Chapter 2

Normal modes

David Morin, morin@physics.harvard.edu

In Chapter 1 we dealt with the oscillations of one mass. We saw that there were various possible motions, depending on what was influencing the mass (spring, damping, driving forces). In this chapter we'll look at oscillations (generally without damping or driving) involving more than one object. Roughly speaking, our counting of the number of masses will proceed as: two, then three, then infinity. The infinite case is relevant to a continuous system, because such a system contains (ignoring the atomic nature of matter) an infinite number of infinitesimally small pieces. This is therefore the chapter in which we will make the transition from the oscillations of one particle to the oscillations of a continuous object, that is, to waves.

The outline of this chapter is as follows. In Section 2.1 we solve the problem of two masses connected by springs to each other and to two walls. We will solve this in two ways – a quick way and then a longer but more fail-safe way. We encounter the important concepts of *normal modes* and *normal coordinates*. We then add on driving and damping forces and apply some results from Chapter 1. In Section 2.2 we move up a step and solve the analogous problem involving three masses. In Section 2.3 we solve the general problem involving N masses and show that the results reduce properly to the ones we already obtained in the $N = 2$ and $N = 3$ cases. In Section 2.4 we take the $N \to \infty$ limit (which corresponds to a continuous stretchable material) and derive the all-important *wave equation*. We then discuss what the possible waves can look like.

2.1 Two masses

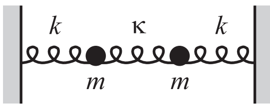

Figure 1

For a single mass on a spring, there is one natural frequency, namely $\sqrt{k/m}$. (We'll consider undamped and undriven motion for now.) Let's see what happens if we have two equal masses and three spring arranged as shown in Fig. 1. The two outside spring constants are the same, but we'll allow the middle one to be different. In general, all three spring constants could be different, but the math gets messy in that case.

Let x_1 and x_2 measure the displacements of the left and right masses from their respective equilibrium positions. We can assume that all of the springs are unstretched at equilibrium, but we don't actually have to, because the spring force is linear (see Problem [to be added]). The middle spring is stretched (or compressed) by $x_2 - x_1$, so the $F = ma$ equations on the two masses are

$$\begin{aligned} m\ddot{x}_1 &= -kx_1 - \kappa(x_1 - x_2), \\ m\ddot{x}_2 &= -kx_2 - \kappa(x_2 - x_1). \end{aligned} \tag{1}$$

Concerning the signs of the κ terms here, they are equal and opposite, as dictated by Newton's third law, so they are either both right or both wrong. They are indeed both right, as can be seen by taking the limit of, say, large x_2. The force on the left mass is then in the positive direction, which is correct.

These two $F = ma$ equations are "coupled," in the sense that both x_1 and x_2 appear in both equations. How do we go about solving for $x_1(t)$ and $x_2(t)$? There are (at least) two ways we can do this.

2.1.1 First method

This first method is quick, but it works only for simple systems with a sufficient amount of symmetry. The main goal in this method is to combine the $F = ma$ equations in well-chosen ways so that x_1 and x_2 appear only in certain unique combinations. It sometimes involves a bit of guesswork to determine what these well-chosen ways are. But in the present problem, the simplest thing to do is add the $F = ma$ equations in Eq. (1), and it turns out that this is in fact one of the two useful combinations to form. The sum yields

$$m(\ddot{x}_1 + \ddot{x}_2) = -k(x_1 + x_2) \implies \frac{d^2}{dt^2}(x_1 + x_2) = -\frac{k}{m}(x_1 + x_2). \tag{2}$$

The variables x_1 and x_2 appear here only in the unique combination, $x_1 + x_2$. And furthermore, this equation is simply a harmonic-motion equation for the quantity $x_1 + x_2$. The solution is therefore

$$x_1(t) + x_2(t) = 2A_s \cos(\omega_s t + \phi_s), \qquad \text{where} \quad \boxed{\omega_s \equiv \sqrt{\frac{k}{m}}} \tag{3}$$

The "s" here stands for "slow," to be distinguished from the "fast" frequency we'll find below. And we've defined the coefficient to be $2A_s$ so that we won't have a bunch of factors of $1/2$ in our final answer in Eq. (6) below.

No matter what complicated motion the masses are doing, the quantity $x_1 + x_2$ always undergoes simple harmonic motion with frequency ω_s. This is by no means obvious if you look at two masses bouncing back and forth in an arbitrary manner.

The other useful combination of the $F = ma$ equations is their difference, which conveniently is probably the next thing you might try. This yields

$$m(\ddot{x}_1 - \ddot{x}_2) = -(k + 2\kappa)(x_1 - x_2) \implies \frac{d^2}{dt^2}(x_1 - x_2) = -\frac{k + 2\kappa}{m}(x_1 - x_2). \tag{4}$$

The variables x_1 and x_2 now appear only in the unique combination, $x_1 - x_2$. And again, we have a harmonic-motion equation for the quantity $x_1 - x_2$. So the solution is (the "f" stands for "fast")

$$x_1(t) - x_2(t) = 2A_f \cos(\omega_f t + \phi_f), \qquad \text{where} \quad \boxed{\omega_f \equiv \sqrt{\frac{k + 2\kappa}{m}}} \tag{5}$$

As above, no matter what complicated motion the masses are doing, the quantity $x_1 - x_2$ always undergoes simple harmonic motion with frequency ω_f.

We can now solve for $x_1(t)$ and $x_2(t)$ by adding and subtracting Eqs. (3) and (5). The result is

$$\begin{aligned} x_1(t) &= A_{\rm s}\cos(\omega_{\rm s}t+\phi_{\rm s}) + A_{\rm f}\cos(\omega_{\rm f}t+\phi_{\rm f}), \\ x_2(t) &= A_{\rm s}\cos(\omega_{\rm s}t+\phi_{\rm s}) - A_{\rm f}\cos(\omega_{\rm f}t+\phi_{\rm f}). \end{aligned} \tag{6}$$

The four constants, $A_{\rm s}$, $A_{\rm f}$, $\phi_{\rm s}$, $\phi_{\rm f}$ are determined by the four initial conditions, $x_1(0)$, $x_2(0)$, $\dot{x}_1(0)$, $\dot{x}_1(0)$.

The above method will clearly work only if we're able to guess the proper combinations of the $F = ma$ equations that yield equations involving unique combinations of the variables. Adding and subtracting the equations worked fine here, but for more complicated systems with unequal masses or with all the spring constants different, the appropriate combination of the equations might be far from obvious. And there is no guarantee that guessing around will get you anywhere. So before discussing the features of the solution in Eq. (6), let's take a look at the other more systematic and fail-safe method of solving for x_1 and x_2.

2.1.2 Second method

This method is longer, but it works (in theory) for any setup. Our strategy will be to look for simple kinds of motions where both masses move with the same frequency. We will then build up the most general solution from these simple motions. For all we know, such motions might not even exist, but we have nothing to lose by trying to find them. We will find that they do in fact exist. You might want to try to guess now what they are for our two-mass system, but it isn't necessary to know what they look like before undertaking this method.

Let's guess solutions of the form $x_1(t) = A_1e^{i\omega t}$ and $x_2(t) = A_2e^{i\omega t}$. For bookkeeping purposes, it is convenient to write these solutions in vector form:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} e^{i\omega t}. \tag{7}$$

We'll end up taking the real part in the end. We can alternatively guess the solution $e^{\alpha t}$ without the i, but then our α will come out to be imaginary. Either choice will get the job done. Plugging these guesses into the $F = ma$ equations in Eq. (1), and canceling the factor of $e^{i\omega t}$, yields

$$\begin{aligned} -m\omega^2A_1 &= -kA_1 - \kappa(A_1 - A_2), \\ -m\omega^2A_2 &= -kA_2 - \kappa(A_2 - A_1). \end{aligned} \tag{8}$$

In matrix form, this can be written as

$$\begin{pmatrix} -m\omega^2 + k + \kappa & -\kappa \\ -\kappa & -m\omega^2 + k + \kappa \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{9}$$

At this point, it seems like we can multiply both sides of this equation by the inverse of the matrix. This leads to $(A_1, A_2) = (0, 0)$. This is obviously a solution (the masses just sit there), but we're looking for a nontrivial solution that actually contains some motion. The only way to escape the preceding conclusion that A_1 and A_2 must both be zero is if the inverse of the matrix doesn't exist. Now, matrix inverses are somewhat messy things (involving cofactors and determinants), but for the present purposes, the only fact we need to know about them is that they involve dividing by the determinant. So if the determinant is zero, then the inverse doesn't exist. This is therefore what we want. Setting the determinant equal to zero gives the quartic equation,

$$\begin{vmatrix} -m\omega^2 + k + \kappa & -\kappa \\ -\kappa & -m\omega^2 + k + \kappa \end{vmatrix} = 0 \implies (-m\omega^2 + k + \kappa)^2 - \kappa^2 = 0$$
$$\implies -m\omega^2 + k + \kappa = \pm\kappa$$
$$\implies \omega^2 = \frac{k}{m} \quad \text{or} \quad \frac{k + 2\kappa}{m}. \tag{10}$$

The four solutions to the quartic equation are therefore $\omega = \pm\sqrt{k/m}$ and $\omega = \pm\sqrt{(k + 2\kappa)/m}$.

For the case where $\omega^2 = k/m$, we can plug this value of ω^2 back into Eq. (9) to obtain

$$\kappa \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{11}$$

Both rows of this equation yield the same result (this was the point of setting the determinant equal to zero), namely $A_1 = A_2$. So (A_1, A_2) is proportional to the vector $(1, 1)$.

For the case where $\omega^2 = (k + 2\kappa)/m$, Eq. (9) gives

$$\kappa \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{12}$$

Both rows now yield $A_1 = -A_2$. So (A_1, A_2) is proportional to the vector $(1, -1)$.

With $\omega_s \equiv \sqrt{k/m}$ and $\omega_f \equiv \sqrt{(k + 2\kappa)/m}$, we can write the general solution as the sum of the four solutions we have found. In vector notation, $x_1(t)$ and $x_2(t)$ are given by

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{i\omega_s t} + C_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-i\omega_s t} + C_3 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{i\omega_f t} + C_4 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-i\omega_f t}. \tag{13}$$

We now perform the usual step of invoking the fact that the positions $x_1(t)$ and $x_2(t)$ must be real for all t. This yields that standard result that $C_1 = C_2^* \equiv (A_s/2)e^{i\phi_s}$ and $C_3 = C_4^* \equiv (A_f/2)e^{i\phi_f}$. We have included the factors of 1/2 in these definitions so that we won't have a bunch of factors of 1/2 in our final answer. The imaginary parts in Eq. (13) cancel, and we obtain

$$\boxed{\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = A_s \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cos(\omega_s t + \phi_s) + A_f \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cos(\omega_f t + \phi_f)} \tag{14}$$

Therefore,

$$\begin{aligned} x_1(t) &= A_s \cos(\omega_s t + \phi_s) + A_f \cos(\omega_f t + \phi_f), \\ x_2(t) &= A_s \cos(\omega_s t + \phi_s) - A_f \cos(\omega_f t + \phi_f). \end{aligned} \tag{15}$$

This agrees with the result in Eq. (6).

As we discussed in Section 1.1.5, we could have just taken the real part of the $C_1(1, 1)e^{i\omega_s t}$ and $C_3(1, -1)e^{i\omega_f t}$ solutions, instead of going through the "positions must be real" reasoning. However, you should continue using the latter reasoning until you're comfortable with the short cut of taking the real part.

Remark: Note that Eq. (9) can be written in the form,

$$\begin{pmatrix} k + \kappa & -\kappa \\ -\kappa & k + \kappa \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = m\omega^2 \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}. \tag{16}$$

So what we did above was solve for the *eigenvectors* and *eigenvalues* of this matrix. The eigenvectors of a matrix are the special vectors that get carried into a multiple of themselves what acted on by the matrix. And the multiple (which is $m\omega^2$ here) is called the eigenvalue. Such vectors are indeed special, because in general a vector gets both stretched (or shrunk) *and* rotated when acted on by a matrix. Eigenvectors don't get rotated at all. ♣

A third method of solving our coupled-oscillator problem is to solve for x_2 in the first equation in Eq. (1) and plug the result into the second. You will get a big mess of a fourth-order differential equation, but it's solvable by guessing $x_1 = Ae^{i\omega t}$.

2.1.3 Normal modes and normal coordinates

Normal modes

Having solved for $x_1(t)$ and $x_2(t)$ in various ways, let's now look at what we're found. If $A_\mathrm{f} = 0$ in Eq. (15), then we have

$$x_1(t) = x_2(t) = A_\mathrm{s}\cos(\omega_\mathrm{s} t + \phi_\mathrm{s}). \tag{17}$$

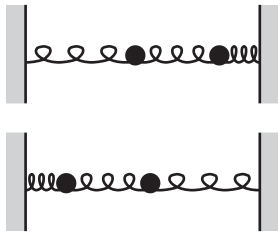

Figure 2

So both masses move in exactly the same manner. Both to the right, then both to the left, and so on. This is shown in Fig. 2. The middle spring is never stretched, so it effectively isn't there. We therefore basically have two copies of a simple spring-mass system. This is consistent with the fact that ω_s equals the standard expression $\sqrt{k/m}$, independent of κ. This nice motion, where both masses move with the same frequency, is called a *normal mode*. To specify what a normal mode looks like, you have to give the frequency and also the relative amplitudes. So this mode has frequency $\sqrt{k/m}$, and the amplitudes are equal.

If, on the other hand, $A_\mathrm{s} = 0$ in Eq. (15), then we have

$$x_1(t) = -x_2(t) = A_\mathrm{f}\cos(\omega_\mathrm{f} t + \phi_\mathrm{f}). \tag{18}$$

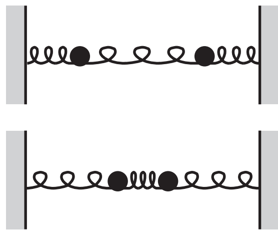

Figure 3

Now the masses move oppositely. Both outward, then both inward, and so on. This is shown in Fig. 3. The frequency is now $\omega_\mathrm{f} = \sqrt{(k + 2\kappa)/m}$. It makes sense that this is larger than ω_s, because the middle spring is now stretched or compressed, so it adds to the restoring force. This nice motion is the other normal mode. It has frequency $\sqrt{(k + 2\kappa)/m}$, and the amplitudes are equal and opposite. The task of Problem [to be added] is to deduce the frequency ω_f in a simpler way, without going through the whole process above.

Eq. (15) tells us that any arbitrary motion of the system can be thought of as a linear combination of these two normal modes. But in the general case where both coefficients A_s and A_f are nonzero, it's rather difficult to tell that the motion is actually built up from these two simple normal-mode motions.

Normal coordinates

By adding and subtracting the expressions for $x_1(t)$ and $x_1(t)$ in Eq. (15), we see that for *any* arbitrary motion of the system, the quantity $x_1 + x_2$ oscillates with frequency ω_s, and the quantity $x_1 - x_2$ oscillates with frequency ω_f. These combinations of the coordinates are known as the *normal coordinates* of the system. They are the nice combinations of the coordinates that we found advantageous to use in the first method above.

The $x_1 + x_2$ normal coordinate is associated with the normal mode $(1, 1)$, because they both have frequency ω_s. Equivalently, any contribution from the other mode (where $x_1 = -x_2$) will vanish in the sum $x_1 + x_2$. Basically, the sum $x_1 + x_2$ picks out the part of the motion with frequency ω_s and discards the part with frequency ω_f. Similarly, the $x_1 - x_2$ normal coordinate is associated with the normal mode $(1, -1)$, because they both have frequency ω_f. Equivalently, any contribution from the other mode (where $x_1 = x_2$) will vanish in the difference $x_1 - x_2$.

Note, however, that the association of the normal coordinate $x_1 + x_2$ with the normal mode $(1, 1)$ does *not* follow from the fact that the coefficients in $x_1 + x_2$ are both 1. Rather, it follows from the fact that the *other* normal mode, namely $(x_1, x_2) \propto (1, -1)$, gives no contribution to the sum $x_1 + x_2$. There are a few too many 1's floating around in the present example, so it's hard to see which results are meaningful and which results are coincidence. But the following example should clear things up. Let's say we solved a problem using the determinant method, and we found the solution to be

$$\begin{pmatrix} x \\ y \end{pmatrix} = B_1 \begin{pmatrix} 3 \\ 2 \end{pmatrix} \cos(\omega_1 t + \phi_1) + B_2 \begin{pmatrix} 1 \\ -5 \end{pmatrix} \cos(\omega_2 t + \phi_2). \tag{19}$$

Then $5x + y$ is the normal coordinate associated with the normal mode $(3, 2)$, which has frequency ω_1. This is true because there is no $\cos(\omega_2 t + \phi_2)$ dependence in the quantity $5x + y$. And similarly, $2x - 3y$ is the normal coordinate associated with the normal mode $(1, -5)$, which has frequency ω_2, because there is no $\cos(\omega_1 t + \phi_1)$ dependence in the quantity $2x - 3y$.

2.1.4 Beats

Let's now apply some initial conditions to the solution in Eq. (15). We'll take the initial conditions to be $\dot{x}_1(0) = \dot{x}_2(0) = 0$, $x_1(0) = 0$, and $x_2(0) = A$. In other words, we're pulling the right mass to the right, and then releasing both masses from rest. It's easier to apply these conditions if we write the solutions for $x_1(t)$ and $x_2(t)$ in the form,

$$\begin{aligned} x_1(t) &= a \cos \omega_s t + b \sin \omega_s t + c \cos \omega_f t + d \sin \omega_f t \\ x_2(t) &= a \cos \omega_s t + b \sin \omega_s t - c \cos \omega_f t - d \sin \omega_f t. \end{aligned} \tag{20}$$

This form of the solution is obtained by using the trig sum formulas to expand the sines and cosines in Eq. (15). The coefficients a, b, c, d are related to the constants A_s, A_f, ϕ_s, ϕ_f. For example, the cosine sum formula gives $a = A_s \cos \phi_s$. If we now apply the initial conditions to Eq. (20), the velocities $\dot{x}_1(0) = \dot{x}_1(0) = 0$ quickly give $b = d = 0$. And the positions $x_1(0) = 0$ and $x_2(0) = A$ give $a = -c = A/2$. So we have

$$\begin{aligned} x_1(t) &= \frac{A}{2}\big(\cos \omega_s t - \cos \omega_f t\big), \\ x_2(t) &= \frac{A}{2}\big(\cos \omega_s t + \cos \omega_f t\big). \end{aligned} \tag{21}$$

For arbitrary values of ω_s and ω_f, this generally looks like fairly random motion, but let's look at a special case. If $\kappa \ll k$, then the ω_f in Eq. (5) is only slightly larger than the ω_s in Eq. (3), so something interesting happens. For frequencies that are very close to each other, it's a standard technique (for reasons that will become clear) to write ω_s and ω_f in terms of their average and (half) difference:

$$\begin{aligned} \omega_s &= \frac{\omega_f + \omega_s}{2} - \frac{\omega_f - \omega_s}{2} \equiv \Omega - \epsilon, \\ \omega_f &= \frac{\omega_f + \omega_s}{2} + \frac{\omega_f - \omega_s}{2} \equiv \Omega + \epsilon, \end{aligned} \tag{22}$$

where

$$\Omega \equiv \frac{\omega_f + \omega_s}{2}, \qquad \text{and} \qquad \epsilon \equiv \frac{\omega_f - \omega_s}{2}. \tag{23}$$

Using the identity $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$, Eq. (21) becomes

$$\begin{aligned} x_1(t) &= \frac{A}{2}\Big(\cos(\Omega-\epsilon)t - \cos(\Omega+\epsilon)t\Big) = A\sin\Omega t\sin\epsilon t, \\ x_2(t) &= \frac{A}{2}\Big(\cos(\Omega-\epsilon)t + \cos(\Omega+\epsilon)t\Big) = A\cos\Omega t\cos\epsilon t. \end{aligned} \quad (24)$$

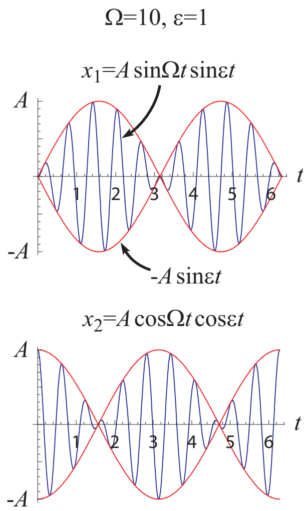

Figure 4

If ω_s is very close to ω_f, then $\epsilon \ll \Omega$, which means that the ϵt oscillation is much slower than that Ωt oscillation. The former therefore simply acts as an envelope for the latter. $x_1(t)$ and $x_2(t)$ are shown in Fig. 4 for $\Omega = 10$ and $\epsilon = 1$. The motion sloshes back and forth between the masses. At the start, only the second mass is moving. But after a time of $\epsilon t = \pi/2 \Longrightarrow t = \pi/2\epsilon$, the second mass is essentially not moving and the first mass has all the motion. Then after another time of $\pi/2\epsilon$ it switches back, and so on.

This sloshing back and forth can be understood in terms of driving forces and resonance. At the start (and until $\epsilon t = \pi/2$), x_2 looks like $\cos\Omega t$ with a slowly changing amplitude (assuming $\epsilon \ll \Omega$). And x_1 looks like $\sin\Omega t$ with a slowly changing amplitude. So x_2 is 90° *ahead* of x_1, because $\cos\Omega t = \sin(\Omega t + \pi/2)$. This 90° phase difference means that the x_2 mass basically acts like a driving force (on resonance) on the x_1 mass. Equivalently, the x_2 mass is always doing positive work on the x_1 mass, and the x_1 mass is always doing negative work on the x_2 mass. Energy is therefore transferred from x_2 to x_1.

However, right after x_2 has zero amplitude (instantaneously) at $\epsilon t = \pi/2$, the $\cos\epsilon t$ factor in x_2 *switches sign*, so x_2 now looks like $-\cos\Omega t$ (times a slowly-changing amplitude). And x_1 still looks like $\sin\Omega t$. So now x_2 is 90° *behind* x_1, because $-\cos\Omega t = \sin(\Omega t - \pi/2)$. So the x_1 mass now acts like a driving force (on resonance) on the x_2 mass. Energy is therefore transferred from x_1 back to x_2. And so on and so forth.

In the plots in Fig. 4, you can see that something goes a little haywire when the envelope curves pass through zero at $\epsilon t = \pi/2$, π, etc. The x_1 or x_2 curves skip ahead (or equivalently, fall behind) by half of a period. If you inverted the second envelope "bubble" in the first plot, the periodicity would then return. That is, the peaks of the fast-oscillation curve would occur at equal intervals, even in the transition region around $\epsilon t = \pi$.

The classic demonstration of beats consists of two identical pendulums connected by a weak spring. The gravitational restoring force mimics the "outside" springs in the above setup, so the same general results carry over (see Problem [to be added]). At the start, one pendulum moves while the other is nearly stationary. But then after a while the situation is reversed. However, if the masses of the pendulums are different, it turns out that not all of the energy is transferred. See Problem [to be added] for the details.

When people talk about the "beat frequency," they generally mean the frequency of the "bubbles" in the envelope curve. If you're listening to, say, the sound from two guitar strings that are at nearly the same frequency, then this beat frequency is the frequency of the waxing and waning that you hear. But note that this frequency is 2ϵ, and not ϵ, because two bubbles occur in each of the $\epsilon t = 2\pi$ periods of the envelope.[1]

2.1.5 Driven and damped coupled oscillators

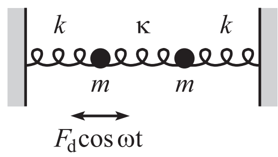

Figure 5

Consider the coupled oscillator system with two masses and three springs from Fig. 1 above, but now with a driving force acting on one of the masses, say the left one (the x_1 one); see Fig. 5. And while we're at it, let's immerse the system in a fluid, so that both masses have a drag coefficient b (we'll assume it's the same for both). Then the $F = ma$ equations are

$$m\ddot{x}_1 = -kx_1 - \kappa(x_1 - x_2) - b\dot{x}_1 + F_d\cos\omega t,$$

[1]If you want to map the spring/mass setup onto the guitar setup, then the x_1 in Eq. (21) represents the amplitude of the sound wave at your ear, and the ω_s and ω_f represent the two different nearby frequencies. The second position, x_2, doesn't come into play (or vice versa). Only one of the plots in Fig. 4 is relevant.

$$m\ddot{x}_2 = -kx_2 - \kappa(x_2 - x_1) - b\dot{x}_2. \tag{25}$$

We can solve these equations by using the same adding and subtracting technique we used in Section 2.1.1. Adding them gives

$$\begin{aligned} m(\ddot{x}_1 + \ddot{x}_2) &= -k(x_1 + x_2) - b(\dot{x}_1 + \dot{x}_2) + F_{\text{d}}\cos\omega t \\ \Longrightarrow \quad \ddot{z}_{\text{s}} + \gamma\dot{z}_{\text{s}} + \omega_{\text{s}}^2 z_{\text{s}} &= F\cos\omega t, \end{aligned} \tag{26}$$

where $z_{\text{s}} \equiv x_1 + x_2$, $\gamma \equiv b/m$, $\omega_{\text{s}}^2 \equiv k/m$, and $F \equiv F_{\text{d}}/m$. But this is our good ol' driven/dampded oscillator equation, in the variable z_{s}. We can therefore just invoke the results from Chapter 1. The general solution is the sum of the homogeneous and particular solutions. But the let's just concentrate on the particular (steady state) solution here. We can imagine that the system has been oscillating for a long time, so that the damping has made the homogeneous solution decay to zero. For the particular solution, we can simply copy the results from Section 1.3.1. So we have

$$x_1 + x_2 \equiv z_{\text{s}} = A_{\text{s}}\cos(\omega t + \phi_{\text{s}}), \tag{27}$$

where

$$\tan\phi_{\text{s}} = \frac{-\gamma\omega}{\omega_{\text{s}}^2 - \omega^2}, \qquad \text{and} \qquad A_{\text{s}} = \frac{F}{\sqrt{(\omega_{\text{s}}^2 - \omega^2)^2 + \gamma^2\omega^2}}. \tag{28}$$

Similarly, subtracting the $F = ma$ equations gives

$$\begin{aligned} m(\ddot{x}_1 - \ddot{x}_2) &= -(k + 2\kappa)(x_1 - x_2) - b(\dot{x}_1 - \dot{x}_2) + F_{\text{d}}\cos\omega t \\ \Longrightarrow \quad \ddot{z}_{\text{f}} + \gamma\dot{z}_{\text{f}} + \omega_{\text{f}}^2 z_{\text{f}} &= F\cos\omega t, \end{aligned} \tag{29}$$

where $z_{\text{f}} \equiv x_1 - x_2$ and $\omega_{\text{f}}^2 \equiv (k + 2\kappa)/m$. Again, this is a nice driven/dampded oscillator equation, and the particular solution is

$$x_1 - x_2 \equiv z_{\text{f}} = A_{\text{f}}\cos(\omega t + \phi_{\text{f}}), \tag{30}$$

where

$$\tan\phi_{\text{f}} = \frac{-\gamma\omega}{\omega_{\text{f}}^2 - \omega^2}, \qquad \text{and} \qquad A_{\text{f}} = \frac{F}{\sqrt{(\omega_{\text{f}}^2 - \omega^2)^2 + \gamma^2\omega^2}}. \tag{31}$$

Adding and subtracting Eqs. (27) and (30) to solve for $x_1(t)$ and $x_2(t)$ gives

$$\begin{aligned} x_1(t) &= C_{\text{s}}\cos(\omega t + \phi_{\text{s}}) + C_{\text{f}}\cos(\omega t + \phi_{\text{f}}), \\ x_2(t) &= C_{\text{s}}\cos(\omega t + \phi_{\text{s}}) - C_{\text{f}}\cos(\omega t + \phi_{\text{f}}), \end{aligned} \tag{32}$$

where $C_{\text{s}} \equiv A_{\text{s}}/2$, and $C_{\text{f}} \equiv A_{\text{f}}/2$.

We end up getting *two* resonant frequencies, which are simply the frequencies of the normal modes, ω_{s} and ω_{f}. If γ is small, and if the driving frequency ω equals either ω_{s} or ω_{f}, then the amplitudes of x_1 and x_2 are large. In the $\omega = \omega_{\text{s}}$ case, x_1 and x_2 are approximately in phase with equal amplitudes (the C_{s} terms dominate the C_{f} terms). And in the $\omega = \omega_{\text{s}}$ case, x_1 and x_2 are approximately out of phase with equal amplitudes (the C_{f} terms dominate the C_{s} terms, and there is a relative minus sign). But these are the normal modes we found in Section 2.1.3. The resonances therefore cause the system to be in the normal modes.

In general, if there are N masses (and hence N modes), then there are N resonant frequencies, which are the N normal-mode frequencies. So for complicated objects with more than two pieces, there are *lots* of resonances.

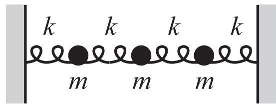

Figure 6

2.2 Three masses

As a warmup to the general case of N masses connected by springs, let's look at the case of three masses, as shown in Fig. 6. We'll just deal with undriven and undamped motion here, and we'll also assume that all the spring constants are equal, lest the math get intractable. If x_1, x_2, and x_3 are the displacements of the three masses from their equilibrium positions, then the three $F = ma$ equations are

$$\begin{aligned} m\ddot{x}_1 &= -kx_1 - k(x_1 - x_2), \\ m\ddot{x}_2 &= -k(x_2 - x_1) - k(x_2 - x_3), \\ m\ddot{x}_3 &= -k(x_3 - x_2) - kx_3. \end{aligned} \tag{33}$$

You can check that all the signs of the $k(x_i - x_j)$ terms are correct, by imagining that, say, one of the x's is very large. It isn't so obvious which combinations of these equations yield equations involving only certain unique combinations of the x's (the normal coordinates), so we won't be able to use the method of Section 2.1.1. We will therefore use the determinant method from Section 2.1.2 and guess a solution of the form

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} e^{i\omega t}, \tag{34}$$

with the goal of solving for ω, and also for the amplitudes A_1, A_2, and A_3 (up to an overall factor). Plugging this guess into Eq. (33) and putting all the terms on the lefthand side, and canceling the $e^{i\omega t}$ factor, gives

$$\begin{pmatrix} -\omega^2 + 2\omega_0^2 & -\omega_0^2 & 0 \\ -\omega_0^2 & -\omega^2 + 2\omega_0^2 & -\omega_0^2 \\ 0 & -\omega_0^2 & -\omega^2 + 2\omega_0^2 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{35}$$

where $\omega_0^2 \equiv k/m$. As in the earlier two-mass case, a nonzero solution for (A_1, A_2, A_3) exists only if the determinant of this matrix is zero. Setting it equal to zero gives

$$\begin{aligned} (-\omega^2 + 2\omega_0^2)\Big((-\omega^2 + 2\omega_0^2)^2 - \omega_0^4\Big) + \omega_0^2\Big(-\omega_0^2(-\omega^2 + 2\omega_0^2)\Big) &= 0 \\ \Longrightarrow \quad (-\omega^2 + 2\omega_0^2)(\omega^4 - 4\omega_0^2\omega^2 + 2\omega_0^4) &= 0. \end{aligned} \tag{36}$$

Although this is technically a 6th-order equation, it's really just a cubic equation in ω^2. But since we know that $(-\omega^2 + 2\omega_0^2)$ is a factor, in the end it boils down to a quadratic equation in ω^2.

Remark: If you had multiplied everything out and lost the information that $(-\omega^2 + 2\omega_0^2)$ is a factor, you could still easily see that $\omega^2 = 2\omega_0^2$ must be a root, because an easy-to-see normal mode is one where the middle mass stays fixed and the outer masses move in opposite directions. In this case the middle mass is essentially a brick wall, so the outer masses are connected to two springs whose other ends are fixed. The effective spring constant is then $2k$, which means that the frequency is $\sqrt{2}\omega_0$. ♣

Using the quadratic formula, the roots to Eq. (36) are

$$\omega^2 = 2\omega_0^2, \qquad \text{and} \qquad \omega^2 = (2 \pm \sqrt{2})\omega_0. \tag{37}$$

Plugging these values back into Eq. (35) to find the relations among A_1, A_2, and A_3 gives the three normal modes:[2]

$$\omega = \pm\sqrt{2}\,\omega_0 \quad\Longrightarrow\quad \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

$$\omega = \pm\sqrt{2+\sqrt{2}}\,\omega_0 \quad\Longrightarrow\quad \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix},$$

$$\omega = \pm\sqrt{2-\sqrt{2}}\,\omega_0 \quad\Longrightarrow\quad \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}. \tag{38}$$

The most general solution is obtained by taking an arbitrary linear combination of the six solutions corresponding to the six possible values of ω (don't forget the three negative solutions):

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = C_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} e^{i\sqrt{2}\omega_0 t} + C_2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} e^{-i\sqrt{2}\omega_0 t} + \cdots. \tag{39}$$

However, the x's must be real, so C_2 must be the complex conjugate of C_1. Likewise for the two C's corresponding to the $(1, -\sqrt{2}, 1)$ mode, and also for the two C's corresponding to the $(1, \sqrt{2}, 1)$ mode. Following the procedure that transformed Eq. (13) into Eq. (14), we see that the most general solution can be written as

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= A_{\rm m} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \cos\left(\sqrt{2}\,\omega_0 t + \phi_{\rm m}\right) \\ &\quad + A_{\rm f} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix} \cos\left(\sqrt{2+\sqrt{2}}\,\omega_0 t + \phi_{\rm f}\right) \\ &\quad + A_{\rm s} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix} \cos\left(\sqrt{2-\sqrt{2}}\,\omega_0 t + \phi_{\rm s}\right). \end{aligned} \tag{40}$$

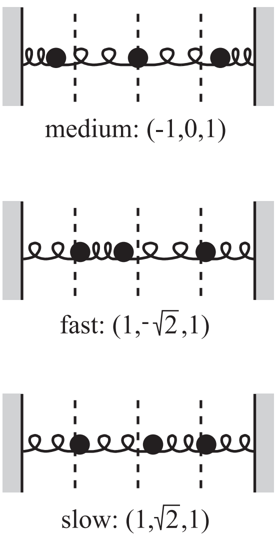

Figure 7

The subscripts "m," "f," and "s" stand for middle, fast, and slow. The six unknowns, $A_{\rm m}$, $A_{\rm f}$, $A_{\rm s}$, $\phi_{\rm m}$, $\phi_{\rm f}$, and $\phi_{\rm s}$ are determined by the six initial conditions (three positions and three velocities). If $A_{\rm m}$ is the only nonzero coefficient, then the motion is purely in the middle mode. Likewise for the cases where only $A_{\rm f}$ or only $A_{\rm s}$ is nonzero. Snapshots of these modes are shown in Fig. 7. You should convince yourself that they qualitatively make sense. If you want to get quantitative, the task of Problem [to be added] is to give a force argument that explains the presence of the $\sqrt{2}$ in the amplitudes of the fast and slow modes.

2.3 N masses

2.3.1 Derivation of the general result

Let's now consider the general case of N masses between two fixed walls. The masses are all equal to m, and the spring constants are all equal to k. The method we'll use below will

[2]Only two of the equations in Eq. (35) are needed. The third equation is redundant; that was the point of setting the determinant equal to zero.

actually work even if we don't have walls at the ends, that is, even if the masses extend infinitely in both directions. Let the displacements of the masses relative to their equilibrium positions be x_1, x_2,..., x_N. If the displacements of the walls are called x_0 and x_{N+1}, then the boundary conditions that we'll eventually apply are $x_0 = x_{N+1} = 0$.

The force on the nth mass is

$$F_n = -k(x_n - x_{n-1}) - k(x_n - x_{n+1}) = kx_{n-1} - 2kx_n + kx_{n+1}. \tag{41}$$

So we end up with a collection of $F = ma$ equations that look like

$$m\ddot{x}_n = kx_{n-1} - 2kx_n + kx_{n+1}. \tag{42}$$

These can all be collected into the matrix equation,

$$m\frac{d^2}{dt^2}\begin{pmatrix} \vdots \\ x_{n-1} \\ x_n \\ x_{n+1} \\ \vdots \end{pmatrix} = \begin{pmatrix} & & \vdots & & & & \\ \cdots & k & -2k & k & & & \\ & & k & -2k & k & & \\ & & & k & -2k & k & \cdots \\ & & & & \vdots & & \end{pmatrix}\begin{pmatrix} \vdots \\ x_{n-1} \\ x_n \\ x_{n+1} \\ \vdots \end{pmatrix}. \tag{43}$$

In principle, we could solve for the normal modes by guessing a solution of the form,

$$\begin{pmatrix} \vdots \\ x_{n-1} \\ x_n \\ x_{n+1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ A_{n-1} \\ A_n \\ A_{n+1} \\ \vdots \end{pmatrix} e^{i\omega t}, \tag{44}$$

and then setting the resulting determinant equal to zero. This is what we did in the $N = 2$ and $N = 3$ cases above. However, for large N, it would be completely intractable to solve for the ω's by using the determinant method. So we'll solve it in a different way, as follows.

We'll stick with the guess in Eq. (44), but instead of the determinant method, we'll look at each of the $F = ma$ equations individually. Consider the nth equation. Plugging $x_n(t) = A_n e^{i\omega t}$ into Eq. (42) and canceling the factor of $e^{i\omega t}$ gives

$$\begin{aligned} -\omega^2 A_n &= \omega_0^2\big(A_{n-1} - 2A_n + A_{n+1}\big) \\ \Longrightarrow \quad \frac{A_{n-1} + A_{n+1}}{A_n} &= \frac{2\omega_0^2 - \omega^2}{\omega_0^2}, \end{aligned} \tag{45}$$

where $\omega_0 = \sqrt{k/m}$, as usual. This equation must hold for all values of n from 1 to N, so we have N equations of this form. For a given mode with a given frequency ω, the quantity $(2\omega_0^2 - \omega^2)/\omega_0^2$ on the righthand side is a constant, independent of n. So the ratio $(A_{n-1}+A_{n+1})/A_n$ on the lefthand side must also be independent of n. The problem therefore reduces to finding the general form of a string of A's that has the ratio $(A_{n-1} + A_{n+1})/A_n$ being independent of n.

If someone gives you three adjacent A's, then this ratio is determined, so you can recursively find the A's for all other n (both larger and smaller than the three you were given). Or equivalently, if someone gives you two adjacent A's and also ω, so that the value of $(2\omega_0^2 - \omega^2)/\omega_0^2$ is known (we're assuming that ω_0 is given), then all the other A's can be determined. The following claim tells us what form the A's take. It is this claim that allows us to avoid using the determinant method.

Claim 2.1 *If $\omega \leq 2\omega_0$, then any set of A_n's satisfying the system of N equations in Eq. (45) can be written as*

$$A_n = B\cos n\theta + C\sin n\theta, \tag{46}$$

for certain values of B, C, and θ. (The fact that there are three parameters here is consistent with the fact that three A's, or two A's and ω, determine the whole set.)

Proof: We'll start by defining

$$\cos\theta \equiv \frac{A_{n-1} + A_{n+1}}{2A_n}. \tag{47}$$

As mentioned above, the righthand side is independent of n, so θ is well defined (up to the usual ambiguities of duplicate angles; $\theta + 2\pi$, and $-\theta$, etc. also work).[3] If we're looking at a given normal mode with frequency ω, then in view of Eq. (45), an equivalent definition of θ is

$$2\cos\theta \equiv \frac{2\omega_0^2 - \omega^2}{\omega_0^2}. \tag{48}$$

These definitions are permitted only if they yield a value of $\cos\theta$ that satisfies $|\cos\theta| \leq 1$. This condition is equivalent to the condition that ω must satisfy $-2\omega_0 \leq \omega \leq 2\omega_0$. We'll just deal with positive ω here (negative ω yields the same results, because only its square enters into the problem), but we must remember to also include the $e^{-i\omega t}$ solution in the end (as usual). So this is where the $\omega \leq 2\omega_0$ condition in the claim comes from.[4]

We will find that with walls at the ends, θ (and hence ω) can take on only a certain set of discrete values. We will calculate these below. If there are no walls, that is, if the system extends infinitely in both directions, then θ (and hence ω) can take on a continuous set of values.

As we mentioned above, the N equations represented in Eq. (45) tell us that if we know two of the A's, and if we also have a value of ω, then we can use the equations to successively determine all the other A's. Let's say that we know what A_0 and A_1 are. (In the case where there are walls, we know know that $A_0 = 0$, but let's be general and not invoke this constraint yet.) The rest of the A_n's can be determined as follows. Define B by

$$A_0 \equiv B\cos(0\cdot\theta) + C\sin(0\cdot\theta) \implies A_0 \equiv B. \tag{49}$$

(So $B = 0$ if there are walls.) Once B has been defined, define C by

$$A_1 \equiv B\cos(1\cdot\theta) + C\sin(1\cdot\theta) \implies A_1 \equiv B\cos\theta + C\sin\theta, \tag{50}$$

For any A_0 and A_1, these two equations uniquely determine B and C (θ was already determined by ω). So to sum up the definitions: ω, A_0, and A_1 uniquely determine θ, B and C. (We'll deal with the multiplicity of the possible θ values below in the "Nyquist" subsection.) By construction of these definitions, the proposed $A_n = B\cos n\theta + C\sin n\theta$ relation holds for $n = 0$ and $n = 1$. We will now show inductively that it holds for all n.

[3] The motivation for this definition is that the fraction on the righthand side has a sort of second-derivative feel to it. The more this fraction differs from 1, the more curvature there is in the plot of the A_n's. (If the fraction equals 1, then each A_n is the average of its two neighbors, so we just have a straight line.) And since it's a good bet that we're going to get some sort of sinusoidal result out of all this, it's not an outrageous thing to define this fraction to be a sinusoidal function of a new quantity θ. But in the end, it does come a bit out of the blue. That's the way it is sometimes. However, you will find it less mysterious after reading Section 2.4, where we actually end up with a true second derivative, along with sinusoidal functions of x (the analog of n here).

[4] If $\omega > 2\omega_0$, then we have a so-called *evanescent* wave. We'll discuss these in Chapter 6. The $\omega = 0$ and $\omega = 2\omega_0$ cases are somewhat special; see Problem [to be added].

If we solve for A_{n+1} in Eq. (47) and use the inductive hypothesis that the $A_n = B\cos n\theta + C\sin n\theta$ result holds for $n-1$ and n, we have

$$\begin{aligned}
A_{n+1} &= (2\cos\theta)A_n - A_{n-1} \\
&= 2\cos\theta\Big(B\cos n\theta + C\sin n\theta\Big) - \Big(B\cos(n-1)\theta + C\sin(n-1)\theta\Big) \\
&= B\Big(2\cos n\theta\cos\theta - (\cos n\theta\cos\theta + \sin n\theta\sin\theta)\Big) \\
&\quad\quad + C\Big(2\sin n\theta\cos\theta - (\sin n\theta\cos\theta - \cos n\theta\sin\theta)\Big) \\
&= B\Big(\cos n\theta\cos\theta - \sin n\theta\sin\theta)\Big) + C\Big(\sin n\theta\cos\theta + \cos n\theta\sin\theta)\Big) \\
&= B\cos(n+1)\theta + C\sin(n+1)\theta, \qquad (51)
\end{aligned}$$

which is the desired expression for the case of $n+1$. (Note that this works independently for the B and C terms.) Therefore, since the $A_n = B\cos n\theta + C\sin n\theta$ result holds for $n = 0$ and $n = 1$, and since the inductive step is valid, the result therefore holds for all n.

If you wanted, you could have instead solved for A_{n-1} in Eq. (51) and demonstrated that the inductive step works in the negative direction too. Therefore, starting with two arbitrary masses anywhere in the line, the $A_n = B\cos n\theta + C\sin n\theta$ result holds even for an infinite number of masses extending in both directions. ■

This claim tells us that we have found a solution of the form,

$$x_n(t) = A_n e^{i\omega t} = (B\cos n\theta + C\sin n\theta)e^{i\omega t}. \qquad (52)$$

However, with the convention that ω is positive, we must remember that an $e^{-i\omega t}$ solution works just as well. So another solution is

$$x_n(t) = A_n e^{-i\omega t} = (D\cos n\theta + E\sin n\theta)e^{-i\omega t}. \qquad (53)$$

Note that the coefficients in this solution need not be the same as those in the $e^{i\omega t}$ solution. Since the $F = ma$ equations in Eq. (42) are all linear, the sum of two solutions is again a solution. So the most general solution (for a given value of ω) is the sum of the above two solutions (each of which is itself a linear combination of two solutions).

As usual, we now invoke the fact that the positions must be real. This implies that the above two solutions must be complex conjugates of each other. And since this must be true for all values of n, we see that B and D must be complex conjugates, and likewise for C and E. Let's define $B = D^* \equiv (F/2)e^{i\beta}$ and $C = E^* \equiv (G/2)e^{i\gamma}$. There is no reason why B, C, D, and E (or equivalently the A's in Eq. (44)) have to be real. The sum of the two solutions then becomes

$$x_n(t) = F\cos n\theta\cos(\omega t + \beta) + G\sin n\theta\cos(\omega t + \gamma) \qquad (54)$$

As usual, we could have just taken the real part of either of the solutions to obtain this (up to a factor of 2, which we can absorb into the definition of the constants). We can make it look a little more symmetrical be using the trig sum formula for the cosines. This gives the result (we're running out of letters, so we'll use C_i's for the coefficients here),

$$\boxed{x_n(t) = C_1\cos n\theta\cos\omega t + C_2\cos n\theta\sin\omega t + C_3\sin n\theta\cos\omega t + C_4\sin n\theta\sin\omega t} \qquad (55)$$

where θ is determined by ω via Eq. (48), which we can write in the form,

$$\boxed{\theta \equiv \cos^{-1}\left(\frac{2\omega_0^2 - \omega^2}{2\omega_0^2}\right)} \qquad (56)$$

The constants C_1, C_2, C_3, C_4 in Eq. (55) are related to the constants F, G, β, γ in Eq. (54) in the usual way ($C_1 = F\cos\beta$, etc.). There are yet other ways to write the solution, but we'll save the discussion of these for Section 2.4.

Eq. (55) is the most general form of the positions for the mode that has frequency ω. This set of the $x_n(t)$ functions (N of them) satisfies the $F = ma$ equations in Eq. (42) (N of them) for any values of C_1, C_2, C_3, C_4. These four constants are determined by four initial values, for example, $x_0(0)$, $\dot{x}_0(0)$, $x_1(0)$, and $\dot{x}_1(0)$. Of course, if $n = 0$ corresponds to a fixed wall, then the first two of these are zero.

Remarks:

1. Interestingly, we have found that $x_n(t)$ varies sinusoidally with position (that is, with n), as well as with time. However, whereas time takes on a continuous set of values, the position is relevant only at the discrete locations of the masses. For example, if the equilibrium positions are at the locations $z = na$, where a is the equilibrium spacing between the masses, then we can rewrite $x_n(t)$ in terms of z instead of n, using $n = z/a$. Assuming for simplicity that we have only, say, the $C_1 \cos n\theta \cos\omega t$ part of the solution, we have

$$x_n(t) \implies x_z(t) = C_1 \cos(z\theta/a)\cos\omega t. \tag{57}$$

   For a given values of θ (which is related to ω) and a, this is a sinusoidal function of z (as well as of t). But we must remember that it is defined only at the discrete values of z of the form, $z = na$. We'll draw some nice pictures below to demonstrate the sinusoidal behavior, when we discuss a few specific values of N.

2. We should stress the distinction between z (or equivalently n) and x. z represents the equilibrium positions of the masses. A given mass is associated with a unique value of z. z doesn't change as the mass moves. $x_z(t)$, on the other hand, measures the position of a mass (the one whose equilibrium position is z) relative to its equilibrium position (namely z). So the total position of a given mass is $z + x$. The function $x_z(t)$ has dependence on both z and t, so we could very well write it as a function of two variables, $x(z, t)$. We will in fact adopt this notation in Section 2.4 when we talk about continuous systems. But in the present case where z can take on only discrete values, we'll stick with the $x_z(t)$ notation. But either notation is fine.

3. Eq. (55) gives the most general solution *for a given value of* ω, that is, for a given mode. While the most general motion of the masses is certainly not determined by $x_0(0)$, $\dot{x}_0(0)$, $x_1(0)$, and $\dot{x}_1(0)$, the motion for a single mode is. Let's see why this is true. If we apply the $x_0(0)$ and $x_1(0)$ boundary conditions to Eq. (55), we obtain $x_0(0) = C_1$ and $x_1(0) = C_1\cos\theta + C_3\sin\theta$. Since we are assuming that ω (and hence θ) is given, these two equations determine C_1 and C_3. But C_1 and C_3 in turn determine all the other $x_n(0)$ values via Eq. (55), because the $\sin\omega t$ terms are all zero at $t = 0$. So for a given mode, $x_0(0)$ and $x_1(0)$ determine all the other initial positions. In a similar manner, the $\dot{x}_0(0)$ and $\dot{x}_1(0)$ values determine C_2 and C_4, which in turn determine all the other initial velocities. Therefore, since the four values $x_0(0)$, $\dot{x}_0(0)$, $x_1(0)$, and $\dot{x}_1(0)$ give us all the initial positions and velocities, and since the accelerations depend on the positions (from the $F = ma$ equations in Eq. (42)), the future motion of all the masses is determined. ♣

2.3.2 Wall boundary conditions

Let us now see what happens when we invoke the boundary conditions due to fixed walls at the two ends. The boundary conditions at the walls are $x_0(t) = x_{N+1}(t) = 0$, for all t. These conditions are most easily applied to $x_n(t)$ written in the form in Eq. (54), although the form in Eq. (55) will work fine too. At the left wall, the $x_0(t) = 0$ condition gives

$$\begin{aligned} 0 = x_0(t) &= F\cos(0)\cos(\omega t + \beta) + G\sin(0)\cos(\omega t + \gamma) \\ &= F\cos(\omega t + \beta). \end{aligned} \tag{58}$$

If this is to be true for all t, we must have $F = 0$. So we're left with just the $G \sin n\theta \cos(\omega t + \gamma)$ term in Eq. (54). Applying the $x_{N+1}(t) = 0$ condition to this then gives

$$0 = x_{N+1}(t) = G \sin(N+1)\theta \, \cos(\omega t + \gamma). \tag{59}$$

One way for this to be true for all t is to have $G = 0$. But then all the x's are identically zero, which means that we have no motion at all. The other (nontrivial) way for this to be true is to have the $\sin(N+1)\theta$ factor be zero. This occurs when

$$(N+1)\theta = m\pi \quad \Longrightarrow \quad \theta = \frac{m\pi}{N+1}, \tag{60}$$

where m is an integer. The solution for $x_n(t)$ is therefore

$$\boxed{x_n(t) = G \sin\left(\frac{nm\pi}{N+1}\right) \cos(\omega t + \gamma)} \tag{61}$$

The amplitudes of the masses are then

$$\boxed{A_n = G \sin\left(\frac{nm\pi}{N+1}\right)} \tag{62}$$

We've made a slight change in notation here. The A_n that we're now using for the amplitude is the *magnitude* of the A_n that we used in Eq. (44). That A_n was equal to $B \cos n\theta + C \sin n\theta$, which itself is some complex number which can be written in the form, $|A_n|e^{i\alpha}$. The solution for $x_n(t)$ is obtained by taking the real part of Eq. (52), which yields $x_n(t) = |A_n| \cos(\omega t + \alpha)$. So we're now using A_n to stand for $|A_n|$, lest we get tired if writing the absolute value bars over and over.[5] And α happens to equal the γ in Eq. (61).

If we invert the definition of θ in Eq. (48) to solve for ω in terms of θ, we find that the frequency is given by

$$\begin{aligned} 2\cos\theta \equiv \frac{2\omega_0^2 - \omega^2}{\omega_0^2} \quad \Longrightarrow \quad \omega^2 &= 2\omega_0^2(1 - \cos\theta) \\ &= 4\omega_0^2 \sin^2(\theta/2) \\ \Longrightarrow \quad \omega &= \boxed{2\omega_0 \sin\left(\frac{m\pi}{2(N+1)}\right)} \end{aligned} \tag{63}$$

We've taken the positive square root, because the convention is that ω is positive. We'll see below that m labels the normal mode (so the "m" stands for "mode"). If $m = 0$ or $m = N + 1$, then Eq. (61) says that all the x_n's are identically zero, which means that we don't have any motion at all. So only m values in the range $1 \le m \le N$ are relevant. We'll see below in the "Nyquist" subsection that higher values of m simply give repetitions of the modes generated by the $1 \le m \le N$ values.

The most important point in the above results is that if the boundary conditions are walls at both ends, then the θ in Eq. (60), and hence the ω in Eq. (63), can take on only a certain set of *discrete* values. This is consistent with our results for the $N = 2$ and $N = 3$ cases in Sections 2.1 and 2.2, where we found that there were only two or three (respectively) allowed values of ω, that is, only two or three normal modes. Let's now show that for $N = 2$ and $N = 3$, the preceding equations quantitatively reproduce the results from Sections 2.1 and 2.2. You can examine higher values of N in Problem [to be added].

[5] If instead of taking the real part, you did the nearly equivalent thing of adding on the complex conjugate solution in Eq. (53), then $2|A_n|$ would be the amplitude. In this case, the A_n in Eq. (62) stands for $2|A_n|$.

The $N = 2$ case

If $N = 2$, there are two possible values of m:

- $m = 1$: Eqs. (62) and (63) give

$$A_n \propto \sin\left(\frac{n\pi}{3}\right), \qquad \text{and} \qquad \omega = 2\omega_0 \sin\left(\frac{\pi}{6}\right). \tag{64}$$

  So this mode is given by

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \propto \begin{pmatrix} \sin(\pi/3) \\ \sin(2\pi/3) \end{pmatrix} \propto \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \text{and} \qquad \omega = \omega_0. \tag{65}$$

  These agree with the first mode we found in Section 2.1.2. The frequency is ω_0, and the masses move in phase with each other.

- $m = 2$: Eqs. (62) and (63) give

$$A_n \propto \sin\left(\frac{2n\pi}{3}\right), \qquad \text{and} \qquad \omega = 2\omega_0 \sin\left(\frac{\pi}{3}\right). \tag{66}$$

  So this mode is given by

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \propto \begin{pmatrix} \sin(2\pi/3) \\ \sin(4\pi/3) \end{pmatrix} \propto \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \text{and} \qquad \omega = \sqrt{3}\,\omega_0. \tag{67}$$

  These agree with the second mode we found in Section 2.1.2. The frequency is $\sqrt{3}\,\omega_0$, and the masses move exactly out of phase with each other.

To recap, the various parameters are: N (the number of masses), m (the mode number), and n (the label of each of the N masses). n runs from 1 to N, of course. And m effectively also runs from 1 to N (there are N possible modes for N masses). We say "effectively" because as we mentioned above, although m can technically take on any integer value, the values that lie outside the $1 \leq m \leq N$ range give duplications of the modes inside this range. See the "Nyquist" subsection below.

In applying Eqs. (62) and (63), things can get a little confusing because of all the parameters floating around. And this is just the simple case of $N = 2$. Fortunately, there is an extremely useful graphical way to see what's going on. This is one situation where a picture is indeed worth a thousand words (or equations).

If we write the argument of the sin in Eq. (62) as $m\pi \cdot n/(N + 1)$, then we see that for a given N, the relative amplitudes of the masses in the mth mode are obtained by drawing a sin curve with m half oscillations, and then finding the value of this curve at equal "$1/(N + 1)$" intervals along the horizontal axis. Fig. 8 shows the results for $N = 2$. We've drawn either $m = 1$ or $m = 2$ half oscillations, and we've divided each horizontal axis into $N + 1 = 3$ equal intervals. These curves look a lot like snapshots of beads on a string oscillating *transversely* back and forth. And indeed, we will find in Chapter 4 that the $F = ma$ equations for transverse motion of beads on a string are *exactly* the same as the equations in Eq. (42) for the longitudinal motion of the spring/mass system. But for now, all of the displacements indicated in these pictures are in the longitudinal direction. And the displacements have meaning only at the discrete locations of the masses. There isn't anything actually happening at the rest of the points on the curve.

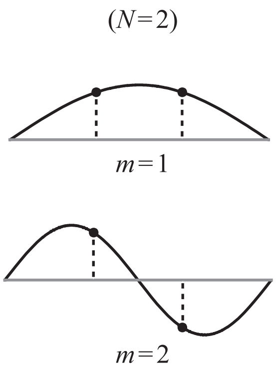

Figure 8

We can also easily visualize what the frequencies are. If we write the argument of the sin in Eq. (63) as $\pi/2 \cdot m/(N+1)$ then we see that for a given N, the frequency of the mth mode is obtained by breaking a quarter circle (with radius $2\omega_0$) into "$1/(N+1)$" equal intervals, and then finding the y values of the resulting points. Fig. 9 shows the results for $N = 2$. We've divided the quarter circle into $N + 1 = 3$ equal angles of $\pi/6$, which results in points at the angles of $\pi/6$ and $\pi/3$. It is *much* easier to see what's going on by looking at the pictures in Figs. 8 and 9 than by working with the algebraic expressions in Eqs. (62) and (63).

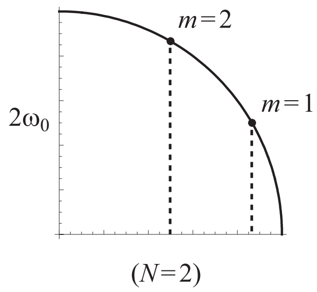

Figure 9

The $N = 3$ case

If $N = 3$, there are three possible values of m:

- $m = 1$: Eqs. (62) and (63) give

$$A_n \propto \sin\left(\frac{n\pi}{4}\right), \qquad \text{and} \qquad \omega = 2\omega_0 \sin\left(\frac{\pi}{8}\right). \tag{68}$$

So this mode is given by

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} \sin(\pi/4) \\ \sin(2\pi/4) \\ \sin(3\pi/4) \end{pmatrix} \propto \begin{pmatrix} 1 \\ \sqrt{2} \\ -1 \end{pmatrix}, \qquad \text{and} \qquad \omega = \sqrt{2-\sqrt{2}}\,\omega_0, \tag{69}$$

where we have used the half-angle formula for $\sin(\pi/8)$ to obtain ω. (Or equivalently, we just used the first line in Eq. (63).) These results agree with the "slow" mode we found in Section 2.2.

- $m = 2$: Eqs. (62) and (63) give

$$A_n \propto \sin\left(\frac{2n\pi}{4}\right), \qquad \text{and} \qquad \omega = 2\omega_0 \sin\left(\frac{2\pi}{8}\right). \tag{70}$$

So this mode is given by

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} \sin(2\pi/4) \\ \sin(4\pi/4) \\ \sin(6\pi/4) \end{pmatrix} \propto \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \text{and} \qquad \omega = \sqrt{2}\,\omega_0. \tag{71}$$

These agree with the "medium" mode we found in Section 2.2.

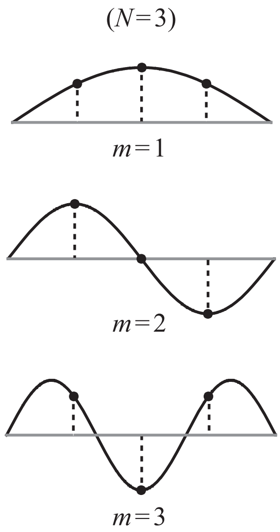

Figure 10

- $m = 3$: Eqs. (62) and (63) give

$$A_n \propto \sin\left(\frac{3n\pi}{4}\right), \qquad \text{and} \qquad \omega = 2\omega_0 \sin\left(\frac{3\pi}{8}\right). \tag{72}$$

So this mode is given by

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \propto \begin{pmatrix} \sin(3\pi/4) \\ \sin(6\pi/4) \\ \sin(9\pi/4) \end{pmatrix} \propto \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix}, \qquad \text{and} \qquad \omega = \sqrt{2+\sqrt{2}}\,\omega_0. \tag{73}$$

These agree with the "fast" mode we found in Section 2.2..

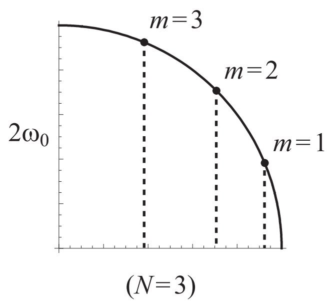

Figure 11

As with the $N = 2$ case, it's much easier to see what's going on if we draw some pictures. Fig. 10shows the relative amplitudes within the three modes, and Fig. 11shows the associated frequencies. Each horizontal axis in Fig. 10 is broken up into $N+1=4$ equal segments, and the quarter circle in Fig. 11 is broken up into $N+1=4$ equal arcs.

As mentioned above, although Fig. 10 looks like transverse motion on a string, remember that all the displacements indicated in this figure are in the longitudinal direction. For example, in the first $m=1$ mode, all three masses move in the same direction, but the middle one moves farther (by a factor of $\sqrt{2}$) than the outer ones. Problem [to be added] discusses higher values of N.

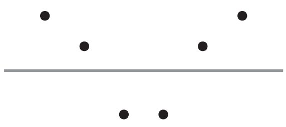

Figure 12

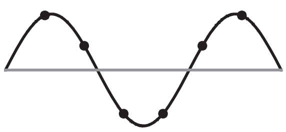

Figure 13

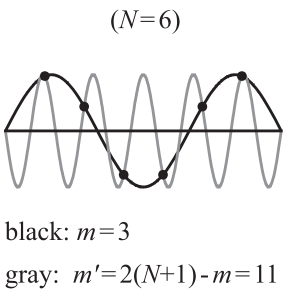

Figure 14

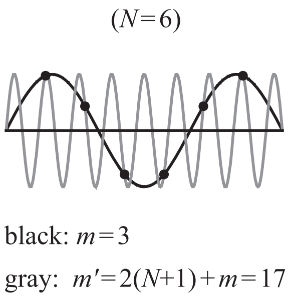

Figure 15

Aliasing, Nyquist frequency

Consider the $N=6$ case shown in Fig. 12. Assuming that this corresponds to a normal mode, which one is it? If you start with $m=1$ and keep trying different values of m until the (appropriately scaled) sin curve in Eq. (62) matches up with the six points in Fig. 12 (corresponding to n values from 1 to N), you'll find that $m=3$ works, as shown in Fig. 13.

However, the question then arises as to whether this is the only value of m that allows the sin curve to fit the given points. If you try some higher values of m, and if you're persistent, then you'll find that $m=11$ also works, as long as you throw in a minus sign in front of the sin curve (this corresponds to the G coefficient in Eq. (61) being negative, which is fine). This is shown in Fig. 14. And if you keep going with higher m values, you'll find that $m=17$ works too, and this time there's no need for a minus sign. This is shown in Fig. 15. Note that the m value always equals the number of bumps (local maxima or minima) in the sin curve.

It turns out that there is an infinite number of m values that work, and they fall into two classes. If we start with particular values of N and m (6 and 3 here), then all m' values of the form,

$$m' = 2a(N+1) - m, \tag{74}$$

also work (with a minus sign in the sin curve), where a is any integer. And all m' values of the form,

$$m' = 2a(N+1) + m, \tag{75}$$

also work, where again a is any integer. You can verify these claims by using Eq. (62); see Problem [to be added]. For $N=6$ and $m=3$, the first of these classes contains $m'=11, 25, 39, \ldots$, and the second class contains $m'=3, 17, 31, 45, \ldots$. Negative values of a work too, but they simply reproduce the sin curves for these m' values, up to an overall minus sign.

You can show using Eq. (63) that the frequencies corresponding to *all* of these values of m' (in both classes) are equal; see Problem [to be added] (a frequency of $-\omega$ yields the same motion as a frequency of ω). So as far as the motions of the six masses go, all of these modes yield *exactly* the same motions of the masses. (The other parts of the various sin curves don't match up, but all we care about are the locations of the masses.) It is impossible to tell which mode the masses are in. Or said more accurately, the masses aren't really in any one particular mode. There isn't one "correct" mode. Any of the above m or m' values is just as good as any other. However, by convention we label the mode with the m value in the range $1 \le m \le N$.

The above discussion pertains to a setup with N discrete masses in a line, with massless springs between them. However, if we have a continuous string/mass system, or in other words a massive spring (we'll talk about such a system in Section 2.4), then the different m' values *do* represent physically different motions. The $m=3$ and $m=17$ curves in Fig. 15 are certainly different. You can think of a continuous system as a setup with $N \longrightarrow \infty$ masses, so all the m values in the range $1 \le m \le N \Longrightarrow 1 \le m \le \infty$ yield different modes. In other words, each value of m yields a different mode.

However, if we have a continuous string, and if we only look at what is happening at equally spaced locations along it, then there is no way to tell what mode the string is really in (and in this case it really *is* in a well defined mode). If the string is in the $m = 11$ mode, and if you only look at the six equally-spaced points we considered above, then you won't be able to tell which of the $m = 3, 11, 17, 25, \ldots$ modes is the correct one.

This ambiguity is known as *aliasing*, or the *nyquist* effect. If you look at only discrete points in space, then you can't tell the true spatial frequency. Or similarly, If you look at only discrete moments in time, then you can't tell the true temporal frequency. This effect manifests itself in many ways in the real world. If you watch a car traveling by under a streetlight (which emits light in quick pulses, unlike an ordinary filament lightbulb), or if you watch a car speed by in a movie (which was filmed at a certain number of frames per second), then the "spokes" on the tires often appear to be moving at a different angular rate than the actual angular rate of the tire. They might even appear to be moving backwards. This is called the "strobe" effect. There are also countless instances of aliasing in electronics.

2.4 $N \to \infty$ and the wave equation

Let's now consider the $N \to \infty$ limit of our mass/spring setup. This means that we'll now effectively have a continuous system. This will actually make the problem easier than the finite-N case in the previous section, and we'll be able to use a quicker method to solve it. If you want, you can use the strategy of taking the $N \to \infty$ limit of the results in the previous section. This method will work (see Problem [to be added]), but we'll derive things from scratch here, because the method we will use is a very important one, and it will come up again in our study of waves.

First, a change of notation. The equilibrium position of each mass will now play a more fundamental role and appear more regularly, so we're going to label it with x instead of n (or instead of the z we used in the first two remarks at the end of Section 2.3.1). So x_n is the equilibrium position of the nth mass (we'll eventually drop the subscript n). We now need a new letter for the displacement of the masses, because we used x_n for this above. We'll use ξ now. So ξ_n is the displacement of the nth mass. The x's are constants (they just label the equilibrium positions, which don't change), and the ξ's are the things that change with time. The actual location of the nth mass is $x_n + \xi_n$, but only the ξ_n part will show up in the $F = ma$ equations, because the x_n terms don't contribute to the acceleration (because they are constant), nor do they contribute to the force (because only the displacement from equilibrium matters, since the spring force is linear).

Instead of the ξ_n notation, we'll use $\xi(x_n)$. And we'll soon drop the subscript n and just write $\xi(x)$. All three of the ξ_n, $\xi(x_n)$, $\xi(x)$ expressions stand for the same thing, namely the displacement from equilibrium of the mass whose equilibrium position is x (and whose numerical label is n). ξ is a function of t too, of course, but we won't bother writing the t dependence yet. But eventually we'll write the displacement as $\xi(x, t)$.

Let $\Delta x \equiv x_n - x_{n-1}$ be the (equal) spacing between the equilibrium positions of all the masses. The x_n values don't change with time, so neither does Δx. If the $n = 0$ mass is located at the origin, then the other masses are located at positions $x_n = n\Delta x$. In our new notation, the $F = ma$ equation in Eq. (42) becomes

$$\begin{aligned} m\ddot{\xi}_n &= k\xi_{n-1} - 2k\xi_n + k\xi_{n+1} \\ \Longrightarrow \quad m\ddot{\xi}(x_n) &= k\xi(x_n - \Delta x) + 2k\xi(x_n) + k\xi(x_n + \Delta x) \\ \Longrightarrow \quad m\ddot{\xi}(x) &= k\xi(x - \Delta x) + 2k\xi(x) + k\xi(x + \Delta x). \end{aligned} \tag{76}$$

In going from the second to the third line, we are able to drop the subscript n because the value of x uniquely determines which mass we're looking at. If we ever care to know the value of n, we can find it via $x_n = n\Delta x \Longrightarrow n = x/\Delta x$. Although the third line holds only for x values that are integral multiples of Δx, we will soon take the $\Delta x \to 0$ limit, in which case the equation holds for essentially all x.

We will now gradually transform Eq. (76) into a very nice result, which is called the *wave equation*. The first step actually involves going backward to the $F = ma$ form in Eq. (41). We have

$$\begin{aligned} m\frac{d^2\xi(x)}{dt^2} &= k\left[\Big(\xi(x+\Delta x) - \xi(x)\Big) - \Big(\xi(x) - \xi(x-\Delta x)\Big)\right] \\ \Longrightarrow \frac{m}{\Delta x}\frac{d^2\xi(x)}{dt^2} &= k\Delta x\left(\frac{\frac{\xi(x+\Delta x)-\xi(x)}{\Delta x} - \frac{\xi(x)-\xi(x-\Delta x)}{\Delta x}}{\Delta x}\right). \end{aligned} \quad (77)$$

We have made these judicious divisions by Δx for the following reason. If we let $\Delta x \to 0$ (which is indeed the case if we have $N \to \infty$ masses in the system), then we can use the definitions of the first and second derivatives to obtain (with primes denoting spatial derivatives)[6]

$$\begin{aligned} \frac{m}{\Delta x}\frac{d^2\xi(x)}{dt^2} &= (k\Delta x)\frac{\xi'(x) - \xi'(x-\Delta x)}{\Delta x} \\ &= (k\Delta x)\xi''(x). \end{aligned} \quad (78)$$

But $m/\Delta x$ is the mass density ρ. And $k\Delta x$ is known as the elastic modulus, E, which happens to have the units of force. So we obtain

$$\rho\frac{d^2\xi(x)}{dt^2} = E\xi''(x). \quad (79)$$

Note that $E \equiv k\Delta x$ is a reasonable quantity to appear here, because the spring constant k for an infinitely small piece of spring is infinitely large (because if you cut a spring in half, its k doubles, etc.). The Δx in the product $k\Delta x$ has the effect of yielding a finite and informative quantity. If various people have various lengths of springs made out of a given material, then these springs have different k values, but they all have the same E value. Basically, if you buy a spring in a store, and if it's cut from a large supply on a big spool, then the spool should be labeled with the E value, because E is a property of the material and independent of the length. k depends on the length.

Since ξ is actually a function of both x and t, let's be explicit and write Eq. (79) as

$$\boxed{\rho\frac{\partial^2\xi(x,t)}{\partial t^2} = E\frac{\partial^2\xi(x,t)}{\partial x^2}} \qquad \text{(wave equation)} \quad (80)$$

This is called the *wave equation*. This equation (or analogous equations for other systems) will appear repeatedly throughout this book. Note that the derivatives are now written as partial derivatives, because ξ is a function of two arguments. Up to the factors of ρ and E, the wave equation is symmetric in x and t.

The second *time* derivative on the lefthand side of Eq. (80) comes from the "a" in $F = ma$. The second *space* derivative on the righthand side comes from the fact that it is the *differences* in the lengths of two springs that yields the net force, and each of these lengths is itself the *difference* of the positions of two masses. So it is the difference of the differences that we're concerned with. In other words, the second derivative.

[6]There is a slight ambiguity here. Is the $(\xi(x+\Delta x)-\xi(x))\Delta x$ term in Eq. (77) equal to $\xi'(x)$ or $\xi'(x+\Delta x)$? Or perhaps $\xi'(x+\Delta x/2)$? It doesn't matter which we pick, as long as we use the same convention for the $(\xi(x)-\xi(x-\Delta x))\Delta x$ term. The point is that Eq. (78) contains the first derivatives at two points (whatever they may be) that differ by Δx, and the difference of these yields the second derivative.

How do we solve the wave equation? Recall that in the finite-N case, the strategy was to guess a solution of the form (using ξ now instead of x),

$$\begin{pmatrix} \vdots \\ \xi_{n-1} \\ \xi_n \\ \xi_{n+1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ a_{n-1} \\ a_n \\ a_{n+1} \\ \vdots \end{pmatrix} e^{i\omega t}. \tag{81}$$

If we relabel $\xi_n \to \xi(x_n, t) \to \xi(x, t)$, and $a_n \to a(x_n) \to a(x)$, we can write the guess in the more compact form,

$$\xi(x,t) = a(x)e^{i\omega t}. \tag{82}$$

This is actually an infinite number of equations (one for each x), just as Eq. (81) is an infinite number of equations (one for each n). The $a(x)$ function gives the amplitudes of the masses, just as the original normal mode vector $(A_1, A_2, A_3, \ldots)$ did. If you want, you can think of $a(x)$ as an infinite-component vector.

Plugging this expression for $\xi(x,t)$ into the wave equation, Eq. (80), gives

$$\begin{aligned} \rho \frac{\partial^2}{\partial t^2}\left(a(x)e^{i\omega t}\right) &= E\frac{\partial^2}{\partial x^2}\left(a(x)e^{i\omega t}\right) \\ \Longrightarrow \quad -\omega^2 \rho\, a(x) &= E\frac{d^2}{dx^2}a(x) \\ \Longrightarrow \quad \frac{d^2}{dx^2}a(x) &= -\frac{\omega^2\rho}{E}a(x). \end{aligned} \tag{83}$$

But this is our good ol' simple-harmonic-oscillator equation, so the solution is

$$a(x) = Ae^{\pm ikx} \qquad \text{where} \quad \boxed{k \equiv \omega\sqrt{\frac{\rho}{E}}} \tag{84}$$

k is called the *wave number*. It is usually defined to be a positive number, so we've put in the $\pm$ by hand. Unfortunately, we've already been using k as the spring constant, but there are only so many letters! The context (and units) should make it clear which way we're using k. The wave number k has units of

$$[k] = [\omega]\sqrt{\frac{[\rho]}{[E]}} = \frac{1}{\text{s}}\sqrt{\frac{\text{kg/m}}{\text{kg m/s}^2}} = \frac{1}{\text{m}}\,. \tag{85}$$

So kx is dimensionless, as it should be, because it appears in the exponent in Eq. (84).

What is the physical meaning of k? If λ is the wavelength, then the kx exponent in Eq. (84) increases by 2π whenever x increases by λ. So we have

$$k\lambda = 2\pi \quad \Longrightarrow \quad k = \frac{2\pi}{\lambda}\,. \tag{86}$$

If k were just equal to $1/\lambda$, then it would equal the number of wavelengths (that is, the number of spatial oscillations) that fit into a unit length. With the 2π, it instead equals the number of radians of spatial oscillations that fit into a unit length.

Using Eq. (84), our solution for $\xi(x,t)$ in Eq. (82) becomes

$$\xi(x,t) = a(x)e^{i\omega t} = Ae^{i(\pm kx + \omega t)}. \tag{87}$$

As usual, we could have done all this with an $e^{-i\omega t}$ term in Eq. (81), because only the square of ω came into play (ω is generally assumed to be positive). So we really have the four different solutions,

$$\xi(x,t) = Ae^{i(\pm kx \pm \omega t)}. \tag{88}$$

The most general solution is the sum of these, which gives

$$\xi(x,t) = A_1 e^{i(kx+\omega t)} + A_1^* e^{i(-kx-\omega t)} + A_2 e^{i(kx-\omega t)} + A_2^* e^{i(-kx+\omega t)}, \tag{89}$$

where the complex conjugates appear because ξ must be real. There are many ways to rewrite this expression in terms of trig functions. Depending on the situation you're dealing with, one form is usually easier to deal with than the others, but they're all usable in theory. Let's list them out and discuss them. In the end, each form has four free parameters. We saw above in the third remark at the end of Section 2.3.1 why four was the necessary number in the discrete case, but we'll talk more about this below.

- If we let $A_1 \equiv (B_1/2)e^{i\phi_1}$ and $A_2 \equiv (B_2/2)e^{i\phi_2}$ in Eq. (89), then the imaginary parts of the exponentials cancel, and we end up with

$$\boxed{\xi(x,t) = B_1 \cos(kx + \omega t + \phi_1) + B_2 \cos(kx - \omega t + \phi_2)} \tag{90}$$

  The interpretation of these two terms is that they represent *traveling waves*. The first one moves to the left, and the second one moves to the right. We'll talk about traveling waves below.

- If we use the trig sum formulas to expand the previous expression, we obtain

$$\boxed{\xi(x,t) = C_1 \cos(kx + \omega t) + C_2 \sin(kx + \omega t) + C_3 \cos(kx - \omega t) + C_4 \sin(kx - \omega t)} \tag{91}$$

  where $C_1 = B_1 \cos\phi_1$, etc. This form has the same interpretation of traveling waves. The sines and cosines are simply $90°$ out of phase.

- If we use the trig sum formulas again and expand the previous expression, we obtain

$$\boxed{\xi(x,t) = D_1 \cos kx \cos\omega t + D_2 \sin kx \sin\omega t + D_3 \sin kx \cos\omega t + D_4 \cos kx \sin\omega t} \tag{92}$$

  where $D_1 = C_1 + C_3$, etc. These four terms are four different *standing waves*. For each term, the masses all oscillate in phase. All the masses reach their maximum position at the same time (the $\cos\omega t$ terms at one time, and the $\sin\omega t$ terms at another), and they all pass through zero at the same time. As a function of time, the plot of each term just grows and shrinks as a whole. The equality of Eqs. (91) and (92) implies that any traveling wave can be written as the sum of standing waves, and vice versa. This isn't terribly obvious; we'll talk about it below.

- If we collect the $\cos\omega t$ terms together in the previous expression, and likewise for the $\sin\omega t$ terms, we obtain

$$\boxed{\xi(x,t) = E_1 \cos(kx + \beta_1)\cos\omega t + E_2 \cos(kx + \beta_2)\sin\omega t} \tag{93}$$

  where $E_1 \cos\beta_1 = D_1$, etc. This form represents standing waves (the $\cos\omega t$ one is $90°$ ahead of the $\sin\omega t$ one in time), but they're shifted along the x axis due to the β phases. The spatial functions here could just as well be written in terms of sines, or one sine and one cosine. This would simply change the phases by $\pi/2$.

- If we collect the $\cos kx$ terms together in Eq. (92) and likewise for the $\sin kx$ terms, we obtain

$$\boxed{\xi(x,t) = F_1\cos(\omega t+\gamma_1)\cos kx + F_2\cos(\omega t+\gamma_2)\sin kx} \tag{94}$$

where $F_1\cos\gamma_1 = D_1$, etc. This form represents standing waves, but they're not 90° separated in time in this case, due to the γ phases. They are, however, separated by 90° (a quarter wavelength) in space. The time functions here could just as well be written in terms of sines.

Remarks:

1. If there are no walls and the system extends infinitely in both directions (actually, infinite extent in just one direction is sufficient), then ω can take on any value. Eq. (84) then says that k is related to ω via $k = \omega\sqrt{\rho/E}$. We'll look at the various effects of boundary conditions in Chapter 4.
2. The fact that each of the above forms requires four parameters is probably most easily understood by looking at the first form given in Eq. (90). The most general wave with a given frequency ω consists of two oppositely-traveling waves, each of which is described by two parameters (magnitude and phase). So two times two is four.

   You will recall that for each of the modes in the $N = 2$ and $N = 3$ cases we discussed earlier (and any other value of N, too), only two parameters were required: an overall factor in the amplitudes, and a phase in time. Why only two there, but four now? The difference is due to the fact that we had walls in the earlier cases, but no walls now. (The difference is *not* due to the fact that we're now dealing with infinite N.) The effect of the walls (actually, only one wall is needed) is most easily seen by working with the form given in Eq. (92). Assuming that one wall is located at $x = 0$, we see that the two $\cos kx$ terms can't be present, because the displacement must always be zero at $x = 0$. So $D_1 = D_4 = 0$, and we're down to two parameters. We'll have much more to say about such matters in Chapter 4.
3. Remember that the above expressions for $\xi(x,t)$, each of which contains four parameters, represent the general solution for a *given mode* with frequency ω. If the system is undergoing arbitrary motion, then it is undoubtedly in a linear combination of many different modes, perhaps even an infinite number. So four parameters certainly don't determine the system. We need four times the number of modes, which might be infinite. ♣

Traveling waves

Consider one of the terms in Eq. (91), say, the $\cos(kx-\omega t)$ one. Let's draw the plot of $\cos(kx-\omega t)$, as a function of x, at two times separated by Δt. If we arbitrarily take the lesser time to be $t = 0$, the result is shown in Fig. 16. Basically, the left curve is a plot of $\cos kx$, and the right curve is a plot of $\cos(kx-\phi)$, where ϕ happens to be $\omega\Delta t$. It is shifted to the right because it takes a larger value of x to obtain the same phase.

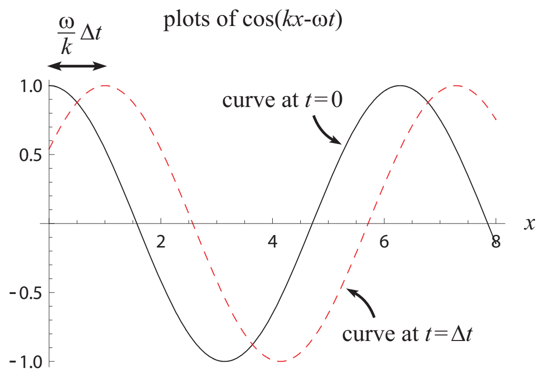

Figure 16

What is the horizontal shift between the curves? We can answer this by finding the distance between the maxima, which are achieved when the argument $kx - \omega t$ equals zero (or a multiple of 2π). If $t = 0$, then we have $kx - \omega \cdot 0 = 0 \implies x = 0$. And if $t = \Delta t$, then we have $kx - \omega \cdot \Delta t = 0 \implies x = (\omega/k)\Delta t$. So $(\omega/k)\Delta t$ is the horizontal shift. It takes a time of Δt for the wave to cover this distance, so the velocity of the wave is

$$v = \frac{(\omega/k)\Delta t}{\Delta t} \implies \boxed{v = \frac{\omega}{k}} \tag{95}$$

Likewise for the $\sin(kx - \omega t)$ function in Eq. (91). Similarly, the velocity of the $\cos(kx + \omega t)$ and $\sin(kx + \omega t)$ curves is $-\omega/k$.

We see that the wave $\cos(kx - \omega t)$ keeps its shape and travels along at speed ω/k. Hence the name "traveling wave." But note that none of the masses are actually moving with this speed. In fact, in our usual approximation of small amplitudes, the actual velocities of the masses are very small. If we double the amplitudes, then the velocities of the masses are doubled, but the speed of the waves is still ω/k.

As we discussed right after Eq. (92), the terms in that equation are standing waves. They don't travel anywhere; they just expand and contract in place. All the masses reach their maximum position at the same time, and they all pass through zero at the same time. This is certainly *not* the case with a traveling wave. Trig identities of the sort, $\cos(kx - \omega t) = \cos kx \cos \omega t + \sin kx \sin \omega t$, imply that any traveling wave can be written as the sum of two standing waves. And trig identities of the sort, $\cos kx \cos \omega t = \big(\cos(kx - \omega t) + \cos(kx + \omega t)\big)/2$, imply that any standing wave can be written as the sum of two opposite traveling waves. The latter of these facts is reasonably easy to visualize, but the former is trickier. You should convince yourself that it works.

A more general solution

We'll now present a much quicker method of finding a much more general solution (compared with our sinusoidal solutions above) to the wave equation in Eq. (80). This is a win-win combination.

From Eq. (84), we know that $k = \omega\sqrt{\rho/E}$. Combining this with Eq. (95) gives $\sqrt{E/\rho} = \omega/k = v \implies E/\rho = v^2$. In view of this relation, if we divide the wave equation in Eq. (80) by ρ, we see that it can be written as

$$\boxed{\frac{\partial^2 \xi(x,t)}{\partial t^2} = v^2 \frac{\partial^2 \xi(x,t)}{\partial x^2}} \tag{96}$$

Consider now the function $f(x - vt)$, where f is an *arbitrary* function of its argument. (The function $f(x + vt)$ will work just as well.) There is no need for f to even vaguely resemble a sinusoidal function. What happens if we plug $\xi(x,t) \equiv f(x - vt)$ into Eq. (96)? Does it satisfy the equation? Indeed it does, as we can see by using the chain rule. In what follows, we'll use the notation f'' to denote the second derivative of f. In other words, $f''(x - vt)$ equals $d^2 f(z)/dz^2$ evaluated at $z = x - vt$. (Since f is a function of only one variable, there is no need for any partial derivatives.) Eq. (96) then becomes (using the chain rule on the left, and also on the right in a trivial sense)

$$\begin{aligned} \frac{\partial^2 f(x - vt)}{\partial t^2} &\stackrel{?}{=} v^2 \frac{\partial^2 f(x - vt)}{\partial x^2} \\ \iff (-v)^2 f''(x - vt) &\stackrel{?}{=} v^2 \cdot (1)^2 f''(x - vt), \end{aligned} \tag{97}$$

which is indeed true.

There is a fairly easy way to see graphically why any function of the form $f(x - vt)$ satisfies the wave equation in Eq. (96). The function $f(x-vt)$ represents a wave moving to the right at speed v. This is true because $f(x_0 - vt_0) = f\big((x_0 + v\Delta t) - v(t_0 + \Delta t)\big)$, which says that if you increase t by Δt, then you need to increase x by $v\Delta t$ in order to obtain the same value of f. This is exactly what happens if you take a curve and move it to the right at speed v. This is basically the same reasoning that led to Eq. (95).

We now claim that any curve that moves to the right (or left) at speed v satisfies the wave equation in Eq. (96). Consider a closeup view of the curve near a given point x_0, at two nearby times separated by Δt. The curve is essentially a straight line in the vicinity of x_0, so we have the situation shown in Fig. 17.

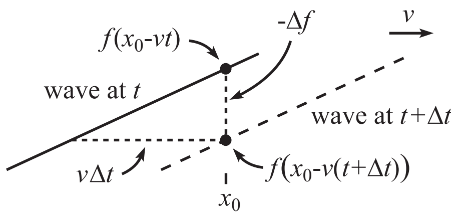

Figure 17

The solid line shows the curve at some time t, and the dotted line shows it at time $t+\Delta t$. The slope of the curve, which is by definition $\partial f/\partial x$, equals the ratio of the lengths of the legs in the right triangle shown. The length of the vertical leg equals the magnitude of the change Δf in the function. Since the change is negative here, the length is $-\Delta f$. But by the definition of $\partial f/\partial t$, the change is $\Delta f = (\partial f/\partial t)\Delta t$. So the length of the vertical leg is $-(\partial f/\partial t)\Delta t$. The length of the horizontal leg is $v\Delta t$, because the curve moves at speed v. So the statement that $\partial f/\partial x$ equals the ratio of the lengths of the legs is

$$\frac{\partial f}{\partial x} = \frac{-(\partial f/\partial t)\Delta t}{v\Delta t} \quad \Longrightarrow \quad \frac{\partial f}{\partial t} = -v\frac{\partial f}{\partial x}\,. \tag{98}$$

If we had used the function $f(x+vt)$, we would have obtained $\partial f/\partial t = v(\partial f/\partial x)$. Of course, these results follow immediately from applying the chain rule to $f(x \pm vt)$. But it's nice to also see how they come about graphically.

Eq. (98) then implies the wave equation in Eq. (96), because if we take $\partial/\partial t$ of Eq. (98), and use the fact that partial differentiation commutes (the order doesn't matter), we obtain

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\frac{\partial f}{\partial t}\right) &= -v\frac{\partial}{\partial t}\left(\frac{\partial f}{\partial x}\right) \\
\Longrightarrow \quad \frac{\partial^2 f}{\partial t^2} &= -v\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial t}\right) \\
&= -v\frac{\partial}{\partial x}\left(-v\frac{\partial f}{\partial x}\right) \\
&= v^2\frac{\partial^2 f}{\partial x^2}\,,
\end{aligned} \tag{99}$$

where we have used Eq. (98) again to obtain the third line. This result agrees with Eq. (96), as desired.

Another way of seeing why Eq. (98) implies Eq. (96) is to factor Eq. (96). Due to the fact that partial differentiation commutes, we can rewrite Eq. (96) as

$$\left(\frac{\partial}{\partial t} - v\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right) f = 0. \tag{100}$$

We can switch the order of these "differential operators" in parentheses, so either of them can be thought of acting on f first. Therefore, if either operator yields zero when applied to f, then the lefthand side of the equation equals zero. In other words, if Eq. (98) is true (with either a plus or a minus on the right side), then Eq. (96) is true.

We have therefore seen (in various ways) that *any* arbitrary function that takes the form of $f(x - vt)$ satisfies the wave equation. This seems too simple to be true. Why did we go through the whole procedure above that involved guessing a solution of the form $\xi(x, t) = a(x)e^{i\omega t}$? Well, that has always been our standard procedure, so the question we should be asking is: Why does an arbitrary function $f(x - vt)$ work?

Well, we gave a few reasons in Eqs. (97) and (98). But here's another reason, one that relates things back to our original sinusoidal solutions. $f(x - vt)$ works because of a combination of *Fourier analysis* and *linearity*. Fourier analysis says that any (reasonably well-behaved) function can be written as the integral (or discrete sum, if the function is periodic) of exponentials, or equivalently sines and cosines. That is,

$$f(z) = \int_{-\infty}^{\infty} C(r)e^{irz}dr. \tag{101}$$

Don't worry about the exact meaning of this; we'll discuss it at great length in the following chapter. But for now, you just need to know that any function $f(z)$ can be considered to be built up out of e^{irz} exponential functions. The coefficient $C(r)$ tells you how much of the function comes from a e^{irz} term with a particular value of r.

Let's now pretend that we haven't seen Eq. (97), but that we *do* know about Fourier analysis. Given the result in Eq. (101), if someone gives us the function $f(x-vt)$ out of the blue, we can write it as

$$f(x - vt) = \int_{-\infty}^{\infty} C(r)e^{ir(x-vt)}dr. \tag{102}$$

But $e^{ir(x-vt)}$ can be written as $e^{i(kx-\omega t)}$, where $k \equiv r$ and $\omega \equiv rv$. Since these values of k and ω satisfy $\omega/k = v$, and hence satisfy Eq. (84) (assuming that v has been chosen to equal $\sqrt{E/\rho}$), we know that all of these $e^{ir(x-vt)}$ terms satisfy the wave equation, Eq. (80). And since the wave equation is *linear* in ξ, it follows that *any* sum (or integral) of these exponentials also satisfies the wave equation. Therefore, in view of Eq. (102), we see that any arbitrary function $f(x - vt)$ satisfies the wave equation. As stated above, both Fourier analysis and linearity are essential in this result.

Fourier analysis plays an absolutely critical role in the study of waves. In fact, it is so important that we'll spend all of Chapter 3 on it. We'll then return to our study of waves in Chapter 4. We'll pick up where we left off here.