Waves and Optics

Dhruva Karkada

Spring 2018

Contents

1 Simple Harmonic Oscillator

Consider an object moving in a quadratic potential well. Then the force from the potential is

$$\mathbf{F} = -\frac{dU}{dx}.$$

Since the potential is quadratic, the resulting force is linear.

By convention, we let the coordinate $x = 0$ where $\mathbf{F} = 0$, which is the bottom of the potential well. This is the equilibrium position for our object. Notice that the force always points towards the equilibrium position, so we call it a *restoring force*. To calculate the equation of motion, we determine the net force acting on the object in terms of the given constants. Then, by Newton's 2nd law, we equate this to ma. Since acceleration is the second derivative of position, we can denote this as $m\ddot{x}$.

In general, we usually end up with a differential equation of the form

$$a\ddot{x} + bx = 0,$$

where a is related to the inertia of the object, and b is related to the restoring force. We can rewrite this in the form of

$$\ddot{x} + (\omega_0)^2 x = 0.$$

where $\omega_0 = \sqrt{b/a}$. The solution to this differential equation is given by

$$x(t) = A\cos(\omega_0 t + \phi),$$

where ω_0 is the natural frequency of oscillation, in radians per second. A and ϕ are the amplitude and phase of the oscillation, and are the undetermined constants for the general solution, and require initial conditions to determine.

An alternate formulation of the solution is

$$x(t) = A_1\cos(\omega_0 t) + A_2\sin(\omega_0 t),$$

where the phase is encoded in the sine term. This formulation is consistent with the idea that the general solution of a differential equation of this form is the linear combinations of the fundamental solution set for the equation.

The period of oscillation T is related to the angular frequency by

$$\omega_0 T = 2\pi.$$

1.1 Examples: Equations of motion

Spring

For a mass on a spring, the force is given by $\mathbf{F} = -kx$. Then the equation of motion is $-kx = m\ddot{x}$, or

$$m\ddot{x} + kx = 0.$$

The general solution of this system has a natural frequency

$$\omega_0 = \sqrt{\frac{k}{m}}.$$

Pendulum

For a pendulum, we have a slightly more complicated setup. First, we notice that the net force on a pendulum is given by $\mathbf{F} = -mg\sin\theta$. First of all, this is not in terms of x; we avoid this by simply changing our coordinates to θ.

But this is still not linear in θ. This is because the potential is not quadratic (in this case, $U = mgl(1-\cos\theta)$). However, at small θ values, the potential can be *approximated* as quadratic (using a Taylor series), which means that the force can be *approximated* as linear: $\mathbf{F} \approx -mg\theta$.

We will also need to express the tangential acceleration of the pendulum in terms of θ. Since $x = l\theta$, we have $\ddot{x} = l\ddot{\theta}$. Then the equation of motion is $ml\ddot{\theta} + mg\theta = 0$, which is simplified to

$$l\ddot{\theta} + g\theta = 0.$$

The general solution of this system has a natural frequency

$$\omega_0 = \sqrt{\frac{g}{l}}.$$

Torsion Pendulum

The torque is linear with the rotation angle ϕ (Hooke's law). We use the torque equivalent for Newton's 2nd law: $\tau = I\ddot{\phi}$. This yields

$$I\ddot{\phi} + c\phi = 0$$

with natural frequency

$$\omega_0 = \sqrt{\frac{c}{I}}.$$

2 Damped Oscillator

The general solution of the simple harmonic oscillator is a sinusoidal function at constant amplitude. This implies oscillation at full amplitude ad infinitum. In reality, oscillators decay with time, due to the effects of friction. The damping force is, in general, given by

$$\mathbf{F}_d = -bv - cv^2$$

where v is the magnitude of velocity, and the force opposes the direction of velocity.

The linear term is known as the "viscous term" and is related to the energy lost to a viscous fluid via friction. The shearing of the fluid is a laminar flow, and some of the energy is returned to the moving object. Thus, the term is linear. On the other hand, the quadratic term is related to the energy lost to a fluid due to turbulence. Because the flow is turbulent, the energy lost to the fluid (which is proportional to v^2) is lost permanently. Thus, the term is quadratic.

At low velocities, the viscous term dominates and the quadratic term is negligible. We will only consider systems in this small-velocity limit.

We will consider a damped spring system, where the net force is the sum of the spring force and the viscous damping force:

$$\mathbf{F}_{net} = m\ddot{x} = -kx - bv = -kx - b\dot{x}.$$

Dividing both sides by m and rearranging, we have

$$\ddot{x} + \gamma\dot{x} + (\omega_0)^2 x = 0.$$

where γ is called the damping constant, and ω_0 is our natural oscillation frequency. This form is the general form for the damped oscillator.

The differential equation is a homogeneous second-order linear equation with constant coefficients, so we find the roots of the characteristic equation:

$$r = -\frac{\gamma}{2} \pm \sqrt{\left(\frac{\gamma}{2}\right)^2 - (\omega_0)^2}$$

which, for convenience, we'll rewrite as

$$r = -\frac{\gamma}{2} \pm i\omega.$$

Note that this ω is different from the natural frequency; instead, it is given by

$$\omega = \sqrt{(\omega_0)^2 - \left(\frac{\gamma}{2}\right)^2}.$$

Now, we enumerate the cases for the solution of the differential equation.

1. UNDERDAMPED: When ω is real and nonzero, we have $r \in \mathbb{C}$. Then the solution takes the form

$$x(t) = Ae^{(-\gamma/2)t}\cos(\omega t + \phi).$$

   Since the restoring force overpowers the damping force, we have oscillatory motion, bound by an exponential decay envelope. Note that the frequency ω is not the same frequency as natural oscillation.

   We can clearly see now that the damping constant is related to the decay of oscillations. We are often interested in how long it takes for the amplitude to decay by a factor of $1/e$; this is called the amplitude damping time. We can see that the amplitude damping time is $2/\gamma$.

2. OVERDAMPED: When ω is imaginary, $r \in \mathbb{R}$. Then the solution takes the form

$$x(t) = Ae^{r_1 t} + Be^{r_2 t}$$

   with

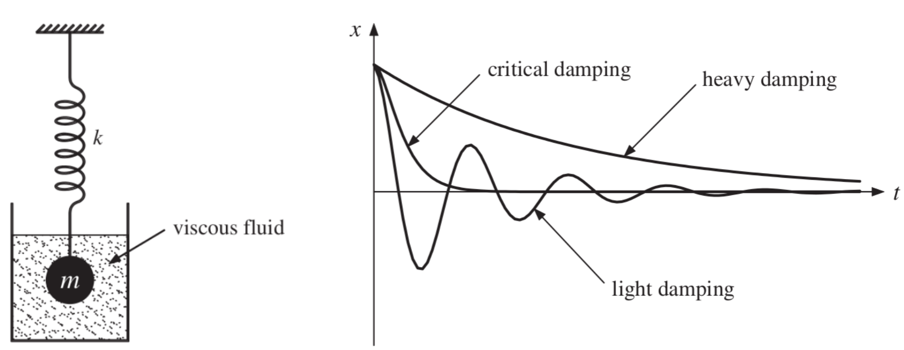
$$r_1 = -\frac{\gamma}{2} + |\omega| \qquad r_2 = -\frac{\gamma}{2} - |\omega|.$$

   There is no oscillatory motion; the damping force overpowers, and the object slowly returns to equilibrium.

3. CRITICALLY DAMPED: When $\omega = 0$, we have a double root for r. So the solution is

$$x(t) = (A + Bt)e^{(-\gamma/2)t}.$$

   The critically damped oscillator represents a balance between the restoring force and the damping force. There are no oscillations, but the object returns to equilibrium quickly.



2.1 Energy and Quality factor

Here we will consider some properties of an underdamped oscillator. First, we'll look at the energy. It can be shown that the energy in an underdamped oscillator is

$$E = E_0 e^{-\gamma t}\left(1 + \frac{\gamma}{2\omega_0}\sin(2(\omega t + \phi))\right).$$

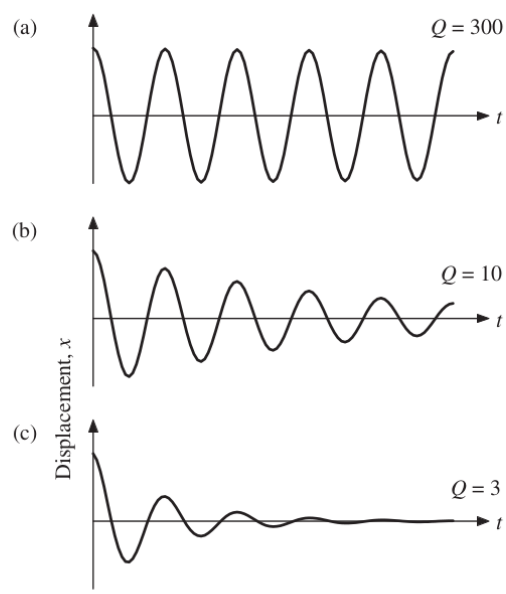
In the weak damping limit, where $\omega_0 \gg \gamma$, the sine term is negligible, and we can approximate the energy as

$$E \approx E_0 e^{-\gamma t}.$$

Notice that the decay factor is the square of the amplitude decay factor. This makes sense since energy is proportional to the square of amplitude. As a result, the energy damping time $\tau = 1/\gamma$, which is half the amplitude damping time.

Let's define the "quality factor" of a damped oscillator to be the number of radians it goes through in one energy damping period. If it oscillates much faster than it decays, it has a high quality factor. Quantitatively,

$$Q = \omega_0/\gamma.$$



The power lost to friction oscillates at twice the rate of the oscillator:

$$P = \mathbf{F_d} \cdot \mathbf{v} = -b\dot{x}^2.$$

3 Driven Oscillator

Now we consider a damped oscillator which is driven by a periodic force. For example, we could have a mass on a spring, where the top of the spring is repeatedly moved up and down, or a magnetic mass on a spring which is driven up and down by an oscillating magnetic field. For simplicity, we will assume that the driving force is sinusoidal.

The equation of motion is

$$m\ddot{x} = F_0 \cos(\omega t) - kx - b\dot{x}$$

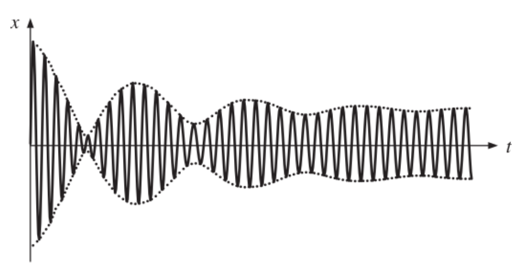
which yields

$$\ddot{x} + \gamma\dot{x} + (\omega_0)^2 x = \frac{F_0}{m}\cos(\omega t)$$

where F_0 is the amplitude of the driving force and ω is the driving frequency. This is a linear non-homogeneous second-order equation, so the solution is the sum of the homogeneous general solution (see damped oscillator) and a particular solution. Since the homogeneous solution decays exponentially, the particular solution ends up being a *steady-state*:

$$x_{steady} = A\cos(\omega t - \delta)$$

where δ is the phase offset between the driving force and the resulting oscillation. Note that the steady-state oscillation is at the driving frequency, not the natural frequency. The initial period during which the decay solution is not negligible is called the *transient*. During the transient, the oscillations due to the driving frequency and the natural frequency can interfere with each other.



In addition, both A and δ are functions of the driving frequency. This is because the additional energy imparted by the driving force is strongly dependent on the relationship between the driving frequency and the natural frequency of oscillation. Since the steady-state is governed by the energy imparted by driving, it makes sense that its parameters are governed by the driving frequency.

3.1 Amplitude and Phase

The amplitude of the steady state oscillation ends up being a function of the driving frequency:

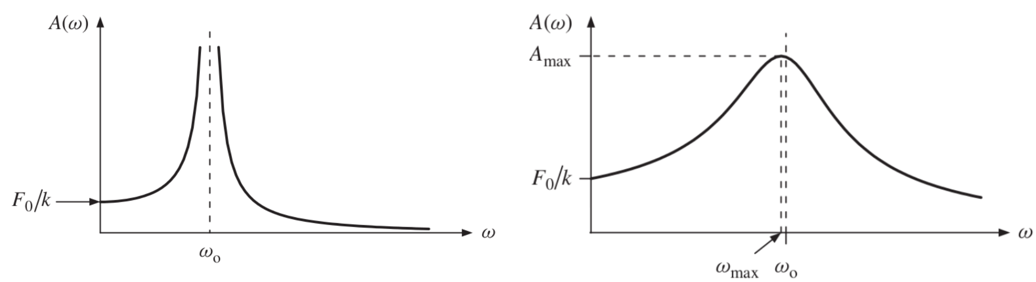
$$A = \frac{F_0}{m}\left((\omega^2 - \omega_0^2)^2 + (\omega\gamma)^2\right)^{-\frac{1}{2}}$$

This relation exhibits *resonance* near the natural frequency. In other words, as the driving frequency approaches the resonant frequency, the amplitude of oscillation increases. This resonance is bound by damping; if we neglect damping, then the amplitude of oscillation increases without bound.

More specifically, when our system is undamped, there is a vertical asymptote at the natural frequency. When we add damping, we lose unbounded behavior, and reach a peak at a frequency

$$\omega_{max} = \omega_0\left(1 - \frac{\gamma^2}{2(\omega_0)^2}\right)^{\frac{1}{2}} = \omega_0\left(1 - \frac{1}{2Q^2}\right)^{\frac{1}{2}}$$

which is negligibly different from the natural frequency at sufficiently high Q.



The phase offset between the driving force and the actual oscillation is also dependent on the driving frequency:

$$\tan\delta = \frac{\omega\gamma}{(\omega_0)^2 - \omega^2}$$

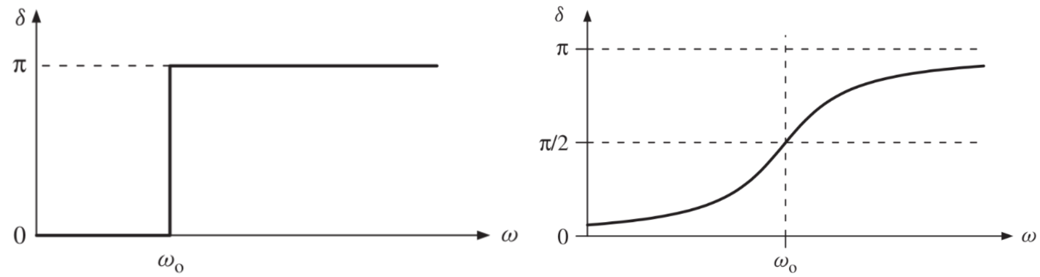
At low frequencies, the oscillating mass reacts relatively quickly to changes in the driving force, so there is very little phase lag.

At the natural frequency, there is a significant, quarter-cycle phase lag. This phase lag is what allows the amplitude to be amplified to its resonant peak – the oscillator is receiving energy from the driving force exactly when it's needed.

As we exceed the natural frequency, the system tends towards being completely out-of-phase. The driving force is always opposing the motion of the oscillator, and the oscillator cannot "catch up" with the high oscillation rate.

This is why the amplitude tends towards zero – the oscillator is not able to absorb energy from the driving force.

The rate of change of phase is related to the amount of damping. At low damping, the phase makes a very fast transition from in-phase to out-of-phase near resonance. At high damping, the phase changes gradually.



3.2 Power

The power absorbed from the driving force is given by

$$P = \mathbf{F_d} \cdot \mathbf{v} = \cdots = F_0 v_0 \left(\sin\delta\cos^2(\omega t) - \cos\delta\cos(\omega t)\sin(\omega t)\right)$$

where $v_0 = \omega A$ is the maximum steady-state velocity of the oscillator. This expression is cumbersome, and often we are only interested in the time-averaged power (i.e. the average power absorbed per cycle). Then we can show

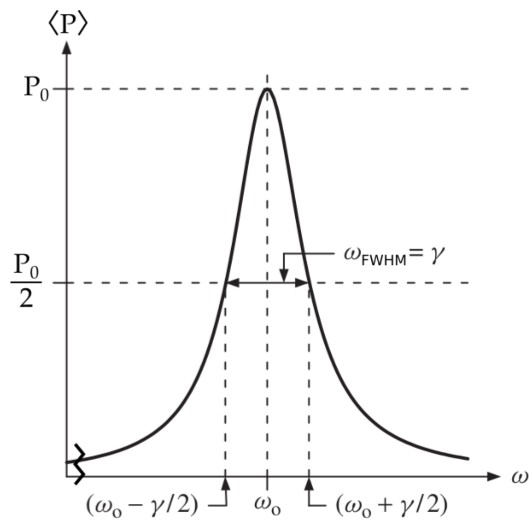
$$\begin{aligned}\langle P\rangle &= F_0 v_0 \left(\sin\delta \cdot \frac{1}{2} - \cos\delta \cdot 0\right) = \frac{1}{2}F_0(\omega A)\sin\delta \\ &= \cdots \\ &= \frac{\gamma F_0^2}{2m} \cdot \frac{\omega^2}{(\omega^2-\omega_0^2)^2 + (\omega\gamma)^2}\end{aligned}$$

We can look at the graph of this curve and notice some important properties. There is a resonant peak near the natural frequency; when the oscillator is being driven at resonance, it is absorbing maximal energy from the driving force. This is why we see an amplitude peak. However, at low and high frequencies, very little power is absorbed. This is why we see much smaller amplitudes.

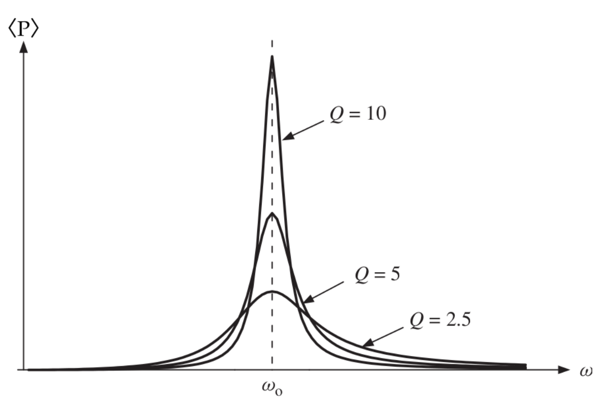
However, this formulation is still pretty cumbersome and resists analysis. We can make a near-resonance approximation using the Lorentzian lineshape. By expanding the difference of squares and replacing all ω with ω_0 except for the $(\omega - \omega_0)$ term, and rearranging, we get the final power equation:

$$\langle P \rangle = \frac{P_0}{1 + \left(\dfrac{\omega - \omega_0}{\gamma/2}\right)^2} \quad \text{where} \quad P_0 = \frac{F_0^2}{2m\gamma}.$$

The Lorentzian lineshape has certain useful properties. The peak power is given by P_0, and the full width of the curve at half the maximum value (FWHM) is equal to γ. These particular properties actually hold true for both the lineshape approximation and the actual power.



However, the approximation as a whole is only valid near resonance; as we move away from resonance, the approximation diverges from the actual time-averaged power at a rate inversely related to Q. In other words, the higher the Q value, the better the approximation.



4 Coupled Oscillators

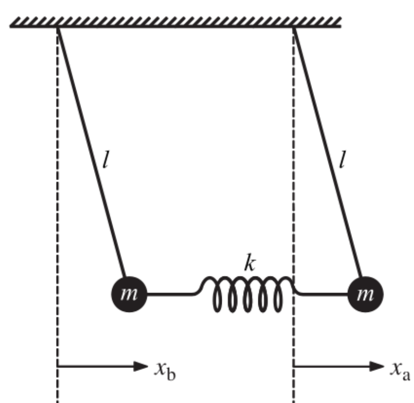
Suppose we have two or more oscillators which are somehow linked or connected. Then, the motion of one oscillator is then tied to the motion of the others. Clearly, the motion of the system because much more complex. However, it can be shown that for many such systems of coupled oscillators, the motion of the system obeys the *superposition principle*, which means that the general solution for the motion of the system can be expressed as a linear combination of a much smaller set of *normal mode* solutions.

A normal mode solution for coupled oscillators has the property that all oscillators oscillate at the same frequency and with constant amplitude. Crucially, however, the normal frequencies for different normal mode solutions are not the same. Since the motion of each bob is a superposition of normal mode solutions, they should take the form

$$\begin{aligned} x_a &= A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2) + \ldots \\ x_b &= A_{n+1} \cos(\omega_1 t + \phi_{n+1}) + A_{n+2} \cos(\omega_2 t + \phi_{n+2}) + \ldots \\ &\vdots \\ x_n &= \ldots \end{aligned}$$

where ω_1 is the first normal mode frequency, ω_2 is the second normal mode frequency, etc. for each of the n normal modes.

4.1 Coupled Pendulum



In the above example of the springed pendulum, you can visualize two normal mode solutions: one where the bobs oscillate together (in-phase) and another where the bobs oscillate in opposite directions (completely out-of-phase). Since the bobs are the same mass, their amplitudes of oscillation must be equal for each normal mode. So, the solutions are restricted to

$$x_a = A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2)$$
$$x_b = A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2 + \pi)$$

where A_1, A_2, ϕ_1, and ϕ_2 are resolved using the initial conditions. So now we only have to find the normal mode frequencies ω_1 and ω_2.

To do this, we can set up an equation of motion that considers both the force of gravity and the force exerted by the spring. Since the spring force is dependent on the separation between the bobs, this component of the force is coupled – it links the motion of the bobs.

$$m_a \frac{d^2 x_a}{dt^2} = -\frac{m_a g x_a}{l} - k(x_a - x_b)$$

By taking advantage of the fact that we know that the oscillations are either in-phase or out-of-phase for the normal mode solutions, we can plug in $x_b = x_a$ and $x_b = -x_a$ to solve for the two normal mode solutions.

The resulting normal mode frequencies are

$$\omega_1 = \sqrt{\frac{g}{l}} \quad \text{and} \quad \omega_2 = \sqrt{\frac{g}{l} + \frac{2k}{m}}$$

which makes sense. For the in-phase oscillations, the spring is never stretched nor compressed, so the oscillation frequency is just that of the pendulum. In other words, there is no real coupling. In the out-of-phase case, the force of gravity compounds with the spring force, so the greater restoring force results in a higher frequency.

In a situation where the oscillators oscillate in a superposition of normal modes, the motion of each oscillator can be expressed as

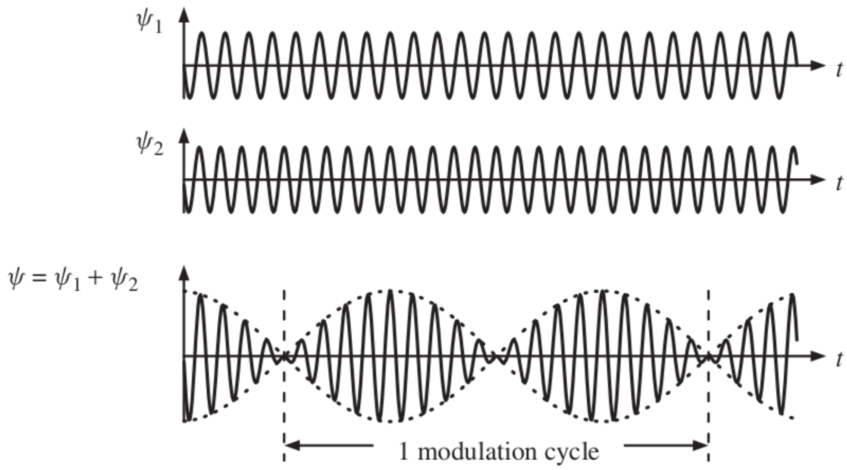
$$x = A\cos(\omega_1 t) + A\cos(\omega_2 t) = 2A\cos\left(\frac{\omega_2 + \omega_1}{2} t\right)\cos\left(\frac{\omega_2 - \omega_1}{2} t\right)$$

where I drop the phase angle for clarity and assume that the amplitude of each mode are equal[1].

Now let's assume that the restoring force of the spring is relatively small compared to the restoring force due to gravity. Then the normal mode oscillations will have similar frequencies, and $\omega_2 - \omega_1$ will be small. In this regime,

[1]Including the phase angle only changes the phase of the beats, and having a different amplitude ratio changes the minimum amplitude of the beating.

the $\cos\left(\frac{\omega_2-\omega_1}{2} \cdot t\right)$ term acts as an envelope for the other higher-frequency term, and we observe a "beats" phenomenon:



4.2 General Case

We made several simplifying assumptions about the coupled pendulum which allowed us to solve for the normal frequencies. In a general system with n oscillators, these assumptions may not hold. We know that a system of n oscillators will have n normal modes, each with its own normal frequency.

In a given normal mode, the oscillators oscillate with different amplitudes, and these precise values are determined by the initial conditions. However, the amplitude ratios between oscillators can be determined by analysis of the system. In other words, for a given normal mode, all amplitudes can be expressed in terms of the amplitude for one of the oscillators.

In addition, for a given normal mode, some oscillators will be oscillating completely in-phase, and the rest are guaranteed to be oscillating completely out-of-phase (relative to the first set). In other words, for a given normal mode, there are only two possible phases, and they are completely out-of-phase. For this reason, we can omit the phase factor and encode the phase factor in the sign of the amplitude (negative amplitude signifies out-of-phase). The actual numerical phases can be added in at the end, and are determined by the initial conditions.

So, the question of solving for the equations of motion is reduced to a question of finding normal frequencies and the corresponding set of amplitude ratios. To do this, we first observe that for normal mode solutions, we expect

$$x_i = A_i \cos(\omega_j t).$$

where i denotes different oscillators and j denotes different normal modes. We also observe that the equation of motion for each oscillator will be a differential equation which is coupled with at least one other oscillator:

$$\begin{aligned} m_a \ddot{x}_a &= f_a(x_a,\ x_b,\ \ldots,\ x_n) \\ m_b \ddot{x}_b &= f_b(x_a,\ x_b,\ \ldots,\ x_n) \\ &\vdots \\ m_n \ddot{x}_n &= f_n(x_a,\ x_b,\ \ldots,\ x_n). \end{aligned}$$

If the coupling relations are linear with constant coefficients, then we can express this system of differential equations in matrix form (after dividing out the masses). But for our normal mode solutions, we can directly differentiate and find an equivalent expression:

$$\ddot{\mathbf{x}} = M\mathbf{x} = \omega^2 \mathbf{x}$$

I can omit the subscript on ω because this system of equations represents the restraints on a single normal mode solution. Finding different ω which satisfy this equation (given a matrix M which codifies the equations of motions for each oscillator) amounts to finding the various normal mode frequencies.

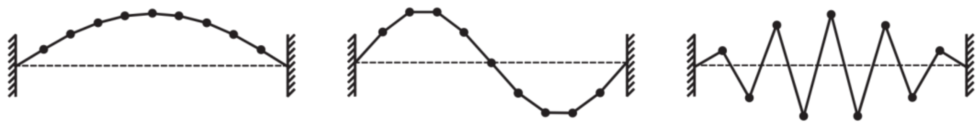
Of course, this is an eigenvalue-eigenvector problem, and can be solved using the canonical techniques. Since the system of equations represents a single normal mode, we can divide the $\cos(\omega t)$ term from all the components of $\mathbf{x}$, leaving a scalar-valued vector of amplitudes:

$$\begin{bmatrix} C_{aa} & C_{ab} & \cdots \\ C_{ba} & C_{bb} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} A_a \\ A_b \\ \vdots \end{bmatrix} = \omega^2 \begin{bmatrix} A_a \\ A_b \\ \vdots \end{bmatrix}$$

Finding eigenvalues is finding (the square of) normal frequencies; the corresponding eigenvector gives the amplitude ratios.

4.3 Transverse Oscillators

Suppose we have a line of n beads of equal mass m equally spaced at length l, connected to each neighbor with identical springs. The end beads are connected to rigid walls. We can describe the oscillations of the system when the beads move in the same axis as the springs (longitudinal); or, we can consider motion perpendicular to the axis (transverse).



We can treat this system of oscillators in the same way as before, and we find that we have n normal modes, which appear to be standing oscillations. As we take the limit as $n \to \infty$, $m \to 0$, and $l \to 0$, we depart from the realm of discrete coupled oscillators and into the realm of continuous waves. The normal modes correspond to the infinite harmonic series for the standing waves of a string bound on both ends.

5 One-Dimensional Travelling Waves

As we take the limit as $n \to \infty$, $m \to 0$, and $l \to 0$ of a series of coupled oscillators in a line[2], we end up with a string exhibiting wave behavior. We've seen that the normal modes of the coupled oscillator correspond to standing waves on a finite, taut string. However, when we take the total length of the string $L \to \infty$, we can consider *travelling waves*.

One dimensional non-dispersive[3] waves satisfy the *wave equation*:

$$\frac{\partial^2 s}{\partial t^2} = v^2 \frac{\partial^2 s}{\partial x^2}$$

where $s(x,t)$ is the displacement of the medium and v is the velocity of the wave through space. Generally, solutions of this equation can be written in the form

$$s(x,t) = f(x - vt) + g(x + vt).$$

where f and g describe the shape of a wave moving with constant velocity in the $+x$ and to the $-x$ directions respectively. We will first discuss the special case where f and g are sinusoidal functions with a well-defined frequency; then we will see that any other choice of f and g can be decomposed into a sum of sinusoidal functions.

On an arbitrarily long string (or any other medium), we can describe a sinusoidal wave travelling in the $+x$ direction as

$$s(x,t) = \sin(k(x - vt)) = \sin(kx - \omega t)$$

[2]where the total mass nm remains constant

[3]Dispersion will be discussed later.

where k is the *wavenumber* and describes the spatial frequency of the wave, and ω is the frequency of oscillation of a single point on the string in time. These relations are given by

$$k\lambda = 2\pi \quad \text{and} \quad \omega T = 2\pi$$

where λ is wavelength and T is period. We also have the crucial relation to the velocity of the wave

$$\omega = kv \quad \text{which yields} \quad v = \lambda f.$$

The quantities k and ω are enough to fully characterize a sinusoidal wave. However, other useful properties can be characterized as well. For example, the velocity v of a mechanical wave depends on both the "tension" and the mass density of the medium. The impedance Z of a medium is a measure of its resistance to propagate energy. In a string under tension T and with linear mass density μ, these are given by

$$v = \sqrt{\frac{T}{\mu}} \quad \text{and} \quad Z = \sqrt{\mu T} = \frac{T}{v}.$$

5.1 String

The derivation of wave velocity on a string is as follows. Each segment experiences a net force which is equal to the sum of the tension forces on each end. Since the magnitude of the tension force is assumed to be uniform throughout the string, the important quantity to consider is the *direction* of the forces. Because we're in the small angle approximation regime (low amplitude relative to wavelength), you can show that the direction of the tension force on one end of the segment is given by the first derivative of the string displacement, dy/dx. Then the change in the direction of the tension force over the segment is related to the second derivative of the string displacement. We assume that the longitudinal force differential is negligble, so the net force is completely in the transverse direction:

$$F_{net} = T\left(\frac{d^2y}{dx^2}\right)\Delta x$$

By equating this with Newton's second law, we can derive the wave equation, proving that it obeys the wave equation, and that the velocity of the wave is indeed $\sqrt{T/\mu}$.

The transverse force that a point on a string experiences (assuming small amplitude wave motion) is

$$F = -T\frac{dy}{dx} = -Z\frac{dy}{dt}.$$

The second equality leads to another definition for impedance: the ratio between force applied and resulting velocity.

The potential energy of a string segment is

$$dU = \frac{1}{2}T\left(\frac{dy}{dx}\right)^2 dx$$

while the kinetic energy is

$$dE_K = \frac{1}{2}\mu\left(\frac{dy}{dt}\right)^2 dx.$$

The power carried by the string is $P = F \cdot (dy/dt)$ or total energy divided by time.

5.2 Sound

A sound wave consists of compressions and rarefactions in a material, where the displacement from equilibrium obeys the wave equation. Since the medium for sound is a compressible fluid, we differentiate between adiabatic compression/rarefaction (which is fast enough for energy to not be dissipated as heat) and isothermal (which is slow enough for all heat to be dissipated, thus maintaining the same temperature everywhere). Sound waves, in the range of human hearing, occur fast enough to qualify as adiabatic.

This has implications on the "tension" of the medium – in the case of sound, it's called the *bulk modulus*. The bulk modulus is a property of the medium[4], but its numerical value changes depending on whether the vibration is adiabatic or isothermal. The adiabatic bulk modulus is given by $B = \gamma P_0$ where γ is the adiabatic index (for air, $\gamma = 1.41$) and P_0 is the equilibrium pressure of the fluid.

The velocity of a wave and impedance of the medium is given by

$$v = \sqrt{\frac{B}{\rho}} \quad \text{and} \quad Z = \sqrt{\rho B} = \frac{B}{v}$$

[4]The bulk modulus is defined as the ratio between the change in pressure to the fractional change in volume of the fluid.

where ρ is the density of the medium. Notice the parallels with the case of the string. The speed of sound in air comes out to be about $343\ \mathrm{m\,s^{-1}}$, although this value fluctuates with temperature and humidity. The force that a sound wave applies to a normal surface is given by $F = pA$ where A is the normal area of the sound wave striking the surface, and p is the change in pressure from the equilibrium pressure, given by

$$p(\mathbf{r}, t) = -B(\nabla \cdot s(\mathbf{r}, t))$$

for a 3-dimensional wave. Then the power is

$$P = F \cdot \frac{\partial s}{\partial t} = -B(\nabla \cdot s(\mathbf{r}, t))A \cdot \frac{\partial s}{\partial t}.$$

But often we're more interested in power per unit area, or *intensity*. Moreover, since power and intensity fluctuate with time, we're also more interested in the time-averaged intensity. For a sinusoidal wave:

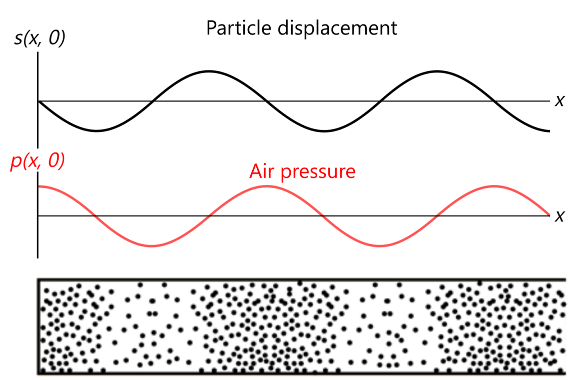
$$\bar{I} = \frac{1}{2} B\omega k s_{max}^2.$$

As always, the time-averaged power and intensity are proportional to the square of the amplitude. Note that for a 3-dimensional spherical wave emanating from a source, the intensity of the wave decreases with distance as an inverse square law.

Although in the general case, sound is a 3-dimensional wave, a sound wave travelling through a pipe (with diameter much smaller than the wavelength) can be safely approximated as a 1-dimensional travelling wave. In this case, the pressure is given by

$$p(x, t) = -B\frac{\partial s}{\partial x}.$$

All the other previous results still apply in the one-dimensional case.



5.3 Coaxial Cable

A coaxial cable carrying an alternating current can be viewed as a medium carrying a travelling wave. Since a coaxial cable can be approximated as a series of LCR circuits, we can derive the following relation between the potential difference V between the two conductors, the inductance L per unit length, and the capacitance C per unit length:

$$\frac{\partial^2 V}{\partial t^2} = \frac{1}{LC}\frac{\partial^2 V}{\partial x^2}.$$

This is another form of the non-dispersive wave equation, where

$$v = \sqrt{\frac{1}{LC}} \quad \text{and} \quad Z = \sqrt{\frac{L}{C}}.$$

In the context of alternating currents, the impedance Z is considered a complex number, where the imaginary part is the *reactance* and the real part is the *resistance*. So for a wire with no reactance, the impedance is the same as the resistance. The power transmitted on a coaxial transmission cable is $P = ZI^2$ where I is the current on the wire.

6 Boundary Conditions and Standing Waves

6.1 Impedance discontinuities

The wave equation describes energy propagation in a medium that is infinite in length. What happens if the medium is terminated somehow? Similarly, what happens if the medium changes? At such discontinuities, the wave is partially transmitted and partially reflected.

We can characterize all discontinuities in the medium as impedance discontinuities. The impedance of a taut string is given by $Z = \sqrt{\mu T}$, so if we attach this string to another string with a different impedance, we can expect to have any wave be partially reflected and partially transmitted. As an extreme case, if the string is attached to a wall, we can consider the wall to be another "string" with infinite impedance.

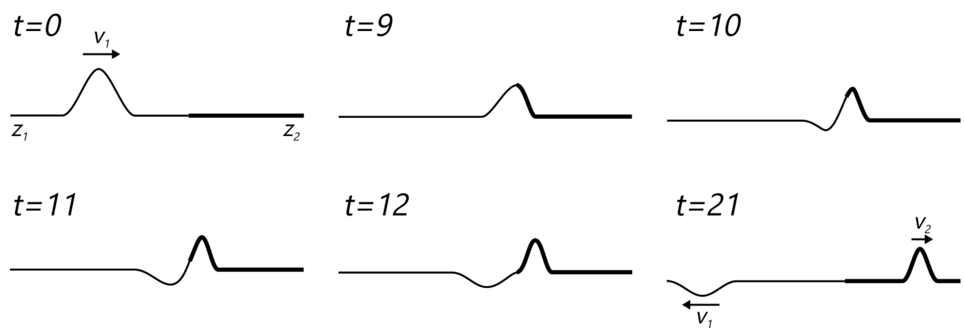
Suppose an incoming wave has amplitude A and is approaching an impedance discontinuity. The amplitude of the reflected wave, in terms of the incident wave, is the reflection coefficient; an analogous definition applies for the transmission coefficient. If the current medium has impedance Z_1 and the adjacent medium has an impedance Z_2, then the reflection and transmission coefficients for amplitude are[5]

$$R_A = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad \text{and} \quad T_A = 1 + R_A = \frac{2Z_1}{Z_1 + Z_2}$$

and the reflected/transmitted coefficients for power is given by

$$R_P = (R_A)^2 \quad \text{and} \quad T_P = \frac{Z_2}{Z_1}(T_A)^2.$$

The picture below shows frames from an animation of a wave pulse propagating from a string of low impedance to high impedence.



Note that the reflected wave is "inverted" if the impedance discontinuity is from low to high. Also, the sum of the reflected and transmitted amplitude coefficients does not always equal 1, but the sum of reflection and transmission power coefficients does (due to conservation of energy). Finally, we can again see that power is proportional to the square of the amplitude.

We can see that in the case of a "closed" termination[6] (i.e. a string clamped to a wall, a sound wave hitting a rigid barrier, or a coaxial cable terminated by a short circuit[7]), we have $Z_2 = \infty$, so the wave is fully reflected (with an inversion), and no energy is transmitted through the termination. On the other hand, a $Z_2 = 0$ termination (i.e. a string with a free end, a sound wave

[5]These formulae are derived from the fact that a) the displacement must be continuous across the impedance discontinuity, and b) the transverse force must be continuous, which implies that ds/dx must be continuous.

[6]These are crude examples because we aren't keeping the "tension" term constant between the two media, which we're supposed to. In these cases, it doesn't really matter.

[7]It turns out that the equations for the reflection and transmission amplitude for coaxial cables apply to the current, not the voltage. For voltage, the equations are modified by switching each Z_1 with Z_2 and vice versa. So a short circuit actually corresponds to a "closed" termination.

hitting a vacuum, or a coaxial cable terminated by an open circuit) reflects a wave with the same amplitude as the incident wave. However, there is actually no transmitted wave, because $Z_2 = 0$ implies that $T_P = 0$.

Notice that waves with terminations impose a boundary condition on the waves. In particular, a closed termination has the condition that displacement $s(t) = 0$ for all t at that point, while an open termination has the condition that $ds/dt = 0$ for all t at that point. These observations are crucial for understanding one-dimensional waves with two boundary conditions.

6.2 Standing Waves

Let's suppose we have a medium with two boundaries – for example, a taut string which is clamped on both ends (this is very similar to the many-coupled-oscillators system we considered earlier). If we send a pulse down the string, it will reflect off each end and continue travelling up and down the string. However, we can also excite normal mode oscillations, just like we did with the coupled oscillators. Again, each point in the string oscillates with the same frequency and with the same phase – this is a very natural continuation of the finite-element limit of the coupled oscillators into the continuous regime.

But now we can look at this phenomenon in a new light: they are *standing waves* which are a superposition of identical waves propagating in opposite directions. In fact, we can view this as a single wave, reflecting on the boundaries and interfering with itself on the return journey. This continuous interference of the opposite-travelling waves is what causes standing waves.

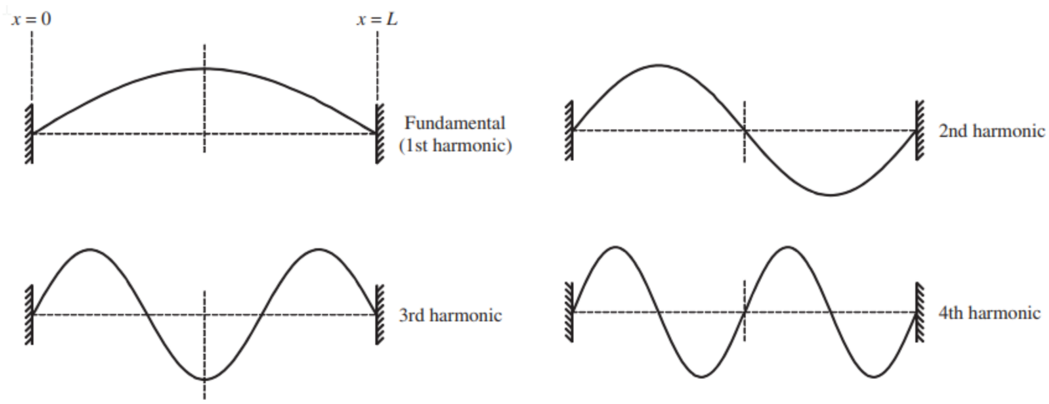
If we're looking for sinusoidal solutions, only certain wavelengths can satisfy the boundary conditions. If the string has length L and both boundaries are closed, then $s(0, t) = s(L, t) = 0$, so the normal mode oscillations are

$$s(x, t) = \sin\left(\frac{2\pi x}{\lambda}\right)\cos(\omega t) \quad \text{where} \quad n\lambda = 2L \quad \text{for} \quad n \in 1, 2, 3 \ldots$$

and $\omega = 2\pi v/\lambda$. Similar formulae can be derived for the case of open terminations; the main difference is that the displacement at the boundaries will be an antinode rather than a node. These cases correspond to strings with clamped/free ends; sound in a pipe with closed and open ends; and coaxial cables with short-circuit/open-circuit terminations.

The largest λ which satisfies the boundary conditions corresponds to the lowest frequency f. This is known as the $fundamental frequency$. All normal modes of a one-dimensional wave have frequencies which are integer multiples of the fundamental frequency, and for this reason, they are called *harmonics*.

When we look at higher-dimensional regimes, this property does not hold; normal mode frequencies are not necessarily multiples of the lowest frequency. This translates to the idea that plucking a string or exciting waves in a pipe can have a very harmonic sound, whereas striking a bell can sound much more dissonant.



Of course, by the superposition principle, any linear combination of such sinusoidal normal modes is a solution to the wave equation:

$$s(x,t) = \sum_{n=1}^{\infty} A_n \sin\Big(\frac{n\pi x}{L}\Big)\cos(\omega_n t)$$

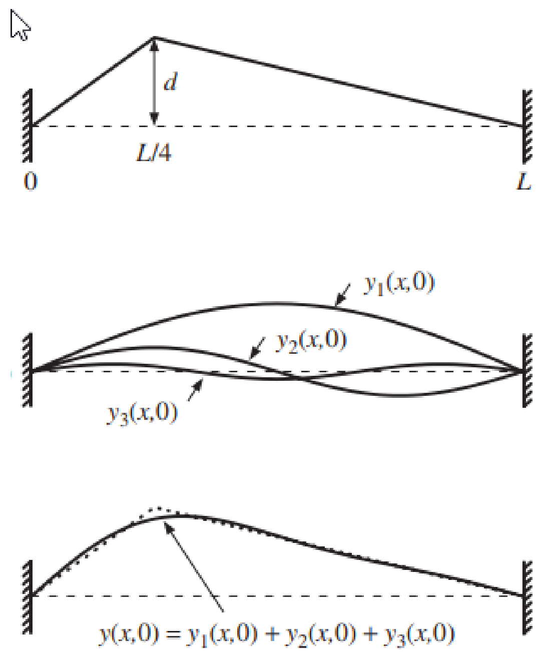
Fourier analysis shows that any function $s(x)$ satisfying $s(0,t) = s(L,t) = 0$ can be resolved as a sum of sinusoidal normal modes; the sum of these normal mode solutions then gives the future evolution of any disturbance on a medium with boundary conditions. To find the coefficients A_n, we can use the formula for the Fourier sine series:

$$A_n = \frac{2}{L}\int_0^L \sin\Big(\frac{n\pi x}{L}\Big) s(x,0)dx$$

In a system which doesn't lose energy to its surroundings, the harmonic content (i.e. the values of each A_n) doesn't change over time. In real physical systems, such as a plucked string, energy is lost as heat and sound; the high harmonics are typically lost before the low harmonics, which is what gives the plucking sound.

7 Fourier Analysis

On a string of fixed length, clamped on both ends, we've seen that there are certain discrete frequencies that characterize the wave. Fourier showed that *any* function $f(x)$ on the interval 0 to L can be decomposed into an infinite sum of sines, as shown below: the first three harmonics already give a decent approximation to the triangular wave.



It turns out that this just a specific case of a more general idea.

Let's suppose that $f(x)$ is a periodic function of wavelength $\lambda = 2L$. Fourier's general theorem says that any such function can be exactly represented as a series of sines and cosines with discrete-valued frequencies:

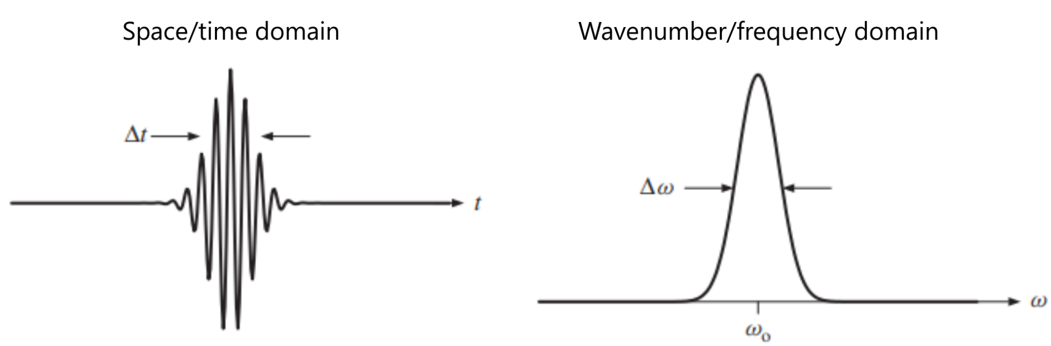
$$f(x) = \sum_{n=0}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right).$$

You can show that if $f(x)$ happens to be an odd function, then all the cosine coefficients vanish; similarly, if $f(x)$ happens to be an even function, then all the sine coefficients vanish.

This applies to any periodic function with $\lambda = 2L$, which is clearly a very general result. However, let's impose the additional restriction that $f(0) = f(L) = 0$. Now, if we look at the interval $0 \leq x \leq L$, it suddenly looks very similar to the clamped string! In fact, earlier when we were trying to find the normal mode frequencies in an arbitrary clamped-string vibration, we really just extended our function $f(x)$ from $[0,\ L]$ to $[-L,\ L]$ by making the arbitrary decision that our extended waveform would be an odd function; then we found the Fourier expansion of our waveform. Of course, since we made our function odd, all the cosine terms dropped out, leaving a sine series as shown above. However, we could have just as easily made our function even and gotten a cosine series, or we could've made our function extend to $[-L,\ L]$ in any other arbitrary way and gotten a Fourier series which included both sines and cosines. We just like to use the odd extension and get a sine expansion, for the sake of convenience.

Regardless of whether or not we choose to use sines or cosines or both, the important quantities don't change: the relative amplitudes of each component frequency. Any clamped-string vibration (in fact, any vibration at all) can be characterized by information in the frequency domain (rather than the spatial domain). We have seen how to convert spatial information to frequency information in the context of normal mode superpositions in a clamped string; what about other signals? Is it possible to analyze non-periodic signals in the frequency domain?

It turns out, the answer is yes. From a mathematical viewpoint, we start by imagining a series of periodic wave pulses (or wavepackets). Now, we can imagine increasing the separation between consecutive wavepackets; in other words, we increase the value of L while keeping $f(x)$ the same in the interval. As the separation goes to infinity, we're left with a single wavepacket. What happens to the Fourier series? As L increases, the discrete "allowed" values for the frequencies start to approach each other; as L goes to infinity, the frequency distribution becomes a continuous function. The end result is that a single wavepacket is comprised or a continuous range of frequencies:



What we've shown is that we can now consider *any* function $f(x)$ – not necessarily periodic – and express it in terms of its component frequencies. The relationship between the spatial representation $f(x)$ and the frequency (technically wavenumber) representation $\hat{f}(k)$ is given by the *Fourier transform*:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(k) e^{-ikx}\, dk$$

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{ikx}\, dx$$

If we're using Fourier analysis to describe a wavepacket, we're limited by a fundamental lack of precision. If a pulse is short, then we can describe its location in space accurately, but its frequency resolution will not be as precise. Conversely, if a pulse is many wavelengths long, then we can more accurately determine its frequency content, but its location in space is less well defined. Of course, it is possible to be imprecise in both, but it's impossible to be precise in both. This is a general property of Fourier pairs, so this applies not only to transforms between $f(x)$ and $\hat{f}(k)$ but also between $f(t)$ and $\hat{f}(\omega)$. A pulse which minimizes the imprecision is said to be *transform-limited* – an example would be the Gaussian wavepacket we've seen.

8 3-Dimensional Waves and Dispersive Waves

The non-dispersive wave equation for 3-dimensional waves is given by

$$\frac{\partial^2 \Psi}{\partial t^2} = v^2 \nabla^2 \Psi$$

where $\nabla^2 \Psi$ is the Laplacian of the wave displacement. As we increase dimension, we change our position x into a position vector $\mathbf{r}$ and our wavenumber k into a wavevector $\mathbf{k}$, which has magnitude equal to the spatial frequency and points in the direction of wave propagation.

Two important solutions are the plane wave and the spherical wave. The plane wave is a straightforward generalization of the one-dimensional wave; in fact, the sound wave in a pipe was really a plane wave in a pipe. The equation for a sinusoidal plane wave is

$$\Psi = A \cos(\mathbf{k} \cdot \mathbf{r} - \omega t)$$

The spherical wave is a wave emanating from a source; by conservation of energy, the amplitude of oscillation must decrease with distance from the source. The equation for a sinusoidal spherical wave centered at the origin is

$$\Psi = \frac{A}{r}\cos(kr - \omega t)$$

where $k = |\mathbf{k}|$ and $r = |\mathbf{r}|$.

8.1 Dispersion

So far, the waves we've considered have obeyed the linear relation $\omega = vk$. This relation is valid in many media, but there are some situations in which the relation between ω and k may not be linear. Examples of dispersive media include light travelling in glass, surface waves on deep water, and waves in plasma. In this case, the velocity of a wave depends on its frequency; we call this velocity the *phase velocity* because it tracks the velocity of a crest (or any other phase) of the wave:

$$v_{ph} = \frac{\omega(k)}{k}.$$

But what if we have a wavepacket containing multiple frequencies? As we'd expect, each frequency in the wavepacket travels at a different velocity, which gives the impression of the wavepacket "spreading out" as it travels. Each crest in the wavepacket travels at a phase velocity, but the wavepacket as a whole travels at a different velocity – the *group velocity*:

$$v_{gr} = \frac{d\omega}{dk}$$

where the derivative is usually evaluated at the mean wavenumber. These equations are actually valid for non-dispersive waves as well; it just happens that since the dispersion relation is linear, $v_{ph} = v_{gr} = \omega/k$.

9 Optics

The electromagnetic wave, unlike many of the waves we've discussed, is not mechanical. It travels through the electromagnetic field which permeates space, and interacts with matter (specifically, electrons). It is these interactions with matter which cause the interesting properties which we study in optics. In some cases, light is absorbed by the electrons in materials, and in other cases, it is reflected. However, transparent materials (sometimes called dielectrics) are often more interesting and important to optics.

9.1 Refraction

To a first approximation, electrons bound to an atom behave as a simple harmonic oscillator. In transparent materials, the natural frequency of the oscillator is much higher than that of visible light. Oscillations in the electromagnetic field (i.e. an incident plane wave of light) act as a driving force for these electrons. Since they are being driven much under resonance, they oscillate in phase, and do not absorb much energy from the light. Importantly, oscillating electrons generate an spherical electromagnetic wave, and since light obeys the superposition principle, the resultant light wave is the linear combination of the incident light and the generated wave.

Specifically, the resultant wave is a plane wave everywhere in space; its wavelength outside the material is unchanged, but its wavelength within the material is different. This change of wavelength is refraction, and the amount of change is measured by the index of refraction n:

$$n = \frac{\lambda_0}{\lambda}$$

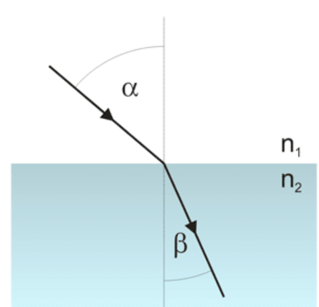
where λ_0 is the vacuum wavelength (approximately the same as the wavelength in air) and λ is the wavelength within the dielectric. The change in wavelength is accompanied by a change in propagation velocity, such that the frequency of the plane wave doesn't change:

$$\frac{c}{\lambda_0} = \frac{v}{\lambda} = \frac{\omega}{2\pi} \quad \text{and} \quad n = \frac{c}{v}$$

where c is the vacuum speed of light and v is the speed of light in the dielectric. Because of this change in velocity, light also bends when it transitions into a dielectric. The angle of refraction θ_r is given by *Snell's law*:

$$n_1 \sin\theta_i = n_2 \sin\theta_r$$

where n_1 and n_2 are the refractive indices of the incident medium and the dielectric respectively, and θ_i is the angle of incidence. Some of the power is also reflected; the angle of reflection is always equal to the angle of incidence.



If the light wave is transitioning from a medium of higher refractive incidence, then there will be a *critical angle* for the angle of incidence at which Snell's law cannot be satisfied. At this angle and higher, the light cannot be refracted, and all of the light is reflected. This is known as *total internal reflection.*

Light is also slightly dispersive in dielectrics, which means that the index of refraction depends on wavelength. Generally speaking, n increases with frequency, which means that bluer light slows down more, and therefore bent more compared to red light. This variation of refractive angle is what allows prisms to separate white light into a spectrum of colors.

A discontinuity in refractive index is analogous to an impedance discontinuity, and one can calculate the reflection and transmission coefficients as before.

9.2 Huygens-Fresnel Principle

Classically, diffraction is best understood as a consequence of the *Huygens-Fresnel principle.* In the late 1600s, Huygens imagined that as light propagates, each new point it reaches becomes a source for a spherical wave. These new point sources help propagate the wave further. When a plane wave propagates, the infinitely many new point sources mostly interfere with each other and cancel, except for the propagated plane wave. However, Newton advocated a particle theory of light, and consequently wave interpretations were largely ignored.

In the early 1800s, Fresnel added to Huygen's principle by introducing an obliquity factor, which requires that the intensity of the spherical wave reduces as you move away from the propagation direction. This wave theory was beginning to gain traction because it made good predictions about diffraction effects. Poisson criticized the wave theory because it made seemingly absurd claims: for example, a beam of light obstructed by a circular disk was predicted to have a bright spot at the center of the shadow. However, Arago then showed that this actually happens; the bright spot is now known as the Arago spot (or the Poisson spot).

The Huygens-Fresnel principle is no longer considered a valid model of reality; however, it does make good predictions and allows for easy visualization of classical optics.

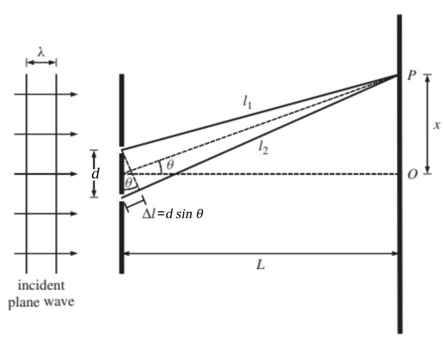
9.3 Diffraction

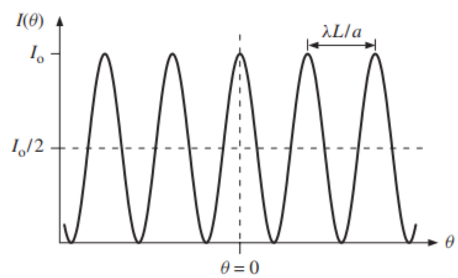
Diffraction is the apparent bending of light around a corner or obstacle. This can be likened to a two people standing outside the corner of a building, one on each side, unable to see each other around the corner. However, if one person speaks, the other will easily be able to hear. Barring the effects of reflection from the ground, this should not be possible, unless the sound wave is somehow bending around the corner. Huygens' idea for waves explains this: when the wave reaches the corner, each point acts as a source for a spherical wave, and so the wave can reach the second person. A similar principle applies to light (although to a much less extreme degree).

If a monochromatic plane wave of light is incident on two slits separated by d, where each slit has an infinitesimally small width, the slits act as point sources for a spherical wave. The two waves will then interfere with each other, with points of constructive and destructive interference extending radially. In other words, interference peaks and troughs occur at specific angles respective to angle of incidence. For the double slit, the time-averaged intensity I at a screen sufficiently far from the slits is given by

$$I = 4I_0 \cos^2\left(\frac{1}{2}kd\sin\theta\right)$$

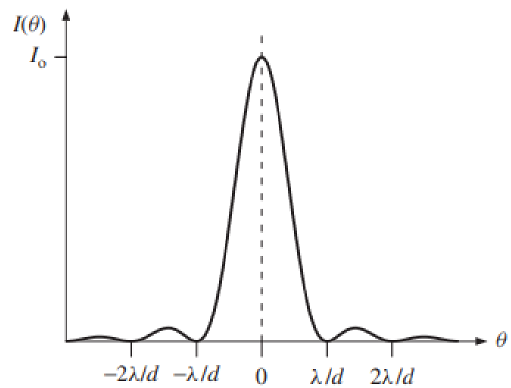
where I_0 is the intensity of the beam without interference. The factor of four comes because the constructive interference will double the amplitude, and the intensity is proportional to the square of the amplitude. The term $kd\sin\theta$ represents to the phase difference between the two spherical waves at angle θ.



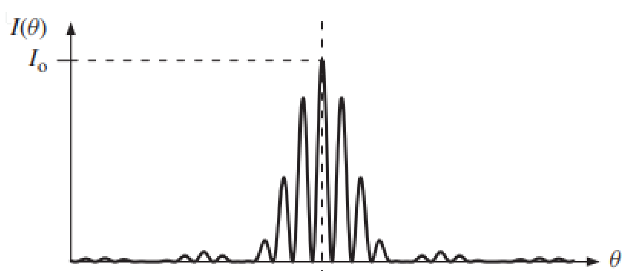


Various other configurations exist as well. For a single slit whose width a is on the order of the wavelength, we can consider each point in the slit to be a source of a spherical wave. Then each source will interfere with the others, causing an interference pattern. The intensity on a screen sufficiently far from the slits is

$$I = I_0 \frac{\sin^2\left(\frac{1}{2}ka\sin\theta\right)}{\left(\frac{1}{2}ka\sin\theta\right)^2}.$$



In the more realistic case of the double slit where each slit has a finite width, the interference pattern will be the normal double slit pattern, modulated by the shape of the single slit pattern.



For multiple evenly-spaced small slits, there is much more controlled interference. As a result, the peaks become narrow (inversely proportional to the number of slits N) and tall (proportional to N^2). There are secondary peaks between the primary peaks, and these are much smaller in intensity. The interference pattern is given by

$$I = \frac{1}{2} I_0 \frac{\sin^2\left(\frac{1}{2}Nkd\sin\theta\right)}{\sin^2\left(\frac{1}{2}kd\sin\theta\right)}$$

and is modulated by the single slit pattern if the slit width is large. The advantage of multi-slit diffraction is that as N increases, so does the precision of the peaks.

Because the location of the peaks is dependent on k, shining white light through a diffraction grating will actually separate the colors – not through bending the light rays as in refraction, but by interference. When we have many slits, the narrow precision of the peaks results in a cleanly-separated spectrum. A device which contains many diffraction slits (on the order of thousands per millimeter) is called a diffraction grating. These gratings are the crucial component of spectrographs.

Other optical phenomena which won't be discussed are: lenses, virtual and real images, aberrations, polarization, Malus' law, coherence (temporal and spatial), Babinet's principle, Michelson interferometers, the Airy disk, diffraction by a circular aperture, thin film diffraction, Fresnel vs. Fraunhofer diffraction, and the Rayleigh criterion for diffraction-limited instruments.