Chapter 1

Modelling

Main concepts: In this first chapter we introduce some basic definitions and a series of model problems of increasing complexity. Keywords are: ODE, IVP, solutions

1.1 Modelling with dynamical systems

Suppose one day you, as a working mathematician, are asked to model the infestation of a rose farm by aphids, a tiny plant-feeding insect (Fig. 1.1a), and assess the feasibility of controlling the aphid population by introduction of a natural predator.

Being a mathematician, you decide that you need a mathematical model to describe the dynamics of the aphid population. Assuming you know nothing at all about population dynamics and can't be bothered to go to the library or search the internet, you might start reasoning as follows: Let $N(t)$ denote the number of aphids in the rose farm at time t. Although we would generally think of N as taking discrete (integer) values, we make the modelling assumption that it can be viewed as a continuous variable, taking values on the (positive) real line $N \in \mathbb{R}$. This assumption is reasonable when the number of aphids is so large that addition or subtraction of a fractional aphid won't matter. If the average life span of an aphid is, say, ten days, then on the average one-tenth of the population will die on any given day. Therefore the death rate is linearly proportional to size of the population: *death rate* $= mN(t)$, $m > 0$. The birth rate works the same way. Although we do not want to delve into the rather exotic reproductive activities of the aphid, suffice it to say that most aphids are females, able to reproduce asexually, giving birth instead of laying eggs, and born pregnant. The birth rate is then some average number b of young per aphid per day: *birth rate* $= bN(t)$, $b > 0$. Based purely on considerations of birth rate and death rate, you could construct a model for the aphid population by setting the rate of change in

Figure 1.1: *Prey and predator.* a. aphids devouring a rosebud. b. ladybug devouring an aphid. (source: Wikipedia)

population equal to the difference in birth and death rates, $dN/dt = bN(t) - mN(t)$. This results in the *exponential growth/decay model*

$$\frac{dN(t)}{dt} = rN(t), \tag{1.1}$$

where $r \in \mathbb{R}$ is the rate of population increase or decrease, depending on whether there are more births or deaths. For aphids, all else being equal, the births far outweigh the deaths.

The model (1.1) is an example of a differential equation, and it is to this type of model—and in particular to its solution on the computer—that this course is devoted.

Formally, a **differential equation** of the general form

$$\frac{dy}{dt} = f(t, y) \tag{1.2}$$

specifies the relation between an unknown differentiable function $y(t)$ and its derivative dy/dt. Equation (1.2) is termed an **ordinary differential equation** (ODE) because it involves derivatives with respect to a single independent variable (t). This is in contrast to a **partial differential equation**, which is one involving derivatives with respect to several independent variables, e.g.

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} \qquad \text{or} \qquad \frac{\partial^2 u}{\partial x^2} + a(x, y)\frac{\partial^2 u}{\partial y^2} = b(x, y).$$

For ODEs, time is often the independent variable, which is suggested by the letter t. We will often use the shorthand notation $y'(t)$ and occasionally $\dot{y}$ to denote dy/dt.

A **solution** of the differential equation (1.2) is a differentiable function $\tilde{y}(t)$ that satisfies the differential equation,

$$\frac{\partial \tilde{y}(t)}{\partial t} = f(t, \tilde{y}(t)),$$

at least for all t on some interval of interest. For example, for the aphid population model (1.1), the function

$$\tilde{N}(t) = ce^{rt}$$

is a solution for each constant $c \in \mathbb{R}$, since $d\tilde{N}/dt = cr\exp(rt) = rN(t)$ conform (1.1). We will usually drop the tilde when there is no chance of confusion.

To determine a unique solution, it is necessary to specify the value of the function at some point t. An **initial value problem** combines a differential equation, an initial value, and an interval

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0, \quad t \in [t_0, t_0 + T]. \tag{1.3}$$

Assuming an initial aphid population N_0 at $t = 0$, the population at any later time t is $N(t) = N_0 \exp(rt)$ If $r < 0$, the death rate exceeds the birth rate and the model predicts exponential decay of the aphid population. If $r > 0$ (typical for aphids), the births outnumber the deaths and the population grows exponentially.

The model (1.1) might be useful to describe the extinction of aphids when treated by a pesticide that reduces the birth rate, or to describe the initial phase of aphid infestation, but one would not expect the exponential growth to go unchecked. Limited food resources will cause an increase in death rate or a decrease in birth rate or both when the aphid population gets too large. The rose farm can only sustain a certain number of aphids, say N^*. A more sophisticated attempt is the model

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{N^*}\right).$$

This equation can be normalized by dividing both sides by N^*, and letting $n(t) = N(t)/N^*$. In terms of n, this becomes the *logistic model*

$$\frac{dn}{dt} = rn(1 - n). \tag{1.4}$$

We can generally solve any of a wide variety of differential equations of the form

$$\frac{dx}{dt} = f(x), \qquad x, f : \mathbb{R} \to \mathbb{R},$$

as follows:

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

then

$$\int \frac{dx}{f(x)} = t + C,$$

where C is an arbitrary constant. In practice there are some difficulties that can obstruct this procedure. For example it may be difficult to evaluate the indicated integral. Even when an antiderivative $G(x)$ for $1/f(x)$ can be written down, we must solve in some way the (typically nonlinear) equation

$$G(x) = t + C.$$

However, these problems can often be overcome and the solution computed as accurately as required at any point.

Let us solve the logistic model (1.4) using a decomposition into partial fractions:

$$\begin{aligned} r\,dt &= \frac{dn}{n(1-n)} \\ rt + C &= \int \frac{dn}{n(1-n)} = \int \frac{1}{n} + \frac{1}{1-n}\,dn \\ &= \ln n - \ln(1-n). \end{aligned}$$

Taking the exponential of both sides, and denoting $c = \exp C$,

$$\frac{n}{1-n} = ce^{rt} \quad \Rightarrow \quad n(t) = \frac{ce^{rt}}{1 + ce^{rt}}.$$

If $n(0) = n_0$, then $c = n_0/(1 - n_0)$ and the solution can be written

$$n(t) = \frac{n_0 e^{rt}}{(1 - n_0) + n_0 e^{rt}}.$$

If $r > 0$ it can be checked that

$$\lim_{t\to\infty} n(t) = 1$$

for any initial state $n_0 > 0$. That is, the population eventually saturates ($N \to N^*$) for any initial population. This behavior is shown in the left plot of Figure 1.2.

Next, suppose that in attempt to control the aphid population, a predator is introduced into the rose farm. One nasty predator is the ladybug (Fig. 1.1b. If you are an aphid, the ladybug is anything but a symbol of nonviolence!). Normally lady bugs will not bother eating aphids unless there is sufficient supply. The logistic model can be extended to incorporate *predation* by lady bugs as follows:

$$\frac{dn}{dt} = rn\left(1 - \frac{n}{s}\right) - \frac{n^2}{1 + n^2}. \tag{1.5}$$

The last term gradually "turns on" ladybug predation as the population n grows greater than 1. We have re-introduced the saturation level s because there are now two important populations levels—the natural saturation level, and the level for which ladybugs get "interested". Some solutions of this model are shown in the right plot of Figure 1.2. It is interesting to note that for $s = 20$ and either $r = 0.1$ or $r = 0.6$, the aphid population eventually tends to a constant value, independent of the initial population. This is similar to the case of the logistic equation without predation (1.4). For the case $s = 20$, $r = 0.25$, however, the final aphid population depends on the starting population. For a starting population of about $n_0 < 5$, the aphids settle down to a low level of about $n = 0.5$, whereas for a starting population of $n_0 > 5$, the aphid population eventually stabilizes at a relatively high level of around $n = 14.5$. Such a change in the quality of solutions with a change in parameter values is called *bifurcation*.

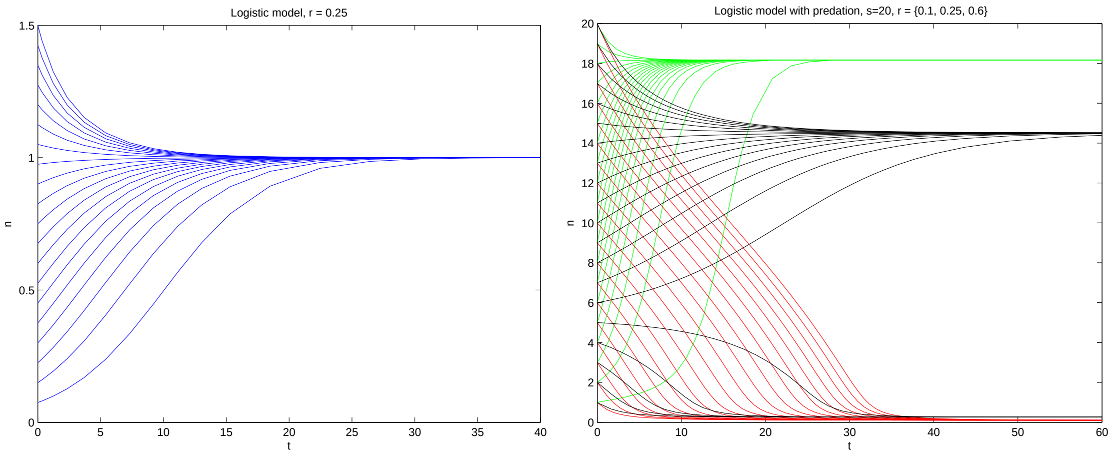

Figure 1.2: Solutions of the logistic model. On the left, solutions of (1.4) with $r = 0.25$ for different initial conditions. On the right, with predation (1.5), $s = 20$ and $r = 0.6$ (green), $r = 0.25$ (black), and $r = 0.1$ (red). For $r = 0.25$, there are two steady states, depending on the initial condition

This model is still rather crude, because it assumes an endless supply of ladybugs. An alternative model might consider an isolated system of aphids and ladybugs, and model the populations of both. Let $p(t)$ denote the (normalized) population of ladybugs. In the absence of aphids, the ladybugs would die out exponentially: $dp/dt = -mp$ for some $m > 0$. But with an increasing number of aphids, the ladybugs will flourish, so the ladybug population should satisfy

$$\frac{dp}{dt} = anp - mp$$

for positive constants a and m. Similarly, in the absence of ladybugs, the aphid population would grow exponentially (at least until they saturate the environment). And their death rate should increase proportional to the number of predators, so

$$\frac{dn}{dt} = bn - cnp,$$

for positive constants b and c. It is possible to rescale these equations such that only one constant remains, and the result is the *Lotka-Volterra predator-prey model*

$$\frac{dn}{dt} = rn(1-p) \tag{1.6a}$$

$$\frac{dp}{dt} = p(n-1), \tag{1.6b}$$

where $r > 0$.

The Lotka-Volterra model is an example of a set of two coupled ordinary differential equations. More generally one may have a **system of d differential equations**:

$$\frac{dy^{(1)}}{dt} = f^{(1)}\left(t, y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right) \tag{1.7}$$

$$\frac{dy^{(2)}}{dt} = f^{(2)}\left(t, y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right) \tag{1.8}$$

$$\vdots \tag{1.9}$$

$$\frac{dy^{(d)}}{dt} = f^{(d)}\left(t, y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right) \tag{1.10}$$

$$\tag{1.11}$$

When dealing with systems of differential equations, it is convenient to use vector notation. Let $y(t) : \mathbb{R} \to \mathbb{R}^d$ denote the vector with components $y^{(1)}(t), \ldots, y^{(d)}(t)$ and $f(t, y) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ denote the vector with components $f^{(1)}(t, y), \ldots f^{(d)}(t, y)$, i.e.

$$y(t) = \left(y^{(1)}(t), \cdots, y^{(d)}(t)\right)^T, \qquad f(t, y(t)) = \left(f^{(1)}(t, y), \cdots, f^{(d)}(t, y)\right)^T.$$

Then the system of differential equations above can be written using the same notation as (1.2), and the same goes for the initial value problem (1.3), where it is understood that the initial condition y_0 is now a vector in $\mathbb{R}^d$. The special case $d = 1$ is referred to as a *scalar* differential equation. Scalar differential equations are special because they can often be solved in quadratures. In this text we will often consider systems with small dimension d, say $d < 10$. However, many real applications have very large dimensions: millions of variables are not uncommon. It is important to keep this in mind when one is designing numerical methods.

The reproduction rate of aphids is actually not constant, as suggested by the constant parameter r in the above models. In fact aphid reproduction depends on temperature. One way to incorporate the effects of a typical yearly temperature variation is to allow r to depend explicitly on time. For example, the logistic model with predation (1.5) and time dependent reproduction becomes

$$\frac{dn}{dt} = r(t)\, n\left(1 - \frac{n}{s}\right) - \frac{n^2}{1+n^2}, \tag{1.12}$$

where $r(t)$ is a specified function, say $r(t) = r_0(\cos\varepsilon t + c_0)$. A system of differential equations for which $f(t, y)$ can be written as a function of y only is an **autonomous differential equation**

$$\frac{dy}{dt} = f(y), \qquad y, f \in \mathbb{R}^d. \tag{1.13}$$

It is useful to observe that autonomous differential equations are invariant under a translation of time. That means that if we define $\tau = t - t_0$, then $dt/d\tau = 1$, and

$$\frac{dy}{d\tau} = \frac{dy}{dt}\frac{dt}{d\tau} = \frac{dy}{dt}.$$

so the initial value problem $dy/dt = f(y)$, $y(t_0) = y_0$ can be replaced by $dy/d\tau = f(y)$, $y(0) = y_0$. Thus without loss of generality we can take the initial value of time to be $t = 0$, as long as the system is autonomous.

For notational convenience, it is sometimes useful to write a *nonautonomous* ODE (i.e. one with explicit dependence on t) as an autonomous one. This is always possible by increasing the dimension d by one. We define $y^{(d+1)} \equiv t$ and introduce a new independent variable s such that $dt/ds = 1$ and $t(0) = t_0$. The system (1.7)–(1.10) becomes

$$\begin{aligned}
\frac{dy^{(1)}}{ds} &= f^{(1)}\left(y^{(d+1)}, y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right)\\
&\vdots\\
\frac{dy^{(d)}}{ds} &= f^{(d)}\left(y^{(d+1)}, y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right)\\
\frac{dy^{(d+1)}}{ds} &= 1
\end{aligned}$$

where the independent variable s no longer appears explicitly on the right side.

From a modelling point of view, nonautonomous terms (such as $r(t)$ in the example above) can often be avoided by extending the system considered. For example instead of defining an explicit function $r(t)$, we could couple our predation model to a climate prediction model in which the average temperature r appears as dependent variable. Typically such a climate model will have additional variables as well, and the coupled system may look like

$$\begin{aligned}
\frac{dx}{dt} &= f(x, r)\\
\frac{dr}{dt} &= g(x, r)\\
\frac{dn}{dt} &= rn\left(1 - \frac{n}{s}\right) - \frac{n^2}{1 + n^2},
\end{aligned}$$

where r and n are scalars and $x \in \mathbb{R}^d$.[1] On the other hand, an accurate climate model may have $d = 10^9$, so this approach could constitute overkill.

1.2 References

For a thorough introduction to mathematical modelling and population dynamics in ecological systems, see Chapters 1–3 of

> James D. Murray, *Mathematical Biology I: An Introduction*, Third edition, Interdisciplinary Applied Mathematics series, Springer, 2002.

1.3 Exercises

1. Consider (1.12) and suppose the aphid reproduction rate is given by

$$r(t) = r_0(\cos \varepsilon t + c).$$

---

[1] The influence of our aphid population on the climate has been neglected, as indicated by the absence of n dependence in f and g.

Write a Matlab function to implement the differential equation (see the example function `lv.m` for the Lotka-Volterra model). Solve this with the Matlab toolbox function `ode23` for $r_0 = 0.32$, $c_0 = 1$, $\varepsilon = 2\pi/40$, $s = 20$. On the same axes, plot the solutions on the interval $t \in [0, 160]$ for initial aphid populations of $n = 1$, $n = 2$, $n = 3$ and $n = 10$. Describe what you see. What can you conclude about the numerical solution? about the model? about the phenomenon being modelled? (Use the Matlab help utility to learn about `ode23`. You can use the default tolerances.)