Introduction to Applied Mathematics MAT00003C

February 2023

Niall MacKay niall.mackay@york.ac.uk, room G/N/125

Contents

Foreword

> *Strange to say, it is to the advantage of realism that mathematicians customarily replace the actual world by various idealized models. For they choose models that can be analyzed with ease; and thus they are free to think about the resemblances or misfits between the model and the actual world. If, with a solemn feeling of the importance of things as they really are, we were to admit the irregularities of the actual world into the statement of our problems, we should in consequence have to attend to enormous elaborations of mathematics in the process of solution, whereby our attention would for a long time be distracted away from the actual world.*
>
> Lewis Fry Richardson

What you will *not* find in the ensuing pages is complex models, with many parameters and variables,[1] tuned to provide precise real-world predictions. Such models are very important, and they are certainly the final destination of applied mathematics. But this is an *introduction* to modelling, and we do not want to become mired in detail.

Some of the simple models we are going to look at will (you may well think) go beyond 'simple' and become 'simplistic' (that is, oversimplified so as to conceal or distort difficulties). That is not the intention, which is rather to begin to get a sense of some of the essential dynamical processes of the real world, such as growth (unbounded or within limits), decay, cycles and oscillations.

For example:

'I understand negative feedback, a "vicious circle",' says a chemist/doctor/biologist. 'Things decay in proportion to themselves, but there's always a bit left; the substance never entirely disappears.'

'Well, in fact it *can* disappear entirely – it's just that the proportionality needs to be a bit more subtle, more indirect,' says the mathematician.

'What do you mean?'

'Well, imagine a water droplet evaporating. In a sense it decays in proportion to itself, but the water *volume* evaporates in proportion not to itself but to its *surface area*. With *this* process the droplet doesn't just decay, it disappears completely.'

'How do you see that?'

– and the answer is among our simplest cases in this course, Example 8.

[1] We call constant quantities 'parameters' and (typically time-)varying quantities 'variables'. Back in the real-world, most things change; it's just a matter of the timescale.

1 Dimensional Analysis and Scaling

To begin with, you should have a working knowledge of the first part of the module. In particular, **you will need to know** the dimensions (in terms of M, L, T) of area (L^2), volume (L^3), speed (LT^{-1}), acceleration (LT^{-2}), frequency and angular frequency (T^{-1}, because cycles and angles are numbers) momentum (MLT^{-1}), force (MLT^{-2}), energy (ML^2T^{-2}, which you can deduce from work's being force multiplied by distance, or from kinetic energy) and density (ML^{-3}).

What you should be able to do by the end of this section,
after working through the notes, examples and problems sheets:

- Write a quantity as a product of powers of others, equate the underlying dimensions, and solve the resulting linear equations for the powers;
- thereby identify the equation as a relationship between **dimensionless numbers**
- Identify and apply resulting **scaling laws** to understand how one quantity changes when another is scaled up or down;
- Recognize scaling laws through the linearity of log-log plots;
- Identify the **dimensionless numbers** (or **dimensionless groups**) in a problem;
- Write an equation in terms of them ('non-dimensionalizing' it);
- Know and apply Buckingham's Π theorem.

1.1 Two stories

Simply by thinking about units and dimensions we can obtain surprising results about important matters.

Video:
G I Taylor and the bomb

Example 1: The atom bomb

Shortly after World War Two, the US Atomic Energy Commission released a film of the 1945 Trinity atom bomb test. The energy yield remained secret, having been estimated by the Americans only with difficulty. So they were most surprised when the British fluid dynamicist Geoffrey Taylor published, in 1950, an accurate estimate merely by studying the AEC's pictures [2].

At its simplest, the crucial result follows on dimensional grounds alone. Suppose that, at time t after detonation, the radius r of the blast wave depends only on time t, air density ρ and the energy E released in the blast. (That is, once the blast wave is propagating it knows nothing of the nature of the explosion which caused it.)

Figure 1: the Trinity atom bomb test.

Suppose

$$r = Ct^a E^b \rho^c,$$

where a, b, c are real constants, and C too is a number, with no dimensions. We write '$[Q]$' for 'the dimensions of Q are', and M, L, T for 'mass', 'length' and 'time'. Energy is a mass multiplied by the square of a speed, $[E] = ML^2T^{-2}$, and $[\rho] = ML^{-3}$. So

$$[r] = L = [t^a E^b \rho^c] = T^a (ML^2T^{-2})^b (ML^{-3})^c,$$

and then, for the dimensions to match, we must have

$$L : 1 = 2b - 3c, \qquad T : 0 = a - 2b, \qquad M : 0 = b + c.$$

Solving these, $c = -b$ and then $1 = 2b + 3b$ so that $b = 1/5$. Then $c = -1/5$ and $a = 2/5$, so

$$r = C\left(\frac{Et^2}{\rho}\right)^{\frac{1}{5}}, \tag{1.1}$$

where C is a dimensionless constant, universal for such waves, which can be estimated from conventional explosive blasts. Used with pictures such as that in Fig. 1, which helpfully gives length and time scales, this formula enabled Taylor to estimate E. The calculation is striking especially for the unexpected appearance of the one-fifth power—after all, when we clap our hands we are used to the effects propagating very differently, with a fixed 'speed of sound' v, and $r \propto vt$. □

Video: Kolmogorov scaling

Example 2: Kolmogorov scaling of ocean waves

Imagine a storm at sea. A few days ago the sea was flat calm. A light wind began to stir tiny ripples on the water. The wind grew, and, as it did so, so did the waves. Now the storm is in full swing, and staying that way: there are waves of all lengths, right up to some monsters of many meters' height and tens of meters' length.

Figure 2: Storm at sea.

How is all that energy distributed among the different wavelengths λ? We want to know $e(\lambda)$, where $e(\lambda)\,\delta\lambda$ is the amount of energy per unit mass (of water) which is in the form of waves of length between λ and $\lambda + \delta\lambda$. So

$$[e\,\delta\lambda] = [e]L = \frac{ML^2T^{-2}}{M} = L^2T^{-2} \qquad \Rightarrow \qquad [e] = LT^{-2}.$$

We assume that this depends only on two things: the wavelength λ itself, with $[\lambda] = L$, and the way in which energy somehow passes up the wavelengths, building tiny ripples up into great ocean waves. For this we define an 'energy flow' per unit mass s, whose dimensions are energy per unit mass per unit time, and which (the crucial physical assumption) is independent of wavelength. Then

$$[s] = \frac{ML^2T^{-2}}{MT} = L^2T^{-3}.$$

Then, if

$$e(\lambda) = C's^a\lambda^b,$$

where C' is again a (dimensionless) constant and a, b are real powers to be determined, we have

$$[e] = LT^{-2} = [C's^a\lambda^b] = (L^2T^{-3})^aL^b.$$

So

$$L : 1 = 2a + b, \qquad T : -2 = -3a.$$

Solving these, $a = 2/3$ and $b = 1 - 4/3 - 3 = -1/3$. So

$$e = C' \left(\frac{s^2}{\lambda} \right)^{1/3}. \tag{1.2}$$

So in the picture in Figure 2, for example, there will be about three times as much energy in waves of length 1-2 m (metres) as in waves of length 27-28 m, since if $\lambda = 1m$ and $\lambda' = 27m$ then $e/e' = (\lambda'/\lambda)^{1/3} = 27^{1/3} = 3$. □

1.2 Scaling laws

Video: Scaling laws

What we've learned so far suggests that there may be a special role in nature, among all the functions you've seen – polynomial, exponential, logarithmic, trigonometric, hyperbolic – for the humble power function $y = Cx^a$, where y is a simple monomial in x. And there is indeed something special about such functions: they are the *only* functions whose form does not alter when we re-scale the variables (see appendix). That is, if we replace y by ky and x by lx then $y = Cx^a$ becomes

$$ky = C(lx)^a \qquad \text{or} \qquad y = \left(\frac{Cl^a}{k} \right) x^a.$$

The constant has changed, but the relationship is still $y =$(something times)x^a: the form of the function has not changed.

So if you believe that a physical law takes the same functional form at different scales – of length, of mass; whatever – then it must be described by a simple power. When one quantity is proportional to a simple power of another, we call this relationship a **scaling law**.

Look at (1.1), our scaling law for an atomic bomb. Suppose we're really only interested in how r increases with t for a particular bomb. Think of the scaling as

$$r = Kt^{2/5},$$

where K combines the dependence on C, E and ρ. (In fact $K = CE^{1/5}\rho^{-1/5}$.) This is dimensionally rather weird: it's better to include our units, to think of it as

$$\frac{r}{r_0} = K' \left(\frac{t}{t_0} \right)^{2/5} \tag{1.3}$$

where r_0 is our unit of length (such as one metre) and t_0 is our unit of time (such as one second). That way, r/r_0 and t/t_0 are automatically dimensionless numbers and so K' is also dimensionless.

(Let's check that: we have

$$r = (K' r_0 t_0^{-2/5}) t^{2/5} = C \left(\frac{E t^2}{\rho} \right)^{\frac{1}{5}} \qquad \text{so} \qquad K' = C t_0^{2/5} r_0^{-1} E^{1/5} \rho^{-1/5}$$

and then

$$[K'] = T^{2/5} L^{-1} (ML^2 T^{-2})^{1/5} (ML^{-3})^{-1/5} = 1$$

and K' is dimensionless. So no nasty surprises.)

A scaling law in the form (1.3) makes it absolutely clear that one is free to change the *units* (*e.g.* metres to miles, seconds to hours) without changing the form of the equation (because a scaling law describes a phenomenon which looks the same at different scales, or using different units.) It is also the *only* form in which we can take its logarithm (which must have for its argument a number, not a dimensionful quantity[2]):

$$\log \left(\frac{r}{r_0} \right) = \log K' + \frac{2}{5} \log \left(\frac{t}{t_0} \right). \tag{1.4}$$

Figure 3 is from Taylor's original paper, and is how he verified his result. Kolmogorov's 1940s predictions were first verified in the 1950s – see for example [4]. The straight downward slope is clearly visible in Figure 4, until on the right, at very small wavelengths, the viscosity of water becomes important and ϕ falls off even more quickly.

A change of units r_0 and t_0 appears in (1.4) as a shift of both axes, a translation of the origin in the log-log plot – where the lack of preferred units appears as the fact that a straight line has no preferred origin.

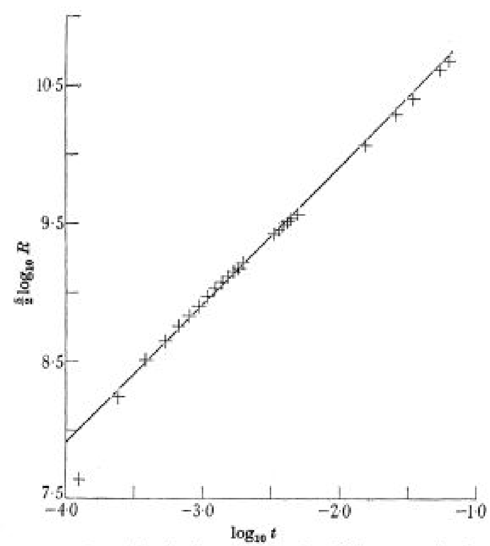


FIGURE 1. Logarithmic plot showing that $R^{\frac{5}{2}}$ is proportional to t.

[2]OK, let's suppose you disagree. You say "Of course I can take the logarithm of a mass. Look, I weigh 60 kilos. I put '60' into my calculator and press 'log'." But what you've just done is *not* to compute the logarithm of a mass. You've computed the logarithm of the number 60, which was the ratio of your mass to a 1 kilogram mass.

Figure 3: G. I. Taylor's log-log plot of r against t for the atom bomb blast.

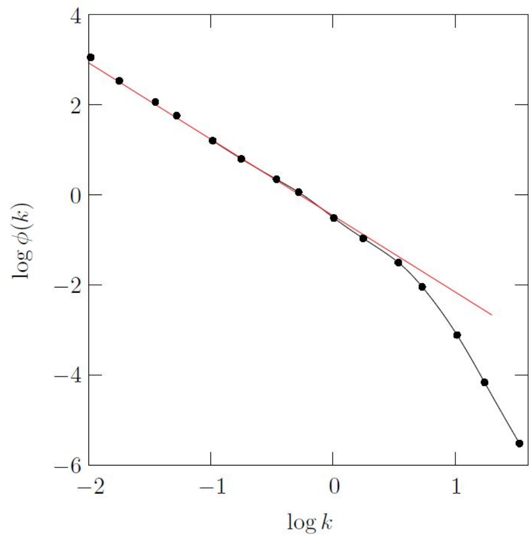


Figure 4: Log-log plot of Kolmogorov scaling. ϕ is here given more conventionally in terms of wavenumber $k = 1/\lambda$, which produces a slope of $-5/3$ instead of $-1/3$. The fall-off on the right is due to the onset of viscosity at very small lengths. From Bob Stewart and Chris Jarrett, *Kolmogorov, turbulence and British Columbia*.

Video: Metabolic scaling

Example 3: Metabolic scaling

Mammalian physiology is remarkably similar over many orders of magnitude: in Figure 5, the mouse and the bear are similar-looking given that one weighs about 10,000 times the other.

Figure 5: a dormouse (mass $\sim$ 20g) and a bear (200kg).

Is there a power law which describes their metabolic rate (their power consumption) as a function of their mass? We might begin by noticing that, to the extent to which the bear is a scaled-up mouse, its surface area will be about $(10000)^{2/3}$ times that of the mouse, since the two animals presumably have similar densities, and mass$\sim$length^3 while surface area$\sim$length^2.[3] So we might guess that, where μ is metabolic rate and m is mass,

$$\mu = Cm^{2/3}.$$

(Biologists call such scaling **allometry**, in contrast to self-similar **isometry**.) But look at Figure 6. For the eutheria (the true mammals, the 'good beasts') the power is about 0.72, although there's a bit of curvature – it's not quite a straight line. For the marsupials, which are less physiologically variable, it's a strikingly straight line at a power of 0.75. This is the (purely empirical) **Kleiber's law**: the power is not 2/3 but 3/4.

The (approximate) relationship

$$\mu = Cm^{3/4}$$

tells us how the bear's and the mouse's energy requirements are related:

$$\frac{\mu_{\text{bear}}}{\mu_{\text{mouse}}} = \left(\frac{m_{\text{bear}}}{m_{\text{mouse}}}\right)^{3/4},$$

so that if the ratio of their masses is 10000 then the ratio of their energy usage is about 1000. Notice that there is no need to compute C, which remains unknown. We simply divide left- and right-hand sides of the equation for the two instances, and one ratio is then a power of the other. A scaling law tells us how things scale up (or down)!

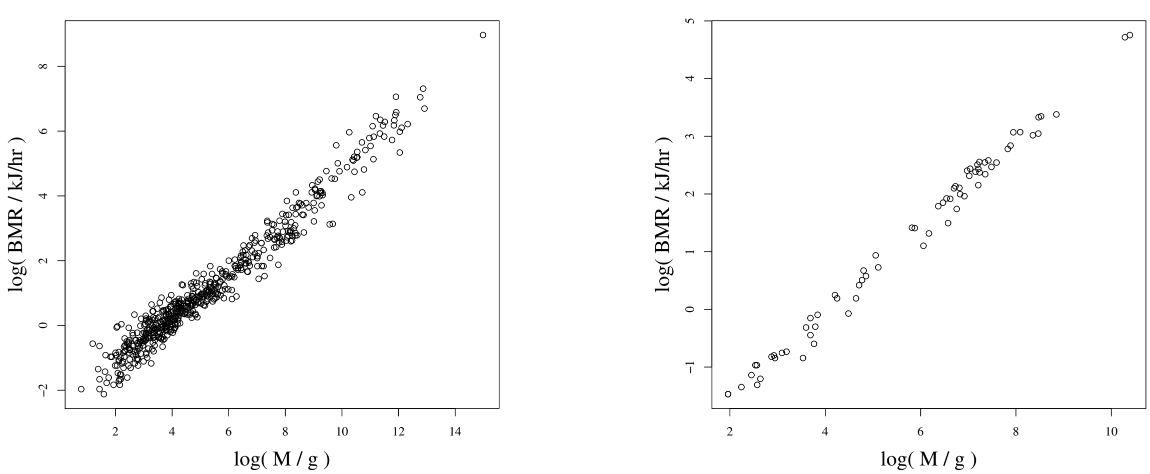


Figure 6: How metabolic rate μ scales with mass m for: (a) Eutheria: $\mu \sim m^{0.72\pm0.01}$; (a) Marsupials: $\mu \sim m^{0.75\pm0.01}$. Data from B. K. McNab, 'An analysis of the factors that influence the level and scaling of mammalian BMR', *Comparative Biochemistry and Physiology* **A151** (2008) 5-28.

[3] Read $\sim$ as 'scales like'.

What could possibly explain this 3/4 power scaling? We do not, in this case, have an easy argument from dimensional analysis. (The nearest I know of is [5].) The answer is still controversial, but one possibility [6] follows from noticing that mammalian physiology – our metabolism, at least, although not our skeleton or shape – is mostly about branching networks: our blood vessels and our lungs. Our lungs, for example, are somewhere between two- and three-dimensional: it is their surfaces which allow us to absorb oxygen, but they largely fill the space of the lung cavities. So if a grown man is twice the height of a small boy, would you expect his lung capacity to be $4 = 2^2$ or $8 = 2^3$ times larger? Probably somewhere in between. □

1.3 Dimensionless numbers

We've been solving systems of simultaneous linear equations for our powers $a, b, c, \ldots$. We know that such systems of equations sometimes have more than one solution: instead of a point, the solution may be a line, plane or higher-dimensional subspace. What happens to our dimensional analysis then? In fact that's where it *really* gets interesting...

Video: Swimming microorganisms

Example 4: Swimming micro-organisms

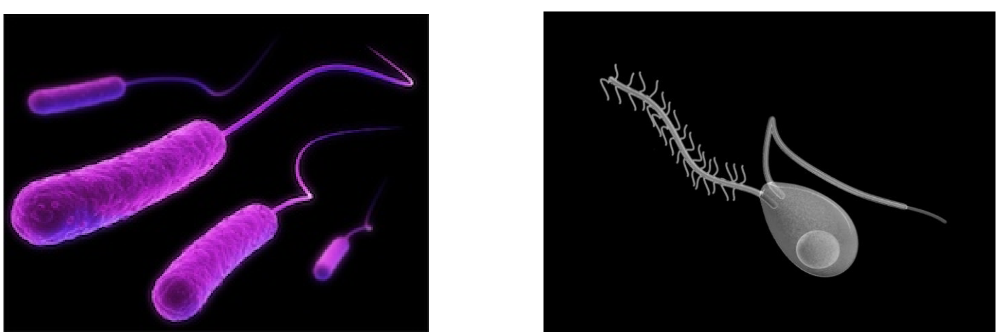


Figure 7: Two swimming micro-organisms with flagellae: (a) *E. Coli* bacteria, (b) plant fungus zoospores.

A creature of length-scale l moving at speed u swims in a fluid of density ρ and **kinematic viscosity** ν, which has dimensions $[\nu] = L^2T^{-1}$.[4] How does the drag force F it experiences depend upon ρ, ν, l and u?

We write

$$F = C\rho^a \nu^b l^c u^d.$$

[4] Viscosity is a measure of a fluid's tendency to resist deformation, and corresponds to our intuitive idea of the thickness or stickiness of a liquid. So, for example, treacle is thicker, and has a higher viscosity, than water.

We have

$$[F] = MLT^{-2}, \quad [\rho] = ML^{-3}, \quad [\nu] = L^2T^{-1}, \quad [l] = L, \quad [u] = LT^{-1}.$$

So

$$\begin{aligned} M: &\quad 1 = a \\ L: &\quad 1 = -3a + 2b + c + d \\ T: &\quad -2 = -b - d. \end{aligned}$$

Then $a = 1$ and $b + d = 2$, $2b + c + d = 4$. But these do not have a unique solution: we must leave one power undetermined. Let's make it (quite arbitrarily) b. Then $c = d = 2 - b$. So

$$F = C\rho l^2 u^2 \left(\frac{\nu}{ul}\right)^b,$$

where b can take any value. Why is this? Notice that

$$\left[\frac{\nu}{ul}\right] = \frac{L^2T^{-1}}{LT^{-1}L} = 1.$$

This quantity is dimensionless! Of course it must be, for we can raise it to an arbitrary power without changing the dimensions of F. And if we can raise it to an arbitrary power, then we can combine these (as a Taylor series) to make any function we like. In fact the most general solution is

$$F = \rho l^2 u^2 f\left(\frac{ul}{\nu}\right), \tag{1.5}$$

where f could be any function whatsoever (and we've absorbed the constant C into it). We've also used ul/ν rather than its reciprocal because this quantity is rather famous: it is the **Reynolds number**. Such **dimensionless numbers** or **dimensionless groups** are fundamental in applied mathematics.

There are many different possibilities for the quantity which multiplies f and which has the dimensions of force, $[\rho l^2 u^2] = MLT^{-2}$, because it can be multiplied by any power of the Reynolds number and still have the dimensions of force. For example, if we'd let d (rather than b) be arbitrary, we'd have arrived at

$$F = C\rho\nu^2 \left(\frac{ul}{\nu}\right)^d.$$

The dimensionless number is unchanged, and $\rho\nu^2 = \rho l^2 u^2.(\nu/lu)^2$.

It is the dimensionless numbers, such as the Reynolds number, that capture the description of a phenomenon. Further, whenever in applied mathematics you find a non-monomial function – a function which is not a single, simple power – then its argument must be a dimensionless number. (To see this, imagine a polynomial or a Taylor series in x. Both are the sums of various powers of x. If x is not dimensionless then the different powers will have different dimensions and so cannot be added together.)

The nature of the fluid flow around the swimming body – the way it feels to swim – will be determined by the Reynolds number. For given u and l, high Reynolds number corresponds to small viscosity (that is, to an inviscid fluid such as water) and low Reynolds number to high viscosity (treacle).

But that's a human point of view, for typical human lengths l and speeds u. In fact it makes no sense at all to say that a dimensionful quantity is (in absolute terms) large or small. For example, a mass of 500 kg is large only relative to another mass – it would take 500 unit masses of 1 kg to balance it. And swimming in water can feel like swimming in treacle, if you're small and slow enough. The Reynolds number will be low if u and l are small enough, whatever the viscosity ν. A micro-organism experiences swimming in water much as a human would experience swimming in treacle. It requires very different techniques, often involving one or more whip-like flagellae (see Figure 7).

The function f is simple at the extremes of large and small Reynolds number. It turns out that $f(x) \sim x^{-1}$ (that is, $b = 1$ in our first expression or $d = 1$ in the second), and $F \propto \rho\nu ul$, when x is small; and $f(x) \sim$ constant ($b = 0$ or $d = 2$), with $F \propto \rho u^2 l^2$, when x is large. The first case is **Stokes drag**, applicable at high viscosity or for small and/or slow objects, and the second **Rayleigh drag**, applicable at low viscosity or for large and/or fast objects. The latter situation is what you are used to experiencing if you go swimming. The former is what a tiny swimming creature experiences in water, and what you would experience in treacle. You would not be able to swim, as you know it – rather you would need to wriggle and squirm your way through the medium – and the mathematics of such 'swimming' is complex and beautiful. □

1.4 Non-dimensionalization

Look again at (1.5). There's actually another dimensionless number there: we could write the equation as

$$\varphi = f(\lambda) \tag{1.6}$$

where

$$\varphi = \frac{F}{\rho l^2 u^2} \qquad \text{and} \qquad \lambda = \frac{ul}{\nu},$$

with both φ and λ dimensionless.

Dimensional analysis gets us as far as this. Beyond this point lies all of the interesting science, in the form of the relationship between the dimensionless numbers, here expressed by the function f.

This is true for any piece of applied mathematics, whether in physics, chemistry, biology, finance, epidemiology; whatever. Thus, if you are studying, or have arrived at, an equation in applied mathematics, (1) check that each of the terms[5] has the same dimensions, and (2) check that any function which is non-monomial – which is not a single power – has for its argument a dimensionless number.

You can then do what we did in arriving at (1.6): you can recast the equation, by a change of variables, as a relationship among dimensionless numbers. This is called 'non-dimensionalizing the equation' or 'making the equation scale-free' or even 'scaling the equation'.

Equations in applied mathematics can often look rather complex and frightening, with many parameters and variables. Non-dimensionalizing makes them look a lot simpler: all of the dimensionful parameters are bundled together into dimensionless numbers (sometimes called 'dimensionless groups') and everything looks cleaner.

Video: Freak waves

Example 5: Freak Waves

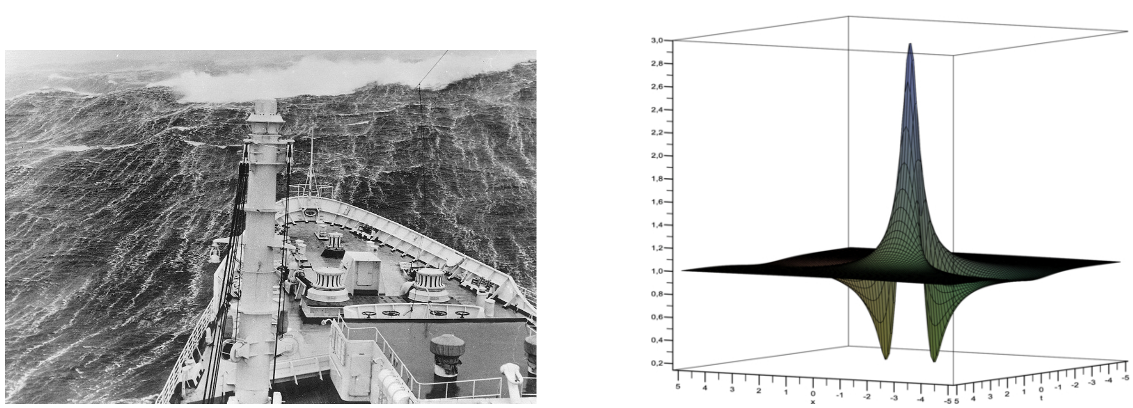


Figure 8: (a) a freak ocean wave, viewed over the bow of a ship; (b) the 'Peregrine breather' of the NLSE, due to D. H. Peregrine and here plotted by Pierre Gaillard.

This example uses the 'nonlinear Schrödinger equation' (NLSE). This equation describes aspects of ocean waves, and some of its strange solutions may be the origin of 'rogue' or 'freak' waves, which lie well outside the Kolmogorov spectrum. Figure 8(a) shows such a wave, and Figure 8(b) shows a relevant solution of the NLSE.)

[5] I mean here the summands, the things which are added or subtracted to make up the equation.

The NLSE is

$$i\left\{\frac{\partial A}{\partial t}+\frac{1}{2}\frac{\omega}{k}\frac{\partial A}{\partial x}\right\}-\frac{\omega}{8k^2}\frac{\partial^2 A}{\partial x^2}-\frac{1}{2}\omega k^2|A|^2A=0. \tag{1.7}$$

You are not meant to understand it, still less to know how to solve it. It is a **partial differential equation**: it includes derivatives with respect to both a time variable t and a space variable x. The dependent variable is a complex number $A(x,t)$, and there are parameters ω (the angular frequency) and k (the 'wave number', which is 2π divided by the wavelength) as well.

The goal is to remove the (dimensionful) parameters from the equation, leaving only three dimensionless numbers proportional to the three variables t, x and A.

To begin with we know that

$$[\omega]=T^{-1}, \qquad [k]=L^{-1}, \qquad [x]=L, \qquad [t]=T,$$

but we don't know the dimensions of A.[6] But these are easy to work out. All the terms must have the same dimensions, and so (for example)

$$\left[\frac{\partial A}{\partial t}\right]=[A]T^{-1}=\left[\omega k^2|A|^2A\right]=[A]^3T^{-1}L^{-2},$$

so

$$[A]^2L^{-2}=1 \quad \text{and} \quad [A]=L.$$

(In fact $|A|$ is the wave height.)

Next we see that the dimensionless combinations of ω, k, x, t are $[\omega t]=1$ and $[kx]=1$, so let us define two dimensionless numbers proportional to x and t,

$$\eta=kx \quad \text{and} \quad \tau=\omega t,$$

and a dimensionless dependent variable proportional to A,

$$\varphi=Ak.$$

Then, using

$$\frac{\partial}{\partial t}=\omega\frac{\partial}{\partial\tau}, \qquad \frac{\partial}{\partial x}=k\frac{\partial}{\partial\eta},$$

(1.7) becomes

$$i\left\{\omega\frac{\partial A}{\partial\tau}+\frac{1}{2}\frac{\omega}{k}k\frac{\partial A}{\partial\eta}\right\}-\frac{\omega}{8k^2}k^2\frac{\partial^2 A}{\partial\eta^2}-\frac{1}{2}\omega k^2|A|^2A=0$$

[6] Note that frequency and angular frequency have dimensions of inverse time. They are, (respectively) revolutions or radians per unit time, but both number of revolutions and number of radians are precisely that: a number. One radian, for example, is defined to be a ratio of two lengths.

or

$$i\left\{\frac{\omega}{k}\frac{\partial\varphi}{\partial\tau}+\frac{1}{2}\frac{\omega}{k}\frac{\partial\varphi}{\partial\eta}\right\}-\frac{\omega}{8k}\frac{\partial^2\varphi}{\partial\eta^2}-\frac{1}{2}\frac{\omega k^2}{k^3}|\varphi|^2\varphi=0.$$

The factor of ω/k drops out, and, writing the partial derivatives as subscripts, we have

$$i\left(\varphi_\tau+\frac{1}{2}\varphi_\eta\right)-\frac{1}{8}\varphi_{\eta\eta}-\frac{1}{2}|\varphi|^2\varphi=0\,. \tag{1.8}$$

This is the **non-dimensional** or **scale-free** form of the equation. Of course we have learned nothing about how to actually solve the equation, but our calculations may be made easier by not having to carry the parameters ω and k through them.

Now suppose we wish to build a $1:100$ scale model of this phenomenon. We want the same equation, in its non-dimensional form (1.8), to govern the model as applies in the ocean. So, with x divided by 100, for η to be unchanged k must be multiplied by 100, and so the wavelengths must be 100 times smaller. Then for φ to be unchanged A must be 100 times smaller — which means that the steepness of the waves is unchanged.

Actually, as we shall see in Q1.9, which deals with the more fundamental 'Navier-Stokes' equation of fluid flow, other things (like pressure) need to be considered too when constructing a scale model. □

1.5 Buckingham's Π theorem

Video: Buckingham Pi theorem

Suppose we don't even have an equation, just a belief that m dimensionful quantities Q_i, $i=1,\ldots,m$ (which may vary or not; for our purposes it doesn't matter) are related by some scientific law, some equation $f(Q_1,Q_2,\ldots,Q_m)=0$. Suppose further that there are n underlying dimensions q_i, $i=1,\ldots,n$. (So far we've had $n\leq 3$, the dimensions being M, L and T, but more are possible.)

The **Buckingham Π (Pi) Theorem** states that the equation

$$f(Q_1,Q_2,\ldots,Q_m)=0$$

may be rewritten in the form

$$g(\pi_1,\pi_2,\ldots,\pi_p)=0$$

where the π_i are $p=m-n$ dimensionless numbers. (It was Buckingham's choice of notation that led to the term 'Π theorem'.) The **proof** follows by thinking about the algebra of solving the simultaneous linear equations for the powers. We will not give it here. Rather we'll see how the theorem has been at work in our examples.

The theorem could be stated in words as:
"A relationship between the quantities describing a phenomenon may be restated as a relationship among dimensionless groups, the number of these being the total number of the original quantities, minus the number of independent dimensions involved."

Example 1 had $m = 4$ quantities r, E, t, ρ in our primitive 3 dimensions of M, L, T. So we should have $p = 1$: and indeed (1.1) may be written as

$$\frac{r^5\rho}{Et^2} = C^5.$$

The left-hand side is a dimensionless number, and a scaling law is just its constancy.

Example 2 expressed the $m = 3$ quantities e, s, λ in terms of just the two dimensions L, T, and again $p = 1$: (1.2) may be written as

$$\frac{e^3\lambda}{s^2} = C'^3.$$

Again a scaling law is really just an equation of the form 'dimensionless number=constant'.

Example 4 had $m = 5$, with F, ρ, μ, l and ν all expressed in terms of M, L, T. So $p = 5 - 3 = 2$: and indeed (1.6) is an equation relating the two dimensionless numbers e and λ.

Example 5 had $m = 5$, with A, x, t, ω, k all written in terms of L, T. So $p = 5 - 2 = 3$, and (1.8) is indeed an equation for three dimensionless numbers, u, η and τ. So the equation doesn't have to be algebraic – it could be a differential or partial differential equation, an iterative (difference) equation, anything – it doesn't matter.

For introductions to dimensional analysis and many more examples see [7, 8]. You might also read the famously entertaining essay [9].

1.6 A note on approximations

Video:
A note on approximations

Applied Mathematics is full of approximations – it has to be, for otherwise we would have to include every phenomenon in every model. Suppose a dimensionless number x is small, very much less than one, written $x \ll 1$, or $1 \gg x$. "Very much less" is not precisely defined, but certainly means "at least an order of magnitude [a factor of 10] less". Then it may be possible to make approximations for $f(x)$ by using the Taylor-MacLaurin series. For example,

$$e^x = 1 + x + \ldots$$

would mean that we are discarding terms "of order x^2". If $x < 10^{-2}$ then what we are discarding will be of order 10^{-4}. A notation which makes this more precise is

$$e^x = 1 + x + \mathcal{O}(x^2).$$

We've replaced the "+..." by a statement of what the dots indicate: terms which, when divided by x^2, do not grow as x becomes small.

Moral of this Section:

Dimensional analysis is the tool which simplifies applied mathematics, cutting through the slew of parameters and variables to get at the heart of the problem. The full, precise statement of the simplification is Buckingham's Pi Theorem. Essentially, all scientific laws are relationships among dimensionless groups, equations of numbers.

The science is the same when the dimensionless numbers are the same. And a fixed dimensionless number – a constant value for a dimensionless group – defines a scaling law, in which the same science happens as the constituents of the group are scaled (up or down, in such a way that the group is unchanged). Any scientific law that has no intrinsic scale or unit (of a given primitive dimension, *e.g.* length) will take the form of a power, 'scaling' law.

The parameters in a dimensional problem correspond to the science involved. For example, motion on the surface of water involves gravitational acceleration g; motion below the surface (*e.g.* of submarines) does not. Or: you may have heard that Einstein's theory of relativity sets a maximum speed in nature, the speed of light c. This doesn't affect most everyday problems, so their dimensional analysis does not include c as a parameter. But if you were investigating 'relativistic' effects, at very high speeds, you would include c. So the key to a dimensional problem is identifying the relevant variables (which are usually clear) and parameters (which are due to the scientific processes at work).

You will probably not, every time you approach an applied-maths problem, conduct a full dimensional analysis, identifying the dimensionless groups and constructing the non-dimensional form of the equation. But two very quick rules (already stated, on p12) will always stand you in good stead:

In any applied-maths equation (and its solution),

1. all of the terms being added together *must* have the same dimensions, and

2. any non-monomial function (any function which is not simply a single power) *must* have for its argument a dimensionless number.

This is equivalent to the requirement that any natural law must be independent of the units in which it is stated. For example, one must not reach different conclusions about the nature of metabolic scaling depending on whether one measures the mass of the animals in grams or kilograms.

OK, now let's start to actually construct some models.

2 Growth and Decay

What you should be able to do by the end of this section:

- Solve simple 1st order **differential equations** (DEs) in applied contexts using separation of variables or an integrating factor;
- Understand the contextual meaning, and check the correctness of the dimensions, of the equation and its solutions;
- For a DE $\dot{x} = f(x)$, sketch $f(x)$ for simple functions and mark the x-axis with arrows corresponding to the sign of $f(x)$ (the **phase line**), and thereby deduce the **fixed points** of the DE and whether each is stable;
- Be familiar with a range of models for growth (and its limits) and decay, including exponential and power-law growth, the logistic equation, the Bertalanffy, Gompertz and similar equations, and the Allee effect;
- Be familiar with a range of contexts for such growth, including ecology and biology, physical and chemical reactions, and finance;
- Solve a 1st order **difference equation** in a financial context and understand its relation to a DE.

Most of the models we shall look at are **autonomous**: the growth rate of the dependent variable (say $N(t)$) depends on N only, and not explicitly on t. That is to say, the equation doesn't care what time it is. Of course explicit time-dependence *can* occur (via, for instance, daily or yearly cycles), and we'll see this in Problem 3.1.

2.1 Unlimited growth and decay

Video: Exponential growth and decay

Example 6: Population growth, radioactive decay

The simplest equation for growth – here of the number N of people in a population – assumes that reproduction occurs in proportion to the number in the population, so that

$$\dot{N} \equiv \frac{dN}{dt} = rN$$

for some constant $r > 0$. Here and always we will use a dot to denote the time-derivative. The solution is easy: we have a separable first-order differential equation

(DE), and, with $N = N_0$ at $t = 0$,[7] we have

$$\int_{N_0}^{N} \frac{dN'}{N'} = \int_0^t r\, dt' \qquad \Rightarrow \qquad \log N - \log N_0 = rt. \tag{2.9}$$

(Note that we put primes on the dummy, integrated variable to avoid confusion with the values at the upper limit of the integration, N and t.) Then

$$N(t) = N_0 e^{rt}.$$

The population grows without limit. It doubles every t_2 where $e^{rt_2} = 2$. This is 'exponential growth'.[8]

Now let $r = -d$ be negative. Now N could describe radioactive atoms, whose decay rate is proportional to their number. The solution is

$$N(t) = N_0 e^{-dt}.$$

The number halves every $t_{1/2}$ (the **half-life**), where $e^{-dt_{1/2}} = 1/2$ or $t_{1/2} = \log 2/d$. Either way, we can see this phenomenon in data by making a log-linear plot: making the vertical axis logarithmic but not the horizontal axis. Then, as in (2.9), we expect to see a straight line, because

$$N = N_0 e^{rt} \qquad \Rightarrow \qquad \log N = \log N_0 + rt.$$

□

Video: Bacterial growth on a plate

Example 7: The growth of bacteria

Figure 9: Bacteria growing on a Petri dish.

Not all growth of N is proportional to N, and not all growth is exponential. Consider instead a number N of bacteria on a plate, forming a circular (disc-shaped) growth. Ignoring for the moment the fact that bacteria also die, suppose that the bacteria can

[7] We will always use a subscript 0 to denote the initial, $t = 0$ value of a dependent variable.

[8] Bearing in mind our results on dimensions, we note that $[\dot{N}] = [N]T^{-1} = [r][N]$ and thus $[r] = T^{-1}$ so that the argument rt of the exponential is dimensionless, $[rt] = 1$, as it should be.

only increase by obtaining nutrients through the perimeter of the disc. If the disc's radius is R and $N \propto R^2$, then the perimeter is proportional to $R \propto \sqrt{N}$. So

$$\dot{N} = rN^{1/2}$$

with $r > 0$. We again separate and solve:

$$\int_{N_0}^{N} \frac{dN'}{N'^{1/2}} = \int_0^t r\,dt' \qquad \Rightarrow \qquad 2N^{1/2} = rt + 2N_0^{1/2},$$

so

$$N(t) = \left(\frac{rt}{2} + N_0^{1/2}\right)^2 = N_0\left(1 + \frac{rt}{2\sqrt{N_0}}\right)^2.$$

This is growth, and without limit, but it goes at a polynomial ($\sim t^2$), not an exponential rate. In 3D we might imagine that the growth rate would be proportional to the surface area, and hence to $N^{2/3}$. The solution is similar, but cubic rather than quadratic. Instead we'll think about how such a scaling applies not to growth but to decay. $\square$.

Note that the dimensions weren't very tidy here: r, in particular, had strange dimensions indeed. If instead we'd used $k := rN_0^{-1/2}$ as the growth rate, we'd have had $[k] = T^{-1}$, and the result would have been

$$N(t) = N_0\left(1 + \frac{kt}{2}\right)^2.$$

Video: Evaporating water droplet

Example 8: The evaporation of a water droplet

Consider a water droplet of mass m which evaporates from its surface area, and thus in proportion to $m^{2/3}$.[9] Then

$$\dot{m} = -rm^{2/3}.$$

with $r > 0$. Again we can solve this:

$$\int_{m_0}^{m} \frac{dm'}{m'^{2/3}} = -\int_0^t r\,dt' \qquad \Rightarrow \qquad 3m^{1/3} = -rt + 3m_0^{1/3}$$

so that

$$m(t) = \left(-\frac{rt}{3} + m_0^{1/3}\right)^3 = m_0\left(1 - \frac{rt}{3m_0^{1/3}}\right)^3.$$

Now, in exponential decay, the quantity never decayed all the way to zero in finite time. Here it does: at time $t = 3m_0^{1/3}/r$ the droplet has *completely* evaporated.

[9] Note that this applies as long as the droplet's shape remains self-similar during evaporation; it doesn't need to be spherical.

Let's just check that our answer has the correct dimensions. From the original DE we have that $[r] = M^{1/3}T^{-1}$. So indeed $\left[rt/m_0^{1/3}\right] = M^{1/3}T^{-1}TM^{-1/3} = 1$. $\square$

Video: Reation kinetics

Example 9: Reaction kinetics

A simple model of a chemical reaction uses the **mass action** principle – essentially, that the chemicals are homogenized (well-mixed). Suppose the reaction is

$$A \,+\, A \longrightarrow B.$$

Then the rate at which it occurs, and thus the rate at which the concentration A decays, is

$$\dot{A} = -aA^2,$$

because the probability of one molecule (at concentration A) encountering another other (also at concentration A) is proportional to A^2.

Solving,

$$\int -\frac{dA}{A^2} = \int a\,dt \qquad \Rightarrow \qquad \frac{1}{A} = \frac{1}{A_0} + at$$

(here and henceforth we leave implicit the limits on the integrals) and so

$$A(t) = \frac{A_0}{1 + A_0 at}.$$

— which decays, but slowly, the concentration falling off as $1/t$. $\square$

2.2 Limits to growth

Video: Logistic growth

Example 10: A limit to growth: the logistic equation

In reality, growth is not unbounded: in the end, somehow, the world imposes some limit. The simplest way to do this is via the Verhulst or 'logistic' equation, which modifies Example 6 to

$$\dot{N} = rN\left(1 - \frac{N}{N_{\text{max}}}\right) \tag{2.10}$$

where N_{max} is a constant parameter.

Now plot the right-hand side (rhs), as in Figure 10: we see that $\dot{N}$ is positive for $N < N_{\text{max}}$ and negative for $N > N_{\text{max}}$. We have added right- and leftwards arrows on the horizontal axis to indicate this; the axis is then called a **phase line**. Further, $N = N_{\text{max}}$ implies $\dot{N} = 0$ and is called a **fixed point** of the DE, a constant solution.

The fact that the arrows point *towards* the fixed point, on both sides, means that if we move a little away from the fixed point we will tend to return towards it: the fixed point is **stable**.

So we see why we called our new constant $N_{\max}$: it is a stable fixed point of the DE and thus the maximum stable value of the population. It is sometimes called the 'carrying capacity' of the system: compared to Example 6, the growth rate has become $k(1 - N/N_{\max})$, and declines to zero as N approaches $N_{\max}$.

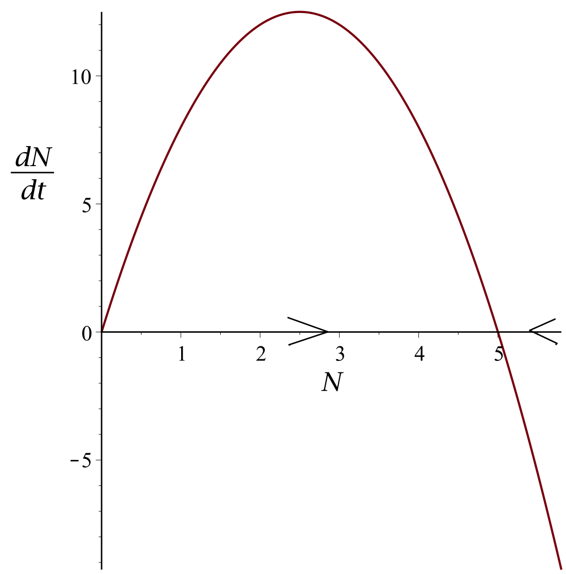


Figure 10: Phase line: dN/dt against N for the logistic equation, with $r = 10$ and $N_{\max} = 5$.

Now let's calculate the solution properly. The DE is again first-order separable: so

$$\int \frac{dN}{N(1 - N/N_{\max})} = \int \left(\frac{1}{N - N_{\max}} - \frac{1}{N} \right) dN = \int r\, dt.$$

Integrating and fixing the constant of integration in terms of the initial value N_0 we have

$$\log \left(\frac{N - N_{\max}}{N} \right) = kt + \log \left(\frac{N_0 - N_{\max}}{N_0} \right).$$

Exponentiating and inverting gives

$$1 - \frac{N_{\max}}{N} = \left(1 - \frac{N_{\max}}{N_0} \right) e^{-rt},$$

and then re-arranging gives

$$N(t) = \frac{N_{\max}}{1 - \left(1 - \frac{N_{\max}}{N_0} \right) e^{-rt}}.$$

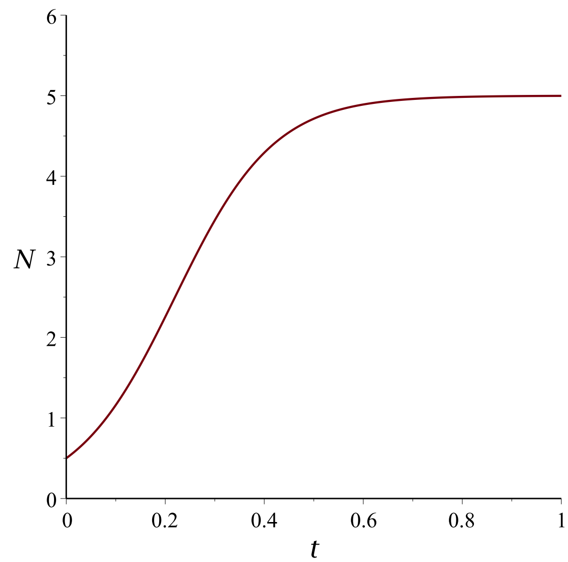


Figure 11: N against t for the logistic equation, with $r = 10$, $N_{\max} = 5$ and $N_0 = 0.5$.

At $t = 0$ the rhs is N_0 (as it must be, if we've done our calculation correctly) while as $t \to \infty$ we have $e^{-rt} \to 0$ and $N \to N_{\max}$, in accordance with our deductions above, and as seen in Figure 11. □

Video: Fixed points and stability

Example 10 demonstrated the power of an essential technique for analysing a first-order DE, that of drawing the **flow diagram** or **phase line**. If

$$\frac{dx}{dt} = f(x)$$

(that is, it does not explicitly depend on t), then first plot $f(x)$ against x, marking a rightward arrow on the x-axis where $f(x)$ is positive and a leftward arrow where $f(x)$ is negative. These arrows tell you the direction of flow of the solution (although not how fast the solution moves, the *rate* of flow).

The zeros of f are the **fixed points**, the constant solutions, of the DE; and a FP is **stable** to small perturbations if the arrows point towards it, and **unstable** if the arrows point away.

The logistic equation of Example 8 gave us a nice way to impose a limit on growth, but it was rather *ad hoc*: we had no good reason for the form of the DE. Here's a better-reasoned alternative.

Video: Bertalanffy's equation

Example 11: Bertalanffy's growth equation, applied to a tumour

Consider the growth of a tumour of mass m [10]. Assume that the tumour has not (yet) created its own blood supply. Cells in the tumour die in proportion to their number. New cells appear only to the extent to which the tumour is able to secure nutrients across its surface. Together, the two effects give us

$$\dot{m} = am^{2/3} - bm,$$

where a, b are positive (dimensionful) constants. Now, the $m^{2/3}$ term is greater than the m term when m is small, and vice versa when m is large. So if we were to draw the phase line, it would look like that of the logistic equation above: we would find a stable fixed point at $m = (a/b)^3$.

Now, we could continue to deal with the fractional powers, but it's easier to change variables to $l = m^{1/3}$ or $m = l^3$.[10] Then $dm = 3l^2\, dl$ and so the DE becomes

$$3l^2 \dot{l} = al^2 - bl^3 \qquad \text{or} \qquad 3\dot{l} = a - bl.$$

It's easy now to see that the DE has a fixed point at $l = a/b$, and that this is stable. (In the original equation we could have seen this by sketching the rhs: as m increases from zero, initially the $am^{2/3}$ term dominates, but later the $-bm$ term does so.)

Let's see the actual solution:

$$\int \frac{dl}{a - bl} = \int \frac{1}{3} dt \qquad \Rightarrow \qquad -\frac{1}{b}\log(a - bl) = \frac{t}{3} - \frac{1}{b}\log(a - bl_0)$$

and so

$$a - bl = (a - bl_0)e^{-bt/3}$$

(in which form it's easy to see that $a - bl$ approaches zero exponentially) and then

$$l = \frac{a}{b} - \left(\frac{a}{b} - l_0\right) e^{-bt/3} \qquad \text{or} \qquad m(t) = m_0 \left(\frac{a}{bm_0^{1/3}}(1 - e^{-bt/3}) + e^{-bt/3} \right)^3.$$

So m begins at m_0 and approaches $(a/b)^3$ at large times.

This is indeed the way tumours grow: and typically they end up with an actively-dividing surface and a necrotic (dead) centre. To grow larger, they have to form their own network of blood vessels (they have to 'vascularize'). □

In fact this also explains why any animal which obeys Kleiber's law (3/4-power scaling of metabolism with mass), or any process which causes sub-linear growth (*i.e.* growth proportional to a power less than 1), has an upper limit. If dm/dt has a growth term proportional to m^a, with $a < 1$, but expends energy in direct proportion to mass m, then

[10] We use the letter l for length because, although l does not have the dimensions of length, it is proportional to the length scale of the tumour.

we can see from the phase line that growth approaches a stable fixed point. Sublinear scaling of available resources imposes a limit on growth [11].

Video: The Allee effect

Example 12: The extinction of the passenger pigeon and the Allee effect

The passenger pigeon used to live in North America in billions, forming enormous flocks. Yet it was hunted to extinction in just a few decades.

The **Allee effect** is that the fitness (to survive) of individual animals may be enhanced when in larger groups. It may have been at work in the extinction of the passenger pigeon, as in this example.

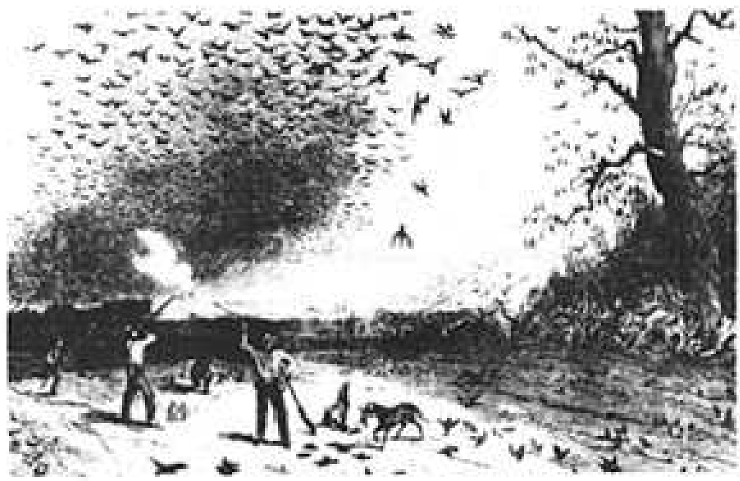


Figure 12: The hunting of passenger pigeons.

Consider a flock of N passenger pigeons. Their number increases exponentially in the absence of hunting and predation, with some logistic limit imposed by their food sources. Now add in some predation, but suppose that this occurs only in proportion to the surface area of the flock – that is, to $N^{2/3}$. (The same effect occurs with large shoals of fish.) Then we have

$$\dot{N} = aN - bN^2 - cN^{2/3}, \tag{2.11}$$

where a, b, c are positive constants.

In the absence of predation (that is, when $c = 0$) there is a stable fixed point at the logistic limit $N = a/b$. This stable fixed point always remains and is positive, even though it may be made smaller by reducing b. The birds cannot simply become extinct.

With $c \neq 0$ there are three fixed points: a stable one at $N = 0$, a stable fixed point at the highest root, and an intermediate unstable fixed point. One way to think about sketching the rhs of (2.11) is to note first the smallest power, $N^{2/3}$, with its negative coefficient: so the curve begins at zero with a negative gradient. Next the positive aN term becomes more important, so the curve moves upwards. It may or may not become positive, depending on a, b and c. Then finally the $-bN^2$ wins and at large N the curve becomes large and negative. Figure 13(a) gives some impression of this, although the behaviour near zero is compressed.

To solve the equation we again set $N = l^3$ so that $dN = 3l^2\, dl$ and

$$3\dot{l} = al - bl^4 - c \equiv f(l),$$

plotted in Figure 13(b).

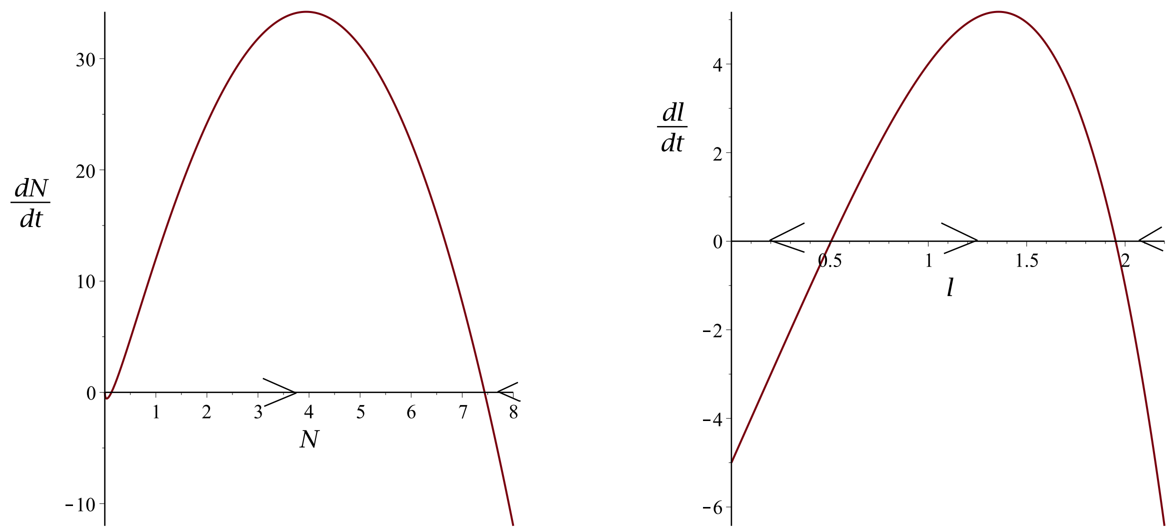


Figure 13: Phase lines for (a) dN/dt against N, with $a = 30, b = 3, c = 15$. It is difficult to see the small range close to the origin for which $dN/dt < 0$, but it is there, and is much more evident in (b) dl/dt against l.

To find the fixed points requires solving $f(l) = 0$, a quartic equation, but we don't need to know their exact values. We can see that $f(l)$, plotted in Figure 12(b), has a single turning point, at $a - 4bl^3 = 0$ and thus $l_{\text{max}} = (a/4b)^{1/3}$, that

$$f(l_{\text{max}}) = \frac{3}{4}a\left(\frac{a}{4b}\right)^{1/3} - c,$$

and that this is a maximum (since $f(0) = -c$ is negative, and the $-bl^4$ term makes $f(l)$ negative for large l). So, provided $f(l_{\text{max}})$ is positive, the graph of $f(l)$ looks like Figure 12(b).

The largest fixed point is stable, but (in contrast to the logistic equation) $N = 0$ is also now stable, with an unstable fixed point in the middle. The crucial point is that, if predation c increases and $f(l_{\text{max}})$ becomes negative, then the unstable fixed point and the positive stable fixed point coalesce and disappear. Then $f(I)$ is everywhere negative, and the population suddenly collapses to zero – and stays there. $\square$

2.3 Finance

Video: Compound interest

Example 13: Compound interest

Suppose an initial sum V_0 of money earns interest at a rate of $100r\%$ annually, and that this is credited once per year. Denote the sum in the account after T years by V_T. Then

$$V_{T+1} = (1+r)V_T \qquad \text{and} \qquad V_T = V_0(1+r)^T.$$

This looks similar to Example 6, continuous exponential growth. What is the relation?

Suppose instead the interest were compounded n times per year: for example, if $n = 2$ then you receive half the interest after six months, and the other half at the end of the year – which then includes interest on the first half-year's interest. The value after T years is then

$$V_T = \left(1 + \frac{r}{n}\right)^{nT} V_0.$$

What happens when n becomes large? – that is, when the interest is continuously-compounded? Well,

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x,$$

so

$$V_T = \lim_{n\to\infty} \left(1 + \frac{r}{n}\right)^{nT} V_0 = V_0 e^{rT}.$$

This, of course, is the solution of the DE for exponential growth,

$$\frac{dV}{dt} = rV. \tag{2.12}$$

The reason is that we can write $V_{T+1} = (1+r)V_T$ as $V_{T+1} - V_T = rV$, and $V_{T+1} = V_T$ is a very simple discrete analogue (the 'forward difference') of the derivative. When one takes the forward interval to be not one year but $1/n$ years, the limit as $n \to \infty$ is (2.12), and the solutions match. $\square$

Video: Bonds, loans and mortgages

Example 14: Bonds, loans and mortgages

– which all share the same underlying equation, namely

$$V_{T+1} = (1+y)V_T - c,$$

where for a
bond: V_T is its value after T days, y is the **yield** and c the daily **coupon**; while for a **mortgage**: V_T is the outstanding amount after T years, y is the interest rate (which is $100y\%$), and c is the annual payment.

This is not a DE but rather a finite **difference equation**. These are used in many areas of applied maths, but in this module (and its assessment) they will only appear in financial problems. A difference equation can be solved in much the same way as we solve the corresponding DE: we make an ansatz (a trial solution) of the form

$$V_T = ak^T + b$$

and then substitute-in to determine a and b. Doing so, we find that

$$ak^{T+1} + b = (1+y)(ak^T + b) - c \qquad \Rightarrow \qquad k = 1+y, \ b = c/y.$$

The initial condition is then $a + b = V_0$ so that $a = V_0 - c/y$ and then, finally,

$$V_T = \left(V_0 - \frac{c}{y}\right)(1+y)^T + \frac{c}{y}.$$

Notice that, if we assign a new dimension V to monetary value, then $[V_T] = V$, $[y] = T^{-1}$, $[c/y] = VT^{-1}T = V$. The only difficulty is that $1+y$, which is an *annual* multiplier, has to share the same dimensions as y.

Suppose this is a mortgage paid off over 25 years, so that $V_{25} = 0$. Then we can solve this for c:

$$0 = \left(V_0 - \frac{c}{y}\right)(1+y)^{25} + \frac{c}{y}$$

becomes

$$c = \frac{yV_0}{1-(1+y)^{-25}}.$$

So the annual payment on a $V_0 =$£200,000 mortgage at a fixed rate of 3% (so $y = 0.03$) is about £11,500 or about £960 per month. $\square$

Moral of this Section:

First-order ordinary differential equations are good for modelling growth and decay (but not oscillations, which require second order, as in simple harmonic motion).

Such growth may be exponential – doubling every fixed period of time, as described by the exponential function – or polynomial (as a power of time).

Corresponding decay processes may cause a quantity to decay exponentially or, if polynomial, to vanish completely.

Limits to growth are imposed by a non-monomial growth rate, in which two different powers compete. As the parameters change, the stable population may be reduced. When there is a third term in the growth rate, the stable population can be very quickly annihilated by a change of parameter.

3 Coupled Systems

What you should be able to do by the end of this section:

- Solve a pair of coupled linear 1st order DEs in applied contexts;
- For a pair of simple coupled nonlinear 1st order DEs, be able to write down the **null clines** and **fixed points**, and sketch the **phase plane**;
- In simple cases, be able to linearize about the fixed point;
- In simple cases, be able to divide one equation by the other, separate, integrate, and thereby obtain a conserved quantity;
- Understand the contextual meaning of both the system and its solutions;
- Be familiar with a range of models and contexts, including predator-prey systems (including Lotka-Volterra), epidemiology (including Kermack-McKendrick), warfare (including Richardson and Lanchester), and the Van der Pol oscillator.

3.1 Introduction: the Simple Harmonic Oscillator

Video: Coupled systems

In the last section we looked at models of the growth and decay of a single dependent dynamical variable. We did not observe any periodic or cyclic behaviour. Now, the simplest system you know which has oscillations is the **simple harmonic oscillator** (SHO)

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

which has oscillatory solutions $x = A \sin \omega(t - t_0)$. This is a *second*-order equation, and for autonomous systems it is only with such **second-order dynamics** that oscillations spontaneously appear. One way to get second-order dynamics is to couple two first-order DEs (as you saw in the Calculus module). For example, the SHO is identical to the coupled system

$$\dot{x} = \omega y, \qquad \dot{y} = -\omega x.$$

One way to describe the SHO is through movements in the xy-plane, where trajectories are circles — because

$$\frac{d}{dt}\left(x^2 + y^2\right) = 2x\dot{x} + 2y\dot{y} = 2\omega xy - 2\omega yx = 0,$$

and the curves of constant $x^2 + y^2$ are circles.

This section is devoted to such coupled **dynamical systems**. A more thorough treatment, which also introduces the strange mathematics of chaos that appears when there are three or more variables, is given in the module 'Dynamical Systems'. A general such system is a pair of equations for $x(t), y(t)$ of the form

$$\dot{x} = f(x, y), \qquad \dot{y} = g(x, y) \tag{3.13}$$

for two smooth functions f and g. What techniques can we apply to understand such systems? Of course we could simulate the solutions on a computer, given some initial values $x(0), y(0)$, but then how should we display the solutions? Taking the SHO above with $x(0) = 0, y(0) = 1$, we can plot both x and y as functions of time t on the same pair of axes, Figure 14(a), here for $0 \le t \le 5$ (which is not quite a full period 2π).

An alternative approach would be to plot y against x: then the equations (3.13) give us a vector $(\dot{x}, \dot{y})$ which indicates the flow at any point (x, y), and the software can include in the plot the directions of this flow at a grid of points. We could also put a sample trajectory into the plot. Such a plot is called a **phase plane** (sometimes a **phase portrait**, Figure 14(b). Later we will see how to sketch a phase plane, capturing the essential insights, without the aid of a computer.

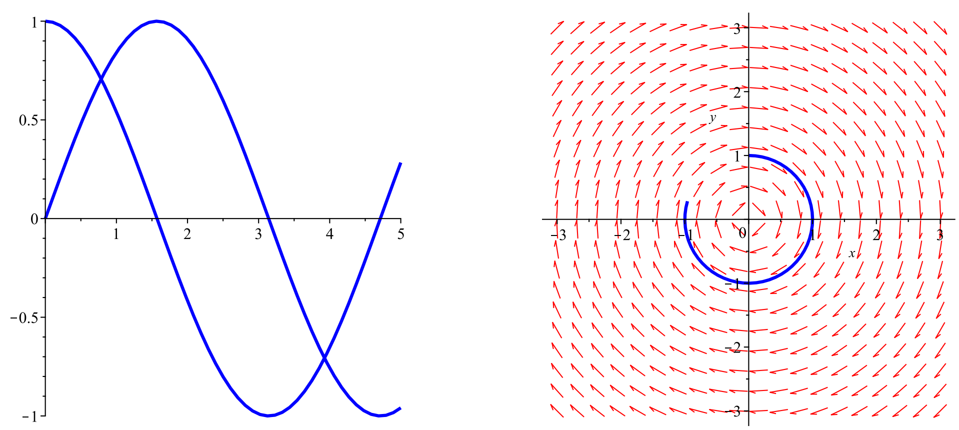


Figure 14: Solutions of the Simple Harmonic Oscillator (3.13) for $0 \le t \le 5$: (a) $x(t)$ and $y(t)$ as functions of t; (b) y plotted against x, with arrows indicating directions of flow.

There are two other things we can do, both exemplified by the SHO. Sometimes there is a **conserved quantity**, a quantity whose time-derivative is zero and which is therefore constant or 'conserved', such as $x^2 + y^2$ for the SHO, and trajectories are then curves on which this quantity is constant, rather like contours (of constant height) on a map. For the SHO, the trajectories are circles. We may be able to find such a quantity by dividing the two equations of (3.13):

$$\frac{\dot{y}}{\dot{x}} = \frac{dy}{dx} = \frac{g(x, y)}{f(x, y)}\,.$$

This is an ODE relating y to x (and note that time no longer explicitly appears). If it is separable, we can integrate it to obtain a function of x and y which is constant (which you may have called the 'constant of integration').

Finally, in some cases we may be able to actually solve the coupled differential equations, most obviously (and as seen in the Calculus module) in cases where f and g are linear functions. We begin with such an example.

3.2 Predator-prey systems

Video: A simplified predator-prey system

Example 15: A simplified predator-prey system

On a tropical island, there are only two species of animal. Both species feed on the abundant supplies of vegetation and species A also feeds on species B, which reproduces at a higher rate than species A. The numbers of individuals $A(t)$ and $B(t)$ can be regarded as continuous functions of time, satisfying, in their non-dimensionalized form,

$$\begin{aligned} \frac{dA}{dt} &= 2A + B \\ \frac{dB}{dt} &= 4B - 2A \,. \end{aligned}$$

At $t = 0$ there are N_0 animals in species A and $4N_0$ in species B.

What happens thereafter? Well, this is exactly the kind of system of two *linear* coupled equations which the Calculus module has taught you to solve. First, let's use the shorthand $d/dt \equiv D$. Then

$$\begin{aligned} (D-2)A &= B \\ (D-4)B &= -2A \end{aligned}$$

$$\Rightarrow \quad (D-2)(D-4)A = (D-4)B = -2A$$

which implies

$$(D^2 - 6D + 10)A = 0 \,.$$

The auxiliary equation is $k^2 - 6k + 10 = 0$, which gives $k = 3 \pm i$, and so the general solution is

$$A = e^{3t}(a \sin t + b \cos t) \,,$$

which yields

$$B = (D-2)A = e^{3t}(a \sin t + b \cos t + a \cos t - b \sin t) \,.$$

Then at $t = 0$, $b = N_0$ and $a + b = 4N_0 \quad \Rightarrow \quad a = 3N_0$.

So

$$\begin{aligned} A &= N_0 e^{3t}(3\sin t + \cos t) \\ B &= N_0 e^{3t}(2\sin t + 4\cos t). \end{aligned}$$

But $B = 2N_0e^{3t}\cos t(\tan t + 2)$ and so at $t = \alpha$, where $\tan\alpha = -2$ and $\frac{\pi}{2} < \alpha < \pi$, species B becomes extinct! Then, for $t > \alpha$, we have $B = 0$, and A will increase exponentially, unchecked, according to

$$\frac{dA}{dt} = 2A\,.$$

□

Video: The Lotka-Volterra island

Example 16: The Lotka-Volterra island

In the previous example, species A fed on species B in proportion to B's numbers. More realistically, the rate at which predation occurs will be proportional to both species' numbers. The classic example of two species on an island – let's make them foxes F and rabbits R – is due to Lotka and Volterra [12], and an example would be

$$\begin{aligned} \frac{dF}{dt} &= -F + FR \\ \frac{dR}{dt} &= R - FR\,. \end{aligned}$$

Thus foxes, in the absence of rabbits (that is, when $R = 0$), die out exponentially, $\dot{F} = -F$; and rabbits, in the absence of foxes, increase exponentially, via $\dot{R} = R$.

(Notice the absence of parameters. In principle all four terms should have independent parameters. Rescaling F, R and t can set three of these to unity: but we have specialized the fourth parameter, too, to one. This does not alter the essential behaviour, but it keeps our equations clean and simple.)

Now, you do not know techniques for solving this equation exactly, but there are some techniques we can use to understand the solution's behaviour. First, we can find the fixed points. They are the solutions of $F(-1 + R) = 0$ and $R(1 - F) = 0$, which are FP1: $R = F = 0$ and FP2: $R = F = 1$. These lie at the intersections of the **null clines**, the curves (in this case, black lines) on which *one* of the two dependent variables' rates of change vanishes: here, either $\dot{F} = 0$ (on the line $R = 1$) or $\dot{R} = 0$ (on $F = 1$).

Next, we can divide the first equation by the second to give

$$\frac{dF}{dR} = \frac{F(R-1)}{R(1-F)}.$$

The time variable t has disappeared, so this won't tell us where the system is at a given t, but it will tell us about a quantity that remains constant as the system evolves. Separate variables:

$$\int \frac{1-F}{F} dF = \int \frac{R-1}{R} dR,$$

so that $\log F - F = R - \log R + \text{constant}$, and so

$$\log F + \log R - F - R \tag{3.14}$$

remains constant. That's just like remaining at a constant height and thereby knowing that you're walking along a contour: the motion will be along curves for which this quantity is constant.

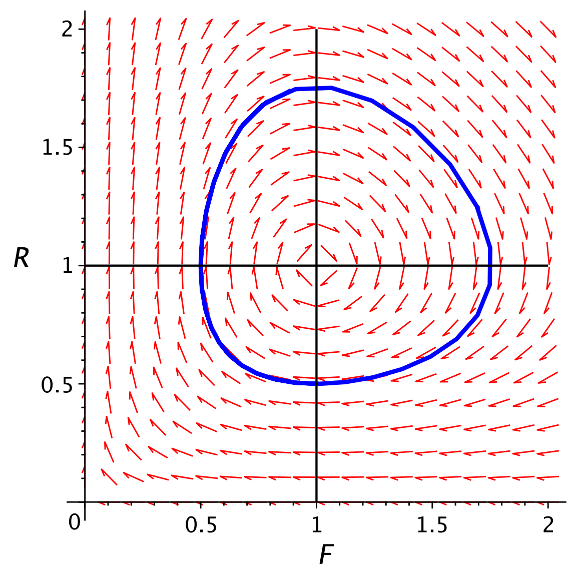


Figure 15: The Lotka-Volterra island.

Let's now have a look at a numerical solution. Figure 15 is the phase plane. It plots F and R on the two axes, so that at a given time the system is at a point in the plane – actually, for our system, it remains in the positive quadrant, $F \geq 0$ and $R \geq 0$. The little red arrows tell you what direction (F, R) is moving in (although not how fast: the arrows are all of the same length). Notice that the arrows are horizontal on $F = 1$ and vertical on $R = 1$: these are the null clines, in black. They help us to get a general sense of the (clockwise, rotational) direction of motion. The thick blue curve is a **trajectory**, the path taken by (F, R), here starting from $(1, 0.5)$. It's a closed curve, because of the constancy of (3.14), so the populations of foxes and rabbits are periodic — they fluctuate in cycles.

Linearization about the fixed point: We can't find these cycles exactly, but notice that around FP2 they're almost exactly circles. We can see this by considering only small departures from FP2: we set

$$F = 1 + \delta, \qquad R = 1 + \epsilon,$$

where δ, ϵ are small. Substituting into the DEs we have

$$\dot{\delta} = F(R-1) = \epsilon(1+\delta), \qquad \dot{\epsilon} = R(1-F) = -\delta(1+\epsilon).$$

Now we ignote the quadratic terms: if δ and ϵ are small, then $\epsilon\delta$ is smaller still. So

$$\dot{\delta} = \epsilon, \qquad \dot{\epsilon} = -\delta,$$

which is linear (so that this process is known as 'linearization'). We differentiate one equation, substitute in the other, and find $\ddot{\delta} = -\delta$, of which the solution is a sine or cosine: we just have simple harmonic motion, in a circle in the (F, R) plane, for

$$\frac{d}{dt}\left(\delta^2 + \epsilon^2\right) = 2(\delta\dot{\delta} + \epsilon\dot{\epsilon}) = 2(\delta\epsilon - \epsilon\delta) = 0,$$

and $\delta^2 + \epsilon^2 =$constant is just a circle. □

What is the correct way to describe the interaction between predator and prey? Example 15 certainly doesn't seem right: A gains in proportion to B by eating them, yet B thereby loses out in proportion to A. By contrast, in Example 16, the Lotka-Volterra island, the interaction is at least symmetric: F gains, and R loses, in proportion to FR.

But there's a problem with this, too. This interaction is *density-dependent*: if the numbers of F and R both double, then the rate of predation increases four-fold. This does not really seem reasonable. A more reasonable interaction term might be

$$\frac{FR}{fF + rR},$$

where f, r are constants. Not only is this density-independent (if F, R both double then so does the interaction) but it also has some other nice properties. If R is much larger than F then the predation is proportional to F: each fox takes as many rabbits as it needs, or can kill. If F is much larger than R then the predation is proportional to R: the foxes have to share the available rabbits between them. Such terms were originally proposed in [13]. But then the equations become harder to descale and harder to solve by any but numerical means. We shall not consider them further.

Video: The Van der Pol oscillator

Example 17: The van der Pol oscillator

Not all oscillations in nature are like simple harmonic motion, or even like the cycles on the Lotka-Volterra island. Some periodic behaviours look more like flips between two nearly-stable states, and it is easy to create electrical circuits with this property. Most importantly, perhaps, the human heartbeat is of similar form.[19] We will study just one case, the prototype for many such oscillators.[20]

We study the following system of equations:

$$\begin{aligned}\dot{U} &= \mu\left(U - \frac{U^3}{3} - W\right)\\ \dot{W} &= \frac{1}{\mu}U,\end{aligned}$$

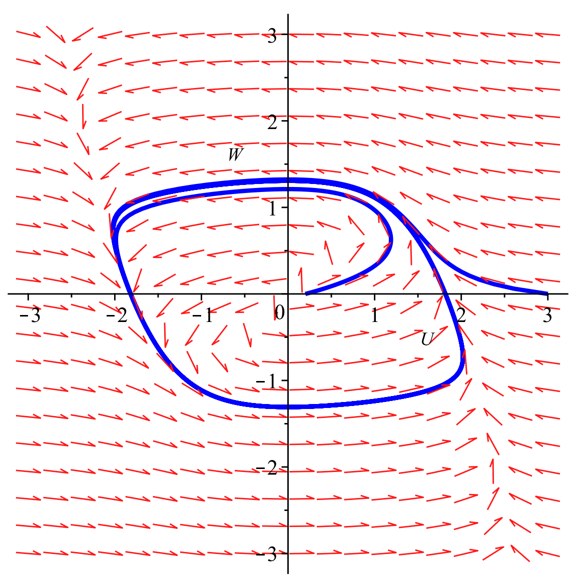
where μ is a positive real parameter. We will not be precise about the meaning of U and W in natural systems, but U may be thought of as an electrical potential.

First, differentiate the first equation and substitute-in the second: then

$$\ddot{U} - \mu(1 - U^2)\dot{U} + U = 0.$$

We see immediately that when $\mu = 0$ the system reduces to the simple harmonic oscillator.[11]

But for $\mu \neq 0$ the system has some very surprising behaviour. To begin with, it has a **limit cycle**: there is a periodic trajectory which other trajectories approach. Trajectories which start inside the cycle spiral out towards it; trajectories which start outside spiral in. There is only one such cycle (see appendix): this is not like the Lotka-Volterra system (Example 16), in which a conserved quantity ensured that there were many closed trajectories, all 'contours' of the constant quantity.



[11]By moving to this second-order equation we have removed the problem of the division by $\mu = 0$ in $\dot{W}$.

Figure 20: The Van der Pol oscillator for $\mu = 2$. Two sample trajectories are shown, beginning at $(0.2, 0)$ and $(3, 0)$.

Further, there is one really surprising feature: when μ is large, there are two very different speeds in the system. We can begin to understand this by drawing the curve $W = U - U^3/3$, the null cline on which $\dot{U} = 0$ and the directional arrows are vertical. As long as a trajectory stays on or very close to the null cline, $\dot{U}$ stays very small. You can see in Figure 20 how the flow pushes the trajectory towards the null cline where it has negative gradient. But then, where the trajectory approaches an extremum of the null cline, it suddenly departs; $\dot{U}$ becomes much larger (by a factor of order μ^2) than $\dot{W}$; and the trajectory 'zaps' across horizontally to the other branch of the null cline, where the 'slow' behaviour then resumes. This repeating 'fast-slow' behaviour occurs in many natural and artificial systems. □

3.3 Epidemiology

We consider here two models of epidemics, one of quick-acting (compared to human lifespan), often non-deadly diseases, and the other of slow-acting, deadly diseases interacting with wider population growth and decay.

The first case has obvious topicality, and I have extended it for 2021 so that you can learn something of how the covid-19 pandemic can be modelled.

Example 18: The Kermack-McKendrick 'SIR' model

This is typically a model [14] of a rapidly-developing epidemic — measles, for example, or influenza (hereafter 'flu'), or covid-19. These diseases *can* kill, of course, but mostly they kill thankfully few. The point of 'rapidly-developing' is that the total population (including deaths) remains approximately constant; we need not consider births and more general deaths.

Video: SIR model, pt 1

The equations of the SIR model split the population into three groups: susceptible S, infected I and removed R (hence the name 'SIR'). The removed group R might be recovered or dead; the point is that these people can no longer infect or be infected. Then

$$\begin{aligned} \dot{S} &= -\beta S I \\ \dot{I} &= \beta S I - \gamma I \\ \dot{R} &= \gamma I. \end{aligned}$$

The γI term is the natural recovery of infecteds, in proportion to their number, while

βSI is the rate of infection, proportional to the number of random meetings between susceptibles and infecteds — just as in the Lotka-Volterra model of meetings between predators and prey. We don't need to worry too much about density-dependence, because the total population is fixed.

Now, there are various assumptions here. First and foremost is that the population is homogeneous – there is no separation of the population by geographical location, or age, or behaviour. The model can easily be extended in each of these directions. Another which you might notice is the exponential decline of I, in the absence of new infections. This implies that some people remain infected for arbitrarily long times: indeed I then follows an exponential distribution. Again, this is easily altered – one could keep the S, I and R bins, but alter the dynamics of the transfer between them. Another is that people who have the disease acquire immunity: they cannot catch it again. Once more, one could easily add a suitable term to describe reinfection.

But this simple model is our starting point, and even such simple a model appears surprisingly often in research papers about the covid-19 pandemic. We should see what we can learn from it.

Now, the model initially appears to be *three*-dimensional, but notice that each term on the right appears once negatively and once positively: if we set $N = S + I + R$ to be the total number the population, then $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$, so in fact one of the variables is redundant. We'll work mostly with S and I; then R is simply calculated as $R = N - S - I$. I won't mention this R again.

Let's ask some questions.

Will there be an epidemic?
Write

$$\dot{I} = (\beta S - \gamma)I = gI,$$

where $g(t) = \beta S - \gamma$. We recognize in this differential equation for I the very first of our models for growth, exponential growth, albeit with a rate g which (via S) will vary with time.

Suppose a disease begins to spread among a susceptible population: that is, $S_0 \simeq N$, $R_0 = 0$, with a small number of initial infecteds $I_0 = \epsilon N$, with $\epsilon \ll 1$ (that is, I_0 is much less than N). Then there will be what we will call an **epidemic**, an outbreak with exponential initial growth, if the initial value $g_0 = \beta N - \gamma > 0$, or

$$\frac{\beta N}{\gamma} > 1. \tag{3.15}$$

This is actually surprisingly simple to interpret: the left-hand side is the initial (sometimes called the 'basic') **reproduction number**, R_0. First, βN is the initial number of new infections per infected person per day. Next, recall that in this SIR model γ was the exponential decline rate (per day) of infected people in the absence of new infections; its inverse $T = \gamma^{-1}$ is the (average) **serial interval time** or **generation time** between successive infections. Thus $\beta N/\gamma = \beta NT$ is the typical number of *new* infections caused by one person with the disease. It is the one dimensionless number (not a 'rate') which characterizes the epidemic, and it's always called R_0 (which is highly unfortunate, because of the obvious potential for confusion with the number of removed individuals). There will be an epidemic if $\mathrm{R}_0 > 1$.

R_0, the fundamental dimensionless number which describes the disease, is a feature of the epidemic itself – of the disease and the social behaviour which spreads it – rather than of the particular model we use to describe it, here the SIR model. That point about social behaviour is crucial: it's the behaviour of the disease in its host society which determines R_0. There are many diseases which can be eradicated by reducing R_0 through public health measures – (oversimplifying somewhat,) plague is a disease of poor hygiene, cholera of poor plumbing, tuberculosis of poor housing. Covid-19 appears to be a disease which spreads principally through people meeting indoors.

Let's explore the relationship between R_0 and the other parameters a little further. Now, β and γ are not very intuitive parameters, but we can rewrite the growth rate in terms of $T = \gamma^{-1}$ and $\mathrm{R} = \beta S/\gamma$:

$$g = \beta N - \gamma = \frac{\mathrm{R}-1}{T}, \quad \text{or} \quad \mathrm{R} = 1 + gT.$$

The growth rate g is easily observed as the gradient of a logarithmic plot of case numbers in the early stages of the epidemic; for covid-19 the cases doubled roughly every 3.5 days, so that $g = \log 2/3.5 \simeq 0.2$ per day. The serial interval time can usually be estimated; let's suppose it's about 7.5 days. Then $\mathrm{R}_0 \simeq 1 + 0.2 \times 7.5 = 2.5$. Note that g is directly observed, T has to be estimated, and R_0 further required a model, here the SIR model. The precise relationship between R, g and T depends on the details of the model.

How large will the epidemic be?
As the number of susceptibles declines, the growth rate $g(t) = \beta S(t) - \gamma$ declines, until when the **effective reproduction number** $\mathrm{R} = \beta S/\gamma$ becomes less than one and the growth rate turns negative. This happens when $\dot{I} = 0$ and thus $S = N/\mathrm{R}_0$, the **herd immunity threshold**. At this time, if there were only a tiny number of infected people, there would then be no epidemic. Unfortunately the opposite holds; this is the time of *peak* I, which then takes a long time to decline, during which many more people are infected. We'll explore this phenomenon a little later.

Video: SIR model, pt 2

You can see how the SIR model behaves in Figure 16. The epidemic is seen in the 'lump' in the curve of infected people. Its final outcome, once the lump in I has declined again, is seen (in the parameters used for this plot) in the removed number having increased from 0 up to about 80% of the population.

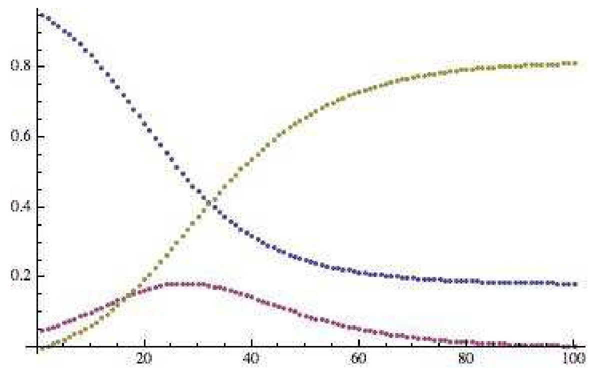


Figure 16: The SIR model. The declining curve is S, the increasing one is R, and the one with the lump is I. C. French, *Simulating a global Ebola outbreak*, Wolfram.

But we can do better than a numerical simulation to find the size of the epidemic; we can find a conserved quantity. Divide the second equation by the first,

$$\frac{dI}{dS} = \frac{\gamma}{\beta S} - 1,$$

separate and integrate:

$$\int dI = \int \left(\frac{\gamma}{\beta S} - 1 \right) dS \qquad \Rightarrow \qquad \frac{\gamma}{\beta} \log S - S - I = \text{constant}.$$

Just as with our other models, knowing that a certain function of the variables remains constant while the variables individually are changing is enormously useful. Here, the constant will tell us the size of the epidemic, which we'll write as a fraction x of the population. Initially, $S \simeq N$ and $I \simeq 0$. After the epidemic (defined by $I \to 0$ as $t \to \infty$), $S = (1 - x)N$ and again $I = 0$. So

$$\frac{\gamma}{\beta} \log S - S - I = \frac{\gamma}{\beta} \log N - N \ \ \text{(initially)} = \frac{\gamma}{\beta} \log(1 - x)N - (1 - x)N \ \ \text{(finally)},$$

which, adding $(1 - x)N - \frac{\gamma}{\beta} \log N$ to both sides and dividing by N, gives

$$\frac{\gamma}{\beta N} \log(1 - x) = -x \quad \text{or} \quad 1 - x = e^{-\mathrm{R}_0 x}. \tag{3.16}$$

The size of the epidemic depends only on (the single dimensionless number) R_0!

We want to solve this for x. A non-zero solution $x(\mathrm{R}_0)$ will exist precisely if the initial downward slope R_0 of the exponential is greater than that of the lhs — that is, precisely if $\mathrm{R}_0 > 1$. However, we cannot find a solution in closed form using functions you know.

This is the sort of situation where numerical help is needed, and Maple plots implicit functions with ease. Figure 17 shows us what happens: and the alarming thing is how rapidly x rises with R_0.

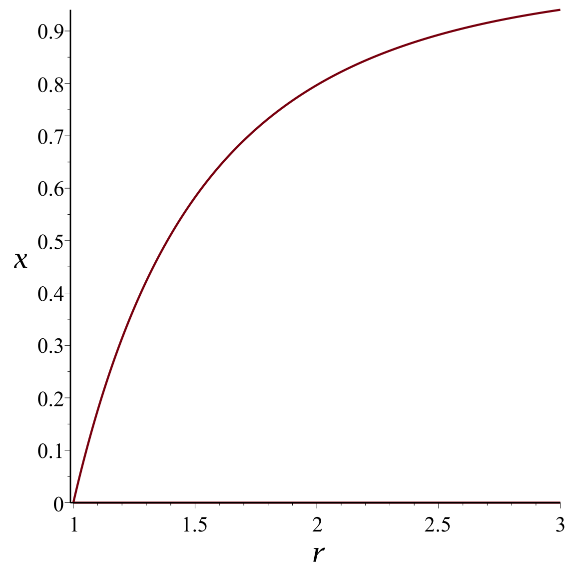


Figure 17: Epidemics in the SIR model: Size of epidemic x as proportion of unvaccinated population versus basic reproduction ratio R_0.

Video: SIR model, pt 3

How can we minimize the epidemic?

As noted above, we reach **herd immunity** or **population immunity**, the point at which an epidemic is no longer inevitable, when S is sufficiently reduced that $R = 1$, which happens when $S = N/R_0$; after this, the number of infections will decline rather than grow. For $R_0 = 1.5$, for example, it happens when $S = N/1.5 = \frac{2}{3}N$, or $1/3$ of people have had the disease. Yet in Fig.17, for this value of R_0 we have $x \simeq 0.6$. This is so large because, at the time at which $1/3$ of the population have had the disease, many of them still have it and are infecting more people, even if this is on average less than one more each. The epidemic still has a long way to run even as it declines.

What can we do to improve things? If we could, at roughly the peak of infections, prevent all infectious interactions, we might (after a few T) reduce I close to zero and have no epidemic thereafter (conceptualized in the covid-19 pandemic as a 'circuit-breaker lockdown'). But a weaker action is to note that this epidemic happens *fast*, and to slow it down, so that the herd immunity threshold is not overshot. We can do this by altering R_0, which you recall is a function of people's behaviour. We temporarily alter social behaviour to effect a new, lower reproduction number R_0', such that $x(R_0')$ is the threshold for herd immunity under the original R_0. We therefore set $x = 1 - 1/R_0$, the desired lower size of the epidemic, lower than $x(\mathrm{R}_0)$, in

$$1-x=e^{-\mathrm{R}_0' x}, \quad \text{or} \quad 1-\left(1-\frac{1}{\mathrm{R}_0}\right)=\frac{1}{\mathrm{R}_0}=\exp\left(-\mathrm{R}_0'\left(1-\frac{1}{\mathrm{R}_0}\right)\right),$$

or

$$\frac{\mathrm{R}_0'}{\mathrm{R}_0}=\frac{\log \mathrm{R}_0}{\mathrm{R}_0-1}.$$

If we allow the epidemic to run its course with this new, 'socially distanced' behaviour R_0', then when it is over and we revert to normal behaviour and R_0, there will not be a renewed epidemic, because each infected person will infect, on average, less than one additional person. For $\mathrm{R}_0=3$,

$$\frac{\mathrm{R}_0'}{\mathrm{R}_0}=\frac{\log 3}{2}\simeq 0.55,$$

so if we can halve our rate of infectious encounters then we will emerge from the socially distanced epidemic into a position of herd immunity.

Video: SIR model, pt 4

How can we avoid an epidemic?
There is only ever one way to escape an infectious disease – through reaching the threshold of herd immunity. Above we took this to be due to immunity acquired through having had the disease. But we can do it instead through vaccination, creating a fraction $1-u$ of the population who cannot get the disease. The unvaccinated fraction u are the only people who can get the disease; of the R_0 new infections which would have been caused by an infected individual in the absence of vaccination, now only a fraction u occur (because the other people encountered have been vaccinated), giving a new effective reproduction number of $\mathrm{R}_V=u\mathrm{R}_0$.

Suppose that, of the initial susceptibles $S_0=uN$, uxN will get the disease. Thus before the epidemic $S_0=uN$, $I_0\simeq 0$. After the epidemic we have $S=u(1-x)N$, $I=0$. The the equation for epidemic size has become

$$\frac{\gamma}{\beta}\log uN-uN=\frac{\gamma}{\beta}\log u(1-x)N-u(1-x)N$$

or

$$\frac{\beta uxN}{\gamma}=\log\left(\frac{1}{1-x}\right) \qquad \Rightarrow \qquad 1-x=e^{-\beta uxN/\gamma}\equiv e^{-\mathrm{R}_V x} \quad \text{where} \quad \mathrm{R}_V:=\beta uN/\gamma=u\mathrm{R}_0.$$

This is as we expected: R_0 has been replaced in the equation for epidemic size by R_V.

The highest known basic reproduction number (*i.e.* in an unvaccinated population with no acquired immunity) is for measles, $\mathrm{R}_0\simeq 15-20$. That's why, to prevent a measles epidemic, R_0 needs to be multiplied by a u which is around $1/20$: that is, the vaccination level needs to be up around 95%. For epidemic flu, R_0 may be around 5-6.

These high R_0 are typical of viral respiratory diseases which spread very effectively by causing their hosts to breathe out large quantities of virus for a long time before making their hosts ill. This works best in crowded, poorly ventilated rooms, which are a feature of modern wealthy societies, but not of most of our evolutionary history – recall that R_0 is determined by society as well as biology. Some people spend a great deal of time in crowded rooms; many do not. Thus an important extension of the SIR model is not to have R_0 the same for everyone, but rather to allow it to vary – some people may be 'superspreaders', not as a result of their biology but rather of their social interactions' creating superspreading events.

Another point to note concerns seasonality. For reasons which are not fully understood, respiratory viruses in temperate latitudes are seasonal. This is surely in part due to spending more time indoors in the winter, in part due to the virus surviving better in winter combinations of temperature and humidity, but possibly also due to aspects of host physiology. Thus *seasonal* flu is different from epidemic flu: a level of past exposure and thereby acquired immunity is reached in the population which is close to herd immunity, creating an effective R which is lower than one in the summer (hence no epidemic) but a little greater than one in winter (allowing a modest epidemic).

Video: SIR model, pt 5

How can we quantify the value of intervention?
Finally, we can do a nice calculation to quantify the effects of infection control when an epidemic is marginal, —that is, when R is only a little greater than one. Let

$$\mathrm{R} = 1 + \delta, \qquad x = \epsilon,$$

with both δ and ϵ small. Substituting into (3.16),

$$1 - \epsilon = 1 - (1 + \delta)\epsilon + \frac{1}{2}(1 + \delta)^2\epsilon^2 + \ldots,$$

the constant and linear terms cancel and the quadratic terms give $\epsilon \simeq 2\delta$. Every 1% reduction in R reduces x by 2%. This result can be seen in the gradient of the curve in Fig. 16, which is approximately 2 when R is small.

Consider ebola, a highly deadly disease with $R_0 \simeq 1.6$ in very poor societies, which can easily be reduced to close to one by changing people's funeral and other societal practices. If there is an ebola epidemic – necessarily with an effective R at least a little greater than one – how much is it worth to reduce it further? If you were a policy analyst and a politician asked you to quantify the human value of extra money spent on infection control, you would now have a (first, simple, approximate) answer.: every 1% reduction in R saves 2% of the population from the disease. □

Video: Modelling HIV/AIDS

Example 19: Modelling HIV/AIDS

In the SIR model, we assumed that the population was constant. Whether the disease was deadly or not — whether R was 'recovered' or 'died' — the point was that the timescale of the disease was much less than one generation. In the next model we impose no constancy of the population N, which can increase or decay. This is suitable for modelling a deadly disease with a very long timescale, comparable with that of population changes. The equations [16] are

$$\begin{aligned}\dot{U} &= cU - \kappa\frac{UI}{U+I} \\ \dot{I} &= -dI + \kappa\frac{UI}{U+I},\end{aligned}$$

where I is the number of infected and U the number of uninfected people, and c, d and κ are positive constants. We set $\kappa = n\epsilon$, where each person has n sexual contacts per year, and each encounter has a probability ϵ of causing infection.

Unlike the previous two examples, these equations can be solved exactly. We do so by setting $R = I/U$ and finding and solving a separable first-order equation for R, thus:

$$R = I/U \quad \Rightarrow \quad \dot{R} = \frac{U\dot{I} - I\dot{U}}{U^2} = \frac{1}{U^2}\left(-dIU + \kappa\frac{U^2 I + UI^2}{U+I} - cIU\right) = (\kappa - c - d)R\,,$$

so $R = R_0 e^{pt}$ where $p := \kappa - c - d$ and $R_0 = I_0/U_0$ is its initial value.

Then we substitute R into the equations for $\dot{U}$ and $\dot{I}$ to find $U(t)$ and $I(t)$, given that $U(0) = U_0$ and $I(0) = I_0$:

$$\dot{I} = -dI + \kappa\frac{I}{1+R} = \left(\frac{\kappa}{1+R_0e^{pt}} - d\right)I = \left(\kappa - \frac{\kappa R_0 e^{pt}}{1+R_0e^{pt}} - d\right)I\,,$$

which is separable and gives

$$\log\left(\frac{I}{I_0}\right) = -dt + \kappa t - \frac{\kappa}{p}\log\left(\frac{1+R_0e^{pt}}{1+R_0}\right)$$

which implies

$$I = I_0 e^{(\kappa-d)t}\left(\frac{1+R_0}{1+R_0e^{pt}}\right)^{\frac{\kappa}{p}}\,.$$

Similarly for U,

$$\dot{U} = cU - \kappa\frac{R}{1+R}U$$

so

$$\log\left(\frac{U}{U_0}\right) = ct - \frac{\kappa}{p}\log\left(\frac{1+R_0e^{pt}}{1+R_0}\right) \quad \Rightarrow \quad U = U_0e^{ct}\left(\frac{1+R_0}{1+R_0e^{pt}}\right)^{\frac{\kappa}{p}}\,.$$

So there's a crucial distinction, for both U and I, between $p > 0$ and $p < 0$. Look at the term in parentheses, where the exponential in the denominator respectively dominates for $p > 0$ and approaches zero for $p < 0$. So the dominant behaviour as $t \to \infty$ is

$$I \sim \begin{cases} I_0(1+R_0)^{\frac{\kappa}{p}} e^{(\kappa-d)t} & p < 0 \\ I_0(1+R_0^{-1})^{\frac{\kappa}{p}} e^{-dt} \to 0 & p > 0 \end{cases}$$

and

$$U \sim \begin{cases} U_0(1+R_0)^{\frac{\kappa}{p}} e^{ct} \to \infty & p < 0 \\ U_0(1+R_0^{-1})^{\frac{\kappa}{p}} e^{(c-\kappa)t} \to 0 & p > 0, \end{cases}$$

this last because $p > 0 \Rightarrow c - \kappa < 0$.

Thus the behaviour is very different for the two cases (i) $\kappa - c - d > 0$ and (ii) $\kappa - c - d < 0$. In case (i) the population will die out, while in case (ii) it will survive and increase, though there will still be a large number of infected people if $\kappa > d$.

Think about what all this means. If we wish to eradicate the disease, we need a high uninfected birth rate c, a low transmission rate κ, and (shockingly) a high infected death rate d. Clearly we must do everything possible to reduce transmission (κ). If we can get κ below $c + d$, the population will survive and increase, but only if we can also get κ below d will the disease die out. □

Just as with the interaction term in predator-prey models, one needs to think carefully about the scaling of the infection term in epidemiological models. In the SIR model, it was SI, which scales as the square of the overall population N: but this did not matter as N was constant. In the HIV/AIDS model, we used $\frac{UI}{U+I}$, which is in proportion to the overall population. This was because we assumed that sexual behaviour does not increase with population density.

Can we imagine a model which is like HIV/AIDS in the sense that it kills on very long timescales, comparable to a generation, but whose infectiousness increases in proportion to population density? Yes, quite easily: tuberculosis (TB) would be the prime example.

3.4 War

Again, two models here, from very different points of view. Richardson's model is born of pacifism, and an attempt to understand the baleful psychology of arms races. Lanchester's are combat models of attrition and annihilation, perhaps even more shocking in their implications.

Video: Richardson's Arms Race

Example 20: Richardson's arms race

Lewis Fry Richardson was a Quaker polymath who went to school in York. He made fundamental contributions to meteorology and numerical analysis. His pacifism led him to try to model the build-up to war. In the classic Richardson arms race — his prototype was the naval arms race before the First World War — two powers build up military forces $R(t)$ and $S(t)$ which vary with time [17]. Each side builds new units in proportion to the other side's overall force (with parameters r and s).

In the absence of antagonists each force tends to decay (with parameters ρ and σ). Of course nations' military forces do *not* dwindle away, so we add parameters k, l for fixed military expenditure, so that in the absence of the other side (that is, with $r = s = 0$) the forces are stable at $R = l/\rho$, $S = k/\sigma$. Then

$$\dot{S} = rR - \sigma S + k \tag{3.17}$$

$$\dot{R} = sS - \rho R + l\,. \tag{3.18}$$

Defining $\Delta := rs - \sigma\rho$, we find that we have an arms race — runaway, exponential growth of both R and S — when $\Delta > 0$. When $\Delta < 0$, we now have a stalemate, in which each side's force approaches a constant value greater than that in the absence of antagonism.

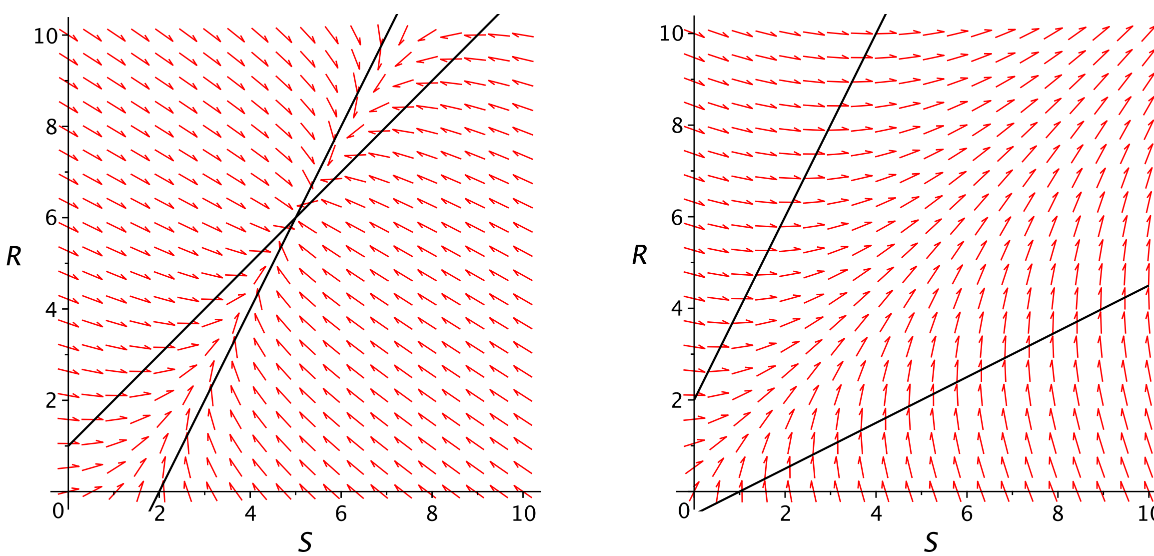


Figure 18: Richardson's arms race: (a) $\Delta < 0$ (low antagonism), fixed point in positive quadrant, stalemate; (b) $\Delta > 0$ (high antagonism), fixed point in negative quadrant, arms race.

The situation is made clear by the phase planes in Figure 18. Rather than show sample trajectories we give the directions of flow (in red) together with the null clines (black), the two curves (in this case, lines) on which $\dot{R} = 0$ and $\dot{S} = 0$. Note that R and S are necessarily positive, since $\dot{R} \geq 0$ on $S = 0$ and vice versa. $\square$

Note that the technique of looking at the null clines is really the extension to 2D of our simple 1D technique of drawing arrows on the real line to indicate the direction of the flow. The null clines, on which the flow is horizontal or vertical, enable us to divide the plane into regions in which the flow is everywhere up & rightwards, up & left, down & right, or down & left. The fixed points are where the null clines intersect.

Video: Lanchester's Linear Law

Example 21: Lanchester's 'ancient' and 'unaimed fire' models

Frederick Lanchester was another polymath, a gifted engineer who built the first British motor car (he invented the carburettor, the accelerator pedal and disc brakes) and made fundamental contributions to the theory of aerodynamics. His model was very different: it was intended to capture the nature of the new forms of warfare that seemed to be emerging when he was writing in 1913 [18].

Lanchester wanted to capture a crucial fact: that combat with long-range aimed weapons is different from hand-to-hand combat. Let R (Reds) and B (Blues) now be the number of units — individuals, perhaps — actively trying to fight on each side.

Before approaching his treatment of long-range weapons, we consider something simpler. In hand-to-hand fighting, in battles before the advent of long-range weapons, the side with the larger numbers cannot typically bring all of them into the fight, for there is only space in the battle line for roughly the same number N of combatants on each side. One might take N to be the lesser of R and B, perhaps. Then

$$\dot{R} = -bN \,, \qquad \dot{B} = -rN \,, \tag{3.19}$$

where the two sides' units' (possibly unequal) kill-rates are b and r. The outcome is determined by the fact that $rR - bB$ is constant throughout the battle, which ends when one side is annihilated. This result is quite intuitive: each side's fighting strength, its capacity to win a war of annihilation, is proportional to its units' numbers multiplied by their effectiveness. This is Lanchester's 'linear law', and (3.19) is his 'ancient warfare' model.

In fact the linear law applies to *any* circumstance in which the ratio of the loss rates $\dot{R}/\dot{B}$ is constant. Lanchester identified two natural such situations: the 'ancient warfare' described above, and 'unaimed fire' in which each side's attrition is proportional not only to its enemy's numbers but also, because the enemy fire is random, to its own density and hence numbers. Then both $\dot{R}$ and $\dot{B}$ are proportional to RB. Such a situation might obtain in an artillery duel, or if both sides were to adopt guerrilla tactics, using cover, concealment and dispersion to reduce their own losses. □

Video: Lanchester's Square Law

Example 22: Lanchester's 'aimed fire' model and Lanchester's Square Law

Lanchester's main insight describes how this situation changes when long-range, aimed weapons are available. Then, he argued, there is no restriction on numbers engaged, and each force can do damage in proportion to its numbers, so that

$$\dot{R} = -bB\,, \qquad \dot{B} = -rR\,. \tag{3.20}$$

This is **Lanchester's aimed-fire model**. It applies when all units of both sides find it easy to acquire targets.

Now we divide, separate variables and integrate: then

$$\int rR\,dR = \int bB\,dB \qquad \Rightarrow \qquad rR^2 - bB^2 = \text{constant.} \tag{3.21}$$

This has remarkable implications. Since this quantity never changes sign, in general only one of R and B can ever be zero. If the initial value of (3.21) is positive, for example, we cannot have $R = 0$ and $B > 0$. Rather only B can equal zero, at a time at which $R > 0$, and so only red can win[12]. If the initial value is negative, then only Blue can win. (In the exceptional case $rR^2 - bB^2 = 0$, the forces are balanced and fight until their mutual annihilation, $R = B = 0$.) So this quantity is key to determining who will win the battle, and a side's fighting strength — we shall call rR^2 and bB^2 the red and blue forces' respective **fighting strengths** — varies as the units' fighting effectiveness times the *square* of their numbers, **Lanchester's square law**.

Consider an example. Suppose red begins with twice as many units as blue, $R_0 = 2B_0$, but the blue units are three times as effective, $b = 3r$. Then

$$rR^2 - bB^2 = r(2B_0)^2 - 3rB_0^2 = rB_0^2 > 0,$$

and (perhaps rather counter-intuitively) the reds win. The tactical conclusion is simple: if your strength is in numbers, then you need to bring all your units to engage with the enemy's as rapidly as possible. Conversely, if your units are fewer in number but more effective, then you need the tactics which will prevent this, allowing you to pick-off opponents, and preventing the enemy from bringing all his units to bear (— like Hannibal at Cannae). This last point becomes clearer if we consider what would have happened had blue been able to divide the reds into two equal forces and engage them sequentially. At the end of the first engagement, between B_0 blues and $R_0 = B_0$ reds, B_1 blues remain, where

$$rB_0^2 - 3rB_0^2 = -3rB_1^2 \qquad \Rightarrow \qquad B_1 = \sqrt{\frac{2}{3}}B_0,$$

[12] To show that red actually does win, we would need to solve the equations explicitly for $R(t)$ and $G(t)$, which is most easily done by differentiating one of (3.20), substituting into the other, and solving the resulting second order ODE.

and then in the second engagement

$$rB_0^2 - 3r\frac{2}{3}B_0^2 = -rB_0^2 < 0.$$

Blue has now won, with $3rB_2^2 = rB_0^2$ and thus $\sqrt{\frac{1}{3}}$ or nearly 60% of its original forces left when all the reds have been destroyed — an amazing turnaround. This becomes even more striking with an N-fold (rather than two-fold) division of red forces: after the simple resulting iteration, blue now wins with a final number B_F of units remaining, where

$$bB_F^2 = bB_0^2 - Nr\left(\frac{R_0}{N}\right)^2,$$

so that the N-fold division has reduced red's fighting strength N-fold. In military parlance this is 'defeat in detail', and the implication is a classic military maxim: you should (almost) never divide your forces.

Now let's go back to the equations of the aimed fire model (3.20). For the sake of argument, set $b = 4r$. What happens if we sketch the phase plane? Well, it looks rather dull, for there are no non-zero constant solutions, no non-trivial fixed points.

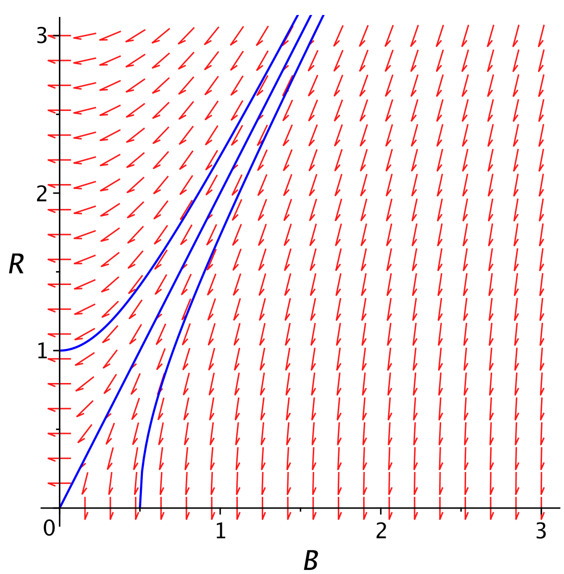


Figure 19: The Lanchester aimed fire model with $r = 0.1, b = 0.4$.

We've drawn three trajectories, however. Each is a particular value of the conserved quantity $4B^2 - R^2$. The straight line is $4B^2 - R^2 = 0$, or (within the positive quadrant) $R = 2B$, and it leads to mutual annihilation; neither side wins. The others are the Blue win $4B^2 - R^2 = 1$ and the Red win $4B^2 - R^2 = -1$. These are hyperbolae: notice that the imbalance between the forces increases as the battle approaches a conclusion.

Actually we can do even better than that: the equations are linear, so we can solve them exactly. Differentiating $\dot{R}$ we find $\ddot{R} = -b\dot{B} = brR$ so that solutions are linear combinations of $e^{\pm\lambda t}$ where $\lambda = \sqrt{br}$. We use the combinations $\sinh \lambda t$ and $\cosh \lambda t$, and find that the correct combination for initial values R_0, B_0 are

$$\begin{aligned} R &= R_0 \cosh \lambda t - \sqrt{\frac{b}{r}} B_0 \sinh \lambda t \\ B &= B_0 \cosh \lambda t - \sqrt{\frac{r}{b}} R_0 \sinh \lambda t. \end{aligned}$$

These describe hyperbolae, and it is straightforward to verify that $rR^2 - bB^2 = rR_0^2 - bB_0^2$.

But the existence of the conserved quantity of the 'square law' helped us to think about the implications of the model more clearly than did the explicit solutions or the phase plane! □

Moral of this Section:

We have begun to use calculus to study *interactions* between two dynamical (changing) variables. This is important in ecology (predator *vs* prey), war (predator *vs* predator), epidemiology of infections (infected *vs* susceptible), and so on.

We still see fixed points, and growth or decay. But now we may also have oscillations — sometimes steady ones like an SHO, but sometimes much stranger, 'jumpy' ones.

In many cases a useful technique, generalizing the phase lines of the previous section, is to draw a phase plane, a 2D plot of the dependent variables indicating the direction of flow, perhaps with some sample trajectories. To do this it is useful to draw the null clines, the curves on which the flow is either horizontal or vertical. One can then see the general direction of the flow in each of the regions separated by the null clines. These intersect at the fixed points, whose stability can then easily be deduced.

We can usefully look at the behaviour very close to a fixed point, discarding any terms which are not linear in the distance from it ('linearization'). This is a useful more general technique in applied maths: consider small variations around a point of interest, discarding higher powers of these variations. This is equivalent to, but simpler to execute than, formal computation of a Taylor series.

Sometimes we can make progress by proving that some function of the dependent variables is a constant (just as $x^2 + y^2$'s being constant under $\dot{x} = y, \dot{y} = -x$ indicates motion in a circle). When variables are changing, it's good to know about the things that do not change.

4 Partial Differential Equations

What you should be able to do by the end of this section:

- Recognize the wave and heat equations, and be able to solve them on the finite interval using the Fourier sine series;
- Solve the wave equation for the initial conditions of a plucked or struck string, and interpret the solutions;
- Solve the heat equation with Dirichlet or Neumann boundary conditions, and interpret the solutions;
- Solve the heat equation on the half-line with a periodic condition at one boundary;
- Be able to extend such results to simple reaction-diffusion systems, including the verification of travelling wave solutions.

Video: Partial Differential Equations

Our world has three space and one time dimension (we call this '3+1D'). To describe continuous processes we therefore need four independent variables, x, y, z and t. To differentiate with respect to these, independently, we need *partial derivatives*, and instead of the 'ordinary' diferential equations which have only one independent variable, we need **partial differential equations** (PDEs).

4.1 The Wave Equation and the Heat Equation

The good news is that one can get a very long way, and describe many phenomena, with just two terms. The first is the **Laplacian**, which, applied to a dependent variable $\varphi(x, y, z, t)$, is

$$\nabla^2\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2}.$$

To understand this, and indeed to do calculus in 3D, you need to use vector methods, and study 'vector calculus'. So we shall look only at PDEs in 1+1D, where the Laplacian of $\varphi(x, t)$ is just the second partial derivative $\partial^2\varphi/\partial x^2$.

The second term is a time-derivative, either first or second order. There are then two prototypical PDEs:

The Wave Equation

$$\frac{\partial^2\varphi}{\partial t^2} = c^2\frac{\partial^2\varphi}{\partial x^2} \tag{4.22}$$

has solutions in which $\varphi(x,t) = f(x-ct) + g(x+ct)$ where f, g may be *any* functions (due to d'Alembert; see problem 3.1). This describes a shape f moving to the right at speed c and another g moving to the left, and these pass through each other. Here φ could take many forms — it might be the transverse (*i.e.* perpendicular to the x-axis) displacement of a string, or of water; or the longitudinal (*i.e.* along the x-axis) displacement of air, in sound waves; or even electric and magnetic fields.

The Heat Equation or **the Diffusion Equation**

$$\frac{\partial \varphi}{\partial t} = \kappa \frac{\partial^2 \varphi}{\partial x^2}, \tag{4.23}$$

with its constant **thermal diffusivity** κ, describes anything which diffuses — 'spreads out' — including, for example, the diffusion backwards in time of certainty about the price of a ('derivative') financial contract [21]. It has a conservation law:

$$\frac{d}{dt}\int_0^L \varphi(x,t)\,dx = \int_0^L \frac{\partial \varphi}{\partial t}\,dx = \kappa \int_0^L \frac{\partial^2 \varphi}{\partial x^2}\,dx = \kappa \frac{\partial \varphi}{\partial x}(L,t) - \kappa \frac{\partial \varphi}{\partial x}(0,t). \tag{4.24}$$

For example, if $\varphi(x,t)$ is the temperature of a rod, then the total heat in the rod is proportional to $\int_0^L \varphi\,dx$, and changes only when the temperature gradient at the ends of the rod is non-zero. (The relative minus sign is because the heat in the rod increases when the temperature goes up as one approaches either end of the rod.)

An essential difference between the two equations is that the wave equation has a second-order time derivative and thus oscillatory solutions, in the form of waves. It is something like having an SHO at every point, coupled to its neighbours. The heat equation, in contrast, has a first-order time derivative, and solutions that merely grow or decay. As we shall see at the end of the module, there can be travelling solutions in variants of the heat equation, but these are in the form of single moving wavefronts (like a tsunami) rather than oscillatory wave-trains.

Boundary Conditions: The two most commonly encountered types of boundary conditions (BC) for these equations are **Dirichlet** conditions, in which φ is held fixed at the boundary, $\varphi = 0$ or =constant, and **Neumann** conditions, in which the derivative of ψ is zero at the boundary, $\partial\varphi/\partial x = 0$. Under Neumann BCs, the two boundary terms in (4.24) are zero, and heat is conserved. Other BCs are possible: for example the **Robin** condition interpolates between Neumann and Dirichlet.

Video: Waves on a Stringed Instrument

Example 23: Standing waves on a stringed instrument

Our goal is to solve

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} \tag{4.25}$$

where the dependent variable is the displacement of a string from its resting position along the x-axis, where it is stretched between $x = 0$ and $x = L$. It is therefore natural to call it $y(x, t)$. Because this is a PDE, with two independent variables, we need boundary conditions (on x; 'BCs') *and* initial conditions (at $t = 0$; 'ICs'). The BCs are that the string is fixed at both ends: $y(0, t) = 0$ and $y(L, t) = 0$ for all t.

This is (in the language of the note above) the Dirichlet condition. For our string the Neumann BC would be rather contrived: instead of fixing the end of the string, we would have to attach a very light ring to it which could move freely up and down a wire. But the Neumann BC does occur naturally in a musical instrument: it applies to the vibrations of the air at the open end of a wind instrument.

We shall now solve the equation, imposing *initial* conditions (such as plucking or hitting the string) later. The key trick is to seek solutions in the form of the product of a function of x with a function of t, a technique called **separation of variables**, so that

$$y(x, t) = X(x)T(t),$$

for then the partial derivatives become ordinary ones: the wave equation becomes

$$X\ddot{T} = c^2 X'' T,$$

where dots denote t-derivatives and primes x-derivatives. Dividing by XT we see that

$$\frac{\ddot{T}}{T} = c^2 \frac{X''}{X},$$

and we then notice that the lhs ('left-hand side') is a function of t only, and does not depend on x, while the rhs is a function of x only (and does not depend on t). But the two sides are equal: so they cannot depend on either t or x, and must be constant.

Suppose that constant is Kc^2. If $K > 0$ then the solution of $X'' = KX$ is $\exp(\pm\sqrt{K}x)$, neither of which is ever zero and so which cannot satisfy the BCs. If $K < 0$ (and we write $K = -c^2\alpha^2$) then the solutions of $X'' = -\alpha^2 X$ are sines and cosines of αx. If we choose $X(x) = \sin\alpha x$ we will satisfy the BC at $x = 0$, because $X(0) = 0$. But what about $x = L$? Well, $\sin n\pi = 0$ for any integer n, so choosing $\alpha = n\pi/L$ will give us $X(L) = 0$.

Then we have to find $T(t)$, which satisfies $\ddot{T} = -(n\pi c/L)^2 T$. The solution of this could be any combination of a sine and a cosine, so we leave the coefficients as a_n and b_n. The wave equation is linear, so we can add together such solutions. Then

$$y(x, t) = \sum_{n=1}^{\infty} \left\{ a_n \cos\frac{n\pi ct}{L} + b_n \sin\frac{n\pi ct}{L} \right\} \sin\frac{n\pi x}{L}.$$

is the general solution for the BCs $y(0,t) = 0 = y(L,t)$.

The initial conditions are what will determine the a_n and b_n. Suppose we know the position and velocity of the string at $t = 0$: that is,

$$y(x,0) \qquad \text{and} \qquad \frac{\partial y}{\partial t}(x,0).$$

Then, comparing with the general solution, we have

$$\begin{aligned} y(x,0) &= \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L}, \\ \frac{\partial y}{\partial t}(x,0) &= \sum_{n=1}^{\infty} \frac{n\pi c}{L} b_n \sin\frac{n\pi x}{L}. \end{aligned}$$

We can then use Fourier series to determine a_n and b_n. To do so we multiply each of these equations by $\sin\frac{m\pi x}{L}$ for some specific m and integrate from 0 to L, which is non-zero for precisely one (the case $n = m$) of the infinitely-many terms on the right-hand side. Then

$$\begin{aligned} a_m &= \frac{2}{L}\int_0^L y(x,0)\sin\frac{m\pi x}{L}\,dx, \qquad (4.26) \\ b_m &= \frac{2}{m\pi c}\int_0^L \frac{\partial y}{\partial t}(x,0)\sin\frac{m\pi x}{L}\,dx. \end{aligned}$$

The proof (non-examinable) is given in an appendix.

We consider two specific initial conditions:

The string is plucked in the middle, pulling it to one side in the middle to make a triangular shape, so that

$$y(x,0) = \begin{cases} dx & 0 \le x \le L/2 \\ d(L-x) & L/2 \le x \le L \end{cases},$$

where d is a positive constant. We release the string from rest in this position, so

$\partial y/\partial x = 0$ at $t = 0$. Thus $b_n = 0$ for all n, while

$$\begin{aligned}
a_n &= \frac{2d}{L}\left\{\int_0^{L/2} x\sin\frac{n\pi x}{L}\,dx + \int_{L/2}^{L}(L-x)\sin\frac{n\pi x}{L}\,dx\right\} \\
&= \frac{2d}{L}\frac{L}{n\pi}\left\{\left[-x\cos\frac{n\pi x}{L}\right]_0^{L/2} + \left[-(L-x)\cos\frac{n\pi x}{L}\right]_{L/2}^{L} + \int_0^{L/2}\cos\frac{n\pi x}{L}\,dx - \int_{L/2}^{L}\cos\frac{n\pi x}{L}\,dx\right\} \\
&= \frac{2d}{L}\left(\frac{L}{n\pi}\right)^2\left\{\left[\sin\frac{n\pi x}{L}\right]_0^{L/2} - \left[\sin\frac{n\pi x}{L}\right]_{L/2}^{L}\right\} \\
&= \frac{4dL}{n^2\pi^2}\sin\frac{n\pi}{2} \\
&= \begin{cases} 0 & n = 2p \\ \frac{4dL}{(2p+1)^2\pi^2}(-1)^p & n = 2p+1 \end{cases}
\end{aligned}$$

The string is hit in the middle, so that, while at $y(x,0) = 0$, it is instantaneously given a velocity v in the interval $(L(1/2-\delta), L(1/2+\delta))$:

$$\frac{\partial y}{\partial t}(x,0) = \begin{cases} v & L(\frac{1}{2}-\delta) \le x \le L(\frac{1}{2}+\delta) \\ 0 & \text{otherwise} \end{cases}$$

Then $a_n = 0$, while

$$\begin{aligned}
b_n &= \frac{2v}{n\pi c}\int_{L(1/2-\delta)}^{L(1/2+\delta)}\sin\frac{m\pi x}{L}\,dx \\
&= \frac{2v}{n\pi c}\frac{L}{n\pi}\left[-\cos\frac{n\pi x}{L}\right]_{L(1/2-\delta)}^{L(1/2+\delta)} \\
&= \frac{2vL}{n^2\pi^2 c}\left(\cos n\pi(1/2-\delta) - \cos n\pi(1/2+\delta)\right) \\
&= \frac{4vL}{n^2\pi^2 c}\sin\frac{n\pi}{2}\sin n\pi\delta.
\end{aligned}$$

Now, if δ is small then, for small n, $n\pi\delta \ll 1$ and $\sin n\pi\delta \simeq n\pi\delta$, so that

$$b_n \simeq \frac{4vL\delta}{n\pi c}\sin\frac{n\pi}{2}.$$

Contrast this with the plucked string. For the plucked string the coefficients fall off as $1/n^2$; for the struck string they fall off more slowly, as $1/n$. Thus one hears a 'purer' note, with more of the fundamental relative to the overtones, for a plucked than for a struck string — on a guitar compared with a piano, say. □

I explain how to use the notes corresponding to these n to build musical scales (non-examinable) in an appendix.

Video: Heat Diffusing in a Rod

Example 24: Heat diffusing in a rod

Here $\varphi(x,t)$ will describe the temperature of a rod of length L at a distance x along its length, with $0 \le x \le L$. The equation is therefore

$$\frac{\partial \varphi}{\partial t} = \kappa \frac{\partial^2 \varphi}{\partial x^2}. \tag{4.27}$$

The boundary conditions describe the heat flow at the end-points. For example, if the end-points are maintained at constant temperatures T_{left} and T_{right} by some external heat source (or sink) then the appropriate boundary conditions would be Dirichlet,

$$\varphi(0,t) = T_{\text{left}}; \qquad \varphi(L,t) = T_{\text{right}}$$

for all t. Alternatively, if the end-points are insulated so that no heat may flow in or out of them then the appropriate boundary conditions would be Neumann,

$$\frac{\partial \varphi}{\partial x}(0,t) = \frac{\partial \varphi}{\partial x}(L,t) = 0$$

for all t, because the heat flow is proportional to temperature gradient $\partial\varphi/\partial x$. Then, as we saw earlier, $\int_0^L \varphi\, dx$ is constant. Of course, one can also imagine having different BCs at the two ends (e.g., left hand end insulated; right hand held at temperature T_{right}). In this example we will hold the ends of the rod in iced water, so that, if φ is measured in degrees Celsius, we have the Dirichlet conditions $\varphi(0,t) = 0 = \varphi(L,t)$.

To begin with we proceed exactly as for the wave equation: we set $\varphi = X(x)T(t)$, separate variables to obtain

$$\frac{\dot{T}}{T} = \kappa \frac{X''}{X} = \text{constant},$$

and insist that that constant be $-\kappa n^2\pi^2/L^2$ so that $X = \sin\frac{n\pi x}{L}$, which is zero at $x = 0$ and $x = L$. But then having the first instead of the second derivative of T with respect to t results in

$$\dot{T} = -\kappa \left(\frac{n\pi}{L}\right)^2 T \qquad \Rightarrow \qquad T(t) = e^{-\kappa n^2\pi^2 t/L^2},$$

so that

$$\varphi(x,t) = \sum_{n=1}^{\infty} a_n e^{-\kappa n^2\pi^2 t/L^2} \sin\frac{n\pi x}{L}.$$

Note that there are no b_n: this is a first-, not a second-order equation, so there is only one free coefficient in the ODE for T for each choice of n.

We can find the a_n by using Fourier methods identical to those in Ex. 23, since $\cos 0 = e^0 = 1$:

$$a_n = \frac{2}{L}\int_0^L \varphi(x,0)\sin\frac{n\pi x}{L}\, dx.$$

Now we move on to implement some initial conditions. In particular, consider

A uniformly hot rod: let the rod initially be at a uniform high temperature $\varphi(x,0) = T$ for $0 < x < L$. Then

$$a_n = \frac{2T}{L}\int_0^L \sin\frac{n\pi x}{L}\,dx = \frac{2T}{L}\left[-\frac{L}{n\pi}\cos\frac{n\pi x}{L}\right]_0^L = \frac{2T}{n\pi}\{1-(-1)^n\},$$

so that

$$\varphi(x,t) = \frac{4T}{\pi}\sum_{n\ \mathrm{odd}}^{\infty}\frac{1}{n}e^{-\kappa n^2\pi^2 t/L^2}\sin\frac{n\pi x}{L}.$$

Notice that all terms decay to zero as t increases: the rod ends up uniformly at $\varphi = 0$, the temperature of the iced water. But some terms decay faster than others: the first term decays most slowly, so, as the rod cools, its temperature distribution looks increasingly like a simple sine-shaped bulge proportional to $\sin\frac{\pi x}{L}$. □

Video: Temperature in a Cellar

Example 25: What is the temperature in a cellar?

Here is another effect which we can determine using the heat equation. The sun shines on the surface of the earth, heating it, in both annual and daily cycles. What is the temperature deep in the earth? We might expect it also to show annual and daily cycles, but the amplitude of the variation will be smaller, and of course we can work this out. But there's also a *delay* in each cycle, as the heat takes time to diffuse into the ground, and this too we can determine.

So we solve the heat equation (4.27) on the half-line, with depth $x \in [0,\infty)$, and the sun heating the ground so that

$$\varphi(0,t) = T\sin\omega t, \qquad \varphi(\infty,t) = 0,$$

where $2\pi/\omega$ is either one year or one day and $2T$ is the difference between highest and lowest (annual or daily) temperatures. We try a solution of the form

$$\varphi(x,t) = Te^{-ax}\sin(\omega t - bx)$$

where a, b are positive real constants to be determined (a will give us the attenuation – decay – of the signal as x increases downwards, while b will give us the time delay). Substituting into (4.27) we obtain

$$T\omega e^{-ax}\cos(\omega t - bx) = \kappa T\left\{a^2e^{-ax}\sin(\omega t - bx) + 2abe^{-ax}\cos(\omega t - bx) - b^2e^{-ax}\sin(\omega t - bx)\right\}.$$

We equate coefficients of $e^{-ax}\sin(\omega t - bx)$ and $e^{-ax}\cos(\omega t - bx)$ to give

$$a^2 - b^2 = 0, \qquad \omega = 2ab\kappa,$$

so that $b = a$ and $a = \sqrt{\omega/2\kappa}$. Thus

$$\varphi(x, t) = Te^{-x\sqrt{\omega/2\kappa}} \sin(\omega t - x\sqrt{\omega/2\kappa}).$$

(Let us pause to check the dimensions. First, $[\varphi] = [T] = \Theta$, a temperature. Then $[\omega] = T^{-1}$, so that $[\omega t] = 1$, while $[\kappa] = L^2 T^{-1}$ as previously. So $[a] = (T^{-1} T L^{-2})^{1/2} = L^{-1}$, and then $[ax] = 1$. Good.)

Now let us put in some realistic values. Let's consider an annual cycle ($\omega = 2\pi$ year^{-1}) in England, with $T \simeq 10°$ Celsius, and in clay soil with $\kappa \simeq 24$ m^2 per year. At a depth of $3m$, then,

$$ax = x\sqrt{\frac{\omega}{2\kappa}} \simeq 3\sqrt{\frac{\pi}{24}} \simeq 1.1,$$

so the annual variation is $e^{-ax} \simeq 0.33$ of that at the surface, or about 7°C instead of 20°C, and the delay is about $12 \times 1.1/2\pi \simeq 2$ months. The delay is of course greater at greater depths: indeed, when $ax = \pi$, which happens at $x \simeq 8.7m$, the temperature is highest in the middle of winter! — although the variation is only $e^{-\pi}$ or about 0.04 of that at the surface. $\square$

4.2 Reaction-diffusion systems

If we had more time, we would put all of this together. We would have the dynamical systems of Section 3 – systems of coupled 1st order ordinary DEs in which the dependent variables' time-evolution is specified as functions of each other – and we would add spatial derivatives to allow what happens at one point to spread out and influence what happens elsewhere.

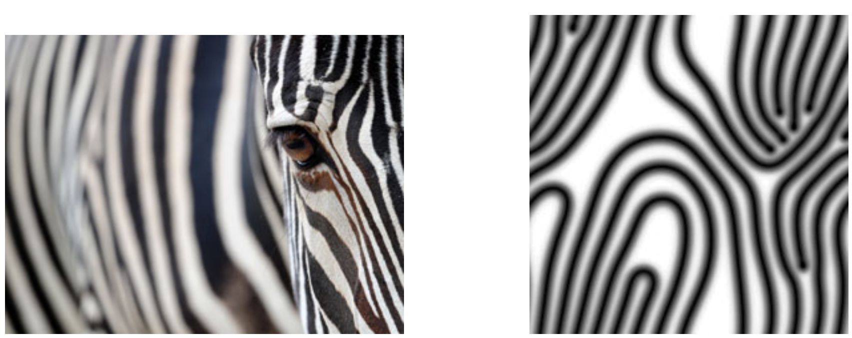


Figure 21: The stripes on a zebra, and stripes in the Turing system.

A classic example is the generation of patterns such as the stripes on a zebra. Alan Turing, more famous for his pioneering work in computing and code-breaking, wrote down the prototypical system which can produce such phenomena, involving the coupling of two reacting chemicals, a fast 'inhibitor' and a slow 'activator'. The result of this system and its extensions is patterns such as those in Figure 21.

Most of the interesting examples appear when one adds spatial diffusion to a system of at least two coupled DEs. Here we can only look at a few of the simplest, single 1+1D reaction-diffusion equations. One phenomenon is the balancing of exponential growth against the diffusion (second spatial derivative) term's tendency to make the dependent variable spread out and thereby decay (see Problems 3). In the following example we'll look at a prototype of the kind of phenomenon one observes in the Turing system, a travelling wave in which a change in stable fixed point travels in the x direction.

Video: A Reaction-Diffusion System

Example 26: Travelling waves of colour

Consider the equation

$$\frac{\partial U}{\partial t} = aU(\alpha - U)(U-2) + b\frac{\partial^2 U}{\partial x^2},$$

where a, b and α are all positive constants. First, we assume that, perhaps by re-scaling the time t and space x variables, we can set $a = 1$ and $b = 1$. Then

$$\dot{U} = U(\alpha - U)(U-2) + U''. \tag{4.28}$$

Now think about this when there is no space dependence, so that $U'' = 0$. The equation is of just the sort we looked at in Section 2 (Ex. 12, the Allee effect). If we sketch the rhs, we have a cubic with zeros at 0, α and 2, and which is negative for $0 < x < \alpha$ (and when $2 < x$), and positive for $\alpha < x < 2$ (and when $x < 0$). So the DE has stable fixed points (FPs) at 0 and 2, and an unstable FP at α. You could imagine these as two colours: 0 for white and 2 for black, say.

Now we allow some space variation, in the form of a diffusion term U'', again. We'll show that there is an exact 'travelling wave' solution

$$U(x,t) = v(x - ct) \qquad \text{where} \qquad v(z) = 1 + \tanh\frac{z}{\sqrt{2}},$$

where c is a speed which we shall determine, and which will depend on α.

First, consider $v(z)$. As $z \to -\infty$, we have $v \to 0$, while as $z \to \infty$, we have $v \to 2$, so this solution is white on the far left, black on the far right. Now substitute it into (4.28), and use

$$v'(z) = \frac{1}{\sqrt{2}}\operatorname{sech}^2\frac{z}{\sqrt{2}}, \qquad v''(z) = -\operatorname{sech}^2\frac{z}{\sqrt{2}}\tanh\frac{z}{\sqrt{2}}.$$

Then (4.28) becomes

$$-\frac{c}{\sqrt{2}}\operatorname{sech}^2\frac{z}{\sqrt{2}} = -\operatorname{sech}^2\frac{z}{\sqrt{2}}\tanh\frac{z}{\sqrt{2}} + (\tanh^2\frac{z}{\sqrt{2}} - 1)(\alpha - 1 - \tanh\frac{z}{\sqrt{2}}).$$

Now $\text{sech}^2 = 1 - \tanh^2$, so the $\text{sech}^2 \tanh$ terms cancel, and the others are equal if

$$c = \sqrt{2}(\alpha - 1).$$

If $\alpha > 1$, and the unstable FP is closer to black than to white, then the wave moves to the right: everything becomes white in the end. If $\alpha < 1$, and the unstable FP is closer to white, then the wave moves left and everything becomes black. In the middle, at $\alpha = 1$, we have $c = 0$ and a stationary solution is possible, in which the 'kink' is static, with white on the left and black on the right. □

Exercise: Show that there is another solution which is black on the far left and white on the far right. This is achieved by changing the sign of the tanh term in v: you can either argue or verify that this is still a solution of the equation, but with α replaced by $-\alpha$.

Moral of this Section:

PDEs have more than one independent variable: here, at their simplest, time t and 1D space x.

Two prototypical PDEs are the wave equation and the heat equation. The former has oscillating solutions; the latter models diffusion, and does not.

To solve a PDE one needs both boundary conditions (typically either fixed, Dirichlet; or free, Neumann) and initial conditions.

We can model the making of music by plucking or striking a string, fixed at both ends; and the diffusion of heat in a rod.

Reaction-diffusion systems put together everything we have learned: growth/decay or coupled systems whose behaviour at each point also diffuses to influence behaviour elsewhere. We've only looked at a couple of the simple travelling single waves that can appear when we add spatial diffusion to one of the differential equations of section 2. But such systems govern a wide-range of real-world behaviour – for example, it is when we add spatial diffusion to the van der Pol oscillator that we get closer to modelling the behaviour of a real heartbeat!

5 Glossary

A guide to terms used in the text, and a basic reference for applied mathematics. Meanings of some terms (marked with an asterisk) need to be known. You will not be asked to recall the others, although you should be familiar with the contexts and models in which they appear.

the **Allee effect** in biology: that mean individual fitness increases with population size. Implemented in a 1D growth model which extends the Bertalanffy model to include a third term, and in which extinction is possible for some ranges of parameters and initial conditions.

allometry: a biological scaling law with a power other than one

an **autonomous** ODE or dynamical system is one in which the independent variable does not explicitly appear.

Bertalanffy's equation: a first-order ODE in which a limit to growth is imposed by having monomial growth and decay terms with differing powers

***boundary condition(s)**: in a differential equation, the condition on a dependent variable at the limits of its spatial domain. For PDEs, the most common are **Dirichlet**, in which the dependent variable is held fixed at the boundary, and **Neumann**, in which the spatial derivative of the dependent variable is zero (*i.e.* the dependent variable is flat).

***Buckingham's Π theorem**: the result of counting physical dimensions and number of variables in a mathematical problem in applied mathematics. The theorem states that if there is an equation relating m quantities with n independent underlying dimensions then it takes the form of a relationship among $m - n$ dimensionless numbers

chaos: complex, structured but unpredictable behaviour found only in *third* or higher order dynamical systems

***conserved quantity** or **constant of the motion**: quantity that remains constant under the evolution of a dynamical system, such as (total, kinetic plus potential) energy in a frictionless pendulum

d'Alembert's solution of the wave equation is *any* function of $x - ct$ or of $x + ct$ or any superposition of the two. These correspond to right- and left-moving (along the positive x-axis) 'travelling waves'. In contrast 'standing waves' are solutions which appear to be oscillating without travelling.

difference equation: a recursion relation which describes changes over discrete time steps, as opposed to **differential equations** in continuous time

*__dimensionless number__ or **dimensionless group**: a dimensionless product of powers of dimensionful quantities, which is therefore a number

drag, the retarding force exerted on a moving object by the medium in which it moves, may specialise to **Stokes drag**, proportional to the speed and length scale and applicable in situations of high viscosity, or **Rayleigh drag**, proportional to the squares of the speed and of the length, and applicable at low viscosity

dynamical system: a set of two or more coupled first-order ODEs, with time as the independent variable

*__exponential__ growth (over time t) has leading term of the form e^{kt}, as opposed to **polynomial** growth in which the leading term has the form t^k

first-order dynamics uses only first time-derivatives; in a single ODE it results in growth and decay (often via polynomial and exponential functions), but oscillations do not emerge unless a periodic function appears explicitly in the ODE. Then the **second-order dynamics** of a single second-order ODE or two coupled first-order ODEs can result in growth, decay or oscillations—perhaps simple, as in SHM; or more subtle, as in the Lotka-Volterra system; or as limit cycles, as in the van der Pol oscillator. Chaos only appears in $\geq$3D dynamical systems.

*__fixed point__: constant solution of a dynamical system or differential equation

Gompertz's equation: a first-order ODE in which the growth rate declines exponentially

the *__half-life__ is the time taken for halving of an exponentially decaying variable; the **doubling time** is the time for an exponentially growing variable to double

heat equation or **diffusion equation**: the PDE, in which the first time derivative of the dependent variable is proportional to its second space derivative, which describes the diffusion of heat and other spatial diffusion processes. Used in the Black-Scholes model of pricing financial contracts.

*__initial condition(s)__: in a differential equation, the condition(s) on a dependent variable at the beginning of its time evolution.

the **Kermack-McKendrick (SIR) model** is a dynamical system which describes the infectious dynamics and thereby the epidemiology of a fast-acting (relative to human life-span) disease

Kleiber's law is the empirical observation that animals' power consumption (metabolism) scales as the 3/4 power of their mass

Lanchester model: one of three models ('ancient', 'unaimed fire', 'aimed fire') which describe attrition between two armed forces by means of a 2D dynamical system. In each case, and with R Red fighting B Blue units, $bB^k - rR^k$ is constant, with $k = 1$ in the first two cases (the 'linear law') and $k = 2$ in the last case (the 'square law').

Laplacian: the second spatial derivative which emerges from Vector Calculus and appears ubiquitously in PDEs to describe diffusion. With a single spatial dimension x it is $\partial^2/\partial x^2$; in Cartesian coordinates for higher dimensions it is the sum of the second spatial derivatives with respect to each of the coordinates. In polar and other coordinates it takes more complicated forms.

limit cycle: a periodic solution of a 2D dynamical system to which other trajectories in the region tend in the large-time limit.

linearization: to *__linearize__ is to consider small variations about a particular solution to a nonlinear equation and then discard all higher-order terms in that variation, thus creating a linear equation. Commonly done about a fixed point of a nonlinear dynamical system to create a linear and thus explicitly solvable dynamical system.

Lotka-Volterra model (sometimes thought of as taking place on a **Lotka-Volterra island**): simple ecosystem of one predator and one prey species described as a 2D dynamical system. The model has a conserved quantity and exhibits periodic behaviour.

mass action principle, usually in chemical reaction kinetics: that the chemicals are well-mixed so that the dynamics do not vary spatially

Michaelis-Menten dynamics impose a limit to the rate of growth: growth that is initially exponential eventually levels off to a constant growth rate.

*__non-dimensional__: starting from an equation relating dimensionful quantities, to **non-dimensionalize** it is to recast it in terms only of dimensionless numbers, in a **scale-free** equation

*__null cline__: in a phase plane, the curves on which one of the two (dependent variables') rates of change vanishes

*__ordinary differential equation (ODE)__: differential equation with one independent variable

oscillator: periodic solution of a (2nd or higher) order ODE or dynamical system

partial differential equation (PDE): differential equation with more than one independent variable

*__phase line__ (or sometimes **flow diagram**): the real line, describing the dependent variable in a first-order growth model, marked with fixed points and an arrow in each region between two points corresponding to the sign of the derivative of the dependent variable and thus to the direction of flow

*__phase plane__ (or sometimes **phase portrait** or **phase diagram**): a Cartesian plot of the dependent variables in a 2D coupled dynamical system, usually with some arrows to indicate the direction of flow, fixed points marked, and possibly some sample trajectories

reaction-diffusion system: a PDE or system of coupled PDEs which mixes a Laplacian diffusion term with other, reaction terms analogous to those found in ODEs and dynamical systems. The interplay between the two can result in complex phenomena such as pattern formation.

Richardson arms race: a 2D dynamical system which describes how antagonism between belligerent nations leads to exponential growth in arms

*__scaling law__: relationship in which one physical quantity varies as a simple power of another. Equivalent to the constancy of a dimensionless number.

simple harmonic motion (SHM) or the **simple harmonic oscillator** (SHO): the second order equation $\ddot{x} + \omega^2 x = 0$ or the dynamical system $\dot{x} = \omega y$, $\dot{y} = -\omega x$, whose solution is the oscillator of the form $x = A\sin(\omega(t - t_0))$

*__stable__ fixed point: one in which a small departure from a fixed point does not become large in the long-term. In contrast a point a small distance away from an **unstable** fixed point will move far from it. For example, this might occur via exponential growth away from an unstable fixed point, or decay towards a stable fixed point.

*__trajectory__: the path traced out by a particular solution to an ODE or dynamical system. In a 2D phase plane, the solution curve generated by the dynamical system from initial conditions corresponding to some particular point in the phase plane

van der Pol equation: a particular one-parameter 2D dynamical system, or equivalently a 2nd order ODE, which for large values of the parameter has a distinctive two-speed (fast-slow) limit cycle (sometimes known as a **relaxation oscillation**).

Verhulst (logistic) equation: a 1st order ODE which imposes a limit to exponential growth via a negative quadratic term to offset the linear growth term and thereby create a stable non-zero fixed point, the **carrying capacity**. Its solution appears in statistics in 'logistic regression', used to model log-odds-ratios.

wave equation: a PDE which describes wave phenomena. There are various of these; if otherwise unexplained the term usually refers to the **linear wave equation**, which is simply (linear) proportionality of the space and time second derivatives

Appendices are **not** examinable.

A Scaling laws are power laws

A *scaling law* is a relationship $y = f(x)$ that 'looks the same' – has the same functional form f – at all scales. More precisely, f has the property that

$$f(\lambda x) = f(x)g(\lambda)$$

for some function g (which depends on f).

This is clearly satisfied by $f(x) = Cx^a$, with $g(\lambda) = \lambda^a$. But is this simple power-law form of f necessary as well as sufficient?

Write everything in terms of logarithms, with $\log y = Y, \log x = X, \log \lambda = \Lambda$, and $\log f(x) = F(\log x) = F(X)$, $\log g(\lambda) = G(\log \lambda) = G(\Lambda)$. Then $Y = F(X)$, and the requirement is that

$$F(X + \Lambda) = F(X) + G(\Lambda).$$

But the only function F with this property is *linear*, $F(X) = aX + b$, with $G(\Lambda) = a\Lambda$. This is $\log f(x) = a \log x + b$ or, exponentiating, $f(x) = Cx^a$ where $b = \log C$.

So the scaling property is satisfied if and only if f is a simple power, and the relationship is $y = f(x) = Cx^a$ for some C and a.

B Separating timescales in the Van der Pol oscillator

Recall the van der Pol oscillator, with its strange dynamics which alternated between fast and slow movements, and whose trajectories rapidly converged to its limit cycle. We don't have an analytic solution, but is there anything analytic we can usefully say?

The answer is 'yes', and we do so by treating the two timescales, a short time t and a long time $\tau = \mu t$, as independent variables. Then

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mu \frac{\partial}{\partial \tau} \quad \text{and} \quad \frac{d^2}{dt^2} = \frac{\partial}{\partial t^2} + 2\mu \frac{\partial}{\partial t}\frac{\partial}{\partial \tau} + \mu^2 \frac{\partial^2}{\partial \tau^2},$$

which turns the Van der Pol oscillator (expressed as a second-order ODE)

$$\ddot{U} - \mu(1 - U^2)\dot{U} + U = 0$$

into

$$U_{tt} + U + \mu \left(2U_{t\tau} + (U^2 - 1)U_t\right) + \mu^2(U_{\tau\tau} + (U^2 - 1)U_\tau = 0,$$

where we write partial derivatives as subscripts.

Now equate terms at different orders of μ. First, at order μ^0, $U_{tt} + U = 0$, so $U = A(\tau)\sin t + B(\tau)\cos t$ – that is, we seek an approximation in which U behaves as a simple oscillator on the fast timescale, but with coefficients A, B which vary on the slow timescale.

At order μ, there is some algebra to do: substitute $U = A(\tau)\sin t + B(\tau)\cos t$ into $2U_{t\tau} + (U^2 - 1)U_t = 0$, then equate coefficients of $\cos t$ and $\sin t$, using

$$\sin^3 t = \frac{3}{4}\sin t - \frac{1}{4}\sin 3t, \qquad \cos^3 t = \frac{3}{4}\cos t + \frac{1}{4}\cos 3t$$

and ignoring the slower-oscillating terms. This gives

$$A_\tau = \frac{1}{2}A - (A^2 + B^2)\frac{A}{8}, \qquad B_\tau = \frac{1}{2}B - (A^2 + B^2)\frac{B}{8}.$$

If we write $E := A^2 + B^2$, we then find that

$$\frac{dE}{d\tau} = E\left(1 - \frac{E}{4}\right)$$

– which is the logistic equation, whose FP characterizes the Van der Pol oscillator's limit cycle. So E converges to 4.

The order μ^2 terms are derivable from the order μ terms, and contribute nothing new.

C Fourier coefficients for the wave equation

Here we derive the Fourier coefficients (4.26),

$$a_m = \frac{2}{L}\int_0^L y(x, 0)\sin\frac{m\pi x}{L}\,dx.$$

These differ from what you learned in calculus in two minor respects. One is that the variable has been rescaled, by L/π. The other is that the series consists of sines only, – no cosines. Recall that cosines are used to approximate even functions on $(-\pi, \pi)$ and sines to approximate odd functions. So what we're doing is to take our initial conditions on $[0, L]$, extend them to an odd function on $[-L, L]$, rescale by L/π, and compute the Fourier series. Rather than do all this, I'll just do the calculation directly.

Take

$$y(x, 0) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L},$$

multiply by $\sin \frac{m\pi x}{L}$, and integrate. Then

$$\int_0^L y(x,0) \sin \frac{m\pi x}{L}\, dx = \sum_{n=1}^{\infty} a_n \int_0^L \sin \frac{m\pi x}{L} \sin \frac{n\pi x}{L}\, dx.$$

The integral on the right-hand side

$$\int_0^L \sin \frac{m\pi x}{L} \sin \frac{n\pi x}{L}\, dx = \frac{1}{2} \int_0^L \left(\cos \frac{(n-m)\pi x}{L} - \cos \frac{(n+m)\pi x}{L} \right) dx.$$

The integrals of the cosines are sines, which evaluate to zero at integer multiples of π, leaving only the term when $n = m$ and $\cos \frac{(n-m)\pi x}{L} = 1$, for which the integral is L. This gives us the Fourier coefficients (4.26) as required. The calculation of the b_n is similar.

D A Musical Offering

For waves on a stretched string, the fundamental ($n = 1$), first harmonic ($n = 2$), second harmonic ($n = 3$) and so on naturally correspond to the notes you can play on a stringed instrument. Plucking or bowing a string will produce a mixture of these, and in general the fundamental note will dominate. Plucking, striking or bowing closer to the end of the string will tend to excite less of the fundamental and more of the harmonics. Placing your finger lightly over the centre of the string will prevent the fundamental from being excited, and you will hear mostly the first harmonic (which has a **node**, where the string remains stationary, at the string's centre).

The harmonics are related to musical notes as follows:

$$\begin{array}{lcccccc} n = & 1 & 2 & 3 & 4 & 5 & \dots \\ & C & C' & G' & C'' & & \end{array}$$

Doubling the frequency (which occurs when we go from $n = 1$ to $n = 2$) increases the note by an **octave**, while $n = 3$ to $n = 4$ is a (perfect) **fourth**, and $n = 2$ to $n = 3$ is a (perfect) **fifth**. A guitar is usually tuned mostly in fourths, a violin in fifths.

A natural way to construct a set of notes within an octave is to take these notes and shift them by octaves. For example: divide 3 by 2 to get 3/2 between 1 and 2. The **Pythagorean** scale builds on this by taking powers of 3/2 and dividing them by powers of 2 (*i.e.* fifths shifted by octaves): the notes are $3^n/2^{n+m}$. A nice set is

$$\begin{array}{rcccccccc} & 1 & \frac{1}{2}\left(\frac{3}{2}\right)^2 & \frac{1}{4}\left(\frac{3}{2}\right)^4 & \frac{4}{3} & \frac{3}{2} & \frac{1}{2}\left(\frac{3}{2}\right)^3 & \frac{1}{4}\left(\frac{3}{2}\right)^5 & 2 \\ & 1 & \frac{9}{8} & \frac{81}{64} & \frac{4}{3} & \frac{3}{2} & \frac{27}{16} & \frac{243}{128} & 2 \\ \text{approx.} & 1 & 1.13 & 1.27 & 1.33 & 1.5 & 1.69 & 1.90 & 2 \end{array}$$

This is an example of a **diatonic** scale; it corresponds (roughly) to the white notes on a piano. Notice that the intervals we earlier referred to as 'fourths' and 'fifths' get their names because they appear fourth and fifth in this scale; they're particularly simple ratios. The ratios between successive notes are all $9/8$ except for the third to fourth and seventh to octave, which are $256/243$, which is close to $\sqrt{9/8}$. So we might take advantage of this to construct a scale which uses only *one* ratio r of about $256/243$, together with r^2. We want the whole scale to consist of two ratios $r \simeq 256/243$ and five ratios of $r^2 \simeq 9/8$, so we set $r = 2^{1/12}$. Indeed, since Bach's time, notes in an octave have been split into twelve evenly-spaced **semitones** in this perfect geometric progression. The diatonic (major) scale consists of TTSTTTS (T=tone, S=semitone). So between $n = 1$ and $n = 2$ above we have the thirteen notes of the **chromatic** scale at ratios $2^{\frac{r}{12}}$, for $r = 0, 1, 2, 3, \ldots, 12$. Set $r = 7$ (note G) and we get $2^{\frac{7}{12}} = 2.997\ldots/2$, which is very close to $3/2$, but not perfect. One might think that such unlovely ratios would produce rather poor music; Bach's 'Evenly-Tuned Piano' (*Das Wohltemperierte Klavier*), a series of studies using this scale, proves otherwise.

References

[1] J. M. Epstein, 'Why model?', *Journal of Artificial Societies and Social Simulation* **11** (2008) 12

[2] Geoffrey I. Taylor, 'The formation of a blast wave by a very intense explosion. I. Theoretical Discussion' and 'II. The atomic explosion of 1945', *Proc. Roy. Soc.* **A201** (1950) 159-174 and 175-186. In fact Taylor had been working on the mathematics of blast waves throughout the war, publishing his results in a special, secret series of the *Proceedings of the Royal Society*.

[3] This example is based on a blog post by Sasha Borovik at Manchester, `https://micromath.wordpress.com/2008/04/04/kolmogorovs-53-law/`

[4] Bob Stewart and Chris Jarrett, *Kolmogorov, turbulence and British Columbia*, `https://www.math.ualberta.ca/pi/current/page22-23.pdf`

[5] A. R. P. Rau, 'Biological scaling and physics', *Journal of Biosciences* **27** (2002) 475-478.

[6] G. B. West, J. H. Brown & B. J. Enquist, 'A general model for the origin of allometric scaling laws in biology', *Science* **276** (1997) 122-126.

[7] A. A. Sonin, *The Physical Basis of Dimensional Analysis* (2001, MIT); `web.mit.edu/2.25/www/pdf/DA_unified.pdf`

[8] K. S. Nielsen, *Scaling: why is animal size so important?* (Cambridge University Press, 1984)

[9] J. B. S. Haldane, 'On being the right size', *Harper's Magazine* **152** (1926) 424-427, and available at many web locations.

[10] P. Gerlee, 'The Model Muddle: in search of tumor growth laws', *Cancer Research* **73** (2015) 2407-2410

[11] Geoffrey West, *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies* (Weidenfeld & Nicholson, 2017)

[12] A. J. Lotka, 'Fluctuations in the abundance of species considered mathematically (with comment by V. Volterra)', *Nature* 119 (1927) 12-13.

[13] R. Arditi and L. R. Ginzburg, 'Coupling in predator-prey dynamics: Ratio-Dependence', *Journal of Theoretical Biology* **139** (1989) 311-326.

[14] W. O.Kermack and A. G. McKendrick, 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society of London* **A115** (1927) 700-721.

[15] J. M. Heffernan, R. J. Smith and L. M. Wahl, 'Perspectives on the basic reproductive ratio', *Journal of the Royal Society Interface* **2** (2005) 281-293.

[16] R.M. May and R.M. Anderson, *The transmission dynamics of Human Immunodeficiency Virus (HIV)*, Philosophical Transactions of the Royal Society of London **B321** (1988) 565-607.

[17] L. F. Richardson, *The Mathematical Psychology of War*, 1919 (Oxford: Hunt); see also 'Mathematical psychology of war', Letter in Supplement to *Nature*, 1935, 830-831; and 'Mathematics of war and foreign politics' and 'Statistics of deadly quarrels', in Newman, *World of Mathematics*, 1956, vol.2, 1240-1263.

[18] F. W. Lanchester,. *Aircraft in Warfare: the Dawn of the Fourth Arm*, 1916 (London: Constable & Co). Based on articles in *Engineering* **98** (1913-14): 422-423 and 452-453. The chapter which deals with the mathematical models is reprinted in Newman J ed., *The World of Mathematics* vol.4, 2138-2157. New York: Simon and Schuster, 1956; also Mineola, NY: Dover, 2000.

[19] B. Van der Pol, J. Van der Mark, 'The heartbeat considered as a relaxation oscillation, and an electrical model of the heart', *Philosophical Magazine* series 7, vol. **6** (1926) 763-775. The FitzHugh-Nagumo model (1961-62), which simultaneously generalizes the Van der Pol oscillator and simplifies the Hodgkin-Huxley model, is nicely described at `http://www.scholarpedia.org/article/FitzHugh-Nagumo_equation`

[20] B. Van der Pol, 'On "relaxation oscillations"', *Philosophical Magazine* series 7, vol. **2** (1926) 978-992.

[21] F. Black and M. Scholes, 'The pricing of options and corporate liabilities', *The Journal of Political Economy* **81** (1973) 637-654.