

Introduction to Mathematical Biology

2008

lecture notes & exercises

Hugo van den Berg

m a t h e m a t i c s i n s t i t u t e
u n i v e r s i t y o f w a r w i c k

Aims These lecture notes support a 15-lecture course for second year students in applied mathematics. The three aims of the course are:

1. To introduce some basic modelling methodology in a practical manner, based on a series of simple case studies.
2. To give a taste of applied mathematics in the life sciences: to be able to understand and critically appreciate the primary literature in this field.
3. To rehearse and reinforce basic material on ordinary differential equations in an applied context.

Learning objectives At the end of the course, you should at least be able:

- to translate a simple biological problem into a mathematical model;
- to give a biological interpretation for simple systems of ODEs and their behaviour;
- to distinguish dimensions from units, and determine the dimensions of quantities;
- to non-dimensionalize a model;
- to assign the correct scale of measurement to data;
- to formulate an informed opinion on mathematical models in the life science;
- to apply the methods of Bernoulli and separation of variables in solving ODEs;
- to solve small linear systems of ODEs with constant coefficients;
- to define those terms which appear printed in **bold type**

Contents

I	Mathematical models in the life sciences	3
1	Getting started	4
1.1	What is a mathematical model?	4
1.2	Why use differential equations?	5
2	A modelling recipe	9
3	Case studies	11
3.1	Alcohol metabolism	11
3.2	Mineral dynamics of a mussel in a tidal zone	13
3.3	The salt balance in a pond snail	15
3.3.1	Eco-physiological background	15
3.3.2	Formulating the model: first attempt	16
3.3.3	Formulating the model: second attempt	17
3.3.4	Is there a ‘set point’?	17
3.4	Fluxes and flows between compartments	18
3.4.1	Blood volumes in a pond snail	18
3.4.2	Circulation of lymphocytes	21
3.5	Metabolite homeostasis	23
3.5.1	The scene of the action	23
3.5.2	Kinetics of a metabolite and an enzyme	24
3.5.3	The equilibrium	25
3.5.4	Stability of the equilibrium	26
3.5.5	The control-system perspective	26
3.5.6	Suggestions for further work	26
3.6	Basic models of population growth	27
3.6.1	Malthusian growth	28
3.6.2	Limits on growth	28
3.6.3	Interference: logistic growth and variations	29
3.6.4	Positive interference	30
3.6.5	Harvesting	31
3.7	Butterflies, again	32

II Exercises	37
4 Dynamics: state transition functions	38
5 Dynamics in continuous time; one-dimensional systems	39
6 Phase plane analysis	45
7 Dynamics in discrete time	51
8 Modelling methodology	54
9 Empirical dimensions and units	58
10 Scales of measurement	64
11 Describing data with models	67
III Past exam papers	71

|

Mathematical models in the life sciences

1

Getting started

In after years I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics, for men thus endowed seem to have an extra sense—Charles Darwin

1.1 What is a mathematical model?

The concept of modelling is familiar from fashion supermodels or scale models of airplanes in a windtunnel. Perhaps less well known is the example of ‘animal models’ in which a disease is induced and studied in animals based on the presumed similarity between the experimental disease and a human disease. In all these instances, the model¹ is an idealized representation or example that provides a benchmark for thinking about a real-life phenomenon. Mathematical models serve exactly the same purpose. To make the most of modelling, we need to be clear about models, their structure, and the role they play in scientific investigations.

Our working definition of **mathematical modelling** will be:

Mathematical modelling is the description of an experimentally delineated phenomenon by means of mathematics, with a view to capturing the salient aspects of the phenomenon at hand.

This definition is rather vague: Would any mathematical description of observations count as a model? To give a trite example of what might count as a mathematical model, we could just list and report what we observed there and then. This often involves mathematical structures, notably measurements: these are numbers with units attached to them (see chapter 9). One might regard the data thus reported as a mathematical model of the phenomena observed, on the basis of the—admittedly perverse—argument that real-world phenomena have been converted into numbers, which are after all mathematical entities.

On the other hand, modern physics affords a superb example of mathematical structures used to represent real-world phenomena. Here a small set of powerful, yet simple, ‘first principles’ allows a wide range of structures and processes in the physical world to be described, sometimes with remarkable success and accuracy (although a considerable mathematical apparatus is needed to link the few simple principles with a plethora of complex phenomena).

The term ‘model’ is usually reserved for the situations somewhere in between these two extremes of the superficial description of a particular occurrence and the deep descriptive ‘first principles’ which aim to be as general as possible (general in the sense of pertaining to as many individual occurrences as possible). Thus, a model is more ambitious—aims for more generality—than the description of a particular occurrence, but is still restricted

¹The word ‘model’ ultimately derives from Latin *modulus*, ‘small measure.’

Figure 1.1: Surviving numbers of butterflies in the hothouse experiment.

to a modest range of similar situations. Models represent the halfway house where first principles are not available, or their application is not feasible, yet some general ideas about the phenomenon at hand are incorporated. More about mathematical models can be found in chapter 8.

1.2 Why use differential equations?

Most published mathematical models in the life sciences take the form of a system of ordinary differential equations. One might conclude that modellers are unimaginative and endowed with a strong herd instinct, but there is also a deeper reason why differential equations are viewed by many as the mathematical tool of choice when it comes to the description of processes.

Consider a nice conservatory full of lush plants and birds. We let loose a host of caterpillars. The animals eat and grow, and ultimately pupate. Adult butterflies emerge out of the chrysalises, fly around, mate, and are caught and eaten by the birds.

Suppose that we have some way of keeping track of each and every butterfly. There are a number of interesting things that can be done with the data thus obtained. We consider a simple example here: we simply record the time between emerging from the chrysalis and death for every butterfly. Furthermore, let us assume for a moment that all butterflies emerged from their chrysalises at the same time. We take this point in time to be our zero time. We plot the number of butterflies still alive as a function of time. Let N denote this number. Clearly, N at $t = 0$ is just the size of the cohort that emerges. Every time a butterfly dies, N is decreased by one. Looking at our plot, Figure 1.1, we notice that the time between subsequent deaths tends to become larger as the number of butterflies still alive dwindles.

How can we describe these data mathematically? An obvious approach is to look for some function f such that

$$N = f(t) .$$

Such a function is called a “state transition function” (see chapter 4). Actually, we do not

expect any particular data set to conform to a simple mathematical function; data sets almost invariably have a somewhat erratic quality. Note that we do not need to require that f maps to integers, even if the N in the data set can only take on integer values. The data set is just a representative of an infinite number of outcomes that the exact same experiment might have had (cf. section 3.7), and we are really interested in the average over all these (imaginary) data sets. One way to get a better picture of this average is of course to repeat the experiment a number of times.

A common notational overloading is to use the symbol N not only for the number of butterflies still alive, represented as a real variable, but for the *function* which relates this real variable to time as well:

$$N = N(t) . \tag{1.1}$$

Now, what function to choose for $N(t)$? A choice which many people seem to find reasonable is the straight line:

$$N(t) = N_0 - \alpha t .$$

One may vary the parameters N_0 and α to obtain a straight line that goes through the ‘cloud of data points’ in a satisfactory manner (for instance, by employing the least-squares method; see chapter 11).

But clearly, the straight line is not a satisfactory description of the data, in the following two respects: the straight line predicts negative numbers from $t = N_0/\alpha$ onwards, and the times between subsequent deaths tend to increase, which is not in keeping with a straight line. Rummaging through our drawer of elementary mathematical functions, we come up with the following alternative, which is more satisfactory:

$$N(t) = N_0 \exp\{-\lambda t\} \tag{1.2}$$

with parameters N_0 and λ . A second experiment provides a strong clue that equation (1.2) is the ‘right’ way to describe the data. In the second experiment, everything is the same, except the initial number, which is twice as large. Looking for the parameter values for which equation (1.2) describes the data best, we come up, unsurprisingly, with a N_0 which is twice the original value, and with a λ -value *that is exactly the same* as in the first experiment. Apparently, λ is capturing some essential property of the process (the butterflies’ proclivity for dying, or something like that) in a way that the straight line does not (for there we would find a best-fit value of α just shy of twice as large as for the original experiment).

Even though we now have a mathematical description that seems to tie in nicely with the data set, we still feel that there is something unsatisfactory about the procedure we followed. After all, this is a data set involving only two variables, and it looks pretty simple. How is our somewhat haphazard strategy of trying out arbitrary functions going to work out in more complicated cases? There has to be a better way.

Let us differentiate equation (1.2) with respect to time:

$$N'(t) = N_0 \exp\{-\lambda t\}(-\lambda) . \tag{1.3}$$

We can replace the factor $N_0 \exp\{-\lambda t\}$ with $N(t)$, by equation (1.2):

$$N'(t) = -\lambda N . \quad (1.4)$$

This is so simple that one can readily overlook how remarkable it really is. The differential equation tells us that the rate at which the cohort of butterflies is dying is proportional to the number still alive at this moment. But that makes perfect sense. If we assume that the butterflies do not influence one another's chances of dying, then all butterflies face an equal chance of dying in the next small increment of time. And the more butterflies there are, the more can be expected to die in the next small increment of time. The parameter λ expresses just this hazard rate per butterfly: the parameter thus admits of a natural interpretation, which is simply the mortality rate.

Starting from an argument along these lines, we could have written down differential equation (1.4) immediately, without guessing equation (1.2) first. This is a fruitful strategy to follow whenever we wish to find a process description. We collect our intuitions and knowledge about what is going on. This immediately leads to a consideration of the rates of change of the process variables of interest in terms of these process variables themselves. In other words: we naturally arrive at differential equations. The process description, equation (1.1), now is no longer the starting point, but rather the solution to a differential equation. Of course, the wish to have a mathematical description of the process, that is, a representation of a quantity of interest as a function of time remains the prime motivation, and, as such, the conceptual starting point. This is easy to forget as modellers usually get down to the business of formulating differential equations straightaway.

The second strategy—first write down differential equations, then solve them to get the desired process description—is superior in a number of related ways. Complicated interactions can be accommodated, which would not be easy in the ‘guess a function of time directly’ approach. Indeed, relevant differential equations or systems of differential equations may not have a solution in terms of elementary analytical functions (intractability, while a nuisance, should not play a role in judging their relevance as a model). The practical problem that an analytical solution cannot be found can often be surmounted by means of numerical approximation of the solution. Moreover, in many cases it is possible to extract interesting information from the study of the differential equations themselves, without bothering to solve them (see chapter 5).

There is one more reason why differential equations (chapters 5, 6) and difference equations (chapter 7) loom large as the mathematical tool of choice in process description. Suppose that, starting with an empty conservatory, we introduce butterflies at a constant rate Φ , and assume for a moment that these are immortal butterflies. Our differential equation then reads

$$N'(t) = \Phi . \quad (1.5)$$

and our solution (using $N(0) = 0$) is:

$$N(t) = \Phi t \quad (1.6)$$

as you can verify directly by differentiation.

Now suppose that we allow the butterflies to be mortal again, with mortality rate. On the original, naïve approach (“we need a process description, let’s find some function f and let N be f of t ”) we might attempt to combine the two processes (introduction of new butterflies Φt and decay by mortality $N_0 \exp\{\lambda t\}$) in some manner or other. We might try adding ($\Phi t + N_0 \exp\{\lambda t\}$) or multiplying ($\Phi t N_0 \exp\{\lambda t\}$), but we would be hard-pressed to motivate or justify our choice, and, anyway, a comparison with experimental data shows that all such attempts are wrong. However, if we try the idea of adding at the level of ordinary differential equations, we find something that does work:

$$N'(t) = \Phi - \lambda N. \quad (1.7)$$

The function $N(t)$ which satisfies this equation is

$$N(t) = \frac{\Phi}{\lambda} + \left(N_0 - \frac{\Phi}{\lambda}\right) \exp\{-\lambda t\} \quad (1.8)$$

as you should be able to verify by differentiation. Equation (1.8) predicts, in agreement with the data, that the number of butterflies ultimately (as $t \rightarrow \infty$) settles on a level $\bar{N} = \Phi/\lambda$ where the attrition rate $-\lambda\bar{N} = -\lambda\Phi/\lambda = -\Phi$ exactly balances the rate Φ at which fresh butterflies are being introduced.

For now, you can accept the moral of the story: “processes can be combined at the level of *rates*, and that is why differential equations are useful in process description.” If you feel that there is more to this, you are right, and you may be able to formulate a deeper reason after you have studied section 3.7.

2

A modelling recipe

This recipe assumes that you are going to formulate a mathematical model in the form of a system of differential equations or difference equations. While the common or garden variety of model does tend to assume this form, you should always first consider whether it is more appropriate to bring another mathematical technique¹ to bear on the problem at hand.

1. Delineate the phenomenon to be modelled. Outline and document the ‘scene of the action.’
2. Identify key quantities in the model.
3. Identify how these quantities interact, how they affect one another. A sketch with arrows showing transformations or causal effects is often helpful.
4. Decide on state variables, the process variables for which you are going to set up differential equations or difference equations.
5. Find appropriate formulations for the state transition functions (i.e. for differential equations, the rates of change of the state variables).
6. Analyse the model. Render the model dimensionless, if necessary. Perform numerical simulations, as required.
7. Relate the behaviour of the model to empirical data on the modelled phenomenon.

During the first five steps you will usually find that you need to adopt a number of assumptions. Try to write down all assumptions explicitly. If you get stuck during these first steps, it often helps to write down exactly what you know, or think you know, about the process to be modelled. In particular, it is always a good idea to jot down elementary, common sense observations. Conservation of matter is a strong guiding principle when you are considering processes in which particles move from one place to another. By imagining yourself in the position of a player in your model—be it a piece of DNA, a neuron, a blood vessel, an organism—you can bring into focus your intuitions about what they can and cannot ‘know’ or ‘sense.’

The model cycle During step 4 in the above procedure you have to extract the salient features of the scheme you set up in steps 1–3. This almost invariably entails some guesswork, since you have to assess the relative importance of various effects and factors. Fortunately, the outcome of steps 6 and 7 will often put you in a better position to judge the saliency

¹There are partial differential equations, integro-differential equations, stochastic differential equations, Markov chains & martingales, branching theory, combinatorics, graph theory, game theory...

Figure 2.1: A more elaborate flowchart outlining the model cycle.

of certain aspects of the problem, and prompt you to revise your earlier decisions. Thus you are led back from step 6 (or step 7) to step 4. You may find yourself going around any number of times in this way, in what is known as the **model cycle** (a.k.a. **empirical cycle**).

As is already apparent from the rather more elaborate version of the model cycle show in Figure 2.1, sound modelling practice requires that due attention be paid to “coherence”, consistency, dimensions, units, scales of measurement and comparison to data. These aspects are treated in more depths in chapters 8—11.

3

Case studies

If you want to apply mathematics, you must act as though the measured magnitudes have precise values. This fiction is very fruitful, as everybody knows; the fact that it is only a fiction does not diminish its value as long as we bear in mind that the precision of the result will be what it will be. To go, with the valid help of mathematics, from approximate premises to approximate conclusions, I must go by way of an exact algorithm, even though I consider it an artifice—Bruno de Finetti

3.1 Alcohol metabolism

A pint of beer contains a few grammes of alcohol (ethanol) which are taken up through the wall of your stomach into your blood. This aliquot of alcohol will not stay in your body forever, though. It is converted in the liver into acetaldehyde which may be further metabolized, or converted via acetate into acetyl-coenzyme A, which may in turn be converted into fat (whence the beer gut); Figure 3.1.

We wish to know how long the alcohol concentration of our blood remains above the legal limit for driving a car (which, in the UK, is a generous 80 mg/dl). As our process variable (or **state variable**, the usual term), we choose this concentration and denote it $A(t)$. We seek a differential equation of the form $\frac{d}{dt}A = \dots$ where in the place of the dots we must put an expression that tells us how fast the enzymes in the liver are breaking down the alcohol. It seems reasonable that such an enzyme will be busier if it meets up with an alcohol molecule more often, and, moreover, one would guess that the frequency with which an enzyme encounters an alcohol molecule is proportional to the alcohol concentration in the tissue where the enzymes reside. Applying the idea of rapid equilibration between blood and liver tissue we assume that the concentration as ‘seen’ by the enzymes is the same as, or at least proportional to the alcohol concentration in the blood. Putting these ideas together, we obtain the following differential equation:

$$\frac{d}{dt}A = -\lambda A \quad (3.1)$$

with a positive parameter λ . The solution of equation (3.1) is:

$$A(t) = A_0 \exp\{-\lambda t\} \quad (3.2)$$

where A_0 is the initial condition, the alcohol concentration just after the alcohol has entered the bloodstream. It can be found by dividing the amount of alcohol Q we drank by the volume V_B of our blood (which is about 10 pints). When the alcohol concentration in the blood is lower than that attained after drinking about a half-pint of beer, the half-life is found to be an hour (thus $\lambda = 0.7 \text{ h}^{-1}$).

Suppose we had a few too many and went to bed at midnight with a blood alcohol concentration of about 250 mg/dl. Would we be allowed to drive a car at 8 o’clock the

Figure 3.1: Major route of alcohol metabolism in the liver.

next morning? With a half life of one hour, our alcohol concentration should be lower than 1 mg/dl by that time, and we should be fit to drive. However, we find that we still feel very much intoxicated at 8 o'clock. Our driving skills (or lack thereof) prompt a police officer to stop us, and an illegal concentration of over 100 mg/dl is found in the subsequent test.

So what went wrong? To get a clue, we need to look at the operation cycle of the enzyme. It waits for some time until it encounters an alcohol molecule in just the right manner to 'catch' it. Presumably, the relative position and speeds of the enzyme and the alcohol molecules need to be just right for this to happen. It seems reasonable to assume that the average time spent waiting for such a catch, T_w , is inversely proportional to the alcohol concentration around the enzyme. Having caught an alcohol molecule, the enzyme must convert it. On average, it takes an enzyme T_a time units to accomplish this. We do not expect this quantity T_a to be dependent on the concentrations of the enzyme or the alcohol molecules, so we can take T_a to be a constant. (In fact, T_a may depend on factors such as the temperature and the concentration of the 'co-factor' required by the enzyme to operate, but we take all such factors to be constant as well.) The average operation time would thus be $T_w + T_a$, and by the law of large numbers, during a time period $\mathcal{T} \gg (T_w + T_a)$, we see that

$$\mathcal{N} = \frac{\mathcal{T}}{T_w + T_a}$$

alcohol molecules will be processed by a single enzyme molecule. But $T_w + T_a$ is much smaller than the time scale of our original problem, and therefore we may take \mathcal{T} to be on the order of the time scale of alcohol clearance from the body. Then

$$\psi = \frac{\mathcal{N}}{\mathcal{T}} = \frac{1}{T_w + T_a} \tag{3.3}$$

becomes a reasonable expression for the rate at which alcohol is cleared per enzyme molecule. When $T_a \ll T_w$, this rate per enzyme becomes $1/T_w$ which is proportional to the alcohol concentration, and we have the original first-order equation (3.1).

However, we know that the condition $T_a \ll T_w$ will not be satisfied when the alcohol

concentration is sufficiently large. Specifically, we have assumed that

$$T_w = \frac{\alpha}{A} \quad (3.4)$$

with some positive parameter α (which makes sense if waiting for a catch is a Poisson process). Thus, $T_a \ll T_w$ does not hold good when $A \sim \alpha/T_a$. In fact, when the alcohol concentration is much larger than α/T_a , the enzyme will spend nearly all of its time splitting alcohol molecules, and hardly any time waiting for the next one to arrive. In such circumstances, the enzyme is said to be saturated. Inserting equation (3.4) into equation (3.3) we get the following for the rate at which alcohol is cleared per enzyme molecule:

$$\psi = T_a^{-1} \frac{A}{\alpha/T_a + A}$$

where the parameter concentration α/T_a represents a concentration: it is the concentration at which ψ equals half its least upper bound T_a^{-1} . This concentration is usually called the saturation constant or affinity constant. To obtain $\frac{d}{dt}A$ we need to multiply ψ by Z , the concentration of enzyme in the liver cells:

$$\frac{d}{dt}A = -\psi Z = -\frac{Z}{T_a} \cdot \frac{A}{\alpha/T_a + A}. \quad (3.5)$$

Equation (3.5) can be solved by separation of variables (chapter 5). The solution reads

$$\frac{T_a}{Z} \left(\frac{Q}{V_B} - A(t) \right) + \frac{\alpha}{Z} \ln \left\{ \frac{Q/V_B}{A(t)} \right\} = t. \quad (3.6)$$

This equation is a bit odd: t is written explicitly in terms of $A(t)$, instead of the other way around. In fact, we cannot get an explicit expression for $A(t)$ here. However, plotting t as a function of $A(t)$ is a useful way of studying this solution. We see that there will be a linear decay phase if Q/V_B is in the saturating region, that is, if $Q/V_B \gg \alpha/T_a$. The linear phase is followed by an exponential decay phase, displayed to the right along the t -axis according to the linear term. There will be a final phase where $A \ll \alpha/T_a$: here the decay will be exponential with decay rate $\lambda = Z/\alpha$. This makes sense because a linear approximation to the right-hand side of equation (3.5) about $A = 0$ is good for small A , and thus the simple kinetics of equation (3.1) apply in this phase.

We can now appreciate what went wrong: at saturating concentrations, well above $\alpha/T_a = 2.5$ mg/dl, alcohol is broken down according to how many enzymes there are. The decay is only linear (at about 17 mg/(dl·h)) and not as fast as it would be according to the exponential decay rate Z/α of about 0.7 h^{-1} .

3.2 Mineral dynamics of a mussel in a tidal zone

Mussels live in estuaries. Their ambient (surrounding) medium is brackish water whose salinity increases as the tide comes in and the water in the estuary mixes with sea-water,

and whose salinity decreases again at low tide, when the water mixes with fresh water from the rivers that feed the estuary. Mussels need to expose quite a bit of skin to the ambient medium. In order to breathe, they must maintain a constant flow of water along their gills. These gills are slightly permeable to salt ions, which diffuse through the skin. When the salinity of the blood is lower than that of the ambient water, the net diffusive flow is inwards, and the blood salinity increases. The net diffusive flow is outwards when the salinity of the blood is higher, so that the blood salinity decreases. The following differential equation captures this diffusive exchange:

$$\frac{d}{dt}X = p(U(t) - X) \quad (3.7)$$

where X denotes the salinity of the blood and $U(t)$ is the salinity of the ambient water. The permeability of the skin is represented as a parameter p . If $U(t)$ were time-constant, $U(t) \equiv \bar{U}$, the solution of equation (3.7) would read,

$$X(t) = \bar{U} + (X(0) - \bar{U}) \exp\{-pt\} \quad (3.8)$$

where $X(0)$ is the salinity of the mussel's blood at $t = 0$. We see that $1/p$ is the characteristic time for the mussel's blood to acclimatize to the surrounding salinity. This time constant is also called the relaxation time, or the recovery time. (The choice depends on the circumstances; 'relaxation time' seems most appropriate here.)

When $U(t)$ is not constant in time, we can solve the equation, using Bernoulli's method (see chapter 5):

$$X(t) = \left(X(0) + \int_0^t pU(\tau) \exp\{p\tau\} d\tau \right) \exp\{-pt\}. \quad (3.9)$$

For instance, for the convenient example $U(t) = U_0 + (\Delta U) \sin \omega t$ we find

$$X(t) = U_0 + K \exp\{-pt\} + \frac{\Delta U}{\sqrt{1 + (\omega/p)^2}} \sin(\omega t - \tan^{-1}(\omega/p)) \quad (3.10)$$

where K depends on the initial condition. We actually want to know how strongly the mussel's blood salinity will follow the tidal movement after the initial transient corresponding to the term with K has died out, so we focus on the term with the sine in it. We see that the mussel is insensitive to external fluctuations when $p \ll \omega$; the small amplitude fluctuations lag behind by about $\pi/2$. For $p \gg \omega$, the internal fluctuations nearly equal the ambient fluctuations, with a vanishing phase shift.

The parameter p is actually a compound parameter. We can build it up from more fundamental parameters as follows. The salt flux is proportional to the area A exposed to the ambient, and the permeability per unit skin area P . To translate this in a term in the kinetics of the salinity X of the blood, we need to divide by the blood volume V . Thus,

$$p = \frac{A}{V}P$$

Figure 3.2: Left: the pond snail, *Lymnaea stagnalis*. Right: steady state blood salinity as a function of external salinity.

where P represents an intrinsic property of mussel gill skin, which we expect to be about the same for very small and very large mussels of the same species. If large mussels are a scaled-up version of small mussels, A will vary as the square of a linear scale factor, whereas V will vary as its cube; therefore p will be inversely proportional to the linear scale factor. Thus smaller animals will have a bigger p , and be more sensitive to external fluctuations in salinity.

3.3 The salt balance in a pond snail

This example again looks at the mineral balance of a mollusk. But this time, we focus on a constant ambient salinity, and we consider an animal that is able to regulate its internal salinity.

3.3.1 Eco-physiological background

The pond snail *Lymnaea stagnalis* is an air-breathing snail that lives in stagnant ponds. It can thrive in both brackish and freshwater environments. Its tissues and internal organs are unable to function in this wide range of salinities, however. When we investigate the salinity of the blood of animals acclimatized to various external salinities ranging from pure water to very brackish water, we find that these blood salinities change very little over a wide range of external conditions, as shown in Figure 3.2.

We know that the salinity of the water in which the animal lives affects the salinity of the blood because the skin is permeable to the various ions dissolved in the ambient water. Thus, diffusion tends to bring internal and external salt concentration to the same value. Because the external environment can be considered as an infinitely large compartment compared to the blood compartment of the snail, which is only about 1 or 2 milliliters, diffusive equilibration in practice means that the animal's blood salinity tends to conform to the ambient salinity.

We also know that there are molecules in the skin of the animal that are able to pump ions from the ambient environment into the blood. This mechanism only works one way: the skin transporters are not capable of taking up ions from the blood and excrete them in the external medium.

3.3.2 Formulating the model: first attempt

Our aim in the present modelling exercise is to find out how the passive processes of diffusion and the active process of salt uptake interact to result in the experimentally found acclimatization response. We only need to consider one state variable, which is the salinity of the blood X . Putting together the gain and loss terms, we find:

$$\frac{d}{dt}X = \Phi + p(U - X) \quad (3.11)$$

where Φ represents the active uptake of salt (sodium ions), the parameter p represents the permeability of the skin, and U represents the salinity of the ambient water. Since we are interested in the acclimatization response, we restrict ourselves to the special case where U is constant in time.

How to model Φ ? The skin cells that take up sodium ions respond to a hormone called ‘Sodium Influx Stimulating peptide’ which is excreted by certain nerve cells in the snail’s tiny brain. Presumably, these cells respond to the blood’s salinity. As the salinity of the blood decreases, these cells stimulate uptake of sodium ions. One way to express this dependency is as follows:

$$\Phi(X) = \Phi^\circ - \alpha X \quad (3.12)$$

that is, a simple straight line, with parameters Φ° and α . The latter parameter expresses how strongly the nerve cells respond to changes in blood salinity.

Substituting equation (3.12) into equation (3.11) we find:

$$\frac{d}{dt}X = \Phi^\circ - \alpha X + p(U - X) = \Phi^\circ + pU - (p + \alpha)X \quad (3.13)$$

which shows that the relaxation time $1/(p + \alpha)$ is composed of both the permeability and the response strength of the nerve cells.

What does the steady state response look like? Putting the left-hand side of equation (3.13) equal to zero we find:

$$\bar{X} = \frac{\Phi^\circ}{p + \alpha} + \frac{p}{p + \alpha}U. \quad (3.14)$$

This is a straight line with slope $p/(p + \alpha)$, which seems a fair model of the middle range in the experimental evidence, Figure 3.2.

3.3.3 Formulating the model: second attempt

For $U > \Phi^\circ/\alpha$ the steady state value \bar{X} predicted by our model lies below the ambient concentration, and Φ would have to be negative. However, we noted that the active uptake systems in the skin can only work one way. Thus, for $X > \Phi^\circ/\alpha$, equation (3.12) should be replaced by $\Phi(X) = 0$. The resulting steady state response for $U > \Phi^\circ/\alpha$ then reads $\bar{X} = U$, in accordance with the experimental evidence.

The other point where our model differs from the experimental response is at very low experimental salinities. Clearly, at near-zero salinities, the uptake molecules spend most of their time idle. Employing the same reasoning as in section 3.1, we are led to the following additional modification of equation (3.12):

$$\Phi(X) = \frac{U}{K+U} (\Phi^\circ - \alpha X) \quad X < \Phi^\circ/\alpha \quad (3.15)$$

where K is the saturation constant of the uptake molecules, which is in fact very small.

The middle region of salinities between $U \gg K$ and $U = \Phi^\circ/\alpha$ is called the ‘regulating’ region. The slope in this region is $p/(p + \alpha)$. Note that the ‘regulation’ becomes more perfect as α becomes larger; for $\alpha \rightarrow \infty$, the slope would be zero. The parameter α is sometimes called the ‘gain.’ Perfect regulation would imply an infinite gain, which is physically impossible.

3.3.4 Is there a ‘set point’?

It is found that the animals appear most healthy and active at a salinity somewhere in the middle of the regulation region. Let U^* denote this salinity. The steady state blood salinity at $U = U^*$ is given by:

$$X^* = \frac{\Phi^\circ + pU^*}{p + \alpha}$$

(since $U^* \gg K$, we ignore the complication represented in equation (3.15)). Now equation (3.12) can be rewritten,

$$\Phi(X) = p(X^* - U^*) + \alpha(X^* - X) \quad (3.16)$$

which can be interpreted as follows. The first term is the active uptake required at the ‘optimal’ point (U^*, X^*) . The second term can be viewed as an additional intervention when X deviates from the ‘desired’ value or ‘set point’ X^* . Some confusion surrounds set points of this kind in biological systems. We saw that we could set up the model without invoking the concept of a set point at all, and that there simply emerged a stable equilibrium value, in the form of a compound parameter $(\Phi^\circ + pU^*)/(p + \alpha)$. The language of set points leads some biologists to wonder where the set point could be found in the brain, to ask which physical structure actually harbours the set point. Finding no sensible answer (to what is indeed not a very sensible question in the first place), they then proceed to challenge the ‘reality’ of set points. But this overlooks the fact that the set point could

Figure 3.3: Blood circulation in the pond snail, *Lymnaea stagnalis*.

simply be a dynamic resultant as shown by equation (3.16). The debate about the ‘reality’ of a set point is moot. One can adopt a control systems point of view and express the apparent set point in terms of the underlying physiological parameters. But the latter are primary, and whenever someone has chosen to set up the model in terms of control systems—with desired set points and the like—we should always check whether the model can be recast in more neutral terms.

3.4 Fluxes and flows between compartments

3.4.1 Blood volumes in a pond snail

The pond snail *Lymnaea stagnalis* has no capillaries. Instead, the arteries are open-ended and give to a body cavity. The blood flows out of the arteries and engulfs the tissues, thus nourishing them and providing them with oxygen. The veins collect the blood from this blood space and conduct it to the two-chambered heart, which pumps it into the aorta, which is the principal artery that branches into various smaller arteries, which end in the blood space.

Figure 3.3 shows a simplified diagram of the circulation of the pond snail. It shows that there are two major blood spaces: the pedal blood space and the visceral blood space: the bit of the snail that sticks out of the shell when the snail is walking or swimming about contains the former, while the shell contains the latter. The pedal blood space contains the muscles of the foot, the brain, the esophagus, and the penis. The visceral blood space contains the gut, the reproductive glands, the kidney, lung and heart. The two spaces are separated by a valve which is normally closed.

The rate at which blood flows through these two blood spaces can be investigated by

introducing a small amount of a radioactive tracer in one of the two spaces (for practical reasons, one needs to select the visceral space for the injection of tracer). Small samples taken from the blood at various times post injection allow the concentration of tracer in the visceral space to be determined¹.

Let F_{CO} denote the flow of blood through the heart. Its dimensions are **volume·time**⁻¹. The quantity F_{CO} is often referred to as the ‘cardiac output.’ Let F_{ped} denote the flow of blood to the pedal sinus. Likewise, let F_{vis} denote the flow of blood to the visceral sinus. Clearly,

$$F_{\text{CO}} = F_{\text{ped}} + F_{\text{vis}} . \quad (3.17)$$

Furthermore, let V_{ped} denote the volume of the pedal blood space, and let V_{vis} be the volume of the visceral blood space. We assume that both V_{ped} and V_{vis} are constant. Hence we conclude that F_{ped} is not only the rate of blood flow from the heart to the pedal space, but also the flow rate of blood from the pedal space to the heart. Similarly, F_{vis} is the flow rate of blood from the visceral space to the heart.

An aliquot Q_0 of radioactive tracer is introduced into the visceral blood space at time 0. If we assume that the visceral blood space is well-mixed, we know the concentration of tracer in the visceral space:

$$C_{\text{vis}} = \frac{Q_{\text{vis}}}{V_{\text{vis}}}$$

where Q_{vis} is the amount of tracer in the visceral space and C_{vis} is its concentration. For the pedal space we assume a homogeneous concentration $C_{\text{ped}} = Q_{\text{ped}}/V_{\text{ped}}$. Of course this pedal tracer concentration equals zero at $t = 0$, but this concentration will increase as the tracer is pumped through the body.

What is the rate at which the tracer enters the heart? The reasoning followed to get at the answer to this question is the standard trick of compartmental analysis, so it’s worthwhile to follow the argument closely. First, note that during an interval ΔT a volume of blood equal to $F_{\text{vis}}\Delta T$ flows to the heart. It helps if one visualizes this as a little parcel of blood. It is the amount of blood received by the heart, every ΔT time units. We know the concentration of tracer in this parcel; it is C_{vis} . Multiplying this concentration by the blood volume we get a quantity of tracer:

$$C_{\text{vis}} \times F_{\text{vis}}\Delta T .$$

This, then, is the amount of tracer that flows to the heart during an interval of duration ΔT . The rate at which tracer is conducted from the visceral space to the heart is this amount divided by ΔT , which is $C_{\text{vis}}F_{\text{vis}}$. The dimensions of this quantity are **amount of tracer·time**⁻¹. Through similar reasoning, we find that the rate at which tracer is conducted from the pedal space to the heart is $C_{\text{ped}}F_{\text{ped}}$.

¹We will pretend that the tracer does not notably decay within the time frame of the experiment. If this assumption is not warranted, it is a simple matter to correct all measurements by multiplying them with $e^{t/\tau}$, with t the time since injection and τ the characteristic time constant of the tracer’s radioactive decay. Notice, however that the measurement error gets multiplied by an exponentially growing factor, which suggests that it is best to choose a tracer whose τ is larger than the duration of the experiment.

The flux of tracer towards the heart (through the veins) has been found to be

$$C_{\text{ped}}F_{\text{ped}} + C_{\text{vis}}F_{\text{vis}} .$$

If we neglect the volume of blood in the blood vessels and the heart, as well as the amount of tracer therein, we see that this expression also is the flux of tracer coming from the heart. This is the second trick of compartmental analysis: ‘neglecting the tubes.’

Of course, we need to distribute this cardiac outflux between the two blood spaces. We may assume that the heart thoroughly mixes the blood returned from the visceral blood space with that returned from the pedal space. In that case, the flux of tracer going through the artery that supplies the visceral blood space is

$$\frac{F_{\text{vis}}}{F_{\text{ped}} + F_{\text{vis}}} (C_{\text{vis}}F_{\text{vis}} + C_{\text{ped}}F_{\text{ped}}) .$$

Adding up the flux of tracer leaving the visceral space and the flux entering it, we find

$$\frac{d}{dt}Q_{\text{vis}} = -C_{\text{vis}}F_{\text{vis}} + \frac{F_{\text{vis}}}{F_{\text{ped}} + F_{\text{vis}}} (C_{\text{vis}}F_{\text{vis}} + C_{\text{ped}}F_{\text{ped}}) . \quad (3.18)$$

Our ‘negligible tubes’ assumption means that $Q_{\text{ped}} = Q_0 - Q_{\text{vis}}$. Also, the volume V_{vis} has been assumed to remain constant, which means that $\frac{d}{dt}Q_{\text{vis}} = V_{\text{vis}}\frac{d}{dt}C_{\text{vis}}$. Assembling all the pieces we find,

$$\frac{d}{dt}C_{\text{vis}} = \frac{Q_0F_{\text{ped}}F_{\text{vis}}}{V_{\text{ped}}V_{\text{vis}}F_{\text{CO}}} - \frac{F_{\text{ped}}F_{\text{vis}}}{F_{\text{CO}}} (V_{\text{ped}}^{-1} + V_{\text{vis}}^{-1}) C_{\text{vis}} \quad (3.19)$$

which is a familiar equation that we can readily solve:

$$C(t) = \bar{C} + \left(\frac{Q_0}{V_{\text{vis}}} - \bar{C} \right) \exp\{-\lambda t\} . \quad (3.20)$$

A data set which contains measurements of C_{vis} taken at various points in time can be used to estimate the parameters \bar{C} and λ . The first parameter is the concentration of tracer finally attained in both blood spaces as $t \rightarrow \infty$,

$$\bar{C} = \frac{Q_0}{V_{\text{ped}} + V_{\text{vis}}} , \quad (3.21)$$

and gives us an estimate of the total blood volume as Q_0/\bar{C} . Moreover, the C -axis intercept at $t = 0$ gives us the volume of the visceral blood space as $Q_0/C(0)$.

The second parameter is the rate parameter λ :

$$\lambda = \frac{F_{\text{ped}}F_{\text{vis}}}{F_{\text{CO}}} (V_{\text{ped}}^{-1} + V_{\text{vis}}^{-1})$$

which is a compound of F s and V s, and which looks rather messy. We can rewrite λ in a more attractive form:

$$\lambda = \frac{\lambda_{\text{ped}}\lambda_{\text{vis}}}{\lambda_{\text{CO}}} \quad (3.22)$$

Figure 3.4: Lymphocyte recirculation routes.

with the following definitions: $\lambda_{\text{ped}} = F_{\text{ped}}/V_{\text{ped}}$, $\lambda_{\text{vis}} = F_{\text{vis}}/V_{\text{vis}}$, and $\lambda_{\text{CO}} = (F_{\text{ped}} + F_{\text{vis}})/(V_{\text{ped}} + V_{\text{vis}})$. The total flux of blood through a blood space is called the ‘turnover rate’ of that compartment. When we make the (not altogether unreasonable) assumption that $\lambda_{\text{ped}} = \lambda_{\text{vis}}$, it follows that λ is this common turnover rate.

However, it is also quite possible that the animal is able to divert more blood to one or the other compartment. For instance, during locomotion, it may well require a greater supply of oxygen and nutrients to the pedal blood space (which engulfs the foot muscles). To investigate whether $\lambda_{\text{ped}} \neq \lambda_{\text{vis}}$, we need an independent estimate of the cardiac output F_{CO} ; this can be obtained by direct observation of the beating heart, which is observable through the shell of the snail. This estimated cardiac output, together with equations (3.17) and (3.22), furnishes estimates of λ_{ped} and λ_{vis} .

3.4.2 Circulation of lymphocytes

Lymphocytes are white blood cells involved in the specific immune response. These cells enter the various lymphoid organs from the blood. In these organs lymphocytes can become activated if an infection has occurred. The lymphoid organs include the spleen and a large number of lymph nodes. The former is a fist-sized organ sitting next to the pancreas, the latter are small bean-shaped organs scattered throughout the body. The lymphocytes eventually return to the blood, from where they may again enter the lymphoid organs, and so on. The situation is depicted schematically in Figure 3.4. This diagram also shows the average residence times of lymphocytes in the various compartments of the lymphocyte pools, as well as the fractions of the total lymphocyte circulation flow Φ that enters the various sites (this diagram has been somewhat simplified). Our aim is to derive, from these data, the distribution of lymphocytes among the various compartments.

Thus, let N_{bl} , N_{sp} , and N_{ln} represent the numbers of lymphocytes in blood, spleen,

and lymph nodes, respectively. We want to find the fractions

$$\frac{N_{\text{bl}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}}, \quad \frac{N_{\text{sp}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}}, \quad \frac{N_{\text{ln}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}}.$$

To relate the mean residence time to the flows of cells through the various sites, we carry out a little thought experiment. Consider one of the compartments, say the spleen. Imagine that at $t = 0$ we mark all the lymphocytes that reside in the spleen at that moment in time. After $t = 0$, unmarked lymphocytes will enter the spleen with a flux 0.45Φ . Since N_{sp} stays constant by assumption (we are looking for the stationary distribution of lymphocytes among the three major compartments), the outflow of lymphocytes also equals 0.45Φ . At $t = 0$, all of these outflowing lymphocytes will be marked, but for $t > 0$ the spleen contains a mixture of marked and unmarked lymphocytes, and the outflow will likewise be such a mixture. Let N_{sp}^* denote the number of marked lymphocytes. Because our imaginary marking does not affect the behaviour of the cells in any way, and assuming a well-mixedness of the lymphocytes, we find that the fraction of marked cells among the exiting cell flow is just equal to the fraction of marked cells in the spleen. Therefore,

$$\frac{d}{dt}N_{\text{sp}}^* = -\frac{N_{\text{sp}}^*}{N_{\text{sp}}}0.45\Phi = -\frac{0.45\Phi}{N_{\text{sp}}}N_{\text{sp}}^*. \quad (3.23)$$

This equation is just like equation (1.4), so we know our solution straightaway:

$$N_{\text{sp}}^*(t) = N_{\text{sp}} \exp\left\{-\frac{0.45\Phi}{N_{\text{sp}}}t\right\}. \quad (3.24)$$

We will now argue that $T_{\text{sp}} = N_{\text{sp}}/0.45\Phi$ is just the average residence time.

The solution $\exp\{-t/T_{\text{sp}}\}$ for the fraction of marked cells $N_{\text{sp}}^*/N_{\text{sp}}$ can be related to probabilities, as follows. It tells us that at $t = 0$ all cells are marked cells. Thus, the probability that any one of the cells present in the spleen at $t = 0$ remains there for at least zero seconds is just 1. This much is the blatantly obvious, of course, but the reasoning applies to any value of t . For instance, at $t = T_{\text{sp}} \ln\{2\}$, the fraction of marked cells equals one-half. Thus, half the cells remain in the spleen for at least a time $T_{\text{sp}} \ln\{2\}$, or: the probability that any cell remains in the spleen for at least $T_{\text{sp}} \ln\{2\}$ time units is $1/2$. At time t , there remains a fraction $\exp\{-t/T_{\text{sp}}\}$ of those cells originally marked at time $t = 0$. These cells, still present at t , must clearly leave the spleen at some time later than t . Therefore, the probability that any cell remains in the spleen for at least t time units is $\exp\{-t/T_{\text{sp}}\}$. From this we can calculate the probability that the residence time is in the interval $[t_1, t_2)$ (with $t_2 > t_1$): it is the probability that a cell remains for at least t_1 units minus the probability that a cell remains for at least t_2 units. The **average residence time** can be approximated by weighing each residence time with the approximate frequency at which that particular residence time occurs:

$$\text{average residence time} \approx \sum_{i=1}^{\infty} t_i (\exp\{-t_i/T_{\text{sp}}\} - \exp\{-t_{i+1}/T_{\text{sp}}\})$$

(with $t_1 = 0$). This is inexact because we associate the residence time t_i with an interval of residence times $[t_i, t_{i+1})$. We therefore let the maximum difference between subsequent values in $\{t_i\}_{i=1}^{\infty}$ go to zero;

$$\text{average residence time} = \int_0^{\infty} (t/T_{\text{sp}}) \exp\{-t/T_{\text{sp}}\} dt \quad (3.25)$$

from which it is easy enough to compute that the average residence time is, indeed, equal to T_{sp} . (The same argument can be applied to the butterfly example above, section 1.2: $1/\lambda$ there is just the life expectancy of the butterflies.)

For the spleen, we have shown that the average residence time T_{sp} of 5 h equals $N_{\text{sp}}/0.45\Phi$. Of course, the same thought experiment can be carried out for the other two compartments. Our data therefore give us the following identities:

$$\frac{N_{\text{bl}}}{\Phi} = 0.5 \text{ h} \quad \frac{N_{\text{sp}}}{0.45\Phi} = 5 \text{ h} \quad \frac{N_{\text{ln}}}{0.55\Phi} = 12 \text{ h}$$

or

$$N_{\text{bl}} = 0.5 \text{ h} \times \Phi \quad N_{\text{sp}} = 5 \text{ h} \times 0.45\Phi \quad N_{\text{ln}} = 12 \text{ h} \times 0.55\Phi$$

from which we find:

$$\frac{N_{\text{bl}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}} = 0.05 \quad \frac{N_{\text{sp}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}} = 0.24 \quad \frac{N_{\text{ln}}}{N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}}} = 0.71 .$$

Thus, about five percent of the lymphocytes are, at any given moment, in the blood. The majority resides in the lymph nodes.

Note that we did not need or use the numerical value of the total circulation Φ in this calculation. The estimate $N_{\text{bl}} + N_{\text{sp}} + N_{\text{ln}} \sim 10^{11}$ implies that Φ is about 10^{10} lymphocytes per hour, which is still only about a few milliliters of cell volume per hour. (These estimates are crude, though, and could be off by about a factor of 10.)

3.5 Metabolite homeostasis

A powerful way to analyse biological systems, from single cells to ecosystems, is by keeping track of the flows of mass through the system. Here we look at the transformations within a cell.

3.5.1 The scene of the action

Matter enters a biological cell through a large number of uptake processes, which can be down a chemical gradient across the cell's membrane, or up such a gradient, in which case special transporter proteins embedded in the cell membrane are required. Some of the matter entering the cell is converted into the molecules and macromolecules that make up the cell. This build-up (anabolism) accounts for the replacement of molecules that

are broken down, and any anabolism in excess of replacement means that the cell grows. Ultimately, cell growth can lead to division into two daughter cells. In this way, cell growth underlies tissue expansion, which in turn underlies growth of the organism itself.

Matter also leaves the cell in the form of products. These generally leave the cell by diffusion down a gradient, but can also leave through active transport or by a vesicle-based pathway. These products are the ‘waste’ of energy metabolism, carbon dioxide when oxidation is complete, fermentation products when it is not. Further kinds of products secreted by a cell include intercellular matrix, hormones, and digestive enzymes.

At the crossroads of cellular metabolism is a pool of small molecules called the intermediate metabolites or ‘core’ metabolites. The whole of cellular metabolism revolves around this pool. These serve as the starting point for the build-up of the ‘structural’ macromolecules (enzymes, DNA, RNA and suchlike) of the cell. They are also the fuel burned in the cell’s energy metabolism. Finally, they may be chemically concatenated into reserve polymers. This pool of metabolites is replenished in various ways. Break-down (turn-over) of the structural macromolecules is one way. Break-down of reserve polymers is another. Uptake of molecules from outside the cell also replenishes this pool. The molecules that are taken up are converted in a series of chemical reactions into the core metabolites.

3.5.2 Kinetics of a metabolite and an enzyme

In this section we focus on a tiny cog in this cellular machinery. We consider only two species of molecules: a single metabolite and the enzyme which catalyses the conversion of this metabolite into another small molecule.

Our state variables are the concentration of the metabolite, M and the concentration of ‘its’ enzyme Z . We assume that the concentration of the metabolite is low enough to warrant a linear approximation for the rate at which it is converted into its follow-up species (see section 3.1). Thus, the kinetics of M feature a loss term of the form $-\beta ZM$. The kinetics of M should also feature one or more gain terms. Each of these gain terms corresponds to a metabolic pathway ‘feeding’ the pool of our particular metabolite. These pathways are outside the present scope of consideration, and we will lump them together into a single term $\Phi(t)$. Thus, this flux $\Phi(t)$ is treated as an unknown ‘forcing’ function, which we assume > 0 . Summing gain and loss terms, we have the kinetics:

$$\frac{d}{dt}M = \Phi(t) - \beta ZM . \quad (3.26)$$

To find the kinetics of the enzyme, we need to consider the gain and loss terms in turn. The loss term is easy. The enzyme is subject to continual turn-over. We can model this term simply as $-\lambda Z$. Absorbed into the parameter λ is the concentration of the ‘turn-over’ enzyme, that is the enzyme that catalyses the break-down of ‘our’ enzyme. By taking λ to be a constant, we implicitly assume that the turn-over enzyme is present at a constant concentration.

As for the gain term, this term represents the synthesis of the enzyme. This synthesis depends on the amount of RNA coding for the enzyme, and also on the availability of

Figure 3.5: A metabolite and its enzyme, both controlling the expression of the enzyme.

ribosomes. The amount of RNA depends on the activity about the gene that codes for our enzyme. We could build a very complicated model indeed to represent these processes. Instead, we will make a grossly simplifying assumption: we will just assume that both our metabolite and our enzyme stimulate the gene of the enzyme (Figure 3.5).

On this assumption, the model has an interesting property which will become apparent shortly. Summing gain and loss terms, we have the kinetics:

$$\frac{d}{dt}Z = \gamma MZ - \lambda Z \quad (3.27)$$

where γ is a parameter that represents the strength of the interaction between enzyme, metabolite and gene.

3.5.3 The equilibrium

Let us consider the equilibrium of the system when the unknown flux Φ acting on the metabolite is constant:

$$\Phi(t) \equiv \bar{\Phi} .$$

Setting the left-hand sides of equations (3.26) and (3.27) equal to zero, we find:

$$\bar{M} = \frac{\lambda}{\gamma} \quad (3.28)$$

$$\bar{Z} = \frac{\gamma}{\beta\lambda} \bar{\Phi} \quad (3.29)$$

which is rather interesting: the steady state concentration of the metabolite is independent of the flux $\bar{\Phi}$.

Does this perfect regulation of the metabolite concentration mean that the system is operating with infinite gain? We saw in section 3.3 that the gain must be finite. I leave the answer to this riddle for you to work out yourself.

Fluctuations in the flux $\Phi(t)$ will of course result in fluctuations in $M(t)$, but—provided the above equilibrium is stable— $M(t)$ will continually tend to the constant value $\bar{M} =$

λ/γ . We see that the arrangement of interactions modelled by equations (3.26) and (3.27) safeguards a nearly constant concentration of the metabolite.

3.5.4 Stability of the equilibrium

To assess the stability of the equilibrium we employ the technique explained in chapter 6. We determine the Jacobian matrix in the equilibrium point:

$$\begin{bmatrix} -\beta\bar{Z} & -\beta\bar{M} \\ \gamma\bar{Z} & 0 \end{bmatrix}$$

and we need to ensure that the eigenvalues have negative real parts. Setting the determinant of the Jacobian matrix equal to zero and working out the resulting second-degree polynomial, we find the eigenvalues μ_1 and μ_2 :

$$\mu_{1,2} = -\frac{\beta\bar{Z}}{2} \pm \sqrt{(\beta\bar{Z}/2)^2 - \gamma\bar{\Phi}}.$$

The real parts of the eigenvalues are always negative, and thus the equilibrium (\bar{M}, \bar{Z}) is stable.

3.5.5 The control-system perspective

Again, it is instructive to consider the model from a different point of view: that of feedback control systems. In the parlance of control systems, the metabolite concentration is seen as part of a controlled process, and an explicit objective is to keep this concentration constant at a desired set point M^* . The enzyme is regarded as a means of removing excess metabolite. Thus, whenever the metabolite concentration $M(t)$ exceeds its set point M^* , the concentration of enzyme is increased. These considerations motivate the following feed-back control kinetics:

$$\frac{d}{dt}Z = \gamma(M - M^*)Z. \quad (3.30)$$

But this equation is equivalent to the previous equation (3.27) if we make the identification,

$$M^* = \frac{\lambda}{\gamma}.$$

For further remarks on the use of the term ‘set point’, see section 3.3.

3.5.6 Suggestions for further work

The model presented here is very simple, but can serve as a starting point for a more elaborate analysis of cellular homeostasis. For instance, the metabolite is just one link in a chain of metabolites, a ‘metabolic pathway.’ Thus, it is formed by conversion of a precursor substance, and it is itself converted into another metabolite. Suppose that all metabolites

in a pathway are controlled in the manner proposed by the present model. Would stability then still hold good?

We noted that the core metabolites form the starting points for many metabolic pathways. Thus, more than one enzyme acts on a metabolite. How do the regulation mechanisms of these metabolites interact? What if a metabolite affects the synthesis rate of an enzyme that does *not* act on that metabolite?

In the simple model, it was assumed that the rate of appearance of the enzyme could be instantaneously linked to the amounts of enzyme and metabolite present. However, changes in the processes at the DNA and RNA levels that underlie the synthesis of the enzyme take some time to occur, and thus one may surmise that the enzyme levels do not respond immediately to changes in the metabolite. Analysis of delays is complicated, and beyond the scope of these notes. However, it is a worthwhile question to ask how such a delay would affect the stability of the homeostasis mechanism described by the model.

Here is another question to explore. RNA molecules have to compete for access to the ribosomes. This is probably simply a passive type of competition: the RNA molecules that happen to be around at higher levels also happen to get a bigger share in the ribosomal processing time available in the cell's synthetic apparatus. Nonetheless, these aspects may come to play a role when the cell is responding to drastic changes in nutrient availabilities, for example.

Finally, we remark upon some fundamental quandaries in modelling cellular metabolism. The first concerns the use of deterministic differential equations. As remarked in section 1.2, such use is justified as these equations describe the expected numbers over a large collection of replicates of the actual process. However, as the number of molecules becomes smaller, random fluctuations that always occur during the actual process may become important for the kinetics.

The second fundamental point concerns the fact that the enzyme itself is made out of building blocks whose pools are replenished by numerous metabolic pathways, among which are those that contain the metabolite that the enzyme helps to convert. Thus, a 'strange loop' connects enzyme and metabolite: the cell's genes, enzymes and structural infrastructure determine the fate of the small molecules, but at the same time these small molecules themselves constitute the fate of the big molecules. How such a structure can maintain itself and, moreover, multiply, is not at all clear.

3.6 Basic models of population growth

A population is an assembly of individual organisms which are similar in some way or other. Usually a population in an area is the sum of all organisms in that area which belong to a certain species.

When everybody is doing what everybody else is doing, the total effort naturally takes on the form of population size times the effort *per capita*. In particular, when the effort in question is the production of new individuals, the rate of change of the population size is

expressed as population size N times *per capita* reproduction rate:

$$\frac{d}{dt}N = N \times \textit{per capita reproduction rate} . \quad (3.31)$$

Accordingly, the *per capita* reproduction rate is often written $\frac{d}{dt} \ln\{N\}$.

3.6.1 Malthusian growth

Perhaps the simplest assumption that one could make about the *per capita* reproduction rate is that it is constant (say, ρ):

$$\frac{d}{dt}N = N \times \rho \quad (3.32)$$

whence $N(t) = N(0) \exp\{\rho t\}$. This is called Malthusian growth². Of course, individuals are not only born, they also die. If we assume that the rate at which individuals enter Kingdom Come is proportional to the population size, we get the following adding the death term $-\mu N$:

$$\frac{d}{dt}N = \rho N - \mu N = (\rho - \mu)N = rN \quad (3.33)$$

where the quantity $(\rho - \mu) = r$ is the net *per capita* growth rate.

3.6.2 Limits on growth

The solution to equation (3.33) does not permit a reasonable interpretation for large t , as N will become as large as you like. Of course, actual populations do not behave in this way. The resources required for reproduction—nutrients, space, water—are limited and will become depleted as the population grows. This suggests that a more reasonable model should multiply r by some dimensionless multiplier $f \in [0, 1]$ to account for this effect:

$$\frac{d}{dt}N = frN \quad (3.34)$$

There are various ways to model this multiplier f . First, the depletion of resources may be modelled explicitly by the addition of one or more state variables that describe the kinetics of the resource and, possibly, of populations of other species which affect the availability of the resource. This approach is aesthetically pleasing, since it offers the opportunity to incorporate mass balances explicitly, and it is mechanistically transparent. A drawback is that the models involve two or more additional state variables, with kinetics (the ‘right-hand sides’ of the differential equations) that generally depend non-linearly on the state variables. Models in this class will not be pursued here.

Second, f might be conceived as dependent on time. In virtually all ecosystems, one finds marked seasonal fluctuations in the availability of resources and the prevalence of

²In fact, ‘Malthusian growth’ was first considered by the philosopher Condorcet, not Malthus.

conditions (sunlight, temperature). A simple way to implement this idea is to use a forcing function (non-autonomous input) for f . Thus, f becomes just a descriptive function of time, which should be supported by measurements in the field.

Third, f might be taken to depend on the population size itself. Population growth models where $\partial f/\partial N < 0$ are said to be models with interference, and $|\partial f/\partial N|$ is just the strength of the within-population interference. Various ways to flesh out this idea are taken up in the next section.

3.6.3 Interference: logistic growth and variations

The most basic way to set up models for population growth with interference is to make f a function of N only. This idea is quite reasonable in various situations. For instance, when the supply of nutrients is steady, the availability per individual decreases with increasing N : $f \propto N^{-1}$ and the population would grow linearly, $N \propto t$. Another example: limited availability of space may be the dominant check on population growth. This might involve the need for space to build nests, or a space-related factor such as access to water or food supplies. A convenient assumption is that f decreases linearly with N in such cases:

$$f = 1 - \alpha N .$$

Note that $1/\alpha$ is just the population size at which $f = 0$ and thus $\frac{d}{dt}N = 0$: accordingly, $1/\alpha$ is called the carrying capacity of the environment and often denoted K :

$$f = 1 - N/K$$

which we plug into equation (3.34):

$$\frac{d}{dt}N = r \left(1 - \frac{N}{K} \right) N = rN - \frac{r}{K}N^2 . \quad (3.35)$$

This equation is called the logistic growth equation or Verhulst equation. It can be solved using separation of variables (chapter 5):

$$N(t) = \frac{K}{1 - (1 - K/N(0)) \exp\{-rt\}} . \quad (3.36)$$

The graph of this solution starts out almost exponentially, but then turns from concave to convex and approaches the asymptote (K). The point of inflection is just at $N = K/2$.

The linear decrease of f with N may not be very realistic. For instance, in a territorial species the interference effect may already be quite strong at lower population sizes, but gradually become less strong as the individuals compromise and settle for smaller territory sizes until the carrying capacity is reached. Alternatively, the individuals may occupy their patches in the habitat without interacting very strongly with other individuals; in that case, interference is weak until almost all available space is taken up, and then interference kicks in strongly as the population size approaches the carrying capacity closely, as might be the

Figure 3.6: Left: $f(N)$ at various σ . Right: The corresponding solutions.

case for duckweed. Both examples are accounted for in the following generalization of the logistic growth equation:

$$f = 1 - \left[\frac{N}{K} \right]^{1/\sigma} \quad \text{so} \quad \frac{d}{dt}N = r \left(1 - \left[\frac{N}{K} \right]^{1/\sigma} \right) N \quad (3.37)$$

where σ is a dimensionless parameter, where $\sigma = 1$ is obviously the original Verhulst case. The territorial species case is represented by $\sigma > 1$, where interference is strongest at low population sizes. The ‘duckweed’ scenario is represented by $\sigma < 1$. The function $f(N)$ is sketched in Figure 3.6 for a number of values of σ .

Again, equation (3.37) can be solved by separation of variables (chapter 5):

$$N(t) = K \left(1 - \left(1 - \left[\frac{N(0)}{K} \right]^{-1/\sigma} \right) \exp\{-rt/\sigma\} \right)^{-\sigma}. \quad (3.38)$$

When σ is very small, growth is virtually exponential (with doubling time $\ln\{2\}/r$) almost right up to the carrying capacity K . The growth curve is sigmoid for all values of σ , with a point of inflection at $N(t) = K(1 + 1/\sigma)^{-\sigma}$. Note that this point of inflection approaches K for $\sigma \rightarrow 0$, while it goes to K/e for $\sigma \rightarrow \infty$.

3.6.4 Positive interference

In sexual populations, females need to find males to fertilize them, which may be hard at very low densities. Thus, at $N = 0$, we may have $\partial f / \partial N > 0$. The multiplier f itself may be zero or even negative at $N = 0$. At some greater population size, interference strength becomes negative again. Thus, the graph of f as a function of N assumes the shape of a ‘hump.’ Ecologists call this the Allee effect.

As long as $f \geq 0$ at $N = 0$, the only qualitative difference the Allee effect makes compared to logistic growth is that the population grows rather more slowly initially, when the population size is still small. When $f < 0$ at $N = 0$, the *per capita* growth rate

Figure 3.7: Above: $f > 0$ at $N = 0$. Below: $f < 0$ at $N = 0$. Left: $f(N)$. Right: growth rate as a function of N .

is negative below a certain critical population size, say N_{crit} with $N_{\text{crit}} < K$. If the initial population size is smaller than N_{crit} , the population will die out; zero is a stable equilibrium in the case of a positive N_{crit} . The population will grow to the carrying capacity if the initial population size is larger than N_{crit} . Note that N_{crit} itself represents an unstable equilibrium. Figure 3.7 shows $f(N)$ and $\frac{d}{dt}N$ as a function of N in both cases.

3.6.5 Harvesting

A growth equation of the following form is often employed in the analysis of harvesting and overfishing:

$$\frac{d}{dt}N = f_{\text{Allee}}(N)rN - \varepsilon N = (f_{\text{Allee}}(N)r - \varepsilon)N \quad (3.39)$$

where f_{Allee} is a multiplier function that incorporates the Allee effect, and the term εN represents an additional loss term due to harvesting. The parameter ε is called the harvesting effort.

Let us denote the largest stable equilibrium at $\varepsilon = 0$ as the carrying capacity K , just as before. What happens when ε is given a slight but positive value, that is, if we try to harvest some of the population? We see that this equilibrium becomes smaller than K , and approaches zero (the unstable equilibrium) as $\varepsilon \rightarrow \infty$.

Figure 3.8: Stable (drawn line) and unstable (dashed line) equilibria as a function of ε . Left: $f > 0$ at $N = 0$. Right: $f < 0$ at $N = 0$.

The extracted harvesting flux εN is called the yield. The steady state yield is just ε times the equilibrium value. Since the latter decreases with ε , there will be an optimal finite value for ε at which the yield can be maximized. Note that the ratio of the extracted harvesting flux (εN) over the effort expended (ε) is, at steady state, just the equilibrium population size, which shows that the harvesting is most effective at low ε . However, the actual costs will consist of a constant term and a term proportional to ε , which complicates the analysis of cost-effectiveness.

When $f < 0$ at $N = 0$, the unstable lower equilibrium is $N_{\text{crit}} > 0$ in the absence of any harvesting effort. As the harvesting effort is increased, this equilibrium becomes larger than N_{crit} . Thus, the lower unstable equilibrium ($> N_{\text{crit}}$) approaches the stable equilibrium ($< K$) as ε increases, until at some ε the two coincide in an unstable equilibrium. Increasing ε any further causes this equilibrium to disappear altogether, leaving only the stable equilibrium of a zero population size. The meeting up of an unstable and a stable equilibrium as a parameter (ε in this case) is increased and the subsequent disappearance of both as the parameter is increased further is called a forward tangent bifurcation (Figure 3.8).

We return to the case where ε is small enough to allow the existence of two distinct equilibria between N_{crit} and K . For any (initial) value of $N(t)$ below the lower unstable equilibrium, the population will tend toward the zero equilibrium. Stepping down ε will bring $N(t)$ above the lower unstable equilibrium, salvaging the population which then tends to the upper stable equilibrium. However, this trick only works as long as $N(t) > N_{\text{crit}}$; for no stepping down the harvesting effort, not even to $\varepsilon = 0$, will prevent the population from perishing (recall that N_{crit} is the lower unstable equilibrium at $\varepsilon = 0$).

3.7 Butterflies, again

Let us return to the example of the declining cohort of butterflies in the hothouse, discussed earlier in chapter 1. Thus far we have been concerned with a deterministic model of the

decline of this cohort:

$$N(t) = N_0 \exp\{-\lambda t\} \quad (3.40)$$

where λ is the death rate of the butterflies and N_0 is the initial amount of butterflies. We justified the use of a real variable N to model the number of butterflies surviving at time t by stating that this N is just the average of a very large number of replications of the hothouse butterfly counting experiment. Thus we dispensed with the objections that actual numbers of butterflies in any given experiment are always integers, and also that for $t \rightarrow \infty$ our model $N_0 \exp\{-\lambda t\}$ approaches zero asymptotically, whereas in a real experiment there comes a definite moment when the number of surviving butterflies drops down from 1 to 0, the moment at which the cohort has died out. In this section we want to show that our deterministic model indeed describes an average.

We begin by looking at the time at which the number of surviving butterflies first attains zero (the moment of extinction). Let us write T_k for the moment at which the k th butterfly dies. Then the moment of extinction is T_{N_0} . We can note straightaway that

$$\mathbb{P}[T_{N_0} \leq t] = \mathbb{P}[\text{all } T_i \text{ are lower than or equal to } t]. \quad (3.41)$$

Also the probability that any given butterfly dies before time t is $1 - \exp\{-\lambda t\}$. If the times of death of the butterflies are statistically independent³, we can then write

$$\mathbb{P}[T_{N_0} \leq t] = (1 - \exp\{-\lambda t\})^{N_0} \quad (3.42)$$

and this is the distribution of the moment of extinction. We can use this result to estimate the death rate λ in the following situation: the experimentalists has kept note *only* of the initial number of butterflies, N_0 , and the time of extinction T_{N_0} (but not of T_k for $k = 1, \dots, N_0 - 1$). The likelihood that the death rate equals λ given these data is

$$L(\lambda \mid N_0, T_{N_0}) = \lambda N_0 (1 - \exp\{-\lambda T_{N_0}\})^{N_0-1} \exp\{-\lambda T_{N_0}\}$$

(the RHS is the probability density function of T_{N_0} , obtained by differentiating its distribution, equation (3.42), with respect to T_{N_0}). We find the λ -value that maximizes this likelihood by working out $dL/d\lambda = 0$, which gives:

$$\lambda = \frac{1 - \exp\{-\lambda T_{N_0}\}}{(1 - N_0 \exp\{-\lambda T_{N_0}\}) T_{N_0}}.$$

This is an implicit equation for the MLE of the death rate, which has to be solved numerically.

The reasoning behind equation (3.42) can be extended to the moment at which the k th death occurs: we note that

$$\mathbb{P}[T_k \leq t] = \mathbb{P}[k \text{ or more times of death are at most } t]$$

³Two events are statistically independent if the probability of both of them happening is the probability that one happens, multiplied by the probability that the other happens.

and furthermore that

$$\mathbb{P}[\text{exactly } k \text{ times of death are at most } t] = \binom{N_0}{k} \mathbb{P}[\text{butterfly dies no later than } t]^k \mathbb{P}[\text{butterfly dies later than } t]^{N_0-k}$$

and combining these two we arrive at the desired formula:

$$\mathbb{P}[T_k \leq t] = \sum_{j=k}^{N_0} \binom{N_0}{j} (1 - \exp\{-\lambda t\})^j \exp\{-\lambda(N_0 - j)t\} . \quad (3.43)$$

Instead of looking at the times of death, we can also look at the probability that the number of butterflies at time t is equal to n . Let us denote this probability as $p(n, t)$ and consider carefully what this quantity represents. Suppose that we could repeat the hothouse experiment under exactly the same conditions an infinite number of times. Then $p(n, t)$ is the fraction of these experiments in which it would be found, at time t , that n butterflies are still surviving. Such an imaginary, infinitely large, collection of replicates of the same physical system is called an **ensemble**. Actual replications of an experiment are only an approximation to this ideal: we can only do a finite number of replications, and we can never guarantee that all experiments were done under exactly the same conditions: for starters, they will typically involve different subjects or the same subject at different points in time.

Our aim is to characterize $p(m, t)$, and it turns out that the idea of writing ODEs for them proves fruitful. If we count the surviving butterflies at time t , come back later at time $t + \Delta t$, and count again, we will notice that the number has stayed the same, or has gone down an integer number. If we choose Δt small enough, the probabilities that the number has gone down more than 1 butterfly become negligibly small, and we can pretend that there are just two possibilities: the number has stayed the same or has gone down by one butterfly. Let p_Δ denote the probability that a butterfly dies between t and $t + \Delta t$. We choose Δt small enough that $p_\Delta \ll 1$; thus a butterfly is very likely to survive a period of time as brief as Δt . Then the probability that all of the n butterflies surviving at time t are still alive at time $t + \Delta t$ is

$$(1 - p_\Delta)^n \approx 1 - np_\Delta ,$$

an approximation which becomes exact as Δt approaches zero. Now the concept on the ensemble comes into play. We can deduce that

$$p(n, t + \Delta t) = p(n, t)(1 - np_\Delta) + p(n + 1, t)(n + 1)p_\Delta$$

which is the fraction of systems in the ensemble that have stayed at n plus the fraction of systems in the ensemble that were at $n + 1$ but have lost a butterfly during the interval Δt . Our final preparatory step is to define the death rate as follows:

$$\lambda = \lim_{\Delta t \rightarrow 0} \frac{p_\Delta}{\Delta t} ; \quad (3.44)$$

a quantity defined in this manner is called a **probability rate**. We are now ready to write down the differential equation for $p(n, t)$:

$$\begin{aligned} \frac{d}{dt}p(n, t) &= \lim_{\Delta t \rightarrow 0} \frac{p(n, t + \Delta t) - p(n, t)}{\Delta t} \\ &= -n\lambda p(n, t) + (n + 1)\lambda p(n + 1, t) . \end{aligned} \quad (3.45)$$

You can solve equations (3.45) for $n = 1, \dots, N_0$. Start with N_0 , noting that $p(n, t) = 0$ for all $n \geq N_0 + 1$, for all t (so that $\frac{d}{dt}p(N_0, t) = -\lambda N_0 p(N_0, t)$), and then work back.

In fact, we do not need to solve the ODEs for $p(n, t)$ to determine the behaviour of interesting statistics. For instance, consider the expectation (over the ensemble):

$$\mathbb{E}[n](t) = \sum_{n=0}^{\infty} np(n, t) = \sum_{n=1}^{\infty} np(n, t) \quad (3.46)$$

(we can skip the $n = 0$ term since it contributes nothing; the terms for $n > N_0$ also make no contribution since $p(n, t) = 0$ for all $n \geq N_0 + 1$, for all t). Using equation (3.45) we can derive an ODE for this expectation:

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[n] &= \sum_{n=1}^{\infty} n \frac{d}{dt}p(n, t) \\ &= \sum_{n=1}^{\infty} n\lambda [(n + 1)p(n + 1, t) - np(n, t)] \\ &= \lambda \left(\sum_{n=1}^{\infty} n(n + 1)p(n + 1, t) - \sum_{n=1}^{\infty} n^2 p(n, t) \right) . \end{aligned} \quad (3.47)$$

To work out the first sum, we reintroduce the term with $n = 0$ and then change to a different index $\tilde{n} = n + 1$:

$$\begin{aligned} \sum_{n=1}^{\infty} n(n + 1)p(n + 1, t) &= \sum_{n=0}^{\infty} n(n + 1)p(n + 1, t) \\ &= \sum_{\tilde{n}=1}^{\infty} \tilde{n}(\tilde{n} - 1)p(\tilde{n}, t) \\ &= \sum_{\tilde{n}=1}^{\infty} \tilde{n}^2 p(\tilde{n}, t) - \sum_{\tilde{n}=1}^{\infty} \tilde{n}p(\tilde{n}, t) \end{aligned} \quad (3.48)$$

Inserting this into equation (3.47) (and renaming the index n again) we obtain:

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[n] &= \lambda \left(\sum_{n=1}^{\infty} n^2 p(n, t) - \sum_{n=1}^{\infty} np(\tilde{n}, t) - \sum_{n=1}^{\infty} n^2 p(n, t) \right) \\ &= -\lambda \sum_{n=1}^{\infty} np(\tilde{n}, t) = -\mathbb{E}[n] \end{aligned} \quad (3.49)$$

and thus we make good on our claim that the original ODE $\frac{d}{dt}N = -\lambda N$ describes the average of a large number of experiments.

||

Exercises

4

Dynamics: state transition functions

The concept of ‘state’ plays a central role in the mathematical description of biological processes. Broadly speaking, the state comprises the totality of information that is required to determine “what happens next”.

§ 4.1 (State-transition functions) The **state-transition function** S of a deterministic dynamical system gives the **state** x at time t_1 given the value of the state at time t_0 :

$$x(t_1) = S(t_1, t_0, \xi) \quad \text{where } x(t_0) = \xi . \quad (4.1)$$

Thus, S “moves the state x forwards in time” by an amount $t_1 - t_0$.

Exercise 4.1 Let $S(t_1, t_0, \xi) = (t_0/t_1)^2 \xi$, and let $t_0 = 1$, $x(t_0) = \xi = 3$. Calculate $x(5)$. Calculate $x(8)$. Sketch a graph of $x(t)$ for $t \in [-3, 10]$. Also sketch a graph of $x(t)$ with the same state-transition function, but with $t_0 = 2$ and $x(t_0) = \xi = 3$.

Exercise 4.2 Motivate the **first consistency condition**:

$$S(t_0, t_0, \xi) = \xi . \quad (4.2)$$

(Hint: how far is the state “moved forwards in time”?)

Exercise 4.3 Motivate the **second consistency condition**:

$$S(t_2, t_0, \xi) = S(t_2, t_1, S(t_1, t_0, \xi)) . \quad (4.3)$$

(Hint: use equation (4.1) to deduce that both sides of equation (4.3) must be expressions for $x(t_2)$; in a *deterministic* system $x(t_2)$ is fixed once $x(t_0)$ has been specified.)

5

Dynamics in continuous time; one-dimensional systems

This chapter covers some basic techniques to solve ordinary differential equations.

§ 5.1 In a **continuous-time** dynamical system, the values for t_0, t_1, t_2, \dots can be arbitrarily close together on the real line. Thus we can consider the partial derivative of the state-transition function with respect to its first argument:

$$S_{\{1\}}(t, \tau, \xi) = \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t, \tau, \xi) - S(t, \tau, \xi)}{\Delta t}. \quad (5.1)$$

Exercise 5.1 Let $S(t_1, t_0, \xi) = (t_0/t_1)^2 \xi$. Verify that

$$S_{\{1\}}(t_1, t_0, \xi) = -2t_0^2 t_1^{-3} \xi.$$

Exercise 5.2 Suppose that, in a dynamical system with a 1-dimensional state $x \in \mathbb{R}$, the state-transition function is defined in terms of two other functions f and g , as follows:

$$S(t_1, t_0, \xi) = f(t_1, t_0, \xi) \exp\{g(t_1, t_0, \xi)\} \quad (5.2)$$

and suppose, furthermore, that $g(t_0, t_0, \xi) = 0$ for all values of ξ . Then show that $f(t_0, t_0, \xi) = \xi$ for all values of ξ . (Hint: notation: $\exp\{x\} \equiv e^x$; S must satisfy the consistency conditions.)

Exercise 5.3 With the state-transition function given by equation (5.2), calculate $S_{\{1\}}$.

Exercise 5.4 Can you explain why we have the following for continuous-time dynamics?

$$\frac{dx(t)}{dt} = S_{\{1\}}(t, t, x(t)). \quad (5.3)$$

(Hint: define the derivative as $\Delta x / \Delta t$ as $\Delta t \rightarrow 0$; first show that $\Delta x = S(t + \Delta t, t, x(t)) - S(t, t, x(t))$.)

§ 5.2 Another common appearance of the differential equation (5.3) is

$$\dot{x}(t) = F(t, x(t)) \quad (5.4)$$

where $F(\tau, \xi) \equiv S_{\{1\}}(\tau, \tau, \xi)$.

Exercise 5.5 (Bernoulli's method) Can you verify that the state-transition function given by equation (5.2) with

$$g(t, t_0, \xi) = \int_{t_0}^t \phi(\tau, t_0, \xi) d\tau \quad (5.5)$$

and

$$f(t, t_0, \xi) = x_0 + \int_{t_0}^t \psi(\tau, t_0, \xi) \exp\{-g(\tau, t_0, \xi)\} d\tau \quad (5.6)$$

is associated with the following differential equation?

$$\dot{x}(t) = \psi(t, t_0, x_0) + \phi(t, t_0, x_0)x(t) \quad (5.7)$$

where $x(t_0) = \xi$. (Hint: use your results in exercises 5.2 and 5.3; notice first that $g_{\{1\}}(t, t_0, x_0)$ corresponds to $\phi(t, t_0, x_0)$ and use the convention that $g(t_0, t_0, \xi) = 0 \forall \xi$.)

Exercise 5.6 Find the state-transition function corresponding to the following differential equation:

$$\dot{x}(t) = \alpha t - \lambda x(t)$$

with $x(t_0) = \xi$. (Hint: refer to equation (5.7), setting $\phi \equiv -\lambda$ and $\psi \equiv \alpha t$, and evaluate f and g according to equations (5.5) and (5.6); $\frac{d}{dt}(te^{\lambda t}/\lambda) = te^{\lambda t} + e^{\lambda t}/\lambda$.)

Exercise 5.7 Find the state-transition function corresponding to the following differential equation:

$$\dot{x}(t) = \alpha t - \beta t x(t)$$

with $x(t_0) = \xi$. (Hint: follow the same procedure as in exercise 5.6 setting $\phi \equiv -\beta t$ and $\psi \equiv \alpha t$; $\frac{d}{dt} \exp\{\beta t^2/2\} = \beta t \exp\{\beta t^2/2\}$.)

Exercise 5.8 Find the state-transition function corresponding to the following differential equation:

$$\dot{x}(t) = \rho e^{-\lambda t} x(t)$$

with $x(t_0) = \xi$. (Hint: follow the same procedure as in exercise 5.6 setting $\phi \equiv \rho e^{-\lambda t}$ and $\psi \equiv 0$.)

§ 5.3 For a wide range of scientific problems involving the quantitative description of “processes”, it turns out that differential equations provide a natural medium for the description of these processes.

Exercise 5.9 Can you explain why this is so?

§ 5.4 Given a differential equation, you will typically want to find (or at least characterize) the accompanying state-transition function. Finding the state-transition function corresponding to a differential equation is called **solving** the differential equation¹.

When a state-transition function $S(t_1, t_0, \xi)$ is presented as a putative solution of a differential equation $\dot{x}(t) = F(t, x(t))$, with boundary condition $x(t_0) = x_0$, two things need to be verified: the first consistency condition $S(t_0, t_0, x_0) = x_0$, and $S_{\{1\}}(t, t_0, \xi) = F(t, S(t, t_0, x_0))$.

¹The ‘unknown’ for which the differential equation is solved is the *function* which maps time t to x at time t . A differential equation specifies a function by giving a relation between that function and its derivative(s).

Exercise 5.10 Perform these two checks on the solutions you found in exercises 5.6—5.8.

Exercise 5.11 Given a dynamical system with state $x \in \mathbb{R}$, $x(t) = S(t, t_0, x(t_0))$, and a strictly increasing function R , define the variable u by

$$x = R(u) + x(t_0) \quad (5.8)$$

and show that u satisfies the following differential equation:

$$\dot{u}(t) = \frac{S_{\{1\}}(t, t, R(u) + x(t_0))}{R'(u(t))} = \frac{\dot{x}(t)}{R'(u(t))}. \quad (5.9)$$

Exercise 5.12 (Separation of variables) Can you show that a differential equation whose right member is a product of two functions, one of which is a function of only the state variable, and the other is a function of only time, like so:

$$\dot{u}(t) = \phi(u)\psi(t) \quad (5.10)$$

has the following solution?

$$\int_{t_0}^t \psi(\tau) d\tau = \int_{u(t_0)}^{u(t)} \frac{ds}{\phi(s)}. \quad (5.11)$$

(Hint: compare equations (5.9) and (5.10); set $\phi(u) = 1/Ru$ and $\psi(t) = \dot{x}(t)$. Integrate these and observe that the left and right members of equation (5.11) are both expressions for $x(t) - x(t_0)$.)

Exercise 5.13 Use formula (5.11) to solve

$$\dot{u}(t) = u^\alpha t^\beta$$

where $t \geq t_0 = 1$, $u(t_0) = 1$, and $\alpha \neq 1$, $\beta \neq -1$.

Exercise 5.14 Do the previous exercise for the case where $\beta = -1$.

Exercise 5.15 Use the formula (5.11) to solve

$$\dot{x}(t) = \rho x(1-x)t^\alpha$$

where $\alpha \neq -1$ and $x(t_0) = \xi < 1$.

§ 5.5 A differential equation is **autonomous** if its right member depends on t through the state x , but *only* through the state:

$$\dot{x}(t) = F(x(t)) \quad (5.12)$$

(cf. equation (5.4)).

Exercise 5.16 Which of the following differential equations is autonomous?

- (i) $\dot{x}(t) = V \frac{x(t)}{K+x(t)}$
- (ii) $\dot{x}(t) = \alpha x(t) + \beta t^2$
- (iii) $\dot{x}(t) = \cos(t) - \lambda x(t)$
- (iv) $\dot{x}(t) = (e^{\mu t} + x(t))^{-1}$

(Hint: look out for appearances of t in the right member other than in $x(t)$; these indicate that the equation is non-autonomous.)

Exercise 5.17 (Barrow's formula) Use formula (5.11) to show that the autonomous differential equation is solved by:

$$t - t_0 = \int_{x(t_0)}^{x(t)} \frac{d\xi}{F(\xi)}$$

whenever $F(\xi) \neq 0$ between the integration limits.

§ 5.6 Solving differential equations is a difficult art. Fortunately, much information about the state-transition function can already be gleaned from the differential equation itself, without the need for an explicit solution.

Exercise 5.18 Consider the autonomous differential equation:

$$\dot{x}(t) = F(x) = x - x^2 = x(1 - x).$$

Sketch $F(x)$ as a function of x . Mark on the x -axis the **critical points** where $F(x) = 0$; color *green* the part(s) of the x -axis where $F(x) > 0$, and *red* the part(s) where $F(x) < 0$. If the initial condition $x(t_0)$ is in a green region, what will the solution look like (i.e. will x go up or down, will it continue to do so or approach an eventual value, *et cetera*)? Ditto for an initial condition in a red region.

Exercise 5.19 Draw a polynomial of your own choosing (just a smooth line that wiggles around the x -axis a couple of times is fine) which you regard as $F(x)$ as in the previous exercise and repeat the red/green segments procedure.

Exercise 5.20 Repeat the previous exercise, but now make sure that a critical point coincides with an extremum of your polynomial.

Exercise 5.21 Consider an autonomous differential equation $\dot{x}(t) = F(x(t))$ where $F(x) > 0$ for $x_\alpha < x < x_\omega$, $F(x_\alpha) = F(x_\omega) = 0$, and $|F(x)| \leq L_\omega |x_\omega - x|$ where L_ω is a finite positive constant. Can you establish the following?

$$\lim_{t \rightarrow \infty} x(t) = x_\omega \quad \text{whenever } x(t_0) \in (x_\alpha, x_\omega).$$

If, in addition, we have $|F(x)| \leq L_\alpha |x_\alpha - x|$ can you similarly prove the following?

$$\lim_{t \rightarrow -\infty} x(t) = x_\alpha \quad \text{whenever } x(t_0) \in (x_\alpha, x_\omega).$$

Exercise 5.22 Consider two distinct state-transition functions ($\xi < 0$ and $t > t_0$ in both cases):

$$S(t, t_0, \xi) = (\xi^{1/3} + \frac{1}{3}(t - t_0))^3$$

and

$$S(t, t_0, \xi) = \begin{cases} (\xi^{1/3} + \frac{1}{3}(t - t_0))^3 & \text{for } t \leq t_0 - 3\xi^{1/3} \\ 0 & \text{for } t > t_0 - 3\xi^{1/3} \end{cases} .$$

Show that *both* state-transition functions satisfy the autonomous differential equation

$$\dot{x}(t) = x^{2/3} . \quad (5.13)$$

§ 5.7 The non-uniqueness in the last exercise arises because the critical point (here: $x = 0$) is reached in finite time. This cannot happen when the condition with L_ω given in exercise 5.21 is satisfied, as was shown in that exercise.

Exercise 5.23 Show that the right member of equation (5.13) fails to satisfy the ‘L’ condition. (Hint: show that there is no finite L such that $|x^{2/3}| \leq L|x|$ for all $x < 0$; it may be helpful to sketch $x^{2/3}$ as a function of x , paying particular attention to the slope at $x = 0$.)

Exercise 5.24 Can you sketch the solutions of

$$\dot{x}(t) = \frac{1}{1 - x(t)}$$

where $x(t_0) < 1$ and where $x(t_0) > 1$?

§ 5.8 Instead of colouring segments of the abscissa red and green, as you did in exercise 5.18, you can alternatively use little arrowheads pointing to the left and to the right, respectively.

Exercise 5.25 Explain how this arrowheads recipe allows you to identify (putative) alpha and omega points.

Exercise 5.26 Consider

$$\dot{x}(t) = F(x(t)) = \lambda - x(t)^2 .$$

Draw sketches of $F(x)$ for (i) $\lambda < 0$; (ii) $\lambda = 0$; (iii) $\lambda > 0$. Carry out the red/green segments (or arrowheads) procedure to identify alpha and omega points (if any).

Exercise 5.27 Repeat the previous exercise for the following differential equation:

$$\dot{x}(t) = F(x) = x(t) (\lambda - x(t)) .$$

Exercise 5.28 Consider

$$\dot{x}(t) = F(x(t)) = \lambda - \alpha x(t) + \beta \frac{x(t)^2}{1 + x(t)^2}$$

where $x \geq 0$, $\lambda \geq 0$, and $0 < \alpha < \beta < 2\alpha$. Draw sketches of $F(x)$ for various values of λ . Can you also sketch a **bifurcation** diagram, which has λ as abscissa², and the x -values for which $F(x) = 0$ (at the given λ) as ordinate?

²The **abscissa** is the horizontal axis, the **ordinate** is the vertical axis. We often also say “ x -axis” and “ y -axis”, knowing full well that the quantities along these axes may bear different names.

6

Phase plane analysis

In the last chapter, the state x was 1-dimensional. Here, you consider 2-dimensional systems.

§ 6.1 (The phase plane) Let the state be 2-dimensional: $x \in \mathbb{R}^2$. We write $x = (x_1, x_2)$, $\xi = (\xi_1, \xi_2)$. Notice that $S(t, t_0, \xi)$ can be represented by a point in the plane (x_1, x_2) (called the **phase plane**); for instance, $S(t_0, t_0, \xi)$ corresponds to the point (ξ_1, ξ_2) . As t ranges over \mathbb{R} , the state-transition function S ‘sweeps out’ a set of points in the phase plane, usually a smooth curve, which is the **(phase) path (passing through ξ)**, also known as the **orbit** or the **trajectory**.

Exercise 6.1 Let $t_0 = 0$ and

$$\begin{cases} x_1(t) &= ae^{-\lambda t} \\ x_2(t) &= be^{-\lambda t} + ce^{-\gamma t} \end{cases}$$

Sketch the phase path for $a = 1$, $b = 3$, $c = 3$, $\gamma = 2\lambda$. (Hint: find x_2 as function of x_1 ; note that $\exp\{-\gamma t\} = (\exp\{-\lambda t\})^{\gamma/\lambda}$; make sure that your curve *only* contains points ‘visited’ by the system at some time t .)

Exercise 6.2 For the system of the previous exercise, verify that $\xi_1 = a$ and $\xi_2 = b + c$ and show that

$$\lim_{t \rightarrow \infty} S(t, t_0, \xi) = (0, 0)$$

for all $\xi = (\xi_1, \xi_2)$. Sketch a few more paths through different choices of ξ . Put arrowheads on your paths to indicate the motion toward the origin.

Exercise 6.3 Show that the system

$$\begin{cases} x_1(t) &= ae^{-\lambda\sqrt{t}} \\ x_2(t) &= be^{-\lambda\sqrt{t}} + ce^{-\gamma\sqrt{t}} \end{cases}$$

has the *same* collection of phase paths as the system of the previous two exercises.

§ 6.2 The entirety of phase paths (which you cannot draw, since you’d fill up your phase plane with ink!) makes up the **phase flow** of the system. The last exercise showed that the phase flow only determines a dynamical system up to a monotone increasing transformation of time. In exercise 6.2, you saw that distinct paths can converge at the **omega point** (here: $(0, 0)$) which is attained in the infinite future. This is generally true of deterministic autonomous systems: paths can only coincide (or intersect) at omega points or **alpha points**, which are attained in the infinite past (that is, limiting points as $t \rightarrow -\infty$).

Exercise 6.4 Can you explain why phase paths can only coincide (intersect) in alpha or omega points?

Exercise 6.5 In exercise 6.2, you drew in arrowheads. In the system

$$\begin{cases} \dot{x}_1(t) &= F_1(x_1(t), x_2(t)) \\ \dot{x}_2(t) &= F_2(x_1(t), x_2(t)) \end{cases} \quad (6.1)$$

relate the direction of these arrows to the signs of F_1 and F_2 . (Hint: consider the ‘direction vector’ $[F_1, F_2]^T$.)

§ 6.3 A **phase portrait** is a picture of the phase plane together with selected phase paths, so as to convey an impression of the phase flow of the autonomous system (6.1). Of special importance in constructing a phase portrait are the critical points, where $F_1(x_1, x_2) = F_2(x_1, x_2) = 0$, and the **nullclines**. The **x_1 -nullcline** is the set of points where $F_1(x_1, x_2) = 0$; the **x_2 -nullcline** is the set of points where $F_2(x_1, x_2) = 0$.

Exercise 6.6 Explain why the critical points are at the intersection of the x_1 -nullcline and the x_2 -nullcline.

Exercise 6.7 Suppose that a path crosses the x_1 -nullcline at a point which is *not* a critical point. What can you say about the arrowhead at that point? (Hint: consider the above definition of the x_1 -nullcline and look at the ‘direction vector’ $[F_1, F_2]^T$.)

Exercise 6.8 What can you say about the arrowheads at the critical points?

Exercise 6.9 Consider the following autonomous system for $x_1 \geq 0$, $x_2 \geq 0$:

$$\begin{cases} \dot{x}_1 &= \rho x_1 - ax_1^2 - bx_1x_2 \\ \dot{x}_2 &= \mu x_2 - cx_2^2 - dx_1x_2 \end{cases} \quad (6.2)$$

(x_1 and x_2 are functions of time t , as before; this is not explicitly indicated here for the sake of brevity). Show that the x_1 -nullcline consists of (i) the x_2 -axis and (ii) a straight line connecting the point $\frac{\rho}{b}$ on the x_2 -axis to the point $\frac{\rho}{a}$ on the x_1 -axis.

Exercise 6.10 For system (6.2), find the x_2 -nullcline. (Hint: it consists of two straight lines, much like the x_1 -nullcline.)

Exercise 6.11 Sketch the phase plane (for $x_1 \geq 0$, $x_2 \geq 0$) and sketch the two nullclines, using different colours; assume $\frac{\rho}{a} > \frac{\mu}{d}$, $\frac{\mu}{c} > \frac{\rho}{b}$. Indicate the critical points.

Exercise 6.12 In your sketch of the previous exercise, draw short vertical (horizontal) lines on the x_1 -nullcline (x_2 -nullcline), and add arrowheads in the appropriate direction. (Hint: consider the ‘direction vector’ $[F_1, F_2]^T$.)

Exercise 6.13 In your sketch of the previous exercises, you should have four regions bounded by axes and nullclines. In each of those, consider the *signs* of F_1 and F_2 , and draw a little arrow pointing NE, SE, SW, or NW to indicate the approximate direction of the phase flow.

Exercise 6.14 In your sketch of the previous exercises, can you tell which critical points are alpha points, and which are omega points? (Hint: three are straightforward. One is a mixed alpha/omega point.)

§ 6.4 The last exercise shows that the classification of alpha and omega points is not always straightforward. Things are somewhat easier when, in system (6.1), F_1 and F_2 depend linearly on their arguments.

Exercise 6.15 Consider the system

$$\begin{cases} x_1 = \xi_1 e^{\lambda_1 t} \\ x_2 = \xi_2 e^{\lambda_2 t} \end{cases} \quad (6.3)$$

Draw a phase portrait for the case where $\lambda_1 = \lambda_2 < 0$.

Exercise 6.16 Draw a phase portrait for system (6.3) in the case where $\lambda_1 = \lambda_2 > 0$.

Exercise 6.17 Draw a phase portrait for system (6.3) in the case where $\lambda_1 \neq \lambda_2$, $\lambda_1 < 0$, $\lambda_2 < 0$.

Exercise 6.18 Draw a phase portrait for system (6.3) in the case where $\lambda_1 \neq \lambda_2$, $\lambda_1 < 0$, $\lambda_2 > 0$.

Exercise 6.19 Verify that system (6.3) satisfies the differential equation

$$\dot{x} = \mathbf{D} \cdot x$$

where $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$.

Exercise 6.20 With x given by system (6.3), let

$$\begin{cases} u_1 = \alpha x_1 + \beta x_2 \\ u_2 = \gamma x_1 + \delta x_2 \end{cases} \quad (6.4)$$

or, briefly $u = \mathbf{R} \cdot x$, where \mathbf{R} is a 2×2 matrix collecting the coefficients of system (6.4). Verify that

$$\dot{u} = \mathbf{R}^{-1} \cdot \mathbf{D} \cdot \mathbf{R} \cdot u. \quad (6.5)$$

Exercise 6.21 Verify that system (6.4) can be rewritten in manifest vector form:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = c_1 \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} e^{\lambda_1 t} + c_2 \begin{bmatrix} \beta \\ \delta \end{bmatrix} e^{\lambda_2 t}$$

where the coefficients c_1 and c_2 are determined by the boundary conditions (can you give an explicit formula for this dependence?).

Exercise 6.22 Can you describe in qualitative terms how the transformation \mathbf{R} in the previous exercise alters the x -phase flow to give the u -phase flow?

§ 6.5 In exercise 6.20 you saw that solving a linear two-dimensional autonomous system of the form

$$\dot{x} = \mathbf{A} \cdot x \quad (6.6)$$

(where \mathbf{A} is a 2×2 matrix called the **system matrix**) is easy if we can rewrite \mathbf{A} as follows:

$$\mathbf{A} = \mathbf{S}^{-1} \cdot \mathbf{D} \cdot \mathbf{S} .$$

The critical point is the origin, and its alpha/omega status depends on the diagonal elements of \mathbf{D} , as illustrated by exercises 6.15–6.18.

Exercise 6.23 Verify that the solution, in manifest vector form, of system (6.6) is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = c_1 \begin{bmatrix} s_{11} \\ s_{21} \end{bmatrix} e^{\lambda_1 t} + c_2 \begin{bmatrix} s_{12} \\ s_{22} \end{bmatrix} e^{\lambda_2 t} \quad (6.7)$$

where the s s are elements of \mathbf{S} , and the λ s are the diagonal elements of \mathbf{D} . (Hint: cf. exercise 6.21.)

Exercise 6.24 Consider system (6.6) with initial conditions such that, in terms of system (6.7), you have $c_1 = 1$ and $c_2 = 0$. Verify that

$$\dot{x} = \lambda_1 \begin{bmatrix} s_{11} \\ s_{21} \end{bmatrix} e^{\lambda_1 t} = \lambda_1 x$$

and hence infer that $\mathbf{A} \cdot x = \lambda_1 x$.

§ 6.6 The result of the last exercise can be written

$$(\mathbf{A} - \lambda_1 \mathbf{I}) \cdot x = 0$$

(\mathbf{I} is the 2×2 identity matrix). In order that x is not the null vector everywhere (which would be a rather boring solution), the matrix on the left has to be non-invertible, which yields the equation

$$(a_{11} - \lambda_1)(a_{22} - \lambda_2) - a_{21}a_{12} = 0 . \quad (6.8)$$

The procedure for λ_2 is entirely analogous and yields the same equation.

Exercise 6.25 Solve equation (6.8) (Hint: it is a quadratic in λ ; solve it for λ using the *abc*-formula.)

Exercise 6.26 Let

$$x(t) = \begin{bmatrix} x_1(0) & \alpha t \\ x_2(0) & \beta t \end{bmatrix} e^{\mu t}$$

and verify that this solution corresponds to a differential equation of the form (6.6), by deriving an expression for the matrix \mathbf{A} and show that the procedure of the previous exercise yields a quadratic with a *repeated root* (i.e., the roots are the same and both equal λ).

§ 6.7 The idea of **linearization** is to approximate the behaviour of the general non-linear system (6.1) in the neighbourhood of a critical point by a linear system.

Exercise 6.27 If \bar{x} is a critical point of system (6.1) (that is, $F_1(\bar{x}_1, \bar{x}_2) = F_2(\bar{x}_1, \bar{x}_2) = 0$) and the deviation from the critical point is z (that is, $z_1 = x_1 - \bar{x}_1$, $z_2 = x_2 - \bar{x}_2$) can you show that a first-order approximation to the dynamics of the deviation is

$$\dot{z} = \begin{bmatrix} F_{1\{r1\}}(\bar{x}_1, \bar{x}_2) & F_{1\{r2\}}(\bar{x}_1, \bar{x}_2) \\ F_{2\{r1\}}(\bar{x}_1, \bar{x}_2) & F_{2\{r2\}}(\bar{x}_1, \bar{x}_2) \end{bmatrix} \cdot z$$

where the matrix elements are the partial derivatives of the functions F_1 and F_2 ?

Exercise 6.28 Consider once more the system of exercise 6.9

$$\begin{cases} \dot{x}_1 = F_1(x_1, x_2) = \rho x_1 - ax_1^2 - bx_1x_2 \\ \dot{x}_2 = F_2(x_1, x_2) = \mu x_2 - cx_2^2 - dx_1x_2 \end{cases} \quad (6.9)$$

Calculate the derivatives of F_1 with respect to x_1 and x_2 ; do the same with F_2 , and collect them in a 2×2 matrix as in the previous exercise.

Exercise 6.29 System (6.2) has the critical point $(\bar{x}_1, \bar{x}_2) = (0, 0)$. Substitute these values in the matrix you obtained in the previous exercise, and solve equation (6.8) for this matrix. Show that both λ s are positive. Can you explain why this means that the critical point $(0, 0)$ is an alpha point?

Exercise 6.30 Repeat the previous exercise for the critical point $(\bar{x}_1, \bar{x}_2) = (0, \mu/c)$. Show that both λ s are negative. Can you explain why this means that this critical point is an omega point?

Exercise 6.31 Can you show, for system (6.2), that the critical point $(\bar{x}_1, \bar{x}_2) = (\rho/a, 0)$ is an omega point?

Exercise 6.32 Can you show, for system (6.2), that the critical point

$$\bar{x}_1 = \frac{b\mu - c\rho}{bd - ac} \quad \bar{x}_2 = \frac{d\rho - a\mu}{bd - ac}$$

is a mixed alpha/omega point?

Exercise 6.33 Draw the phase portrait of the following system

$$\begin{cases} \dot{x}_1 &= \mu x_1 + x_2 - x_1(x_1^2 + x_2^2) \\ \dot{x}_2 &= -x_1 + \mu x_2 - x_2(x_1^2 + x_2^2) \end{cases}$$

for (i) the case $\mu \leq 0$ and (ii) the case $\mu > 0$. (Hint: in polar coordinates the equations become $\dot{r} = r(\mu - r^2)$, $\dot{\theta} = -1$.)

Exercise 6.34 Can you generalize the results of the previous exercise to systems of the following form?

$$\begin{cases} \dot{x}_1 &= \mu x_1 + x_2 - x_1 f(r) \\ \dot{x}_2 &= -x_1 + \mu x_2 - x_2 f(r) \end{cases}$$

where $r = \sqrt{x_1^2 + x_2^2}$, the functions $f(r)$ and $f'(r)$ are continuous for $r \geq 0$, $f(0) = 0$ and $f'(r) > 0$ for $r \geq 0$.

7

Dynamics in discrete time

For some problems, a description of the process in discrete steps is more appropriate.

§ 7.1 In a **discrete-time** dynamical system, only denumerably many values for t_0, t_1, t_2, \dots are considered, together with the associated **state sequence** $x(t_0), x(t_1), x(t_2), \dots$

Exercise 7.1 Contrast continuous-time versus discrete-time dynamical systems. How fundamental is this distinction? Can the same state-transition function arise in both a continuous-time and a discrete-time system?

§ 7.2 The state-transition function is usually written as the **k th iterate map**, defined as follows:

$$F^{[k]}(i, \xi) = S(t_{i+k}, t_i, \xi) \quad (7.1)$$

Exercise 7.2 Complete: $F^{[0]}(i, \xi) = \dots$

Exercise 7.3 Verify that the second consistency condition now reads:

$$F^{[m]}(i, \xi) = F^{[m-k]}(i+k, F^{[k]}(i, \xi))$$

for $k = 0, 1, 2, \dots, m$.

§ 7.3 Again, the autonomous case is of special interest, where $F^{[k]}(i, \xi) \equiv F^{[k]}(\xi) \forall i$ (“the same map for every step”). A **critical point (of the k th iterate)** \bar{x} satisfies $F^{[k]}(\bar{x}) = \bar{x}$. Critical points of first iterate maps are called **fixpoints**.

Exercise 7.4 Draw a graph of $F^{[1]}(x)$ with $x \in \mathbb{R}$ (any one you like). Also draw the diagonal line $y = x$. Indicate the fixpoints (in any) in your graph.

Exercise 7.5 Suppose that, given a fixpoint $\bar{x} \in \mathbb{R}$ and some $\delta > 0$, the first iterate map satisfies the following:

$$\left| \frac{F^{[1]}(\bar{x} + \epsilon) - \bar{x}}{\epsilon} \right| < 1$$

for all $|\epsilon| < \delta$. Deduce that \bar{x} is an omega point for all x_0 within a distance δ of \bar{x} . Can you find a similar condition for alpha points?

Exercise 7.6 Show that a fixpoint $\bar{x} \in \mathbb{R}$, satisfying $F^{[1]}(\bar{x}) = \bar{x}$, is also a critical point of $F^{[2]}$.

Exercise 7.7 Can you show that a fixpoint $\bar{x} \in \mathbb{R}$ is also a critical point of $F^{[k]}$ for $k \geq 2$?

Exercise 7.8 Suppose that $\bar{x} \in \mathbb{R}$ is a critical point of the second iterate. Then $F^{[1]}(\bar{x})$ is also a critical point of the second iterate; show this. (Hint: you have $F^{[2]}(\bar{x}) = \bar{x}$; $F^{[2]}(\bar{x})$ may, but need not, be equal to \bar{x} .)

Exercise 7.9 Show that a critical point $\bar{x} \in \mathbb{R}$ of the second iterate which is *not* also a fixpoint corresponds to a periodic state sequence

$$\dots, \bar{x}, F^{[1]}(\bar{x}), \bar{x}, F^{[1]}(\bar{x}), \bar{x}, F^{[1]}(\bar{x}), \dots$$

Exercise 7.10 Can you show that a critical point $\bar{x} \in \mathbb{R}$ of the p th iterate, where p is a prime number, which is *not* also a fixpoint corresponds to a p -periodic state sequence? (Hint: why the emphasis on primes here?)

Exercise 7.11 Consider a q -dimensional discrete-time dynamical system ($x \in \mathbb{R}^q$) defined by the following first-iterate map:

$$F^{[1]}(x) = \begin{bmatrix} \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_q x_q \\ x_1 \\ x_2 \\ \vdots \\ x_{q-1} \end{bmatrix} \quad (7.2)$$

and let μ be a root of the polynomial equation

$$z^q = \gamma_1 z^{q-1} + \gamma_2 z^{q-2} + \dots + \gamma_{q-1} z + \gamma_q.$$

Then verify that the following equation defines a possible state sequence for this system (indexed by n):

$$k \begin{bmatrix} \mu^n \\ \mu^{n-1} \\ \mu^{n-2} \\ \vdots \\ \mu^{n-q+1} \end{bmatrix}$$

where k is an arbitrary constant.

Exercise 7.12 Suppose that the system of the previous exercise is modified as follows, to become non-autonomous:

$$F^{n,[1]}(x) = \begin{bmatrix} \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_q x_q + \kappa \lambda^n \\ x_1 \\ x_2 \\ \vdots \\ x_{q-1} \end{bmatrix} \quad (7.3)$$

with an additional **forcing term** $\kappa \lambda^n$. Verify that the following equation defines a possible state sequence for this system (indexed by n):

$$k \begin{bmatrix} \kappa \lambda^{n+q} / (\lambda^q - \gamma_1 \lambda^{q-1} - \gamma_2 \lambda^{q-2} - \dots - \gamma_q) \\ \kappa \lambda^{n-1+q} / (\lambda^q - \gamma_1 \lambda^{q-1} - \gamma_2 \lambda^{q-2} - \dots - \gamma_q) \\ \kappa \lambda^{n-2+q} / (\lambda^q - \gamma_1 \lambda^{q-1} - \gamma_2 \lambda^{q-2} - \dots - \gamma_q) \\ \vdots \\ \kappa \lambda^{n+1} / (\lambda^q - \gamma_1 \lambda^{q-1} - \gamma_2 \lambda^{q-2} - \dots - \gamma_q) \end{bmatrix}.$$

Exercise 7.13 If $x_H(n)$ denotes the state sequence given in exercise 7.11 and $x_{PS}(n)$ denotes the state sequence given in exercise 7.12, show that $x_H(n) + cx_{PS}(n)$, where c is an arbitrary constant, is also a possible state sequence for system (7.3).

Exercise 7.14 In continuous-time dynamics, you encounter terms like $e^{\lambda t}$ where $\lambda > 0$ means ‘expanding’ as t increases (while $\lambda < 0$ means ‘shrinking’); whereas in discrete dynamics, you encounter terms like λ^t , where $\lambda > 1$ means ‘expanding’ as t increases (while $\lambda < 1$ means ‘shrinking’). How come the cross-over occurs at zero in one case, and at unity in the other? (Hint: the rule $e^{ab} = (e^a)^b$ might help.)

Exercise 7.15 Consider a 4-dimensional discrete dynamical system ($x \in \mathbb{R}^4$) defined by the following first-iterate map:

$$F^{[1]}(x) = \begin{bmatrix} \alpha x_1 + \gamma x_2 + \gamma x_3 + \gamma x_4 \\ \sigma x_2 \\ \sigma x_3 \\ \sigma x_4 \end{bmatrix}. \quad (7.4)$$

Verify the following matrix equation:

$$\begin{bmatrix} \alpha & \gamma & \gamma & \gamma \\ 0 & \sigma & 0 & 0 \\ 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & \sigma \end{bmatrix}^n = \begin{bmatrix} \alpha^n & \gamma \frac{\alpha^n - \sigma^n}{\alpha - \sigma} & \gamma \frac{\alpha^n - \sigma^n}{\alpha - \sigma} & \gamma \frac{\alpha^n - \sigma^n}{\alpha - \sigma} \\ 0 & \sigma^n & 0 & 0 \\ 0 & 0 & \sigma^n & 0 \\ 0 & 0 & 0 & \sigma^n \end{bmatrix}$$

and hence give an expression for the state-transition function.

Modelling methodology

The modelling recipe must be used with care. In particular, due attention must be made to the assumptions that underlie the equations of the model. The exercises in this chapter highlight some of the caveats and pitfalls.

Exercise 8.1 List the ways in which people use the word “model”. Describe what (if anything) these meanings have in common.

Exercise 8.2 Review the “modelling recipe” and the **empirical cycle** (chapter 2).

§ 8.1 Loosely speaking, a **model**¹ is a piece of mathematics, interpreted in terms of observable real-world phenomena. The mathematical component of the model is called the **model proper**.

Exercise 8.3 Could any “model proper” (uninterpreted mathematics) be made part of a model by means of a suitable interpretation?

Exercise 8.4 Can you define **observable phenomenon**? (Hint: how do the terms of art of (empirical) scientific discourse acquire their meaning? How are these meanings affected by advances in experimental techniques? How theory-laden are data and experimental observations?)

Exercise 8.5 Is it necessary that *everything* (say, every variable and parameter) in the model proper corresponds to an observable phenomenon?

Exercise 8.6 Can you motivate the following statement? The interpretation of a model can be split into two components, (i) a list of identifications between parts of the model proper and parts of the world; (ii) the observational interface, the experimental approach to the world.

Exercise 8.7 Can you explain why the identifications (in the sense of the previous exercise) are said to constitute a **theoretical hypothesis**?

§ 8.2 In practice, you start with (imperfect) knowledge about the real-world, and you attempt to formulate a suitable model proper. This knowledge can take the form of a list of statements, called the **assumptions underlying the model**.

Exercise 8.8 The assumptions are said to be **contingent**, which means that they may be wrong, depending on what is actually the case². Is *wrong* an all-or-nothing property, or are there shades of gray? Is the model proper contingent?

¹‘Model’ will usually mean ‘mathematical model’, although some comments also apply to other types of model such as scale models.

²Die Welt ist alles, was der Fall ist.

Exercise 8.9 Comment on the claim that every model has exactly *one* assumption, namely that: Model proper \mathcal{M} represents real world phenomenon \mathcal{P} according to a list of identifications \mathcal{I} (with specifications of \mathcal{M} , \mathcal{P} and \mathcal{I} supplied).

Exercise 8.10 What are the advantages of breaking down the “monolithic” assumption of the previous exercise into a long list of detailed & specific assumptions? What are the disadvantages?

§ 8.3 An assumption that is not spelled out by the modeller is said to be **implicit** (antonym: **explicit**). When it becomes important, an implicit assumption is usually called a **hidden assumption**.

Exercise 8.11 Why would modellers leave assumptions implicit?

§ 8.4 It could happen that the conclusions of the model analysis may change dramatically if an assumption is altered only slightly; such an assumption is said to be **strong** (antonym: **weak**).

Exercise 8.12 Motivate a preference for weak assumptions.

Exercise 8.13 An assumption may strongly confine or reduce the range of real-world phenomena described by the model; such an assumption is said to be **stringent** (antonym: **relaxed**).

Exercise 8.14 A common strategy is to start with stringent assumptions, and as you go round the model cycle, to relax these assumptions (i.e. replace them by more relaxed versions). Why is this a sound strategy? (Hint: stringent assumptions tend to be easier to state and to analyse.)

§ 8.5 Assumptions (within a given model!) should not contradict one another; this the requirement of **consistency**.

Exercise 8.15 Should the consistency requirement be extended to cover the demand that the assumptions do not contradict the available data?

Exercise 8.16 Explain why it might well happen that the assumptions of a model are consistent with each other as well as the available data, but the conclusions of the model analysis fail to do so. Do you consider this to be a failure or a success of the modelling exercise as a whole?

Exercise 8.17 Discuss (in a group) the thorny issue of **coherence** between the assumptions. (Hint: should the assumptions be equally stringent, equally strong? Why (not)?)

§ 8.6 As a modeller, you have to decide how complex you want your model to be.

Exercise 8.18 Can you define or characterize the concept of complexity for a model?

Exercise 8.19 Consider the following three possible aims of a model:

- (i) to predict the (future) behaviour of a real-world phenomenon;
 - (ii) to control the behaviour of a real-world phenomenon;
 - (iii) to understand the behaviour of a real-world phenomenon;
- and arrange these three aims in the order of required model complexity.

Exercise 8.20 What positive purpose may be served by deliberately *leaving out* certain bits of knowledge (i.e. deliberately not representing certain known mechanisms in the model proper)? Discuss advantages and disadvantages.

Exercise 8.21 Discuss ways of tackling real-world phenomena that would appear, in one sense or another, too complex to capture in a model. (Hint: think of scales in time and space, emergent properties, statistical mechanics. As regards examples of domains where this problem arises in the life sciences; think of neurones÷brains, individuals÷ecosystems, molecules÷cells.)

Exercise 8.22 Discuss (in a group) the following common pitfalls:

- (i) wishful thinking—tailoring model assumptions to attractive equations;
 - (ii) misleading names for model quantities—inviting overinterpretation;
 - (iii) imputing goals, objectives³ and functions⁴ to living things—adopting an engineer’s point of view.
- (Hint: reflect that, if used judiciously, a pitfall may turn into a useful trick.)

§ 8.7 To derive results & conclusions from the model, its behaviour *vis a vis* the real-world phenomenon must be probed. To this end, the model proper is subjected to mathematical analysis. An (almost obligatory) part of this is **simulation**⁵, which is a numerical (and usually graphical) demonstration of the behaviour of the model.

Exercise 8.23 Discuss the distinction between analytical and non-analytical, tractability, and review standard numerical techniques. (Hint: why is it permitted⁶ to define $y = J_0(x)$ as the solution to $x^2\ddot{y} + x\dot{y} = -x^2y$ whereas you cannot very well declare your problem to be solved by $y = \mathcal{S}_{\text{my problem}}(x)$ where $\mathcal{S}_{\text{my problem}}$ is defined to be just that?)

³*Teleology* is the doctrine that everything in nature was or is designed with a view to achieving a (God-given?) set of objectives (from the Greek *telos* meaning ‘purpose’ or ‘end’).

⁴The *biological function* of a certain structure or process in an organism is constituted by whatever effects the presence of that structure or process happens to have within the context of the organism. If these effects tend to aid the lifetime reproductive success of the individuals in which it happens to find itself, it will persist on an evolutionary timescale.

⁵The term ‘simulation’ is usually taken to imply the use of numerical techniques, in particular for the solution of boundary value problems.

⁶Quod licet iovi non licet bovi.

Exercise 8.24 Suppose that ξ , v , and ζ are related by

$$\zeta = \frac{\xi}{v} \tag{8.1}$$

but the modeller is unaware of this simple fact and generates a series graphs in which ζ is plotted against v , perhaps at a variety of values of ξ . Will the modeller be able to infer from his graphics that the three quantities are interrelated according to equation (8.1)?

§ 8.8 The last exercise illustrates a problem that can become very serious when you do not know what relationship you are looking for, and when there are many parameters and parameter values to study. The moral is that it is a good idea to try to penetrate the model by analytical means as far as your mathematical prowess (or that of friendly mathematicians) will take you, because it is not always easy to uncover important interdependencies. **Dimensional analysis** is a useful technique which helps to alleviate this problem. This is the subject of the next chapter.

Empirical dimensions and units

All terms in an equations should have the same empirical dimension. This can often be used as a quick check to see if you have made mistakes. An analysis of the empirical dimensions can be used to eliminate parameters from your model, which often facilitates the exploration of the model's behaviour to a considerable extent.

§ 9.1 (Empirical dimension) Two quantities (P and Q) are said to have the **same (empirical) dimension**, written briefly as $P \overset{d}{\sim} Q$, iff:

1. Quantities P and Q can be meaningfully equated or compared ($=$, $<$, $>$).
2. Quantities P and Q have a meaningful sum or difference.

The **empirical dimension** (briefly: **dimension**¹) as an equivalence class of quantities under $\overset{d}{\sim}$. Finally, we say that the dimension of given quantity Q is the particular equivalence class (under $\overset{d}{\sim}$) to which Q belongs.

Exercise 9.1 Verify that ‘same dimension’ is an equivalence relation.

§ 9.2 To establish whether two quantities have the same dimension, we need to ascertain whether or not $P \overset{d}{\sim} Q$.

Exercise 9.2 Why is it not really a matter of mathematics whether $P \overset{d}{\sim} Q$ is valid? (Hint: the crux is the word ‘meaningful’, which hinges on interpretation, not model proper.)

Exercise 9.3 Motivate the convention

$$\frac{P}{P} \overset{d}{\sim} \frac{Q}{Q}$$

for any two quantities P and Q , regardless of whether $P \overset{d}{\sim} Q$.

§ 9.3 The convention of the last exercise induces an equivalence class denoted by 1: we write $(P/P) \overset{d}{\sim} 1$ and call the equivalence class thus formed **pure number**.

Exercise 9.4 Explain why quantities in this special class are referred to as **dimensionless**.

¹Be careful that you do not confuse this use of the word ‘dimension’ with other meanings in mathematics, such as that of linear algebra; in the proof of Buckingham’s theorem we will be using the term in both senses.

Exercise 9.5 Why is the (commonly heard) statement that “quantities, naturally interpreted as numbers, are always dimensionless” in fact fallacious? (Hint: refer to absolute scales of measurement, explained below.)

Exercise 9.6 Suppose your model has M scalar quantities $\{Q_1, \dots, Q_M\}$; these quantities are the variables and parameters in the model proper. Using the $Q_i \stackrel{d}{\sim} Q_j$ test, you can apportion these quantities among a number of empirical dimensions. Show that the number of distinct empirical dimensions will be anywhere between 1 (inclusive) and M (inclusive).

§ 9.4 (A basis of empirical dimensions) Consider m distinct dimensions, each represented by some quantity D_i , which has the following properties:

$$\text{for every } Q_j, j = 1, \dots, M, \quad Q_j \stackrel{d}{\sim} \prod_{i=1}^m D_i^{r_i}$$

for some vector $[r_1, \dots, r_m]$, while for *no choice* of $\{r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_m\}$ it is true that

$$\text{for any } D_i, i = 1, \dots, m, \quad D_i \stackrel{d}{\sim} \prod_{k \neq i} D_k^{r_k}$$

(the r_i are real numbers). Such a set $\{D_1, \dots, D_m\}$ constitutes a **basis** of empirical dimensions for your model.

Exercise 9.7 Verify that the first condition ensures sufficiency of the basis for the model. whereas the second is an analog of linear independence.

§ 9.5 We may refer to these basis dimensions (which you will recall are equivalence classes) with an appropriate sans serif name that recalls the interpretation of the quantities in the class, for example time. To indicate that the dimension of the quantity t is time one usually writes $\dim\{t\} = \text{time}$ instead of $t \in \text{time}$.

Exercise 9.8 Can you make sense of a formula such as $\text{velocity} = \text{length}/\text{time}$? (Hint: with $v \in \text{velocity}$, we have $v \propto s/t$ where $s \in \text{length}$ and $t \in \text{time}$.)

Exercise 9.9 Show that the following formulæ hold:

$$\dim\{x + y\} = \dim\{x\} = \dim\{y\} \tag{9.1}$$

$$\dim\{xy\} = \dim\{x\} + \dim\{y\} \tag{9.2}$$

$$\dim\{x^n\} = n \dim\{x\} \quad \text{where} \quad \dim\{n\} = 1. \tag{9.3}$$

Exercise 9.10 Show that $\ln\{x\}$ is well-defined only if $\dim\{x\} = 1$.

Exercise 9.11 For chemists: why is the Henderson-Hasselbalch relation dimensionally correct, despite appearances and the previous exercise? Is the claim that “the pK has no units” defensible?

Exercise 9.12 For physicists: verify that {length, time, mass} is a basis of classical mechanics. Can you show that {velocity, force, energy} is a basis, too? What about {velocity, momentum, action}?

§ 9.6 The last exercise shows that a dimension basis need not be unique. A systematic procedure is available to discover a dimension basis given a problem with M quantities apportioned over a set of dimensions, but inspection and intuition suffice in many problems, as the simple example in the following exercises makes clear.

Exercise 9.13 Let the following model properly represent the dynamics of the amount of mRNA molecules transcribed from a given gene in a cell:

$$\frac{d}{dt}N = M - \lambda N \quad (9.4)$$

with the following interpretations: N is the number of mRNA molecules; M is the rate at which mRNA molecules are produced by transcription; and λ is the rate at which the mRNA molecules are being degraded. At time $t = 0$, there are N_0 molecules present in the cell. Name the *three* parameters of this model.

Exercise 9.14 Show equation (9.4) has the following solution:

$$N(t) = \frac{M}{\lambda} + \left(N_0 - \frac{M}{\lambda} \right) \exp\{-\lambda t\}. \quad (9.5)$$

§ 9.7 Equation (9.5) is simple enough for its characteristic properties to be gleaned by inspection. However, the solution of a differential equation is not always that simple, or may not be analytically available and must be numerically evaluated. In both cases, you would study the behaviour of the model by plotting curves for specific numerical settings of the parameter. If you plot equation (9.5) for various settings of M , λ , and N_0 , you get a tangled mess of different curves. In models with more than three parameters, numerically exploring the behaviour at all parameter settings readily becomes a hopeless task. However, the parameter count can be reduced by as many dimensions as there are. The idea is to use the parameters as ‘natural units.’

Exercise 9.15 Choose time and number of molecules as the basis dimensions. Establish the following:

$$\begin{aligned} \dim\{N\} &= \text{number of molecules} \\ \dim\{t\} &= \text{time} \\ \dim\{M\} &= \text{number of molecules} \cdot \text{time}^{-1} \\ \dim\{\lambda\} &= \text{time}^{-1} \\ \dim\{N_0\} &= \text{number of molecules} \end{aligned}$$

(Hint: use the dimension test.)

§ 9.8 A **unit** for an empirical dimension d is an element of d , say u_d , which serves to report the magnitude of all other $x \in d$ in terms of the dimensionless ratio x/u_d .

Exercise 9.16 Verify that x/u_d is dimensionless. (Hint: the thing being measured and the unit must share empirical dimension.)

§ 9.9 It is perfectly legitimate to choose quantities in the model as units.

Exercise 9.17 Show that $\lambda^{-1} \in \text{time}$ and that λ^{-1} can therefore serve as a unit of time.

Exercise 9.18 Show that $M\lambda^{-1}$ can serve as a unit of number of molecules.

Exercise 9.19 Show that $t^* = t\lambda$ and $N^* = NM^{-1}\lambda$ are dimensionless.

Exercise 9.20 Derive the following scaled version of equation (9.4):

$$\frac{d}{dt^*}N^* = 1 - N^* \quad (9.6)$$

Exercise 9.21 Show that one free dimensionless parameter remains: $N_0^* = N_0M^{-1}\lambda$.

Exercise 9.22 Sketch the solution of equation (9.6) for various values of N_0^* .

Exercise 9.23 Show that alternative choice of units are N_0 for number of molecules and $M^{-1}N_0^{-1}$ for time; then show that the remaining free dimensionless parameter is λ^* . (Hint: $M^{-1}N_0^{-1}t$ is dimensionless.)

Exercise 9.24 Can you think of reasons why you might prefer to (not) use the alternative units of the previous exercise.

§ 9.10 It is noteworthy that (scaled versions of) quantities remain in the dimensionless model whenever these quantities have *not* been used to define the units. This gives you a general guideline for choosing units: do *not* involve the parameters which are of interest.

Exercise 9.25 Suppose that the model now also deals with the mRNA transcript of a second gene. Now you are dealing with two genes, say I and II. Would number of mRNA molecules of species I and number of mRNA molecules of species II count as two distinct empirical dimensions, or would both variables belong to the same class number of mRNA molecules?

§ 9.11 (**Buckingham's theorem**) The foregoing example suggests that by scaling, you can reduce the number of independent parameters by as many dimensions as there are independent dimensions in your model. **Buckingham's theorem** asserts that this is indeed always so. *Every relationship*

$$f(Q_1, \dots, Q_M) = 0 \quad (9.7)$$

among the M quantities in a model proper can be rewritten in dimensionless form

$$\bar{f}(Z_1, \dots, Z_n) = 0 \quad Z_i \stackrel{d}{\sim} 1, \quad i = 1, \dots, n \quad (9.8)$$

as a relationship \bar{f} among $n = M - m$ dimensionless quantities.

The following series of exercises establish a proof of this theorem.

Exercise 9.26 Show that, given a vector $\mathbf{r} \in \mathbf{R}^M$, you can form a quantity $W(\mathbf{r})$ which is expressible in the basis dimensions:

$$W(\mathbf{r}) = \prod_{j=1}^M Q_j^{r_j} \stackrel{d}{\sim} \prod_{i=1}^m D_i^{s_i}. \quad (9.9)$$

(Hint: recall the defining properties of a dimension basis, § 9.4.)

Exercise 9.27 Verify that equation (9.9) induces a linear transformation

$$T : \mathbf{R}^M \mapsto \mathbf{R}^m.$$

(Hint: the vector $\mathbf{r} \in \mathbf{R}^M$ is mapped to a vector $\mathbf{s} \in \mathbf{R}^m$.)

Exercise 9.28 Verify that the above transformation T has a kernel $\text{Nul } T$ of dimension $n = M - m$. (Warning: the term ‘dimension’ is used in the sense of linear algebra here!)

Exercise 9.29 Verify that $\mathbf{s} = \mathbf{0}$ means that the quantity $W(\mathbf{r})$ is dimensionless.

Exercise 9.30 Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for $\text{Nul } T$ and verify that this induces n dimensionless quantities

$$Z_k = \prod_{j=1}^M Q_j^{b_{jk}} \stackrel{d}{\sim} 1 \quad k = 1, \dots, n.$$

§ 9.12 Adding vectors $\{\mathbf{b}_{n+1}, \dots, \mathbf{b}_M\}$ to the basis of $\text{Nul } T$, you can form a complete basis for \mathbf{R}^M . With each of these additional basis vectors, associate a quantity

$$Z_k = \prod_{j=1}^M Q_j^{b_{jk}} \not\stackrel{d}{\sim} 1 \quad k = n + 1, \dots, M.$$

Exercise 9.31 Show that none of these quantities Z_k is dimensionless. (Hint: the kernel of T is already spanned by the first n basis vectors.)

Exercise 9.32 Show that each Q_j ($j = 1, \dots, M$) can be expressed uniquely as a product of powers of the $\{Z_k\}_{k=1}^M$. (Hint: the matrix collecting the basis vectors is invertible.)

Exercise 9.33 Show that each relation (9.7) can be rewritten in the form $\bar{f}(Z_1, \dots, Z_M) = 0$. (Hint: the previous exercise.)

§ 9.13 If we vary units arbitrarily, $f(Q_1, \dots, Q_M) = 0$ remains valid in virtue of compensatory changes among all the Q_j ; on the other hand, the first n Z_j are immune from such compensation since they are dimensionless.

Exercise 9.34 Argue that such immunity implies that you can derive $\bar{f} \neq 0$ by varying units, unless \bar{f} only depends on Z_1, \dots, Z_n .

Exercise 9.35 Show that \bar{f} only depends on Z_1, \dots, Z_n . (Hint: $\bar{f} = 0$ is valid, no matter what the units are.)

Exercise 9.36 Show that $\bar{f}(Z_1, \dots, Z_M) = 0$ reduces to $\bar{f}(Z_1, \dots, Z_n) = 0$.

§ 9.14 In the last exercise you proved Buckingham's theorem, which tells us that we need to work with only n quantities rather than M . This greatly simplifies the amount of work to do, and, more importantly, essential relationships are preserved when non-dimensionalizing. In most cases there are at least as many parameters as there are independent dimensions in the unscaled model, and one is left with some remaining scaled parameters.

10

Scales of measurement

All measurements are done on some scale or other. It is important to be aware of the strength of this scale.

Exercise 10.1 Suppose a trainspotter spots a number of engines during the course of a windy autumn day. In the evening, the diligent anorak adds up all these numbers and calculates the average. Comment on the usefulness of this calculation.

§ 10.1 The solecism of the last exercise illustrates the issue of the **strength** of measurements: not all scales are measurements are strong enough to support all possible arithmetical operations.

§ 10.2 On a **nominal scale of measurement**, number is merely a name. No arithmetical operations make empirical sense on a nominal scale. The only comparison possible within a nominal scale is for equality.

Exercise 10.2 Give examples of quantities measured on a nominal scale. (Hint: think of situations where categories of data are represented by integers.)

§ 10.3 On an **ordinal scale of measurement**, the order of the numbers now has a meaning. For instance, every pair of biological species may be assigned the numbers 1, 2, 3, 4, 5, or 6 according to whether they belong to the same genus, family, order, class, phylum, or kingdom.

Exercise 10.3 Verify ordinality: a pair of species characterized by 3 is more closely related than a pair characterized by 5 on this scale.

Exercise 10.4 Explain why the difference between pairs that score 4 and 5 cannot be meaningfully compared to another pair of pairs that score 2 and 3. (Hint: the numbers 1 to 6 might just as well have been 3, 12, 45, 700, 701, and 764980.)

Exercise 10.5 Is a species pair that scores 6 three times less closely related than a pair that scores 2?

Exercise 10.6 Can you meaningfully work with averages on an ordinal scale? Why (not)?

Exercise 10.7 (Transitivity) Let Or represent an ordinal scale, so that $Or(X)$ is the measurement ('score') of object X on this scale. Then verify that $Or(A) > Or(B)$ and $Or(B) > Or(C)$ together imply $Or(A) > Or(C)$.

Exercise 10.8 Can the concept of empirical dimension be applied to quantities that are measured on an ordinal scale?

§ 10.4 On an **interval scale of measurement**, differences between the numbers have an empirical meaning.

Exercise 10.9 Show that you can meaningfully do additions, subtractions and averaging on the interval scale.

Exercise 10.10 Argue why it makes sense to state that $15\text{ }^{\circ}\text{C}$ is the mean of $10\text{ }^{\circ}\text{C}$ and $20\text{ }^{\circ}\text{C}$. (Hint: think about mixing equal quantities of water of different temperatures.)

Exercise 10.11 Is it true that a temperature of $20\text{ }^{\circ}\text{C}$ is twice as high as $10\text{ }^{\circ}\text{C}$? Why (not)?

Exercise 10.12 Explain why proportions (ratios, quotients) between measurements are meaningless on an interval scale. (Hint: the zero of an interval scale is arbitrary.)

§ 10.5 On a **ratio scale of measurement**, differences and proportions (ratios, quotients) between the numbers have an empirical meaning. The zero is no longer arbitrary but expresses a natural null point. All arithmetic makes sense on the ratio scale. Only the unit is still arbitrary on the ratio scale. The ratio scale is the scale of most commonly used quantities in physics, chemistry and biology.

Exercise 10.13 Explain why most mistakes with scale strength are due to an erroneous assumption that the measurement at hand has the strength of a ratio scale.

§ 10.6 Only **absolute scales** are stronger than ratio scales. These scales are like the ratio scale, but the unit is no longer arbitrary.

Exercise 10.14 Can you see why a common mistake with quantities on an absolute scale is to consider them to be dimensionless?

Exercise 10.15 Indicate whether the following scales are nominal, ordinal, interval, ratio, or absolute: (i) mass in kilograms; (ii) Moh's scale for the hardness of minerals; (iii) temperature on the Kelvin scale; (iv) 'semi-quantitative' scales that express the cognitive state of an animal, such as degree of aggressiveness or IQ; (v) the temperature scales of Celsius, Réaumur and Fahrenheit; (vi) population size; (vii) time in seconds; (viii) the Beaufort scale of wind; (ix) the decibel scale of sound intensity; (x) the Richter scale.

Exercise 10.16 Verify that only quantities on an interval scale or stronger have an empirical dimension.

Exercise 10.17 Explain why a quantity that admits measurement on a scale of a given strength also admits measurement on a weaker scale. (Hint: everything can be expressed on a nominal scale, if only as 'measurement number x '.)

Exercise 10.18 Why would you prefer a weaker scale? (Hint: reliability.)

Exercise 10.19 Can you prove the following claims?

- (i) A nominal scale is invariant under any bijection (1-1 transformation).
- (ii) An ordinal scale is invariant under any monotone transformation.
- (iii) An interval scale is invariant under any transformation of the form $y = ax + b$.
- (iv) A ratio scale is invariant under any transformation of the form $y = ax$.
- (iv) An absolute scale is invariant only under the identity $y = x$.

11

Describing data with models

Any serious modelling involves a comparison with experiment.

§ 11.1 The combined result of measurements performed to chart or characterize a real-world phenomenon is called **the experimental data**¹. A major use of mathematical models is to order data and extract information from these data (cf. 8.19).

Exercise 11.1 Contrast and compare the following two ways in which a model interacts with data: data **source material** versus data as **test material**.

§ 11.2 When the model is developed independently of some particular data set, confrontation of the model with that data set amounts to the testing of a prediction.

Exercise 11.2 Is it essential that the test data set is collected only after the model has been developed?

§ 11.3 In **parameter fitting**², data serve both as source and test material, as parameter values are estimated while the model is being confronted with the data. Suppose that a model f describes a process variable y as a function of time t and three parameters α, β, γ :

$$y = f(t; \alpha, \beta, \gamma)$$

and suppose that a number n of data pairs (t_i, y_i) is available (the entire data set may be represented as (\mathbf{t}, \mathbf{y})).

Exercise 11.3 Suppose that there are i, j such that $t_i = t_j$. Do you expect $y_i = y_j$? If not, how do you account for the discrepancy?

Exercise 11.4 Do you expect the data point y_i to equal $f(t_i; \alpha, \beta, \gamma)$? If not, how do you account for the discrepancy?

§ 11.4 To quantify how strongly the model prediction confirms to (or deviates from) the data, in other words, to express the **goodness-of-fit**, the following **sum of squared errors** function S is widely used:

$$S(\alpha, \beta, \gamma; \mathbf{t}, \mathbf{y}) = \sum_{i=1}^n (f(t_i; \alpha, \beta, \gamma) - y_i)^2. \quad (11.1)$$

Exercise 11.5 Is the goodness-of-fit high or low when S is high?

¹The singular is *datum* which means ‘that which is given’.

²Sometimes also called **calibration**.

§ 11.5 The best-fit values $\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\}$ of the three parameters are those for which S is minimized. These are obtained by putting $\partial S/\partial\alpha = \partial S/\partial\beta = \partial S/\partial\gamma = 0$ and solving for the parameters.

Exercise 11.6 Discuss (numerical) procedures to determine the best-fit values.

Exercise 11.7 In general, you are looking for a minimum of S in a parameter space that has as many dimensions as you have parameters. Discuss the difficulties you might expect determining the *global* minimum of S .

§ 11.6 It often happens that the data comprise simultaneous measurements: thus you have data points of the form (t_i, x_i, y_i, z_i) is available. If the model describes all these state variables, you can extend the procedure to find the best fitting set of parameter values using all available data sets at once.

Exercise 11.8 Suppose that you have, as in the above example, three parameters and three state variables. The data can be represented as three graphics, plotting, respectively, data points of the form (t_i, x_i) , (t_i, y_i) , and (t_i, z_i) . How many parameters are you estimating “per curve”? How does this compare to a straight line (often touted as “the simplest possible model”)?

§ 11.7 The last exercise illustrates that **simultaneous fitting** can be very parsimonious.

§ 11.8 The method of minimizing the sum of squares with respect to the parameters can be justified as follows. Assume that the measurement error in y_i is normally distributed about the predicted value, while the t_i are known with perfect accuracy. Furthermore, the errors are taken to be: (i) unbiased relative to a ‘true’ model, that is, the expected value of every measurement is what is predicted by the model for the ‘true’ parameter values; (ii) independent; and of (iii) equal variance σ^2 . Then the joint probability density p of finding the measurements \mathbf{t}, \mathbf{y} given the model and its parameters is

$$p(\mathbf{y}; \mathbf{t}, \alpha, \beta, \gamma) = \prod_{i=1}^n (\sqrt{2\pi}\sigma)^{-1} \exp\{-([f(t_i; \alpha, \beta, \gamma) - y_i]/\sigma)^2/2\}.$$

Now you apply a conceptual twist: you view this p as the **likelihood** of the *parameters*, given the data. You are looking for the parameter values that maximize this likelihood.

Exercise 11.9 Show that you have to maximize

$$L(\alpha, \beta, \gamma; \mathbf{t}, \mathbf{y}) = -\sigma^{-2} \sum_{i=1}^n (f(t_i; \alpha, \beta, \gamma) - y_i)^2$$

with respect to the parameters. (Hint: take logarithms, discard terms that do not depend on the parameters.)

Exercise 11.10 Show that maximizing L amounts to minimizing the sum of squares as defined in equation (11.1).

Exercise 11.11 Verify that you do not need to know (or estimate) the value of σ .

§ 11.9 The same principles can be applied to different statistical models of the error, leading to other goodness-of-fit criteria.

§ 11.10 The log-likelihood argument can be extended to simultaneous data sets, as follows. Let x_i, y_i, z_i be three measurements at time t_i . Suppose that f, g, h are the solutions of the model that was proposed to describe the data. The log-likelihood argument then leads us to minimize

$$L(\alpha, \beta, \gamma; \mathbf{t}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{i=1}^n \left(\frac{(f(t_i; \alpha, \beta, \gamma) - x_i)^2}{\sigma_x^2} + \frac{(g(t_i; \alpha, \beta, \gamma) - y_i)^2}{\sigma_y^2} + \frac{(h(t_i; \alpha, \beta, \gamma) - z_i)^2}{\sigma_z^2} \right). \quad (11.2)$$

The only additional difficulty is that now the variances $\sigma_x^2, \sigma_y^2, \sigma_z^2$ of the measurement error within each of the three simultaneous data sets $\mathbf{x}, \mathbf{y}, \mathbf{z}$ must be known.

Exercise 11.12 How would you tackle this difficulty?

§ 11.11 One solution is as follows: when iteratively solving the set of equations $\partial S/\partial \alpha = 0 \dots$ you use the variance found in the previous iteration. To initialize this procedure (in iteration 1 there is no previous iteration!) you use the within-dataset variances.

Exercise 11.13 Suppose that you obtain a reasonably good fit, or perhaps a poor fit, or an excellent fit. How do you think goodness-of-fit should influence your confidence in the model?

Exercise 11.14 Suppose that in your exploration of your model you have found that, as parameters are varied, the curves depicting the process variables' variation with time assume a wide range of shapes. Explain why in such circumstances even a good agreement between best fit and data does little to confirm the structure of the model. (Hint: how 'easy' is it for the parameters to find fitting values?)

Exercise 11.15 Suppose that, by contrast, your model can only assume a restricted set of shapes. Argue why in this case a good agreement between best fit and data does enhance confidence in the model³.

³Keep in mind, though, that a radically different model, based on quite distinct assumptions, may provide an equally good fit to the data.

Exercise 11.16 Suppose once more that your model is ‘versatile’ in the sense of exercise 11.14, and suppose that you have obtained a good fit. How confident are you that the best-fit values of the parameters are close to the “true” values? Answer the same question when your model is ‘restricted’ in the sense of exercise 11.15, and you have obtained a good fit. (Hint: how large is the region of the parameter space where the fit is any good?)

§ 11.12 The foregoing exercises suggest a trade-off between the confidence a good fit gives you *in the model* and the confidence a good fit gives you *in the fitted values*. You can chart how the model’s behaviour changes over the parameter space to determine on which side of this trade-off your model lies.

Exercise 11.17 Discuss the role of dimensional analysis in this exploration of parameter space.

§ 11.13 A more direct way to assess the confidence you might have in your best-fit parameter values is to determine (by numerical means, if necessary) the second partial derivatives with respect to the parameters:

$$\frac{\partial^2 S}{\partial \alpha^2}, \quad \frac{\partial^2 S}{\partial \alpha \partial \beta} \quad \text{and so on,}$$

evaluated at $\alpha = \hat{\alpha}$, $\beta = \hat{\beta}$, $\gamma = \hat{\gamma}$.

Exercise 11.18 Explain why these quantities express how precisely the parameter values are fixed by the data.

§ 11.14 In the more general setting of likelihood theory, the statistical expectations of the second partial derivatives of the log likelihood with respect to the parameters make up a matrix called the **Fisher information**, an apt name as it tells us how well the parameters are fixed by the data.

Exercise 11.19 Explain the seemingly perverse attitude of a modeller who claims that the large values in the Fisher information matrix tend to *support* the theoretical hypothesis of his model.

§ 11.15 After the best fitting parameter values have been found, you may yet decide to confront the model with an additional data set that was not used in the fitting procedure (the model then gives a so-called **out-of-sample** prediction).

III

Past exam papers