Tomi Johnson

Keble College - Hilary 2015

CP3&4: Mathematical methods I&II

Notes: Waves and normal modes

While the intuitive notion of a wave is quite straight forward, the topic of waves quickly becomes complicated by the numerous examples of physical systems supporting waves and the many ways in which the waves supported by those systems differ from the ideal. Here we develop the concept of a wave as far as we can before discussing any details of the actual physical systems that realise wave behaviour. The aim is for a more unified mathematical description, to complement a more system-by-system description, as is usually provided.

What is a wave? Consider a system with some quantity $f(\mathbf{r}, t)$, e.g. pressure or displacement, that is a function of position $\mathbf{r}$ and time t. A wave is a special type of behaviour of this quantity $f(\mathbf{r}, t)$.

Complementing this, waves must also be the solution to certain equations of motion, which are in general some partial differential equations of the form $\mathcal{D}f(\mathbf{r}, t) = 0$, where $\mathcal{D}$ is a, usually linear, partial differential operator. Therefore the mathematical aspects of the topic of waves is the natural extension what you learnt about (mostly linear) ordinary differential equations with oscillating solutions, to (mostly linear) partial differential equations with oscillating solutions.

We'll therefore come across many familiar results. We'll find that solutions $f(\mathbf{r}, t)$ to a linear equation $\mathcal{D}f(\mathbf{r}, t) = 0$ supporting waves will consist of a linear superpositions of solution modes $f_{\mathbf{k}}(\mathbf{r}, t)$, each of which is labelled by some $\mathbf{k}$ and oscillates at a single frequencie $\omega_{\mathbf{k}}$. This is the general solution to $\mathcal{D}f(\mathbf{r}, t) = 0$, i.e. the complementary function. External forces appear as $\mathcal{D}f(\mathbf{r}, t) = d(\mathbf{r}, t)$ and the solution to such an equation is the complementary function plus a particular integral, which is any single solution to this inhomogeneous equation.

Normal modes are merely a special case of wave behaviour where instead of some $f(\mathbf{r}, t)$ that is a continuous function of $\mathbf{r}$ we have some $f_i(t)$ that is a function of a discrete parameter i. Accordingly, the equation of motion is the discrete analogy of a partial differential equation. But exactly the same thinking can be applied to these cases.

The plan is to start off with the most ideal type of wave behaviour and wave equation and then slowly add complexity (finiteness, inhomogeneity, dissipation and dispersion). Hence, the first systems we'll look at are called ideal waves.

Part I
Ideal systems

1 Defining ideal waves as solutions to the ideal wave equation

Let us start by deducing the mathematical form of what we would consider an ideal wave.

1.1 Travelling wavepackets

We intuitively think of a(n undamped undriven undispersed) travelling wavepacket (wave) that is travelling at a fixed speed c as some quantity $f(\mathbf{r}, t')$ that at a later time $t' > t$ is merely $f(\mathbf{r}, t)$ linearly translated in space by some distance $c(t' - t)$. Let's say the direction of travel is along some vector $\hat{\mathbf{k}}$. In order to conform with our idea of wave, the value of the wave can only depend on the component $\hat{\mathbf{k}} \cdot \mathbf{r}$ of position $\mathbf{r}$ along the direction of travel. To be travelling, this component $\hat{\mathbf{k}} \cdot \mathbf{r}$ must enter into the form of $f(\mathbf{r}, t)$ alongside the negative of ct. Otherwise the above simultaneous translations in space and time would not leave the quantity $f(\mathbf{r},t)$ invariant, as desired. This gives us the form of an ideal wavepacket

$$f_{\hat{\mathbf{k}}}(\mathbf{r},t) = |A_{\hat{\mathbf{k}}}|F(\hat{\mathbf{k}}\cdot\mathbf{r} - ct). \tag{1}$$

In one dimension such an ideal travelling wave in the $\pm x$-direction has the form

$$f_{\pm}(x,t) = |A_{\pm}|F(\pm x - ct). \tag{2}$$

1.2 Sinusoidal travelling waves

Next we want to consider a special type of travelling wave, the sinusoidal travelling wave.

For simplicity, for the next few paragraphs, choose the coordinate system such that $\hat{\mathbf{k}}$ is in the positive x-direction, and so $f(\mathbf{r},t) = f_+(x,t) = |A_+|F(x-ct)$ for the travelling wave.

One example of a travelling wave is a sinusoidal wave $f_{+|k|}(x,t) = |A_{+|k|}|F_{|k|}(x-ct)$, where

$$F_{|k|}(x-ct) = \cos(|k|[x-ct] + \phi_k). \tag{3}$$

Here we have introduced a positive wavevector $|k| > 0$ associated with the periodicity of the wave.

Fourier found that you can break down any continuous function of x into a superposition of periodic functions. For us that means that any travelling wave is actually just the superposition of many sinusoidal travelling waves. So this means an arbitrary wavepacket $f_+(x,t)$ is an integral over all $f_{+|k|}(x,t)$ for different values $|k|$, given by

$$f_+(x,t) = \int_0^\infty \mathrm{d}|k| f_{|k|}(x,t) = \int_0^\infty \mathrm{d}k |A_{+|k|}|\cos(|k|[x-ct] + \phi_{+|k|}). \tag{4}$$

All ideal plane waves in the positive x-direction can be considered as a superposition of ideal sinusoidal plane waves in the positive x-direction (i.e. the direction of $\hat{\mathbf{k}}$). It becomes convenient to write the sinusoidal plane wave equivalently as

$$f_{+|k|}(x,t) = |A_{+|k|}|\cos\left(|k|[x-ct] + \phi_{+|k|}\right) = |A_{+|k|}|\cos\left(|k|x - \omega_k t + \phi_{+|k|}\right), \tag{5}$$

where we have introduced $\omega_{|k|} = c|k|$.

Then going back to the general case, where $\hat{\mathbf{k}}$ can be in any direction, we arrive at the general expression by replacing $|k|x$ by $\mathbf{k}\cdot\mathbf{r}$ to get

$$f_{\hat{\mathbf{k}}}(\mathbf{r},t) = \int_0^\infty \mathrm{d}|k| f_{\mathbf{k}}(\mathbf{r},t) = \int_0^\infty \mathrm{d}|k||A_{\mathbf{k}}|\cos\left(\mathbf{k}\cdot\mathbf{r} - \omega_{|k|}t + \phi_{\mathbf{k}}\right), \tag{6}$$

where $f_{\hat{\mathbf{k}}}(r,t)$ is any plane wave in the $\hat{\mathbf{k}}$ direction. Sinusoidal waves can then rightly be considered a fundamental type of wave and it is worth specifically introducing names for some of their properties.

- Amplitude: $|A_{\mathbf{k}}|$
- Wavevector: $\mathbf{k}$
- Wavenumber: $\nu_{|k|} = |k|/2\pi$
- Wavelength: $\lambda_{|k|} = 2\pi/|k|$
- Speed: c
- Frequency: $f_{|k|} = c|k|/2\pi$
- Angular frequency[1]: $\omega_{|k|} = c|k|$
- Phase: $\phi_{\mathbf{k}}$

You should sketch a one-dimensional sinusoidal wave $f_{+|k|}(x,t)$ at two different times and make sure you know how the above quantities appear in the sketch. In 2D or higher dimensions, these sinusoidal waves, waves that are formed of a single frequency and take values that are uniform over planes perpendicular to the direction of travel, are called plane waves.

[1]Sometimes just called frequency, sorry.

1.3 Travelling sinusoidal (plane) wave superposition

We haven't allowed yet for the possibility that the quantity $f(\mathbf{r}, t)$ is describing two or multiple wave packets. Intuitively we associate wave motion with linearity. When two waves meet they usually pass through each other and carry on unperturbed. What this is telling us is that the multiple waves are really linear superpositions of two waves.

And so our general type of ideal travelling wave we will consider is

$$f(\mathbf{r}, t) = \int \mathrm{d}\hat{\mathbf{k}} f_{\hat{\mathbf{k}}}(\mathbf{r}, t) = \int \mathrm{d}\mathbf{k} f_{\mathbf{k}}(\mathbf{r}, t) = \int \mathrm{d}\mathbf{k} |A_{\mathbf{k}}| \cos{(\mathbf{k} \cdot \mathbf{r} - \omega_{|k|} t + \phi_{\mathbf{k}})}, \tag{7}$$

which is again just a superposition of sinusoidal plane waves, where this time we have allowed the superposition to include $\mathbf{k}$ that are not in the same direction.

1.4 Wave equation

We have the mathematical form for our ideal wave behaviour: a superposition $f(\mathbf{r}, t)$ of some travelling waves $f_{\hat{\mathbf{k}}}(\mathbf{r}, t) = |A_{\hat{\mathbf{k}}}| F(\hat{\mathbf{k}} \cdot \mathbf{r} - ct)$, each composed of many sinusoidal travelling plane waves $f_{\mathbf{k}}(\mathbf{r}, t) = |A_{\mathbf{k}}| \cos{(\mathbf{k} \cdot \mathbf{r} - \omega_{|k|} t + \phi_{\mathbf{k}})}$.

So what is a system that supports ideal wave behaviour? The system will be defined by some partial differential equation of motion $\mathcal{D}(\mathbf{r}, t) f(\mathbf{r}, t) = 0$ and it must be that this equation has solutions $f(\mathbf{r}, t)$ of the form we found above.

How do we go about deducing this equation? To start, we see that the superposition of any two solutions of the form above is also a solution of this form, so we expect there to be a linear equation with such waves as its solution. What form will this equation take? Well, we know the defining property of the wave solutions is the relationship between the $\mathbf{r}$ and t dependences, since they appear together. Because of the isometry with respect to direction, the relationship must be between the second derivatives. As such the simplest[2] possible equation with any sinusoidal wave as a solution is

$$\left(c^2 \nabla^2 - \frac{\partial^2}{\partial t^2} \right) f(\mathbf{r}, t) = 0, \tag{8}$$

which, for understandable reasons, is called the wave equation.

It is possible to show by substitution that sinusoidal waves are solutions to such equations, and by integration with separation of variables that sinusoidal travelling waves (including a superposition of many) are the only solutions. For one dimension, the latter is called d'Alembert's method. You should know this. It is covered well in many textbooks, lecture notes, and websites.

1.5 Plane waves and complex exponentials

To avoid confusion, above we have developed a mathematical description of waves that is carefully built up from real functions only. However, as occurred in our study of differential equations, if we insist on carrying on in this way (which is perfectly possible in principle) things will get very complicated mathematically. So we now move towards allowing complex functions to be present in our mathematical expressions in a way that is mathematically rigorous and totally physical. This is centred around the idea of a plane (complex exponential) wave

$$g_{\mathbf{k}}(\mathbf{r}, t) = A_{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k} \cdot \mathbf{r} - \omega_{|k|} t)}, \tag{9}$$

where $A_{\mathbf{k}} = |A_{\mathbf{k}}| \mathrm{e}^{i\phi_{\mathbf{k}}}$. The real part of this plane wave, by inspection, is the sinusoidal plane wave considered above

$$f_{\mathbf{k}}(\mathbf{r}, t) = |A_{\mathbf{k}}| \cos{(\mathbf{k} \cdot \mathbf{r} - \omega_{|k|} t + \phi_{\mathbf{k}})} = \mathfrak{Re} \left\{ g_{\mathbf{k}}(\mathbf{r}, t) \right\} = \mathfrak{Re} \left\{ A_{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k} \cdot \mathbf{r} - \omega_{|k|} t)} \right\}. \tag{10}$$

[2]Perhaps the only one for dimensions higher than one.

So instead of describing ideal waves by real functions $f(\mathbf{r},t)$ that are solutions to the wave equation we equivalently deal with complex functions $g(\mathbf{r},t)$, solutions to the complex wave equation

$$\left(c^2\nabla^2 - \frac{\partial^2}{\partial t^2}\right) g(\mathbf{r},t) = 0, \tag{11}$$

where any (initial or boundary) conditions on $f(\mathbf{r},t)$ are now conditions on $\mathfrak{Re}\left\{g(\mathbf{r},t)\right\}$ and the solution to the real wave equation $f(\mathbf{r},t)$ at any position or time can be found from the real part $\mathfrak{Re}\left\{g(\mathbf{r},t)\right\}$ of the solution $g(\mathbf{r},t)$ to the complex wave equation.

It follows (from or in a very similar way to the real wave equation) that the general solution to the complex wave equation is

$$g(\mathbf{r},t) = \int \mathrm{d}\mathbf{k} g_{\mathbf{k}}(\mathbf{r},t) = \int \mathrm{d}\mathbf{k} A_{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - \omega_{|k|}t)}, \tag{12}$$

a superposition of plane waves, each representing, by its real part, a travelling sinusoidal wave.

We will use this complex notation throughout what follows, and see that it makes things generally simpler[3]. In your problem sets or exam, you should feel free to answer a question phrased in terms of real functions by taking shortcuts via complex functions, provided you explain yourself. You should be able to use and switch between both notations.

2 Common solutions to the ideal wave equation

We know the general solution to the wave equation but let's consider some common specific solutions. We've already done a single plane wave. Let's look at three more examples.

2.1 Two parallel plane waves of slightly different frequencies; beats

In this case we have two waves with wavevectors $\mathbf{k}$ and $\mathbf{k}' = \mathbf{k} + \Delta\mathbf{k}$ in the same direction so we may simplify notation by assuming they both move in the positive x-direction and moving into one dimension

$$g(x,t) = g_{\bar{k}+\Delta k/2}(x,t) + g_{\bar{k}-\Delta k/2}(x,t) \tag{13}$$

$$= \left[A_{\bar{k}+\Delta k/2}\mathrm{e}^{\mathrm{i}\Delta k(x-ct)/2} + A_{\bar{k}+\Delta k/2}\mathrm{e}^{-\mathrm{i}\Delta k(x-ct)/2}\right]\mathrm{e}^{\mathrm{i}\bar{k}(x-ct)}, \tag{14}$$

where we have introduced the mean frequency $\bar{k} = (k + k')/2$.

We see that this wave superposition consists of two terms. The second consists of oscillations at the central frequency $\omega_{|\bar{k}|} = c|\bar{k}|$ with an amplitude given by the first term that oscillates much more slowly (if $\Delta k \ll \bar{k}$) at the smaller angular frequency $\omega_{|\Delta k|/2} = c|\Delta k|/2$. The slowly oscillating amplitude term is often said to be the envelope of the quickly oscillating term.

If you listen to two musicians playing in near unison, you will hear 'beats' in the amplitude (these envelope functions). If they occur once a second then the musicians are 1 Hz out of tune with each other[4].

One can also do the same calculation as above using real functions, to obtain the real part of the above. This is

$$f(x,t) = \Big(|A_{\bar{k}+\Delta k/2}| + |A_{\bar{k}-\Delta k/2}|\Big)\cos(\Delta k[x-ct]/2 + \Delta\phi/2)\cos(\bar{k}[x-ct] + \bar{\phi}) \tag{15}$$

$$- \Big(|A_{\bar{k}+\Delta k/2}| + |A_{\bar{k}-\Delta k/2}|\Big)\sin(\Delta k[x-ct]/2 + \Delta\phi/2)\sin(\bar{k}[x-ct] + \bar{\phi}), \tag{16}$$

where we have introduced the phase mean and difference $\bar{\phi} = (\phi_{k+\Delta k/2} + \phi_{k-\Delta k/2})/2$ and $\Delta\phi = (\phi_{k+\Delta k/2} - \phi_{k-\Delta k/2})/2$.

It is a good idea to practice the above derivation, plot the result (with help from a plotting tool at first) and mark on the important quantities, just as for a single sinusoidal plane wave.

[3] In some places this simplification is more apparent that in others. You should get a feel for when.

[4] Note the absence of a factor of 2 here, as this has been cancelled by the fact that there will be two antinodes in the amplitude per period of the envelope oscillation.

2.2 Two plane waves, identical but in opposite directions; standing waves

In this case we have both waves $\mathbf{k}$ and $\mathbf{k}' = -\mathbf{k}$ in the same axis, just in opposite directions, so we may simplify notation by moving into one dimension

$$g(x,t) = g_{|k|}(x,t) + g_{-|k|}(x,t) \tag{17}$$

$$= \left[A_{|k|}\mathrm{e}^{\mathrm{i}|k|x} + A_{-|k|}\mathrm{e}^{-\mathrm{i}|k|x}\right]\mathrm{e}^{-\mathrm{i}\omega_{|k|}t}. \tag{18}$$

We see that this wave superposition consists of two terms. A wave that has a periodic spatial structure with wavelength $\lambda_{|k|} = 2\pi/|k|$ whose amplitude oscillates in time with angular frequency $\omega_{|k|} = c|k|$. You might immediately think of this as a standing wave, but you'd be wrong because when we take the real part of the product of two complex exponentials their behaviour combines.

To see this, consider the real part

$$f(x,t) = \left(|A_{|k|}| + |A_{-|k|}|\right)\cos(|k|x + \Delta\phi)\cos(-|k|ct + \bar{\phi}) \tag{19}$$

$$- \left(|A_{|k|}| - |A_{-|k|}|\right)\sin(|k|x + \Delta\phi)\sin(-|k|ct + \bar{\phi}). \tag{20}$$

where we have introduced the phase mean and difference $\bar{\phi} = (\phi_{|k|} + \phi_{-|k|})/2$ and $\Delta\phi = (\phi_{|k|} - \phi_{-|k|})/2$. Each term is a standing wave but both terms together do result in a partly travelling wave i.e. if $|A_{|k|}| \neq |A_{-|k|}|$ and the magnitudes of the two waves are not exactly equal, there will be an excess of one wave travelling in one direction, resulting in some directed wave. But if the two amplitudes are equal it is possible for two travelling waves to produce a wave where there is no movement in space.

2.3 Gaussian wave packet; localisation in real and k space

We have already stated that a superposition of plane waves could be used to construct any shape of travelling wave (and even standing waves as shown immediately above). Let us demonstrate this for a travelling wave with a Gaussian spatial distribution. Again, let's choose the direction to be along the x-axis and so work in one dimension.

$$g(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k g_k(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k A_k \mathrm{e}^{\mathrm{i}(kx - \omega_{|k|}t)}. \tag{21}$$

If we choose the amplitudes of the plane waves to be non-zero only for $k > 0$[5] distributed as Gaussian

$$A_k = \sqrt{\frac{a^2}{2\pi}}\mathrm{e}^{-(ka)^2/2}, \tag{22}$$

with width $1/a$ in k-space, then we find that

$$g(x,t) = \frac{1}{2}\exp\left[-\frac{1}{2}\left(\frac{x-ct}{a}\right)^2\right], \tag{23}$$

is a wave with the shape of a Gaussian with width a in real space.

This result is based on something you'll see lots of: the Fourier transform of a Gaussian is a Gaussian and their widths are inversely related. If the width in real space is large, then it is localised in k-space and vice versa. This is related to the uncertainty principle in quantum mechanics (as k is related to momentum).

3 Initial conditions

As we already know, given a general solution, all we need for a specific solution is some initial conditions. For the 1D wave equation, for example, all we need to know is $f(x,t)$ at some time t. Rather than give a general discussion of it here, you may explore this in the problem set.

[5] We want to have a wave going in the forward direction, not applying this condition would lead to the Gaussian wavepacket we are about to obtain superposed with its mirror image in the $x = 0$ plane.

4 Spatial inhomogeneity

By thinking about what types of functions represent ideal wave-like behaviour we have now identified the wave equation, the equation of motion governing systems that support ideal waves, and discussed some examples of wave-like behaviour. We were assuming the system supporting the wave to be spatially homogeneous, i.e. we assumed the speed of sound was a constant in this system. The benefit of having the equation of motion is that we can start considering what happens in situations beyond this homogeneous case. Let us do this now.

In general a system may be composed of multiple regions each with a different speed of sound. We, of course, start with two regions. The conclusions are easily generalised.

4.1 Transmission and reflection of a plane wave at a boundary

Let us consider two regions $x < 0$ and $x \geq 0$ where the speed of sound in each region is c_- and c_+, respectively.

We want to see what happens when a plane wave from the negative x-region travelling in the positive x-direction scatters from the boundary. All resulting waves will be plane waves travelling in the positive or negative x-direction and so it is essentially a one-dimensional problem.

Since the wave equation governs the motion in both regions we know the general solution is

$$g(x,t) = \begin{cases} \int_{-\infty}^{\infty} \mathrm{d}k A_k^- \mathrm{e}^{\mathrm{i}(kx-|k|c_-t)} & x < 0 \\ \int_{-\infty}^{\infty} \mathrm{d}k A_k^+ \mathrm{e}^{\mathrm{i}(kx-|k|c_+t)} & x \geq 0 \end{cases} . \tag{24}$$

From experience we expect that if the only input into the system is a plane wavevector k_I then there will be a single plane reflected wave with wavevector k_R in the negative direction and a single plane transmission wave with wavevector k_T in the positive direction. The accuracy of this assumption is supported by the consistency of the results we will obtain having made this ansatz:

$$g(x,t) = \begin{cases} A_{k_I} \mathrm{e}^{\mathrm{i}k_I(x-c_-t)} + A_{k_R} \mathrm{e}^{\mathrm{i}-k_R(x+c_-t)} & x < 0 \\ A_{k_T} \mathrm{e}^{\mathrm{i}k_T(x-c_+t)} & x \geq 0 \end{cases} . \tag{25}$$

It would be pretty unphysical if the this function was discontinuous or had a discontinuous derivative in space or time[6]. So we apply these boundary conditions. It quickly becomes obvious that in order for these to be satisfied we must have all angular frequencies equal, and so $k_I = -k_R$ and $k_I c_- = k_T c_+$. Next we see that we ensure continuity at $x = t = 0$ by choosing all phases to be equal, which results in

$$g(x,t) = \begin{cases} |A_{k_I}| \mathrm{e}^{\mathrm{i}k_I(x-c_-t)} + |A_{k_R}| \mathrm{e}^{\mathrm{i}k_I(x+c_-t)} & x < 0 \\ |A_{k_T}| \mathrm{e}^{\mathrm{i}k_I(c_-/c_+)(x-c_+t)} & x \geq 0 \end{cases} . \tag{26}$$

Finally, demanding a continuous derivative at $x = 0$ we find the reflection and transmission amplitudes

$$\frac{|A_{k_R}|}{|A_{k_I}|} = \frac{c_+ - c_-}{c_+ + c_-}, \tag{27}$$

$$\frac{|A_{k_T}|}{|A_{k_I}|} = \frac{2c_+}{c_+ + c_-}, \tag{28}$$

which depend only on the speeds of sound in the two regions.

Consider two special cases. Firstly, when the speed of sound in the second region $c_+ = 0$ is zero (does not support waves) then we get zero transmission and full reflection with an amplitude exactly opposite to that of the incoming wave (and from before we know this sets up a standing wave in the negative x-region). Secondly, when the speed of sound in the second region $c_+ = c_-$ is exactly equal to that of the first then we get zero reflection and full transmission with an amplitude equal to that

[6] When we get around to thinking about the types of quantities represented by $g(x,t)$ this will become obvious.

of the incoming wave. These two predictions are reassuring as any other result would have meant an error had been made.

It is worthwhile drawing these two special cases, as well as the general case. Understanding the relationship between the wavelength ratio and the speed of sound ratio, and other interesting quantities. Note also the transition in the sign of the amplitude of the reflected wave. Get a feel for the behaviour of the equations.

4.2 Transmission and reflection of a wavepacket at a boundary

The above treatment assumed an infinitely long perfectly periodic incident, reflected and transmitted wave. This is not a general wavepacket. But, yet, we do find the same reflection and transmission coefficients for any shape of wavepacket. Why is this?

Well, previously we have seen that any wavepacket will be a superposition of plane waves. And we have noted that the systems supporting ideal waves and the wave equation are linear. So we can split any plane wavepacket into its constituent plane wave parts, treat each one like the above, and put the result back together. What we find is the expected result: a wavepacket moves towards the boundary, and then a ratio $|A_{k_T}|/|A_{k_I}|$ of the amplitude is transmitted while a ratio $|A_{k_R}|/|A_{k_I}|$ is reflected back with a change in sign if $|A_{k_R}|/|A_{k_I}| < 0$.

It might be worth looking up some video clips/applets of this happening.

4.3 Conserved quantities

Note in the above example that $|A_{k_R}|^2/|A_{k_I}|^2 + |A_{k_T}|^2/|A_{k_I}|^2 = 1$ and so that the amplitude squared of the wave is conserved when an incident wave is scattered by an inhomogeneity. This mathematical conserved quantity will have to match a conserved quantity of the physical system, and indeed it will turn out to be energy.

4.4 Driving

Another way to introduce inhomogeneity is to have some driving. In mathematical terms we have been solving the homogeneous equation $\mathcal{D}g(\mathbf{r}, t) = 0$ and finding the complementary function of the system. If we were to allow some external driving, we would need to solve the equation $\mathcal{D}g(\mathbf{r}, t) = d(\mathbf{r}, t)$. The driving $d(\mathbf{r}, t)$ might depend on position $\mathbf{r}$, and can even be localised to a single point. We do not worry about driving in this course, leaving it as something to be covered in your study of partial differential equations next year.

4.5 Impedance

Another way of obtaining inhomogeneity is for there to be some resistance at a point (e.g. mass attached to a string). This changes the meaning of $f(\mathbf{r}, t)$ at that point and thus affects the conditions of continuity of the derivative of $f(\mathbf{r}, t)$ at that point. We will consider examples of this in the problem set.

Part II

Finite systems

Everything we've done above has been for infinite systems. We did introduce a boundary in the previous part, but this was really a boundary between two infinite systems.

So let's explicitly consider what happens with a finite system. And then show in each case how the infinite system behaviour is well approximated for a large enough system when we are not near the boundaries, reassuring us of the correctness of the previous part's analysis on the whole, but also revealing slight inaccuracies.

There are lots of types of finite systems, each specified by a boundary condition. We know the possible solutions to the wave equation. The boundary conditions will only be satisfied by some of them, hence the effect of boundary conditions and therefore finiteness is to reduce the number of valid solutions. The result is a general solution that is a discrete superposition of mode functions, e.g. plane waves, rather than a continuous superposition.

Let's stick to one dimension while we get our heads around this.

5 Periodic boundary conditions

We might imagine that our x-axis does not represent a line, but is in fact on some curved surface (e.g. the surface of a cylinder) that is closed. In this case our system has a finite length L and has periodic boundary conditions $f(x+L,t) = f(x,t)$, which we implement by demanding $g(x+L,t) = g(x,t)$.

By looking at the general solution to the wave equation

$$g(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k A_k \mathrm{e}^{\mathrm{i}(kx-\omega_{|k|}t)}, \tag{29}$$

we see that the only wavevectors allowed by our boundary condition satisfy $k = 2\pi n/L$ where n is some non-negative integer. And so the general solution that satisfies these boundary conditions is

$$g(x,t) = \sum_k A_k \mathrm{e}^{\mathrm{i}(kx-\omega_{|k|}t)}. \tag{30}$$

where in the sum k takes only the allowed values.

Thus we have a general solution that is a discrete superposition of plane waves. It is said that plane waves $g_k(x) \propto \mathrm{e}^{\mathrm{i}kx}$ are the discrete modes of a finite system with periodic boundary conditions.

For a really large L we expect things to get back to what we obtained considering an infinitely large system. They do. The spacing $\Delta k = 2\pi/L$ between the allowed wavevectors goes to zero, and the system supports superpositions of arbitrary plane waves, as assumed for the infinite system.

6 Box boundary conditions

A more physical boundary condition might be that two ends of the system are fixed, e.g. the box boundary conditions $f(0,t) = f(L,t) = 0$, which we implement by demanding $g(0,t) = g(L,t) = 0$.

By looking at the general solution to the wave equation

$$g(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k A_k \mathrm{e}^{\mathrm{i}(kx-\omega_{|k|}t)}, \tag{31}$$

we see that the amplitudes obeying our our first boundary condition $g(0,t) = 0$ satisfy $A_k = -A_{-k}$, allowing us to write

$$g(x,t) = \int_{0}^{\infty} \mathrm{d}k A_k 2i \sin(kx) \mathrm{e}^{-\mathrm{i}\omega_{|k|}t}, \tag{32}$$

where we only need to take the integral over non-negative k. The wavevectors satisfying our second boundary condition $g(L,t) = 0$ are then $k = \pi n/L$ for non-negative integer n. And so the general solution that satisfies these boundary conditions is

$$g(x,t) = \sum_k A_k 2i \sin(kx) \mathrm{e}^{-\mathrm{i}\omega_{|k|}t}. \tag{33}$$

where in the sum k takes only these allowed non-negative values.

Thus we have a general solution that is a discrete superposition of standing waves. It is said that standing waves $g_k(x) \propto \sin(kx)$ are the discrete modes of a finite system with box boundary conditions.

For a really large L we expect things to get back to what we obtained considering an infinitely large system. They do, but to see this is a bit more difficult than for periodic boundary conditions. The spacing $\Delta k = \pi/L$ between the allowed wavevectors goes to zero, but since the modes of the system are standing waves it is unclear how travelling waves arise (even though we know when a string is clamped at two ends, there is still a way to see waves travelling across it). Let's resolve this now.

Consider two modes with wavevectors only separated by a small $\Delta k = \pi/L$, i.e.

$$g(x,t) = A_{k+\Delta k/2} 2i \sin([k+\Delta k/2]x) \mathrm{e}^{-\mathrm{i}[k+\Delta k/2]ct} + A_{k-\Delta k/2} 2i \sin([k-\Delta k/2]x) \mathrm{e}^{-\mathrm{i}[k-\Delta k/2]ct} \tag{34}$$

$$= A_{k+\Delta k/2} 2 \left[\mathrm{e}^{\mathrm{i}\Delta k(x-ct)} \mathrm{e}^{\mathrm{i}k(x+ct)} + \mathrm{e}^{-\mathrm{i}\Delta k(x+ct)} \mathrm{e}^{-\mathrm{i}k(x-ct)} \right] \tag{35}$$

$$+ A_{k-\Delta k/2} 2 \left[\mathrm{e}^{\mathrm{i}\Delta k(x+ct)} \mathrm{e}^{\mathrm{i}k(x+ct)} + \mathrm{e}^{-\mathrm{i}\Delta k(x-ct)} \mathrm{e}^{-\mathrm{i}k(x-ct)} \right]. \tag{36}$$

Next consider some $x \approx L/2$, i.e. far away from the boundary so that the phases $\mathrm{e}^{\mathrm{i}\Delta k(x-ct)}$ and $\mathrm{e}^{-\mathrm{i}\Delta k(x+ct)}$ can have significantly different real parts. Further let's consider only a small region of time and space satisfying $\Delta x, c\Delta t \ll 1/\Delta k = L/\pi$ so that these phases don't vary in the region $\mathrm{e}^{\mathrm{i}\Delta k(x-ct)} \approx \mathrm{e}^{\mathrm{i}\phi_+}$ and $\mathrm{e}^{-\mathrm{i}\Delta k(x+ct)} \approx \mathrm{e}^{\mathrm{i}\phi_-}$. Then

$$g(x,t) = \left[A_{k+\Delta k/2} \mathrm{e}^{\mathrm{i}\phi_+} + A_{k-\Delta k/2} \mathrm{e}^{\mathrm{i}\phi_-} \right] \mathrm{e}^{\mathrm{i}k(x+ct)} + \left[A_{k+\Delta k/2} \mathrm{e}^{\mathrm{i}\phi_-} + A_{k-\Delta k/2} \mathrm{e}^{\mathrm{i}\phi_+} \right] \mathrm{e}^{-\mathrm{i}k(x-ct)}. \tag{37}$$

By choosing $|A_{k+\Delta k/2}| = -|A_{k-\Delta k/2}|$ and $\arg(A_{k+\Delta k/2}/A_{k-\Delta k/2}) = \pm(\phi_+ - \phi_-)$ we see that $g(x,t)$ can look like a plane wave moving in the $\pm x$-direction.

All this is a very complicated way to say that these standing wave modes do support what are essentially plane waves in regions somewhere roughly in the middle of the system that are always far from the boundaries and over times shorter than it would take to reach the boundaries. Our infinite system description is thus a good approximation of this behaviour.

7 Finite-system perspective on infinite systems

The main implication of this part is that, thankfully, even though we know all real systems are finite and affected by boundary conditions, ruling out plane waves with some wavevectors, as long as we consider only a small part of a system (for a short time) and are not right near the boundary then we are basically fine to use the mathematics of infinite systems developed in the previous parts, where arbitrary plane waves are allowed. That's lucky, as it would be a shame to have to retract everything said in the previous parts.

There are, however, some differences due to the discrete nature of the modes. Firstly, if the system is driven at some frequency, then, because of the discrete rather than continuous frequency distribution of the modes of the system, the driving will go through resonances as the frequency nears that of one of the discrete modes.

Secondly, there is an implication will not be touched upon this year. The fact that there are discrete modes of finite systems will be crucial in quantum mechanics, for examples, you'll be solving the wave equation again when considering Sommerfeld's model for electrons in a solid in third year. Both the number of distinct modes increases with system size as does the number of electrons. Each electron must occupy a different mode (Pauli exclusion principle) and so the physically crucially depends on the number of electrons per unit volume. This can't be understood in the context of a continuum of modes.

The take away message is that boundary conditions do have some effect by replacing a continuum by a discrete set of modes. But actually, exactly which boundary conditions you choose aren't important as long as you have a big system and are away from the boundary conditions. So in future years you might often choose periodic boundary conditions, because the plane wave modes are easier to work with, even though the system is not periodic. When you do, you can think back to this section for justification.

Part III
Dispersive systems

We have assumed in all of the above that waves of all shapes travel at the same speed. This reduces to assuming that all modes have a frequency $\omega_{|k|} = c|k|$ that is linear in the wavevector.

What happens if this is no longer the case, as it is known not to be[7]? Dispersion happens, and this part is about highlighting how this occurs, and then finding some equations that predict dispersive solutions.

8 Dispersion of a Gaussian wavepacket

In a dispersive system, plane waves are still the solutions to the equation of motion, i.e.

$$g(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k g_k(x,t) = \int_{-\infty}^{\infty} \mathrm{d}k A_k \mathrm{e}^{\mathrm{i}(kx-\omega_{|k|}t)}, \tag{38}$$

but $\omega_{|k|} \neq c|k|$.

It is clear how this effects the evolution of a single plane wave, it just changes its phase velocity (really, it is a speed, but seems to always get called velocity) $c^{\mathrm{p}}_{|k|} = \omega_{|k|}/|k|$. Let us then consider the effect of this on a superposition of plane waves, i.e. a wavepacket. A relatively easy to study and representative wavepacket is the Gaussian wavepacket. We previously showed that a superposition of modes with Gaussian amplitudes

$$A_k = \sqrt{\frac{a^2}{2\pi}} \mathrm{e}^{-(ka)^2/2}, \tag{39}$$

for $k > 0$ corresponds to a Gaussian wavepacket

$$g(x,t) = \frac{1}{2} \exp\left[-\frac{1}{2}\left(\frac{x-ct}{a}\right)^2\right], \tag{40}$$

travelling with speed c in the x-direction if $\omega_{|k|} = c|k|$.

Let us now consider the effects of dispersion. For a small enough spread in k-space, this will be captured by the MacLaurin series

$$\omega_{|k|} = c^{\mathrm{g}}_0|k| + \tfrac{1}{2}D_0 k^2, \tag{41}$$

where $c^{\mathrm{g}}_0 = \partial\omega_{|k|}/\partial|k|$ is the group velocity at $|k| = 0$ and $D_0 = \partial^2\omega_{|k|}/\partial|k|^2$ is the dispersion parameter at $|k| = 0$.

Inserting this into the general solution Eq. (38) we find that, as before, at $t = 0$ the wavepacket has a Gaussian shape

$$g(x,0) = \frac{1}{2} \exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^2\right], \tag{42}$$

with width a. But at a later time

$$\begin{aligned}
g(x,t) &= \int_{-\infty}^{\infty} \mathrm{d}k \sqrt{\frac{a^2}{2\pi}} \mathrm{e}^{-(ka)^2/2} \mathrm{e}^{\mathrm{i}(kx-[c^{\mathrm{g}}_0|k|+\frac{1}{2}D_0k^2]t)} && (43)\\
&= \int_{-\infty}^{\infty} \mathrm{d}k \sqrt{\frac{a^2}{2\pi}} \mathrm{e}^{-k^2(a^2+\mathrm{i}D_0t)/2} \mathrm{e}^{\mathrm{i}(kx-c^{\mathrm{g}}_0|k|t)} && (44)\\
&= \frac{1}{2\sqrt{1+D_0^2t^2/a^4}} \exp\left[-\frac{1}{2}\frac{(x-c^{\mathrm{g}}_0t)^2}{a^2+\mathrm{i}D_0t}\right] && (45)\\
&= \frac{1}{2\sqrt{1+D_0^2t^2/a^4}} \exp\left[-\frac{1}{2}\frac{(x-c^{\mathrm{g}}_0t)^2}{a^2(1+D_0^2t^2/a^4)}\right] \exp\left[-\mathrm{i}\frac{D_0t}{2}\frac{(x-c^{\mathrm{g}}_0t)^2}{a^2(1+D_0^2t^2/a^4)}\right]. && (46)
\end{aligned}$$

[7]For example, some water waves are known to have a frequency scaling with the root of $|k|$

The second term on the final line is just a change in phase and is not important. The first term describes a Gaussian wavepacket whose centre is moving at the group (not phase) velocity $c_0^{\mathrm{g}} = \partial\omega_{|k|}/\partial|k|$ and whose width increases in time by a factor $\sqrt{1 + D^2t^2/a^4} \approx 1 + D^2t^2/2a^4$ for small time, i.e. the wavepacket disperses.

This example captures everything you really need to know about dispersion. Localised packets move at the group velocity $c^{\mathrm{g}} = \partial\omega_{|k|}/\partial|k|$, which for dispersive systems is not necessarily the same as the phase velocity $c^{\mathrm{p}} = \omega_{|k|}/|k|$ or the speed of sound $c^{\mathrm{s}} = \lim_{|k|\to 0}\omega_{|k|}/|k|$, and the dispersion also causes them to spread out.

9 Superposition of two waves in a dispersive system

The Gaussian example above is a bit too complicated for this course. Instead, often people consider it easier to consider the superposition of only two waves, with wavevectors that are slightly different, as we did before, but this time with dispersion. The envelope of the system is taken as the wavepacket and travels at the group velocity, though there are waves within this travelling at the phase velocity. It's not such a nice example, as you don't see the spreading out of the wavepacket, but nevertheless it's worth trying.

10 Dispersive wave equation

In order to capture dispersion the wave equation must be adjusted so that plane waves are still solutions, but the dispersion relation $\omega_{|k|}(|k|)$ is no longer a straight line. There are many ways of achieving this, and you will study a few examples in the problems.

Part IV
Dissipative systems

Another big topic in waves is dissipation. What this means, as with ODEs, is just some damping of the waves. Mathematically and physically there is nothing particularly novel here, except for complexity. We normally completely ignore dissipation except to say that in the infinite time limit dissipation will always mean that the complementary function is decayed, leaving the particular integral as the description of the long time behaviour, as with the driven damped harmonic oscillator. However, it is worth noting here that ignoring dissipation is an idealisation.

Part V
Non-linear systems

So far we have studied how waves can arise in a linear system. We saw that for an ideal linear system we have waves that do not spread out, but in practice there will be some spreading out. Some of these conclusions change if we allow the equation of motion of a wave to be non-linear. The cornerstone example is the idea of a soliton. A soliton is a solitary wave that maintains its shape and propagates at a constant velocity. Importantly they are not the result of an ideal wave equation, rather they are caused by the cancellation of nonlinear and dispersive effects. They are observed in shallow water waves and mathematically can be found as the solutions of non-linear dispersive partial differential equations. They don't seem to be on syllabus, but see a textbook for an example.

Part VI
Real systems

This is the part that is most intuitive and often covered in lectures and textbooks: how to derive the equations of motion for different real physical systems. After this is done, the solutions and expected behaviours can be found from the mathematical discussions above.

Part VII
Discrete systems

Normal modes are merely a special case of wave behaviour where instead of some $g(\mathbf{r}, t)$ that is a continuous function of $\mathbf{r}$ we have some $g_i(t)$ that is a function of a discrete parameter i. [Note that here we are going straight into a representation based on complex numbers.] Think of replacing a continuum of mass all connected, by a series of point masses connected by springs, where each $g_i(t)$ marks the displacement of a mass in some direction. We might store these N displacements as a column vector $\mathbf{g}(t) = (g_1(t), g_2(t), \ldots, g_N(t))^T$.

Often we have such a discretized system e.g. a lattice of positively-charged atoms interacting by the electric field, and we seem to observe wave behaviour e.g. sound waves in solids. How do we account for this mathematically, as we did for waves in a continuum? A starting point is to look for a fundamental building block of vibration that fits with what we think of as a wave. For continuous media this was a plane wave, characterised by the fact that the value of $g(\mathbf{r}, t)$ at all points in space oscillated at the same frequency, i.e. $g(\mathbf{r}, t) = g(\mathbf{r})\mathrm{e}^{-\mathrm{i}\omega t}$. For waves in discrete degrees of freedom we think of the building blocks in exactly the same way, those where all degrees of freedom oscillate with the same frequency $\mathbf{g}_i(t) = \mathbf{g}_i\mathrm{e}^{-\mathrm{i}\omega t}$, where we have indexed solutions of this type by i. Here ω_i is the frequency of oscillation and $\mathbf{g}_i$ is a vector storing the amplitudes of the oscillations in the various degrees of freedom. These fundamental solutions $\mathbf{g}_i(t)$ are called the normal modes of the system. Once again assuming linearity, we write the general wavelike solutions as $\mathbf{g}(t) = \sum_i q_i\mathbf{g}_i(t) = \sum_i q_i\mathbf{g}_i\mathrm{e}^{-\mathrm{i}\omega_i t} = \sum_i Q_i(t)\mathbf{g}_i$ where $Q_i(t) = q_i\mathrm{e}^{-\mathrm{i}\omega_i t}$ are called the normal coordinates and oscillate sinusoidally at frequency ω_i.

As with continuous media we would like to seek a type of equation for which the normal mode superpositions $\mathbf{g}(t)$ above are solutions. As we are looking for something linear that supports oscillations but contains no dissipation, there are very few options, and we quickly are led to some sets of equations, the simplest being $\ddot{\mathbf{g}}(t) + \mathbf{A}\mathbf{g}(t) = 0$, with some matrix $\mathbf{A}$ that tells us how the different degrees of freedom are coupled. Typically $\mathbf{A}$ is diaongalisable with N linearly independent eigenvectors and real non-negative eigenvalues. Performing such a diagonalisation $\mathbf{A} = \mathbf{R}\mathbf{D}\mathbf{R}^T$ where $\mathbf{D}$ is diagonal with elements $D_{ii} = \omega_i^2$ and orthogonal transformation matrix $\mathbf{R} = (\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N)$ with columns $\mathbf{g}_i$, we arrive at $\ddot{\mathbf{g}}(t) + \mathbf{R}\mathbf{D}\mathbf{R}^T\mathbf{g}(t) = 0$. This is easily transformed to $\mathbf{R}^T\ddot{\mathbf{g}}(t) + \mathbf{D}\mathbf{R}^T\mathbf{g}(t) = 0$, and introducing transformed coordinates $\mathbf{g}'(t) = \mathbf{R}^T\mathbf{g}(t)$, which are just the normal coordinates $\mathbf{g}' = (Q_1, Q_2, \ldots, Q_N)^T$, we arrive at the uncoupled set of equations $\ddot{Q}_i(t) + \omega_i^2 Q_i(t) = 0$ for the normal coordinates that support sinusoidal solutions $Q_i(t) = q_i\mathrm{e}^{-\mathrm{i}\omega_i t}$. Returning to the original coordinates $\mathbf{g}(t) = \mathbf{R}\mathbf{g}'(t)$, we see this can be expanded as $\mathbf{g}(t) = \sum_i Q_i(t)\mathbf{g}_i$ the solution we hypothesised before stating the equation to which it is the general solution.

All this goes to show is that N linearly coupled oscillators (the assumption of positive eigenvalues above is equivalent to assuming that the coupling of the degrees of freedom does support oscillations) will, through a change of basis, look like N uncoupled oscillating degrees of freedom Q_i, which are the normal coordinates.

We have done this for the coupled oscillators $\ddot{\mathbf{g}}(t) + \mathbf{A}\mathbf{g}(t) = 0$ and related the normal mode frequencies and amplitudes to eigenvalues and eigenvectors of $\mathbf{A}$. What one normally does, which amounts to the same thing, is make the ansatz $\mathbf{g}_i(t) = \mathbf{g}_i\mathrm{e}^{-\mathrm{i}\omega t}$ and substitute this into the equation $\ddot{\mathbf{g}}(t) + \mathbf{A}\mathbf{g}(t) = 0$, obtaining $(-\omega_i^2\mathbf{1} + \mathbf{A})\mathbf{g}_i = 0$ for the normal mode frequency ω_i and amplitudes $\mathbf{g}_i$.

A frequency ω_i corresponds to non-trivial solutions iff $|-\omega_i^2\mathbf{1}+\mathbf{A}| = 0$ and so we solve this equation to find the normal mode frequencies ω_i. Once this is done we solve linear equation $(-\omega_i^2\mathbf{1}+\mathbf{A})\mathbf{g}_i = 0$ to obtain normal mode amplitudes $\mathbf{g}_i$ (note the similarities of this to the method finding eigenvalues and eigenvectors). A general solution is thus written $\mathbf{g}(t) = \sum_i Q_i(t)\mathbf{g}_i$ with sinusoidal normal coordinates $Q_i(t) = q_i e^{-i\omega_i t}$.

It is possible, with experience, to find the normal modes and thus the general solution to such equations even more quickly. This is done by exploiting symmetries. Let's go back to the equation $\ddot{\mathbf{g}}(t) + \mathbf{A}\mathbf{g}(t) = 0$ for $N = 2$ degrees of freedom. We might know our system has the symmetry that the equation does not change under the swap $g_1 \leftrightarrow g_2$. We can use this to infer the two eigenvectors $\mathbf{g}_1 = (1,1)^T$ and $\mathbf{g}_1 = (1,-1)^T$ without any more information. This approach is actually how we go about solving some systems for arbitrary large N, where a brute force finding of the eigenvectors and eigenvalues would be intractable.

Above is the general discussion, you will see many specific examples in the problem set.