilon}_B = \hat{k}\times\vec{\epsilon}_E = (0,1,0)$ and we clearly see the orthonormal set.

Notice that $\vec{E}$ and $\vec{B}$ are in phase with each other and we saw their amplitudes are proportional. Thus an electromagnetic wave is an oscillating pair of crossed electric and magnetic fields having amplitudes in a fixed proportion, the same phase, and perpendicular directions. Note that when we draw a picture of an electromagnetic wave, the transverse displacement that we draw is not plotted in real space! The wave is not "oscillating transversely in real space". Rather, there is a transverse electric and magnetic field, and the picture shows the *magnitude* of this field.

We have derived these results starting from Maxwell's equations. Since those equations have a sound experimental basis, one can be confident that these results are correct. But what is the experimental evidence for electromagnetic radiation?And do we know that it is made up of crossed

electric and magnetic fields? Historically, visible light was of course known since pre-history. Infrared and ultraviolet radiations were isolated in the early 19th century using sunlight and a prism. They were shown to have somewhat distinctive properties: ultraviolet rays darkened photo film more rapidly (higher energy) while infrared rays heated up substances more rapidly (lower energy, so more absorbed). But these were chance discoveries that did not by themselves tell us what radiation was made up of. The big event came *after* Maxwell wrote his equations. Heinrich Hertz used Maxwell's equations (in the presence of sources) to design an emitter of radio waves and microwaves. He was able to detect these waves at a distance, and test their reflection and interference properties. He knew their frequency from the production mechanism, and their wavelength from interference patterns. Using both together he could calculate the speed:

$$v = \nu\lambda$$

and found that this speed equalled the speed of light. This was conclusive evidence that the electromagnetic waves predicted by Maxwell's equations really exist.

Exercise (level A): Show that:

$$\vec{\nabla}\,(e^{i\vec{k}\cdot\vec{x}}) = i\vec{k}\,e^{i\vec{k}\cdot\vec{x}}$$
$$\vec{\nabla}^{\,2}(e^{i\vec{k}\cdot\vec{x}}) = -\tfrac{\omega^2}{c^2}e^{i\vec{k}\cdot\vec{x}}$$

where $\omega = c|\vec{k}\,|$. Hence verify that the fields in Eq.(6.2) satisfy the electromagnetic wave equation.

Exercise (level A): For the fields given in Eq.(6.8), show that $\vec{E}\cdot\vec{B} = 0$.

Exercise (level A): Referring to Eq.(6.3), explain clearly (in words) why the first term describes a wave propagating to the right and the second term describes a wave propagating to the left.

Exercise (level B): Suppose we fix a coordinate system on earth such that the $x-y$ plane is parallel to the earth's surface while the z-axis is perpendicular to it. Now consider a radio wave of 10 metres wavelength, beamed into the sky at an angle of 60° to the earth. What are the possible wave vectors $\vec{k}$ of such a beam?

Exercise (level C): If you know Maxwell's equations, write them (in MKS units) and use them to derive the electromagnetic wave equations.

6.2 Standing electromagnetic waves

We can get standing electromagnetic waves much as we get standing waves of other kinds. However there is a slight subtlety involving the magnetic field. First let us consider a (real) plane wave whose electric field is:

$$\vec{E}(\vec{x}\,,t)_{\text{plane}} = E_0\,\hat{x}\cos(kz-\omega t) \tag{6.9}$$

As mentioned above, we are free to superpose any number of these. Suppose we superpose the above with an identical wave of the same amplitude travelling in the opposite direction. Then we have:

$$\begin{aligned}\vec{E}_{\text{standing}} &= E_0\,\hat{x}\Big(\cos(kz-\omega t)+\cos(-kz-\omega t)\Big)\\ &= 2E_0\,\hat{x}\cos kz\cos\omega t\end{aligned} \tag{6.10}$$

(we frequently need to use the fact that $\cos(-A) = \cos A$, i.e. that cos is an even function of its argument). This is a mode of oscillation in which the points $kz = (n+\frac{1}{2})\pi$ have zero amplitude (i.e. zero electric field) for all time. That characterises a standing wave. Of course it is not completely trivial to superpose two oppositely directed electromagnetic waves with the same phase. In practice this can be achieved by confining the radiation in an optical cavity.

We still have to specify the magnetic field. There is a standard mistake one can make here. If one writes: $\vec{B} = \frac{1}{c}\hat{k} \times \vec{E}$ then one will get the wrong answer. The reason is that $\vec{k}$ is different for the two components of $\vec{E}$ that we added. The above relation between $\vec{B}$ and $\vec{E}$ holds only for plane waves and not for their superpositions! Thus to find the magnetic field of a standing wave, we must first find $\vec{B}$ for each of the two counter-propagating waves that we superposed, and then add them. We find:

$$\begin{aligned} \vec{B}_{\text{standing}} &= \frac{E_0}{c}\Big(\hat{z} \times \hat{x}\cos(kz - \omega t) + (-\hat{z}) \times \hat{x}\cos(kz + \omega t)\Big) \\ &= 2\frac{E_0}{c}\hat{y}\sin kz \sin \omega t \end{aligned} \tag{6.11}$$

Since $\sin kz \sin \omega t = \cos(kz + \frac{\pi}{2})\cos(\omega t + \frac{\pi}{2})$, we see that the magnetic field of a standing wave is *out of phase* with the electric field. This is remarkable given that for a plane wave it was always *in phase* with the electric field! Thus in a standing wave, the points where $\vec{E} = 0$ (nodes of the electric field) are the points where $\vec{B}$ is maximum (anti-nodes of the magnetic field) and vice versa.

Exercise (level A): Sketch the electric and magnetic fields of a standing wave as a function of z, one next to the other, at the following fixed times: $t = 0, t = \frac{\pi}{4\omega}, t = \frac{\pi}{2\omega}$.

6.3 Energy of an electromagnetic wave

What is the energy of an electromagnetic wave? For any wave, the energy density is proportional to the square of the amplitude. Thus we expect:

$$\mathcal{E} = a\vec{E}^2 + b\vec{B}^2 \tag{6.12}$$

The precise answer can be derived from Maxwell's equations. Instead of going through the derivation, we first note that the constants a and b must be such that each term in the above expression has dimensions of energy. If ϵ_0 is the permittivity of the vacuum then it turns out that $\epsilon_0\vec{E}^2$ indeed has the dimensions of energy. The dimensions of $\vec{B}^2$ are $\frac{1}{c^2}$ times those of $\vec{E}^2$. Therefore, $\epsilon_0 c^2 \vec{B}^2$ also has the right dimensions. Therefore a, b must be proportional to $\epsilon_0, \epsilon_0 c^2$ respectively, with some numerical coefficients. Indeed the energy density of any configuration of $(\vec{E}, \vec{B})$, as obtained from Maxwell's equations, is:

$$\mathcal{E} = \frac{1}{2}\epsilon_0\left(\vec{E}^2 + c^2\vec{B}^2\right) \tag{6.13}$$

This result is very general and holds for any $\vec{E}, \vec{B}$. For a plane wave we have $|\vec{B}| = \frac{1}{c}|\vec{E}|$, so in this case:

$$\mathcal{E}_{\text{plane}} = \epsilon_0\vec{E}^2 \tag{6.14}$$

Note that since $\vec{E}$ is a function of $(\vec{x}\,, t)$, the energy density is time- and position-dependent. However usually we are more interested in the *average* energy density, where the average is taken over a cycle. By now it should be very familiar that the average of $\cos^2$ over a cycle is $\frac{1}{2}$, and hence the average energy density in a plane wave is:

$$\bar{\mathcal{E}}_{\text{plane}} = \frac{1}{2}\epsilon_0 E_0^2 \tag{6.15}$$

Another relevant quantity is the instantaneous energy flow per unit area per unit time. In any configuration of electric and magnetic fields in vacuum, this is given by the "Poynting vector":

$$\vec{S} = \epsilon_0 c^2 \vec{E}(\vec{x}\,, t) \times \vec{B}(\vec{x}\,, t) \tag{6.16}$$

To see its relation to the energy density, let us insert an electromagnetic plane wave solution into this. Thus we can take:

$$\begin{aligned} \vec{E} &= E_0\,\hat{x}\cos(kz - \omega t) \\ \vec{B} &= \frac{E_0}{c}\hat{y}\cos(kz - \omega t) \end{aligned} \tag{6.17}$$

from which we get:

$$\vec{S}_{\text{ plane}} = \epsilon_0 c E_0^2 \, \hat{z} \cos^2(kz - \omega t) \tag{6.18}$$

Thus the energy flow points along the direction of propagation of the wave, as it should. Taking the time-average, we find that the time-averaged energy transfer per unit area per unit time for a plane wave is:

$$\bar{S}_{\text{plane}} = \tfrac{1}{2} c \, \epsilon_0 E_0^2$$

If we divide the energy flow per unit area per unit time by the velocity of the wave, we get a quantity with dimensions of energy per unit volume, the energy density. Hence we have found that the average energy density of a plane electromagnetic wave is:

$$\bar{\mathcal{E}}_{\text{plane}} = \frac{1}{2} \epsilon_0 E_0^2 \tag{6.19}$$

which agrees with Eq.(6.15).

The units of energy density are (energy)/(length^3) which is the same as that of (force)/(length^2). So in fact the above quantity is also equal to the "radiation pressure", which is the force exerted by the electromagnetic radiation on a unit of transverse area. This is true only if the area is a perfect absorber of radiation. If it is a perfect reflector then the pressure will be double.

Literally, when a big electromagnetic wave hits you, you will feel a pressure! But this is a very small quantity in most daily applications. The radiation pressure exerted by sunlight on the earth is about 9 micro-Pascals (9×10^{-3} N/m^2). It is 7 times higher on Mercury. Radiation pressure is measurable and can be considered one more verification of the theory of electromagnetic waves. It can also be large in the context of astrophysical processes like galaxy formation, dynamics of heavy stars, motion of comets etc. Finally, the "solar sail" is an experimental spacecraft that propels itself in outer space by radiation pressure.

Exercise (level A): Show that the instantaneous (not averaged) value of the Poynting vector for a plane electromagnetic wave is:

$$\vec{S}_{\text{plane}} = \epsilon_0 c^2 |\vec{E}(\vec{x}\,, t)|^2$$

Exercise (level C): It was claimed above that the radiation pressure exerted by sunlight on the earth is about 9 micro-Pascals (9×10^{-3} N/m^2) and that it is 7 times higher on Mercury. Prove these approximate statements.

6.4 Polarisation

Let us again consider the standard plane electromagnetic wave along the z-direction:

$$\vec{E} = E_0 \, \hat{x} \cos(kz - \omega t) \tag{6.20}$$

As usual, $\vec{B}$ is completely specified as $\frac{1}{c}\hat{z} \times \vec{E}$, so in this section, we will confine our attention to $\vec{E}$.

Since the above wave has an electric field purely in the x-direction, we say that it is "plane-polarised along x". Now we will coherently superpose two different plane-polarised light beams, both propagating in the z-direction, and will find that the result depends on the relative direction of the two beams, as well as the relative phase between them.

Let's start by superposing two beams with perpendicular polarisations, and the same phase. One of them will be taken to be polarised along x and the other along y. We find:

$$\vec{E} = E_0 \Big[\alpha \, \hat{x} \cos(kz - \omega t) + \beta \, \hat{y} \cos(kz - \omega t) \Big] \tag{6.21}$$

Here α and β are two constants, which we can conveniently choose to satisfy $\alpha^2 + \beta^2 = 1$ (by scaling them and absorbing the scale into $\vec{E}_0$). We see that the above wave is the same as:

$$\vec{E} = E_0(\alpha\hat{x} + \beta\hat{y})\cos(kz - \omega t) \tag{6.22}$$

As one can easily verify, $\alpha\hat{x} + \beta\hat{y}$ is a unit vector at an angle θ to the x axis, where $\tan\theta = \frac{\beta}{\alpha}$. So, superposing plane-polarised light in the x and the y directions *with the same phase* leads to plane-polarised light at some angle to these axes. We can adjust the angle by suitably choosing α and β.

Things are quite different if we superpose the two waves with perpendicular polarisations as well as a phase difference between them. Thus we replace Eq.(6.21) by:

$$\vec{E} = E_0\Big[\alpha\,\hat{x}\cos(kz - \omega t) + \beta\,\hat{y}\cos(kz - \omega t + \delta)\Big] \tag{6.23}$$

Let us examine this superposed wave as time progresses, at a fixed position z. The wave will be effectively plane-polarised along $\hat{x}$ at the time when $\cos(kz - \omega t + \delta) = 0$. But after a short interval, we will have $\cos(kz - \omega t) = 0$ and then the wave is effectively plane-polarised along $\hat{y}$. So such a wave, said to be "elliptically polarised", behaves like a beam whose plane of polarisation is rotating with time.

Note that if $\delta = \pi$ then the second term is just $-\beta\hat{y}\cos(kz - \omega t)$ so we can again combine the two terms to get plane polarised light with a polarisation vector $\alpha\hat{x} - \beta\hat{y}$. In fact, only for $\delta = 0, \pi$ do we get a linearly polarised wave by combining the original ones. For intermediate values of the phase difference, $0 < \delta < \pi$, the light is not plane polarised but elliptically polarised.

An interesting special case is when $\alpha = \beta = \frac{1}{\sqrt{2}}$ and also the phase difference is $\delta = \frac{\pi}{2}$. In this case, the wave is:

$$\vec{E} = \frac{1}{\sqrt{2}}E_0\Big[\hat{x}\cos(kz - \omega t) - \hat{y}\sin(kz - \omega t)\Big] \tag{6.24}$$

We call this wave "circularly polarised". The characteristic of this is that the plane of polarisation rotates while the amplitude remains constant.

Clearly we have seen all this before, in the study of the two-dimensional isotropic harmonic oscillator. That is basically what an electromagnetic wave is!

6.5 Interference

Suppose we superpose electromagnetic waves of equal amplitude from two sources. We take the two sources to be separated by a distance $2L$ along the x-direction and also let the polarisation be along some common transverse direction (it doesn't matter which one). We treat the original waves from each source as plane waves along the z-direction, but actually they travel in slightly different directions depending on the point where we observe them. In this situation we can write the two waves as:

$$E_1 = E_0\cos(kd_1 - \omega t), \quad E_2 = E_0\cos(kd_2 - \omega t) \tag{6.25}$$

where d_1, d_2 are the distances of a given point from the first and second source respectively. As we will see, after adding them we no longer have a plane wave.

For a point on the z-axis, clearly $d_1 = d_2 = \sqrt{z^2 + L^2}$. However, if the point is off the z-axis then $d_1 \neq d_2$. Let us now add the two waves. We get:

$$\begin{aligned} E_1 + E_2 &= E_0\Big[\cos(kd_1 - \omega t) + \cos(kd_2 - \omega t)\Big] \\ &= 2E_0\cos\Big(\tfrac{1}{2}k(d_1 + d_2) - \omega t\Big)\cos\Big(\tfrac{1}{2}k(d_1 - d_2)\Big) \\ &= 2E_0\cos\Big(kd - \omega t\Big)\cos\tfrac{1}{2}k\Delta \end{aligned} \tag{6.26}$$

where $d = \frac{1}{2}(d_1 + d_2), \Delta = d_1 - d_2$. For a source of fixed average distance from the emitters, as we vary the relative distance the wave appears to be modulated (i.e. to have a varying amplitude) due to the second factor. Thus the effective amplitude is $2E_0 \cos \frac{1}{2}k\Delta$ and the effective intensity is the square, or $4E_0^2 \cos^2 \frac{1}{2}k\Delta$. This varies from a maximum of $4E_0^2$ to a minimum of 0, leading to interference fringes. The dark fringes occur when $\cos \frac{1}{2}k\Delta = 0$, i.e. when $\Delta = (2n+1)\frac{\pi}{k}$, while the bright fringes occur in between the dark ones at $\cos \frac{1}{2}k\Delta = \pm 1$, i.e. $\Delta = \frac{2n\pi}{k}$.

It is an easy exercise, given the coordinates of the point of observation, to calculate the average distance and separation in terms of those coordinates. It is also amusing to note that the the bright and dark points lie on hyperbolas (because given two centres, the set of all points with $d_1 - d_2 =$ constant is a pair of hyperbolas).

6.6 Coherence and bandwidth

Above we have studied ideal electromagnetic waves:

$$\vec{E} = \vec{E}_0 \cos(kz - \omega t + \delta) \tag{6.27}$$

These have a fixed angular frequency ω (equivalently, a fixed wave number, and fixed wavelength) as well as a fixed amplitude and phase. Thus all the parameters $\vec{E}_0, \omega, \delta$ are assumed fixed. However, real radiation is not like this. Consider a lightbulb. It is basically a hot filament that emits light. This departs from the above formula in many ways:

(i) the light does not have a fixed frequency ω. Rather, it has a spread of frequencies $\Delta\omega$ peaked about some particular value. For example a glowing yellow filament will emit radiation peaked at around 5700 Angstroms (5700×10^{-10} metres). The corresponding frequency is:

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^8 \,\mathrm{m/sec}}{5.7 \times 10^{-7}\,\mathrm{m}} \simeq 5.2 \times 10^{14}\,\mathrm{Hz} = 520\,\mathrm{Terahertz}$$

and of course $\omega = 2\pi\nu$. However the range of wavelengths goes from around 300 Angstroms all the way to extremely large values (in the infrared). In fact a bulb filament is roughly a "blackbody" which means it emits over a wide range of frequencies when heated up. This is because at a high temperature (a few thousand degrees Centigrade) the various molecules in the bulb vibrate, not at one frequency but at a range of frequencies. Thus one can think of the emitted light as being a collection of billions of individual beams, each with the above form but with differing values of ω.

(ii) each beam of light does not have the same amplitude and polarisation (encoded in the vector $\vec{E}_0$) since the emitting molecules are moving around randomly.

(iii) each beam is not emitted at the same instant, so the phase δ is not the same.

The above facts are not particularly surprising. The wavelength of light is so incredibly short that any kind of vibration of the emitter will vary the properties of the wave slightly. Besides the randomness of the emitter, interactions with matter can also be expected to make a coherent light beam become incoherent. To reduce the frequency spread of a given beam one can use filters, however it is more difficult to make the phase the same for all components. Radiation that is made up of waves in step with each other is called *coherent radiation*. This is possible to achieve only with a laser. That is why the invention of the laser opened up so many new areas of research.

Remarkably, even in a vacuum and even with a very good emitter, it turns out that light waves inevitably become incoherent after travelling some distance if they have a nonzero frequency spread or *bandwidth*. This distance is called the *coherence length*. Let us try to estimate this length knowing the velocity of the wave (in this case, c) and the bandwidth $\Delta\nu$. Very close to the laser, the light is a coherent superposition (i.e. all the components have the same phase). Now let's first imagine there is no spread of frequencies at all (ideal laser). Then the wave is precisely as in the above equation and it remains coherent indefinitely. In particular the crests and troughs of any component

are perfectly aligned with those of any other. Thus in the ideal situation, light will propagate for an infinite distance without any loss of coherence.

However once there is a spread in frequency, the situation changes. Suppose all the components have exactly the same phase near the emitter, but a frequency spread $\Delta\nu$ among them. Then very close to the emitter they will constructively add. But after some distance ΔL, all the phases will be completely randomised due to the frequency spread. How much is this distance?

To estimate this, assume that the waves are coherent at the initial point $z = z_0$, and their mean wave number is $k = k_0$ with a spread $\Delta k \ll k$. Suppose we work at the time $t = 0$. In this situation we will typically encounter a superposition like:

$$\cos(kz) + \cos\Big((k + \Delta k)z\Big) \simeq 2\cos(kz)\cos\left(\frac{(\Delta k)z}{2}\right) \tag{6.28}$$

Now notice that the second factor, which modulates the wave, changes from maximum to minimum after a shift of z by $L_c = \frac{\pi}{\Delta k}$. If we have a collection of waves whose frequencies are uncertain by an amount upto Δk, then all mutual phase information between them will be lost after traversing the distance L_c. Thus we say that:

$$L_c \simeq \frac{\pi}{\Delta k} = \frac{c}{2\,\Delta\nu} \tag{6.29}$$

is the coherence length of the wave. Here we have used $k = \frac{2\pi\nu}{c}$. Equivalently, we can say that:

$$L_c\,\Delta\nu \simeq \frac{c}{2} \tag{6.30}$$

We can convert this into a *coherence time*:

$$T_c = \frac{L_c}{c} \simeq \frac{1}{2\,\Delta\nu} \tag{6.31}$$

Finally, we can write the coherence length and coherence time in terms of the spread in wavelengths, which is more intuitive. Since $\nu = \frac{c}{\lambda}$, we have:

$$\Delta\nu = -\frac{c}{\lambda^2}\Delta\lambda \tag{6.32}$$

from which we find:

$$L_c = \frac{\lambda^2}{2\Delta\lambda} \tag{6.33}$$

To see some concrete figures, suppose a beam of radiation has a mean wavelength of 5700 Angstroms (yellow) with a spread of 100 Angstroms on each side. Then $L_c \sim 1.6 \times 10^{-3}$ cm. In a laser we can reduce the bandwidth to as little as 0.02 Angstroms and get a coherence length of around 8 cm. Notice that if we take long-wavelength light (red) then, for the same bandwidth, the coherence length is greater.

Exercise (level A): Consider a superposition of an infinite number of electromagnetic waves in the same direction but with different wave numbers:

$$\vec{E}(z,t) = \vec{E}_0 \int_{-\infty}^{\infty} dk\, A(k)\cos(kz - \omega t) \tag{6.34}$$

where $\omega = c|k|$. Show that this wave satisfies the electromagnetic wave equation in free space.

Exercise (level B): Write down the electric and magnetic fields, as vectors, for a monochromatic plane wave of amplitude E_0, frequency ω and phase angle $\delta = 0$ which is travelling along the line from the origin to the point $(x, y, z) = (1, 1, 1)$ and has a polarisation along the $x - z$ plane.

---

7 Elastic Properties of Matter

We have considered a variety of oscillating systems so far. The central point in all these discussions was that the system responds linearly to an external force, when the applied force is small, and the change in the system is small. However, we did not always *quantify* the smallness carefully, especially in the case of systems that are close to real systems – such as vibrating strings or longitudinal waves in a gas. In such systems the restoring force, which is the response of the system to the external force, is determined by the properties of material in question. The restoring property is in general terms referred to as the elasticity of the material. We will examine such properties in detail.

Problem (Level A): What is more elastic, a rubber band or a steel wire? How do cork and block of plastic compare as regards their elasticity?

7.1 Stress and Strain

When a wire is subjected to an elongating force $F^{(a)}$ applied to it, the elongation of the wire is found to be directly proportional to the applied force, as long as the elongation is tiny compared to the length of the wire. Such a force may be realised by either suspending the wire from a rigid support and hanging a weight by the other end, or by pulling the ends of the wire by equal and opposite forces, $F^{(a)}$. The elongation is directly proportional to the length of the wire and inversely proportional to the cross-section area of the wire. In other words,

$$\Delta l \propto F^{(a)}(l/A)$$

An equilibrium is quickly attained, in that the elongation of the wire reaches a fixed value, so that there must be a restoring force which opposes and balances the applied force: $|F^{(r)}| = |F^{(a)}|$. Thus, at equilibrium we have a restoring force proportional to the extension:

$$F^{(r)} \propto \Delta l$$

This proportionality is reminiscent of Hooke's law, which states that the restoring force is proportional to the extension (of a spring). Combining the two equations, we have

$$\frac{\Delta l}{l} \propto \frac{F^{(r)}}{A}$$

The quantity on the left is called the *strain*, the ratio of the extension to the original length, while the quantity on the right is called *stress*, the restoring force per unit area. The proportionality constant in the case of linear elongation (or equivalently, for linear compression) of an elastic material is given a special name *Young's Modulus*, after Thomas Young, who studied elasticity in great detail. Young's relationship, in conformity with Hooke's law, is usually stated as *Stress is proportional to strain for small extensions (or compressions) of a material.* Note, that strain is a dimensionless quantity and stress, as well as Young's modulus, have units of pressure. Since this modulus usually relates to wires or rods under axial tension, the modulus is also called tensile modulus, or tensile strength.

Taking the previous equation forward, we then have at equilibrium

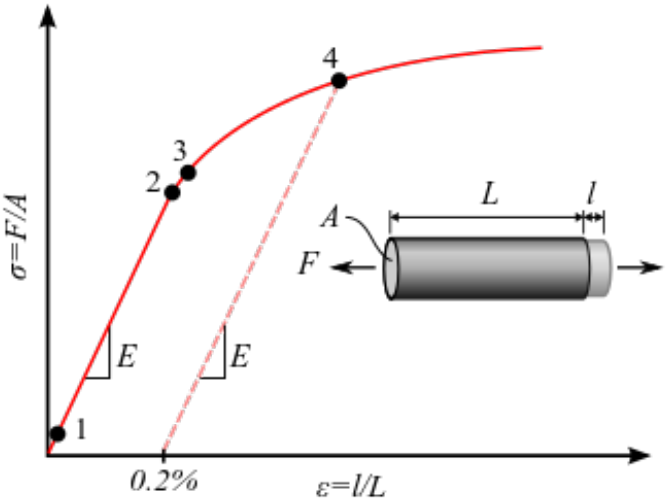
$$\frac{\Delta l}{l} = \frac{1}{Y}\frac{F^{(r)}}{A} = \frac{1}{Y}\frac{F^{(a)}}{A}$$

which is often written as

$$\frac{\Delta l}{l} = \frac{P}{Y},$$

where P is the applied force per unit area *under equilibrium with the restoring force.*

The linear relationship between stress and strain breaks down if the strain is large. The graph below shows stress vs. strain for a steel wire. Note that in the graph $\sigma = F^{(r)}/A$ and $\epsilon = \Delta l/l$. The inset in the graph shows the free body diagram of a rod under longitudinal stress in which the rigid support structure is replaced by an equivalent force. The stress-strain curve is linear up to point 1 on the graph. After this stress increases very rapidly, and the material tends to 'flow'. The material is expected to return to its original shape and size once the stress is removed, as long as it is in the linear region. However, in engineering practice the stress corresponding to an offset strain of 0.2% (refer to point 2 in the figure, and read off the stress corresponding to the linear portion corresponding to a strain of 0.2%) is taken to be the upper limit of the acceptable limit of stress, before the material will be considered to be unsafe for use. If the stress increases beyond this point, the strain builds up rapidly, and the material starts flowing even without an increase in the applied force. The restoring force fails to keep up with the applied force, as the applied force increases.



It is observed that the elongation of a free wire is always accompanied by a concurrent reduction of the diameter of the wire, and in the more general case, a longitudinal extension is accompanied by a transverse compression. In analogy with the longitudinal strain we can define a transverse strain. In the case of a wire, the transverse strain would be defined as $\Delta d/d$, where d is the diameter. Experimentally, the transverse and longitudinal strains are found to be related to each other

$$\frac{\Delta d}{d} = -\sigma \frac{\Delta l}{l} = -\sigma \frac{P}{Y}.$$

The negative sign indicates that longitudinal extension implies transverse compression and vice-versa. The proportionality constant σ is called the *Poisson's ratio*.[3] Had we considered a bar instead of a wire, the transverse strain would have been defined as $\Delta w/w$ and $\Delta t/t$ where w and t refer to width and thickness of the bar. For the bar under longitudinal extension, we have

$$\frac{\Delta w}{w} = \frac{\Delta t}{t} = -\sigma \frac{\Delta l}{l}.$$

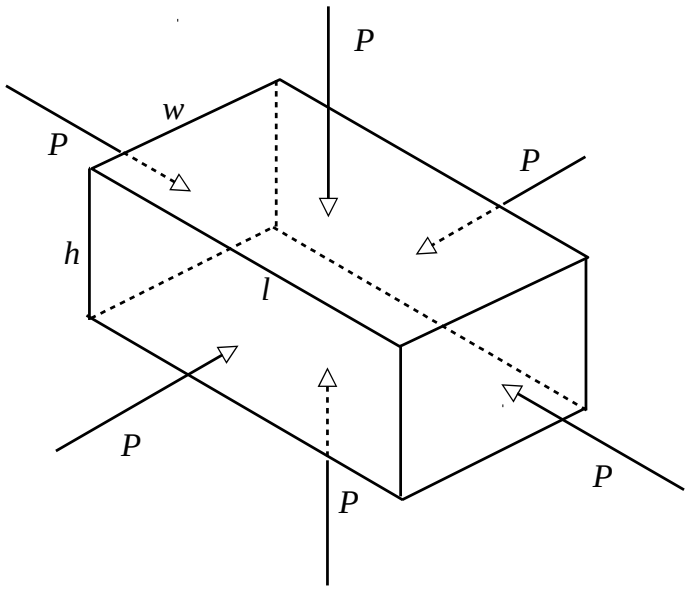
Problem (Level A) : Two identical lifts are installed in two buildings of different heights. The lifts are suspended by identical steel cables of different lengths, $l_A = 2\,l_B$. For identical loads, Which cable is under greater stress? Which is under greater strain?

---

[3]Note, that in engineering practice, Young's modulus is denoted by the symbol E, while Poisson's ratio is denoted by the symbol ν, and the symbols for stress and strain are σ and ϵ.

7.2 Uniform Strain

In the previous section we considered the case where there is elongating strain along one direction and a resultant compressive strength along the other two (transverse) directions. But, in practice we are often faced with a situation where a material is under uniform compressive strain (from all directions). How do we describe the volumetric strain?



Consider a block (rectangular parallelepiped) of sides l, w, h, which is under uniform compressive pressure from all directions. The strain along any one direction (say the l direction) comprises a direct *compressive* strain due to the force applied along the same direction ($\Delta l_{\text{direct}}/l = -P/Y$) and an indirect *elongating* strain due to the forces applied along the two transverse directions. The indirect strain due to the forces along each of the transverse directions is given by $+\sigma(P/Y)$. The sign of the indirect strain is opposite to the sign of the direct strain. Hence the net strain along the l direction is

$$\begin{aligned}\frac{\Delta l}{l} &= -\frac{P}{Y} + 2\sigma\frac{P}{Y} \\ &= -(1-2\sigma)\frac{P}{Y}\end{aligned}$$

Exactly identical arguments lead us to the expressions for strains along the w and the h directions:

$$\begin{aligned}\frac{\Delta w}{w} &= -(1-2\sigma)\frac{P}{Y} \\ \frac{\Delta h}{h} &= -(1-2\sigma)\frac{P}{Y}\end{aligned}$$

The net effect of the forces is a volumetric compression. The volume of the block is $V = lwh$, so the relative change in volume is

$$\frac{\Delta V}{V} = \frac{\Delta l}{l} + \frac{\Delta w}{w} + \frac{\Delta h}{h}$$

Using the expressions for the linear strains along the three directions in the expression for volume strain, we have

$$\frac{\Delta V}{V} = -3(1-2\sigma)\frac{P}{Y}$$

The ratio on the LHS is the volume strain. The ratio $-P/(\Delta V/V)$ is is called the compressibility or bulk modulus, K. Evidently,

$$\frac{1}{K} = (1 - 2\sigma)\frac{3}{Y}$$

7.3 Shear

In addition to longitudinal and bulk strains, we come across another kind of strain, namely shear strain. By shear we mean an elastic deformation of a solid, which results in differential displacements of layers of the material in a direction *parallel* to the applied force. The usual picture of a cube under shear is shown as one in which one face remains fixed while the face opposite to the fixed face moves parallel to it under the action of a force. Notice, that in case of a perfectly rigid body, such a force would result in acceleration along the applied force, but in the present case we take one face to stay fixed while the other moves relative to it, creating a restoring, elastic force, until equilibrium is reached. There is no net force on the body, nor is there a net torque.

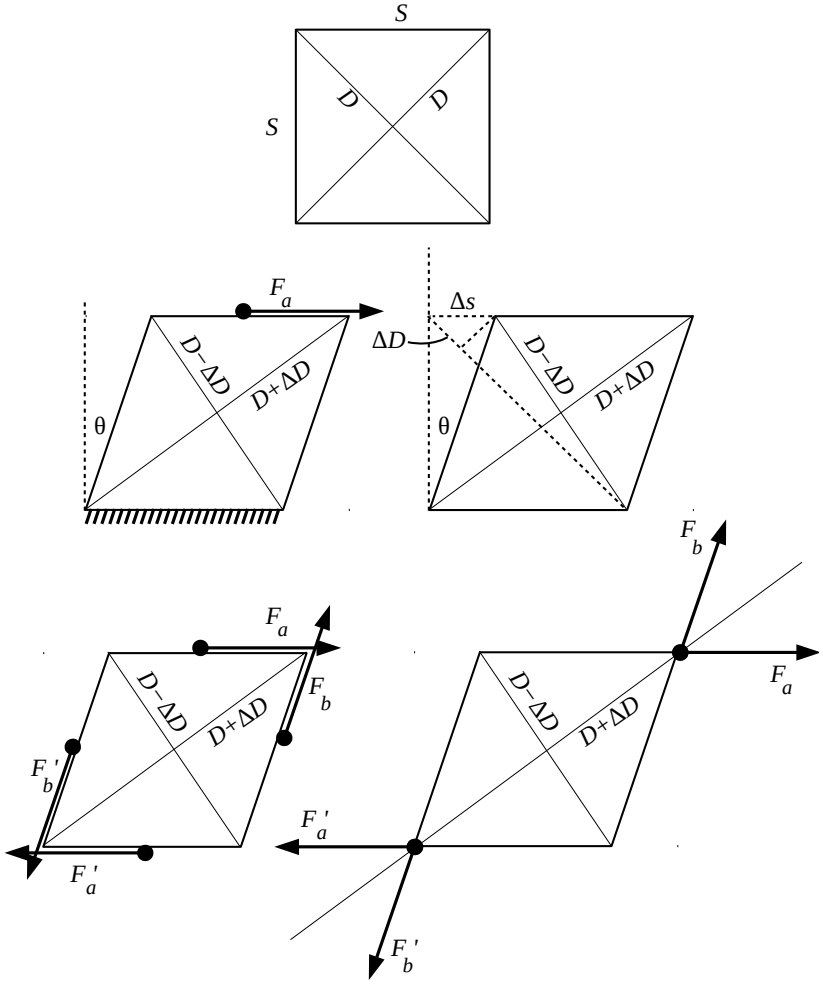
Clearly, the contact between the fixed face and the rigid platform is maintained because of some force due to the contact itself. This force is not known to us directly, but we can derive it by enforcing the conditions of zero net force and zero net torque. Let, the applied force be F_a. Let the area of the face on which the force acts be A_{face}, then the shear stress is (F_a/A_{face}). The effect of the shear can be quantified by a (single) shear angle θ as shown in the figure. A single shear angle indicates a linear strain: layers further away from the fixed face are displaced more than the layers closer to the fixed face. The ratio of the shear stress to the shear strain is called the shear modulus, μ:

$$\mu = \frac{F_a/A_{\text{face}}}{\theta}.$$

Shearing does not change the length of the side or the area of the face, but it does change the length of the diagonal. The shear strain can be written in terms of the change in the lengths of the diagonals (refer to parts 2, 3 of the diagram):

$$\begin{aligned}\theta &\approx \frac{\Delta s}{s} \\ &\approx \frac{\Delta D\sqrt{2}}{D/\sqrt{2}} \\ &= \frac{2\Delta D}{D}\end{aligned}$$

Since there is no net force on the block, the role of the contact in creating the shearing situation must be to provide an equal and opposite force to F'_a. This pair of forces do cancel each other, but since they do not act at the same point, and do not act on the centre of mass, they result in a torque. So the assumed force F'_a at the contact is not adequate for fulfilling the constraints. (See parts 4 of the diagram.) To null the torque resulting from F_a and F'_a, we must consider another pair of opposing forces F_b and F'_b acting along the faces perpendicular to the ones considered before.



For equilibrium we must have

$$F_a = F_a' = F_b = F_b' \quad (= F)$$

The free body diagram of the shear condition is shown in the last part of the figure. The net effect of the shear is to elongate one diagonal and compress the other, while the face diagonals of the cube continue to remain orthogonal to each other. The forces, when resolved along the diagonals, yield

$$\begin{aligned} F_{\text{elong}} &= \frac{F_a}{\sqrt{2}} + \frac{F_b}{\sqrt{2}} &&= \sqrt{2}F \\ F_{\text{compr}} &= \frac{F_a'}{\sqrt{2}} + \frac{F_b'}{\sqrt{2}} &&= \sqrt{2}F \end{aligned}$$

The elongating and compressive forces act on the same cross-section area, $A' = s^2\sqrt{2}$, which is the area of the cross-section perpendicular to the plane of the diagram and containing the shorter diagonal. The strain along the diagonal is a sum of the direct and indirect strains:

$$\begin{aligned} \frac{\Delta D}{D} &= \pm\left[\frac{1}{Y}\frac{\sqrt{2}F}{s^2\sqrt{2}} + \sigma\frac{1}{Y}\frac{\sqrt{2}F}{s^2\sqrt{2}}\right] \\ \frac{\Delta D}{D} &= \pm\frac{1+\sigma}{Y}\frac{F^{(a)}}{A_{\text{face}}} \end{aligned}$$

where the $\pm$ signs correspond to elongation and compression respectively.

Combining the two expressions for the shear strain we get

$$\theta = 2\frac{1+\sigma}{Y}\frac{F^{(a)}}{A_{\text{face}}}$$

We immediately recognise the quantity $Y/[2(1+\sigma)]$ as being equal to the shear modulus μ.

Problem (Level B): Show that the Poisson's ratio of a homogeneous, isotropic, elastic material (in the linear limit) must lie in the range $-1.0 < \sigma < 0.5$.

Problem (Level B): Show that for a cube under pure shear the change in volume is zero to first order in the shear strain. It is easier to to do this if the strain is written in terms of the length of the diagonal.

7.4 Torsion

There is another kind of shear, which is due to a differential angular displacement of layers of a solid – in contrast to differential linear displacement we discussed in the previous section. Imagine a solid cylinder whose base is fixed and whose top is twisted relative to the base. All points lying within a circle parallel to the base, as shown in the diagram, will experience an angular displacement ϕ. The value of ϕ increases linearly with the distance h of this imaginary circle from the fixed base. The base circle experiences no angular displacement, while the circle at the top has the largest angular displacement. The angular displacement of any point within or on the surface of the cylinder can be expressed as

$$\phi = \theta(h/R)$$

where h is the height of the point above the base and R is the radius of the cylinder. And θ is the twist, or torsional angle. Note that this relationship is true, irrespective of whether the point lies on the surface or inside the cylinder.

The angular displacement on account of the torsion results in a tangential linear displacement of the point, so for a small cuboidal element of the cylinder (within or on the surface), the effect is identical to a shear. The shear strain is, as before, given by the angle θ:

$$\theta = \frac{1}{\mu}\frac{\Delta F_a}{A}$$

where ΔA is the area of the face of the small volume element that is parallel to the (fixed) base, and equals $\Delta r \Delta w$, so we have

$$\theta = \frac{1}{\mu}\frac{\Delta F_a}{\Delta r\, \Delta w}$$
$$\text{or } \Delta F_a = \theta \mu\, \Delta r\, \Delta w$$

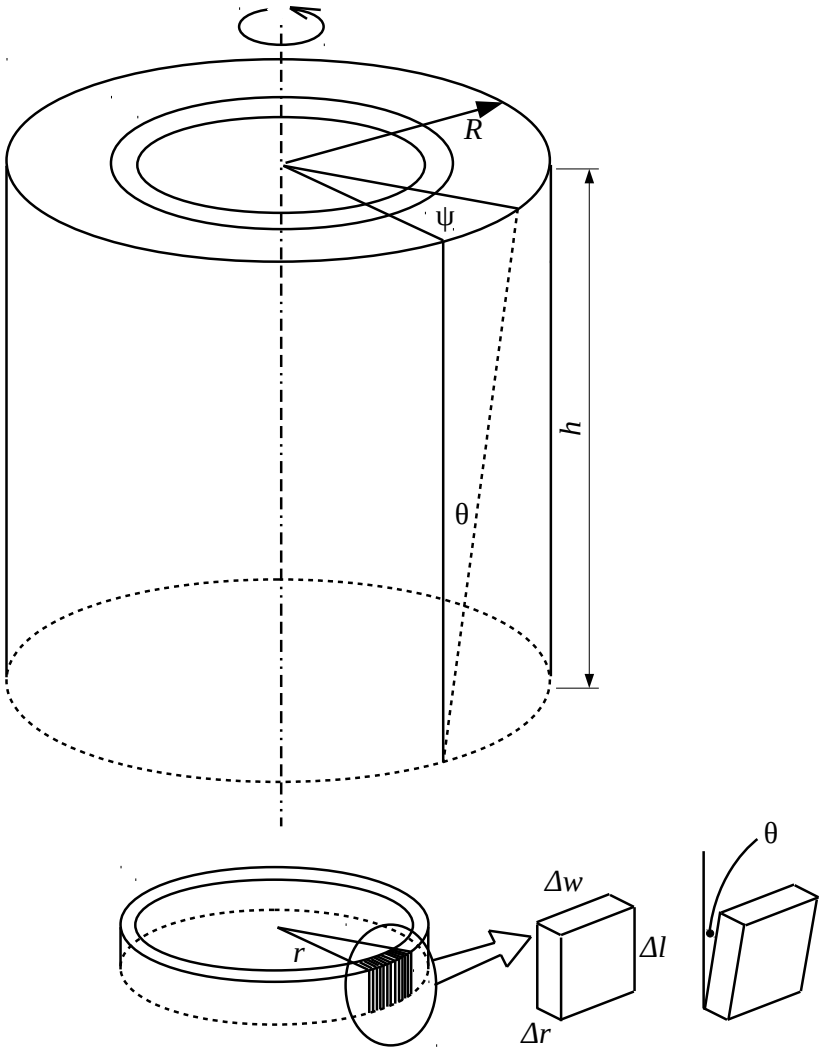
We note that the shearing force on a tiny element is related to the torque which is causing the twist, and the shear force on the tiny volume element is related to the torque by $\Delta\tau_a = r\Delta F_a$, where r is the radial coordinate of the volume element. We also note, that $\theta = \psi r/h$, and $\Delta w = r\Delta\phi$ where ψ is the azimuthal coordinate. We can rearrange and simplify the equation for torsion as

$$\Delta\tau_a = \frac{\mu\psi r}{h}\,(\Delta r)\,(r\Delta\phi)$$
$$\therefore \tau_a = \int_0^R \int_0^{2\pi} \frac{\mu\psi r}{h}\, r^2\, dr\, d\phi$$
$$= \frac{\mu\psi}{h}\,\frac{\pi R^4}{2}$$

If we define $\eta = \pi R^4/2\mu l$, then the last equation takes a form that is similar to the case of a block under shear:

$$\psi = \frac{1}{\eta}\,\tau,$$

showing that the twist, or torsional shear is proportional to the applied torque and the two are related by a factor that depends on the geometry of the cylinder and the modulus of rigidity, μ.

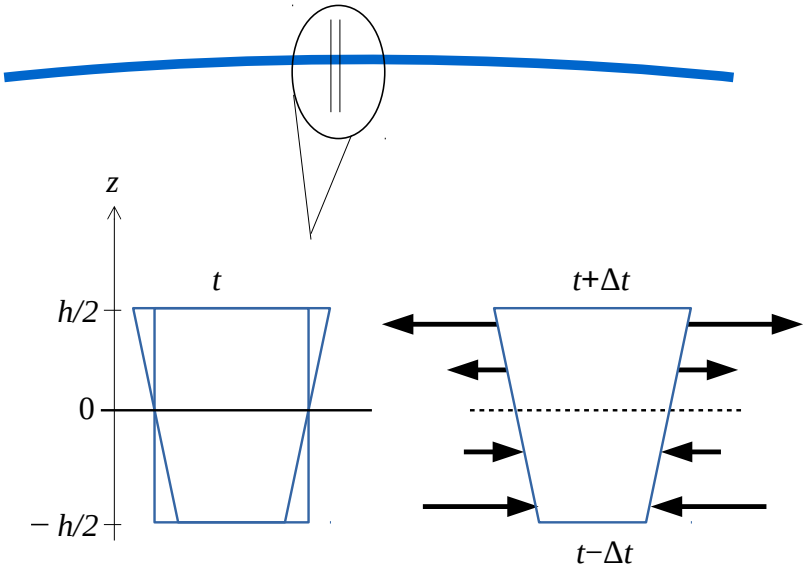


Problem (Level B): Find the frequency of small torsional oscillations for a rod of radius a, length l, and Young's Modulus Y, when it is subject to a small torque about its axis of cylindrical symmetry.

---

8 Bending of Beams

The elastic properties of solids are very important in designing structures of all kinds. A commonly occurring form of structures involves beams and rods of various shapes. We will study the Euler–Bernoulli theory of bending of elastic beams that forms the basis for determining the appropriate shapes of beams in building structures.



Shape of slice before and after bending
Slice has width w into the plane of the drawing

Compressive and extensive forces below and above the neutral line

For simplicity let us take a beam of rectangular cross section (sides $w \times h$), which is bent by a suitable set of forces (we will not worry about their details at this point). Let us take an arbitrary slice of the beam perpendicular to its length, of thickness t in the normal state. In this state the slice is a rectangular parallelipiped. When the beam is bent this slice is deformed and the thin face of the slice changes from being a rectangle to a trapezium. The outer edge is extended, and the inner edge is compressed. There is a certain line along which the thickness remains unchanged, we call this the *neutral line*. On both sides of the neutral line the slice is under strain, and the strain is non-uniform. It increases as we move away from the neutral line. Consistent with our approach of linear response, we assume that the strain varies linearly with the distance z from the neutral line. This assumption automatically takes account of the fact that the strain is negative on the inner side $(-z)$ of the neutral line, and positive on the outer side $(+z)$ of the neutral line. Thus,

$$\Delta t \propto z$$

By comparing similar triangles, we obtain the following relationship

$$\frac{t/2}{R} = \frac{\Delta t/2}{z}$$

The longitudinal strain on a slice of thickness δz at location z is $\Delta t/t$. This strain is due a force δF acting along the local tangent. The stress due to the force is $\delta F/w\,\delta z$. For a rectangular cross-section of the beam, w is constant, independent of the z, but for other shapes, in general, w varies with z. In particular, for a circular cross-section of radius r_0, $w = 2r_0$ for $z = 0$ and $w = 0$ for $z = r_0$. We therefore find the following relationship

$$\frac{\delta F}{w(z)\,\delta z} = Y\frac{\Delta t}{t}$$
$$\text{so, } \delta F = Y\frac{z}{R}w(z)\,\delta z$$

Note that the force is of compressive type below the neutral line and extensive above it, so the net effect of this set of forces in any radial plane is to give rise to a moment (which is the cause of bending). The bending moment can be written as

$$\begin{aligned}\delta M &= \delta F(z)\, z \\ &= Y\frac{z}{R}w(z)\,\delta z\, z \\ M &= \int Y\frac{z^2}{R}w(z)\,dz\end{aligned}$$

It is convenient to define a quantity J which we can understand loosely as the "moment of the cross-section"

$$J = \int z^2 w(z)\,dz.$$

which is a way of describing how the material in the beam is distributed away from the neutral line.

Using this definition of J, the equation for bending will take the simple form:

$$M = JY/R$$

It is more instructive to write the equation in the form

$$\frac{1}{R} = \frac{M}{JY}$$

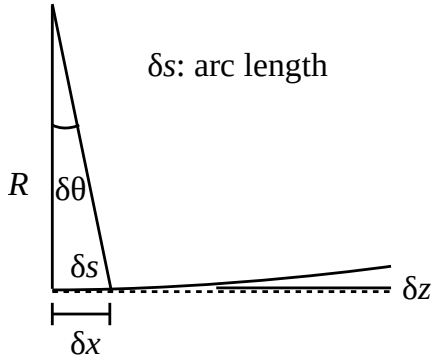
What this equation tells us, is that the curvature to which the rod is bent depends not only on the bending moment, and the Young's modulus of the material of the rod, but also to how the shape of the cross-section is. The information about the shape of the cross-section is contained in J.

Problem (Level A): Compare the bending moments needed to bend two solid beams of circular cross-section, of same length and material, but with the second having twice the radius of the first.

Problem (Level B): Find the moment of the cross-section for a circular cross-section beam and a square cross-section beam, both having the same area of cross section.

We have seen how the cross-section of a beam affects it stiffness, or its resistance to bend. Let us now turn to finding the shape of the bent beam. The starting point is the equation above that relates the (local) radius of curvature $R(x)$ to the moment. What exactly is $R(x)$? It is the radius of the circle which touches the point on the neutral line under consideration and shares a tangent with the curve assumed by the neutral line.

The radius of curvature, for small distortions of the beam can be derived from the the diagram below:



$$\delta s = R\,\delta\theta; \quad \delta x \approx \delta s$$

Hence

$$\frac{1}{R} = \frac{d\theta}{ds}$$

Furthermore,

$$\frac{d\theta}{ds} \approx \frac{d\theta}{dx} \approx \frac{d}{dx}\frac{dy}{dx}$$

Hence the radius of curvature, which is not constant for all x, is given by

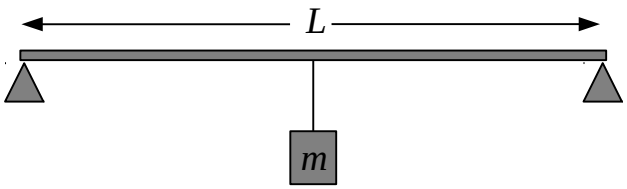
$$\frac{1}{R(x)} = \frac{d^2z}{dx^2}$$

where $z(x)$ is the equation of the neutral locus. This expression is approximate, but good enough for the small deflections we are considering.

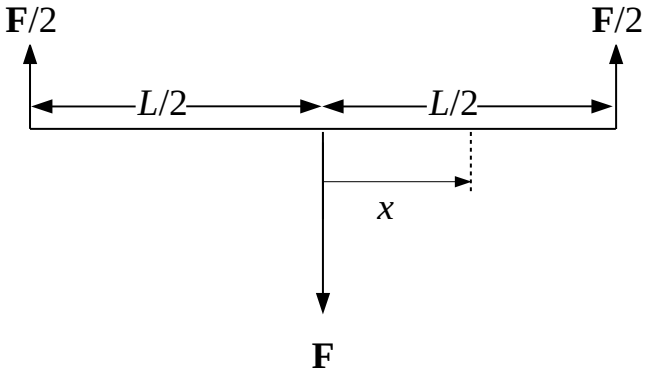
What about M? M is not the same everywhere along the beam, it is zero at the fixed point, and the largest at the point of maximum deflection from the original. The equation for the shape is thus

$$\frac{d^2z}{dx^2} = \frac{M(x)}{JY}$$

Let us see how this equation is applied to a beam supported at its ends and under load F at its mid-point as shown below. An apparatus of this type is often used for determining the Young's modulus of the material of the beam.



We will assume that the load is much greater than the weight of the beam itself, and that the deflection is well within elastic limit. Let the origin of the coordinate be at the mid point. The shape of the beam is symmetric, so it is enough the solve for one half of the beam. Refer to the free body diagram below. The end supports provide an upward force $F/2$ each to balance the load. This pair of forces also provide a null torque, which is consistent with the fact that the beam does neither rotate nor translate.



The moment of the forces applied to the right of the point located at x is given by

$$M(x) = (F/2)(L/2 - x)$$

The moment of the forces applied to the left of the point located at x is given by

$$M(x) = -(F/2)(L/2 + x) + Fx$$

The two moments are equal and opposite as expected on grounds of equilibrium of an infinitesimal slice of the beam located at x in its deformed state. (This is consistent with our analysis that a slice of rectangular section becomes a trapezium due to the couple created by a pair of compressive and extensive forces on the upper and lower sides neutral locus) Hence the equation of the shape of the beam is

$$\frac{d^2z}{dx^2} = \frac{1}{JY}\frac{F}{2}\left(\frac{L}{2} - x\right)$$

Integrating once we get

$$\frac{dz}{dx} = \frac{1}{JY}\frac{F}{2}\left(\frac{Lx}{2} - \frac{x^2}{2} + A\right)$$

The integration constant is found to be zero, by requiring that the tangent to the beam at its mid-point is horizontal, i.e. $dy/dx = 0$ at $x = 0$. Integrating once again we get

$$z = \frac{1}{JY}\frac{F}{2}\left(\frac{Lx^2}{4} - \frac{x^3}{6} + B\right)$$

Since the beam is undeflected at the end-point we have $z = 0$ at $x = L/2$. Substituting this condition, we get $B = -L^3/24$. Hence the shape of the beam is given by

$$z = \frac{F}{JY}\left(\frac{Lx^2}{8} - \frac{x^3}{12} - \frac{L^3}{48}\right)$$

Shapes of beams under different loading and support conditions can be obtained in a manner similar to the one above.

Problem (Level A): Find the deflection of the beam at its centre for the case discussed above.

Problem (Level B): Find the shape of the beam that is rigidly supported at one end and has a load applied at the other.

Problem (Level C): Find the shape of the beam that is supported at its ends, but bends under its own weight.
