2.3 3D waves in a spherical cavity

Having looked at 1D examples in detail, let's now move to the 3D case, and consider waves in a homogeneous spherical cavity. This is not quite the case of pressure waves in stars (where the fluid is not homogeneous), but it is one step in the right direction. The 3D wave equation, in a spherical coordinate system (r, θ, ϕ) is given by

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial p}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 p}{\partial \phi^2} \right] \tag{2.1}$$

For simplicity, we'll take $p = 0$ at $r = R$ as boundary condition, and require that p be bounded at $r = 0$.

As before we will look for basic solutions via separation of variables. Let's first assume separation of the spatial and temporal variables, i.e., $p(r, \theta, \phi, t) = a(r, \theta, \phi) b(t)$. Then,

$$\frac{1}{b} \frac{d^2 b}{dt^2} = \frac{c^2}{a} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial a}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial a}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 a}{\partial \phi^2} \right] \tag{2.2}$$

so, as before, since the LHS is only a function of t while the RHS is only a function of the spatial coordinates, we can set both to a constant. And as before, it can reasonably easily be shown that this constant must be negative or 0, hence we write

$$\begin{aligned} \frac{d^2 b}{dt^2} &= -\omega^2 b \\ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial a}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial a}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 a}{\partial \phi^2} &= -\frac{\omega^2}{c^2} a \end{aligned} \tag{2.3}$$

As always with separation of variables, we start with the spatial problem. We separate a into r, θ and ϕ coordinates, as in $a(r, \theta, \phi) = f(r) g(\theta) h(\phi)$. Then

$$\frac{1}{fr^2} \frac{d}{dr} \left(r^2 \frac{df}{dr} \right) + \frac{1}{gr^2 \sin\theta} \frac{d}{d\theta} \left(\sin\theta \frac{dg}{d\theta} \right) + \frac{1}{hr^2 \sin^2\theta} \frac{d^2 h}{d\phi^2} = -\frac{\omega^2}{c^2} \tag{2.4}$$

Multiplying the whole equation by $r^2 \sin^2\theta$ and rearranging it, we get

$$\frac{\sin^2\theta}{f} \frac{d}{dr} \left(r^2 \frac{df}{dr} \right) + \frac{\sin\theta}{g} \frac{d}{d\theta} \left(\sin\theta \frac{dg}{d\theta} \right) + \frac{\omega^2}{c^2} r^2 \sin^2\theta = -\frac{1}{h} \frac{d^2 h}{d\phi^2} \tag{2.5}$$

Now the LHS is a function of r and θ only, while the RHS is a function of ϕ only, so both have to be constant. Based on the fact that the solutions in ϕ will have to be periodic with period 2π, we can already guess that this constant will have to be the square of an integer number, which we call m^2. Hence

$$\begin{aligned} \frac{d^2 h}{d\phi^2} &= -m^2 h \\ \frac{\sin^2\theta}{f} \frac{d}{dr} \left(r^2 \frac{df}{dr} \right) + \frac{\sin\theta}{g} \frac{d}{d\theta} \left(\sin\theta \frac{dg}{d\theta} \right) + \frac{\omega^2}{c^2} r^2 \sin^2\theta &= m^2 \end{aligned} \tag{2.6}$$

The solutions for h are therefore simply linear combinations of $\cos(m\phi)$ and $\sin(m\phi)$, for each value of m selected.

Let's now divide the remaining (r, θ) equation by $\sin^2\theta$ and rearrange the terms:

$$\frac{1}{f}\frac{d}{dr}\left(r^2\frac{df}{dr}\right)+\frac{\omega^2}{c^2}r^2=-\frac{1}{g\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dg}{d\theta}\right)+\frac{m^2}{\sin^2\theta} \tag{2.7}$$

Again, we have worked this out so the LHS is a function of r only, while the RHS is a function of θ only, so both have to be equal to a constant. This time, it's not as obvious what the sign of that constant has to be, so for the moment let's just call it α. The θ equation is

$$-\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dg}{d\theta}\right)+\frac{m^2}{\sin^2\theta}g=\alpha g \tag{2.8}$$

It is customary in working with equations in spherical coordinate systems to introduce the new variable $x = \cos\theta$, where x varies between -1 and 1 as θ varies between $-\pi/2$ and $\pi/2$. Then with $d/d\theta = -\sin\theta d/dx$, the θ equation becomes

$$\frac{d}{dx}\left((1-x^2)\frac{dg}{dx}\right)-\frac{m^2}{(1-x^2)}g+\alpha g=0 \tag{2.9}$$

In this form, it is now possible to recognize it as the equation for the Legendre functions $P_l^m(x)$ and $Q_l^m(x)$, as long as the constant $\alpha = l(l+1)$. The $Q_l^m(x)$ are singular at $x = \pm 1$, so we can rule these solutions out – but keeping the other solutions, we have found the solution for g:

$$g_{lm}(\theta) = P_l^m(\cos\theta) \tag{2.10}$$

Note that the Legendre functions $P_l^m(\cos\theta)$ are equal to the Legendre Polynomials $P_l(\cos\theta)$ for $m = 0$. Also, $P_l^m(\cos\theta)$ is identically 0 if $m > l$, so the only values of m allowable are $m \le l$. Finally, note that the product of g and h is a well-know set of functions called the spherical harmonics, and usually written $Y_l^m(\theta, \phi)$. For convenience, it is customary to write them in complex form to include both the sine and the cosine part of h, as in:

$$Y_l^m(\theta,\phi) = e^{im\phi}P_l^m(\cos\theta) \tag{2.11}$$

The structure of the spherical harmonics is shown in Figure 2.1. For $l = m = 0$, we simply have the constant function. For $l = 1$ and $m = 0$, the function is invariant in ϕ and has one node in latitude at the equator. For $l = 1$ and $m = 1$, it's exactly the same pattern but rotated by 90°. Note how it's equal to 0 at the poles, and has 2 nodes in longitude. For $l = 2$, $m = 0$, the function has 2 nodes in latitude, but is invariant in longitude. For $l = 2$, $m = 2$, again the function is 0 at the poles, and has 4 nodes in longitude. Generally speaking, if $m = 0$ the function has l nodes in latitude, while if $l = m$ the function is 0 at the poles, and has $2m$ nodes in longitude. In order words the larger l or m, the more complex the spatial structure of the mode on the sphere.

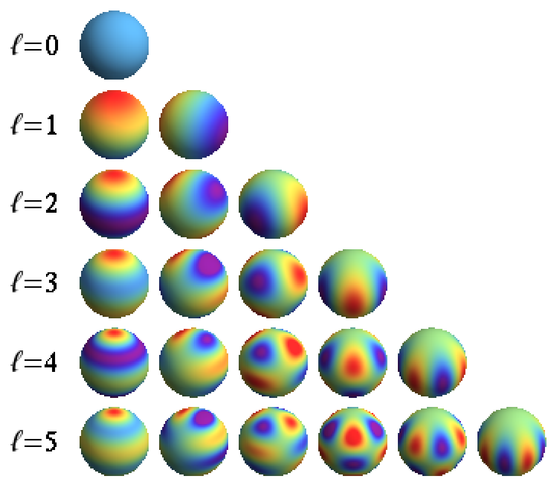

Figure 2.1: Spherical harmonics as a function of l and m. Image ripped from the web. m increases to the right, with $m = 0$ on the first column.

Nextt, we have to solve the remaining radial equation:

$$\frac{d}{dr}\left(r^2 \frac{df}{dr}\right) + \frac{\omega^2}{c^2} r^2 f - l(l+1)f = 0 \tag{2.12}$$

which can be rewritten as

$$r^2 \frac{d^2 f}{dr^2} + 2r\frac{df}{dr} + \frac{\omega^2}{c^2} r^2 f - l(l+1)f = 0 \tag{2.13}$$

This actually looks *very* similar to a spherical Bessel equation, namely

$$z^2 f'' + 2zf + (z^2 - l(l+1))f = 0 \tag{2.14}$$

except for the term in ω^2/c^2 in front of r^2. In fact, we can easily get rid of it by creating the variable $z = \omega r/c$. Then our equation is exactly the spherical Bessel equation, which has the solutions $j_l(z)$ and $y_l(z)$. The $y_l(z)$ solutions are singular at $z = 0$, so we can discard them. We are then simply left with

$$f(r) = j_l\left(\frac{\omega r}{c}\right) \tag{2.15}$$

In order for $f(r)$ to satisfy the boundary conditions, we need $f(R) = 0$. This implies that $\omega R/c$ must be a zero of the spherical Bessel function of degree l. If we call these zeros z_{nl} (for the n-th zero of the j_l function), this sets the eigenfrequency ω to be

$$\omega_{nl} = \frac{z_{nl}c}{R} \tag{2.16}$$

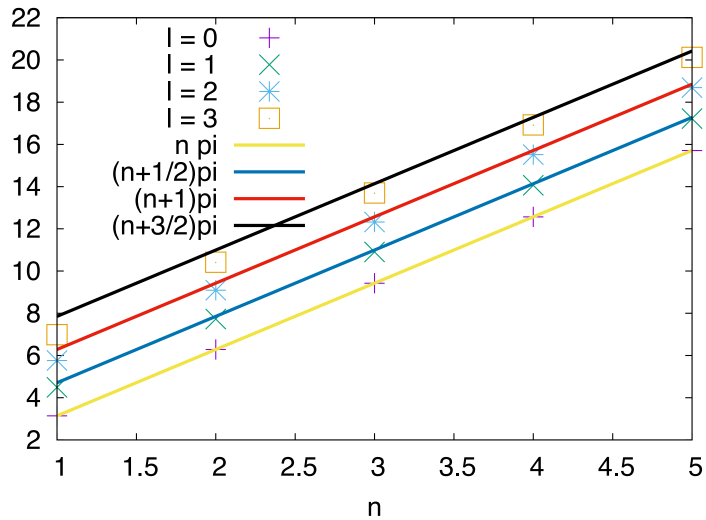

Figure 2.2: First few zeros of the first 4 spherical Bessel functions j_0 to j_3, as well as the corresponding asymptotic approximation. As we can see, the zeros are nearly linear functions of n and increase slowly with l. This is also equal to $\omega_{nl}R/c$ and therefore gives an idea of how the mode frequency changes with both n and l.

The z_{nl} are tabulated numbers, that can easily be found in appropriate textbooks (e.g. Abramowitz & Stegun), using Mathematica/Wolfram Alpha, or in various sites on the web. They are shown in Figure 2.2. For very large n (which correspond to very large zeros), there is an asymptotic approximation to the zeros that can be derived from WKB theory which states that (see Abramowitz and Stegun equation 9.5.12):

$$z_{nl} \simeq \left[n + \frac{l}{2}\right] \pi \text{ for } n \gg l \tag{2.17}$$

In practice, this approximation is exact for $l = 0$, and is already quite good for n that is not that much larger than l for $l \neq 0$. For each of these eigenfrequencies, we have the corresponding eigenfunction

$$f_{nl}(r) = j_l \left(z_{nl} \frac{r}{R} \right) \tag{2.18}$$

Going back to the temporal equation, finally, we see that each of the eigenfrequencies has a corresponding temporal solution as

$$b_{nl}(t) = \alpha_{nl} \cos(\omega_{nl} t) + \beta_{nl} \sin(\omega_{nl} t) \tag{2.19}$$

Figure 2.2 therefore also shows how $\omega_{nl} = z_{nl} c/R$ changes with n and l in this system.

Putting everything together, we have the full solution as

$$p(r,\theta,\phi,t) = \sum_{nlm} j_l\left(z_{nl}\frac{r}{R}\right) P_l^m(\cos\theta)\left(C_{mln}\cos(m\phi) + D_{mln}\sin(m\phi)\right) \times \left[\alpha_{nl}\cos\left(\frac{z_{nl}ct}{R}\right) + \beta_{nl}\sin\left(\frac{z_{nl}ct}{R}\right)\right] \quad (2.20)$$

where the various constants can be found by fitting the initial conditions if necessary. In practice, that's rather messy so we won't do it here.

As in the 1D case, we find that the solutions are a superposition of standing waves, whose spatial form is given by the product of the spherical Bessel function $j_l(z_{nl}r/R)$ with the spherical harmonic $Y_l^m(\theta,\phi)$. This time there are 3 important integer numbers that fully characterize the solution, namely n (called the radial order), l (called the latitudinal wavenumber) and m (the longitudinal wavenumber). The eigenfrequencies, however, only depend on n and l, showing that there is some degeneracy (i.e. different spatial modes can have the same eigenfrequency).

For $l = 0$, the only possible value of m is 0, so the corresponding spherical harmonic is simply $P_0^0(\theta,\phi)$, which is the constant function. In that case, the corresponding spatial eigenmodes have no structure in the latitudinal and horizontal directions, and the radial part is equal to $j_0(z_{n0}r/R)$, which is also

$$j_0\left(\frac{z_{n0}r}{R}\right) = \frac{R}{z_{n0}r}\sin\left(\frac{z_{n0}r}{R}\right) \quad (2.21)$$

Meanwhile, the zeros z_{n0} are simply $z_{n0} = n\pi$. These radial eigenfunctions are shown in Figure 2.3. As for the case of the 1D wave in the tube, we see that the fundamental mode ($n = 1$ here) has no nodes, the first harmonic has 1 node, the second harmonic has 2 nodes, etc... so as before, the larger n, the more complex the spatial structure of the mode. The eigenfrequencies of each of these radial modes are simply $\omega_{0n} = z_{0n}c/R = n\pi c/R$. The fundamental eigenmode has frequency $\omega 01 = \pi c/R$, so the period of that mode is equal to the time it takes for the sound to get from the surface to the center and back (as it was in the case of the tube).

For $l > 0$, there is not always a simple analytical expression for the zeros or for the radial eigenfunctions. However, for large l, we can use WKB theory to approximate the $j_l(z)$ functions and their zeros if needed. As shown in Figure 2.2, and through asymptotic theory, the frequencies corresponding to different n and l increase more-or-less linearly with both n and l.

Having learned about the case of pressure waves in an isothermal sphere, let's now look at pressure waves in a real star, the Sun. We have, thanks to helioseismology, excellent data on the frequency of various pressure modes observed at the surface of the Sun. Because we can only observe the surface, we can measure

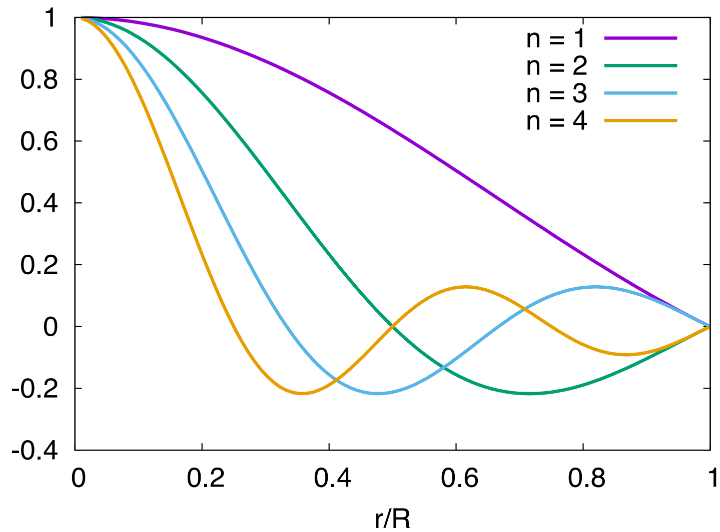

Figure 2.3: The first 4 radial eigenfunctions for $l = m = 0$.

the spatial structure of the mode in latitude and longitude, but not in radius. Hence, it is customary to plot the mode frequency as a function of the latitudinal wavenumber l – since we don't know what n is. The results are shown in Figure 2.4, obtained with the Michelson Doppler Imager (MDI) instrument on board the spacecraft SOHO. We see several ridges, which each correspond to a different radial order n. Because the Sun is rotating, the modes are not completely degenerate in m as it would be for a non-rotating star. Hence different modes with different m have slightly different frequencies, which is why each ridge has a significant thickness. The ridges are not really linear, which can be attributed to the fact that the Sun is not isothermal.

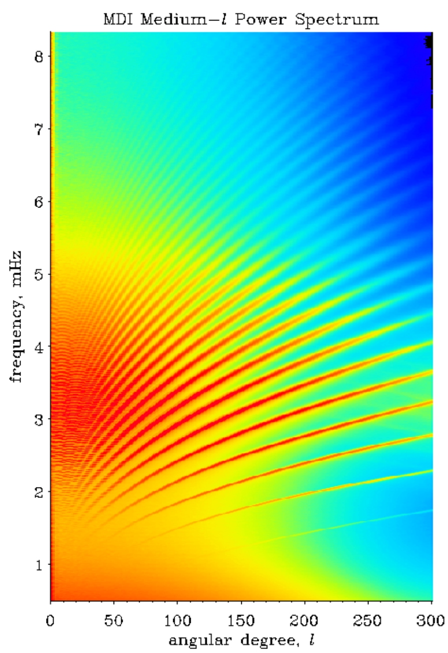

Figure 2.4: Helioseismic data on pressure waves in the Sun. The horizontal axis shows the measured latitudinal wavenumber l, and the vertical axis shows the measured oscillation frequency of the mode. The color shows the power in the mode (i.e. how strong it is).

2.4 Wave packets solutions

We will now switch gear and learn a new method of studying waves that revolves around the use of wave-packets. This will turn out to be a very powerful tool that can very easily be generalized in more than 1D, and, much more importantly, for non-constant sound speed – which we had so far ignored. Let's proceed to build the components of that solution step by step, then look at some examples.

2.4.1 Wave packets in 1D

Consider for the moment just two waves of nearby wavenumbers, and therefore nearby frequencies. For simplicity, let's just take real cosine waves, and assume their amplitudes are similar. The sum of these waves yields

$$\begin{aligned}
&A(k)\cos\left[k(x-ct)\right] + A(k)\cos\left[(k+dk)(x-ct)\right] \\
&= 2A(k)\cos\left[\left(k+\frac{dk}{2}\right)(x-ct)\right]\cos\left[\frac{dk}{2}(x-ct)\right] \\
&\simeq 2A(k)\cos\left[k(x-ct)\right]\cos\left[\frac{dk}{2}(x-ct)\right] \qquad (2.22)
\end{aligned}$$

using the standard trigonometric identity $\cos a + \cos b = 2\cos((a+b)/2)\cos((a-b)/2)$. This shows that the sum of two waves of nearby frequencies can also be written as the product of two waves, one which has the original frequency $\omega = ck$ and one with a much lower one, $d\omega = cdk/2$. This well-known phenomenon is called *beating*, illustrated in Figure 2.5

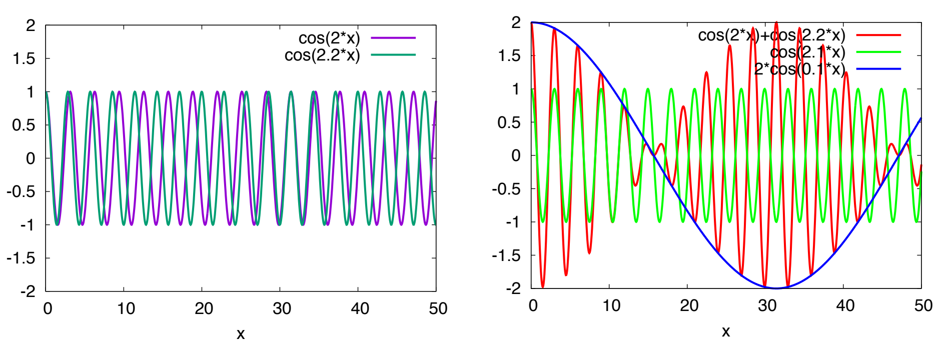

Figure 2.5: Left: 2 waves, one of frequency 2, and one of frequency 2.2. Right: the sum of these two waves, which shows the beating phenomenon. The carrier wave now has frequency 2.1, while the modulation has frequency 0.1 and amplitude 2

However, the key aspect illustrated here is the fact that an exact solution of the wave equation (given by the sum of plane waves) can also be approximated as the product of a rapidly oscillating function times a slowly varying one. This suggests that there may be some circumstances in which functions that are given as the product of a slowly varying function times a rapidly oscillating one are good approximations to the true solution of the wave equation. Let us see in which cases this may happen.

Based on the discussion above, we consider the following general form of solution:

$$p(x,t) = A(\epsilon x, \epsilon t)e^{i\theta(x,t)} \tag{2.23}$$

where $e^{i\theta(x,t)}$ is the carrier plane wave, and A is its modulated amplitude. Note how we have expressed A as a function of "slow variables" $X = \epsilon x$ and $T = \epsilon t$ to imply that they vary slowly with x and t – at least, much more slowly than the variations intrinsic to the plane wave itself[1]. As discussed earlier, A could be complex, and in order to extract the true physical value of p when needed, we shall always take its real part. In what follows, p will now be function of x, t but also of X and T as

$$p(x,X;t,T) = A(X,T)e^{i\theta(x,t)} \tag{2.24}$$

Since the phase of the wave is now defined as the more general function $\theta(x,t)$ instead of the function $kx - \omega t$ that is specific to plane monochromatic waves, we no longer have an obvious explicit definition for k and ω. However, let's remember that

- The period of a wave is defined by how long one needs to wait before it is in the same phase again (modulo $\pm 2\pi$)
- The wavelength of a wave is defined by how far one has to move to see it in the same phase again (modulo $\pm 2\pi$)

In other words, if the period of the wave is $2\pi/\omega$, then

$$\theta\left(x, t + \frac{2\pi}{\omega}\right) = \theta\,(x,t) \pm 2\pi \tag{2.25}$$

Taylor expanding the first term, we then get

$$\frac{2\pi}{\omega}\frac{\partial\theta}{\partial t} = \pm 2\pi \tag{2.26}$$

so that

$$\omega = \pm\frac{\partial\theta}{\partial t} \tag{2.27}$$

It now remains to be seen which of the + or − sign is consistent with the plane wave definition. We see that the plane wave would have

$$\omega = -\frac{\partial\theta}{\partial t} \tag{2.28}$$

[1]To see why $X = \epsilon x$ (and similarly, $T = \epsilon t$) are slow variables, plot the functions $\cos(x)$ and $\cos(X)$ side by side.

and therefore choose the $-$ sign solution. Similarly, we can construct the wavenumber k to be

$$k = \frac{\partial \theta}{\partial x} \tag{2.29}$$

which is also consistent with the plane wave definition. We then have a relationship between k and ω, namely

$$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0 \tag{2.30}$$

Finally, note that *unless* we actually have a plane wave with $\theta = kx - \omega t$, k and ω are generally functions of x and t themselves. In the wave-packet approximation, however, we shall require that they only be *slowly* varying functions of x and t, meaning that they are functions of X and T only. As a result $\partial k/\partial t = (\partial k/\partial T)(\partial T/\partial t) = \epsilon \partial k/\partial T$ and similarly for $\partial \omega/\partial x$. The equation above then becomes, to first order in ϵ,

$$\frac{\partial k}{\partial T} + \frac{\partial \omega}{\partial X} = 0 \tag{2.31}$$

Let's now plug the wave-packet solution into the 1D wave equation. To do so, we need to evaluate partial derivatives of p with respect to t and x, remembering that there is a t-dependence in T, and an $x-$dependence in X. We have:

$$\begin{aligned} \frac{\partial}{\partial t} p(x, X; t, T) &= \frac{\partial p}{\partial t} + \frac{\partial T}{\partial t}\frac{\partial p}{\partial T} = \frac{\partial p}{\partial t} + \epsilon \frac{\partial p}{\partial T} \\ &= i\frac{\partial \theta}{\partial t} p + \epsilon \frac{\partial A}{\partial T} e^{i\theta} = -i\omega p + \epsilon \frac{\partial A}{\partial T} e^{i\theta} \end{aligned} \tag{2.32}$$

using the definition of ω, and, up to first order in ϵ only,

$$\begin{aligned} \frac{\partial^2 p}{\partial t^2} &= \frac{\partial}{\partial t}\left[-i\omega p + \epsilon \frac{\partial A}{\partial T} e^{i\theta}\right] \\ &= -i\frac{\partial \omega}{\partial t} p - i\omega \frac{\partial p}{\partial t} + i\epsilon \frac{\partial \theta}{\partial t}\frac{\partial A}{\partial T} e^{i\theta} \\ &= -\omega^2 p - 2i\epsilon\omega \frac{\partial A}{\partial T} e^{i\theta} - i\epsilon \frac{\partial \omega}{\partial T} p \end{aligned} \tag{2.33}$$

Similarly, we have

$$\frac{\partial p}{\partial x} = ikp + \epsilon \frac{\partial A}{\partial X} e^{i\theta} \tag{2.34}$$

and, up to first order in ϵ only,

$$\frac{\partial^2 p}{\partial x^2} = -k^2 p + 2ik\epsilon \frac{\partial A}{\partial X} e^{i\theta} + i\epsilon \frac{\partial k}{\partial X} p \tag{2.35}$$

Plugging these back in the wave equation, and equating orders, we get:

- To lowest order in ϵ we recover the dispersion relation for sound waves,

$$\omega^2 = k^2 c^2 \tag{2.36}$$

- To the next order, we have:

$$\frac{\partial A}{\partial T} + c^2 \frac{k}{\omega} \frac{\partial A}{\partial X} = -\frac{A}{2\omega} \left[\frac{\partial \omega}{\partial T} + c^2 \frac{\partial k}{\partial X} \right] \tag{2.37}$$

In order words, the evolution of the wave packet can be studied by solving the coupled system of equations

$$\begin{aligned} \omega^2 &= c^2 k^2 \\ \frac{\partial k}{\partial T} + \frac{\partial \omega}{\partial X} &= 0 \\ \frac{\partial A}{\partial T} + c^2 \frac{k}{\omega} \frac{\partial A}{\partial X} &= -\frac{A}{2\omega} \left[\frac{\partial \omega}{\partial T} + c^2 \frac{\partial k}{\partial X} \right] \end{aligned} \tag{2.38}$$

all of which only depend on the slow-variables X and T. In essence, we have filtered out all of the rapid oscillatory behavior of the waves, keeping only the more manageable slow variations! Solving these new equations is often a lot easier than solving the primitive ones.

However, we can do even better. There is another way of re-writing these equations that leads to a much more intuitive interpretation of their solutions. Let's first consider the evolution of ω. Taking the slow-time derivative of the dispersion relation, we have

$$2\omega \frac{\partial \omega}{\partial T} = 2c^2 k \frac{\partial k}{\partial T} = -2c^2 k \frac{\partial \omega}{\partial X} \tag{2.39}$$

using (2.31). We can therefore re-write this as

$$\frac{\partial \omega}{\partial T} + \frac{c^2}{k} \omega \frac{\partial \omega}{\partial X} = \frac{\partial \omega}{\partial T} \pm c \frac{\partial \omega}{\partial X} = 0 \tag{2.40}$$

depending on the branch ($\pm$) of the dispersion relation selected. In other words, the frequency function is advected at velocity $\pm c$ without change of form.

Next, using the spatial derivative of the dispersion relation, we have

$$\frac{\partial k}{\partial T} + \frac{\partial \omega}{\partial X} = \frac{\partial k}{\partial T} + \frac{c^2 k}{\omega} \frac{\partial k}{\partial X} = 0 \tag{2.41}$$

which becomes

$$\frac{\partial k}{\partial T} \pm c \frac{\partial k}{\partial X} = 0 \tag{2.42}$$

which again implies that the wavenumber is advected at velocity $\pm c$ without change of form.

Finally, combining (2.40) and (2.42) with (2.38), we find that the amplitude equation also simplifies, in such a way that

$$\frac{\partial A}{\partial T} \pm c \frac{\partial A}{\partial X} = 0 \tag{2.43}$$

To summarize, an alternative way of looking at the evolution of the wave packet is to solve simultaneously the much more intuitive set of equations

$$\begin{aligned} \frac{\partial \omega}{\partial T} \pm c\frac{\partial \omega}{\partial X} &= 0 \\ \frac{\partial k}{\partial T} \pm c\frac{\partial k}{\partial X} &= 0 \\ \frac{\partial A}{\partial T} \pm c\frac{\partial A}{\partial X} &= 0 \end{aligned} \tag{2.44}$$

where the choice of $\pm$ simply depends on the branch of the dispersion relation we have chosen ($\omega = \pm ck$). Note that the first two equations are equivalent, so one only needs to solve one or the other. The solutions to these equations are simply $A(X,T) = A(X \mp cT)$, $\omega(X,T) = \omega(X \mp cT)$ and $k(X,T) = k(X \mp cT)$.

As written we see that all the properties of the wave packet are advected without change of form at the same velocity, $\pm c$. This velocity is the *group velocity* discussed earlier, and describes the propagation of the packet rather than the phase within the packet. Note how much of the phase information is lost from the wave packet description: this is the approximation made and the price to pay for using this method.

Worked example

What is the exact solution of the right-ward propagating wave for the initial condition given by $p(x,0) = \cos(x)\exp(-x^2/200)$ with $p_t(x,0) = 0$? What is the approximate wave-packet solution? Compare the two.

Finding the exact solution is pretty trivial since we can just use the right-ward propagating component of d'Alembert's solution for instance:

$$p(x,t) = p_0(x - ct) = \cos(x - ct)\exp(-(x - ct)^2/200) \tag{2.45}$$

In the wave-packet solution, we have to identify the carrier wave and the slow amplitude from the initial conditions. In the way it is written, separating the two is fairly obvious: the carrier wave at time $t = 0$ is $\cos(x) = \Re(e^{i\theta(x,0)})$, which has a constant wavenumber, while the slowly varying amplitude function is $\exp(-x^2/200)$. To see this, simply let $X = x/10$, in which case

$$\exp(-x^2/200) = \exp(-X^2/2) = A(X,0) \tag{2.46}$$

In the wave-packet approximation, the both the wavenumber function $k(X,T)$ and the amplitude function $A(X,T)$ are advected with velocity c to the right. Since the wavenumber function is constant at $t = 0$, then this will not change with time since c is constant. The time-dependent carrier wave is then $\Re(e^{i\theta(x,t)}) = \cos(x - ct)$ since ω and k are constant. Finally, the advection of the amplitude function gives $\exp(-(X - cT)^2/2) = \exp(-(x - ct)^2/200)$. As a result, the approximate solution of the wave packet equations is

$$p(x,t) = \cos(x - ct)\exp(-(x - ct)^2/200) \tag{2.47}$$

which is exacly the same as d'Alembert's solution.

This is a fairly simplistic problem, however. In general, the two will not be exactly equal when k (and ω) are not constant, and/or when the sound speed is not constant.

2.4.2 Generalization of the wave packet to multiple dimensions

As we now see, the concept of a wave packet is trivially generalized to multiple dimensions, which makes it a very useful tool! In more than 1D, the wave equation as derived earlier becomes

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{2.48}$$

where we are still assuming that c is constant, and plane wave solutions are of the kind

$$p(\boldsymbol{x}, t) = \hat{p} \exp(i\boldsymbol{k} \cdot \boldsymbol{x} - i\omega t) \tag{2.49}$$

where $\boldsymbol{x} = (x, y, z)$ and $\boldsymbol{k} = (k_x, k_y, k_z)$ are now three-dimensional.

We create a wave packet exactly the same way as before, assuming that p can be written as the plane monochromatic wave times the slowly varying amplitude:

$$p(\boldsymbol{x}, t) = A(\boldsymbol{X}, T) \exp\left(i\theta(\boldsymbol{x}, t)\right) \tag{2.50}$$

where $\boldsymbol{X} = (X, Y, Z) = (\epsilon x, \epsilon y, \epsilon z)$ is a three-dimensional vector and θ is the phase function. For the same reasons as discussed in the 1D case, we can define the local frequency and wave*vector* of the wave to be

$$\omega = -\frac{\partial \theta}{\partial t} \text{ and } \boldsymbol{k} = \nabla \theta \tag{2.51}$$

or in other words

$$k_x = \frac{\partial \theta}{\partial x} \text{ , } k_y = \frac{\partial \theta}{\partial y} \text{ and } k_z = \frac{\partial \theta}{\partial z} \tag{2.52}$$

This then implies that

$$\frac{\partial \boldsymbol{k}}{\partial t} + \nabla \omega = 0 \tag{2.53}$$

and therefore that

$$\frac{\partial \boldsymbol{k}}{\partial T} + \nabla_\epsilon \omega = 0 \tag{2.54}$$

where ∇_ϵ means that the spatial operator only acts on the slow position variables.

Plugging the wave packet solution into the wave equation, and proceeding exactly as before, we now find that

$$\begin{aligned}\omega^2 &= c^2|\boldsymbol{k}|^2 \\ \frac{\partial\omega}{\partial T} + \frac{c^2}{\omega}\boldsymbol{k}\cdot\nabla_\epsilon\omega &= \frac{\partial\omega}{\partial T} + \boldsymbol{c}_g\cdot\nabla_\epsilon\omega = 0 \\ \frac{\partial\boldsymbol{k}}{\partial T} + \frac{c^2}{\omega}\boldsymbol{k}\cdot\nabla_\epsilon\boldsymbol{k} &= \frac{\partial\boldsymbol{k}}{\partial T} + \boldsymbol{c}_g\cdot\nabla_\epsilon\boldsymbol{k} = 0 \end{aligned} \tag{2.55}$$

where $\boldsymbol{c}_g = \frac{c^2}{\omega}\boldsymbol{k} = c\boldsymbol{k}/k$. This shows that the group velocity $\boldsymbol{c}_g$ is in the direction of $\boldsymbol{k}$, and since $\boldsymbol{k} = \nabla\theta$, we see that the waves travel in a direction that is perpendicular to lines of constant phase. While this seems to be pretty obvious for compression waves, we will see that it is also not always the case – some dispersive waves have group velocities that are *not* necessarily perpendicular to their constant-phase surfaces. Figure 2.6 shows examples of constant phase surfaces in various types of configurations and corresponding selected wave-vectors, for non-dispersive waves.

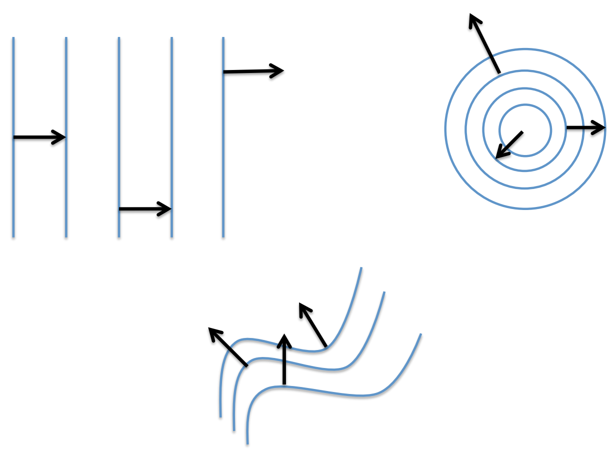

Figure 2.6: Lines of constant θ for sample 2D pressure wave fields, and selected wavenumbers. The wavenumbers are always perpendicular to the lines of constant θ, and the group velocity is parallel to $\boldsymbol{k}$. Hence the wave is propagating to the right in the first case and radially outward in the second case. In the third case it's a little more complicated.

Finally, the amplitude equation becomes

$$\frac{\partial A}{\partial T} + \frac{c^2}{\omega}\boldsymbol{k}\cdot\nabla_\epsilon A = -\frac{A}{2ck}\left[\frac{\partial\omega}{\partial T} + c^2\nabla_\epsilon\cdot\boldsymbol{k}\right] \tag{2.56}$$

This can be rewritten in a clearer form as:

$$\frac{\partial A}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon A = -\frac{A}{2ck}\left[\frac{\partial \omega}{\partial T} + \nabla_\epsilon \cdot (\omega \boldsymbol{c}_g)\right] \tag{2.57}$$

Expanding the divergence, and using (2.55), we then get

$$\frac{\partial A}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon A = -\frac{Ac}{2}\nabla_\epsilon \cdot \left(\frac{\boldsymbol{k}}{k}\right) \tag{2.58}$$

To summarize, in 3D the wave packet equations are:

$$\begin{aligned} \frac{\partial \boldsymbol{k}}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon \boldsymbol{k} &= 0 \\ \frac{\partial \omega}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon \omega &= 0 \\ \frac{\partial A}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon A &= -\frac{Ac}{2}\nabla_\epsilon \cdot \left(\frac{\boldsymbol{k}}{k}\right) \end{aligned} \tag{2.59}$$

This shows that $\boldsymbol{k}$ and ω are advected without change of form or ampiltude by the velocity field $\boldsymbol{c}_g$. The equation for the amplitude function A, on the other hand, now has a non-zero RHS (the fact that we had a zero RHS in the 1D case simply stems from the fact that $\nabla_\epsilon \cdot (\boldsymbol{k}/k) \equiv 0$ in 1D). The physical interpretation of this RHS is that the convergence or divergence of the wavenumber field can focus or de-focus the waves. In that case the total amplitude increases or decreases correspondingly (see the example below of the spherical loudspeaker for instance).

2.4.3 Ray Tracing

The set of 3 equations for the evolution of the frequency, wavevector and wave packet amplitude given in (2.55) and (2.58) shows that all three quantities evolve on the same characteristics (see Method of Characteristics, AMS 212A), which are called the *ray paths*. To find the equations for these ray paths, we look at the evolution of $\boldsymbol{k}$ (we cannot start with ω and A since these equations depend on $\boldsymbol{k}$ via $\boldsymbol{c}_g$). Component by component, we have that

$$\frac{\partial k_i}{\partial T} + \boldsymbol{c}_g \cdot \nabla_\epsilon k_i = 0 \tag{2.60}$$

where k_i is either k_x, k_y or k_z. Using the method of characteristics, we then have

$$\begin{aligned} \frac{\partial T}{\partial \tau} &= 1 \\ \frac{\partial X}{\partial \tau} = (\boldsymbol{c}_g)_x = c\frac{k_x}{k} \ , \ \frac{\partial Y}{\partial \tau} = (\boldsymbol{c}_g)_y = c\frac{k_y}{k} \ , \ \frac{\partial Z}{\partial \tau} = (\boldsymbol{c}_g)_z &= c\frac{k_z}{k} \\ \frac{\partial k_i}{\partial \tau} &= 0 \end{aligned} \tag{2.61}$$

where τ is the "time" variable along a characteristic (not to be mixed up with the τ used in Chapter 1).

We first see that k_i is conserved along a ray path. Since all the components of $\boldsymbol{k}$ are, then so is $\boldsymbol{k}$. This implies that the right-hand-sides of all these characteristic equations are constant, so that the ratios $\partial X/\partial Y = k_x/k_y$, $\partial X/\partial Z = k_x/k_z$ and $\partial Y/\partial Z = k_y/k_z$ along a ray path are all constant – in other words, *the ray path is a straight line until it hits a boundary (see below).* Furthermore, it is easy to show that the ray path is parallel to $\boldsymbol{k}$ (or equivalently, to $\boldsymbol{c}_g$), so that its direction is given by the value $\boldsymbol{k}$ has at time $t = 0$.

Next, by analogy, we see that ω is also constant along a ray path, (since it has the same characteristics, so $\partial\omega/\partial\tau = 0$). Finally, we have that $\partial A/\partial\tau = -(Ac/2)\nabla \cdot (\boldsymbol{k}/k)$, which implies that the amplitude of a sound wave increases if the ray paths converge, and decreases if the ray paths diverge. We will revisit the amplitude equation shortly, but in the meantime, note how the distribution of $\boldsymbol{k}$ at $t = 0$ entirely determines the ray paths, which then entirely determines the full solution $\boldsymbol{k}$, ω and A everywhere along them!

Worked example: the spherical loudspeaker

Consider sound waves being generated by a perfectly spherical loudspeaker of radius R_0 vibrating radially. The sound waves generated have a given constant frequency ω. Suppose the speaker at $r = R_0$ emits a wave packet whose amplitude is a Gaussian function of the slow time T, e.g. $p(R_0, t) = \cos(\omega t)\exp(-T^2/2)$. What is the solution $p(R, t)$ far from the speaker?

To begin with, we must figure out the ray paths of the waves. Being only given the frequency as a function of time at $r = R_0$, we don't a priori know what the wavenumber field $\boldsymbol{k}$ is. However, we know that $\boldsymbol{k}$ must be perpendicular to the phase surfaces, and since the loudspeaker is vibrating radially, it is creating sound waves that only depend on r. Hence the phase θ varies only with r, so $\boldsymbol{k}$ has to be perpendicular to surfaces of constant r. This shows that $\boldsymbol{k}$ must be radial: $\boldsymbol{k} = k\boldsymbol{e}_r$ where $k = \omega/c$.

Based on ray tracing, we know from the initial conditions selected that the rays are straight lines, and so they remain purely radial throughout space. We also know that ω and k are conserved along a ray: hence $\boldsymbol{k} = k\boldsymbol{e}_r = (\omega/c)\boldsymbol{e}_r$ everywhere in space. This implies

$$\frac{\partial A}{\partial T} + c\boldsymbol{e}_r \cdot \nabla_\epsilon A = -\frac{Ac}{2}\nabla_\epsilon \cdot \boldsymbol{e}_r \tag{2.62}$$

By spherical symmetry, we also expect that the amplitude will only depend on the radius R away from the center of the sphere, and time. It is therefore more appropriate to study this equation in a spherical coordinate system than in the Cartesian one used until now. Note how the advantage of having cast the ray path equations in vector form is that they are valid in any coordinate system.

We now merely need to re-express them in a spherical coordinate system:

$$\frac{\partial A}{\partial T} + c\frac{\partial A}{\partial R} = -\frac{Ac}{R} \tag{2.63}$$

This equation can be solved using the method of characteristics.

We first have to create the "initial condition curve". We have that, on $R = R_0$ (the radius of the loudspeaker), $A(R_0, T) = A_p(T)$ where $A_p(T) = \exp(-T^2/2)$ is the slow-time variation of the sound pulse. This can be parametrized as $R_0(s) = R_0$, $T_0(s) = s$, and $A_0(s) = A_p(s)$. The characteristic equations are

$$\frac{\partial T}{\partial \tau} = 1 \ , \ \frac{\partial R}{\partial \tau} = c \ , \ \frac{\partial A}{\partial \tau} = -\frac{Ac}{R} \tag{2.64}$$

The first of these equations has solution $T = \tau + T_0(s) = \tau + s$. The second has solution $R = c\tau + R_0(s) = c\tau + R_0$. The last equation can then be cast in terms of τ only as

$$\frac{\partial A}{\partial \tau} = -\frac{A}{(R_0/c) + \tau} \tag{2.65}$$

This implies that

$$\ln A = -\ln((R_0/c) + \tau) + K(s) \tag{2.66}$$

where $K(s)$ is an integration function. To satisfy the initial conditions, we have to have

$$K = \ln A_0(s) + \ln((R_0/c)) \tag{2.67}$$

so that

$$A(s, \tau) = \frac{A_p(s)}{1 + \tau c/R_0} = \frac{e^{-s^2/2}}{1 + \tau c/R_0} \tag{2.68}$$

To transform this solution back into (R, T) space, we have to write s and τ in terms of R and T. We have

$$\tau = \frac{R - R_0}{c} \text{ and } s = T - \tau = T - \frac{R - R_0}{c} \tag{2.69}$$

so

$$A(R, T) = \frac{R_0}{R} A_p \left(T - \frac{R - R_0}{c} \right) = \frac{R_0}{R} \exp \left[\frac{1}{2} \left(T - \frac{R - R_0}{c} \right)^2 \right] \tag{2.70}$$

With see that the Gaussian pulse propagates radially with velocity c without change of width, but its amplitude decreases away from the speaker as $A \propto R^{-1}$. This is shown in Figure 2.7.

Note also that by suitably taking the limit of an infinitely short pulse, and an infinitely small sphere, we are not far from getting the Green's function solution for sound waves in an infinite homogeneous domain, which can then be used to reconstruct solutions for any distribution of sound-sources and any initial condition (see AMS212A for detail).

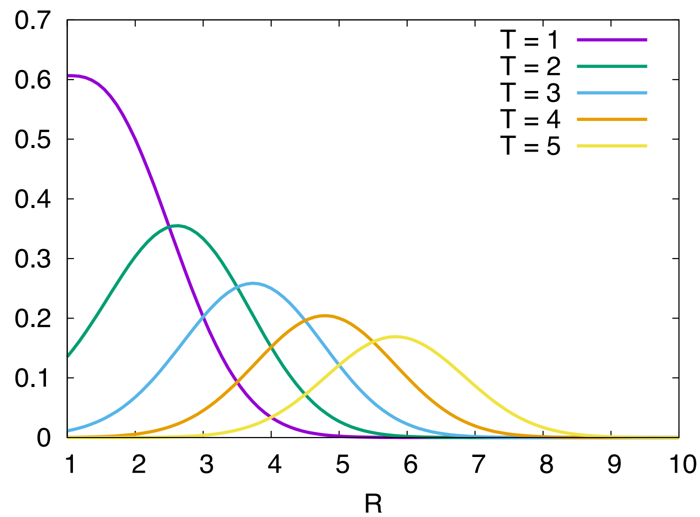

Figure 2.7: Function $A(R,T)$ of the amplitude of the sound emitted by a spherical loudspeaker of radius $R_0 = 1$, assuming $c = 1$.

2.4.4 Reflection near a wall

The only problem left is to address what happens when a ray hits a wall. To do so, let's look close to the wall, in a small region where the wave is well-approximated by a plane wave. We use the same method as we did in 1D, looking at the solution near the wall as the sum of an incident and a reflected wave. Let $\boldsymbol{k}^I$ and ω_I be the wavenumber and frequency of the incident wave, and $\boldsymbol{k}^R$ and ω_R those of the reflected wave. Suppose the wall is at $x = 0$. We have

$$p_I(\boldsymbol{x},t) = A_I e^{i\boldsymbol{k}^I\cdot\boldsymbol{x}-i\omega_I t} \text{ and } p_R(\boldsymbol{x},t) = A_R e^{i\boldsymbol{k}^R\cdot\boldsymbol{x}-i\omega_R t} \tag{2.71}$$

There are two possible cases. Those with boundary conditions $\hat{\boldsymbol{n}}\cdot\nabla p = 0$ (where the derivative of the pressure perpendicular to the wall must be 0) or $p = 0$, where the pressure itself must be 0 at the wall. Here we will look at the $p = 0$ case, and the other case is left as homework.

If, at the wall ($x = 0$), we require that $p = 0$ then

$$p(\boldsymbol{x},t) = p_I(0,y,z,t)+p_R(0,y,z,,t) = A_I e^{ik_y^I y+ik_z^I z-i\omega_I t}+A_R e^{ik_y^R y+ik_z^R z-i\omega_R t} = 0 \tag{2.72}$$

The only way to enforce this for any y, z, and t is to have ω, k_y and k_z be the same for the incident and reflected waves, and $A_R = -A_I$. The first condition implies that the modulus of $\boldsymbol{k}$ must be invariant (since ω is, and ω only depends on the modulus of $\boldsymbol{k}$). This in turn implies that $k_x^R = -k_x^I$. The second condition can be recast as a change of phase by a factor π, since $-1 = e^{i\pi}$. In summary, we have that

$$p_R(\boldsymbol{x},t) = A_I e^{-ik_x^I x+ik_y^I y+ik_z^I z-i\omega_I t-i\pi} \tag{2.73}$$

The ray path of the reflected ray is shown in Figure 2.8

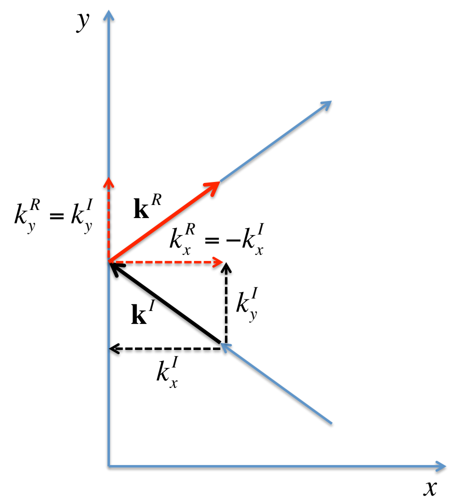

Figure 2.8: Incident and reflected ray paths near a wall at $x = 0$, in the case where $p = 0$ on the wall.

This process can be easily generalized to other geometries, at least when the boundary is smooth, to show that (1) the frequency and amplitude remains unchanged, (2) the component of $\boldsymbol{k}$ parallel to the boundary remains unchanged, (3) the component of $\boldsymbol{k}$ perpendicular to the boundary changes sign and (4) the phase is shifted by a factor of π. What happens at corners is a lot harder, and will be ignored here.

Many examples of application of ray tracing exist, and it is one of the fundamental tools of the theory for acoustic design. Interesting ones involve, for instance, wave guides and sound focussing designs. Also note that it is possible to derive quantization conditions from ray tracing in multiple dimension in a manner analogous to what we did in 1D, to recover the global eigenmodes/eigenfrequencies of oscillation of an acoustic cavity. This is one the techniques used to determine the frequencies of oscillations of stars, for instance. This field holds many interesting mathematical tricks/theorems, some quite fundamental such as the Einstein-Brillouin-Keller quantization, which is equally useful for studying pressure waves in stars *and* in quantum mechanics to calculate energy levels in atoms/molecules! It has also opened the door to another concept called "quantum chaos", first glimpsed in the context of ray tracing by Einstein himself. Finally, remember that all we have done so far is valid, not just for sound waves, but for all non-dispersive waves, such as electromagnetic waves (i.e. light) for example.