School of Mathematical Sciences

Mathematical Biology (Honours)

Cell and tissue, shell and bone, leaf and flower, are so many portions of matter, and it is in obedience to the laws of physics that their particles have been moved, moulded and conformed. They are no exception to the rule that God always geometrizes. Their problems of form are in the first instance mathematical problems, their problems of growth are essentially physical problems, and the morphologist is, ipso facto, a student of physical science.

-Sir D'Arcy Wentworth Thompson, *'On Growth and Form'* [9]

---

1 Introduction: Basic Ideas in Mathematical Modelling

1.1 Mathematical modelling

The aim of mathematical modelling is to improve our understanding of some real-life problem. In areas such as physics, chemistry or engineering, mathematics has been so important in testing hypotheses and making predictions that (to the dismay of many students!) it is now impossible to study these subjects without first learning a significant amount of mathematics. This is not true to the same extent in biology, but their conspicuous success in the physical sciences has meant that there is an increasing readiness to apply mathematical techniques to biological problems. Correspondingly, the field of mathematical biology has grown rapidly in the last 20 years or so.

The general process of mathematical modelling can be summarised as follows:

1. Develop a hypothesis regarding the important mechanisms underlying the problem. This requires a detailed knowledge of the processes at work. In mathematical biology research, the input of experimental collaborators is often very important at this stage.

2. Formulate the hypothesis in mathematical terms. The often involves recasting the assumptions you have made about what is happening in the problem in the form of differential equations.

3. Analyse the resulting mathematical problem so as to understand the behaviour of the solution. (For very simple models, you may be able to write down a closed-form solution; in other cases, more complicated techniques will need to be applied to gain insights into its behaviour.)

4. Interpret the mathematical results in the context of the original problem. Is the solution consistent with what is observed in experiments? If so, are there new experiments you could devise based on your model, which would further test its validity? If not, what assumptions in your model might you need to revisit?

As mathematicians, your training up to now has probably focused mainly on stage 3; however, in many cases stage 2 can be equally if not more important. A problem which has been experimentally intractable to biologists can sometimes be solved quite straightforwardly once it has been formulated in mathematical terms. Similarly, a mathematical formula on a piece of paper is not likely to be much use to a biologist; you need to see the significance (and the limitations) of the result in the context of the problem (stage 4) if any real biological insight is to be gained. One of the aims of this course is to help you develop these additional skills.

1.2 Building mathematical models

One of the golden rules of mathematical modelling is: **keep it simple!** No understanding will be gained by converting an intractable biological problem into a mathematical model that is too complicated to analyse. When building a model, we must to focus on the processes we think are the most important, and neglect the others, at least to begin with. (Once the basic model is fully understood, additional effects can be added to it and their effect on the solution investigated.) Knowing what to put in, and what to leave out, is something of an art, and requires experience. However, there are a couple of basic techniques that can be very helpful.

1.2.1 Dimensional analysis

It is obvious that in order to be physically consistent, we can only equate together quantities which have the same dimensions - *e.g.* in Newton's Second Law, $F = ma$, the quantities on both sides of the equation must have the dimensions of force. Measurements of quantities are taken with respect to a reference value; the particular reference value used will depend upon the system of units adopted. Physical relationships must be true irrespective of the system of units used. In defining a system of units some quantities are considered fundamental - *e.g.* mass, length and time, which we shall denote [M], [L] and [T]. Other units are derived from these - *e.g.* speed is the rate of change of distance, and so has dimensions of [L] $[\mathrm{T}]^{-1}$. Other fundamental units include electric charge, [Q], and temperature, [Θ].

Example: Coulomb's law
The force, F acting on two particles with charges q_1 and q_2, separated by a distance r, is given by

$$F = k_e \frac{q_1 q_2}{r^2}$$

where k_e is a constant. What are the dimensions of k_e?

Both sides of the equation must have the dimensions of force - *i.e.* [M] [L] $[T]^{-2}$. Hence

$$\frac{[M][L]}{[T]^2} = [k_e]\frac{[Q]^2}{[L]^2}, \qquad \Rightarrow \qquad [k_e] = \frac{[M][L]^3}{[Q]^2[T]^2}.$$

A quantity which has no units is said to be **dimensionless**. Recall that the length, a, of an arc of a circle of radius r, subtended by an angle, θ, is given by $a = r\theta$. Since both a and r have dimensions of [L], $[\theta] = [1]$ (*i.e.* θ is dimensionless). Note that we can create dimensionless quantities by making appropriate combinations of dimensional quantities. For example, in the case of a simple pendulum of length, l, which oscillates with a frequency ω in a gravitational field of strength, g, the quantity $\dfrac{g}{l\omega^2}$ is dimensionless (exercise).

We can exploit dimensionless quantities to help us deduce model equations. Suppose we are considering a problem with dimensional variables x_1, $x_2, \ldots, x_n$ and dimensional parameters (constants) $\alpha_1, \ldots, \alpha_m$, and want to determine the relationship between these quantities. Since we can only equate quantities of the same dimension, this limits the combinations of the x_i and α_j which are possible. Obviously, dimensionless quantities can be equated in this way, so if the x_i and α_j can be combined to create k dimensionless groups, β_1, $\beta_2, \ldots, \beta_k$, the relationship between these can be written in the format

$$f(\beta_1, \beta_2, \ldots, \beta_k) = 0.$$

It is frequently the case that only a small number of dimensionless groups can be created from the dimensional variables and parameters which we believe to influence the problem. This will simplify the modelling very considerably. For example, if there is only one dimensionless group, we have must have a relationship of the form $f(\beta) = 0$. This means β is a zero of the function f - *i.e.* the physical relationship between our variables can be expressed as $\beta = constant$.

Example: How powerful is an atomic bomb?

In 1945, the USA test-detonated the world's first atomic bomb, a device code-named Trinity, in the New Mexico desert. Information concerning the test was highly classified by the US government at the time (in fact, the full technical report on the explosion was not published until 1976), though a series of photographs of the explosion (which included the time since detonation) were declassified in 1947. The UK government was intensely interested in the results of the US test, as they wanted to develop their own atomic weapons; in particular, they wanted to know how much energy such a device might release. They asked Sir Geoffrey Taylor (G. I. Taylor) an applied mathematician at the University of Cambridge, to work on the problem. For explosions which took place in the open, he assumed the blast wave created would be spherical in shape. He reasoned that, in the early stages of an explosion (before energy could be radiated as heat), the radius of the blast, R, could only depend on the energy of the explosion, E, the time since detonation, t, and the density of the air, ρ. By using dimensional analysis, he realised that only one dimensionless number could be created from these quantities, and so

$$\frac{Et^2}{\rho R^5} = constant.$$

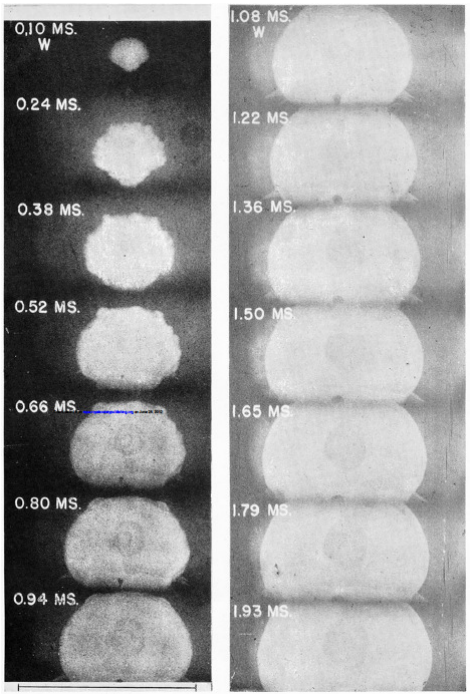


Figure 1: Photographs of the Trinity atomic bomb test used by G. I. Taylor (from [8]).

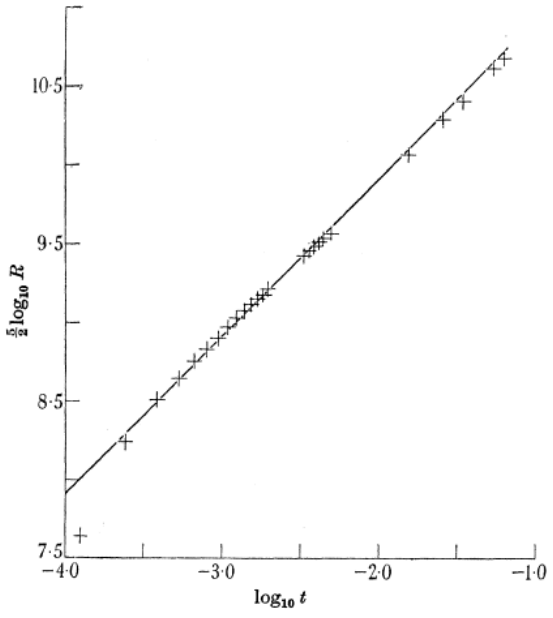


Figure 2: Graph of data from the photographs, showing clearly that $R^5 \propto t^2$ (from [8]).

By means of other arguments, Taylor determined that the value of the constant would be approximately one.

Assuming a typical value for the density of air, ρ, and making measurements from the published photographs of the test to estimate R at each value of t, Taylor was able to obtain a value for E, the energy of the blast, of $16.8 - 23.7$ kT (*i.e.* equivalent to 16,800-23,700 tons of TNT). The official test determined it was 20 kT. Taylor published his results in 1950 [8], two years before the first UK atomic weapons test.

1.2.2 Nondimensionalisation and scaling

Although dimensional analysis can sometimes help us to deduce the equations governing a process, another important reason for dealing with dimensionless quantities is that it allows us to compare the relative importance of various effects very easily. This is not really true with dimensional equations.

Example: Suppose for a moment, we are interested in the spread of a pollutant being released from a factory chimney into the air over a small town. The chemical will diffuse, and will also be carried on the wind. We are told the typical wind speed in the area is 5 km h^{-1}, and the diffusion coefficient of the chemical in air at 25 °C has been measured as 0.3 cm^2 s^{-1}. Is the pollutant transported mainly by diffusion, or on the wind?

In order to answer this question, we first need to think about how the pollutant concentration, c evolves. For simplicity, let us just consider the spread of the pollutant in one dimension, and assume that the concentration profile is steady (*i.e.* does not change with time). We let x be the distance downwind of the chimney. Then the concentration obeys

$$U\frac{\partial c}{\partial x} = D\frac{\partial^2 c}{\partial x^2}, \qquad c = c_0 \quad \text{at } x = 0, \qquad c \to 0 \quad \text{as } x \to \infty,$$

where U is the wind speed and D is the diffusion coefficient. (You can just accept this equation for now, we will derive it later in the course.) We are interested in how the pollutant spreads over a town, so a typical lengthscale (distance) we be around a kilometre; let L be the distance from the chimney to the town centre. Then we notice there are some obvious natural scales in the problem. It is convenient to specify the concentration in terms of the fraction of the value at the site of release, c_0, and similarly to measure length in terms of the fraction of the distance to the town centre. Hence we set

$$\tilde{x} = \frac{x}{L}, \qquad \tilde{c} = \frac{c}{c_0},$$

where the tildes indicate dimensionless quantities. We then note that UL/D is a dimensionless quantity, which is called the Péclet number, $\mathcal{P}$. Hence the dimensionless concentration obeys

$$\mathcal{P}\frac{\partial \tilde{c}}{\partial \tilde{x}} = \frac{\partial^2 \tilde{c}}{\partial \tilde{x}^2}, \qquad \tilde{c} = 1 \quad \text{at } x = 0, \qquad \tilde{c} \to 0 \quad \text{as } \tilde{x} \to \infty.$$

This equation has some obvious superficial advantages over the dimensional version: there are fewer constants, so there is less chance of us making a mistake. But more

importantly, we now see that if $\mathcal{P} \ll 1$, diffusion is the dominant transport mechanism, whilst for $\mathcal{P} \gg 1$, transport on the wind is the more important factor. If $\mathcal{P} = O(1)$, then both mechanisms play an equal role. For our situation, we find that $U \approx 1.4$ m s^{-1}, $D = 3 \times 10^{-5}$ m^2 s^{-1} and $L = 10^3$ m. Hence

$$\mathcal{P} \approx \frac{1.4 \times 1000}{3 \times 10^{-5}} \approx 5 \times 10^7 \gg 1.$$

Therefore, transport by the wind is vastly more important than diffusion, so we can make our lives easier by neglecting the diffusion term in the equation.

To summarise, it is very helpful to work with dimensionless models for the following reasons:

- The equations involve fewer symbols, so we are less likely to make mistakes in calculations, and it is often easier to recognise the type of equations involved. Since the coefficients are real numbers, rather than dimensional quantities, their magnitudes can be directly compared, which is useful for determining the most important effects in the problem, and making simplifications where appropriate.
- Reducing the number of parameters means results can be investigated more quickly and presented in more compact form. Above, we reduced the number of parameters from 3 (c_0, U and D) to one. Hence the behaviour of the solution depends on only one parameter, $\mathcal{P}$ - we do not have to give separate plots for different values of c_0, U and D). This can be particularly important when we have to solve a problem numerically - especially if the simulation takes a long time to run.
- Solutions obtained for one system can be applied to another which obeys the same equation, but with different parameter values - there is no need to recalculate the solution.
- Often the dimensionless equations can help in the design of experiments - *e.g.* we could investigate the pollutant dispersal problem experimentally by building a scale model in the lab, such that $\mathcal{P}$ takes the same value as for the real problem.

1.3 Types of models

In this course, we will concentrate on models based on ordinary and partial differential equations. As these are so numerous in the literature, it might be tempting to think that they are a good way to attack any real-life problem: this is certainly not the case (we will discuss the cautionary example from [1] in class). It is useful to remind ourselves at this point that there may be other more appropriate modelling approaches for certain problems, such as:

- **Difference equations** (also known as recurrence relations, iterated maps, or just maps): The dependent variable can be discrete or continuous; time is always discrete; suitable for seasonal events; can have deterministic and stochastic difference equations.

- **Stochastic processes:** A family of random variables $\{X(t)\}$, indexed by a parameter t, is called a stochastic process; Markov-chain models are one class of (memory-less) stochastic model; particularly useful for small populations.
- **Cellular automata:** Fully discrete models, all independent variables and all dependent variables are discrete; analysis is mainly restricted to computer analysis and numerical simulation; can be either deterministic or stochastic, using a random number generator.

2 Biochemical reactions

In the next two sections, we will consider mathematical models consisting of ordinary differential equations, with which you will already be familiar. As much of the challenge in mathematical modelling is simply knowing what equations to write down, we give considerable attention to this aspect. We consider how the important variables are identified, and how we can use the processes of scaling and inspectional analysis introduced earlier to identify appropriate simplifying assumptions. Since only the simplest models can be solved analytically, we also introduce techniques which allow us to gain insight into the qualitative behaviour of the models, without actually having to solve them.

2.1 Chemical reactions

Chemical reactions are important in many biological processes - *e.g.* digestion, photosynthesis, respiration. We first consider an irreversible reaction process in which reactants A and B produce C - *i.e.*

$$A + B \xrightarrow{k} C, \quad k = \text{the reaction constant.}$$

Let $a = [A]$, $b = [B]$, $c = [C]$ denote the concentrations of A, B and C, respectively. (The use of square brackets to indicate a concentration is standard in biology and chemistry, and hence appears in many papers and books. However, I think it can result in equations that look very cluttered, and so will try to avoid it in this course.) The SI units of concentration are moles m^{-3} (abbreviated to mol m^{-3}), where a *mole* is simply a number of molecules - 6.023×10^{23}. Since a cubic metre is often an inconveniently large volume, moles per litre (also called *molar*, M) is a common alternative unit.

We argue

$$\left\{ \begin{array}{c} \text{change of} \\ \text{the product} \\ \text{over time} \end{array} \right\} = \left\{ \begin{array}{c} \text{number of} \\ \text{collisions of} \\ \text{molecules } A \\ \text{and } B \end{array} \right\} \cdot \left\{ \begin{array}{c} \text{probability that a} \\ \text{collision has enough} \\ \text{kinetic energy to} \\ \text{initiate a reaction} \end{array} \right\},$$

which yields the ODE

$$\frac{dc}{dt} = k\, a\, b.$$

This is the **Law of Mass Action**. While called a law, it is really just a mathematical model, which is useful in many, but not all situations. Reactions which obey mass action kinetics are called *elementary reactions*. However, there are many biological reactions which proceed through complex mechanisms consisting of several elementary steps which are not known in sufficient detail for the law of mass action to be applied.

It is also worth remarking that the assumption that the reaction rate, k, is constant is strictly only true if the reaction occurs at a constant temperature. Many chemical

reactions give off, or take in, significant amounts of heat. (In modelling such a case, we would need to know the reaction rate as a function of temperature, and either measure, or have a model for, the temperature throughout the reaction. This would make the modelling much more complicated.) However, for reactions occurring under physiological conditions, the assumption of constant temperature is usually quite appropriate - *e.g.* body temperature in mammals is maintained very close to constant under a wide range of external conditions.

Extension: What equation would we write down for a reaction involving m molecules of A and n molecules of B reacting to produce a molecule of C?

Next, consider a reversible reaction

$$A + B \underset{k_-}{\overset{k_+}{\rightleftharpoons}} C.$$

Here, as for many biological processes it is necessary to follow the time evolution of more than one factor, leading to systems of ODEs. Balancing the production and consumption terms for each participating chemical species gives

$$\begin{aligned} \frac{dc}{dt} &= k_+ a\, b - k_- c, \\ \frac{da}{dt} &= -k_+ a\, b + k_- c, \\ \frac{db}{dt} &= -k_+ a\, b + k_- c. \end{aligned}$$

Note that this system is reducible to just one equation, since

$$\frac{d}{dt}(c + a) = \frac{d}{dt}(c + b) = 0,$$

and given initial concentrations a_i, b_i, c_i we have $a = c_i + a_i - c$ and $b = c_i + b_i - c$.

2.2 Enzymes

A *catalyst* is a substance that speeds up a chemical reaction without itself being consumed in the reaction. An *enzyme* is a biological catalyst: most are proteins. The increase in the rate of reaction they induce is staggering - often of the order of several million-fold. The substances on which they operate are called *substrates*.

Enzymes regulate a vast array of processes in the body, particularly those related to digestion and metabolism. For example, amylase in saliva starts to break down starch in our food into sugars, which are in turn further broken down into glucose, the body's primary source of energy. The release of energy from glucose also occurs through a chain of enzyme-mediated reactions. However, they are also important industrially. An everyday example would be the use of enzymes in laundry powder, which help to

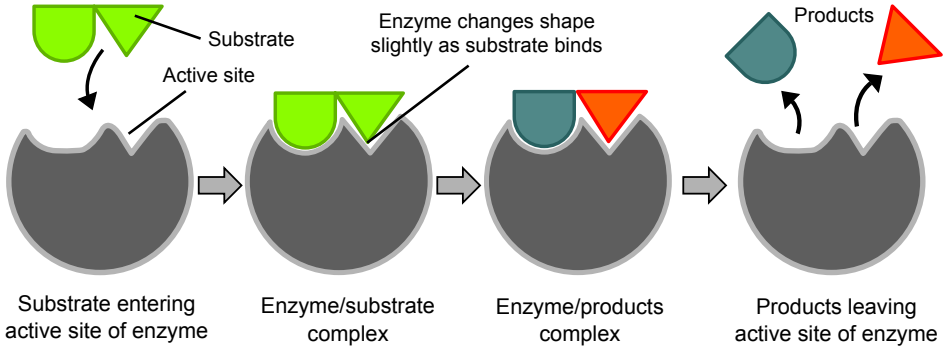


Figure 3: Schematic of the induced fit theory of enzyme action. (Diagram credit: TimVickers, Wikimedia.)

break down the fats in stains.

Each enzyme catalyses a specific reaction. This specificity is usually explained in terms of the 'lock and key' theory (or its slight modification- the induced fit theory, shown diagrammatically in Fig. 3). The idea is that the enzyme and its substrates both have specific, complementary shapes. The substrate must bind to the enzyme in order for the reaction to occur, and this is only possible when the two fit together like a 'lock and key'. (The induced fit theory allows for some flexibility in the enzyme, which helps the binding become tighter.) Thus, the enzyme is unable to bond with other substrates, as they do not have the correct shape.

Since enzymes are proteins, they tend to work best within a narrow range of temperature and pH. At low temperatures, molecules have low kinetic energy, so the number of collisions is reduced, and hence the rate of reaction. At temperatures above around $40°C$ (or extremes of pH), proteins become *denatured*, meaning their structure is changed (*e.g.* think of what happens to egg white as it is cooked). This makes the enzyme less effective at catalysing the reaction.

2.3 Michaelis-Menten kinetics

Consider the enzymatic reaction

$$E + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} E + P,$$

involving the reaction of a substrate S with an enzyme E to gain a product P, and the enzyme. The enzyme is a catalyst.

To allow use of the Law of Mass Action, we assume E and S form an intermediate complex C which then decays into P and E:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} E + P, \quad \text{the } \textit{Michaelis-Menten scheme.}$$

Letting $s = [S]$, $e = [E]$, $c = [C]$, $p = [P]$, we can describe the process using ODEs:

$$\begin{aligned}
\frac{ds}{dt} &= -k_1 s\, e + k_{-1} c, \\
\frac{de}{dt} &= -k_1 s\, e + k_{-1} c + k_2 c, \\
\frac{dc}{dt} &= k_1 s\, e - k_{-1} c - k_2 c, \\
\frac{dp}{dt} &= k_2 c.
\end{aligned}$$

Note that the units of the constant k_1 differ from those of k_-1 and k_2. We take the initial conditions for the system to be

$$s(0) = s_i > 0, \qquad e(0) = e_i > 0, \qquad c(0) = p(0) = 0. \tag{2.1}$$

We note that by adding the second and third equations above, integrating and applying the initial conditions, we find that

$$e(t) + c(t) = e_i. \tag{2.2}$$

Hence we can eliminate e from the equations for s and c. Since p depends only on c, we need only consider the reduced system

$$\frac{ds}{dt} = -k_1 s\,(e_i - c) + k_{-1} c = -k_1 s e_i + (k_1 s + k_{-1}) c, \tag{2.3a}$$

$$\frac{dc}{dt} = k_1 s\,(e_i - c) - k_{-1} c - k_2 c = k_1 s e_i - (k_1 s + k_{-1} + k_2) c, \tag{2.3b}$$

with initial conditions $s(0) = s_i$, $c(0) = 0$.

These have no known exact solution but approximate solutions have been obtained. The best known and most commonly used is the approximation introduced by Michaelis and Menten in 1913, which is motivated by the observation that the availability of the enzyme is often the limiting factor in the reaction. In order to develop our approximate solution, we must use the techniques of scaling and inspectional analysis we met earlier in the course.

We translate the assumption that the reaction is limited by the availability of the enzyme into the mathematical assumption that $e_i \ll s_i$. We let T be the timescale over which the substrate is observed to be converted into product (we will specify this in terms of the system parameters shortly). Since we start with a finite amount of substrate and no product, we must have $0 \le s \le s_i$ and $0 \le p \le s_i$. From (2.2) we have $0 \le c \le e_i$, so $c \ll s_i$. We hence nondimensionalise our variables as follows (where tildes indicate dimensionless variables):

$$t = T\tilde{t}, \qquad s = s_i \tilde{s}, \qquad c = \epsilon s_i \tilde{c},$$

where we have introduced the dimensionless parameter $\epsilon = e_i/s_i \ll 1$. Equation (2.3a) then becomes (dropping tildes):

$$\frac{ds}{dt} = \epsilon k_1 s_i T(cs - s) + \epsilon k_{-1} T c, \tag{2.4}$$

Since, by definition, T is the timescale for s to undergo an $O(1)$ change, we expect the parameters on the RHS of (2.4) to be $O(1)$. Hence we choose $T = 1/\epsilon k_1 s_i$, which implies that the timescale on which the substrate is converted to product is much longer than the timescale of complex formation. We also assume $A = \epsilon k_{-1} T = k_{-1}/k_1 s_i = O(1)$, so the rates of complex formation and dissociation are of the same order of magnitude. Hence (2.3) simplifies to

$$\frac{ds}{dt} = cs - s + Ac, \tag{2.5a}$$

$$\epsilon \frac{dc}{dt} = s - cs - (A + B)c, \tag{2.5b}$$

where $B = k_2/k_1 s_i$ is the ratio of the rate of product formation to complex formation, and the initial conditions become $s(0) = 1$, $c(0) = 0$.

We now expand s and c as power series in the small parameter, ϵ, so

$$s = s_0 + \epsilon s_1 + O(\epsilon^2), \qquad c = c_0 + \epsilon c_1 + O(\epsilon^2).$$

Substituting the above into (2.5), at leading order we find

$$c_0 = \frac{s_0}{K_m + s_0} \qquad \text{where} \quad K_m = A + B = \frac{k_{-1} + k_2}{k_1 s_i} \tag{2.6}$$

$$\frac{ds_0}{dt} = \frac{-B s_0}{K_m + s_0}, . \tag{2.7}$$

This is the most common Michaelis-Menten function used, among other things, for rate of depletion of nutrients in biological systems. B is the maximum uptake rate; K_m is the concentration at which the uptake rate is one half of the maximum.

Remarks:

- Note how the different timescales in the reaction have emerged naturally from the scaling analysis. In fact, it is possible to show that the approximate solution remains valid under the weaker assumption that $\epsilon/(K_m + 1) \ll 1$ (*i.e.* either there is little enzyme compared to substrate, or the rates of complex dissociation and product formation are slow compared to the rate of complex formation). Full details are given in [7].

- Our leading order solution for c, c_0 as given in equation (2.6) does not obey the initial condition $c(0) = 0$. This is because, in obtaining it, we neglected the time derivative in equation (2.5b). Situations like this, where the small parameter multiplies the highest derivative in an equation are called **singular perturbation problems**, and generally give rise to a leading-order solution that does not obey one or more of the initial or boundary conditions. If, in equation (2.5b) we introduce the new, shorter timescale $t = \epsilon\tau$, then the time derivative term becomes $O(1)$. Physically, this implies that there is a short timescale in the problem on which c changes from zero to the value given in (2.6). See Chapter 6 of [6] for full details.

3 Excitable systems

3.1 Background

The cell membrane is a *phospholipid bilayer* separating the cell interior (the cytoplasm) from the extracellular environment. The membrane contains numerous proteins, and is approximately 7.5nm thick. The most important property of the cell membrane is its selective permeability: it allows the passage of some molecules but restricts the passage of others, thereby regulating the passage of materials into and out of the cell. Many substances penetrate the cell membrane at rates reflected by their diffusive behaviour in a pure phospholipid bilayer. However, certain molecules and ions such as glucose, amino acids and Na^+ pass through cell membranes much more rapidly, indicating that the membrane proteins selectively facilitate transport.

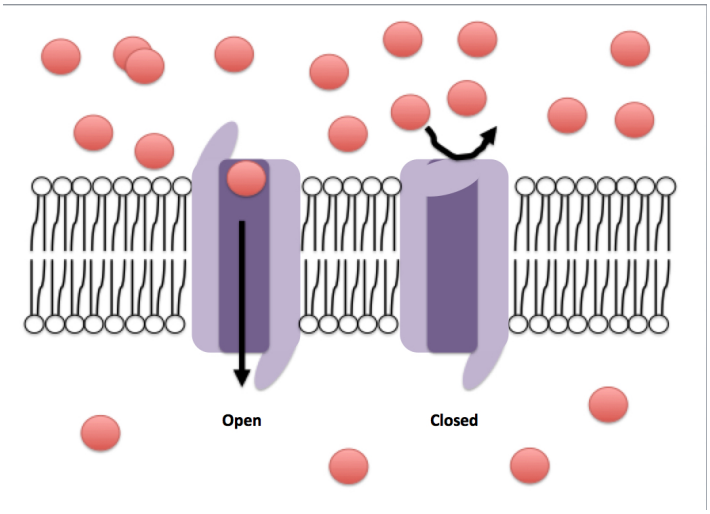


Figure 4: Illustration of a channel in the cell membrane (Credit: Efazzari [CC BY-SA 4.0])

The membrane contains water-filled pores with diameters of about 0.8nm, and protein lined pores, called channels or gates, which allow the passage of specific molecules. Both the intracellular and extracellular environments comprise (among other things) a dilute aqueous solution of dissolved salts, mainly NaCl and KCl, which dissociate into Na^+, K^+ and Cl^- ions. The cell membrane acts as a barrier to the free flow of these ions and to the flow of water.

The mechanisms that facilitate transport across the cellular membrane can be divided into active and passive processes. Active processes requires energy expenditure, while passive processes result solely from the random motion of molecules, for example, diffusion.

Action potentials, or 'nerve impulses', are brief changes in the membrane potential of a cell produced by the flow of ionic current across the cell membrane. They enable communication by many cell types, including neurons, cardiac and muscle cells.

First we note that:

- Numerous fundamental particles, ions and molecules have an electric charge, *e.g.* the electron, e^-, and the sodium ion, Na^+. The SI unit of electrical charge is the Coulomb (C).
- It is an empirical fact that total charge is conserved.
- Electric charges exert electrical forces on one another such that like charges repel and unlike charges attract. The electric potential, denoted V, is the potential energy of a unit of charge due to such forces and is measured in volts (V) or Joules per Coulomb (JC^{-1}).
- A concentration of positive particles has a large *positive* potential, while a concentration of negative particles has a large, but *negative* potential.
- Electric current is defined to be the rate of flow of electric charge, measured in Amperes, A (also known as Amps; equivalently, Cs^{-1}).

3.1.1 Revision of electrical circuits

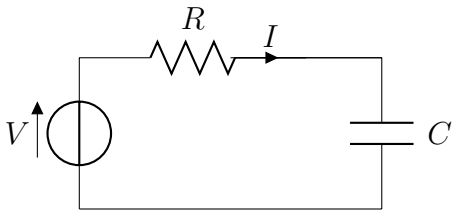


Figure 5: A simple electrical circuit with a voltage source, resistor and capacitor

Before we consider modelling electrical behaviour in cells, recall the simple electrical circuit shown in Figure 5 where:

- $q(t)$ is the charge;
- $I(t) = \dfrac{dq}{dt}$ is the current (rate of flow of charge);
- $V(t)$ is the potential difference / voltage (which causes the movement of charge - somewhat analogous to pressure in fluid mechanics);
- R is the resistance (measured in Ohms, Ω);
- $g = \dfrac{1}{R}$ is the conductance;
- C is the capacitance (ability to store charge) - measured in Farads, F.

We have:

- **Ohm's law*** - the potential difference (voltage drop) across a resistor is proportional to the current through the resistor.

$$V_R(t) = I(t)R = \frac{I(t)}{g}.$$

- **Faraday's law** - the potential difference across a capacitor is proportional to the charge stored.

$$V_c(t) = \frac{q(t)}{C}.$$

- **Kirchoff's law** - the voltage supplied is equal to the total voltage drop

$$V(t) = V_R(t) + V_C(t).$$

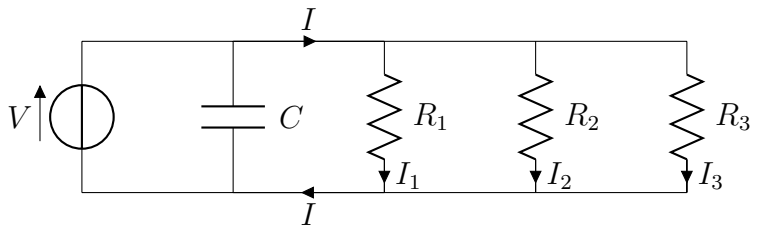


Figure 6: A more complex electrical circuit with a voltage source, three resistors and a capacitor

- For elements in parallel (see Figure 6), total current is the sum of the currents in each branch, whilst the potential difference across each is the same - *e.g.*

$$I(t) = I_1(t) + I_2(t) + I_3(t) = g_1V + g_2V + g_3V.$$

Now since $V(t) = q/C$

$$\frac{dV}{dt} = \frac{1}{C}\frac{dq}{dt} = \frac{I}{C},$$

and so

$$\frac{dV}{dt} = \frac{V}{C}(g_1 + g_2 + g_3).$$

In fact, when we are thinking about ionic currents across the cell membrane, things are a little more complicated than in electrical circuits. Instead of the current being directly proportional to the potential ($I = gV$), we usually have a relationship of the form

$$I = g(V)(V - V^*),$$

where $g(V)$ is the ion-specific membrane conductance, and V^* is the *Nernst potential*. We will explain in more detail how they arise in the next two sections.

*Similar to many others, this 'law', is in fact just a model.

3.2 The membrane potential

Active and passive transport of ions into and out of the cell can result in differences in ion concentrations between the interior and exterior of a cell. This induces a potential difference across the membrane, and in turn affects ion transport.

Suppose we have two reservoirs containing different concentrations of a positively charged ion X^+. We suppose that both reservoirs are electrically neutral to begin with, so that there is an equal concentration of a negatively charged ion Y^-. Now suppose that the reservoirs are separated by a semi-permeable membrane which is permeable to X^+ but not to Y^-. Then the difference in concentration of X^+ on each side will lead to the flow of X^+ across the membrane. However, because Y^- cannot diffuse through the membrane this will lead to a build up of charge on one side. This charge imbalance sets up an electric field, which produces a force on the ions opposing further diffusion of X^+. (The actual amount of X^+ which diffuses through the membrane is small, and the excess charge all accumulates near the interface, so that to a good approximation the solutions on either side remain electrically neutral.) The potential difference at which equilibrium is established and diffusion and electric-field-generated fluxes balance is known as the *Nernst potential*.

Instead of using the Ohm's law relation $I = gV$ for the ionic current across the cell membrane, it is common to assume a relationship of the form

$$I = g(V - V^*),$$

V^* is the *Nernst potential*. This gives an Ohm's law-like linear dependence of current on V, but takes into account the fact that there is no net flux of ions across the membrane when $V = V^*$.

3.3 Gating

It is found experimentally that the conductance g is not constant but depends on both V and time t. One proposed explanation for this is that the channels are not always open, but may be open or closed, and that the transition rates between open and closed states depends on the potential difference, V . The membrane conductance may then be written as ng, where g is the constant conductance that would result if all the channels were open, and n is the proportion of open channels.

3.3.1 Simple gates

Consider a generic ion, with n being the proportion of open ion channels. Denoting the open channels by O and the closed channels by C, the reaction scheme is simply

$$C \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftharpoons}} O.$$

where $\alpha(V)$ and $\beta(V)$ represent voltage dependent rates of switching between the closed and open states. Using the law of mass action we obtain

$$\frac{dn}{dt} = \alpha(V)(1-n) - \beta(V)n,$$

or equivalently,

$$\tau_n(V)\frac{dn}{dt} = n_\infty(V) - n,$$

where $n_\infty(V) = \alpha/(\alpha+\beta)$is the equilibrium value of n and $\tau_n(V) = 1/(\alpha+\beta)$ is the timescale for approach to this equilibrium (both of which can be determined experimentally).

3.3.2 Multiple gates

We can now extend the simple model above to channels composed of multiple subunits (gates), each one of which can be in either the open or closed state.

We start by assuming that the channel consists of two gates, which may both exist in open or closed states. The ion channel is open only if both gates are open; the ion channel is closed if any one gate within the ion channel is closed. Let S_i $(i = 0, 1, 2)$ denote the proportion of channels with i gates open. Then, the reaction scheme is

$$S_0 \underset{\beta(V)}{\overset{2\alpha(V)}{\rightleftharpoons}} S_1 \underset{2\beta(V)}{\overset{\alpha(V)}{\rightleftharpoons}} S_2,$$

where the factors of two arise because there are two possible states with one gate open and one gate closed (since each gate is identical we lump these two states into one variable S_1). Note we also have the overall constraint that

$$S_0 + S_1 + S_2 = 1. \tag{3.1a}$$

Using mass action kinetics we have

$$\frac{dS_0}{dt} = \beta(V)(1 - S_0 - S_2) - 2\alpha(V)S_0, \tag{3.1b}$$

$$\frac{dS_2}{dt} = \alpha(V)(1 - S_0 - S_2) - 2\beta(V)S_2, \tag{3.1c}$$

where, in general, V could itself be a function of time (and space). Note, we could have written down a third equation for S_1 but this is not needed, since it can be found using (3.1a).

Now, the proportion of open gates n is given by

$$n = \frac{1}{2}S_1 + S_2 = \frac{1}{2}(1 - S_0 + S_2),$$

so, by taking the appropriate linear combination of the equation for S_0 and S_1 we find that

$$\frac{dn}{dt} = \alpha(V)(1-n) - \beta(V)n. \tag{3.2}$$

The system (3.1) is satisfied by

$$S_0 = (1-n)^2, \quad S_1 = 2n(1-n), \quad S_2 = n^2, \tag{3.3}$$

(which can be verified by simple substitution). In fact, by writing $S_0 = (1-n)^2 + z_0$, $S_2 = n^2 + z_2$ (so that $S_1 = 2n(1-n) - z_0 - z_2$) and substituting in equation (3.1b) and (3.1c) we find

$$\frac{dz_0}{dt} = -2\alpha z_0 - \beta(z_0 + z_2), \qquad \frac{dz_2}{dt} = -\alpha(z_0 + z_2) - 2\beta z_2.$$

This is a linear system with eigenvalues $-(\alpha+\beta)$ and $-2(\alpha + beta)$, so z_0 and z_2 will decay exponentially to zero. Hence, the solution of the system will always approach exponentially that given by equations (3.2) and (3.3).

The analysis of a two-gated channel generalised easily to channels containing more gates. In the case of k identical gates the fraction of open channels is n^k, where n again satisfies (3.2). It has been found that a model with 4 gates agrees with empirical observations of K^+ channels, a fact we will exploit later.

3.3.3 Non-identical gates

Often channels are controlled by more than one protein, with each protein controlling a set of identical gates, but with the gates controlled by each protein different and independent. Consider, for example, the case of a channel with two types of gate, m and h, each of which may be open or closed. For the purposes of illustration, we will assume that the channel has two m subunits and one h subunit. Then, we let S_{ij} denote the proportion of channels with i of the of the m-gates open and j of the h-gates open (where $i = 0, 1, 2$ and $j = 0, 1$). Note, there are thus six possible configurations of the channel.

If m and h denote the proportions of each of the two types of gate that are open, then the law of mass action equations can be shown to be satisfied by

$$S_{00} = (1-m)^2(1-h), \quad S_{10} = 2m(1-m)(1-h), \quad S_{20} = m^2(1-h),$$

$$S_{01} = (1-m)^2 h, \quad S_{11} = 2m(1-m)h, \quad S_{21} = m^2 h,$$

so that the proportion of open channels is m^2h where m and h satisfy

$$\frac{dm}{dt} = \alpha(V)(1-m) - \beta(V)m, \qquad \frac{dh}{dt} = \gamma(V)(1-h) - \delta(V)h.$$

Note that here γ and δ are the rates of switching between the closed and open states for the h-gates (analogous to α and β for the m-gates).

We now have the required background to consider a model of nerve signal propagation.

3.4 The Hodgkin-Huxley model

The Hodgkin-Huxley model was developed by British scientists Alan Lloyd Hodgkin and Andrew Huxley in 1952 to explain the mechanisms underlying the initiation and propagation of action potentials (electrical signals) in the squid giant axon (the part of the squid's nervous system that controls its water jet propulsion system). It is considered one of the major achievements of mathematical biology, and Hodgkin and Huxley were awarded the Nobel prize in 1963 for their work.

More generally, the model can be used to describe the behaviour of a variety of excitable cells, including neurons, the cells which compose the electrochemical communication system that constituents our nervous system.

3.4.1 Structure of a neuron

Inputs detected by the dendrites are conducted to the soma. A nerve signal is then initiated at the axon hillock, which travels down the axon to terminal branches where the signal is passed to the next cells in the network. These signals are transmitted by action potentials, and cells which can transmit action potentials are called excitable cells (*e.g.* cardiac cells, smooth and skeletal muscle, secretory cells and neurons).

Neuronal signals travel along the cell membrane of the axon in the form of a local voltage difference across the cell membrane. In the inactivated state the cytoplasm (fluid inside the cell) in the axon is slightly negative in potential compared to the outside (-50mV difference) *i.e.* the cell is polarised, due to a difference in ionic composition, maintained by actively pumping sodium ions out of the cell and potassium ions into the cell, as well as differences in other ionic concentrations across the membrane.

It is tempting to think of neurons as a long electrical cable, but this is not quite right. The current in a neuron is made up of an ionic (rather than electron) flow, and the flow is across the membrane - i.e. transverse not longitudinal.

When the cell becomes partially depolarised a series of events takes place:

1. **The upstroke phase.** Sodium channels open in response to the depolarisation, allowing positively charged sodium ions to enter the cell, increasing the depolarisation further, till the cell becomes positively charged.
2. **The excited phase.** Over a slower timescale, the potassium channels open, allowing potassium ions to leave. Sodium ions continue to enter the cells and the potential difference slowly falls.
3. **The downstroke phase** The potassium ions make the cell negatively charged, which closes the sodium channels making the cell more negatively charged. The cell becomes hyperpolarised - it has overshot.
4. **The refractory and recovery phases** The sodium channels are now mostly inactive so cannot respond to any further stimulus. They gradually become active again, and the cell returns to its original state.

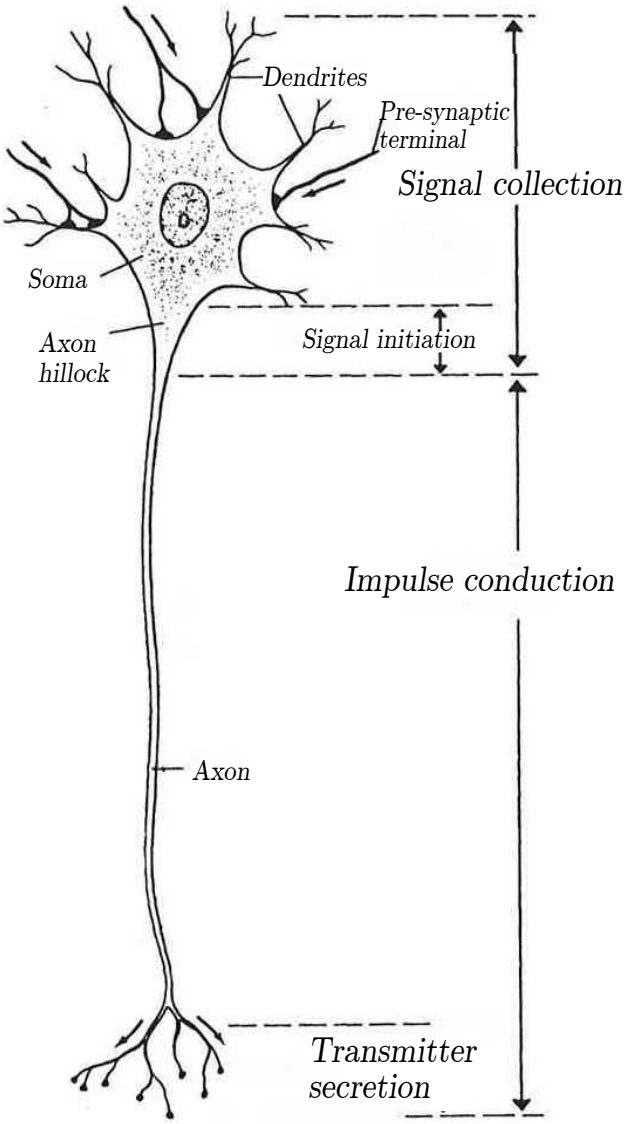


Figure 7: The structure of a neuron (reproduced from [2]).

5. **Propagation.** The nerve impulse propagates down the axon via the same process, since the sodium ions move down the axon along the potential gradient after entering the cell, depolarising the next bit of the cell and hence triggering the same process. We have a **travelling wave**.

3.4.2 Hodgkin-Huxley model for a clamped neuron

We make the simplifying assumption that the axon is can be represented as a cylinder.

We then introduce the further simplifying assumption that the axon is *space clamped*. This essentially means we place a conducting wire along the centre of the cylindrical axon, so there will be no spatial variations in the potential difference or current . An axon can then be represented schematically as in Figure 8.

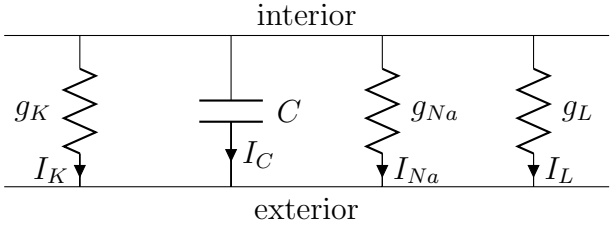


Figure 8: Schematic representation of the axon

We have conductances g_K, g_{Na} and g_L (due to the flow of potassium and sodium ions, and 'leakage' -which captures all other ions) and a capacitance, C due to the thickness of the membrane. We define:

- q - charge density (*i.e.* charge per unit length) on the interior of the axon;
- C - capacitance of the membrane per unit area;
- a - radius of axon;
- I_i - the outward current (rate of movement of ions) per unit membrane area;
- V - membrane potential inside the axon.

Hence, $q = 2\pi aCV$ and $\dfrac{dq}{dt} = 2\pi aC\dfrac{dV}{dt}$.

Now, by conservation of charge, the total current flowing across the membrane must be zero, so summing the contributions (assuming no externally applied currents) gives

$$I_C + \underbrace{I_{Na}}_{\text{sodium current}} + \underbrace{I_K}_{\text{potassium current}} + \underbrace{I_L}_{\text{leakage current (other ions)}} = 0.$$

$$2\pi aC\frac{dV}{dt} = -2\pi a\left(I_{Na} + I_K + I_L\right).$$

For the reasons discussed earlier, we take the ionic currents to be given by

$$I_{Na} = g_{Na}(V)(V - V_{Na}), \qquad I_K = g_K(V)(V - V_K), \qquad I_L = g_L(V - V_L).$$

Thus, on substituting we obtain

$$\frac{dV}{dt} = -\frac{1}{C}\left(g_{Na}(V)(V - V_{Na}) + g_K(V)(V - V_K) + g_L(V - V_L)\right),$$

where we have assumed the conductances of sodium and potassium ions to depend on the potential difference whilst the conductance of other ions (leakage) is constant. This gives the first equation of the Hodgkin-Huxley model. It remains to specify g_{Na} and g_K.

We introduce the sodium activation variable, m, the sodium inactivation variable, h, and the potassium activation, n, and write:

$$\tau_m(V)\frac{dm}{dt} = m_\infty(V) - m,$$

$$\tau_h(V)\frac{dh}{dt} = h_\infty(V) - h,$$

$$\tau_n(V)\frac{dn}{dt} = n_\infty(V) - n,$$

so for constant V, $m \to m_\infty(V)$ exponentially with time constant, τ_m, *etc.* Note that these equations take the same form as those modelling gated channels that we met earlier. The equilibrium value $n_\infty(V)$ is found to be an increasing function of V, which is why n is called the potassium activation. The naming of m and h follows the same reasoning - *i.e.* m_∞ increases with V, but h_∞ decreases with V. The forms of these functions, and the $\tau_i(V)$, chosen by Hodgkin and Huxley are quite complicated (see *e.g.* [2], p323). For our purposes, it will be enough to consider their qualitative behaviour, which we will do in the next section.

The potassium conductance depends on n^4, which suggests that the potassium channel is a four-gate channel. However, the functional form was chosen by fitting to data, rather than from any detailed knowledge of the channel, as was the sodium conductance. We take

$$g_{Na} = \bar{g}_{Na}m^3h, \qquad g_K = \bar{g}_K n^4.$$

The Hodgkin-Huxley system consists of four highly nonlinear coupled ODEs and so is very difficult to understand / analyse (especially considering digital computers were still in their infancy in the 1950s). We shall apply phase-plane methods to a simplified version of the system.

3.4.3 Simplification of the Hodgkin-Huxley model

We now reduce our four-equation model to a pair of coupled equations. We begin by writing

$$C\frac{dV}{dt} = -\left(\bar{g}_{Na}m^3h + \bar{g}_K n^4 + \bar{g}_L\right)(V - V_{eq}),$$

where

$$V_{eq} = \frac{\bar{g}_{Na}m^3hV_{Na} + \bar{g}_K n^4 V_K + \bar{g}_L V_L}{\bar{g}_{Na}m^3h + \bar{g}_K n^4 + \bar{g}_L},$$

is the *resting potential* or equilibrium potential ($V_{eq} \approx -70mV$). It turns out to be convenient to write everything in terms of the deviation from the resting potential, $\nu = V - V_{eq}$.

$$\frac{dV}{dt} = -\frac{1}{C}\left(g_{Na}(\nu)(\nu - \nu_{Na}) + g_K(\nu)(\nu - \nu_K) + g_L(\nu - \nu_L)\right),$$

$$\tau_m(\nu)\frac{dm}{dt} = m_\infty(\nu) - m,$$

$$\tau_h(\nu)\frac{dh}{dt} = h_\infty(\nu) - h,$$

$$\tau_n(\nu)\frac{dn}{dt} = n_\infty(\nu) - n,$$

where

$$\nu_K = V_K - V_{eq} = -12\text{mV}, \quad \nu_{Na} = V_{Na} - V_{eq} = 115\text{mV}, \quad \nu_L = V_L - V_{eq} = 10.6\text{mV}$$

$$g_{Na} = 0.12\Omega^{-1}\text{cm}^{-3}, \quad g_K = 0.036\Omega^{-1}\text{cm}^{-3}, \quad g_L = 3\times10^{-4}\Omega^{-1}\text{cm}^{-3}, \quad C = 1\mu\,\text{Farad}\,\text{cm}^{-2}.$$

We have also introduced:

$$m_\infty(\nu) = \frac{\alpha_m(\nu)}{\alpha_m(\nu)+\beta_m(\nu)}, \quad \tau_m(\nu) = \frac{1}{\alpha_m(\nu)+\beta_m(\nu)},$$

$$\alpha_m(\nu) = \alpha_m(V), \quad \beta_m(\nu) = \beta_m(V),$$

and similarly for h_∞, n_∞, τ_h, τ_n.

We plot the behaviour of the functions τ_h, τ_n and τ_m in Figure 9 below.

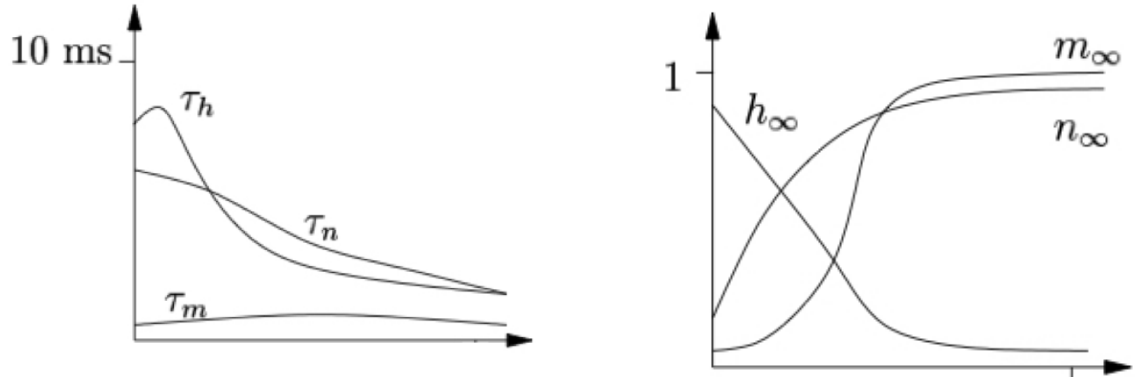


Figure 9: Sketches of the function $\tau_h(\nu)$, $\tau_m(\nu)$, and $\tau_n(\nu)$ (left) and $h_\infty(\nu)$, $m_\infty(\nu)$ and $n_\infty(\nu)$ (right)

We observe that

$$\tau_m \ll \tau_h,\, \tau_n.$$

Hence, we can make the approximation $m \approx m_\infty(\nu)$. It can also be shown numerically that $n + h \approx n_\infty + h_\infty \approx const = \bar{h} \approx 0.85$; hence, we can eliminate h. The system is then

$$C\frac{d\nu}{dt} = -\left[\bar{g}_{Na}m_\infty^3(\nu)(\bar{h}-n)(\nu-\nu_{Na}) + \bar{g}_K n^4(\nu-\nu_K) + \bar{g}_L(\nu-\nu_L)\right],$$

$$\tau_n(\nu)\frac{dn}{dt} = n_\infty(\nu) - n.$$

We shall take τ_n to be (approximately) constant (with $\tau_n \approx \bar{\tau}_n = 5$ ms), and nondimensionalise as follows:

$$\nu = \nu_{Na}\tilde{\nu}, \qquad t = \bar{\tau}_n\tilde{t},$$

where tildes indicate dimensionless quantities (note that n is already dimensionless). The dimensionless system is then (dropping tildes)

$$\frac{dn}{dt} = n_\infty(\nu) - n = k(\nu, n),$$

$$\epsilon\frac{d\nu}{dt} = -g(\nu, n) = -\left[\gamma_K n^4(\nu + \nu_K^*) + \gamma_L(\nu - \nu_L^*) + m_\infty^3(\nu)(\bar{h} - n)(\nu - 1)\right],$$

where the dimensionless parameters (and their typical experimental values) are:

$$\epsilon = \frac{C}{g_{Na}\bar{\tau}_n} = 2 \times 10^{-3}, \quad \gamma_K = \frac{g_K}{g_{Na}} \sim 0.3, \quad \gamma_L = \frac{g_L}{g_{Na}} \sim 3 \times 10^{-3},$$

$$\nu_K^* = -\frac{\nu_K}{\nu_{Na}} \sim 0.1, \quad \nu_L^* = \frac{\nu_L}{\nu_{Na}} \sim 0.1.$$

We will neglect γ_L except when ν is close to ν_K^*, as it is small.

3.4.4 Phase plane analysis

We now investigate the behaviour of the simplified system by considering the phase plane - *i.e.* the (ν, n) plane.

We begin by noting that we have one fixed point. As previously noted, the system is at equilibrium when $\nu = 0$ (corresponds to V_{eq} in the original dimensional variables). Hence, the fixed point is

$$\nu^* = 0, \qquad n^* = n_\infty(0) > 0.$$

If we linearise around this fixed point by writing $\nu = \hat{\nu}$ and $n = n_\infty(0) + \hat{n}$ with $|\hat{\nu}|, |\hat{n}| \ll 1$, by Taylor expanding we get

$$\frac{d}{dt}\begin{pmatrix} \hat{n} \\ \hat{\nu} \end{pmatrix} = \begin{pmatrix} k_n & k_\nu \\ -g_n/\epsilon & -g_v/\epsilon \end{pmatrix}\begin{pmatrix} \hat{n} \\ \hat{\nu} \end{pmatrix} = \mathbf{J}\begin{pmatrix} \hat{n} \\ \hat{\nu} \end{pmatrix},$$

where the subscripts indicate a partial derivative which is to be evaluated at the fixed point. The Jacobian matrix is given by

$$J = \begin{pmatrix} -1 & n_{\infty_\nu} \\ -g_n/\epsilon & -g_\nu/\epsilon \end{pmatrix}.$$

Then

$$\operatorname{tr} J = -1 - \frac{g_\nu}{\epsilon}, \qquad \det J = \frac{g_\nu}{\epsilon} + \frac{n_{\infty_\nu} g_n}{\epsilon}.$$

We note that $n_{\infty_\nu} > 0$ (from the earlier graph) and by inspection, it is clear that g is an increasing function of ν, so $g_\nu > 0$. By partial differentiation, we have

$$g_n = m_\infty^3(\nu)(1 - \nu) + 4n^3\gamma_K(\nu + \nu_K^*).$$

Evaluating this derivative at $\nu = 0$, $n = n_\infty(0)$, we obtaoin

$$g_n(0, n_\infty(0)) = m_\infty^3(0) + 4n_\infty^3(0)\gamma_K\nu_K^* > 0.$$

Hence $\operatorname{tr} J < 0$, $\det J > 0$ and so we have a stable fixed point. This might lead us to think that any perturbation to the potential would quickly die away, as we expect

to return to the stable fixed point. In fact, depending on where we start from, we might end up going on a long excursion before we come back to the fixed point. More precisely, a sufficiently large perturbation to ν excites an *action potential* (significant, transient change in the membrane potential). To understand why this happens, we need to sketch the behaviour of trajectories in the phase plane.

We plot the nullclines, which are given by

$$\frac{dn}{dt} = 0, \quad \Rightarrow \quad n = n_\infty(\nu),$$

$$\frac{d\nu}{dt} = 0, \quad \Rightarrow \quad g(\nu, n) = 0.$$

We hence need to sketch the curves

$$n = n_\infty(\nu), \qquad g = 0 \quad \Rightarrow \quad \frac{n^4}{\bar{h} - n} = \frac{(1-\nu)m^3(\nu)}{\gamma_K \nu_K^* (1 + \frac{\nu}{\nu_K^*})},$$

where we have neglected γ_L which is small. (Space for diagrams below).

.

Description of trajectories

- We start at a point below the ν- and n-nullclines - both ν and n are increasing in this region. Since $\epsilon \ll 1$, away from the ν-nullcline, ν is increasing very rapidly. Hence, the trajectory is almost horizontal until it reaches $g = 0$.
- After reaching the ν-nullcline, the trajectory moves upwards (as n is increasing in this region) remaining close to the ν nullcline.
- Close to the turning point of the ν-nullcline, the trajectory is in a region where ν is decreasing. It leaves the nullcline, and moves almost horizontally leftwards, until reaching the ν-nullcline again.
- The trajectory now moves downwards along the ν-nullcline, since n is decreasing in this region. It approaches the stable fixed point.

We can use this information to sketch the variation in ν over time.

3.5 Fitzhugh-Nagumo equations

Fitzhugh (1961) and Nagumo *et al.* (1962) independently considered idealised systems which were simpler to analyse than Hodgkin-Huxley, but could still demonstrate excitable behaviour. The general form of the system is

$$\epsilon\frac{d\nu}{dt} = A\nu(\nu-a)(1-\nu) - w + I^*_{ext}, \quad \frac{dw}{dt} = -w + b\nu$$

where we have now introduced the possibility of an applied inward current I^*_{ext} (which would be controlled by the experimenter), and set $w = n - n_{eq} = n - n_\infty(0)$. In the regimes of interest $0 < A <\sim O(10)$, $0 < a < 1$ and for excitable behaviour, we require b to be sufficiently large that there is only one fixed point.

When $I^*_{ext} = 0$ the nullclines satisfy

$$w = A\nu(\nu-a)(1-\nu), \qquad w = b\nu.$$

The behaviour of trajectories is then qualitatively the same as we saw in the last section. If, however, when $I^* > 0$ the ν-nulllcline is shifted upwards. We can get different behaviour depending on the shapes of the graphs (changing I^* gives rise to bifurcations).

4 Partial differential equation models

4.1 Conservation equations

In biology, we are often concerned with how the population density or concentration of various species evolves in space and time, within some domain $\mathcal{D}$. In principle, deriving appropriate model equations in this situation is quite straightforward. For the sake of argument, say that we are interested in the evolution of a chemical concentration, $c(\boldsymbol{x}, t)$ (the same principles apply to a cell or population density, or a mass density). Let V be an arbitrary closed volume within $\mathcal{D}$, and let S be the surface of V. Then

$$\text{total amount of chemical in } V = \iiint_V c \, dV$$

Any change in the amount of c in V must be due either to production or loss of the chemical inside V, or chemical entering or leaving V by crossing S. Hence

$$\frac{\partial}{\partial t} \iiint_V c \, dV = - \iint_S \boldsymbol{J} \cdot \hat{\boldsymbol{n}} \, dS + \iiint_V f(\boldsymbol{x}, t) \, dV,$$

where:

- $\boldsymbol{J}$ is the flux of c (number of molecules crossing a unit area per unit time),
- $\hat{\boldsymbol{n}}$ is the unit out ward normal to S,
- $f(\boldsymbol{x}, t)$ is the rate of production of c per unit volume.

Note the minus sign in the flux term; if the flux is in the direction of the unit outward normal, the amount of c in V decreases.

We can apply the divergence theorem to obtain

$$\frac{\partial}{\partial t} \iiint_V c \, dV = - \iiint_V \nabla \cdot \boldsymbol{J} \, dV + \iiint_V f(\boldsymbol{x}, t) \, dV.$$

We then note that, since V was arbitrary, we must have

$$\frac{\partial c}{\partial t} + \nabla \cdot \boldsymbol{J} = f(x, t) \tag{4.1}$$

at all points within $\mathcal{D}$.

Many models have this basic form. The tricky part is choosing the right form for the $\boldsymbol{J}$ and f terms; here, again, we need to use our physical intuition and our modelling skills.

4.2 The physical processes of advection and diffusion

Diffusion: If you add a drop of dye to a glass of water but don't stir it, it will over time become mixed throughout the water. This is due to the random movement of the water and dye molecules. Diffusion is more rapid at higher temperature, because the molecules are moving faster. Diffusion results in the movement of a substance (dye) from regions of high concentration to regions of lower concentration and, hence, in the reduction of concentration differences.

This implies that, if a chemical is diffusing, then the flux will depend on the **concentration gradient**. The simplest assumption would be that the flux is proportional to the concentration gradient - *i.e.*

$$\boldsymbol{J} = -D\nabla c, \tag{4.2}$$

where D is the diffusion coefficient (this will depend on temperature, and possibly other variables). Note the minus sign, since the chemical moves **down** concentration gradients.

Equation (4.2) is called **Fick's Law**. As with the Law of Mass Action, the term 'law' is slightly misleading. It is in fact just a simple model - there is no reason to believe it will be true for all types of chemicals; however, for most substances of interest, it has been shown to provide excellent agreement with experimental observations.

Advection: If you add a drop of dye to a glass of water and stir, the dye will mix through the water much faster than by diffusion only. This is because it moves with the water. Advection is then the movement of something due to a flow. Hence transport by advection occurs in the direction of the flow.

If a chemical is transported only by advection (no diffusion) with a fluid flow, then

$$\boldsymbol{J} = c\boldsymbol{v},$$

where $\boldsymbol{v}$ is the fluid velocity.

Of course, in reality, a chemical can diffuse within a fluid that is flowing; hence both types of transport occur simultaneously.

4.3 Reaction-advection-diffusion equations

In cases where transport is by diffusion and advection, our general conservation law (4.1) becomes a **reaction-advection-diffusion equation**, which has the form:

$$\frac{\partial c}{\partial t} + \nabla \cdot (c\boldsymbol{v}) = \nabla \cdot (D\nabla c) + f. \tag{4.3}$$

- $c(\boldsymbol{x}, t)$ is the dependent variable for which we want to solve (chemical concentration, cell density, *etc*)
- $\boldsymbol{v}$ is the velocity vector (usually given)

- $D(\boldsymbol{x}, t)$ is the diffusion coefficient (usually in applications, D is a constant)
- $f(c, \boldsymbol{x}, t)$ is the source/degradation/reaction term.

There are two commonly-encountered special cases:
$\boldsymbol{v} = 0$ — a reaction-diffusion model; $D = 0$ — a reaction-advection model.

If D is constant, equation (4.3) can be nondimensionalised as follows (where tildes indicate dimensionless variables)

$$\boldsymbol{x} = L\tilde{\boldsymbol{x}}, \qquad t = \frac{L^2}{D}\tilde{t}, \qquad c = c_0\tilde{c}, \qquad \boldsymbol{v} = U\tilde{\boldsymbol{v}}, \qquad f = \frac{Dc_0}{L^2}\tilde{f},$$

where U is a typical flow velocity and c_0 is a typical value of c.

The dimensionless equation is then (dropping tildes)

$$\frac{\partial c}{\partial t} + \mathcal{P}\nabla \cdot (c\boldsymbol{v}) = \nabla^2 c + f, \tag{4.4}$$

where $\mathcal{P} = UL/D$ is the Péclet number, which represents the relative importance of transport by advection, compared to diffusion. If, furthermore, the fluid is **incompressible** (which is the case for most liquids), it must obey the continuity equation

$$\nabla \cdot \boldsymbol{v} = 0.$$

Equation (4.4) then reduces to

$$\frac{\partial c}{\partial t} + \mathcal{P}(\boldsymbol{v} \cdot \nabla)c = \nabla^2 c + f. \tag{4.5}$$

Many many biological models may be approximated by a 1D reaction-advection-diffusion equation, which involves only one spatial dimension, say x.

4.4 Random motion and diffusion

Before we move on to consider a range of reaction-advection diffusion models, it is worth pointing out the important mathematical connection between random motion and the diffusion equation. Intuitively, this is clear, since the diffusion of *e.g.* a chemical through the air is due to the random motion of the chemical molecules. However, we can formalise this as follows.

Consider a collection of particles that move randomly along a line. We divide time into discrete steps of length δt, and assume that during each timestep, each particle jumps a distance δx, either to the right (with probability p_r) or to the left (with probability p_l). Let us assume the particles' movements are not biased, so $p_l = p_r = \frac{1}{2}$. Now divide the line into segments of length δx, and let $c(x,t)\delta x$ be the number of particles in the segment $[x, x + \delta x]$ at time t. Then, by conservation of particles

$$c(x, t + \delta t) = c(x, t) + p_r c(x - \delta x, t) - p_r c(x, t) + p_l c(x + \delta x, t) - p_l c(x, t).$$

Taylor expanding gives:

$$c(x, t+\delta t) = c(x,t) + \delta t \frac{\partial c}{\partial t} + \frac{1}{2}(\delta t)^2 \frac{\partial^2 c}{\partial t^2} + \ldots,$$

$$c(x \pm \delta x, t) = c(x,t) \pm \delta x \frac{\partial c}{\partial x} + \frac{1}{2}(\delta x)^2 \frac{\partial^2 c}{\partial x^2} + \ldots.$$

Substituting the Taylor expansions into the conservation equation, and using the fact that $p_r = p_l = \frac{1}{2}$, we have'

$$\delta t \frac{\partial c}{\partial t} + O(\delta t^2) = \frac{1}{2}(\delta x)^2 \frac{\partial^2 c}{\partial x^2} + O(\delta x)^4.$$

We now divide through by δt and take the limit $\delta t \to 0$ assuming

$$\lim_{\delta t \to 0} \frac{(\delta x)^2}{2\delta t} = D \,(\text{constant}).$$

Our conservation equation for c is thus approximated, at leading order, by the diffusion equation

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}, \quad \text{where } D = \frac{(\delta x)^2}{2\delta t}.$$

Note, that we can use the same argument in two or three dimensions, and similarly obtain the diffusion equation as our leading-order approximation (though the algebra becomes much messier!).

In fact, it is possible to take into account the possibility that $p_r \neq p_l$, or that a particle may not move into a position which is already occupied, *etc.* (though the resulting equations become more complicated than the diffusion equation). Relating the collective behaviour of a large number of individuals (*e.g.* cells) to the rules governing the behaviour of each individual is an active research area in mathematical biology - see *e.g.* the work of Professor Kerry Landman (Melbourne), Professor Matthew Simpson (QUT) and Associate Professor Ben Binder (Adelaide).

4.5 Reaction-advection models: the membrane oxygenator

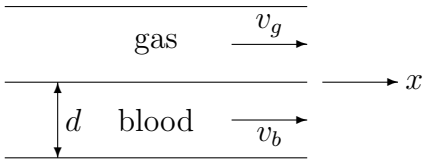
When diffusion is unimportant, the reaction-advection-diffusion equation reduces to the simpler reaction-advection equation. We consider an application of this below.

4.5.1 Motivation: Heart surgery

In cardiac surgery, it may be necessary to bypass the heart in order to perform the procedure. During this time, the patient's circulation must be maintained artificially. This is done using a heart-lung machine, which pumps the blood through a membrane oxygenator to allow oxygenation and carbon dioxide removal, before returning it to the body.

An idealised membrane oxygenator is depicted below; a thin gas permeable membrane separates the blood and gas flows in the cardiopulmonary bypass circuit. Oxygen

diffuses from the gas side into the blood, and carbon dioxide diffuses from the blood into the gas for disposal (*e.g.* see 'membrane oxygenator' on Wikipedia). An important issue in designing the machine will be to understand how long it takes for the blood to be reoxygenated - *i.e.* how long does the gas exchange channel have to be? We will try to provide a answer to this question by deriving a simple 1D model of the oxygenation process.



4.5.2 Model development

We will base our model on the simplified set-up shown above, which captures the essential mechanism of the process. For simplicity, we make the following assumptions:

- Blood is an incompressible fluid
- The gas flows much faster than the blood ($v_g \gg v_b$) so that the concentration of oxygen in the gas (C_g) is approximately constant.
- The concentration of oxygen in the blood is a function of position x and time t only.
- Advection is the dominant transport process in the blood; neglect diffusion.
- Transport of oxygen from gas to blood is a proportional to the concentration difference.

By conservation of mass, the oxygen concentration in the blood, $C(x,t)$ is given by

$$\frac{\partial C}{\partial t} + v_b \frac{\partial C}{\partial x} = h(C_g - C(x,t)),$$

where h is a transfer coefficient. Note that h has units of $[T]^{-1}$, and so gives a timescale for oxygen transfer. To complete this model we need an initial condition and a boundary condition, *i.e.* $C(x,0) = f(x)$, and $C(0,t) = g(t)$. We will assume that blood enters the channel at $x = 0$ with some constant oxygen concentration, C_0, where $C_0 < C_g$, since oxygen will have been consumed by the body. Similarly, we will assume that at $t = 0$ all the blood in the channel is at this level of oxygenation.

4.5.3 Solution

We nondimensionalise, based on a general timescale T, with the relevant lengthscale being the distance travelled by the blood on the oxygen transfer timescale, *i.e.*

$$t = T\tilde{t}, \qquad x = \frac{v_b}{h}\tilde{x}, \qquad C = C_g\tilde{C}.$$

The dimensionless governing equation is then

$$\tau \frac{\partial C}{\partial t} + \frac{\partial C}{\partial x} = 1 - C,$$

where $\tau = 1/hT$ is the ratio of the timescale of oxygen transfer to the timescale of interest. The boundary and initial conditions now become $C(0,t) = C(x,0) = C_i$, where $C_i = C_0/C_g < 1$.

Based on physical intuition, we would expect that after a short start-up time, the oxygen concentration in the device would reach steady state. Hence we assume $\tau \ll 1$ and look for a solution C that is a function of x only. The PDE then reduces to an ODE for $C(x)$:

$$\frac{dC}{dx} = 1 - C(x), \quad \text{subject to } C(0) = C_i.$$

The steady-state solution is

$$C(x) = 1 + (C_i - 1)e^{-x}. \tag{4.6}$$

Note that as $x \to \infty$, $C \to 1$ (*i.e.* the oxygen concentration tends to that in the gas), which makes physical sense.

Now, suppose that by the time it leaves the channel, we want the oxygen concentration in the blood to have reached some fraction C^* of its value in the gas (we would need to be advised by medics what a suitable value for C^* might be. Then, the required length of the channel satisfies

$$C^* = 1 + (C_i - 1)e^{-L}.$$

Hence,

$$L = \log \left| \frac{1 - C_i}{1 - C^*} \right|. \tag{4.7}$$

If we let l be the dimensional length required, re-dimensionalising (4.7) gives

$$l = \frac{v_b}{h} \log \left| \frac{C_g - C_0}{C_g(1 - C^*)} \right|.$$

Thus, the required length of the channel will depend on the flow rate of the blood, the channel depth, and the transfer coefficient (which will depend on the material used for the membrane), as well as C_i and C^*.

4.5.4 How happy are we with our answer?

Whenever we consider the results of a model, we need to think carefully about what factors have been left out, and how they might affect the results. In our model above we have neglected diffusion of oxygen in the blood, and we have also assumed that the blood velocity is constant throughout the channel. These assumptions certainly made our life easier mathematically, but are they physically sensible?

In fact, for real fluids the velocity varies across the channel, with a maximum in the centre, and zero velocity at the walls (this is called **Poiseuille flow**). Hence blood near the wall will become reoxygenated over a shorter distance than that near the centre of the channel, which is moving faster. This in turn will set up a gradient in oxygen concentration within the channel, which can mean that diffusion within the blood is important. This phenomenon is known as Taylor dispersion (after G. I. Taylor, whose nuclear explosion analysis we met at the start of the course).

4.6 An age-structured model for chemotherapy

In this section, we will study a version of a model by Himmelstein and Bischoff [3] for chemotherapy of leukemia cells. This turns out to be another application of the reaction-advection equation, although this is perhaps not obvious initially. The aim of the model is to understand how chemotherapy might change the age distribution of the malignant cells, and takes account of the possibility that the rate at which the drug kills the cells might depend on their age, as well as the drug concentration.

4.6.1 Model formulation

We determine an evolution equation for $n(a,t)$, where $n(a,t)$ is the number of malignant cells with ages between a and $a+\delta a$ at time t. The age of the cell might more properly be called its 'maturity', as in this application the important factor is the phase of the cell cycle which the cell has reached (cells only divide in one phase of the cell cycle, whilst some drugs are only effective against cells that are in a particular phase). Hence, in general we will have a parameter, τ, in our model, which is the timescale over which cells mature. However, for simplicity, in this example, we will set $\tau = 1$, so the maturity corresponds to chronological age (we can always rescale time to achieve this).

After a small time δt, the maturity or age of a cell increases by $\delta a = \delta t$. Then, by conservation of the cell population

$$n(a, t+\delta t) = n(a,t) + n(a-\delta a, t) - n(a,t) - \mu(a) n(a,t)\delta t,$$

where $\mu(a,t)$ is the age and time dependent death rate of the cells. The form of this term will depend on the particular action of the drug used. Dividing by δt we obtain

$$\begin{aligned}
\frac{n(a,t+\delta t) - n(a,t)}{\delta t} &= \frac{n(a-\delta a,t) - n(a,t)}{\delta t} - \mu(a,t)n(a,t) \\
&= -\frac{n(a,t) - n(a-\delta a,t)}{\delta a}\frac{\delta a}{\delta t} - \mu(a,t)n(a,t) \\
&= -\frac{n(a,t) - n(a-\delta a,t)}{\delta a} - \mu(a,t)n(a,t).
\end{aligned}$$

Then, a little manipulation and taking limits as $\delta t \to 0$, gives **von Foerster's equation**

$$\frac{\partial n}{\partial t} + \frac{\partial n}{\partial a} = -\mu(a,t)n,$$

a reaction-advection equation; age a takes the place of spatial position; the 'velocity' $v = 1$ is the rate of change of age with time.

To complete the model description we need an initial condition, and a boundary condition. The initial condition supplies the age distribution $f(a)$ of the population at time $t = 0$

$$n(a, 0) = f(a).$$

The boundary condition at $a = 0$ tells us about how the population reproduces. Let the rate of reproduction of individuals of age a be $b(a)$ (we can think of this as the birth rate). Then, the number of new individuals or cells born at time t is given by

$$n(0, t) = \int_0^\infty b(a)n(a, t)\, da.$$

Note that often we will have situations where $b(a) = 0$ if $a < a_1$ or $a > a_2$ (where $a_1 < a_2$), indicating that individuals have to reach some minimum age before they are able to reproduce, and are no longer able to reproduce once they become too old.

For the particular case of interest in [3], it is assumed that when cells reach a level of maturity equivalent to having age $a = 1$, they divide, giving two new cells of age zero. Hence, our boundary condition is

$$n(0, t) = 2n(1, t). \tag{4.8}$$

The initial condition is

$$n(a, 0) = N_0\gamma(a). \tag{4.9}$$

Here N_0 is the initial number of cells, and $\gamma(a)$ gives their age distribution at the start of the experiment. Note that this implies $\int_0^1 \gamma(a)\, da = 1$. Note the upper limit on the integral is due to cell division occurring at age 1.

Solution for a drug with low cell-cycle specificity

In general, solving this type of model requires the use of numerical methods. However, for the simple case where $\mu = \mu(t)$ (*i.e.* the drug concentration, and hence the death rate, is a known function of time, but does not depend on cell age) we try the method of separation of variables (this would not work for all choices of $\mu(a, t)$). We set

$$n(a, t) = f(t)w(a).$$

Then, substituting into the PDE yields

$$wf' = -fw' - \mu(t)fw,$$

where the primes indicate differentiation. Dividing through by fw yields

$$\frac{f'}{f} + \mu(t) = -\frac{w'}{w} = \lambda \text{ (constant)}.$$

Hence we have

$$f' = (\lambda - \mu(t))f \quad \Rightarrow \quad f = Ae^{\lambda t}e^{-M(t)}$$

where $M(t) = \int_0^t \mu(\tau)\, d\tau$ and A is an arbitrary constant. Our equation for w is then

$$w' = -\lambda w \quad \Rightarrow \quad w = Be^{-\lambda a}.$$

After rearranging, and tidying up constants, our solution is

$$n(a,t) = Ce^{\lambda t}e^{-M(t)}e^{-\lambda a}$$

where C and λ are constants to be determined. We must use the initial and boundary conditions to fix these. We require

$$n(a,0) = N_0\gamma(a) = Ce^{-\lambda a}.$$

Integrating both side from $a = 0$ to $a = 1$ gives

$$N_0 = -\frac{C}{\lambda}(e^{-\lambda} - 1) \quad \Rightarrow \quad C = \frac{\lambda N_0}{1 - e^{-\lambda}}.$$

Now we impose the boundary condition, which implies

$$2e^{-\lambda} = 1 \quad \Rightarrow \quad \lambda = \log 2.$$

Hence our solution is

$$n(a,t) = 2\log 2 N_0 e^{\log 2(t-a)}e^{-M(t)}.$$

Integrating the above from $a = 0$ to $a = 1$ gives $N(t)$, the total cell population at time t. Thus

$$N(t) = N_0 e^{\log 2t}e^{-M(t)} = N_0 2^t e^{-M(t)}.$$

We note that

$$n(0,t) = 2\log 2 N(t),$$

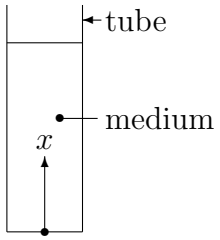
so that the new-born cells make up a constant proportion of the total cell population.

4.7 Reaction-diffusion models

4.7.1 Embryo culture *in vitro*

In animal breeding programs, as well as *in vitro* fertilisation technology, embryos must be kept alive in culture media in the laboratory. Typically the culture environment is static, *i.e.* the medium is stationary, and nutrient transport is by diffusion.

Suppose that a tube of medium contains a number of embryos. The tube is open at the top, the oxygen supply for the embryos is the surrounding air and the transport mechanism through the medium to the embryos is diffusion.



Assuming that the oxygen concentration in the medium is a function of position x and time t, *i.e.* at a given time t, the oxygen concentration is constant across any cross section of the tube, we derive the 1D reaction-diffusion equation.

We assume the flux J to be given by **Fick's law**, so it is proportional to the concentration gradient, - *i.e.*

$$J(x,t) = -D(x,t)\frac{\partial C}{\partial x}, \quad D = \text{diffusion coefficient.}$$

Substituting into the PDE:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right) - Q.$$

To complete the model we need an initial condition, $C(x,0) = C_0 = $ constant, and **two** boundary conditions: no oxygen flux at $x = 0$ and continuity of partial pressure at $x = h$, *i.e.*

$$\frac{\partial C}{\partial x}(0,t) = 0, \quad \gamma C(h,t) = P_a.$$

Note: Henry's law gives $P = \gamma C$; P = partial pressure, C = concentration, γ = Henry's law constant; partial pressure in air P_a is assumed to be constant.

As in the membrane oxygenator example, we assume that after a short period, the oxygen concentration in the medium will reach a steady state and to find this we solve the steady-state reaction-advection equation (an ODE). The solution $C(x)$ of course depends on $Q(x)$. If $Q(x) = Q = $ constant, solving for C is easy.

Note that, depending on the values of the various parameters, we could have $C(0) < 0$ which would be non-physical, in which case the model would need modification. Actually, we might expect Q to be concentration dependent. For example Q might be proportional to concentration $Q = kC$, or given by the Michaelis-Menten function

$$Q = \frac{Q_{max}C}{K_m + C}.$$

Either way we have $Q = 0$ if $C = 0$ which is sensible. Is $Q = kC$ a good choice? Probably fine so long as C_0 is not too large, otherwise the Michaelis-Menten function is preferable or there may be a better functional form for Q.

If $Q = kC$ is appropriate we can still easily find an analytic solution to the ODE — exercise.

5 Travelling waves

5.1 Motivation: A wound healing assay

Medical scientists aim to improve our understanding of the wound healing process to improve outcomes for patients who suffer injuries, *e.g.* reduced recovery times, reduction of scarring. For both practical and ethical reasons, they use *in vitro* assays for their experiments. Cells are grown to confluence on a substance that mimics the cells' natural environment, *e.g.* collagen, laminin, fibronectin, *etc.* . A 'wound' is then made by scraping away the cells in a certain area. The cells then migrate and proliferate, forming a moving front which invades the now unoccupied region, closing up the wound [4]. The effects of, for example, different drugs on the wound healing process can then be assessed by adding them to the cells culture medium. We are interested in understanding how the effects of cell movement and proliferation both contribute to wound healing. Can we predict how quickly the wound will close?

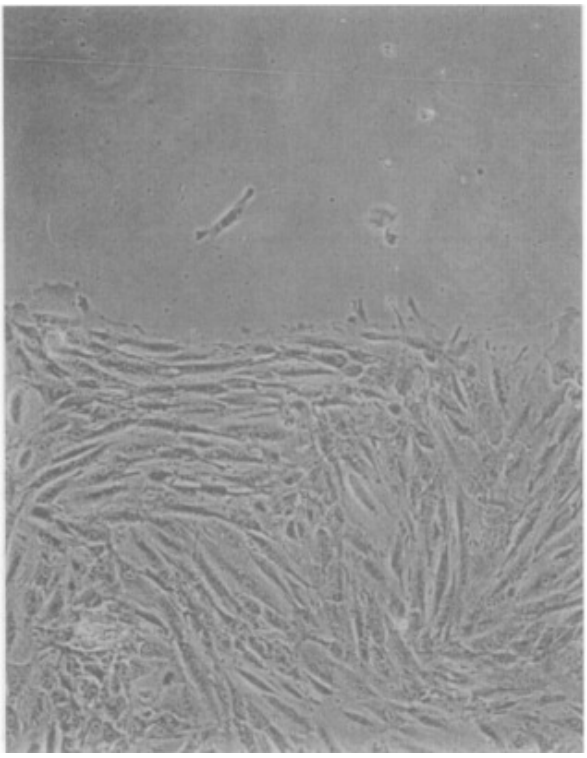


Figure 10: The cell front in the *in vitro* wound assay, ten hours after 'wounding' (from [4]).

5.2 Mathematical model

We model the proliferation and migration of the cells using **Fisher's equation**:

$$\frac{\partial n}{\partial t} = D\nabla^2 n + \mu n(1 - \frac{n}{K}).$$

It describes the change in the cell density $n(\boldsymbol{x}, t)$ due to (logistic) growth ($\mu n(1-n/K)$) and random motion ($D\nabla^2 n$). This equation was originally proposed by Fisher in 1937 to model the spread of an advantageous gene in a population. It has also been used to model of a species invading a new habitat in ecology.

We will consider just one spatial dimension (perpendicular to the cell front), and nondimensionalise the model as follows

$$t = \frac{\tilde{t}}{\mu}, \qquad x = \sqrt{\frac{D}{\mu}}\tilde{x}, \qquad n = K\tilde{n},$$

which, on dropping the tildes, reduces the equation to

$$\frac{\partial n}{\partial t} = \frac{\partial^2 n}{\partial x^2} + n(1-n).$$

For $x \to -\infty$ we assume the cell population has already reached the carrying capacity; for $x \to \infty$ the cells have not yet arrived. Hence we impose $n(-\infty) = 1$, $n(\infty) = 0$.

5.3 Travelling wave solution

Next, we make the **travelling wave ansatz** (educated guess verified later by its results)

$$n(x,t) = \phi(x - ct), \quad \phi(-\infty) = 1, \quad \phi(\infty) = 0.$$

The parameter $c > 0$ is the **wave speed**, the new variable $z = x - ct$ is the **wave variable**, and the function $\phi(z)$ is the **wave profile**.

Substituting into Fisher's equation yields a 2nd-order ODE for $\phi(z)$ which can be written as a system of two 1st-order ODEs: Now,

$$\frac{\partial n}{\partial t} = -c\frac{d\phi}{dz}, \quad \frac{\partial^2 n}{\partial x^2} = \frac{d^2\phi}{dz^2},$$

which we can substitute into Fisher's equation to give

$$-c\frac{d\phi}{dz} = \frac{d^2\phi}{dz^2} + \phi(1-\phi)$$

As previously we write this 2nd-order ODE as a system of two 1st-order ODEs by defining $\psi = d\phi/dz$. Then

$$\begin{aligned} \phi' &= \psi, \\ \psi' &= -c\psi - \phi(1-\phi). \end{aligned}$$

We find the equilibria (steady states) and linearise to examine the behaviour in their near vicinity.

There are two equilibria:

- $P_1 = (0,0)$: stable (both eigenvalues of $J(0,0)$ have negative real part) for $c > 0$. For $c < 2$ a stable spiral; for $c > 2$ a stable node.

- $P_2 = (1, 0)$: a saddle.

Examining the vector fields shows that the solutions for $c < 2$ are not biologically relevant (see Figs. 11-11). However, if $c > 2$, the solutions are biologically meaningful (see Figs. 13-14). The minimum wave speed for a biologically relevant wave front solution is $c^* = 2$.

(Note: to plot the form of $\phi(x - ct)$ versus t at a fixed point x observe that $z = x - ct$ decreases as t increases, so we follow the curve in the phase plane in the direction of decreasing z, *i.e.* the population density increases from $\phi = 0$ to $\phi = 1$.)

5.4 The linear conjecture

The minimum wave speed c^* is exactly that value where $(0, 0)$ changes from a spiral into a node. If we consider the travelling wave solution close to $(0, 0)$, then the behaviour is described by the linearisation around $(0, 0)$, in particular the eigenvalues of $J(0, 0)$:

$$\lambda_{1/2} = -\frac{c}{2} \pm \frac{c}{2}\sqrt{1 - \frac{4}{c^2}}.$$

For $c = c^* = 2$ we have an eigenvalue of multiplicity 2, $\lambda = -1$.

Solving the linearised problem we find

$$\begin{bmatrix} \phi \\ \psi \end{bmatrix} = e^{-z}(C_1 + C_2 z) \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The solution near $(0, 0)$, i.e. for $z \to \infty$, behaves like e^{-z} and, hence, like e^{-x} for $x \to \infty$. Then, on converting x back to dimensional form $-\sqrt{\dfrac{\mu}{D}} = -\dfrac{c^*}{2D}$ (since $c^* = 2\sqrt{\mu D}$) is the decay rate of the solution as $x \to \infty$.

In many cases, it is enough to measure the decay rate of the profile for large x to get a good approximation for the minimum wave speed c^*. This is known as the *linear conjecture*.

General Fisher Equation

The result on the minimum wave speed of travelling fronts can be generalized to general Fisher equations

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + f(u)$$

where $f(u)$ is a function with parameter $K > 0$ such that

$$\begin{aligned} f(0) = 0 \qquad & f(K) = 0, \\ f(u) > 0 \qquad & \text{for all } 0 < u < K, \\ f'(0) > 0 \qquad & f'(K) < 0 \end{aligned}$$

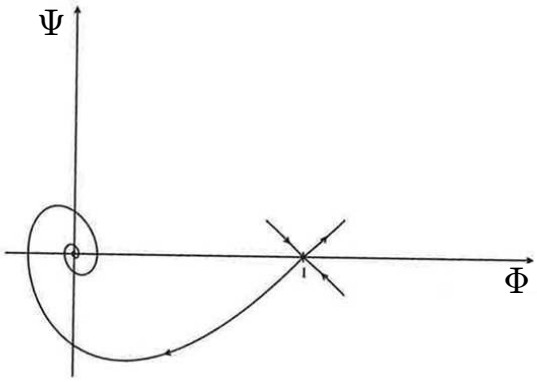


Figure 11: The phase plane for $c < 2$. (Note: $\phi < 0$ along the trajectory connecting the fixed points, so this case is not biologically relevant.)

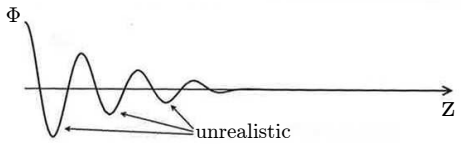


Figure 12: The cell front for $c < 2$.

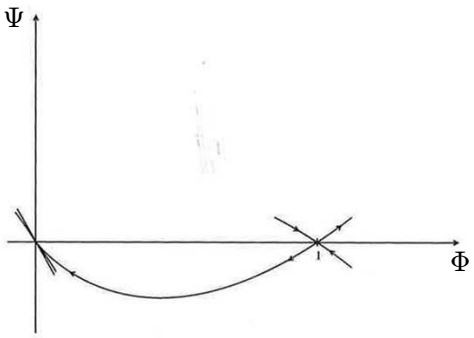


Figure 13: The phase plane for $c > 2$; note $\phi > 0$ along the trajectory connecting the two fixed points.

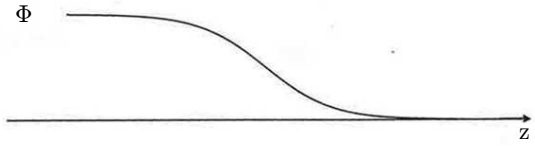


Figure 14: The cell front for $c > 2$.

(The logistic term, $\mu u(1-u)$, satisfies these conditions.)

If we also assume that $f(u)$ satisfies the *subtangential condition* $f'(0)u > f(u)$ for all $0 < u < \infty$ (which means that $f(u)$, $u \geq 0$ has maximum slope at $u = 0$), then the minimum wave speed is

$$c^* = 2\sqrt{Df'(0)}.$$

6 The Keller-Segel Model

6.1 Motivation: How to build a slug

Slime moulds may not sound like particularly appealing organisms, but they are fascinating to biologists as they represent a bridge between primitive, unicellular life forms (such as bacteria) and sophisticated multicellular organisms like ourselves. One of the best-studied slime moulds is *Dictyostelium discoideum*, which commonly lives as hundreds of thousands of single amoeboid cells in the soil. However, when food supplies run low, a fascinating series of events is triggered. Firstly, an initially uniform distribution of cells develops centres of organisation called **aggregation sites**. Cells move towards these foci, sometimes in a pulsating, wave-like manner. Contacts form between neighbours and the cells converge on a single site, eventually forming a multicellular 'blob', known as a slug. The slug can move about as a single body, although of course, each cell is an independent organism. Within the slug, changes occur, with cells towards the front becoming **prestalk** cells, whilst those at the rear become **prespores**. After crawling around for a while, until it finds a suitable environment, the slug undergoes a series of shape changes. This ends with a cellular streaming event which resembles a 'reverse fountain', which brings all the prestalk cells around the outside and down through the centre of the cell mass. The results is a slender stalk, supporting a spore-filled capsule. The stalk cells harden and die, but in doing so, they provide an opportunity for the spore cells to be carried on air currents to a more favourable environment, and thus survive and propagate the species. This cooperation between cells gives *disctyostelium* an advantage compared to other single-celled organisms, which would simply starve to death; however, the price is that part of the colony must die to form the stalk.

Understanding this process of aggregation and organisation would help us to understand how multicellular life began, and might provide insights into how cells organise themselves into tissues during development. The central question is: how do the cells 'know' how to build a slug? Since all the single-celled amoebae are the same, there is no obvious candidate for a 'leader' to organise the others. How, then, does the initially uniform distribution of cells break up into clusters around the aggregation centres?

It turns out that the starving slime mould cells secrete a chemical called cyclic AMP (cAMP), which is attractive to other cells. Free cAMP in the extracellular environment promotes further cAMP secretion by the cells. The cells also secrete an enzyme called *phosphodisterase* that degrades cAMP. We will now aim to build a mathematical model to try to detemine if these simple mechanisms are sufficient to cause **dictyostelium** to aggregate.

6.2 Modelling cell aggregation

In order to keep our model simple enough to analyse, we make the following assumptions:

- In the absence of chemical signals, *dictyostelium* cells move around at random;
- Cells are attracted to cAMP, so will tend to move from areas of low cAMP concentration to those of higher concentration;
- Cell proliferation and death is negligible during the aggregation process;
- cAMP diffuses freely in the environment;
- Each cell produces cAMP at a constant rate;
- The rate of degradation of cAMP in the environment is proportional to its concentration.

It is worth noting that only the last two assumptions are drastic simplifications of the biology.

Now, for simplicity, let us consider a one dimensional domain $0 \leq x \leq L$, and let $a(x,t)$ and $c(x,t)$ be, respectively, the cell density and chemical concentration. Then, our model equations will be of the form:

$$\frac{\partial a}{\partial t} = \underbrace{\mu \frac{\partial^2 a}{\partial x^2}}_{\text{random motion}} - \underbrace{\frac{\partial}{\partial x}(J_{chem})}_{\text{attraction to cAMP}}, \tag{6.1a}$$

$$\frac{\partial c}{\partial t} = \underbrace{D\frac{\partial^2 c}{\partial x^2}}_{\text{diffusion}} + \underbrace{pa}_{\text{production}} - \underbrace{kc}_{\text{degradation}}, \tag{6.1b}$$

where:

- μ is the random motility (diffusion) coefficient for the cells
- D is the diffusion coefficient for the chemical
- J_{chem} is the cell flux due to chemical attraction
- p and k are the rate of cAMP production and degradation, respectively.

It only remains to specify the form of the J_{chem} or **chemotaxis** term (chemotaxis= chemo (chemical) + taxis (movement) = movement up a chemical gradient). Since the cells are attracted to cAMP, they will move in the direction of $\nabla c = \dfrac{\partial c}{\partial x}$. Hence we set

$$J_{chem} = \chi a \frac{\partial c}{\partial x},$$

where χ is a constant called the **chemotactic coefficient**, which describes how attracted the cells are to cAMP (if $\chi = 0$ the cells are not influenced by the chemical).

On substituting the above into (6.1), we obtain the **Keller-Segel model** of chemotaxis:

$$\frac{\partial a}{\partial t} = \mu \frac{\partial^2 a}{\partial x^2} - \chi \frac{\partial}{\partial x}\left(a \frac{\partial c}{\partial x}\right), \tag{6.2a}$$

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + pa - kc. \tag{6.2b}$$

We assume the cells are grown in a confined region, so that there is no flux of either chemical or cells at $x = 0,\ L$. Hence the boundary conditions are

$$\frac{\partial a}{\partial x} = \frac{\partial c}{\partial x} = 0, \quad \text{at } x = 0,\ L.$$

Nondimensionalisation

We can nondimensionalise the model in the following way

$$x = L\tilde{x}, \qquad t = T\tilde{t}, \qquad a = a_i\tilde{a}, \qquad c = c^*\tilde{c},$$

where a_i is the initial density of cells, and the timescale, T, and chemical concentration scale c^* are yet to be determined.

The dimensionless equations are then (dropping tildes)

$$\frac{\partial a}{\partial t} = \frac{\mu T}{L^2}\frac{\partial^2 a}{\partial x^2} - \frac{\chi T c^*}{L^2}\frac{\partial}{\partial x}\left(a\frac{\partial c}{\partial x}\right),$$

$$\frac{L^2}{TD}\frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial x^2} + \frac{pa_iL^2}{Dc^*}a - \frac{kL^2}{D}c.$$

We are interested in cell motion in this model, so we set our timescale to that for cell diffusion across the domain (*i.e.* we set $T = L^2/\mu$). We then note that $L^2/TD = \mu/D$ represents the ratio of the diffusivities of the cells and cAMP. We assume that chemical diffusion is much faster than cell movement, so $\delta = \mu/D \ll 1$.We set $c^* = pa_iL^2/D$, noting that this is the chemical concentration level due to production by the cells on the diffusion timescale for the chemical (since L^2/D is has dimensions of time). This leaves us with two dimensionless parameters $\tilde{\chi} = pa_i\chi L^2/D\mu$ and $\tilde{k} = kL^2/D$.

Our dimensionless system is then (dropping tildes):

$$\frac{\partial a}{\partial t} = \frac{\partial^2 a}{\partial x^2} - \chi\frac{\partial}{\partial x}\left(a\frac{\partial c}{\partial x}\right), \tag{6.3a}$$

$$\delta\frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial x^2} + a - kc. \tag{6.3b}$$

The boundary conditions are:

$$\frac{\partial a}{\partial x} = \frac{\partial c}{\partial x} = 0, \quad \text{at } x = 0,\ 1. \tag{6.3c}$$

Note that henceforth, we shall set $\delta = 0$.

6.3 Linear stability analysis

Our modelling was motivated by the question of whether the attraction of the *dicytostelium* cells to cAMP was a sufficient mechanism to cause the break up of the initially uniform distribution of cells into clumps, as seen in the early stages of slug formation. Now, we need to pose this question in a mathematical form that we can answer using our model.

A **spatially uniform** or **spatially homogeneous steady state** of our model (6.3) is a solution that is constant in both space and time - *i.e.*

$$a(x,t) = \bar{a}, \qquad c(x,t) = \bar{c},$$

where $\bar{a}$ and $\bar{c}$ are constants.

On substituting into (6.3), we see the first equation is automatically satisfied, whilst the second yields the condition

$$\bar{a} - k\bar{c} = 0, \qquad \Rightarrow \bar{c} = \frac{\bar{a}}{k}.$$

This means that for a spatially homogenous steady state to exist, the rate of chemical degradation must be exactly balanced by its rate of secretion by the cells.

We now wish to determine if this solution is stable or not. If it is, then if we start with a relatively uniform distribution of cells, we would expect this to persist for all time. If it is unstable, then we expect that any small non-uniformities in the initial cell distribution will grow in time, and we will then see some regions where the cell density increases, and other where it decreases - *i.e.* the beginnings of aggregation.

We consider small amplitude perturbations to the spatially homogenous steady state, so that

$$a = \bar{a} + \epsilon a_1(x,t) + O(\epsilon^2), \qquad c = \bar{c} + \epsilon c_1(x,t) + O(\epsilon^2),$$

where $\epsilon \ll 1$. Substituting into (6.3) (with $\delta = 0$) and neglecting terms of $O(\epsilon^2)$ or smaller, gives

$$\frac{\partial a_1}{\partial t} = \frac{\partial^2 a_1}{\partial x^2} - \chi\bar{a}\frac{\partial^2 c_1}{\partial x^2} \tag{6.4a}$$

$$\frac{\partial^2 c_1}{\partial x^2} + a_1 - kc_1 = 0. \tag{6.4b}$$

From our earlier studies of the diffusion equation, we anticipate the solutions will be of the form $a_1 \propto e^{\lambda t}\cos qx$ where λ and q are to be determined. Note that the no flux boundary conditions at $x = 0$ have led us to reject the possibility of a $\sin qx$ term. Since the chemical is produced by the cells, we make a similar ansatz for c_1. In order to satisfy the boundary condition at $x = 1$, we require $q = m\pi$, where m is an integer. We call q the **wavenumber**; it provides a measure of the 'waviness' of the pattern, since $2\pi/q$ is the wavelength. λ is the growth rate of the perturbation: if $\lambda > 0$, small perturbations to the initial uniform state will grow in amplitude, whilst if $\lambda < 0$ they will decay. Thus the spatially uniform state is stable if $\lambda < 0$ and unstable if $\lambda > 0$.

We thus set

$$a_1 = Ae^{\lambda t}\cos qx, \qquad c_1 = Ce^{\lambda t}\cos qx,$$

which, on substituting into (6.4) leads to

$$A\lambda = -q^2 A + q^2 C\chi\bar{a},$$

$$-q^2 C + A - kC = 0.$$

Hence, eliminating between these two equations we obtain:

$$\lambda = q^2\left(\frac{\chi\bar{a}}{k+q^2} - 1\right) \tag{6.5}$$

This equation for the growth rate λ in terms of q is called the **dispersion relation**. (The dependence on q^2 is due to the fact that $\cos qx = \cos(-qx)$.)

Definitions:

- A **necessary condition** for instability is a condition that holds for any case in which instability occurs - *i.e.* instability occurs $\Rightarrow$ condition holds. (However, there may be situations in which the condition holds, but instability does not occur.)
- A **sufficient condition** for instability is a condition such that, if it holds, instability occurs - *i.e.* condition holds $\Rightarrow$ instability occurs. (The possibility that instability may occur when the condition is not satisfied is not excluded - *i.e.* we may not have found all the conditions that give instability.)
- A **necessary and sufficient condition** for instability is a condition such that instability occurs if and only if the condition holds - *i.e.* condition holds $\Leftrightarrow$ instability occurs.

With these definitions in place, we now consider the dispersion relation (6.5) in more detail. For instability to occur, we need $\lambda > 0$. Thus, a necessary and sufficient condition for instability is that there exists at least one permissible wavenumber, $q \neq 0$, such that

$$\frac{\chi\bar{a}}{k+q^2} - 1 > 0 \quad \Rightarrow \frac{\chi\bar{a}}{k+q^2} > 1.$$

We note that the LHS of the second inequality is maximised when q^2 is minimised. Since the values of q are restricted to integer multiples of π, the smallest non-zero value q^2 can take is $q^2 = \pi^2$. Hence the condition can be re-written as

$$\frac{\chi\bar{a}}{k+\pi^2} > 1.$$

Based on the definition of the dimensionless parameters, we interpret this condition as telling us that aggregation is more likely to occur when:

- Cells movement is mainly due to chemotaxis (relative to diffusion)

- The initial cell density is high
- The rate of production of the chemical is high.

Aggregation is less likely to occur when:

- There is too much random motion of the cells
- The rate of decay of the chemical is high
- The chemical diffuses rapidly (relative to its rate of production).

Note that is the smallest wavenumber (*i.e.* longest wavelength) perturbation which is most likely to be unstable. This is because the effects of diffusion operate most strongly to smooth large gradients on small lengthscales, so long-range, gradual variations in cell density and chemical concentration are least likely to be effaced by diffusion. However, there may be more than one permissible wavenumber, q, which gives a value of $\lambda > 0$ in equation (6.5). In that case, we would expect the cell clumps to have a lengthscale corresponding to the wavenumber, q, which gives the largest value of λ; we call this wavenumber the fastest growing mode.

Example: Travelling waves of chemotactic bacteria

Consider a colony of bacteria, which are attracted towards a nutrient that they consume. We let the bacterial cell density be b, and the nutrient concentration c. We consider a simple model for this situation, in one dimension, given by

$$\frac{\partial b}{\partial t} = D\frac{\partial^2 b}{\partial x^2} - \frac{\partial}{\partial x}\left(\chi\frac{b}{c}\frac{\partial c}{\partial x}\right), \tag{6.6a}$$

$$\frac{\partial c}{\partial t} = -kb, \tag{6.6b}$$

where D is the random motility (diffusion) coefficient of the bacteria and k is the rate of nutrient consumption. Note the form of the chemotaxis term: by comparing to the standard Keller-Segel model, we see the chemotactic flux now includes an additional factor of $1/c$; this models the fact that the bacteria are less inclined to move up a nutrient concentration gradient if they have plenty of nutrient at their current location.

We look for travelling wave solutions, just as in the case of Fisher's equation by writing $b(x,t) = B(z)$, $c(x,t) = C(z)$, where $z = x - vt$ (in this case, we denote the wavespeed by v). The boundary conditions are:

$$B \to 0 \quad \text{as} \quad |z| \to \infty,$$

$$C \to 0 \quad \text{as} \quad z \to -\infty, \qquad C \to 1 \quad \text{as} \quad z \to \infty.$$

Substituting the travelling wave ansatz into the model equations (6.6) gives

$$-v\frac{dB}{dz} = D\frac{d^2 B}{dz^2} - \chi\frac{d}{dz}\left(\frac{B}{C}\frac{dC}{dz}\right), \tag{6.7a}$$

$$-v\frac{dC}{dz} = -kB. \tag{6.7b}$$

Integrating (6.7a) with respect to z gives

$$-vB = D\frac{dB}{dz} - \chi\left(\frac{B}{C}\frac{dC}{dz}\right) + K_1,$$

where K_1 is the constant of integration. As $z \to \infty$, $B \to 0$ and hence $\frac{dB}{dz} \to 0$, and $C \to 1$, so $\frac{dC}{dz} \to 0$. Thus $K_1 = 0$. After a little re-arrangement, we have

$$B' = \frac{B}{D}\left(\chi\frac{C'}{C} - v\right), \qquad C' = \frac{kB}{v},$$

where we have used primes to indicate differentiation with respect to z. Note that we can now re-write the first of these equations again, in the form

$$\frac{d}{dz}\log B = \frac{\chi}{D}\frac{d}{dz}\log C - \frac{v}{D},$$

which can be integrated to obtain

$$\log B = \frac{\chi}{D}\log C - \frac{vz}{D} + K_2, \tag{6.8}$$

where K_2 is a constant of integration.

For the special case $\chi = 2D$, we can solve the system analytically as follows. We rewrite equation (6.8) in the form

$$B = K_3 C^2 e^{\frac{-v}{D}z},$$

where K_3 is an arbitary constant. We then substitute for B in equation (6.7b) and integrate to obtain

$$-\frac{1}{C} = \frac{kK_3}{v}\left(-\frac{D}{v}e^{\frac{-v}{D}z} + K_4\right).$$

Using the boundary conditions as $z \to \infty$ implies $K_4 = -v/(kK_3)$. Hence the solution is

$$B = \frac{Kv^2}{Dk}\left(\frac{e^{\frac{-v}{D}z}}{(1 + Ke^{\frac{-v}{D}z})^2}\right), \qquad C = \frac{1}{(1 + Ke^{\frac{-v}{D}z})},$$

where $K = kDK_3/v^2$. Note that K is an arbitrary positive constant equivalent to a linear translation; it may be set to 1.

Hence we see that biologically, the bacteria form a travelling band, which moves into 'new territory' where the nutrient is abundant, consumes it all, and then continues moving on.

7 Turing patterns

Alan Turing (1912-1954) was one of the most important scientists of the 20^{th} century. A mathematician by training, he laid the theoretical basis for much of modern computer science and artificial intelligence. During the Second World War, he worked at Bletchley Park, the British code-breaking centre, where he was instrumental in cracking the German enigma codes. It was during his time at Bletchley that he was involved in developing electro-mechanical devices to carry out the calculations involved in code breaking. After the war, he continued this work leading to the development of stored-programme computers. His name is probably most familiar now from the Turing test - the test of a machine's ability to exhibit intelligent behaviour.

In 1952 Turing published a seminal paper on mathematical biology - 'The Chemical Basis of Morphogenesis' [10]. This gave a theoretical explanation for the way patterns can emerge in living organisms - *e.g.* the patterns of spots or stripes on the skin of animals - through the interaction of two chemical **morphogens**. This was an enormous breakthrough biologically, as it suggested a mechanism by which neighbouring cells might be programmed to do different things, essential in the development of organs and tissues. It was revolutionary in theoretical terms too, as the mechanism that produces the patterns relies on diffusion, which usually evens out variations in chemical concentrations. We investigate this mechanism in the following section.

7.1 Motivation: Pattern formation in morphogenesis

Morphogenesis is the development of the form and structure of an organism during its life. Exactly how this is accomplished is still not fully understood, although both chemical and mechanical effects are known to contribute. One idea is that spatially inhomogeneous concentration profiles of various chemicals, called **morphogens** are set up, under the control of genes. These chemical concentrations then influence the behaviour of the cells - *e.g.* their proliferation rate, differentiation, *etc.* , which results in the characteristic form of the tissue. Turing considered the initial stages of morphogenesis, by developing a models for the interaction of two chemical morphogens, A and B, which diffuse and react with each other, within a domain $\mathcal{D}$.

7.2 Governing equations

The governing equations of the system are

$$\frac{\partial A}{\partial t} = D_A \nabla^2 A + F(A,B), \tag{7.1a}$$

$$\frac{\partial B}{\partial t} = D_B \nabla^2 B + G(A,B), \tag{7.1b}$$

where the reaction terms F and G are nonlinear. We impose **no flux** conditions on the boundary of the domain - *i.e.*

$$\hat{\boldsymbol{n}} \cdot \nabla A = \hat{\boldsymbol{n}} \cdot \nabla B = 0, \qquad \text{on } \partial\mathcal{D}, \tag{7.2}$$

and let $A(x,0) = A_i(x)$, $B(x,0) = B_i(x)$.

For the sake of simplicity, we shall only consider a one-dimensional domain, although the ideas can easily be generalised to higher dimensions. We nondimensionalise the system (7.1) to yield

$$\frac{\partial a}{\partial t} = \frac{\partial^2 a}{\partial x^2} + \gamma f(a,b), \tag{7.3a}$$

$$\frac{\partial b}{\partial t} = d\frac{\partial^2 b}{\partial x^2} + \gamma g(a,b). \tag{7.3b}$$

The above is to be solved on $0 < x < L$, with no flux boundary conditions. Full details of the scalings are given in [5]; for now it is sufficient to note that γ is a parameter related to the domain length, and $d = D_B/D_A$.

We assume that the system (7.3) has a spatially uniform steady state $(a,b) = (a_0, b_0)$ where a_0, $b_0 > 0$. We are interested in situations where the instability of the system is driven by diffusion; hence, we require that in the absence of diffusion (a_0, b_0) is linearly stable. We therefore begin by determining the conditions under which this is the case.

7.3 Linear stability analysis in the absence of diffusion

If there is no diffusion, the system (7.3) reduces to

$$\frac{\partial a}{\partial t} = \gamma f(a,b), \tag{7.4a}$$

$$\frac{\partial b}{\partial t} = \gamma g(a,b). \tag{7.4b}$$

We now consider small perturbations to the steady state, so that

$$a = a_0 + \epsilon a_1 + O(\epsilon^2), \qquad b = b_0 + \epsilon b_1 + O(\epsilon^2), \tag{7.5}$$

where $\epsilon \ll 1$ is the typical amplitude of the perturbation. Substituting this into equation (7.4), at leading order we obtain

$$\frac{\partial}{\partial t}\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \gamma \begin{pmatrix} f_a & f_b \\ g_a & g_b \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \tag{7.6}$$

where $f_a = \left.\dfrac{\partial f}{\partial a}\right|_{(a_0,b_0)}$, *etc.*.

For convenience, we set

$$\boldsymbol{w} = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}, \qquad M = \begin{pmatrix} f_a & f_b \\ g_a & g_b \end{pmatrix}$$

We then note that (7.6) can be written as

$$\frac{\partial \boldsymbol{w}}{\partial t} = \gamma M\, \boldsymbol{w}.$$

From earlier in the course, we know the stability of the steady state (a_0, b_0) just depends on the eigenvalues, λ, of the matrix γM (which is the Jacobian matrix for the diffusion-free system) evaluated at (a_0, b_0). Thus linear stability, $Re\{\lambda\} < 0$, is guaranteed provided

$$(f_a + g_b) = \text{tr}\, M < 0, \qquad (f_a g_b - f_b g_a) = \det M > 0. \tag{7.7}$$

7.4 Determining conditions for diffusion-driven instability

We now return to considering the full system from equation (7.3). Once again, we consider small perturbations to the spatially-homogeneous steady state (a_0, b_0). Substituting (7.5) into (7.3) yields, at leading order

$$\frac{\partial \boldsymbol{w}}{\partial t} = D\,\frac{\partial^2 \boldsymbol{w}}{\partial x^2} + \gamma M\,\boldsymbol{w}, \qquad \text{where} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}. \tag{7.8}$$

From our earlier studies of the diffusion equation, we anticipate the solutions will be of the form $\boldsymbol{w} \propto e^{\lambda t}\cos qx$ (as in the Keller-Segel model). Note that the no flux boundary conditions at $x = 0$ have again led us to reject the possibility of a $\sin qx$ term. In order to satisfy the boundary condition at $x = L$, we require $q = n\pi/L$, where n is an integer.

Since the problem (7.8) is linear, we can write the solution as

$$\boldsymbol{w} = \sum_n C_n e^{\lambda_n t}\cos q_n x = \sum_n \boldsymbol{w}_n,$$

where $q_n = n\pi/L$ and the constants C_n are determined by the initial conditions. Substituting this into equation (7.8) yields

$$\lambda_n \boldsymbol{w}_n = -q_n^2 D\,\boldsymbol{w}_n + \gamma M\,\boldsymbol{w}_n. \tag{7.9}$$

For nontrivial solutions $\boldsymbol{w}_k$, we require

$$\det(\lambda_n I - \gamma M + q_n^2 D) = 0. \tag{7.10}$$

Henceforth, we shall drop the subscript n, and take it as read that λ is the growth rate corresponding to a particular value of q. Equation (7.10) can be re-written as

$$\lambda^2 + \lambda[q^2(1+d) - \gamma(f_a + g_b)] + h(q^2) = 0, \tag{7.11}$$

where

$$h(q^2) = dq^4 - \gamma(df_a + g_b)q^2 + \gamma^2 \det M. \tag{7.12}$$

The steady state (a_0, b_0) will be linearly stable if both roots of (7.11) have $Re\{\lambda\} < 0$. We have already imposed the constraint that in the absence of spatial effects, the steady state is stable - *i.e.* for $q = 0$, the corresponding growth rates obey $Re\{\lambda\} < 0$. (The quadratic in that case is $\det(\lambda I - \gamma M) = 0$, and the condition $Re\{\lambda\} < 0$ gives conditions (7.7).) For the steady state to be unstable to spatially varying disturbances, we require $Re\{\lambda(q)\} > 0$ for some $q \neq 0$. Recall from the quadratic formula that the sum of the roots of (7.11) is given by minus the coefficient of λ and the product of the roots is $h(q^2)$. Hence, instability can occur if **either**:

1. The coefficient of λ in equation (7.11) is negative, *or*

2. $h(q^2) < 0$.

However, from equation (7.7), we have $f_a + g_b < 0$, and so

$$q^2(1+d) - \gamma(f_a + g_b) > 0, \tag{7.13}$$

since d, $\gamma > 0$. This rules out the first possibility. Hence instability is only possible if $h(q^2) < 0$ for some q. But, from equation (7.7), we have $\det M > 0$. Hence, it is only possible for $h(q^2) < 0$ is if $df_a + g_b > 0$. Since $f_a + g_b < 0$ from (7.7), this implies we require f_a and g_b to be of opposite sign, and $d \neq 0$. Thus the conditions for a diffusion driven instability are those of (7.7) plus

$$df_a + g_b > 0, \qquad \Rightarrow \quad d \neq 1. \tag{7.14}$$

In fact, the inequality in (7.14) is a **necessary condition** for instability, but it is not **sufficient** (if $df_a + g_b$ is positive, but very small, $h(q^2)$ could still be positive). For $h(q^2)$ to be negative for some $q \neq 0$, then clearly its minimum value, h_{min} must be negative. Let q_c be the value of q such that $h(q_c^2) = h_{min}$. At a minimum of h, we have

$$\frac{dh}{dq^2} = 0 \qquad \Rightarrow \qquad q_c^2 = \frac{\gamma(df_a + g_b)}{2d}.$$

Substituting q_c into (7.12) gives

$$h_{min} = \gamma^2 \left[\det M - \frac{(df_a + g_b)^2}{4d} \right].$$

Hence for instability we require

$$\det M < \frac{(df_a + g_b)^2}{4d}.$$

Note however, that this condition is still not sufficient for instability, for reasons which will shortly become clear.

The bifurcation from stable to unstable behaviour occurs when $h_{min} = 0$. Taking the kinetics functions f and g to be given, this defines a critical diffusion ratio, d_c, which is the root of $h_{min} = 0$ - *i.e.*

$$d_c^2 f_a^2 + 2(2f_b g_a - f_a g_b) d_c + g_b^2 = 0.$$

We can sketch the dispersion relation, based on this information (since only q^2 appears in the equation, we can restrict attention to $q \geq 0$). For $d < d_c$, there are no unstable wavenumbers, whilst for d_c, there is one neutrally-stable wavenumber, $q = q_c$. For $d > d_c$, we note that, as $q \to \infty$, $h \to dq^4 > 0$ by equation (7.12), and so instability can only occur for wavenumbers in a certain range $q_1^2 < q^2 < q_2^2$ (where q_1^2 and q_2^2 are the two roots of $h(q^2) = 0$). However, at this point it is important to recall the fact that we required $q = n\pi/L$ ($n = 1, 2, \ldots$) in order to satisfy the boundary condition

at $x = 1$. Hence if instability is to occur, there must be at least one multiple of π/L between q_1 and q_2; this is more likely to be the case if L is large.

Now, since for stable wavenumbers $\lambda(q^2) < 0$, these will decay exponentially fast. Hence, after a short time, the solution will be dominated by the unstable wavenumbers (assuming some exist) - *i.e.*

$$(a_1, b_1) \sim \sum_{q_1}^{q_2} (\alpha_q, \beta_q) \exp\left(\lambda(q^2)t\right) \cos qx$$

for some constants α_q and β_q. Note that, as for the case of the Keller-Segel model, we hence expect the pattern to be dominated by the fastest-growing mode (the value of q which gives the largest possible value of λ).

The necessary conditions for diffusion driven instability are thus

$$f_a + g_b < 0, \tag{7.15a}$$

$$f_a g_b - f_b g_a > 0, \tag{7.15b}$$

$$d f_a + g_b > 0, \tag{7.15c}$$

$$(d f_a + g_b)^2 - 4d(f_a g_b - f_b g_a) > 0, \tag{7.15d}$$

where all derivatives are evaluated at (a_0, b_0).

7.4.1 Physical explanation

- By condition (7.15a) at least one of f_a or g_b is negative. For the sake of argument assume it is g_b. Then b inhibits its own rate of formation; we call it an **inhibitor**.
- By condition (7.15c), if g_b is negative, f_a must be positive. Hence a promotes its own rate of formation; we call this species an **activator**.
- The two previous steps imply $f_a g_b < 0$.
- Hence condition (7.15b) can only be true if $f_b g_a < 0$. This means one of f_b and g_a must be negative, and the other positive. We have two posibilities:
  - Activator-inhibitor
  $$f_b < 0, \quad g_a > 0, \quad \begin{pmatrix} + & - \\ + & - \end{pmatrix}.$$
  - Positive feedback
  $$f_b > 0, \quad g_a < 0, \quad \begin{pmatrix} + & + \\ - & - \end{pmatrix}.$$
- Comparing (7.15a) and (7.15c) we see we require $d \neq 1$ (the diffusion coefficients must not be the same). In fact, we must have $d > 1$ in order to satisfy (7.15c), since we have $f_a > 0$ and $g_b < 0$. Hence the inhibitor diffuses faster than the activator.

Now if we consider the activator-inhibitor case, we can build up the following picture of the pattern formation mechanism. Suppose, as a result of random perturbations, a small peak in the activator concentration, a, occurs at some location. This causes a locally enhanced rate of inhibitor production, which, were it not for diffusion, would halt and reverse any increase in a. However, the inhibitor, b, diffuses away more quickly than a, so it cannot control the local activator production, and the peak grows. However, since near the peak there is increased inhibitor production as well (with the inhibitor diffusing away), in the region close to the peak, there will be enough inhibitor to prevent any more peaks of a forming. This is how the fastest growing wavenumber of the pattern is determined.

7.4.2 How the leopard got his spots (but not on the tail)

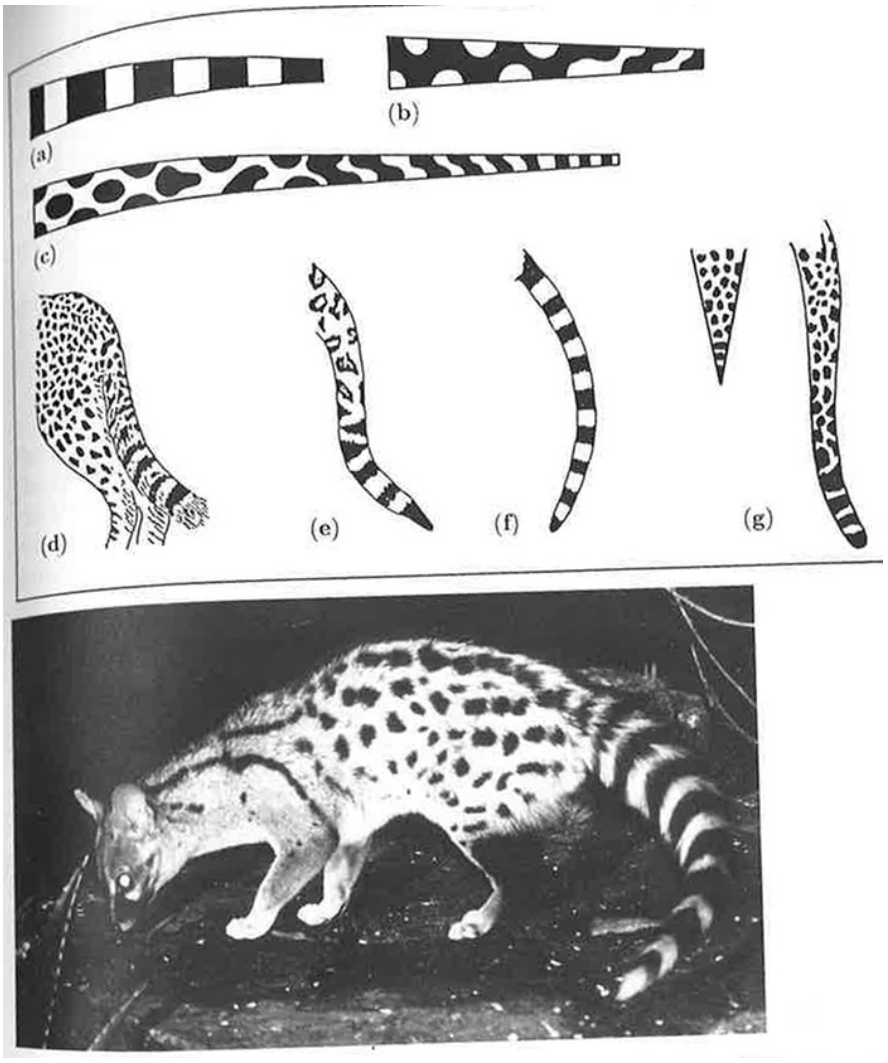


Figure 15: Examples of tail patterns predicted using Turing's theory (a-c) compared with real animal tails (d-g) (from [5])

One interesting consequence of the model is a hypothetical explanation for the observation that several spotted animals, such as the leopard, have a striped tail. However, no

striped animal has a spotted tail. The idea here is that if coat patterns are determined by a Turing-type mechanism, then there will be a fastest growing wavenumber which will determine the lengthscale of the coat pattern. On the body of the animal, where the dimensions of the skin are roughly the same in both directions, this wavenumber can 'fit' in the domain in both directions. The spots occur when peaks coincide. However, on a narrow region like the tail, there is not room enough to fit a whole wavelength in the narrow direction. Hence, we see variations in pattern only along the longitudinal direction - stripes.

8 Growing tissues

8.1 Modelling avascular tumours

Unfortunately, cancer is one of the leading causes of death in the Western world. Intensive research is going on worldwide to develop new and better treatments, and increasingly, researchers are turning to mathematical models in order to understand the complex web of interactions between tumour cells, other cell populations, growth factors, the extracellular matrix, *etc.* , and how these factors influence the progression of the disease. In this section, we will investigate a simple model for the early stages of tumour growth, which illustrates how the typical structure of a thin rim of proliferating cells surrounding a dead core arises, and allows us to calculate the tumour's growth rate.

In order to grow, tumour cells require oxygen and other nutrients. Most tissues have blood vessels running through them, which supply these nutrients and take away waste products. However, in the early stages of growth, the tumour has not yet acquired its own blood supply, and is termed **avascular**. At this stage, nutrients have to diffuse to the tumour cells from nearby normal tissues. As the tumour grows larger, at some point, the supply of nutrients by diffusion becomes insufficient to keep up with the demand of the proliferating cells; the oxygen concentration at the centre of the tumour falls below a critical level, and the cells there start to die. As the tumour continues to grow, this region of dead cells, called the **necrotic core**, also becomes larger. The debris from the dead cells breaks down, so at some point the tumour reaches a diffusion-limited size, in which the rate of cell proliferation at the tumour's edge exactly balances the rate of cell death and breakdown in the core. The tumour can then grow no more until has become **vascularised** (*i.e.* gained its own system of blood vessels).

8.1.1 Mathematical modelling of nutrient transport

For simplicity, we will consider a spherical tumour, with the radial coordinate denoted by r. We let the radius of the necrotic core (when it exists) by a, and the radius of the tumour be R; in general, these will both be functions of time, t. We consider transport of a single critical nutrient (*e.g.* oxygen) within the tumour, and let its concentration be denoted by c. Live tumour cells will consume oxygen, whilst the dead cells in the necrotic core will not, so we will have different equations in the two regions. We assume that oxygen diffusion through the tumour occurs on a much shorter timescale that the timescale of tumour growth (days). Since we are primarily interested in what happens on the latter timescale, we can make the quasi-steady assumption in our equations for oxygen diffusion. Thus our model equations are:

$$D\nabla^2 c = D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c}{\partial r}\right) = 0 \quad \text{for } 0 < r < a, \tag{8.1a}$$

$$D\nabla^2 c - k = D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c}{\partial r}\right) - k = 0 \quad \text{for } a < r < R, \tag{8.1b}$$

where k is the (constant) rate of oxygen consumption by the cells. We assume that in normal tissue the oxygen concentration is c^*, and so impose the boundary condition $c = c^*$ at $r = R$. However, if the oxygen concentration drops below some critical level, c_n, then the cells become necrotic (die). Hence we define a to be the value of r which satisfies $c(a) = c_n$.

Note that a will thus depend on the oxygen concentration, and so has to be found as part of the solution. This type of problem, where the position of the boundaries depends upon the solution, is called a **free boundary problem**. They are often quite complicated to solve (compared to the more familiar type of problems where the boundaries are fixed). However, to begin with we will avoid this difficulty by considering the earliest stages of growth, when the tumour is still too small to have a necrotic core.

Solution for a non-necrotic tumour

In this case, $a = 0$, and so our model reduces to

$$D\nabla^2 c - k = D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c}{\partial r}\right) - k = 0, \tag{8.2}$$

subject to $c(R) = c^*$.

Integrating (8.2) yields the general solution

$$c(r) = \frac{k}{6D}r^2 + \frac{A}{r} + B,$$

where A and B are constants to be determined. Since we require the oxygen concentration at $r = 0$ to be bounded, we set $A = 0$. Then, imposing $c(R) = c^*$ gives

$$c(r) = c^* - \frac{k}{6D}(R^2 - r^2).$$

Note that this expression for the oxygen concentration is valid provided the minimum value of c (which will occur at $r = 0$) is greater than c_n - *i.e.*

$$c(0) = c^* - \frac{kR^2}{6D} > c_n.$$

This implies that R must be less than some critical value R_c, given by

$$R^2 < R_c^2 = \frac{6(c^* - c_n)D}{k}.$$

Solution for a necrotic tumour

Now let us turn to the case where $R > R_c$, so we need to solve both of equations (8.1), with a to be determined. At the boundary of the necrotic and living regions, we assume that the concentration and the flux $-D\dfrac{\partial c}{\partial r}$, must be continuous.

Solving equation and imposing $c = c_n$ at $r = a$ together with boundedness of the solution at $r = 0$, we see that, in fact, the oxygen concentration is constant in the necrotic region:

$$c = c_n \quad \text{for } 0 < r < a.$$

Hence the flux at $r = a$ is zero. Solving equation in the live region gives (as before)

$$c(r) = \frac{k}{6D}r^2 + \frac{A}{r} + B,$$

where the three unknowns, A, B and a are found by imposing

$$c(a) = c_n, \qquad \left.\frac{\partial c}{\partial r}\right|_{r=a} = 0, \qquad c(R) = c^*.$$

Doing so yields

$$c_n = \frac{ka^2}{6D} + \frac{A}{a} + B, \qquad c^* = \frac{kR^2}{6D} + \frac{A}{R} + B, \qquad \frac{ka}{3D} - \frac{A}{a^2} = 0.$$

From the third equation above, we find $A = \frac{ka^3}{3D}$, and on subtracting the first equation from the second and substituting for A we obtain

$$\begin{aligned} c^* - c_n &= \frac{k}{6D}\left[R^2 - a^2 - 2a^3\left(\frac{1}{a} - \frac{1}{R}\right)\right] \\ &= \frac{kR^2}{6D}\left(1 + \frac{2a}{R}\right)\left(1 - \frac{a}{R}\right)^2 = \frac{k}{6D}\left(1 + \frac{2a}{R}\right)(R-a)^2, \end{aligned} \tag{8.3}$$

from which a can be found in terms of the tumour radius, R. In the limit $R \to \infty$, the second version of the equation implies $a/R \to 1$, and so, from the third, $R - a \to h$, a constant, where

$$h^2 = \frac{2D(c^* - c_n)}{k}.$$

Hence, in a large tumour there is a shell of proliferating cells at the edge of the tumour, the thickness of which depends upon the excess oxygen concentration over the critical level, and the rates of oxygen consumption and diffusion, but not on the tumour's size itself.

Note that, in the above we take R as a parameter (*i.e.* if we know the tumour's size from observations, our model tells us the size of the necrotic region); however, we have not yet considered how the tumour develops.

8.1.2 Modelling tumour growth

In the preceding section, we have solved for the nutrient concentration in a tumour of known size, R, and shown how the radius of the necrotic core, a relates to this size. However, tumours are growing tissues, so we we now consider how the tumour radius will evolve in time. Hence, in the following section we develop a model which allows us to determine $a(t)$ and $R(t)$.

Let n be the tumour cell density, and let v be the radial velocity of the cells. Within the necrotic core, there is no cell proliferation, but the dead cells break down at some rate, β, which we shall assume to be constant. In the region $a < r < R$, the cells are proliferating; in general, the rate of proliferation will depend on the concentration of oxygen and other nutrients; however, for simplicity, here we shall assume that, provided $c > c_n$, the cells proliferate at a constant rate, γ. Conservation of cells then gives

$$\frac{\partial n}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 n v\right) = -\beta n \quad \text{in } 0 < r < a, \tag{8.4a}$$

$$\frac{\partial n}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 n v\right) = \gamma n \quad \text{in } a < r < R. \tag{8.4b}$$

Now let us assume that the cell density in a tumour is constant; without loss of generality, we can set this constant to be unity. Then our equations above reduce to

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 n v\right) = -\beta \quad \text{in } 0 < r < a,$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 n v\right) = \gamma \quad \text{in } a < r < R.$$

Upon integrating and apply continuity of the velocity at $r = a(t)$ we have

$$v = -\frac{1}{3}\beta r \quad \text{in } 0 < r < a,$$

$$v = \frac{1}{3}\gamma r - \frac{1}{3}(\beta + \gamma)\frac{a^3}{r^2} \quad \text{in } a < r < R.$$

In order to solve for $R(t)$ we need one more equation. We note that the cells on the tumour's boundary are moving at the velocity of expansion of the tumour, so

$$\frac{dR}{dt} = v(R(t), t).$$

Thus, using our solution for v, we find

$$\frac{dR}{dt} = \frac{1}{3}\gamma R\left[1 - \frac{\beta + \gamma}{\gamma}\left(\frac{a^3}{R^3}\right)\right]. \tag{8.5}$$

In principle, we can now substitute for a in terms of R, using (8.3) and integrate the resulting ODE to obtain $R(t)$; however, in practice, the resulting equation is rather complicated, and we would need to resort to numerical methods.

However, it is still possible to extract some useful information from this equation analytically. Usually, in a tumour the rate of cell death, β will be much smaller than the rate of cell proliferation, γ, and hence we anticipate that the tumour will grow large, and therefore, from our earlier results, that h/R will be small, (where $h = R - a$) - *i.e.* the proliferating rim of cells will be very thin.

From equation (8.5) we see that, as $t \to \infty$, R will tend to a steady state value such that

$$\frac{\beta + \gamma}{\gamma}\left(\frac{a}{R}\right)^3 = 1 \quad \Rightarrow \quad \left(1 - \frac{h}{R}\right)^3 = \frac{1}{1 + \epsilon},$$

where $\epsilon = \beta/\gamma \ll 1$. We now expand as follows

$$\left(1 - \frac{h}{R}\right)^3 = 1 - 3\frac{h}{R} + O\left(\frac{h^2}{R^2}\right), \qquad \frac{1}{1+\epsilon} = 1 - \epsilon + O(\epsilon^2).$$

Hence, balancing the leading order terms we have

$$3\frac{h}{R} = \epsilon, \quad \Rightarrow \quad R = 3\frac{h\gamma}{\beta}.$$

Hence the tumour cannot grow beyond a certain size whilst its nutrient supply is still diffusion limited.

References

[1] NJL Brown, AD Sokal, and HL Friedman. The complex dynamics of wishful thinking: The critical positivity ratio. *Am. Psychologist*, 68(9):801813, 2013.

[2] L. Edelstein-Keshet. *Mathematical models in biology.* SIAM, 2005.

[3] K. J. Himmelstein and K. B. Bischoff. Mathematical representations of cancer chemotherapy effects. *Journal of Pharmacokinetics and Pharmacodynamics*, 1:51–68, 1973.

[4] P. K. Maini, D. L. S. McElwain, and D. Leavesley. Travelling waves in a wound healing assay. *Appl. Math. Lett.*, 17:575–580, 2004.

[5] J. D. Murray. *Mathematical Biology.* Springer-Verlag, New York, 2nd edition, 1993.

[6] JD Murray. *Mathematical Biology: I. An Introduction.* Springer, 2003.

[7] L. A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bulletin of Mathematical Biology*, 50:579–593, 1988.

[8] G. I. Taylor. The formation of a blast wave by a very intense explosion. ii. the atomic explosion of 1945. *Proc. R. Soc. (Lond.) A*, 201:175–186, 1950.

[9] D. W. Thompson. *On Growth and Form.* Cambridge University Press, 2nd edition, 1942.

[10] A. M. Turing. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. B*, 237:37–72, 1952.