DIFFERENCE AND DIFFERENTIAL EQUATIONS IN MATHEMATICAL MODELLING

J. Banasiak

Contents

Chapter 1

Basic ideas about mathematical modelling

1.1 Introduction: what is mathematical modelling?

Engineers, natural scientists and, increasingly, researchers and practitioners working in economical and social sciences, use mathematical models of the systems they are investigating. Models give simplified descriptions of real-life problems so that they can be expressed in terms of mathematical equations which can be, hopefully, solved in one way or another. Mathematical modelling is a subject difficult to teach but it is what applied mathematics is about. The difficulty is that there are no set rules, and the understanding of the 'right' way to model can be only reached by familiarity with a number of examples. This, together with basic techniques for solving the resulting equations, is the main content of this course.

Despite these difficulties, applied mathematicians have a procedure that should be applied when building models. First of all, there must be a phenomenon of interest that one wants to describe or, more importantly, to explain and make predictions about. Observation of this phenomenon allows to make hypotheses about which quantities are most relevant to the problem and what are the relations between them so that one can devise a hypothetical mechanism that can explain the phenomenon. The purpose of a model is then to formulate a description of this mechanism in quantitative terms, that is, as mathematical equations, and the analysis of the resulting equations. It is then important to interpret the solutions or other information extracted from the equations as the statements about the original problem so that they can be tested against the observations. Ideally, the model also leads to predictions which, if verified, lend authenticity to the model. It is important to realize that modelling is usually an iterative procedure as it is very difficult to achieve a balance between simplicity and meaningfulness of the model: often the model turns out to be too complicated to yield itself to an analysis, and often it is over-simplified so that there is insufficient agreement between the actual experimental results and the results predicted from the model. In both these cases we have to return to the first step of modelling and try to remedy the ills.

The first step in modelling is the most creative but also the most difficult, involving often a concerted effort of specialists in many diverse fields. Hence, though we describe a number of models in detail, starting from first principles, the main emphasis of the course is on the later stages of the modelling process, that is: introducing mathematical symbols and writing assumptions as equations, analysing and/or solving these equations and interpreting their solutions in the language of the original problem and reflecting on whether the answers seem reasonable.

In most cases discussed here a model is a representation of a process, that is, it describes a change of the states of some system in time. There are two ways of describing what happens to a system: discrete and continuous. Discrete models correspond to the situation in which we observe a system in regular finite time intervals, say, every second or every year and relate the observed state of the system to the states at the previous instants. Such a system is modelled through difference equations. In the continuous cases we treat time as a continuum allowing observation of the system at any instant. In such a case the model expresses relations between the rates of change of various quantities rather than between the states at various times and, as rates of change are given by derivatives, the model is represented by differential equations.

In the next two sections of this chapter we shall present some simple discrete and continuous models. These models are presented here as an illustration of the above discussion. Their analysis, and discussion of more advanced models, will appear later in the course.

1.2 Simple difference equation models

1.2.1 Basic difference equations of finance mathematics

Compound interest
Compound interest is relevant to loans or deposits made over longer periods. The interest is added to the initial sum at regular intervals, called conversion periods, and the new amount, rather than the initial one, is used for calculating the interest for the next conversion period. The fraction of a year occupied by the conversion period is denoted by α so that the conversion pe-

riod of 1 month is given by $\alpha = 1/12$. Instead of saying that the conversion period is 1 month we say that the interest is compounded monthly.

For an annual interest rate of $p\%$ and conversion period equal to α, the interest earned for the period is equal to $\alpha p\%$ of the amount on deposit at the start of the period, that is

$$\left\{\begin{array}{c}\text{amount on}\\ \text{deposit}\\ \text{after } k+1\\ \text{conversion}\\ \text{periods}\end{array}\right\} = \left\{\begin{array}{c}\text{amount on}\\ \text{deposit}\\ \text{after } k\\ \text{conversion}\\ \text{periods}\end{array}\right\} + \frac{\alpha p}{100}\left\{\begin{array}{c}\text{amount on}\\ \text{deposit}\\ \text{after } k\\ \text{conversion}\\ \text{periods}\end{array}\right\}$$

To express this as a difference equation, for each k let S_k denote the amount on deposit after k conversion periods. Thus

$$S_{k+1} = S_k + \frac{\alpha p}{100} S_k = S_k\left(1 + \frac{\alpha p}{100}\right)$$

which is a simple first-order (that is, expressing the relation only between the consecutive values of the unknown sequence) difference equation. Here, S_k follows the geometric progression so that

$$S_k = \left(1 + \frac{\alpha p}{100}\right)^k S_0 \qquad (1.2.1)$$

gives the so-called compound interest formula. However, as we shall see below, in general this is not the case, even for first order equations.

Loan repayments

A slight modification of the above argument can be used to find the equation governing a loan repayment. The scheme described here is usually used for the repayment of house or car loans. Repayments are made at regular intervals and usually in equal amounts to reduce the loan and to pay the interest on the amount still owing.

It is supposed that the compound interest at $p\%$ is charged on the outstanding debt with the conversion period equal to the same fraction α of the year as the period between the repayments. Between payments, the debt increases because of the interest charged on the debt still outstanding after the last repayment. Hence

$$\left\{\begin{array}{c}\text{debt after}\\ k+1 \text{ payments}\end{array}\right\} = \left\{\begin{array}{c}\text{debt after}\\ k \text{ payments}\end{array}\right\} + \left\{\begin{array}{c}\text{interest}\\ \text{on this debt}\end{array}\right\} - \{\text{payment}\}$$

To write this as a difference equation, let D_0 be the initial debt to be repaid, for each k let the outstanding debt after the kth repayment be D_k, and let the payment made after each conversion period be R. Thus

$$D_{k+1} = D_k + \frac{\alpha p}{100} D_k - R = D_k\left(1 + \frac{\alpha p}{100}\right) - R. \qquad (1.2.2)$$

This equation is more difficult to solve. We shall discuss general methods of solving first order difference equations in Section 4.1.

The modelling process in these two examples was very simple and involved only translation of given rules into mathematical symbols. This was due to the fact that there was no need to discover these rules as they are explicitly stated in bank's regulations. In the next subsection we shall attempt to model behaviour of living organisms and then we shall have to make some hypotheses about the rules.

1.2.2 Difference equations of population theory

In many fields of human endeavour it is important to know how populations grow and what factors influence their growth. Knowledge of this kind is important in studies of bacterial growth, wildlife management, ecology and harvesting.

Many animals tend to breed only during a short, well-defined, breeding season. It is then natural to thing of the population changing from season to season and therefore time is measured discretely with positive integers denoting breeding seasons. Hence the obvious approach for describing the growth of such a population is to write down a suitable difference equation. Later we shall also look at populations that breed continuously (e.g. human populations).

We start with population models that are very simple and discuss some of their more realistic variants.

Exponential growth – linear first order difference equations
In nature, species typically compete with other species for food and are sometimes preyed upon. Thus the population of different species interact with each other. In the laboratory, however, a given species can be studied in isolation. We shall therefore concentrate, at first, on models for a single species.

We are looking at large populations in which individuals give birth to new offspring but also die after some time. Since we deal with large populations, we can treat population as a whole and therefore we can assume that the population growth is governed by the average behaviour of its individual members. Thus, we make the following assumptions:

- Each member of the population produces in average the same number of offspring.

- Each member has an equal chance of dying (or surviving) before the next breeding season.

- The ratio of females to males remains the same in each breeding season

We also assume

- Age differences between members of the population can be ignored.
- The population is isolated - there is no immigration or emigration.

Suppose that on average each member of the population gives birth to the same number of offspring, α, each season. The constant α is called per-capita birth rate. We also define β as the probability that an individual will die before the next breeding season and call it the per-capita death rate. Thus

(a) the number of individuals born in a particular breeding season is directly proportional to the population at the start of the breeding season, and

(b) the number of individuals who have died during the interval between the end of consecutive breeding seasons is directly proportional to the population at the start of the breeding season.

Denoting by N_k the number of individuals of the population at the start of the kth breeding season, we obtain

$$N_{k+1} = N_k - \beta N_k + \alpha N_k,$$

that is

$$N_{k+1} = (1 + \alpha - \beta)N_k. \tag{1.2.3}$$

We have seen this equation before, modelling the compound interest. Since it is the equation for the geometric progression, we immediately obtain

$$N_k = (1 + \alpha - \beta)^k N_0, \qquad k = 0, 1, 2 \ldots \tag{1.2.4}$$

Note first that though our modelling referred to the size of the population, that is an integer, the number N_k given by (1.2.4) usually is not an integer. This is due to the fact that we operate with average rates α and β. To circumvent this apparent paradox, we can always round N_k to the nearest integer. Another look at this is that in modelling we often use the population density, that is, the number of individuals per unit area. Population density usually is not an integer.

Returning to (1.2.4), we see that the behaviour of the model depends on the combination

$$r = \alpha - \beta \tag{1.2.5}$$

that is called the growth rate. If $r < 0$, then the population decreases towards extinction, but with $r > 0$ it grows indefinitely. Such a behaviour over long periods of time is not observed in any population so that we see that the model is over-simplified and requires corrections.

Another over-simplification is lack of the age structure in the model – we assume that offspring immediately enter into the breeding cycle. In the next two examples we shall present two the population models that take the above aspects into account.

Fibonacci sequence and difference equations of second order
Suppose that we have a rabbit population and that in this population each pair of rabbits produces a new pair every month and the pair of newborns becomes productive two month after birth. Assuming that no deaths occur, we can write for the end of month $k+1$

$$\left\{\begin{array}{c}\text{number present}\\ \text{in month } k+1\end{array}\right\} = \left\{\begin{array}{c}\text{number present}\\ \text{in month } k\end{array}\right\} + \left\{\begin{array}{c}\text{number born}\\ \text{in month } k+1\end{array}\right\}$$

Since rabbits become productive only two months after birth and produce only one pair per month, we can write

$$\left\{\begin{array}{c}\text{number born}\\ \text{in month } k+1\end{array}\right\} = \left\{\begin{array}{c}\text{number present}\\ \text{in month } k-1\end{array}\right\}$$

Denoting by N_k the number of pairs at the end of month k and combining the two equations above, we obtain the so-called Fibonacci equation

$$N_{k+1} = N_k + N_{k-1}, \qquad k = 1, 2, \ldots. \tag{1.2.6}$$

This is a second order equation as it gives the value of N_k at time k in terms of its values at two times immediately preceding k.

Restricted growth - non-linear difference equations
As we said earlier, the linear difference equation is not generally suitable as a model for population growth since it predicts unbounded growth if we have an expanding population. This is not what is observed in nature. However, over some periods of time populations tend to follow an exponential growth. Therefore, rather than reject the model outright we shall try to build into it modifications so that it better approximates the observed behaviour.

Studies show that typically as a population increases, the per-capita death rate goes up and the per-capita birth rate goes down. This is due to overcrowding and competition for food. Typically, for each population and habitat there is a number of individuals that a given environment can support. This number is known as the carrying capacity of the environment. It can be alternatively described as the number of individuals in the population when the birth and death rate are equal. Recalling the linear difference equation for population growth

$$N_{k+1} = N_k + rN_k,$$

where the constant r is the growth rate, we can incorporate the discussion above by writing

$$N_{k+1} = N_k + R(N_k)N_k \tag{1.2.7}$$

where $R(N_k)$ is the population dependent growth rate. This equation is an example of a non-linear difference equation since the unknown function appears in the equation as the argument of a function that is not linear: in this case as the argument of the function $xR(x)$.

Though the function R can have different forms, it must satisfy the following requirements

(a) Due to overcrowding, $R(N_k)$ must decrease as N_k increases until N_k equals the carrying capacity K; then $R(K) = 0$.

(b) Since for N_k much smaller than K we observe an exponential growth of the population so that $R(N_k) \to r$ as $N_k \to 0$, where r is a constant called the unrestricted growth rate.

Constants r and N are usually determined experimentally.

In the spirit of mathematical modelling we start with the simplest function satisfying these requirements. The simplest function is a linear function which, to satisfy (a) and (b), must be chosen as

$$R(N_k) = -\frac{r}{K}N_k + r.$$

Substituting this formula into (1.2.7) yields the so-called discrete logistic equation

$$N_{k+1} = N_k + rN_k\left(1 - \frac{N_k}{K}\right), \tag{1.2.8}$$

which is still one of the most often used discrete equations of population dynamics.

Coupled model of an epidemic: system of difference equations
In nature, various population interact with each other coexisting in the same environment. This leads to systems of difference equations. As an illustration we consider a model for spreading of measles epidemic.

Measles is a highly contagious disease, caused by a virus and spread by effective contact between individuals. It affects mainly children. Epidemic of measles have been observed in Britain and the US roughly every two or three years.

Let us look at the development of measles in a single child. A child who has not yet been exposed to measles is called a susceptible. Immediately after the child first catches the disease, there is a latent period where the child is not contagious and does not exhibit any symptoms of the disease. The latent period lasts, on average, 5 to 7 days. After this the child enters the contagious period. The child is now called infective since it is possible for another child who comes in contact with the infective to catch the disease. This period last approximately one week. After this the child recovers, becomes immune to the disease and cannot be reinfected.

For simplicity we assume that both latent and contagious periods last one week. Suppose also that most interactions between children occur on week-end so that the numbers of susceptibles and infectives remains constant over the rest of the week. Since the typical time in the model is one week, we shall model the spread of the disease using one week as the unit of time.

To write down the equations we denote

$$I_k = \left\{\begin{array}{c}\text{number of}\\ \text{infectives in week } k\end{array}\right\}$$

and

$$S_k = \left\{\begin{array}{c}\text{number of}\\ \text{susceptibles during week } k\end{array}\right\}$$

To develop an equation for the number of infectives we consider the number of infectives in week $k+1$. Since the recuperation period is one week after which an infective stops to be infective, no infectives from week k will be present in week $k+1$. Thus we have

$$I_{k+1} = \left\{\begin{array}{c}\text{number of}\\ \text{infectives in week } k+1\end{array}\right\} = \left\{\begin{array}{c}\text{number of susceptibles}\\ \text{who caught measles in week } k\end{array}\right\}$$

It is generally thought that the number of new births is an important factor in measles epidemic. Thus

$$\begin{aligned} S_{k+1} &= \left\{\begin{array}{c}\text{number of}\\ \text{susceptibles in week } k\end{array}\right\} - \left\{\begin{array}{c}\text{number of susceptibles}\\ \text{who caught measles in week } k\end{array}\right\}\\ &\quad + \left\{\begin{array}{c}\text{number of}\\ \text{births in week } k+1\end{array}\right\}\end{aligned}$$

We assume further that the number of births each week is a constant B. Finally, to find the number of susceptibles infected in a week it is assumed that a single infective infects a constant fraction f of the total number of susceptibles. Thus, if fS_k is the number of susceptibles infected by a single infective so, with a total of I_k infectives, then

$$\left\{\begin{array}{c}\text{number of susceptibles}\\ \text{who caught measles in week } k\end{array}\right\} = fS_kI_k.$$

Combining the obtained equations we obtain the system

$$\begin{aligned} I_{k+1} &= fS_kI_k,\\ S_{k+1} &= S_k - fS_kI_k + B, \end{aligned} \qquad (1.2.9)$$

where B and f are constant parameters of the model.

1.3 Basic differential equations models

As we observed in the previous section, the difference equation can be used to model quite a diverse phenomena but their applicability is limited by the fact that the system should not change between subsequent time steps. These steps can vary from fraction of a second to years or centuries but they must stay fixed in the model. There are however numerous situations when the changes can occur instantaneously. These include growth of populations in which breeding is not restricted to specific seasons, motion of objects where the velocity and acceleration changes every instant, spread of epidemic with no restriction on infection times, and many others. In such cases it is not feasible to model the process by relating the state of the system at a particular instant to the earlier states (though this part remains as an intermediate stage of the modelling process) but we have to find relations between the rates of change of quantities relevant to the process. Rates of change are typically expressed as derivatives and thus continuous time modelling leads to differential equations that express relations between the derivatives rather than to difference equations that express relations between the states of the system in subsequent moments of time.

In what follows we shall derive basic differential equations trying to provide continuous counterparts of some discrete systems described above.

1.3.1 Continuously compounded interest

Many banks now advertise continuous compounding of interest which means that the conversion period α of Subsection 1.2.1 tends to zero so that the interest is added to the account on the continual basis. If we measure now time in years, that is, Δt becomes the conversion period, and p is the annual interest rate, then the increase in the deposit between time instants t and $t+\Delta t$ will be

$$S(t+\Delta t) = S(t) + \Delta t \frac{p}{100} S(t),$$

which, dividing by Δt and passing with Δt to zero, as suggested by the definition of continuously compounded interest, yields the differential equation

$$\frac{dS}{dt} = \bar{p}S, \tag{1.3.1}$$

where $\bar{p} = p/100$. This is a first order (only the first order derivative of the unknown function occurs) linear (the unknown function appears only by itself, not as an argument of any function) equation. It is easy to check that it has the solution

$$S(t) = S_0 e^{\bar{p}t} \tag{1.3.2}$$

where S_0 is the initial deposit made at time $t = 0$.

To compare this formula with the discrete one (1.2.1) we note that in t years we have $k = t/\alpha$ conversion periods

$$S(t) = N_k = (1 + \bar{p}\alpha)^k S_0 = (1 + \bar{p}\alpha)^{t/\alpha} S_0 = \left((1 + \bar{p}\alpha)^{1/\bar{p}\alpha}\right)^{\bar{p}t}.$$

From Calculus we know that

$$\lim_{x\to 0^+} (1 + x)^{1/x} = e,$$

and the sequence is monotonically increasing. Thus, if the interest is compounded very often (almost continuously), then practically

$$S(t) = S_0 e^{\bar{p}t},$$

which is exactly (1.3.2). Typically, the exponential can be calculated even on a simple calculator, contrary to (1.2.1). Due to monotonic property of the limit, the continuously compounded interest rate is the best one can get, and that is why banks are advertising it. However, the differences in return are negligible. A short calculation reveals that if one invests R10000 at $p =$ 15% in banks with conversion periods 1 year, 1 day and with continuously compounded interest, then the return will be, respectively, R11500, R11618 and R11618.3.

1.3.2 Continuous population models: first order equations

In this subsection we will study first order differential equations which appear in the population growth theory. At first glance it appears that it is impossible to model the growth of species by differential equations since the population of any species always change by integer amounts. Hence the population of any species can never be a differentiable function of time. However, if the population is large and it increases by one, then the change is very small compared to a given population. Thus we make the approximation that large populations changes continuously (and even differentiable)in time and, if the final answer is not an integer, we shall round it to the nearest integer. A similar justification applies to our use of t as a real variable: in absence of specific breeding seasons, reproduction can occur at any time and for sufficiently large population it is then natural to think of reproduction as occurring continuously.

Let $N(t)$ denote the size of a population of a given isolated species at time t and let Δt be a small time interval. Then the population at time $t + \Delta t$ can be expressed as

$$N(t + \Delta t) - N(t) = \text{number of births in } \Delta t - \text{number of deaths in } \Delta t.$$

It is reasonable to assume that the number of births and deaths in a short time interval is proportional to the population at the beginning of this interval and proportional to the length of this interval. Taking $r(t, N)$ to be the difference between the birth and death rate coefficients at time t for the population of size N we obtain

$$N(t + \Delta t) - N(t) = r(t, N(t))\Delta t N(t).$$

Dividing by Δt and passing with $\Delta t \to 0$ gives the equation

$$\frac{dN}{dt} = r(t, N)N. \tag{1.3.3}$$

Because of the unknown coefficient $r(t, N)$, depending on the unknown function N, this equation is impossible to solve. The form of r has to be deduced by other means.

The simplest possible $r(t, N)$ is a constant and in fact such a model is used in a short-term population forecasting. So let us assume that $r(t, N(t)) = r$ so that

$$\frac{dN}{dt} = rN. \tag{1.3.4}$$

It is exactly the same equation as (1.3.1). A little more general solution to it is given by

$$N(t) = N(t_0)e^{r(t-t_0)}, \tag{1.3.5}$$

where $N(t_0)$ is the size of the population at some fixed initial time t_0.

To be able to give some numerical illustration to this equation we need the coefficient r and the population at some time t_0. We use the data of the U.S. Department of Commerce: it was estimated that the Earth population in 1965 was 3.34 billion and that the population was increasing at an average rate of 2% per year during the decade 1960-1970. Thus $N(t_0) = N(1965) = 3.34 \times 10^9$ with $r = 0.02$, and (1.3.5) takes the form

$$N(t) = 3.34 \times 10^9 e^{0.02(t-1965)}. \tag{1.3.6}$$

To test the accuracy of this formula let us calculate when the population of the Earth is expected to double. To do this we solve the equation

$$N(T + t_0) = 2N(t_0) = N(t_0)e^{0.02T},$$

Fig 1.1. Comparison of actual population figures (points) with those obtained from equation (1.3.6).

thus

$$2 = e^{0.02T}$$

and

$$T = 50 \ln 2 \approx 34.6 \text{ years.}$$

This is an excellent agreement with the present observed value of the Earth population and also gives a good agreement with the observed data if we don't go too far into the past. On the other hand, if we try to extrapolate this model into a distant future, then we see that, say, in the year 2515, the population will reach 199980 $\approx$ 200000 billion. To realize what it means, let us recall that the Earth total surface area 167400 billion square meters, 80% of which is covered by water, thus we have only 3380 billion square meters to our disposal and there will be only $0.16m^2$ ($40cm \times 40cm$) per person. Therefore we can only hope that this model is not valid for all times. Indeed, as for discrete models, it is observed that the linear model for the population growth is satisfactory as long as the population is not too large. When the population gets very large (with regard to its habitat), these models cannot be very accurate, since they don't reflect the fact that the individual members have to compete with each other for the limited living space, resources and food available. It is reasonable that a given habitat can sustain only a finite number K of individuals, and the closer the population is to this number, the slower is it growth. Again, the simplest way to take this into account is to take $r(t, N) = r(K - N)$ and then we obtain the so-called *continuous logistic model*

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), \tag{1.3.7}$$

which proved to be one of the most successful models for describing a single species population. This equation is still first order equation but a non-linear one (the unknown function appears as an argument of the non-linear (quadratic) function $rx(1 - x/K)$. Since this model is more difficult to solve, we shall discuss it in detail later. However, even now we can draw from (1.3.7) a conclusion that is quite important in fishing (or other animal) industry. The basic idea of sustainable economy is to find an optimal level between too much harvesting, that would deplete the animal population beyond a possibility of recovery and too little, in which case the human population would not get enough return from the industry. It is clear that to maintain the animal population at a constant level, only the increase in population should be harvested during any one season. Hence, to maximize the harvest, the population should be kept at the size N for which the rate of increase dN/dt is a maximum. However, dN/dt is given by the right-hand side of (1.3.7) which is a quadratic function of N. It is easy to find that the maximum is attained at $N = K/2$, that is, the population should be kept at around half of the carrying capacity. Further, maximum of dN/dt is then given by

$$\left(\frac{dN}{dt}\right)_{max} = \frac{rK}{4}$$

and this is the maximum rate at which fish can be harvested, if the population is to be kept at a constant size $N/2$.

It is interesting to note that the same (mathematically) equation (1.3.7) can be obtained as a model of a completely different process: spreading information in a fixed size community. Let us suppose that we have a community of constant size C and N members of this community have some important information. How fast this information is spreading? To find an equation governing this process we state the following assumptions: the information is passed when a person knowing it meets a person that does not know it. Let us assume that the rate at which one person meets other people is a constant f so that in a time interval Δt this particular person will meet $f\Delta t$ people and, in average, $\Delta t f(C - N)/C$ people who do not know the news. If N people had the information at time t, then the increase will be

$$N(t + \Delta t) - N(t) = fN(t)\left(1 - \frac{N(t)}{C}\right)\Delta t$$

so that, as before,

$$\frac{dN}{dt} = fN\left(1 - \frac{N}{C}\right).$$

1.3.3 Equations of motion: second order equations

Second order differential equations appear often as equations of motion. This is due to the Newton's law of motion that relates the acceleration of the body, that is, the second derivative of the position y with respect to time t, to the body's mass m and the forces F acting on it:

$$\frac{d^2y}{dt^2} = \frac{F}{m}. \tag{1.3.8}$$

We confined ourselves here to a scalar, one dimensional case with time independent mass. The modelling in such cases concern the form of the force acting on the body. We shall consider two such cases in detail.

A waste disposal problem
In many countries toxic or radioactive waste is disposed by placing it in tightly sealed drums that are then dumped at sea. The problem is that these drums could crack from the impact of hitting the sea floor. Experiments confirmed that the drums can indeed crack if the velocity exceeds $12m/s$ at the moment of impact. The question now is to find out the velocity of a drum when it hits the sea floor. Since typically the waste disposal takes place at deep sea, direct measurement is rather expensive but the problem can be solved by mathematical modelling.

As a drum descends through the water, it is acted upon by three forces W, B, D. The force W is the weight of the drum pulling it down and is given by mg, where g is the acceleration of gravity and m is the mass of the drum. The buoyancy force B is the force of displaced water acting on the drum and its magnitude is equal to the weight of the displaced water, that is, $B = g\rho V$, where ρ is the density of the sea water and V is the volume of the drum. If the density of the drum (together with its content) is smaller that the density of the water, then of course the drum will be floating. It is thus reasonable to assume that the drum is heavier than the displaced water and therefore it will start drowning with constant acceleration. Experiments (and also common sense) tell us that any object moving through a medium like water, air, etc. experiences some resistance, called the drag. Clearly, the drag force acts always in the opposite direction to the motion and its magnitude increases with the increasing velocity. Experiments show that in a medium like water for small velocities the drag force is proportional the velocity, thus $D = cV$. If we now set $y = 0$ at the sea level and let the direction of increasing y be downwards, then from (1.3.8)

$$\frac{d^2y}{dt^2} = \frac{1}{m}\left(W - B - c\frac{dy}{dt}\right). \tag{1.3.9}$$

This is a second order (the highest derivative of the unknown function is of second order) and linear differential equation.

Motion in a changing gravitational field

According to Newton's law of gravitation, two objects of masses m and M attract each other with force of magnitude

$$F = G\frac{mM}{d^2}$$

where G is the gravitational constant and d is the distance between objects' centres. Since at Earth's surface the force is equal (by definition) to $F = mg$, the gravitational force exerted on a body of mass m at a distance y above the surface is given by

$$F = -\frac{mgR^2}{(y+R)^2},$$

where the minus sign indicates that the force acts towards Earth's centre. Thus the equation of motion of an object of mass m projected upward from the surface is

$$m\frac{d^2y}{dt^2} = -\frac{mgR^2}{(y+R)^2} - c\left(\frac{dy}{dt}\right)^2$$

where the last term represents the air resistance which, in this case, is taken to be proportional to the square of the velocity of the object. This is a second order nonlinear differential equation.

1.3.4 Equations coming from geometrical modelling

Satellite dishes

In many applications, like radar or TV/radio transmission it is important to find the shape of a surface that reflects parallel incoming rays into a single point, called the focus. Conversely, constructing a spot-light one needs a surface reflecting light rays coming from a point source to create a beam of parallel rays. To find an equation for a surface satisfying this requirement we set the coordinate system so that the rays are parallel to the x-axis and the focus is at the origin. The sought surface must have axial symmetry, that is, it must be a surface of revolution obtained by rotating some curve C about the x-axis. We have to find the equation $y = y(x)$ of C. Using the notation of the figure, we let $M(x, y)$ be an arbitrary point on the curve and denote by T the point at which the tangent to the curve at M intersects the x-axis. It follows that the triangle TOM is isosceles and

$$\tan\sphericalangle OTM = \tan\sphericalangle TMO = \frac{dy}{dx}$$

where the derivative is evaluated at M. On the other hand

$$\tan\sphericalangle OTM = \frac{|MP|}{|TP|},$$

Fig 1.2. Geometry of a reflecting surface.

but $|MP| = y$ and, since the triangle is isosceles, $|TP| = |OT| - |OP| = |OM| - |OP| = \sqrt{x^2+y^2} + x$. Thus, the differential equation of the curve C is

$$\frac{dy}{dx} = \frac{y}{\sqrt{x^2+y^2}+x}. \tag{1.3.10}$$

This is a nonlinear, so called homogeneous, first order differential equation. As we shall see later, it is not difficult to solve, if one knows appropriate techniques, yielding a parabola, as expected from the Calculus course.

The pursuit curve

What is the path of a dog chasing a rabbit or the trajectory of self-guided missile trying to intercept an enemy plane? To answer this question we must first realize the principle used in controlling the chase. This principle is that at any instant the direction of motion (that is the velocity vector) is directed towards the chased object.

To avoid technicalities, we assume that the target moves with a constant speed v along a straight line so that the pursuit takes place on a plane. We introduce the coordinate system in such a way that the target moves along the y-axis in the positive direction, starting from the origin at time $t = 0$, and the pursuer starts from a point at the negative half of the x-axis, see Figure 1.3. We also assume that the pursuer moves with a constant speed u.

Fig 1.3. The pursuit curve.

Let $M = M(x(t), y(t))$ be a point at the curve C, having the equation $y = y(x)$, corresponding to time t of the pursuit. At this moment the position of the target is $(0, vt)$. Denoting $y' = \frac{dy}{dx}$, from the principle of the pursuit we obtain

$$y' = -\frac{vt - y}{x} \tag{1.3.11}$$

In this equation we have too many variables and we shall eliminate t as we are looking for the equation of the trajectory in x, y variables. Solving (1.3.11) with respect to t we obtain

$$t = \frac{y - xy'}{v},$$

whereupon, using the assumption that v is a constant,

$$\frac{dt}{dx} = -\frac{1}{v}xy''$$

or, using the formula for differentiation of the inverse,

$$\frac{dx}{dt} = -\frac{v}{xy''}. \tag{1.3.12}$$

On the other hand, since we know that the speed of an object moving according to parametric equation $(x(t), y(t))$ is given by

$$u = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{1 + (y')^2}\left|\frac{dx}{dt}\right|, \tag{1.3.13}$$

where we used the formula for parametric curves

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}},$$

whenever $dx/dt \neq 0$. From the problem it is reasonable to expect that $dx/dt > 0$ so that we can drop the absolute value bars in (1.3.12). Thus, combining (1.3.12) and (1.3.13) we obtain the equation of the pursuit curve

$$xy'' = -\frac{v}{u}\sqrt{1+(y')^2}. \tag{1.3.14}$$

This is a nonlinear second order equation having, however, a nice property of being reducible to a first order equation and thus yielding a closed form solutions. We shall deal with such equations later on.

1.3.5 Modelling interacting quantities – systems of differential equations

In many situations we have to model evolutions of two (or more) quantities that are coupled in the sense that the state of one of them influences the other and conversely. We have seen this type of interactions in the discrete case when we modelled the spread of a measles epidemic. It resulted then in a system of difference equations. Similarly, in the continuous case the evolution of interacting populations will lead to a system of differential equations. In this subsection we shall discuss modelling of such systems that result in both linear and non-linear systems.

Two compartment mixing – a system of linear equations
Let us consider a system consisting of two vats of equal capacity containing a diluted dye: the concentration at some time t of the dye in the first vat is $c_1(t)$ and in the second is $c_2(t)$. Suppose that the pure dye is flowing into the first vat at a constant rate r_1 and water is flowing into the second vat at a constant rate r_2 (in, say, litres per minute). Assume further that two pumps exchange the contents of both vats at constant rates: p_1 from vat 1 to vat 2 and conversely at p_2. Moreover, the diluted mixture is drawn off vat 2 at a rate R_2. The flow rates are chosen so that the volumes of mixture in each vat remain constant, equal to V, that is $r_1+p_2-p_1 = r_2-p_2-R_2 = 0$. We have to find how the dye concentration in each vat changes in time.

Let $x_1(t)$ and $x_2(t)$ be the volumes of dye in each tank at $t \geq 0$. Thus, the concentrations c_1 and c_2 are defined by $c_1 = x_1/V$ and $c_2 = x_2/V$. We shall consider what happens to the volume of the dye in each vat during a small time interval from t to Δt. In vat 1

$$\begin{aligned} x_1(t+\Delta t) - x_1(t) &= \left\{\begin{array}{c}\text{volume of}\\ \text{of pure dye}\\ \text{flowing into vat 1}\end{array}\right\} + \left\{\begin{array}{c}\text{volume of}\\ \text{dye in mixture 2}\\ \text{flowing into vat 1}\end{array}\right\}\\ &\quad -\left\{\begin{array}{c}\text{volume of}\\ \text{dye in mixture 1}\\ \text{flowing out of vat 1}\end{array}\right\}\\ &= r_1\Delta t + p_2\frac{x_2(t)}{V}\Delta t - p_1\frac{x_1(t)}{V}\Delta t, \end{aligned}$$

and in vat 2, similarly,

$$\begin{aligned} x_2(t+\Delta t) - x_2(t) &= \left\{\begin{array}{c}\text{volume of}\\ \text{dye in mixture 1}\\ \text{flowing into vat 2}\end{array}\right\} - \left\{\begin{array}{c}\text{volume of}\\ \text{dye in mixture 2}\\ \text{flowing out of vat 2}\end{array}\right\}\\ &\quad -\left\{\begin{array}{c}\text{volume of}\\ \text{dye in mixture 2}\\ \text{flowing from vat 2 vat 1}\end{array}\right\}\\ &= p_1\frac{x_1(t)}{V}\Delta t - R_2\frac{x_2(t)}{V}\Delta t - p_2\frac{x_2(t)}{V}\Delta t. \end{aligned}$$

Dividing by Δt and passing with it to zero we obtain the following simultaneous system of linear differential equations

$$\begin{aligned} \frac{dx_1}{dt} &= r_1 + p_2\frac{x_2}{V} - p_1\frac{x_1}{V}\\ \frac{dx_2}{dt} &= p_1\frac{x_1}{V} - (R_2+p_2)\frac{x_2}{V}. \end{aligned} \tag{1.3.15}$$

Continuous model of epidemics – a system of nonlinear differential equations

A measles epidemic discussed earlier was modelled as a system of non-linear difference equations. The reason for the applicability of difference equations was the significant latent period between catching the disease and becoming contagious. If this period is very small (ideally zero) it it more reasonable to construct a model involving coupled differential equations. For the purpose of formulating the model we divide the population into three groups: susceptibles (who are not immune to the disease), infectives (who are capable of infecting susceptibles) and removed (who have previously had the disease and may not be reinfected because they are immune, have been quarantined or have died from the disease). The symbols S, I, R will be used to denote the number of susceptibles, infectives and removed, respectively, in the population at time t. We shall make the following assumptions on the character of the disease:

(a) The disease is transmitted by close proximity or contact between an infective and susceptible.

(b) A susceptible becomes an infective immediately after transmission.

(c) Infectives eventually become removed.

(d) The population of susceptibles is not altered by emigration, immigration, births and deaths.

(e) Each infective infects a constant fraction β of the susceptible population per unit time.

(f) The number of infectives removed is proportional to the number of infectives present.

As mentioned earlier, it is assumption (b) that makes a differential rather than difference equation formulation more reasonable. Diseases for which this assumption is applicable include diphtheria, scarlet fever and herpes. Assumption (e) is the same that used in difference equation formulation. It is valid provided the number of infectives is small in comparison to the number of susceptibles.

To set up the differential equations, we shall follow the standard approach writing first difference equations over arbitrary time interval and then pass with the length of this interval to zero. Thus, by assumptions (a), (c) and (d), for any time t

$$S(t+\Delta t) = S(t) - \left\{ \begin{array}{c} \text{number of susceptibles} \\ \text{infected in time } \Delta\text{t} \end{array} \right\},$$

by assumptions (a), (b) and (c)

$$I(t+\Delta t) = I(t) + \left\{ \begin{array}{c} \text{number of susceptibles} \\ \text{infected in time } \Delta\text{t} \end{array} \right\} - \left\{ \begin{array}{c} \text{number of infectives} \\ \text{removed in time } \Delta\text{t} \end{array} \right\},$$

and by assumptions (a), (c) and (d)

$$R(t+\Delta t) = R(t) + \left\{ \begin{array}{c} \text{number of infectives} \\ \text{removed in time } \Delta\text{t} \end{array} \right\}.$$

However, from assumptions (c) and (f)

$$\begin{aligned} \left\{ \begin{array}{c} \text{number of susceptibles} \\ \text{infected in time } \Delta\text{t} \end{array} \right\} &= \beta S I \Delta t \\ \left\{ \begin{array}{c} \text{number of infectives} \\ \text{removed in time } \Delta\text{t} \end{array} \right\} &= \gamma I \Delta t. \end{aligned}$$

Combining all these equations and dividing by Δt and passing with it to 0 we obtain the coupled system of nonlinear differential equations

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI,\\ \frac{dI}{dt} &= \beta SI - \gamma I,\\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{1.3.16}$$

where α, β are proportionality constants. Note that R does not appear in the first two equations so that we can consider separately and then find R by direct integration. The first two equations are then a continuous analogue of the system (1.2.9) with $B = 0$. Note that a simpler form of the equation for I in the discrete case follows from the fact that due to precisely one week recovering time the number of removed each week is equal to the number of infectives the previous week so that these two cancel each other in the equation for I.

Predator–prey model – a system of nonlinear equations
Systems of coupled nonlinear differential equations similar to (1.3.16) appear in numerous applications. One of the most famous is the Lotka-Volterra, or predator-prey model, created to explain why in a period of reduced fishing during the World War I, the number of sharks substantially increased. We shall describe it on the original example of small fish – shark interaction.

To describe the model, we consider two populations: of smaller fish and sharks, with the following influencing factors taken into account.

(i) Populations of fish and sharks display an exponential growth when considered in isolation. However, the growth rate of sharks in absence of fish is negative due to the lack of food.

(ii) Fish is preyed upon by sharks resulting in the decline in fish population. It is assumed that each shark eats a constant fraction of the fish population.

(iii) The population of sharks increases if there is more fish. The additional number of sharks is proportional to the number of available fish.

(iv) Fish and sharks are being fished indiscriminately, that is, the number of sharks and fish caught by fishermen is directly proportional to the present populations of fish and sharks, respectively, with the same proportionality constant.

If we denote by x and y the sizes of fish and shark populations, then an argument, similar to that leading to (1.3.16), gives the following system

$$\begin{aligned} \frac{dx}{dt} &= (r-f)x - \alpha xy, \\ \frac{dy}{dt} &= -(s+f)y + \beta xy \end{aligned} \tag{1.3.17}$$

where α, β, r, s, f are positive constants.

Chapter 2

Basic differential equations: models and solutions

2.1 Basic information on differential equations

What precisely do we mean by a differential equation? The more familiar notion of an algebraic equation, like for example the quadratic equation $x^2 - 4x - 5 = 0$, states something about a number x. It is sometimes called an open statement since the number x is left unspecified, and the statement's truth depends on the value of x. Solving the equation then amounts to finding values of x for which the open statement turns into a true statement.

Algebraic equations arise in modelling processes where the unknown quantity is a number (or a collection of numbers) and all the other relevant quantities are constant. As we observed in the first chapter, if the data appearing in the problem are variable and we describe a changing phenomenon, then the unknown will be rather a function (or a sequence). If the changes occur over very short interval, then the modelling usually will have to balance small increments of this function and the data of the problem and will result typically in an equation involving the derivatives of the unknown function. Such an equation is called a differential equation.

Differential equations are divided into several classes. The main two classes are ordinary differential equations (ODEs) and partial differential equations (PDEs). As suggested by the name, ODEs are equations where the unknown function is a function of one variable and the derivatives involved in the equation are ordinary derivatives of this function. A partial differential equation involves functions of several variables and thus expresses relations between partial derivatives of the unknown function.

In this course we shall be concerned solely with ODEs and systems of ODEs. Symbolically, the general form of ODE is

$$F(y^{(n)}, y^{(n-1)}, \ldots y', y, t) = 0, \tag{2.1.1}$$

where F is a given function of $n+2$ variables. For example, the equation of exponential growth can be written as $F(y', y, t) = y' - ry$ so that the function F is a function of two variables (constant with respect to t) and acting into r. Systems of differential equations can be also written in the form (2.1.1) if we accept that both F and y (and all the derivatives of y) can be vectors. For example, in the case of the epidemic spread (1.3.16) we have a system of ODEs which can be written as

$$\mathbf{F}(\mathbf{y}, t) = 0,$$

with three-dimensional vector $\mathbf{y} = (S, I, R)$ and the vector $\mathbf{F} = (F_1, F_2, F_3)$ with $F_1(S, I, R, t) = -\beta SI, F_2(S, I, R, t) = \beta SI - \gamma I$ and $F_3(S, I, R, t) = \gamma I$.

What does it mean to solve a differential equation? For algebraic equations, like the one discussed at the beginning, we can apply the techniques learned in the high school finding the discriminant of the equation $\Delta = (-4)^2 - 4 \cdot 1 \cdot (-5) = 36$ so that $x_{1,2} = 0.5(4 \pm 6) = 5, -1$. Now, is this the solution to our equation? How can we check it? The answer is given above – the solution is a number (or a collection of numbers) that turns the equation into a true statement. In our case, $5^2 - 20 - 5 = 0$ and $(-1)^2 - 4(-1) - 5 = 0$, so both numbers are solutions to the equation.

Though presented in a simple context, this is a very important point.

To solve a problem is to find a quantity that satisfies all the conditions of this problem.

This simple truth is very often forgotten as students tend to apply mechanically steps they learned under say, ”techniques for solving quadratic equations” or ”techniques of integration” labels and looking for answers or model solutions ”out there” though the correctness of the solution in most cases can be checked directly.

The same principle applies to differential equations. That is, to solve the ODE (2.1.1) means to find an n-times continuously differentiable function $y(t)$ such that for any t (from some interval)

$$F(y^{(n)}(t), y^{(n-1)}(t), \ldots y'(t), y(t), t) \equiv 0.$$

Once again, there are many techniques for solving differential equations. Some of them give only possible candidates for solutions and only checking that these suspects really turn the equation into the identity can tell us whether we have obtained the correct solution or not.

Example 2.1.1. As an example, let us consider which of these functions $y_1(t) = 30e^{2t}, y_2(t) = 30e^{3t}$ and $y_3(t) = 40e^{2t}$ solves the equation $y' = 2y$. In the first case, LHS is equal to $60e^{2t}$ and RHS is $2 \cdot 30e^{2t}$ so that $LHS = RHS$ and we have a solution. In the second case we obtain LHS $= 90e^{3t} \neq 2 \cdot 30e^{3t} =$ RHS so that y_2 is not a solution. In the same way we find that y_3 satisfies the equation.

Certainly, being able to check whether a given function is a solution is not the same as actually finding the solution. Thus, this example rises the following three questions.

1. Can we be sure that a given equation possesses a solution at all?
2. If we know that there is a solution, are there systematic methods to find it?
3. Having found a solution, can we be sure that there are no other solutions?

Question 1 is usually referred to as the **existence problem** for differential equations, and Question 3 as the **uniqueness problem**. Unless we deal with very simple situations these should be addressed before attempting to find a solution. After all, what is the point of trying to solve equation if we do not know whether the solution exists, and whether the solution we found is the one we are actually looking for, that is, the solution of the real life problem the model of which is the differential equation.

Let us discuss briefly Question 1 first. Roughly speaking, we can come across the following situations.

1. No function exists which satisfies the equation.
2. The equation has a solution but no one knows what it looks like.
3. The equation can be solved in a closed form, either

   in elementary functions,

   or in quadratures.

Case 1 is not very common in mathematics and it should never happen in mathematical modelling. In fact, if a given equation was an exact reflection of a real life phenomenon, then the fact that this phenomenon exists would ensure that the solution to this equation exists also. For example, if we have an equation describing a flow of water, then the very fact that water flows would be sufficient to claim that the equation must have a solution.

However, in general, models are imperfect reflections of real life and therefore it may happen that in the modelling process we missed a crucial fact, rendering thus the final equation unsolvable. Thus, checking that a given equation is solvable serves as an important first step in validation of the model. Unfortunately, these problems are usually very difficult and require quite advanced mathematics that is beyond the scope of this course. On the other hand, all the equations we will be dealing with are classical and the fundamental problems of existence and uniqueness for them have been positively settled at the beginning of the 20th century.

Case 2 may look somewhat enigmatic but, as we said above, there are advanced theorems allowing to ascertain the existence of solution without actually displaying them. This should be not surprising: after all, we know that the Riemann integral of any continuous function exists though in many cases we cannot evaluate it explicitly.

Even if we do not know a formula for the solution, the situation is not hopeless. Knowing that the solution exists, we have an array of approximate, numerical methods at our disposal. Using them we are usually able to find the numerical values of the solution with arbitrary accuracy. Also, very often we can find important features of the solution without knowing it. These feature include e.g. the long time behaviour of the solution, that is, whether it settles at a certain equilibrium value or rather oscillates, whether it is monotonic etc. These questions will be studied by in the final part of our course.

Coming now to Case 3 and to an explanation of the meaning of the terms used in the subitems, we note that clearly an ideal situation is if we are able to find the solution as an algebraic combination of elementary functions

$$\begin{aligned} y(t) &= \text{combination of elementary functions like :} \\ &\quad \sin t, \cos t, \ln t, \text{exponentials, polynomials...} \end{aligned}$$

Unfortunately, this is very rare for differential equation. Even the simplest cases of differential equations involving only elementary functions may fail to have such solutions.

Example 2.1.2. For example, consider is the equation

$$y' = e^{-t^2}.$$

Integrating, we find that the solution must be

$$y(t) = \int e^{-t^2} dt$$

but, on the other hand, it is known that this integral cannot be expressed as a combination of elementary functions.

This brings us to the definition of *quadratures*. We say that an equation is *solvable in quadratures* if a solution to this equation can be written in terms of integrals of elementary functions (as above). Since we know that every continuous function has an antiderivative (though often we cannot find this antiderivative explicitly), it is almost as good as finding the explicit solution to the equation.

Having dealt with Questions 1 and 2 above, that is, with existence of solutions and solvability of differential equations, we shall move to the problem of uniqueness. We have observed in Example 2.1.1 that the differential equation by itself defines a family of solutions rather than a single function. In this particular case this class depend on an arbitrary parameter. Another simple example of a second order differential equation $y'' = t$, solution of which can be obtained by a direct integration as $y = \frac{1}{6}t^3 + C_1t + C_2$, shows that in equations of the second order we expect the class of solutions to depend on 2 arbitrary parameters. It can be then expected that the class of solutions for an nth order equation will contain n arbitrary parameters. Such a full class is called the *general solution* of the differential equation. By imposing the appropriate number of *side conditions* we can specify the constants obtaining thus a *special solution* - ideally one member of the class.

A side condition may take all sorts of forms, like "at $t = 15$, y must have the value of 0.4" or "the area under the curve between $t = 0$ and $t = 24$ must be 100". Very often, however, it specifies the initial value of $y(0)$ of the solution and the derivatives $y^k(0)$ for $k = 1, \ldots, n-1$. In this case the side conditions are called the *initial conditions*.

After these preliminaries we shall narrow our consideration to a particular class of problems for ODEs.

2.2 Cauchy problem for first order equations

In this section we shall be concerned with *first order* ordinary differential equations which are solved with respect to the derivative of the unknown function, that is, with equations which can be written as

$$\frac{dy}{dt} = f(t, y), \tag{2.2.1}$$

where f is a given function of two variables.

In accordance with the discussion of the previous session, we shall be looking for solutions to the following Cauchy problem

$$\begin{aligned} y' &= f(t, y), \\ y(t_0) &= y_0 \end{aligned} \tag{2.2.2}$$

where we abbreviated $\frac{dy}{dt} = y'$, and t_0 and y_0 are some given numbers.

Several comments are in place here. Firstly, even though in such a simplified form, the question of solvability of the problem (2.2.2) is almost as difficult as that of (2.1.1). Before we embark on studying this problem, we again emphasize that to solve (2.2.2) is to find a function $y(t)$ that is continuously differentiable at least in some interval (t_1, t_2) containing t_0, that satisfies

$$\begin{array}{rcl} y'(t) & \equiv & f(t, y(t)) \quad \text{for all} \quad t \in (t_1, t_2) \\ y(t_0) & = & y_0. \end{array}$$

Let consider the following example.

Example 2.2.1. Check that the function $y(t) = \sin t$ is a solution to the problem

$$\begin{array}{rcl} y' & = & \sqrt{1 - y^2}, \quad t \in (0, \pi/2), \\ y(\pi/2) & = & 1 \end{array}$$

Solution. LHS: $y'(t) = \cos t$, RHS: $\sqrt{1 - y^2} = \sqrt{1 - \sin^2 t} = |\cos t| = \cos t$ as $t \in (0, \pi/2)$. Thus the equation is satisfied. Also $\sin \pi/2 = 1$ so the "initial" condition is satisfied.

Note that the function $y(t) = \sin t$ is not a solution to this equation on a larger interval $(0, a)$ with $a > \pi/2$ as for $\pi/2 < t > 3\pi/2$ we have LHS: $y'(t) = \cos t$ but RHS: $\sqrt{1 - y^2} = |\cos t| = -\cos t$, since $\cos t < 0$.

How do we know that a given equation has a solution? For an equation in the (2.2.1) form the answer can be given in relatively straightforward terms, though it is still not easy to prove.

Theorem 2.2.2. [Peano] *If the function f in (2.2.2) is continuous in some neighbourhood of the point (t_0, y_0), then the problem (2.2.2) has at least one solution in some interval (t_1, t_2) containing t_0.*

Thus, we can safely talk about solutions to a large class of ODEs of the form (2.2.1) even without knowing their explicit formulae.

As far as uniqueness is concerned, we know that the equation itself determines a class of solutions; for first order ODE this class is a family of functions depending on one arbitrary parameter. Thus, in principle, imposing one additional condition, as e.g. in (2.2.2), we should be able to determine this constant so that the Cauchy problem (2.2.2) should have only one solution. Unfortunately, in general this is no so as demonstrated in the following example.

Example 2.2.3. The Cauchy problem

$$\begin{aligned} y' &= \sqrt{y}, \quad t > 0 \\ y(0) &= 0, \end{aligned}$$

has at least two solutions: $y \equiv 0$ and $y = \frac{1}{4}t^2$.

Fortunately, there is a large class of functions f for which (2.2.2) has exactly one solution. This result is known as the Picard Theorem which we state below.

Theorem 2.2.4. [Picard] *Let f and $\partial f/\partial y$ be continuous in some neighbourhood of (t_0, y_0). Then the Cauchy problem (2.2.2) has exactly one solution defined on some neighbourhood of t_0.*

Example 2.2.5. We have seen in Example 2.2.3 that there are two solutions to the problem

$$\begin{aligned} y' &= \sqrt{y}, \quad t \geq 0 \\ y(0) &= 0. \end{aligned}$$

In this case $f(t, y) = \sqrt{y}$ and $f_y = 1/2\sqrt{y}$; obviously f_y is not continuous in any neighbourhood of 0 and we may expect troubles.

Another example of a nonuniqueness is offered by

$$\begin{aligned} y' &= (\sin 2t)y^{1/3}, \quad t \geq 0 \\ y(0) &= 0, \end{aligned} \tag{2.2.3}$$

Direct substitution shows that we have at least 3 different solutions to this problem: $y_1 \equiv 0, y_2 = \sqrt{8/27}\sin^3 t$ and $y_3 = -\sqrt{8/27}\sin^3 t$. These are shown at the picture below.

In the next few sections we shall discuss some cases when an ODE can be solved explicitly, either in elementary functions, or in quadratures.

2.3 Equations admitting closed form solutions

2.3.1 Separable equations

The simplest differential equation is of the form

$$\frac{dy}{dt} = g(t). \tag{2.3.1}$$

Its simplicity follows from the fact that the function $f(t, y)$ of (2.2.1) here is a function of the independent variable t only so that both sides can be

Fig 2.1 Multiple solutions of the problem (2.2.3).

integrated with respect to t. Note, that this is impossible if the right hand side of (2.2.1) depends on y is an unknown function of t. There is, however, a class of equations for which a simple modification of the above procedure works.

Consider an equation that can be written in the form

$$\frac{dy}{dt} = \frac{g(t)}{h(y)}, \tag{2.3.2}$$

where g and h are known functions. Equations that can be put into this form are called *separable* equations. Firstly, we note that any constant function $y = y_0$, such that $1/h(y_0) = 0$, is a special solution to (2.3.2), as the derivative of a constant function is equal to zero. We call such solutions *stationary or equilibrium solutions.*

To find a general solution, we assume that $1/h(y) \neq 0$, that is $h(y) \neq \infty$. Multiplying then both sides of (2.3.2) by $h(y)$ to get

$$h(y)\frac{dy}{dt} = g(t) \tag{2.3.3}$$

and observe that, denoting by $H(y) = \int h(y)dy$ the antiderivative of h, we can write (2.3.2) in the form

$$\frac{d}{dt}(H(y(t))) = g(t),$$

that closely resembles (2.3.1). Thus, upon integration we obtain

$$H(y(t)) = \int g(t)dt + c, \tag{2.3.4}$$

where c is an arbitrary constant of integration. The next step depends on the properties of H: for instance, if $H : \mathbb{R} \to \mathbb{R}$ is monotonic, then we can find y explicitly for all t as

$$y(t) = H^{-1}\left(\int g(t)dt + c\right).$$

Otherwise, we have to do it locally, around the initial values. To explain this, we solve the initial value problem for separable equation.

$$\begin{aligned} \frac{dy}{dt} &= \frac{g(t)}{h(y)}, \\ y(t_0) &= y_0, \end{aligned} \tag{2.3.5}$$

Using the general solution (2.3.4) (with definite integral) we obtain

$$H(y(t)) = \int\limits_{t_0}^{t} g(s)ds + c,$$

we obtain

$$H(y(t_0)) = \int\limits_{t_0}^{t_0} a(s)ds + c,$$

which, due $\int\limits_{t_0}^{t_0} a(s)ds = 0$, gives

$$c = H(y(t_0)),$$

so that

$$H(y(t)) = \int\limits_{t_0}^{t} g(s)ds + H(y(t_0)).$$

We are interested in the existence of the solution at least close to t_0, which means that H should be invertible close to y_0. From the Implicit Function Theorem we obtain that this is possible if H is differentiable in a neighbourhood of y_0 and $\frac{\partial H}{\partial y}(y_0) \neq 0$. But $\frac{\partial H}{\partial y}(y_0) = h(y_0)$, so we are back at Picard's theorem: if $h(y)$ is differentiable in the neighbourhood of y_0 with $h(y_0) \neq 0$ (if $h(y_0) = 0$, then the equation (2.3.2) does not make sense at y_0, and g is continuous, then $f(t, y) = g(t)/h(y)$ satisfies the assumptions of the theorem in some neighbourhood of (t_0, y_0).

Example 2.3.1. Find the general solution of the equation

$$y' = t^2/y^2.$$

This equation is equivalent to

$$\frac{d}{dt}\left(\frac{y(t)^3}{3}\right) = t^2,$$

hence $y^3(t) = t^3 + c$, where c is an arbitrary constant and, since the cubic function is monotonic,

$$y(t) = (t^3 + c)^{1/3}.$$

Remark 2.3.2. Note that Picard's theorem gives only a sufficient condition for the existence of the unique solution. In the example above the assumptions are obviously violated at $t = 0$ and $y = 0$, that is, the theorem doesn't give an answer as to whether there exists a unique solution to the problem

$$y' = t^2/y^2, \qquad y(0) = 0.$$

However, direct computation shows that $y(t) = t$ is the unique solution to this problem. This is possible due to the cancellation of singularities.

On the other hand, consider a similar problem

$$y' = t/y, \qquad y(0) = 0.$$

Then the general solution is given by

$$y^2(t) = t^2 + c,$$

and with the initial condition we obtain

$$y^2(t) = t^2.$$

Quadratic function is not invertible close to 0 and it produces two solutions $y(t) = t$ and $y(t) = -t$.

Example 2.3.3. Solve the initial value problem

$$y' = 1 + y^2, \qquad y(0) = 0.$$

We transform the equation as

$$\frac{d}{dt}\tan^{-1} y(t) = 1$$

which gives

$$\tan^{-1} y = t + c,$$

and from the initial condition $c = 0$. Therefore, the solution is given by

$$y = \tan t.$$

We point out that $y \to \pm\infty$ as $t \to \pm\pi/2$. In other words the solution exists only on the interval $(-\pi/2, \pi/2)$ but there is seemingly nothing at all in the form of the equation which would suggest such a behaviour.

In the previous example the solution ceased to exist beyond $\pi/2$ and $-\pi/2$ as it becomes infinite at these point. We call such a situation a blow up. The next example shows that there might be another way for a solution to cease to exist.

Example 2.3.4. Find the solution to the following initial value problem

$$yy' + (1+y^2)\sin t = 0, \qquad y(0) = 1.$$

In a standard way we obtain

$$\int\limits_1^y \frac{rdr}{1+r^2} = -\int\limits_0^t \sin s ds,$$

which gives

$$\frac{1}{2}\ln(1+y^2) - \frac{1}{2}\ln 2 = \cos t - 1.$$

Solving this equation for $y(t)$ gives

$$y(t) = \pm(2e^{-4\sin^2 t/2} - 1)^{1/2}.$$

To determine which sign we should take we note that $y(0) = 1 > 0$, thus the solution is given by

$$y(t) = (2e^{-4\sin^2 t/2} - 1)^{1/2}.$$

Clearly, this solution is only defined when

$$2e^{-4\sin^2 t/2} - 1 \geq 0,$$

that is

$$e^{4\sin^2 t/2} \leq 2.$$

Since the natural logarithm is increasing we may take logarithms of both sides preserving the direction of inequality. We get this way

$$4\sin^2 t/2 \leq \ln 2$$

and consequently

$$\left|\frac{t}{2}\right| \leq \sin^{-1}\frac{\sqrt{\ln 2}}{2}.$$

Therefore, the solution $y(t)$ exists only on the open interval $(-2\sin^{-1}\frac{\sqrt{\ln 2}}{2}, 2\sin^{-1}\frac{\sqrt{\ln 2}}{2})$. However, contrary to the previous example, the solution does not blow up at the end-points, but simply vanishes.

In the last example we shall see that sometimes it is sensible to adopt a little different approach to the solutions of a differential equation.

Fig 2.2 The graph of the solution in Example 2.3.4.

Example 2.3.5. Find all solutions of the differential equation

$$\frac{dy}{dt} = -\frac{t}{y}. \tag{2.3.6}$$

Standard approach gives

$$\int \frac{d}{dt}\left(\frac{1}{2}y^2(t)\right)dt = -\int tdt$$

which gives

$$y^2 + t^2 = c^2. \tag{2.3.7}$$

The curves described by equation (2.3.7) are closed (they are, in fact, circles), thus we don't have single valued solutions. However, if (2.3.6) describes a motion of a point in (t, y)-plane, then we can interpret (2.3.7) as traces of this motion and therefore solution in such an implicit form has a physical sense.

Thus, it is not always necessary or desirable to look for the solution in functional form $y = y(t)$ as, depending on the problem, the solution in the implicit form $F(y, t) = c$ may be the proper one. In such a case curves $F(y, t) = c$ are called *solution curves* of the equation.

2.3.2 Linear ordinary differential equations of first order

Definition 2.3.6. *The general first order linear differential equation is*

$$\frac{dy}{dt} + a(t)y = b(t). \tag{2.3.8}$$

Functions a and b are known continuous functions of t.

Let us recall that we call this equation ”linear” because the dependent variable y appears by itself in the equation. In other words, y' and y appear in the equation only possibly multiplied by a known function and not in the form $yy', \sin y$ or $(y')^3$.

It is not immediate how to solve (2.3.8), therefore we shall simplify it even further by putting $b(t) = 0$. The resulting equation

$$\frac{dy}{dt} + a(t)y = 0, \tag{2.3.9}$$

is called the *reduced* first order linear differential equation. We observe that the reduced equation is a separable equation and thus can be solved easily. As in (2.3.3) we obtain that if $y(t) \neq 0$ for any t, then

$$\frac{1}{y(t)}\frac{dy}{dt} = \frac{d}{dt}\ln|y(t)|,$$

so that

$$\frac{d}{dt}\ln|y(t)| = -a(t),$$

and, by direct integration,

$$\ln|y(t)| = -\int a(t)dt + c_1$$

where c_1 is an arbitrary constant of integration. Taking exponentials of both sides yields

$$|y(t)| = \exp\left(-\int a(t)dt + c_1\right) = c_2\exp\left(-\int a(t)dt\right)$$

where c_2 is an arbitrary positive constant: $c_2 = \exp c_1 > 0$. We have to get rid of the absolute value bars at $y(t)$. To do this observe that in the derivation we required that $y(t) \neq 0$ for any t, thus y, being a continuous function, must be of a constant sign. Hence,

$$y(t) = \pm c_2\exp\left(-\int a(t)dt\right) = c_3\exp\left(-\int a(t)dt\right) \tag{2.3.10}$$

where this time c_3 can be either positive ore negative.

Are these all the possible solutions to (2.3.9)? Solution (2.3.10) was derived under provision that $y \neq 0$. We clearly see that $y \equiv 0$ is a solution to (2.3.9) but, fortunately, this solution can be incorporated into (2.3.10) by allowing c_3 to be zero.

However, we still have not ruled out the possibility that the solution can cross the x-axis at one or more points. To prove that this is impossible, we must resort to the Picard theorem. First of all we note that the function $f(t, y)$ is here given by

$$f(t, y) = a(t)y$$

and $\partial f/\partial y = a(t)$ so that, if a is a continuous function, the assumptions of Picard's theorem are satisfied in any neighbourhood of any pair (t_0, y_0). If there was a solution satisfying $y(t_0) = 0$ for some t_0, then from uniqueness part of Picard's theorem, this solution should be identically zero, as $y(t) \equiv 0$ is a solution to this problem.. In other words, if a solution to (2.3.9) is zero at some point, then it is identically zero.

After this considerations we can claim that all the solutions to (2.3.9) are of the form

$$y(t) = c\exp\left(-\int a(t)dt\right), \tag{2.3.11}$$

where c is an arbitrary real constant.

How this solution can help us with solving the nonhomogeneous equation

$$\frac{dy}{dt} + a(t)y = b(t)? \tag{2.3.12}$$

If we could repeat the trick used in the solution of (2.3.9) and write the above equation in the form

$$\frac{d}{dt}(\text{"}something\text{"}) = b(t),$$

then the solution would be easy. However, the expression $dy/dt + a(t)y$ does not appear to be a derivative of any simple expression and we have to help it a little bit. We shall multiply both sides of (2.3.12) by some continuous nonzero function μ (for a time being, unknown) to get the equivalent equation

$$\mu(t)\frac{dy}{dt} + \mu(t)a(t)y = \mu(t)b(t), \tag{2.3.13}$$

and ask the question: for which function μ the left-hand side of (2.3.13) is a derivative of some simple expression? We note that the first term on the left-hand side comes from

$$\frac{d\mu(t)y}{dt} = \mu(t)\frac{dy}{dt} + \frac{d\mu(t)}{dt}y,$$

thus, if we find μ in such a way that

$$\mu(t)\frac{dy}{dt} + \frac{d\mu(t)}{dt}y = \mu(t)\frac{dy}{dt} + \mu(t)a(t)y,$$

that is

$$\frac{d\mu(t)}{dt}y = \mu(t)a(t)y,$$

then we are done. Note that an immediate choice is to solve the equation

$$\frac{d\mu(t)}{dt} = \mu(t)a(t),$$

but this is a separable equation, the general solution of which is given by (2.3.11). Since we need only one such function, we may take

$$\mu(t) = \exp\left(\int a(t)dt\right).$$

The function μ is called an *integrating factor* of the equation (2.3.12). With such function, (2.3.12) can be written as

$$\frac{d}{dt}\mu(t)y = \mu(t)b(t),$$

thus

$$\mu(t)y = \int \mu(t)b(t)dt + c$$

where c is an arbitrary constant of integration. Finally

$$\begin{aligned} y(t) &= \frac{1}{\mu(t)}\left(\int \mu(t)b(t)dt + c\right) \\ &= \exp\left(-\int a(t)dt\right)\left(\int b(t)\exp\left(\int a(t)dt\right)dt + c\right) \end{aligned} \qquad (2.3.14)$$

It is worthwhile to note that the solution consists of two parts: the general solution to the reduced equation associated with (2.3.12)

$$c\exp\left(-\int a(t)dt\right)$$

and, what can be checked by direct differentiation, a particular solution to the full equation.

If we want to find a particular solution satisfying $y(t_0) = y_0$, then we write (2.3.14) using definite integrals

$$y(t) = \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right)\left(\int\limits_{t_0}^{t} b(s)\exp\left(\int\limits_{t_0}^{s} a(r)dr\right)ds + c\right)$$

and use the fact that $\int\limits_{t_0}^{t_0} f(s)ds = 0$ for any function f. This shows that the part of the solution satisfying the nonhomogeneous equation:

$$y_b(t) = \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right)\int\limits_{t_0}^{t} b(s)\exp\left(\int\limits_{t_0}^{s} a(r)dr\right)ds$$

takes on the zero value at $t = t_0$. Thus

$$y_0 = y(t_0) = c$$

and the solution to the initial value problem is given by

$$y(t) = y_0 \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right) + \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right)\int\limits_{t_0}^{t} b(s)\exp\left(\int\limits_{t_0}^{s} a(r)dr\right)ds. \tag{2.3.15}$$

Once again we emphasize that the first term of the formula above solves the reduced ($b(t) = 0$) equation with the desired initial value ($y(0) = y_0$) whereas the second solves the full equation with the initial value equal to zero.

Again, Picard's theorem shows that there are no more solutions to than those given by (2.3.15). Why? Let $y_1(t)$ be a solution to (2.3.12). For some $t = t_0$ this will take on the value $y_1(t_0)$. But we know that there is a solution to (2.3.12) given by

$$y(t) = y_1(t_0) \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right) + \exp\left(-\int\limits_{t_0}^{t} a(s)ds\right)\int\limits_{t_0}^{t} b(s)\exp\left(\int\limits_{t_0}^{s} a(r)dr\right)ds,$$

and by Picard's theorem (this time $f(t, y) = -a(t)y + b(t)$ but the assumptions are still satisfied), $y_1(t) = y(t)$.

Example 2.3.7. Find the general solution of the equation

$$y' - 2ty = t.$$

Here $a(t) = -2t$ so that

$$\mu(t) = \exp\left(-2\int t dt\right) = e^{-t^2}.$$

Multiplying both sides of the equation by μ we obtain

$$e^{-t^2}y' - 2te^{-t^2} = te^{-t^2},$$

which can be written as

$$\frac{d}{dt}(e^{-t^2}y) = te^{-t^2}.$$

Upon integration we get

$$e^{-t^2}y = \int te^{-t^2}dt = -\frac{e^{-t^2}}{2} + c.$$

Thus the general solution is given by

$$y(t) = -\frac{1}{2} + ce^{t^2}.$$

Example 2.3.8. Find the solution of the equation

$$y' + 2ty = t,$$

satisfying $y(1) = 2$. Here $a(t) = 2t$ so that

$$\mu(t) = \exp\left(2\int t dt\right) = e^{t^2}.$$

As above, we obtain

$$e^{t^2}y' + 2te^{t^2} = \frac{d}{dt}(e^{t^2}y) = te^{t^2}. \tag{2.3.16}$$

General solution is

$$y(t) = \frac{1}{2} + ce^{-t^2}$$

and to find c we have

$$2 = y(1) = \frac{1}{2} + ce^{-1}$$

which gives

$$c = \frac{3e}{2}$$

and

$$y(t) = \frac{1}{2} + \frac{3}{2}e^{1-t^2}.$$

Alternatively, we can integrate (2.3.16) from 1 to t to get

$$e^{s^2}y(s)\Big|_1^t = \frac{e^{s^2}}{2}\Bigg|_1^t$$

and further

$$e^{t^2}y - 2e = e^{t^2}2 - \frac{e}{2}.$$

Thus again

$$y(t) = \frac{1}{2} + \frac{3}{2}e^{1-t^2}.$$

2.3.3 Equations of homogeneous type

In differential equations, as in integration, a smart substitution can often convert a complicated equation into a manageable one. For some classes of differential equations there are standard substitutions that transform them into separable equations. We shall discuss one such a class in detail.

A differential equation that can be written in the form

$$\frac{dy}{dt} = f\left(\frac{y}{t}\right), \tag{2.3.17}$$

where f is a function of the single variable $z = y/t$ is said to be of *homogeneous type*. Note that in some textbooks such equations are called *homogeneous equations* but this often creates confusion as the name homogeneous equation is generally used in another context.

How one can recognize that given a function $F(t, y)$ can be written as a function of y/t only. We note that such functions have the following property of homogeneity: for any constant $\lambda \neq 0$

$$f\left(\frac{\lambda y}{\lambda t}\right) = f\left(\frac{y}{t}\right).$$

The converse is also true: if for any $\lambda \neq 0$

$$F(\lambda t, \lambda y) = F(t, y) \tag{2.3.18}$$

then $F(t, y) = f(y/t)$ for some function f. In fact, taking in (2.3.18) $\lambda = 1/t$ we obtain

$$F(t, y) = F(\lambda t, \lambda y) = F\left(1, \frac{y}{t}\right) = f\left(\frac{y}{t}\right),$$

that is, $f(z) = F(1, z)$.

To solve (2.3.17) let us make substitution

$$y = tz \tag{2.3.19}$$

where z is the new unknown function. Then, by the product rule for derivatives

$$\frac{dy}{dt} = z + t\frac{dz}{dt}$$

and (2.3.17) becomes

$$z + t\frac{dz}{dt} = f(z),$$

or

$$t\frac{dz}{dt} = f(z) - z. \tag{2.3.20}$$

In (2.3.20) the variables are separable so it can be solved as in Subsection 2.3.1.

Example 2.3.9. Find the general solution of the equation

$$(3t - y)y' + t = 3y.$$

Writing this equation in the form (2.2.1) we obtain

$$\frac{y}{dt} = \frac{3y-t}{3t-y}$$

and $F(\lambda t, \lambda y) = \frac{3\lambda y - \lambda t}{3\lambda t - \lambda y} = \frac{3y-t}{3t-y} = F(t,y)$ so that the equation is of homogeneous type. Using the substitution (2.3.19), we find

$$z + t\frac{dz}{dt} = \frac{3z-1}{3-z},$$

so that

$$t\frac{dz}{dt} = \frac{z^2-1}{3-z}.$$

We observe that $z = \pm 1$ (or $y = \pm t$) are stationary solutions. Assuming $z \neq \pm 1$ we find the general solution is

$$\int \frac{3-z}{z^2-1} dz = \int \frac{dt}{t}.$$

Using partial fraction decomposition

$$\frac{3}{z^2-1} = \frac{3}{2}\left(\frac{1}{z-1} - \frac{1}{z+1}\right),$$

we obtain

$$\frac{3}{2}\ln\left|\frac{z-1}{z+1}\right| - \frac{1}{2}\ln|z^2-1| - \ln t + C$$

for some constant C. Consequently,

$$\frac{|z-1|}{(z+1)^2} = Dt$$

for the constant D defined as $D = e^C$. Returning to the original variable y we get

$$y = t + D(y+t)^2, \tag{2.3.21}$$

where we dropped absolute value bars allowing D to be an arbitrary real constant. Note that the special solution $y = t$ can be recovered from (2.3.21) by putting $D =$ whilst $y = -t$ cannot be recovered from it (unless we agree to write (2.3.21) as $(y+t)^2 = \frac{1}{D}(y-t)$ and put $D = \infty$).

2.3.4 Equations that can be reduced to first order equations

Some higher order equations can be reduced to equations of the first order. We shall discuss two such cases for second order equations.

Equations that do not contain the unknown function

If we have the equation of the form

$$F(y'', y', t) = 0, \tag{2.3.22}$$

then the substitution $z = y'$ reduces this equation to an equation of the first order

$$F(z', z, t) = 0. \tag{2.3.23}$$

If we can solve this equation

$$z = \phi(t, C),$$

where C is an arbitrary constant, then, returning to the original unknown function y, we obtain another first order equation

$$y' = \phi(t, C),$$

which is immediately solvable as

$$y(t) = \int \phi(t, C)dt + C_1.$$

Example 2.3.10. Find solutions to the equation

$$(y'')^2 = 4(y' - 1) \tag{2.3.24}$$

satisfying initial conditions: a) $y(0) = 0$ and $y'(0) = 2$, b) $y(0) = 0$ and $y'(0) = 1$.

Eq. (2.3.24) does not contain y so it can be integrated following the method described above. Setting $z = y'$ we get

$$(z')^2 = 4(z - 1)$$

which gives two equations

$$z' = \pm 2\sqrt{z - 1}.$$

Each of these is a separable equation. There is a particular stationary solution $z = 1$ that is common to both equations. For $z \neq 1$ we obtain

$$\pm \int \frac{dz}{2\sqrt{z - 1}} = \int dt,$$

integrating which yields

$$\pm\sqrt{z - 1} = t + C_1.$$

Squaring, we find

$$z = 1 + (t + C_1)^2$$

Returning to the original unknown function y we find the particular solution solving $y' = 1$, that is,

$$y = t + C. \tag{2.3.25}$$

The general solution is obtained by solving

$$y' = 1 + (t + C_1)^2 \tag{2.3.26}$$

which gives

$$y = t + \frac{1}{3}(t + C_1)^3 + C_2. \tag{2.3.27}$$

To find solutions to the Cauchy problem a), we take the general solution (2.3.27) and substitute $t = 0$ getting

$$0 = y(0) = \frac{1}{3}C_1^3 + C_2,$$

and, using the condition for the derivative

$$2 = y'(0) = 1 + C_1^2.$$

Solving, we obtain $C_1^2 = \pm 1$ and $C_2 = \mp\frac{1}{3}$. Therefore we obtain two solutions to the Cauchy problem

$$\begin{aligned} y &= t + \frac{1}{3}(t+1)^3 - \frac{1}{3}, \\ y &= t + \frac{1}{3}(t-1)^3 + \frac{1}{3}. \end{aligned}$$

There are no other solutions, as there is no constant C in (2.3.25) that allows both initial conditions to be satisfied.

To find solution for the Cauchy problem b), we obtain from (2.3.27) and (2.3.26)

$$\begin{aligned} 0 &= \frac{1}{3}C_1^3 + C_2, \\ 1 &= 1 + C_1^2. \end{aligned}$$

This gives $C_1 = C_2 = 0$ which yields the solution

$$y = t + \frac{1}{3}t^3.$$

Moreover, we can check that the particular solution (2.3.25) can be made to satisfy these conditions by taking $C = 0$. Hence, the second solution to the Cauchy problem is given by

$$y = t.$$

Equations that do not contain the independent variable

Let us consider the equation

$$F(y'', y', y) = 0, \tag{2.3.28}$$

that does not involve the independent variable t. Such an equation can be also reduced to a first order equation, the idea, however, is a little more complicated. Firstly, we note that the derivative y' is uniquely defined by the function y. This means that we can write $y' = g(y)$ for some function g. Using the chain rule we obtain

$$y'' = \frac{d}{dt}y' = \frac{dg}{dy}(y)\frac{dy}{dt} = y'\frac{dg}{dy}(y) = g(y)\frac{dg}{dy}(y). \tag{2.3.29}$$

Substituting (2.3.29) into (2.3.28) gives the first order equation with y as an independent variable

$$F\left(g\frac{dg}{dy}, g, y\right) = 0. \tag{2.3.30}$$

If we solve this equation in the form $g(y) = \phi(y, C)$, then to find y we have to solve one more first order equation with t as the independent variable

$$\frac{dy}{dt} = \phi(y, C).$$

Example 2.3.11. Find the general solution to the equation

$$yy'' = (y')^2. \tag{2.3.31}$$

As the equation does not contain t, we can use the technique described above. Denoting $y' = g(y)$ we get $y'' = g\frac{dg}{dy}$, so that the equation (2.3.31) turns into

$$yg\frac{dg}{dy} = g^2.$$

We obtain the particular solution in the form $g = 0$. If $g \neq 0$, we can separate the variables, getting

$$\frac{1}{g}\frac{dg}{dy} = \frac{1}{y}.$$

Integration gives

$$\ln|g| = \ln|y| + C$$

or

$$g = \pm Cy$$

for some constant $C \neq 0$. Since $g(y) = y'$, we have either the solution $y = C_1$ for some constant C_1, or we have to solve the following equation

$$y' = \pm Cy.$$

Assuming $y \neq 0$, we separate variables getting

$$\int \frac{dy}{y} = \pm C \int dt,$$

so that

$$y = C_2 e^{\pm Ct}.$$

Note that for $C_2 = 0$ we recover the particular solution $y \equiv 0$ and for $C = 0$ we obtain the particular constant solutions coming from $g = 0$.

2.4 Miscellaneous applications

Here we shall use the theory developed in the previous subsections to provide solutions to some models introduced in Chapter 1. We start with the logistic equation.

Logistic equation

Let as recall the logistic equation

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), \qquad (2.4.1)$$

where r denotes the unrestricted growth rate and K the carrying capacity of the environment. Since the right-hand side does not contain t, we immediately recognize (2.4.1) as a separable equation. Let us consider the related Cauchy problem

$$\begin{aligned} \frac{dN}{dt} &= rN\left(1 - \frac{N}{K}\right), \\ N(t_0) &= N_0 \end{aligned} \qquad (2.4.2)$$

Separating variables and integrating we obtain

$$\frac{K}{r}\int\limits_{N_0}^{N} \frac{ds}{(K-s)s} = t - t_0.$$

To integrate the left-hand side we use partial fractions

$$\frac{1}{(K-s)s} = \frac{1}{K}\left(\frac{1}{s} + \frac{1}{K-s}\right)$$

which gives

$$\begin{aligned} \frac{K}{r}\int\limits_{N_0}^{N} \frac{ds}{(K-s)s} &= \frac{1}{r}\int\limits_{t_0}^{t}\left(\frac{1}{s} + \frac{1}{K-s}\right) ds \\ &= \frac{1}{r}\ln\frac{N}{N_0}\left|\frac{K-N_0}{K-N}\right|. \end{aligned}$$

From the above equation we see that $N(t)$ cannot reach K in any finite time, so if $N_0 < K$, then $N(t) < K$ for any t, and if $N_0 > K$, then $N(t) > K$ for all $t > 0$ (note that if $N_0 = K$, then $N(t) = K$ for all t – this follows from Picard's theorem). Therefore $(K - N_0)/(K - N(t))$ is always positive and

$$r(t - t_0) = \ln \frac{N}{N_0} \frac{K - N_0}{K - N}.$$

Exponentiating, we get

$$e^{r(t-t_0)} = \frac{N(t)}{N_0} \frac{K - N_0}{K - N(t)}$$

or

$$N_0(K - N(t))e^{r(t-t_0)} = N(t)(K - N_0).$$

Bringing all the terms involving N to the left-hand side and multiplying by -1 we get

$$N(t)\left(N_0 e^{r(t-t_0)} + K - N_0\right) = N_0 K e^{r(t-t_0)},$$

thus finally

$$N(t) = \frac{N_0 K}{N_0 + (K - N_0)e^{-r(t-t_0)}}. \qquad (2.4.3)$$

Let us examine (2.4.3) to see what kind of population behaviour it predicts. First observe that we have

$$\lim_{t\to\infty} N(t) = K,$$

hence our model correctly reflects the initial assumption that K is the maximal capacity of the habitat. Next, we obtain

$$\frac{dN}{dt} = \frac{rN_0K(K - N_0)e^{-r(t-t_0)}}{(N_0 + (K - N_0)e^{-r(t-t_0)})^2}$$

thus, if $N_0 < K$, the population monotonically increases, whereas if we start with the population which is larger then the capacity of the habitat, then such a population will decrease until it reaches K. Also

$$\frac{d^2N}{dt^2} = r\frac{d}{dt}(N(K - N)) = N'(K - 2N) = N(K - N)(K - 2N)$$

from which it follows that, if we start from $N_0 < K$, then the population curve is convex down for $N < K/2$ and convex up for $N > K/2$. Thus, as long as the population is small (less then half of the capacity), then the rate of growth increases, whereas for larger population the rate of growth decreases. This results in the famous *logistic* or *S-shaped* curve which is presented below for particular values of parameters $r = 0.02, K = 10$ and $t_0 = 0$ resulting in the following function:

$$N(t) = \frac{10N_0}{N_0 + (10 - N_0)e^{-0.2t}}.$$

Fig 2.3 Logistic curves with $N_0 < K$ (dashed line) and $N_0 > K$ (solid line) for $K = 10$ and $r = 0.02$.

To show how this curve compare with the real data and with the exponential growth we take the experimental coefficients $K = 10.76$ billion and $r = 0.029$. Then the logistic equation for the growth of the Earth population will read

$$N(t) = \frac{N_0(10.76 \times 10^9)}{N_0 + ((10.76 \times 10^9) - N_0)e^{-0.029(t-t_0)}}.$$

We use this function with the value $N_0 = 3.34 \times 10^9$ at $t_0 = 1965$. The comparison is shown on Fig. 2.4.

The waste disposal problem

Let us recall that the motion of a drum of waste dumped into the sea is governed by the equation (1.3.9)

$$\frac{d^2y}{dt^2} = \frac{1}{m}\left(W - B - c\frac{dy}{dt}\right). \tag{2.4.4}$$

The drums are dropped into the 100m deep sea. Experiments show that the drum could brake if its velocity exceeds 12m/s at the moment of impact. Thus, our aim is to determine the velocity of the drum at the sea bed level. To obtain numerical results, the mass of the drum is taken to be 239 kg, while its volume is 0.208 m^3. The density of the sea water is 1021 kg/m^3 and the drag coefficient is experimentally found to be $c = 1.18$kg/s. Thus, the mass of water displaced by the drum is 212.4 kg.

Fig 2.4 Human population on Earth. Comparison of observational data (points), exponential growth (solid line) and logistic growth (dashed line).

Equation (2.4.4) can be re-written as the first order equation for the velocity $V = dy/dt$.

$$V' + \frac{c}{m}V = g - \frac{B}{m}. \tag{2.4.5}$$

Since the drum is simply dumped into the sea, its initial velocity $V(0) = 0$. Since (2.4.5) is a linear equation, we find the integration factor $\mu(t) = e^{tc/m}$ and the general solution of the full equation is obtained as

$$V(t) = e^{-tc/m}\left(g - \frac{B}{m}\right)\int e^{tc/m}dt = \frac{mg - B}{c}(1 + Ce^{-tc/m})$$

for some constant C. Using the initial condition $V(0) = 0$, we find $C = -1$ so that

$$V(t) = \frac{mg - B}{c}(1 - e^{-tc/m}). \tag{2.4.6}$$

Integrating once again, we find

$$y(t) = \frac{mg - B}{c}\left(t + \frac{m}{c}e^{-tc/m}\right) + C_1.$$

To determine C_1 we recall that the coordinate system was set up in such a way that $y = 0$ was at the sea surface so we can take the initial condition to be $y(0) = 0$. Thus we obtain the equation

$$0 = y(0) = \frac{mg - B}{c}\frac{m}{c} + C_1,$$

so that

$$y(t) = \frac{mg - B}{c}\left(t + \frac{m}{c}e^{-tc/m}\right) - \frac{m(mg - B)}{c^2}. \tag{2.4.7}$$

Equation (2.4.6) expresses the velocity of the drum as a function of time t. To determine the impact velocity, we must compute the velocity at time t at which the drum hits the ocean floor, that is we have to solve for t the equation (2.4.7) with $y(t) = 100$m. Explicit solution of this equation is obviously impossible so let us try some other method.

As a first attempt, we notice from (2.4.6) that $V(t)$ is an increasing function of time and that it tends to a finite limit as $t \to \infty$. This limit is called the terminal velocity and is given by

$$V_T = \frac{mg - B}{c}. \tag{2.4.8}$$

Thus, for any time t the velocity is smaller that V_T and if $V_T < 12$m/s, we can be sure that the velocity of the drum when it hits the sea floor is also smaller that 12 m/s and it will not crack upon the impact. Substituting the data to (2.4.8) we obtain

$$V_T = \frac{(239 - 212.4)9.81}{1.18} \approx 221\text{m/s},$$

which is clearly way too large.

However, the approximation that gave the above figure is far too crude - this is the velocity the drum would eventually reach if it was allowed to descend indefinitely. As this is clearly not the case, we have to find the way to express the velocity as a function of the position y. This velocity, denoted by $v(y)$, is very different from $V(t)$ but they are related through

$$V(t) = v(y(t)).$$

By the chain rule of differentiation

$$\frac{dV}{dt} = \frac{dv}{dy}\frac{dy}{dt} = V\frac{dv}{dy} = v\frac{dv}{dy}.$$

Substituting this into (2.4.5) we obtain

$$mv\frac{dv}{dy} = (mg - B - cv)\,. \tag{2.4.9}$$

We have to supplement this equation with an appropriate initial condition. For this we have

$$v(0) = v(y(0)) = V(0) = 0.$$

This is a separable equation which we can solve explicitly. Firstly, we note that since $v < V_T = (mg - B)/c$, $mg - B - cv > 0$ all the time. Thus, we can divide both sides of (2.4.9) by $mg - B - cv$ and integrate, getting

$$\int\limits_0^v \frac{rdr}{mg - B - cr} = \frac{1}{m}\int\limits_0^y ds = \frac{y}{m}.$$

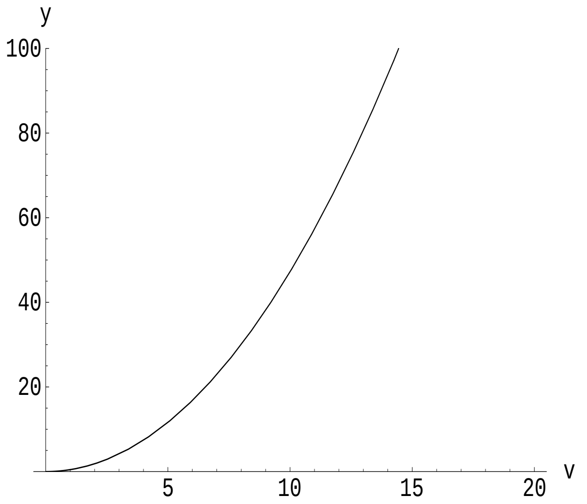

Fig 2.5 The depth as a function of velocity of the drum.

To find the left-hand side integral, we note that the degree of the numerator is the same as the degree of the denominator so that we have to decompose

$$\begin{aligned}\frac{r}{mg-B-cr} &= -\frac{1}{c}\frac{-cr}{mg-B-cr} = -\frac{1}{c}\frac{-mg+B+mg-Bcr}{mg-B-cr}\\ &= -\frac{1}{c}\left(\frac{-mg+B}{mg-B-cr}+1\right).\end{aligned}$$

Thus

$$\begin{aligned}\int\limits_0^v \frac{rdr}{mg-B-cr} &= -\frac{1}{c}\int\limits_0^v dr + \frac{mg-B}{c}\int\limits_0^v \frac{dr}{mg-B-cr}\\ &= -\frac{v}{c} - \frac{mg-B}{c^2}\ln\frac{mg-B-cv}{mg-B},\end{aligned}$$

and we obtain the solution

$$\frac{y}{m} = -\frac{v}{c} - \frac{mg-B}{c^2}\ln\frac{mg-B-cv}{mg-B}. \tag{2.4.10}$$

It seems that the situation here is as hopeless as before as we have $y = y(v)$ and we cannot find $v(y)$ explicitly. However, at least we have a direct relation between the quantities of interest, and not through intermediate parameter t that is irrelevant for the problem, as before. Thus, we can easily graph y as a function of v and estimate $v(100)$ from the graph shown at the Figure 2.5. We can also answer the question whether the velocity at $y = 100$m is higher that the critical velocity $v = 12$m/s. To do this, we note that from

(2.4.9) and the fact that $v < V_T$ we can infer that v is an increasing function of y. Let us find what y corresponds to $v = 12$m/s. Using the numerical data, we obtain

$$\begin{aligned} y(12) &= 239\left(-\frac{12}{1.18} - \frac{(239-212.4)9.81}{(1.18)^2}\ln\frac{(239-212.4)9.81 - 1.18\cdot 12}{(239-212.4)9.81}\right) \\ &\approx 68.4\text{m}, \end{aligned}$$

that is, the drum will reach the velocity of 12m/s already at the depth of 68.4m. Since v is a strictly increasing function of y, the velocity at 100m will be much higher and therefore the drum could crack on impact.

The satellite dish

In Chapter 1 we obtained the equation (1.3.10) for a reflecting surface:

$$\frac{dy}{dx} = \frac{y}{\sqrt{x^2+y^2}+x}. \tag{2.4.11}$$

Now we shall solve this equation. We observe that the right-hand side can be written as

$$\frac{\frac{y}{x}}{\sqrt{1+(\frac{y}{x})^2}+1},$$

for $x > 0$. This suggest the substitution used for homogeneous equations $z = y/x$. Since $y' = z'x + z$, we obtain

$$z'x\sqrt{1+z^2} + z'x + z\sqrt{1+z^2} + z = z,$$

which, after a simplification, can be written as

$$z'\left(\frac{1}{z} + \frac{1}{z\sqrt{z^2+1}}\right) = -\frac{1}{x}.$$

Integrating, we obtain

$$\ln|z| + \int\frac{dz}{z\sqrt{1+z^2}} = -\ln|x| + C'. \tag{2.4.12}$$

To integrate the second term, we use the hyperbolic substitution $z = \sinh\xi$ so that $dz = \cosh\xi d\xi$ and $\sqrt{1+z^2} = \sqrt{1+\sinh^2\xi} = \sqrt{\cosh^\xi} = \cosh\xi$, as $\cosh\xi$ is always positive. Thus we obtain

$$\begin{aligned} \int\frac{dz}{z\sqrt{1+z^2}} &= \int\frac{d\xi}{\sinh\xi} = \frac{1}{2}\int\frac{d\xi}{\sinh\xi/2\cosh\xi/2} \\ &= \frac{1}{2}\int\frac{d\xi}{\tanh\xi/2\cosh^2\xi/2} = \frac{1}{2}\int\frac{du}{u}, \end{aligned}$$

where in the last integral we used the change of variables $u = \tanh \xi/2$ so that $du = d\xi/2\cosh^2 \xi$. Continuing, we obtain

$$\int \frac{du}{u} = \ln |u| = \ln |\tanh \xi/2| + C.$$

But

$$\tanh \xi/2 = \frac{\sinh \xi/2}{\cosh \xi/2} = \frac{\sinh \xi/2 \cosh \xi/2}{\cosh^2 \xi/2}.$$

But

$$\sinh \xi/2 \cosh \xi/2 = \frac{1}{4}(e^{\xi/2} + e^{-\xi/2})(e^{\xi/2} - e^{-\xi/2}) = \frac{1}{4}(e^{\xi} + e^{-\xi}) = \frac{1}{2}\sinh \xi,$$

and

$$\cosh^2 \xi/2 = \frac{1}{4}(e^{\xi/2} + e^{-\xi/2})^2 = \frac{1}{4}(e^{\xi} + e^{-\xi} + 2) = \frac{1}{2}(\cosh \xi + 1),$$

so that returning to the original variable $z = \sinh \xi$ we obtain

$$\tanh \xi/2 = \frac{z}{1 + \sqrt{z^2 + 1}}.$$

Thus

$$\int \frac{dz}{z\sqrt{1 + z^2}} = \ln |z| - \ln(1 + \sqrt{z^2 + 1}) + C'.$$

Returning to (2.4.12) we obtain

$$\ln \frac{z^2}{1 + \sqrt{z^2 + 1}} = -\ln x/C$$

for some constant $C > 0$. Thus

$$\frac{z^2}{1 + \sqrt{z^2 + 1}} = \frac{C}{x},$$

and, returning to the original unknown function $z = y/x$,

$$\frac{y^2}{x + \sqrt{y^2 + x^2}} = C,$$

which, after some algebra, gives

$$y^2 - 2Cx = C^2. \tag{2.4.13}$$

This is an equation of the parabola with the vertex at $x = -C/2$ and with focus at the origin.

We note that this equation was obtained under the assumption that $x > 0$ so, in fact, we do not have the full parabola at this moment. The assumption

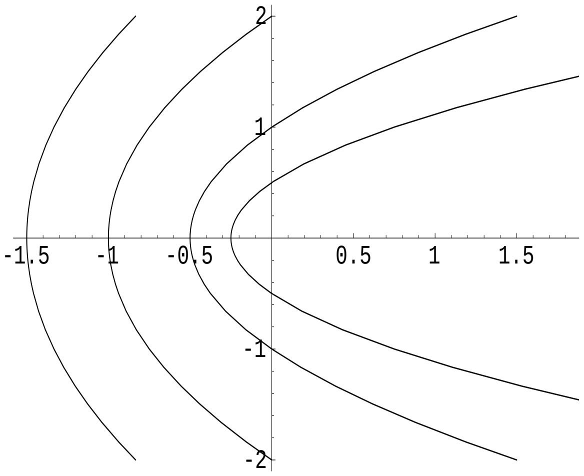

Fig 2.6 Different shapes of parabolic curves corresponding to various values of the constant C. In each case the focus is at the origin.

$x > 0$ was, however, purely technical and by direct substitution we can check that (2.4.13) determines the solution also for $-C/2 \leq x \leq 0$. In fact, $y = \pm\sqrt{2Cx + C^2}$ so that

$$LHS = \frac{dy}{dx} = \pm\frac{C}{\sqrt{2Cx + C^2}},$$

and

$$\begin{aligned} RHS &= \frac{y}{\sqrt{y^2 + x^2} + x} = \frac{\pm\sqrt{2Cx + C^2}}{\sqrt{x^2 + 2Cx + C^2} + x} = \frac{\pm\sqrt{2Cx + C^2}}{\sqrt{(x + C)^2} + x} \\ &= \pm\frac{C}{\sqrt{2Cx + C^2}}, \end{aligned}$$

where we used the fact that $x \geq -C/2$ so that $x + C > 0$. Thus LHS = RHS for any $x \geq -C/2$ and (2.4.13) gives the solution to the equation in the whole range of independent variables.

Pursuit equation

In this paragraph we shall provide the solution to the pursuit equation

$$xy'' = -\frac{v}{u}\sqrt{1 + (y')^2}. \qquad (2.4.14)$$

Firstly, we observe that this a second order equation that, however, does not contain the unknown function but only its higher derivatives. Thus, following the approach of Subsection 2.3.4 we introduce the new unknown $z = y'$ reducing thus (2.4.14) to a first order equation:

$$xz' = -k\sqrt{1 + z^2}$$

where we denoted $k = v/u$. This is a separable equation with non-vanishing right-hand side, so that we do not have stationary solutions. Separating variables and integrating, we obtain

$$\int \frac{dz}{1+z^2} = -k\ln(-C'x)$$

for some constant $C' > 0$, where we used the fact that in the model $x < 0$. Integration (for example as in the previous paragraph) gives

$$\ln(z + \sqrt{z^2+1}) = \ln C(-x)^{-k},$$

with $C = (C')^{-k}$, hence

$$z + \sqrt{z^2+1} = C(-x)^{-k},$$

from where, after some algebra,

$$z = \frac{1}{2}\left(C(-x)^{-k} - \frac{1}{C}(-x)^{k}\right). \tag{2.4.15}$$

Returning to the original unknown function y, where $y' = z$, and integration the above equation, we find

$$y(x) = \frac{1}{2}\left(\frac{1}{C(k+1)}(-x)^{k+1} - \frac{1C}{(1-k)}(-x)^{-k+1}\right) + C_1.$$

Let us express the constants C_1 and C_2 through initial conditions. We assume that the pursuer started from the position $(x_0, 0)$, $x_0 < 0$ and that at the initial moment the target was at the origin $(0,0)$. Using the principle of the pursuit, we see that the initial direction was along the x-axis, that is, we obtain the initial conditions in the form

$$y(x_0) = 0, \qquad y'(x_0) = 0.$$

Since $y' = z$, substituting $z = 0$ and $x = x_0$ in (2.4.15), we obtain

$$0 = y'(x_0) = z(x_0) = C(-x_0)^{-k} - \frac{1}{C}(-x_0)^{k}$$

which gives

$$C = (-x_0)^{k},$$

so that

$$y(x) = -\frac{x_0}{2}\left(\frac{1}{k+1}\left(\frac{x}{x_0}\right)^{k+1} - \frac{1}{1-k}\left(\frac{x}{x_0}\right)^{-k+1}\right) + C_1.$$

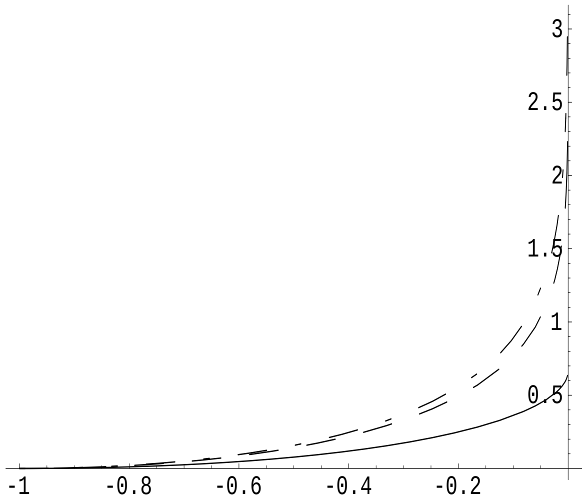

Fig 2.7 Pursuit curve for different values of k. $k = 0.5$ (solid line), $k = 0.9$ (dashed line), $k = 0.99$ (dot-dashed line).

To determine C_1 we substitute $x = x_0$ and $y(x_0) = 0$ above getting

$$0 = -\frac{x_0}{2}\left(\frac{1}{k+1} + \frac{1}{k-1}\right) + C_1$$

thus

$$C_1 = \frac{kx_0}{k^2-1}.$$

Finally,

$$y(x) = -\frac{x_0}{2}\left(\frac{1}{k+1}\left(\frac{x}{x_0}\right)^{k+1} - \frac{1}{1-k}\left(\frac{x}{x_0}\right)^{-k+1}\right) + \frac{kx_0}{k^2-1}.$$

This formula can be used to obtain two important pieces of information: the time and the point of interception. The interception occurs when $x = 0$. Thus

$$y(0) = \frac{kx_0}{k^2-1} = \frac{vux_0}{v^2-u^2}.$$

Since $x_0 < 0$ and the point of interception must by on the upper semi-axis, we see that for the interception to occur, the speed of the target v must be smaller that the speed of the pursuer u. This is of course clear from the model, as the pursuer moves along a curve and has a longer distance to cover.

The duration of the pursuit can be calculated by noting that the target moves with a constant speed v along the y axis from the origin to the interception point $(0, y(0))$ so that

$$T = \frac{y(0)}{v} = \frac{ux_0}{v^2 - u^2}.$$

Escape velocity

The equation of motion of an object of mass m projected upward from the surface of a planet was derived at the end of Subsection 1.3.3. The related Cauchy problem reads

$$\begin{aligned} m\frac{d^2y}{dt^2} &= -\frac{mgR^2}{(y+R)^2} - c(y)\left(\frac{dy}{dt}\right)^2 \\ y(0) &= R, \qquad y'(0) = v_0, \end{aligned}$$

where the initial conditions tell us that the missile was shot from the surface with initial velocity v_0 and we allow the air resistance coefficient to change with height. Rather than solve the full Cauchy problem, we shall address the question of the existence of the *escape velocity*, that is, whether there exists an initial velocity which would allow the object to escape from planet's gravitational field.

The equation is of the form (2.3.28), that is, it does not contain explicitly the independent variable. To simplify calculations, firstly we shall change the unknown function according to $z = y + R$ (so that z is the distance from the centre of the planet) and next introduce $F(z) = z'$ so that $z'' = F_z F$, see (2.3.29). Then the equation of motion will take the form

$$F_z F + C(z)F^2 = -\frac{gR^2}{z^2}, \tag{2.4.16}$$

where $C(z) = c(z - R)/m$. Noting that

$$F_z F = \frac{1}{2}\frac{d}{dz}F^2$$

and denoting $F^2 = G$ we reduce (2.4.16) to the linear differential equation

$$G_z + 2C(z)G = -\frac{2gR^2}{z^2}. \tag{2.4.17}$$

We shall consider three forms for C.

Case 1. $C(z) \equiv 0$ (airless moon).

In this case (2.4.17) becomes

$$G_z = -\frac{2gR^2}{z^2}.$$

which can be immediately integrated from R to z giving

$$G(z) - G(R) = 2gR^2\left(\frac{1}{z} - \frac{1}{R}\right).$$

Returning to the old variables $G(z) = F^2(z) = v^2(z)$, where v is the velocity of the missile at the distance z from the centre of the moon, we can write

$$v^2(z) - v^2(R) = 2gR^2\left(\frac{1}{z} - \frac{1}{R}\right).$$

The missile will escape from the moon if it's speed remains positive for all times – if it stops at any finite z, then the gravity pull will bring it back to the moon. Since $v(z)$ is decreasing, its minimum value will be the limit at infinity so that, passing with $z \to \infty$, we must have

$$v^2(R) \geq 2gR$$

and the escape velocity is

$$v(R) = \sqrt{2gR}.$$

Case 2. Constant air resistance.

If we are back on Earth, it is not reasonable to assume that there is no air resistance during motion. Let us investigate the next simple case with $c = \mathit{constant}$. Then we have

$$G_z + 2CG = -\frac{2gR^2}{z^2}, \tag{2.4.18}$$

where $C = c/m$. The integrating factor equals e^{2cz} so that we obtain

$$\frac{d}{dz}\left(e^{2cz}G(z)\right) = -2gR^2\frac{e^{2Cz}}{z^2},$$

and, upon integration,

$$e^{2Cz}v^2(z) - e^{2CR}v_0^2 = -2gR^2\int\limits_R^z e^{2Cs}s^{-2}ds,$$

or

$$v^2(z) = e^{-2Cz}\left(e^{2CR}v_0^2 - 2gR^2\int\limits_R^z e^{2Cs}s^{-2}ds\right). \tag{2.4.19}$$

Consider the integral

$$I(z) = \int\limits_R^z e^{2Cs}s^{-2}ds.$$

Since $\lim\limits_{s\to\infty} e^{2Cs}s^{-2} = \infty$, we have also

$$\lim_{s\to\infty} \int\limits_R^z e^{2Cs}s^{-2} = \infty.$$

Since $\int\limits_R^R e^{2Cs}s^{-2}ds = 0$ and because $e^{2CR}v^2(R)$ is independent of z, from the Darboux theorem we see that, no matter what the value of v_0 is, for some $z_0 \in [R, \infty)$ the right-hand side of (2.4.19) becomes 0 and thus $v^2(z_0) = 0$. Thus, there is no initial velocity v_0 for which the missile will escape the planet.

Case 3. Variable air resistance.

By passing from no air resistance at all ($c = 0$) to a constant air resistance we definitely overshot since the air becomes thinner with height and thus its resistance decreases. Let us consider one more case with $C(z) = k/z$ where k is a proportionality constant. Then we obtain

$$G_z + \frac{2k}{z}G = -\frac{2gR^2}{z^2}. \tag{2.4.20}$$

The integrating factor equals z^{2k} so that we obtain

$$\frac{d}{dz}\left(z^{2k}G(z)\right) = -2gR^2z^{2k-2},$$

and, upon integration,

$$z^{2k}v^2(z) - R^{2k}v_0^2 = -2gR^2 \int\limits_R^z s^{2k-2}ds.$$

Using the same argument, we see that the escape velocity will exist if and only if

$$\lim_{z\to\infty} \int\limits_R^z s^{2k-2}ds < +\infty$$

and from the properties of improper integral we infer that we must have $2k - 2 < -1$ or

$$k < \frac{1}{2}.$$

Of course, from physics $k \geq 0$. Thus, the escape velocity is given by

$$v_0 = \sqrt{\frac{2gR}{1-2k}}.$$

Chapter 3

Simultaneous systems of equations and higher order equations

3.1 Systems of equations

3.1.1 Why systems?

Two possible generalizations of the first order scalar equation

$$y' = f(t, y)$$

are: a differential equation of a higher order

$$y^{(n)} = F(t, y', y'', \ldots, y^{(n-1)}) = 0, \tag{3.1.1}$$

(where, for simplicity, we consider only equations solved with respect to the highest derivative), or a system of first order equations, that is,

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}) \tag{3.1.2}$$

where,

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix},$$

and

$$\mathbf{f}(t, \mathbf{y}) = \begin{pmatrix} f_1(t, y_1, \ldots, y_n) \\ \vdots \\ f_n(t, y_1, \ldots, y_n) \end{pmatrix},$$

is a nonlinear function of t and $\mathbf{y}$. It turns out that, at least from the theoretical point of view, there is no need to consider these two cases separately as any equation of a higher order can be always written as a system (the converse, in general, is not true). To see how this can be accomplished, we introduce new unknown variables $z_1(t) = y(t), z_2(t) = y'(t), z_n = y^{(n-1)}(t)$ so that $z_1'(t) = y'(t) = z_2(t), z_2'(t) = y''(t) = z_3(t), \ldots$ and (3.1.1) converts into

$$\begin{array}{rcl} z_1' &=& z_2, \\ z_2' &=& z_3, \\ \vdots & \vdots & \vdots \\ z_n' &=& F(t, z_1, \ldots, z_n) \end{array}$$

Clearly, solving this system, we obtain simultaneously the solution of (3.1.1) by taking $y(t) = z_1(t)$.

3.1.2 Linear systems

At the beginning we shall consider only systems of first order differential equations that are solved with respect to the derivatives of all unknown functions. The systems we deal with in this section are linear, that is, they can be written as

$$\begin{array}{rcl} y_1' &=& a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n + g_1(t), \\ \vdots & \vdots & \vdots, \\ y_n' &=& a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n + g_n(t), \end{array} \tag{3.1.3}$$

where $y_1, \ldots, y_n$ are unknown functions, $a_{11}, \ldots a_{nn}$ are constant coefficients and $g_1(t) \ldots, g_n(t)$ are known continuous functions. If $g_1 = \ldots = g_n = 0$, then the corresponding system (3.1.3) is called the associated homogeneous system. The structure of (3.1.3) suggest that a more economical way of writing it is to use the vector-matrix notation. Denoting $\mathbf{y} = (y_1, \ldots, y_n)$, $\mathbf{g} = (g_1, \ldots, g_n)$ and $\mathcal{A} = \{a_{ij}\}_{1\le i,j\le n}$, that is

$$\mathcal{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix},$$

we can write (3.1.3) in a more concise way as

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}. \tag{3.1.4}$$

Here we have n unknown functions and the system involves first derivative of each of them so that it is natural to consider (3.1.4) in conjunction with the following initial conditions

$$\mathbf{y}(t_0) = \mathbf{y^0}, \tag{3.1.5}$$

or, in the expanded form,

$$y_1(t_0) = y_1^0, \ldots, y_n(t_0) = y_n^0, \tag{3.1.6}$$

where t_0 is a given argument and $\mathbf{y^0} = (y_1^0, \ldots, y_n^0)$ is a given vector.

As we noted in the introduction, systems of first order equations are closely related to higher order equations. In particular, any nth order linear equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1 y' + a_0 y = g(t) \tag{3.1.7}$$

can be written as a linear system of n first order equations by introducing new variables $z_1 = y$, $z_2 = y' = z_1'$, $z_3 = y'' = z_2', \ldots z_n = y^{(n-1)} = z_{n-1}'$ so that $z_n' = y^{(n)}$ and (3.1.7) turns into

$$\begin{array}{rcl} z_1' &=& z_2, \\ z_2' &=& z_3, \\ \vdots & & \vdots \\ z_n' &=& -a_{n-1}z_n - a_{n-2}z_{n-1} - \ldots - a_0 z_1 + g(t). \end{array}$$

Note that if (3.1.7) was supplemented with the initial conditions $y(t_0) = y_0, y'(t_0) = y_1, \ldots y^{(n-1)} = y_{n-1}$, then these conditions will become natural initial conditions for the system as $z_1(t_0) = y_0, z_2(t_0) = y_1, \ldots z_n(t_0) = y_{n-1}$. Therefore, all the results we shall prove here are relevant also for nth order equations.

In some cases, especially when faced with simple systems of differential equations, it pays to revert the procedure and to transform a system into a single, higher order, equation rather than to apply directly a heavy procedure for full systems. We illustrate this remark in the following example.

Example 3.1.1. Consider the system

$$\begin{array}{rcl} y_1' &=& a_{11}y_1 + a_{12}y_2, \\ y_2' &=& a_{21}y_1 + a_{22}y_2. \end{array} \tag{3.1.8}$$

Firstly, note that if either a_{12} or a_{21} equal zero, then the equations are uncoupled, e.g., if $a_{12} = 0$, then the first equation does not contain y_2 and can be solved for y_1 and this solution can be inserted into the second equation which then becomes a first order nonhomogeneous equation for y_2.

Assume then that $a_{12} \neq 0$. We proceed by eliminating y_2 from the first equation. Differentiating it, we obtain

$$y_1'' = a_{11}y_1' + a_{12}y_2',$$

so that, using the second equation,

$$y_1'' = a_{11}y_1' + a_{12}(a_{21}y_1 + a_{22}y_2).$$

To get rid of the remaining y_2, we use the first equation once again obtaining

$$y_2 = a_{12}^{-1}(y_1' - a_{11}y_1), \tag{3.1.9}$$

$$y_1'' = (a_{11} + a_{22})y_1' + (a_{12}a_{21} - a_{22}a_{11})y_1$$

which is a second order linear equation. If we are able to solve it to obtain y_1, we use again (3.1.9) to obtain y_2.

However, for larger systems this procedure becomes quite cumbersome unless the matrix $\mathcal{A}$ of coefficients has a simple structure.

3.1.3 Algebraic properties of systems

In this subsection we shall prove several results related to the algebraic structure of the set of solutions to

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(t_0) = \mathbf{y^0}. \tag{3.1.10}$$

An extremely important rôle here is played by the uniqueness of solutions. In Section 2.2 we discussed Picard' theorem, Theorem 2.2.4, that dealt with the existence and uniqueness of solution to the Cauchy problem

$$y' = f(t, y), \qquad y(t_0) = y_0$$

where y and f were scalar valued functions. It turns out that this theorem can be easily generalized to the vector case, that is to the case where f is a vector valued function $\mathbf{f}(t, \mathbf{y})$ of a vector valued argument $\mathbf{y}$. In particular, it can be applied to the case when $\mathbf{f}(t, \mathbf{y}) = \mathcal{A}\mathbf{y} + \mathbf{g}(t)$. Thus, we can state

Theorem 3.1.2. *Let $\mathbf{g}(t)$ be a continuous function from $\mathbb{R}$ to $\mathbb{R}^n$. Then there exists one and only one solution of the initial value problem*

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}(t), \qquad \mathbf{y}(t_0) = \mathbf{y}^0. \tag{3.1.11}$$

Moreover, this solution exists for all $t \in \mathbb{R}$.

One of the important implications of this theorem is that if $\mathbf{y}$ is a non-trivial, that is, not identically equal to zero, solution to the homogeneous equation

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \tag{3.1.12}$$

then $\mathbf{y}(t) \neq 0$ for any t. In fact, as $\mathbf{y}^* \equiv 0$ is a solution to (3.1.12) and by definition $\mathbf{y}^*(\bar{t}) = 0$ for any $\bar{t}$, the existence of other solution satisfying $\mathbf{y}(\bar{t}) = 0$ for some $\bar{t}$ would violate Theorem 3.1.2.

Let us denote by $\mathbf{X}$ the set of all solutions to (3.1.12). Due to linearity of differentiation and multiplication by $\mathcal{A}$, it is easy to see that $\mathbf{X}$ is a vector space. Moreover

Theorem 3.1.3. *The dimension of* $\mathbf{X}$ *is equal to* n*.*

Proof. We must exhibit a basis of $\mathbf{X}$ that contains exactly n elements. Thus, let $\mathbf{z_j}(t)$, $j = 1, \ldots, n$ be solutions of special Cauchy problems

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{e_j}, \tag{3.1.13}$$

where $\mathbf{e_j} = (0, 0, \ldots, 1, \ldots, 0)$ with 1 at jth place is a versor of the coordinate system. To determine whether the set $\{\mathbf{z_1}, \ldots, \mathbf{z_n}\}$ is linearly dependent, we ask whether from

$$c_1\mathbf{z_1}(t) + \ldots + c_n\mathbf{z_n}(t) = 0,$$

it follows that $c_1 = \ldots = c_n = 0$. If the linear combination vanishes for any t, then it must vanish in particular for $t = 0$. Thus, using the initial conditions $\mathbf{z_j}(0) = \mathbf{e_j}$ we see that we would have

$$c_1\mathbf{e_1} + \ldots + c_n\mathbf{e_2} = 0,$$

but since the set $\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ is a basis in $\mathbb{R}^n$, we see that necessarily $c_1 = \ldots = c_n = 0$. Thus $\{\mathbf{z_1}(t), \ldots, \mathbf{z_n}(t)\}$ is linearly independent and $\dim \mathbf{X} \geq n$. To show that $\dim \mathbf{X} = n$ we must show that $\mathbf{X}$ is spanned by $\{\mathbf{z_1}(t), \ldots, \mathbf{z_n}(t)\}$, that is, that any solution $\mathbf{y}(t)$ can be written as

$$\mathbf{y}(t) = c_1\mathbf{z_1}(t) + \ldots + c_n\mathbf{z_n}(t)$$

for some constants $c_1, \ldots, c_n$. Let $\mathbf{y}(t)$ be any solution to (3.1.12) and define $\mathbf{y}^0 = \mathbf{y}(0) \in \mathbb{R}^n$. Since $\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ is a basis $\mathbb{R}^n$, there are constants $c_1, \ldots, c_n$ such that

$$\mathbf{y}^0 = c_1\mathbf{e_1} + \ldots + c_n\mathbf{e_2}.$$

Consider

$$\mathbf{x}(t) = c_1\mathbf{z_1}(t) + \ldots + c_n\mathbf{z_n}(t).$$

Clearly, $\mathbf{x}(t)$ is a solution to (3.1.12), as a linear combination of solutions, and $\mathbf{x}(0) = c_1\mathbf{e_1} + \ldots + c_n\mathbf{e_n} = \mathbf{y}^0 = \mathbf{y}(0)$. Thus, $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are both solutions to (3.1.12) satisfying the same initial condition and therefore $\mathbf{x}(t) = \mathbf{y}(t)$ by Theorem 3.1.2. Hence,

$$\mathbf{y}(t) = c_1\mathbf{z_1}(t) + \ldots + c_n\mathbf{z_n}(t).$$

and the set $\{\mathbf{z_1}(t), \ldots, \mathbf{z_n}(t)\}$ is a basis for $\mathbf{X}$. ■

Next we present a convenient way of determining whether solutions to (3.1.12) are linearly independent.

Theorem 3.1.4. *Let $\mathbf{y_1}, \ldots, \mathbf{y_k}$ be k linearly independent solutions of $\mathbf{y}' = \mathcal{A}\mathbf{y}$ and let $t_0 \in \mathbb{R}$ be an arbitrary number. Then, $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ form a linearly independent set of functions if and only if $\{\mathbf{y_1}(t_0), \ldots, \mathbf{y_k}(t_0)\}$ is a linearly independent set of vectors in $\mathbb{R}$.*

Proof. If $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ are linearly dependent functions, then there exist constants $c_1, \ldots, c_k$, not all zero, such that for all t

$$c_1\mathbf{y_1}(t) + \ldots + c_n\mathbf{y_k}(t) = 0.$$

Taking this at a particular value of t, $t = t_0$, we obtain that

$$c_1\mathbf{y_1}(t_0) + \ldots + c_n\mathbf{y_k}(t_0) = 0,$$

with not all c_i vanishing. Thus the set $\{\mathbf{y_1}(t_0), \ldots, \mathbf{y_k}(t_0)\}$ is a set of linearly dependent vectors in $\mathbb{R}^n$.

Conversely, suppose that $\{\mathbf{y_1}(t_0), \ldots, \mathbf{y_k}(t_0)\}$ is a linearly dependent set of vectors. Then for some constants

$$c_1\mathbf{y_1}(t_0) + \ldots + c_n\mathbf{y_k}(t_0) = 0,$$

where not all c_i are equal to zero. Taking these constants we construct the function

$$\mathbf{y}(t) = c_1\mathbf{y_1}(t) + \ldots + c_n\mathbf{y_k}(t),$$

which is a solution to (3.1.12) as a linear combination of solutions. However, since $\mathbf{y}(t_0) = 0$, by the uniqueness theorem we obtain that $\mathbf{y}(t) = 0$ for all t so that $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ is a linearly dependent set of functions. ■

Remark 3.1.5. To check whether a set of n vectors of $\mathbb{R}^n$ is linearly independent, we can use the determinant test: $\{\mathbf{y_1}, \ldots, \mathbf{y_k}\}$ is linearly independent if and only if

$$det\{\mathbf{y_1}, \ldots, \mathbf{y_k}\} = \begin{vmatrix} y_1^1 & \ldots & y_1^n \\ \vdots & & \vdots \\ y_n^1 & \ldots & y_n^n \end{vmatrix} \neq 0.$$

If $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ is a set of solution of the homogeneous system, then the determinant

$$det\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\} = \left| \begin{array}{ccc} y_1^1(t) & \ldots & y_1^n(t) \\ \vdots & & \vdots \\ y_n^1(t) & \ldots & y_n^n(t) \end{array} \right|$$

is called *wronskian.* The theorems proved above can be rephrased by saying that the wronskian is non-zero if it is constructed with independent solutions of a system of equations and, in such a case, it is non-zero if and only if it is non-zero at some point.

Example 3.1.6. Consider the system of differential equations

$$\begin{array}{rcl} y_1' & = & y_2, \\ y_2' & = & -y_1 - 2y_2, \end{array} \qquad (3.1.14)$$

or, in matrix notation

$$\mathbf{y}' = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix} \mathbf{y}.$$

Let us take two solutions:

$$\mathbf{y^1}(t) = (y_1^1(t), y_2^1(t)) = (\phi(t), \phi'(t)) = (e^{-t}, -e^{-t}) = e^{-t}(1, -1)$$

and

$$\mathbf{y^2}(t) = (y_1^2(t), y_2^2(t)) = (\psi(t), \psi'(t)) = (te^{-t}, (1-t)e^{-t}) = e^{-t}(1, 1-t).$$

To check whether these are linearly in dependent solutions to the system and thus whether they span the space of all solutions, we use Theorem 3.1.4 and check the linear dependence of vectors $\mathbf{y^1}(0) = (1, -1)$ and $\mathbf{y^2}(0) = (0, 1)$. Using e.g. the determinant test for linear dependence we evaluate

$$\left| \begin{array}{cc} 1 & -1 \\ 0 & 1 \end{array} \right| = 1 \neq 0,$$

thus the vectors are linearly independent. Consequently, all solutions to (3.1.14) can be written in the form

$$\mathbf{y}(t) = C_1 \begin{pmatrix} e^{-t} \\ -e^{-t} \end{pmatrix} + C_2 \begin{pmatrix} te^{-t} \\ (1-t)e^{-t} \end{pmatrix} = \begin{pmatrix} (C_1 + C_2 t)e^{-t} \\ (C_2 - C_1 - C_2 t)e^{-t} \end{pmatrix}.$$

Assume now that we are given this $\mathbf{y}(t)$ as a solution to the system. The system is equivalent to the second order equation

$$y'' + 2y' + y = 0 \qquad (3.1.15)$$

under identification $y(t) = y_1(t)$ and $y'(t) = y_2(t)$. How can we recover the general solution to (3.1.15) from $\mathbf{y}(t)$? Remembering that y solves (3.1.15) if and only if $\mathbf{y}(t) = (y_1(t), y_2(t)) = (y(t), y'(t))$ solves the system (3.1.14), we see that the general solution to (3.1.15) can be obtained by taking first components of the solution of the associated system (3.1.14). We also note the fact that if $\mathbf{y^1}(t) = (y_1^1(t), y_2^1(t)) = (y^1(t), \frac{dy^1}{dt}(t))$ and $\mathbf{y^1}(t) = (y_1^2(t), y_2^2(t)) = (y^2(t), \frac{dy^2}{dt}(t))$ are two linearly independent solutions to (3.1.14), then $y^1(t)$ and $y^2(t)$ are linearly independent solutions to (3.1.15). In fact, otherwise we would have $y^1(t) = Cy^2(t)$ for some constant C and therefore also $\frac{dy^1}{dt}(t) = C\frac{dy^2}{dt}(t)$ so that the wronskian, having the second column as a scalar multiple of the first one, would be zero, contrary to the assumption that $\mathbf{y^1}(t)$ and $\mathbf{y^2}(t)$ are linearly independent.

3.1.4 The eigenvalue-eigenvector method of finding solutions

We start with a brief survey of eigenvalues and eigenvectors of matrices. Let $\mathcal{A}$ be an $n \times n$ matrix. We say that a number λ (real or complex) is an *eigenvalue* of $\mathcal{A}$ is there exist a non-zero solution of the equation

$$\mathcal{A}\mathbf{v} = \lambda\mathbf{v}. \tag{3.1.16}$$

Such a solution is called an *eigenvector* of $\mathcal{A}$. The set of eigenvectors corresponding to a given eigenvalue is a vector subspace. Eq. (3.1.16) is equivalent to the homogeneous system $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$, where $\mathcal{I}$ is the identity matrix, therefore λ is an eigenvalue of $\mathcal{A}$ if and only if the determinant of $\mathcal{A}$ satisfies

$$det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} a_{11} - \lambda & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{3.1.17}$$

Evaluating the determinant we obtain a polynomial in λ of degree n. This polynomial is also called the characteristic polynomial of the system (3.1.3) (if (3.1.3) arises from a second order equation, then this is the same polynomial as the characteristic polynomial of the equation). We shall denote this polynomial by $p(\lambda)$. From algebra we know that there are exactly n, possibly complex, roots of $p(\lambda)$. Some of them may be multiple, so that in general $p(\lambda)$ factorizes into

$$p(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdot \ldots \cdot (\lambda_k - \lambda)^{n_k}, \tag{3.1.18}$$

with $n_1 + \ldots + n_k = n$. It is also worthwhile to note that since the coefficients of the polynomial are real, then complex roots appear always in conjugate pairs, that is, if $\lambda_j = \xi_j + i\omega_j$ is a characteristic root, then so is $\bar{\lambda}_j =$

$\xi_j - i\omega_j$. Thus, eigenvalues are roots of the characteristic polynomial of $\mathcal{A}$. The exponent n_i appearing in the factorization (3.1.18) is called the *algebraic multiplicity* of λ_i. For each eigenvalue λ_i there corresponds an eigenvector $\mathbf{v_i}$ and eigenvectors corresponding to distinct eigenvalues are linearly independent. The set of all eigenvectors corresponding to λ_i spans a subspace, called the *eigenspace* corresponding to λ_i which we will denote by E_{λ_i}. The dimension of E_{λ_i} is called the *geometric multiplicity* of λ_i. In general, algebraic and geometric multiplicities are different with geometric multiplicity being at most equal to the algebraic one. Thus, in particular, if λ_i is a single root of the characteristic polynomial, then the eigenspace corresponding to λ_1 is one-dimensional.

If the geometric multiplicities of eigenvalues add up to n, that is, if we have n linearly independent eigenvectors, then these eigenvectors form a basis for $\mathbb{R}^n$. In particular, this happens if all eigenvalues are single roots of the characteristic polynomial. If this is not the case, then we do not have sufficiently many eigenvectors to span $\mathbb{R}^n$ and if we need a basis for $\mathbb{R}^n$, then we have to find additional linearly independent vectors. A procedure that can be employed here and that will be very useful in our treatment of systems of differential equations is to find solutions to equations of the form $(\mathcal{A} - \lambda_i\mathcal{I})^k\mathbf{v} = 0$ for $1 < k \leq n_i$, where n_i is the algebraic multiplicity of λ_i. Precisely speaking, if λ_i has algebraic multiplicity n_i and if

$$(\mathcal{A} - \lambda_i\mathcal{I})\mathbf{v} = 0$$

has only $\nu_i < n_i$ linearly independent solutions, then we consider the equation

$$(\mathcal{A} - \lambda_i\mathcal{I})^2\mathbf{v} = 0.$$

It follows that all the solutions of the preceding equation solve this equation but there is at least one more independent solution so that we have at least $\nu_i + 1$ independent vectors (note that these new vectors are no longer eigenvectors). If the number of independent solutions is still less than n_i, we consider

$$(\mathcal{A} - \lambda_i\mathcal{I})^3\mathbf{v} = 0,$$

and so on, till we get a sufficient number of them. Note, that to make sure that in the step j we select solutions that are independent of the solutions obtained in step $j-1$ it is enough to find solutions to $(\mathcal{A} - \lambda_i\mathcal{I})^j\mathbf{v} = 0$ that satisfy $(\mathcal{A} - \lambda_i\mathcal{I})^{j-1}\mathbf{v} \neq 0$.

Now we show how to apply the concepts discussed above to solve systems of differential equations. Consider again the homogeneous system

$$\mathbf{y}' = \mathcal{A}\mathbf{y}. \tag{3.1.19}$$

Our goal is to find n linearly independent solutions of (3.1.19). We have seen that solutions of the form $e^{\lambda t}$ play a basic rôle in solving first order linear equations so let us consider $\mathbf{y}(t) = e^{\lambda t}\mathbf{v}$ for some vector $\mathbf{v} \in \mathbb{R}^n$. Since

$$\frac{d}{dt}e^{\lambda t}\mathbf{v} = \lambda e^{\lambda t}\mathbf{v}$$

and

$$\mathcal{A}(e^{\lambda t}\mathbf{v}) = e^{\lambda t}\mathcal{A}\mathbf{v}$$

as $e^{\lambda t}$ is a scalar, $\mathbf{y}(t) = e^{\lambda t}\mathbf{v}$ is a solution to (3.1.19) if and only if

$$\mathcal{A}\mathbf{v} = \lambda\mathbf{v}. \tag{3.1.20}$$

Thus $\mathbf{y}(t) = e^{\lambda t}\mathbf{v}$ is a solution if and only if $\mathbf{v}$ is an eigenvector of $\mathcal{A}$ corresponding to the eigenvalue λ.

Thus, for each eigenvector $\mathbf{v^j}$ of $\mathcal{A}$ with eigenvalue λ_j we have a solution $\mathbf{y^j}(t) = e^{\lambda_j t}\mathbf{v^j}$. By Theorem 3.1.4 these solutions are linearly independent if and only if the eigenvectors $\mathbf{v^j}$ are linearly independent in $\mathbb{R}^n$. Thus, if we can find n linearly independent eigenvectors of $\mathcal{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct), then the general solution of (3.1.19) is of the form

$$\mathbf{y}(t) = C_1 e^{\lambda_1 t}\mathbf{v^1} + \ldots + C_n e^{\lambda_n t}\mathbf{v^n}. \tag{3.1.21}$$

Distinct real eigenvalues

The simplest situation is of course if the characteristic polynomial $p(\lambda)$ has n distinct roots, that is, all roots are single and in this case the eigenvectors corresponding to different eigenvalues (roots) are linearly independent, as we mentioned earlier. However, this can also happen if some eigenvalues are multiple ones but the algebraic and geometric multiplicity of each is the same. In this case to each root of multiplicity n_1 there correspond n_1 linearly independent eigenvectors.

Example 3.1.7. Find the general solution to

$$\mathbf{y}' = \begin{pmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{pmatrix}\mathbf{y}.$$

To obtain the eigenvalues we calculate the characteristic polynomial

$$\begin{aligned} p(\lambda) &= det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} 1-\lambda & -1 & 4 \\ 3 & 2-\lambda & -1 \\ 2 & 1 & -1-\lambda \end{vmatrix} \\ &= -(1+\lambda)(1-\lambda)(2-\lambda) + 12 + 2 - 8(2-\lambda) + (1-\lambda) - 3(1+\lambda) \\ &= -(1+\lambda)(1-\lambda)(2-\lambda) + 4\lambda - 4 = (1-\lambda)(\lambda-3)(\lambda+2), \end{aligned}$$

so that the eigenvalues of $\mathcal{A}$ are $\lambda_1 = 1$, $\lambda_2 = 3$ and $\lambda_3 = -2$. All the eigenvalues have algebraic multiplicity 1 so that they should give rise to 3 linearly independent eigenvectors.

(i) $\lambda_1 = 1$: we seek a nonzero vector $\mathbf{v}$ such that

$$(\mathcal{A} - 1\mathcal{I})\mathbf{v} = \begin{pmatrix} 0 & -1 & 4 \\ 3 & 1 & -1 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus

$$-v_2 + 4v_3 = 0, \qquad 3v_1 + v_2 - v_3 = 0, \qquad 2v_1 + v_2 - 2v_3 = 0$$

and we get $v_2 = 4v_3$ and $v_1 = -v_3$ from the first two equations and the third is automatically satisfied. Thus we obtain the eigenspace corresponding to $\lambda_1 = 1$ containing all the vectors of the form

$$\mathbf{v^1} = C_1 \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$$

where C_1 is any constant, and the corresponding solutions

$$\mathbf{y^1}(t) = C_1 e^t \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}.$$

(ii) $\lambda_2 = 3$: we seek a nonzero vector $\mathbf{v}$ such that

$$(\mathcal{A} - 3\mathcal{I})\mathbf{v} = \begin{pmatrix} -2 & -1 & 4 \\ 3 & -1 & -1 \\ 2 & 1 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hence

$$-2v_1 - v_2 + 4v_3 = 0, \qquad 3v_1 - v_2 - v_3 = 0, \qquad 2v_1 + v_2 - 4v_3 = 0.$$

Solving for v_1 and v_2 in terms of v_3 from the first two equations gives $v_1 = v_3$ and $v_2 = 2v_3$. Consequently, vectors of the form

$$\mathbf{v^2} = C_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

are eigenvectors corresponding to the eigenvalue $\lambda_2 = 3$ and the function

$$\mathbf{y^2}(t) = e^{3t} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

is the second solution of the system.

(iii) $\lambda_3 = -2$: We have to solve

$$(\mathcal{A}+2\mathcal{I})\mathbf{v} = \begin{pmatrix} 3 & -1 & 4 \\ 3 & 4 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus

$$3v_1 - v_2 + 4v_3 = 0, \qquad 3v_1 + 4v_2 - v_3 = 0, \qquad 2v_1 + v_2 + v_3 = 0.$$

Again, solving for v_1 and v_2 in terms of v_3 from the first two equations gives $v_1 = -v_3$ and $v_2 = v_3$ so that each vector

$$\mathbf{v^3} = C_3 \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to the eigenvalue $\lambda_3 = -2$. Consequently, the function

$$\mathbf{y^3}(t) = e^{-2t} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

is the third solution of the system. These solutions are linearly independent since the vectors $\mathbf{v^1}, \mathbf{v^2}, \mathbf{v^3}$ are linearly independent as eigenvectors corresponding to distinct eigenvalues. Therefore, every solution is of the form

$$\mathbf{y}(t) = C_1 e^t \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix} + C_2 e^{3t} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + C_3 e^{-2t} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}.$$

Distinct complex eigenvalues

If $\lambda = \xi + i\omega$ is a complex eigenvalue, then also its complex conjugate $\bar{\lambda} = \xi - i\omega$ is an eigenvalue, as the characteristic polynomial $p(\lambda)$ has real coefficients. Eigenvectors $\mathbf{v}$ corresponding to a complex complex eigenvalue λ will be complex vectors, that is, vectors with complex entries. Thus, we can write

$$\mathbf{v} = \begin{pmatrix} v_1^1 + iv_1^2 \\ \vdots \\ v_n^1 + iv_n^2 \end{pmatrix} = \begin{pmatrix} v_1^1 \\ \vdots \\ v_n^1 \end{pmatrix} + i \begin{pmatrix} v_1^2 \\ \vdots \\ v_n^2 \end{pmatrix} = \Re\mathbf{v} + i\Im\mathbf{v}.$$

Since $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$, taking complex conjugate of both sides and using the fact that matrices $\mathcal{A}$ and $\mathcal{I}$ have only real entries, we see that

$$\overline{(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}} = (\mathcal{A} - \bar{\lambda}\mathcal{I})\bar{\mathbf{v}} = \mathbf{0}$$

so that the complex conjugate $\bar{\mathbf{v}}$ of $\mathbf{v}$ is an eigenvector corresponding to the eigenvalue $\bar{\lambda}$. Since $\lambda \neq \bar{\lambda}$, as we assumed that λ is complex, the eigenvectors $\mathbf{v}$ and $\bar{\mathbf{v}}$ are linearly independent and thus we obtain two linearly independent complex valued solutions

$$\mathbf{z^1}(t) = e^{\lambda t}\mathbf{v}, \qquad \mathbf{z^2}(t) = e^{\bar{\lambda} t}\bar{\mathbf{v}} = \overline{\mathbf{z^1}}(t).$$

Since the sum and the difference of two solutions are again solutions, by taking

$$\mathbf{y^1}(t) = \frac{\mathbf{z^1}(t) + \mathbf{z^2}(t)}{2} = \frac{\mathbf{z^1}(t) + \overline{\mathbf{z^1}}(t)}{2} = \Re \mathbf{z^1}(t)$$

and

$$\mathbf{y^2}(t) = \frac{\mathbf{z^1}(t) - \mathbf{z^2}(t)}{2i} = \frac{\mathbf{z^1}(t) - \overline{\mathbf{z^1}}(t)}{2i} = \Im \mathbf{z^1}(t)$$

we obtain two real valued (and linearly independent) solutions. To find explicit formulae for $\mathbf{y^1}(t)$ and $\mathbf{y^2}(t)$, we write

$$\begin{aligned}
\mathbf{z^1}(t) &= e^{\lambda t}\mathbf{v} = e^{\xi t}(\cos\omega t + i\sin\omega t)(\Re\mathbf{v} + i\Im\mathbf{v}) \\
&= e^{\xi t}(\cos\omega t\,\Re\mathbf{v} - \sin\omega t\,\Im\mathbf{v}) + ie^{\xi t}(\cos\omega t\,\Im\mathbf{v} + \sin\omega t\,\Re\mathbf{v}) \\
&= \mathbf{y^1}(t) + i\mathbf{y^2}(t)
\end{aligned}$$

Summarizing, if λ and $\bar{\lambda}$ are single complex roots of the characteristic equation with complex eigenvectors $\mathbf{v}$ and $\bar{\mathbf{v}}$, respectively, then the we can use two real linearly independent solutions

$$\begin{aligned}
\mathbf{y^1}(t) &= e^{\xi t}(\cos\omega t\,\Re\mathbf{v} - \sin\omega t\,\Im\mathbf{v}) \\
\mathbf{y^2}(t) &= e^{\xi t}(\cos\omega t\,\Im\mathbf{v} + \sin\omega t\,\Re\mathbf{v})
\end{aligned} \tag{3.1.22}$$

Example 3.1.8. Solve the initial value problem

$$\mathbf{y}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}\mathbf{y}, \qquad \mathbf{y}(0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

The characteristic polynomial is given by

$$\begin{aligned}
p(\lambda) &= det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & -1 \\ 0 & 1 & 1-\lambda \end{vmatrix} \\
&= (1-\lambda)^3 + (1-\lambda) = (1-\lambda)(\lambda^2 - 2\lambda + 2)
\end{aligned}$$

so that we have eigenvalues $\lambda_1 = 1$ and $\lambda_{2,3} = 1 \pm i$.

It is immediate that

$$\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = 1$ and thus we obtain a solution to the system in the form

$$\mathbf{y^1}(t) = e^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Let us take now the complex eigenvalue $\lambda_2 = 1 + i$. We have to solve

$$(\mathcal{A} - (1+i)\mathcal{I})\mathbf{v} = \begin{pmatrix} -i & 0 & 0 \\ 0 & -i & -1 \\ 0 & 1 & -i \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus

$$-iv_1 = 0, \qquad -iv_2 - v_3 = 0, \qquad v_2 - iv_3 = 0.$$

The first equation gives $v_1 = 0$ and the other two yield $v_2 = iv_3$ so that each vector

$$\mathbf{v^2} = C_2 \begin{pmatrix} 0 \\ i \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to the eigenvalue $\lambda_2 = 1 + i$. Consequently, we obtain a complex valued solution

$$\mathbf{z}(t) = e^{(1+i)t} \begin{pmatrix} 0 \\ i \\ 1 \end{pmatrix}.$$

To obtain real valued solutions, we separate $\mathbf{z}$ into real and imaginary parts:

$$\begin{aligned} e^{(1+i)t} \begin{pmatrix} 0 \\ i \\ 1 \end{pmatrix} &= e^t(\cos t + i \sin t) \left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + i \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right) \\ = \quad e^t &\left(\cos t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \sin t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + i \sin t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + i \cos t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right) \\ = \quad e^t &\begin{pmatrix} 0 \\ -\sin t \\ \cos t \end{pmatrix} + ie^t \begin{pmatrix} 0 \\ \cos t \\ \sin t \end{pmatrix}. \end{aligned}$$

Thus, we obtain two real solutions

$$\begin{aligned} \mathbf{y^1}(t) &= e^t \begin{pmatrix} 0 \\ -\sin t \\ \cos t \end{pmatrix} \\ \mathbf{y^2}(t) &= e^t \begin{pmatrix} 0 \\ \cos t \\ \sin t \end{pmatrix} \end{aligned}$$

and the general solution to our original system is given by

$$\mathbf{y}(t) = C_1 e^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + C_2 e^t \begin{pmatrix} 0 \\ -\sin t \\ \cos t \end{pmatrix} + C_3 e^t \begin{pmatrix} 0 \\ \cos t \\ \sin t \end{pmatrix}.$$

We can check that all these solutions are independent as their initial values

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

are independent. To find the solution to our initial value problem we set $t = 0$ and we have to solve for C_1, C_2 and C_3 the system

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = C_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + C_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}.$$

Thus $C_1 = C_2 = C_3 = 1$ and finally

$$\mathbf{y}(t) = e^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + e^t \begin{pmatrix} 0 \\ -\sin t \\ \cos t \end{pmatrix} + e^t \begin{pmatrix} 0 \\ \cos t \\ \sin t \end{pmatrix} = e^t \begin{pmatrix} 1 \\ \cos t - \sin t \\ \cos t + \sin t \end{pmatrix}.$$

Multiple eigenvalues

If not all roots of the characteristic polynomial of $\mathcal{A}$ are distinct, that is, there are multiple eigenvalues of $\mathcal{A}$, then it may happen that $\mathcal{A}$ has less than n linearly independent eigenvectors. Precisely, let us suppose that an $n \times n$ matrix $\mathcal{A}$ has only $k < n$ linearly independent solutions. Then, the differential equation $\mathbf{y}' = \mathcal{A}\mathbf{y}$ has only k linearly independent solutions of the form $e^{\lambda t}\mathbf{v}$. Our aim is to find additional $n - k$ independent solutions. We approach this problem by introducing an abstract framework for solving systems of differential equations.

Recall that for a single equation $y' = ay$, where a is a constant, the general solution is given by $y(t) = e^{at}C$, where C is a constant. In a similar way, we would like to say that the general solution to

$$\mathbf{y}' = \mathcal{A}\mathbf{y},$$

where $\mathcal{A}$ is an $n \times n$ matrix, is $\mathbf{y} = e^{\mathcal{A}t}\mathbf{v}$, where $\mathbf{v}$ is any constant vector in $\mathbb{R}^n$. The problem is that we do not know what it means to evaluate the exponential of a matrix. However, if we reflect for a moment that the exponential of a number can be evaluated as the power (Maclaurin) series

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots + \frac{x^k}{k!} + \ldots,$$

where the only involved operations on the argument x are additions, scalar multiplications and taking integer powers, we come to the conclusion that the above expression can be written also for a matrix, that is, we can define

$$e^{\mathcal{A}} = \mathcal{I} + \mathcal{A} + \frac{1}{2}\mathcal{A}^2 + \frac{1}{3!}\mathcal{A}^3 + \ldots + \frac{1}{k!}\mathcal{A}^k + \ldots. \tag{3.1.23}$$

It can be shown that if $\mathcal{A}$ is a matrix, then the above series always converges and the sum is a matrix. For example, if we take

$$\mathcal{A} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \lambda\mathcal{I},$$

then

$$\mathcal{A}^k = \lambda^k \mathcal{I}^k = \lambda^k \mathcal{I},$$

and

$$\begin{aligned} e^{\mathcal{A}} &= \mathcal{I} + \lambda\mathcal{I} + \frac{\lambda^2}{2}\mathcal{I} + \frac{\lambda^3}{3!}\mathcal{I} + \ldots + \frac{\lambda^k}{k!} + \ldots \\ &= \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \ldots + \frac{\lambda^k}{k!} + \ldots\right)\mathcal{I} \\ &= e^{\lambda}\mathcal{I}. \end{aligned} \tag{3.1.24}$$

Unfortunately, in most cases finding the explicit form for $e^{\mathcal{A}}$ directly is impossible.

Matrix exponentials have the following algebraic properties

$$\left(e^{\mathcal{A}}\right)^{-1} = e^{-\mathcal{A}}$$

and

$$e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{A}}e^{\mathcal{B}} \tag{3.1.25}$$

provided the matrices $\mathcal{A}$ and $\mathcal{B}$ commute: $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$.

Let us define a function of t by

$$e^{t\mathcal{A}} = \mathcal{I} + t\mathcal{A} + \frac{t^2}{2}\mathcal{A}^2 + \frac{t^3}{3!}\mathcal{A}^3 + \ldots + \frac{t^k}{k!}\mathcal{A}^k + \ldots\ . \tag{3.1.26}$$

It follows that this function can be differentiated with respect to t by termwise differentiation of the series, as in the scalar case, that is,

$$\begin{aligned} \frac{d}{dt}e^{\mathcal{A}t} &= \mathcal{A} + t\mathcal{A}^2 + \frac{t^2}{2!}\mathcal{A}^3 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^k + \ldots \\ &= \mathcal{A}\left(\mathcal{I} + t\mathcal{A} + \frac{t^2}{2!}\mathcal{A}^2 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^{k-1} + \ldots\right) \\ &= \mathcal{A}e^{t\mathcal{A}} = e^{t\mathcal{A}}\mathcal{A}, \end{aligned}$$

proving thus that $y(t) = e^{t\mathcal{A}}\mathbf{v}$ is a solution to our system of equations for any constant vector $\mathbf{v}$.

As we mentioned earlier, in general it is difficult to find directly the explicit form of $e^{t\mathcal{A}}$. However, we can always find n linearly independent vectors $\mathbf{v}$ for which the series $e^{t\mathcal{A}}\mathbf{v}$ can be summed exactly. This is based on the following two observations. Firstly, since $\lambda\mathcal{I}$ and $\mathcal{A} - \lambda\mathcal{I}$ commute, we have by (3.1.24) and (3.1.25)

$$e^{t\mathcal{A}}\mathbf{v} = e^{t(\mathcal{A}-\lambda\mathcal{I})}e^{t\lambda\mathcal{I}}\mathbf{v} = e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v}.$$

Secondly, if $(\mathcal{A} - \lambda\mathcal{I})^m\mathbf{v} = \mathbf{0}$ for some m, then

$$(\mathcal{A} - \lambda\mathcal{I})^r\mathbf{v} = \mathbf{0}, \tag{3.1.27}$$

for all $r \geq m$. This follows from

$$(\mathcal{A} - \lambda\mathcal{I})^r\mathbf{v} = (\mathcal{A} - \lambda\mathcal{I})^{r-m}[(\mathcal{A} - \lambda\mathcal{I})^m\mathbf{v}] = \mathbf{0}.$$

Consequently, for such a $\mathbf{v}$

$$e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v} = \mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} + \ldots + \frac{t^{m-1}}{(m-1)!}(\mathcal{A} - \lambda\mathcal{I})^{m-1}\mathbf{v}.$$

and

$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} + \ldots + \frac{t^{m-1}}{(m-1)!}(\mathcal{A} - \lambda\mathcal{I})^{m-1}\mathbf{v}\right). \tag{3.1.28}$$

Thus, to find all solutions to $\mathbf{y}' = \mathcal{A}\mathbf{y}$ it is sufficient to find n independent vectors $\mathbf{v}$ satisfying (3.1.27) for some scalars λ. To check consistency of this method with our previous consideration we observe that if $\lambda = \lambda_1$ is a single eigenvalue of $\mathcal{A}$ with a corresponding eigenvector $\mathbf{v^1}$, then $(\mathcal{A} - \lambda_1\mathcal{I})\mathbf{v^1} = 0$, thus m of (3.1.27) is equal to 1. Consequently, the sum in (3.1.28) terminates after the first term and we obtain

$$\mathbf{y_1}(t) = e^{\lambda_1 t}\mathbf{v^1}$$

in accordance with (3.1.21). From our discussion of eigenvalues and eigenvectors it follows that if λ_i is a multiple eigenvalue of $\mathcal{A}$ of algebraic multiplicity n_i and the geometric multiplicity is less then n_i, that is, there is less than n_i linearly independent eigenvectors corresponding to λ_i, then the missing independent vectors can be found by solving successively equations $(\mathcal{A} - \lambda_i\mathcal{I})^k\mathbf{v} = \mathbf{0}$ with k running at most up to n_1. Thus, we have the following algorithm for finding n linearly independent solutions to $\mathbf{y}' = \mathcal{A}\mathbf{y}$:

1. Find all eigenvalues of $\mathcal{A}$;

2. If λ is a single real eigenvalue, then there is an eigenvector $\mathbf{v}$ so that the solution is given by
$$\mathbf{y}(t) = e^{\lambda t}\mathbf{v} \tag{3.1.29}$$

3. If λ is a single complex eigenvalue $\lambda = \xi + i\omega$, then there is a complex eigenvector $\mathbf{v} = \Re\mathbf{v} + i\Im\mathbf{v}$ such that two solutions corresponding to λ (and $\bar{\lambda}$) are given by
$$\begin{aligned} \mathbf{y^1}(t) &= e^{\xi t}(\cos\omega t\, \Re\mathbf{v} - \sin\omega t\, \Im\mathbf{v}) \\ \mathbf{y^2}(t) &= e^{\xi t}(\cos\omega t\, \Im\mathbf{v} + \sin\omega t \Re\mathbf{v}) \end{aligned} \tag{3.1.30}$$

4. If λ is a multiple eigenvalue with algebraic multiplicity k (that is, λ is a multiple root of the characteristic equation of multiplicity k), then we first find eigenvectors by solving $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$. For these eigenvectors the solution is again given by (3.1.29) (or (3.1.30), if λ is complex). If we found k independent eigenvectors, then our work with this eigenvalue is finished. If not, then we look for vectors that satisfy $(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v} = \mathbf{0}$ but $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} \neq \mathbf{0}$. For these vectors we have the solutions
$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}\right).$$
If we still do not have k independent solutions, then we find vectors for which $(\mathcal{A} - \lambda\mathcal{I})^3\mathbf{v} = \mathbf{0}$ and $(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v} \neq \mathbf{0}$, and for such vectors we construct solutions
$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} + \frac{t^2}{2}(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v}\right).$$
This procedure is continued till we have k solutions (by the properties of eigenvalues we have to repeat this procedure at most k times).

   If λ is a complex eigenvalue of multiplicity k, then also $\bar{\lambda}$ is an eigenvalue of multiplicity k and we obtain pairs of real solutions by taking real and imaginary parts of the formulae presented above.

Remark 3.1.9. Once we know that all solutions must be of the form (3.1.28) with the degree of the polynomial being at most equal to the algebraic multiplicity of λ, we can use the method of undetermined coefficients to find the solutions. Namely, if λ is an eigenvalue of multiplicity k, the we can look for a solutions in the form
$$\mathbf{y}(t) = e^{\lambda t}\left(\mathbf{a_0} + \mathbf{a_1}t + \ldots \mathbf{a_{k-1}}t^{k-1}\right)$$
where unknown vectors $\mathbf{a_0}, \ldots, \mathbf{a_{k-1}}$ are to be determined by inserting $\mathbf{y}(t)$ into the equation and solving the resulting simultaneous systems of algebraic equations.

Example 3.1.10. Find three linearly independent solutions of the differential equation

$$\mathbf{y}' = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \mathbf{y}.$$

To obtain the eigenvalues we calculate the characteristic polynomial

$$\begin{aligned} p(\lambda) &= det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} 1-\lambda & 1 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix} \\ &= (1-\lambda)^2(2-\lambda) \end{aligned}$$

so that $\lambda_1 = 1$ is eigenvalue of multiplicity 2 and $\lambda_2 = 2$ is an eigenvalue of multiplicity 1.

(i) $\lambda = 1$: We seek all non-zero vectors such that

$$(\mathcal{A} - 1\mathcal{I})\mathbf{v} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This implies that $v_2 = v_3 = 0$ and v_1 is arbitrary so that we obtain the corresponding solutions

$$\mathbf{y}^{\mathbf{1}}(t) = C_1 e^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

However, this is only one solution and $\lambda_1 = 1$ has algebraic multiplicity 2, so we have to look for one more solution. To this end we consider

$$\begin{aligned} (\mathcal{A} - 1\mathcal{I})^2\mathbf{v} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

so that $v_3 = 0$ and both v_1 and v_2 arbitrary. The set of all solutions here is a two-dimensional space spanned by

$$\begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

We have to select from this subspace a vector that is not a solution to $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$. Since for the later the solutions are scalar multiples of the vector $(1,0,0)$ we see that the vector $(0,1,0)$ is not of this form and consequently can be taken as the second independent vector corresponding to the eigenvalue $\lambda_1 = 1$. Hence

$$\begin{aligned}\mathbf{y^2}(t) &= e^t\left(\mathcal{I}+t(\mathcal{A}-\mathcal{I})\right)\begin{pmatrix}0\\1\\0\end{pmatrix} = e^t\left(\begin{pmatrix}0\\1\\0\end{pmatrix}+t\begin{pmatrix}0&1&0\\0&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix}\right)\\ &= e^t\begin{pmatrix}0\\1\\0\end{pmatrix}+te^t\begin{pmatrix}1\\0\\0\end{pmatrix}=e^t\begin{pmatrix}t\\1\\0\end{pmatrix}\end{aligned}$$

(ii) $\lambda = 2$: We seek solutions to

$$(\mathcal{A}-2\mathcal{I})\mathbf{v}=\begin{pmatrix}-1&1&0\\0&-1&0\\0&0&0\end{pmatrix}\begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

This implies that $v_1 = v_2 = 0$ and v_3 is arbitrary so that the corresponding solutions are of the form

$$\mathbf{y^3}(t) = C_3e^{2t}\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Thus we have found three linearly independent solutions.

Fundamental solutions and nonhomogeneous problems

Let us suppose that we have n linearly independent solutions $\mathbf{y^1}(t),\ldots,\mathbf{y^n}(t)$ of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$, where $\mathcal{A}$ is an $n\times n$ matrix, like the ones constructed in the previous paragraphs. Let us denote by $\mathcal{Y}(t)$ the matrix

$$\mathcal{Y}(t)=\begin{pmatrix}y_1^1(t)&\ldots&y_1^n(t)\\\vdots&&\vdots\\y_n^1(t)&\ldots&y_n^n(t)\end{pmatrix},$$

that is, the columns of $\mathcal{Y}(t)$ are the vectors $\mathbf{y^i}$, $i = 1,\ldots,n$. Any such matrix is called a *fundamental matrix* of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$.

We know that for a given initial vector $\mathbf{y^0}$ the solution is given by

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{y^0}$$

on one hand, and, by Theorem 3.1.3, by

$$\mathbf{y}(t) = C_1\mathbf{y^1}(t)+\ldots+C_n\mathbf{y^n}(t) = \mathcal{Y}(t)\mathbf{C},$$

on the other, where $\mathbf{C} = (C_1, \ldots, C_n)$ is a vector of constants to be determined. By putting $t = 0$ above we obtain the equation for $\mathbf{C}$

$$\mathbf{y^0} = \mathcal{Y}(0)\mathbf{C}$$

Since $\mathcal{Y}$ has independent vectors as its columns, it is invertible, so that

$$\mathbf{C} = \mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Thus, the solution of the initial value problem

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{y^0}$$

is given by

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Since $e^{t\mathbf{A}}\mathbf{y^0}$ is also a solution, by the uniqueness theorem we obtain explicit representation of the exponential function of a matrix

$$e^{t\mathcal{A}} = \mathcal{Y}(t)\mathcal{Y}^{-1}(0). \tag{3.1.31}$$

Let us turn our attention to the non-homogeneous system of equations

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}(t). \tag{3.1.32}$$

The general solution to the homogeneous equation ($\mathbf{g}(t) \equiv 0$) is given by

$$\mathbf{y_h}(t) = \mathcal{Y}(t)\mathbf{C},$$

where $\mathcal{Y}(t)$ is a fundamental matrix and $\mathbf{C}$ is an arbitrary vector. Using the technique of variation of parameters, we will be looking for the solution in the form

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{u}(t) = u_1(t)\mathbf{y^1}(t) + \ldots + u_n(t)\mathbf{y^n}(t) \tag{3.1.33}$$

where $\mathbf{u}(t) = (u_1(t), \ldots, u_n(t))$ is a vector-function to be determined so that (3.1.33) satisfies (3.1.32). Thus, substituting (3.1.33) into (3.1.32), we obtain

$$\mathcal{Y}'(t)\mathbf{u}(t) + \mathcal{Y}(t)\mathbf{u}'(t) = \mathcal{A}\mathcal{Y}(t)\mathbf{u}(t) + \mathbf{g}(t).$$

Since $\mathcal{Y}(t)$ is a fundamental matrix, $\mathcal{Y}'(t) = \mathcal{A}\mathcal{Y}(t)$ and we find

$$\mathcal{Y}(t)\mathbf{u}'(t) = \mathbf{g}(t).$$

As we observed earlier, $\mathcal{Y}(t)$ is invertible, hence

$$\mathbf{u}'(t) = \mathcal{Y}^{-1}(t)\mathbf{g}(t)$$

and

$$\mathbf{u}(t) = \int^t \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds + \mathbf{C}.$$

Finally, we obtain

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{C} + \mathcal{Y}(t)\int^t \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds \tag{3.1.34}$$

This equation becomes much simpler if we take $e^{t\mathcal{A}}$ as a fundamental matrix because in such a case $\mathcal{Y}^{-1}(t) = \left(e^{t\mathcal{A}}\right)^{-1} = e^{-t\mathcal{A}}$, that is, to calculate the inverse of $e^{t\mathcal{A}}$ it is enough to replace t by $-t$. The solution (3.1.34) takes then the form

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{C} + \int e^{(t-s)\mathcal{A}}\mathbf{g}(s)ds. \tag{3.1.35}$$

Example 3.1.11. Find the general solution to

$$\begin{aligned} y_1' &= 5y_1 + 3y_2 + 2te^{2t}, \\ y_2' &= -3y_1 - y_2 + 4. \end{aligned}$$

Writing this system in matrix notation, we obtain

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}(t)$$

with

$$\mathcal{A} = \begin{pmatrix} 5 & 3 \\ -3 & -1 \end{pmatrix}$$

and

$$\mathbf{g}(t) = \begin{pmatrix} 2te^{2t} \\ 4 \end{pmatrix}.$$

We have to find $e^{t\mathcal{A}}$. The first step is to find two independent solutions to the homogeneous system. The characteristic polynomial is

$$p(\lambda) = \begin{vmatrix} 5-\lambda & 3 \\ -3 & -1-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 4 = (\lambda-2)^2$$

We have double eigenvalue $\lambda = 2$. Solving

$$(\mathcal{A} - 2\mathcal{I})\mathbf{v} = \begin{pmatrix} 3 & 3 \\ -3 & -3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

we obtain $v_1 = -v_2$ so that we obtain the eigenvector $\mathbf{v}^1 = (1, -1)$ and the corresponding solution

$$\mathbf{y}^{\mathbf{1}}(t) = C_1 e^{2t}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Since $\lambda_1 = 2$ has algebraic multiplicity 2, we have to look for another solution. To this end we consider

$$\begin{aligned}(\mathcal{A}-2\mathcal{I})^2\mathbf{v} &= \begin{pmatrix} 3 & 3 \\ -3 & -3 \end{pmatrix}\begin{pmatrix} 3 & 3 \\ -3 & -3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix},\end{aligned}$$

so that v_1 and v_2 arbitrary. We must simply select a vector linearly independent of $\mathbf{y^1}$ – to make things simple we can take

$$\mathbf{y^2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so that the second solution is given as

$$\begin{aligned}\mathbf{y^2}(t) &= e^{2t}\left(\mathcal{I}+t(\mathcal{A}-\mathcal{I})\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = e^{2t}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 3 & 3 \\ -3 & -3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \\ &= e^{2t}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + te^{2t}\begin{pmatrix} 3 \\ -3 \end{pmatrix} = e^{2t}\begin{pmatrix} 1+3t \\ -3t \end{pmatrix}\end{aligned}$$

Thus, the fundamental matrix is given by

$$\mathcal{Y}(t) = e^{2t}\begin{pmatrix} 1 & 1+3t \\ -1 & -3t \end{pmatrix}$$

with

$$\mathcal{Y}(0) = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.$$

The discriminant of $\mathcal{Y}(0)$ is equal to 1 and we immediately obtain

$$\mathcal{Y}^{-1}(0) = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}.$$

so that

$$e^{t\mathcal{A}} = \mathcal{Y}(t)\mathcal{Y}^{-1}(0) = e^{2t}\begin{pmatrix} 1+3t & 3t \\ -3t & 1-3t \end{pmatrix}.$$

Thus

$$e^{-t\mathcal{A}} = e^{-2t}\begin{pmatrix} 1-3t & -3t \\ 3t & 1+3t \end{pmatrix}$$

and

$$e^{-t\mathcal{A}}\mathbf{g}(t) = e^{-2t}\begin{pmatrix} 1-3t & -3t \\ 3t & 1+3t \end{pmatrix}\begin{pmatrix} 2te^{2t} \\ 4 \end{pmatrix} = \begin{pmatrix} 2t-6t^2-12te^{-2t} \\ 6t^2+4e^{-2t}+12te^{-2t} \end{pmatrix}.$$

To find the particular solution, we integrate the above, getting

$$\int e^{-t\mathcal{A}}\mathbf{g}(t)dt = \begin{pmatrix} \int(2t-6t^2-12te^{-2t})dt \\ \int(6t^2+4e^{-2t}+12te^{-2t})dt \end{pmatrix} = \begin{pmatrix} t^2-2t^3+3(2t+1)e^{-2t} \\ 2t^3-(6t+5)e^{-2t} \end{pmatrix},$$

and multiply the above by $e^{t\mathcal{A}}$ to obtain

$$e^{2t}\begin{pmatrix} 1+3t & 3t \\ -3t & 1-3t \end{pmatrix}\begin{pmatrix} t^2-2t^3+3(2t+1)e^{-2t} \\ 2t^3-(6t+5)e^{-2t} \end{pmatrix} = \begin{pmatrix} (t^2+t^3)e^{2t}+3 \\ -t^3e^{2t}-5 \end{pmatrix}.$$

Therefore, the general solution is given by

$$\mathbf{y}(t) = e^{2t}\begin{pmatrix} 1+3t & 3t \\ -3t & 1-3t \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} + \begin{pmatrix} (t^2+t^3)e^{2t}+3 \\ -t^3e^{2t}-5 \end{pmatrix}$$

where C_1 and C_2 are arbitrary constants.

3.1.5 An application

In Subsection 1.3.5 we have derived the system

$$\begin{aligned} \frac{dx_1}{dt} &= r_1 + p_2\frac{x_2}{V} - p_1\frac{x_1}{V} \\ \frac{dx_2}{dt} &= p_1\frac{x_1}{V} - (R_2+p_2)\frac{x_2}{V}. \end{aligned} \tag{3.1.36}$$

describing mixing of components in two containers. Here, x_1 and x_2 are the amount of dye in vats 1 and 2, respectively. We re-write these equations using concentrations $c_1 = x_1/V$ and $c_2 = x_2/V$, getting

$$\begin{aligned} \frac{dc_1}{dt} &= \frac{r_1}{V} + \frac{p_2}{V}c_2 - \frac{p_1}{V}c_1 \\ \frac{dc_2}{dt} &= \frac{p_1}{V}c_1 - \frac{R_2+p_2}{V}c_1. \end{aligned} \tag{3.1.37}$$

We solve this equations for numerical values of the flow rates $r_1/V = 0.01$, $p_1/V = 0.04$, $p_2/V = 0.03$ and $(R_2+p_2)/V = 0.05$,

$$\begin{aligned} \frac{dc_1}{dt} &= 0.01 - 0.04c_1 + 0.03c_2 \\ \frac{dc_2}{dt} &= 0.04c_1 - 0.05c_2, \end{aligned} \tag{3.1.38}$$

and assume that at time $t = 0$ there was no dye in either vat, that is, we put

$$c_1(0) = 0, \qquad c_2(0) = 0.$$

To practice another technique, we shall solve this system by reducing it to a second order equations, as described in Example 3.1.1. Differentiating the first equation and using the second we have

$$\begin{aligned} c_1'' &= -0.04c_1' + 0.03c_2' \\ &= -0.04c_1' + 0.03(0.04c_1 - 0.05c_2) \\ &= -0.04c_1' + 0.03\left(0.04c_1 - 0.05\cdot\frac{100}{3}(c_1' - 0.01 + 0.04c_1)\right) \end{aligned}$$

so that, after some algebra,

$$c_1'' + 0.09c_1' + 0.008c_1 = 0.005.$$

We have obtained second order non-homogenous equation with constant coefficients. To find the characteristic roots we solve the quadratic equation

$$\lambda^2 + 0.09\lambda + 0.008 = 0$$

getting $\lambda_1 = -0.08$ and $\lambda_2 = -0.01$. Thus, the space of solutions of the homogeneous equations is spanned by $e^{-0.01t}$ and $e^{-0.08t}$. The right hand side is a constant and since zero is not a characteristic root, we can look for a solution to the nonhomogeneous problem in the form $y_p(t) = A$, which immediately gives $y_p(t) = 5/8$ so that the general solution of the non-homogeneous equation for c_1 is given by

$$c_1(t) = C_1e^{-0.08t} + C_2e^{-0.01t} + \frac{5}{8},$$

where C_1 and C_2 are constants whose values are to be found from the initial conditions.

Next we find c_2 by solving the first equation with respect to it, so that

$$\begin{aligned} c_2 &= \frac{100}{3}\left(c_1' + 0.04c_1 - 0.01\right) \\ &= \frac{100}{3}\Big(-0.08C_1e^{-0.08t} - 0.01C_2e^{-0.01t} \\ &\quad + 0.04\left(C_1e^{-0.08t} + C_2e^{-0.01t} + \frac{5}{8}\right) - 0.01\Big) \end{aligned}$$

and

$$c_2(t) = -\frac{4}{3}C_1e^{-0.08t} + C_2e^{-0.01t} + 0.5.$$

Finally, we use the initial conditions $c_1(0) = 0$ and $c_2(0) = 0$ to get the system of algebraic equations for C_1 and C_2

$$\begin{aligned} C_1 + C_2 &= -\frac{5}{8}, \\ \frac{4}{3}C_1 - C_2 &= \frac{1}{2}. \end{aligned}$$

From these equations we find $C_1 = -3/56$ and $C_2 = -4/7$. Hence

$$\begin{aligned} c_1(t) &= -\frac{3}{56}e^{-0.08t} - \frac{4}{7}e^{-0.01t} + \frac{5}{8} \\ c_2(t) &= \frac{1}{14}e^{-0.08t} - \frac{4}{7}e^{-0.01t} + \frac{1}{2}. \end{aligned}$$

From the solution formulae we obtain that $\lim\limits_{t\to\infty} c_1(t) = \frac{5}{8}$ and $\lim\limits_{t\to\infty} c_2(t) = \frac{1}{2}$. This means that the concentrations approach the steady state concentration as t becomes large. This is illustrated in Figures 2.10 and 2.11

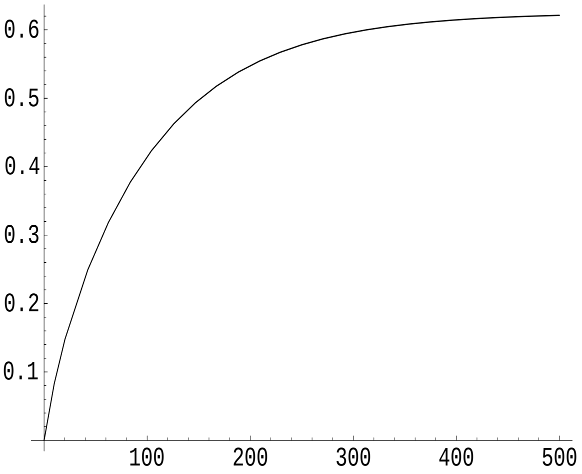

Fig 2.10 Approach to the steady-state of the concentration c_1*.*

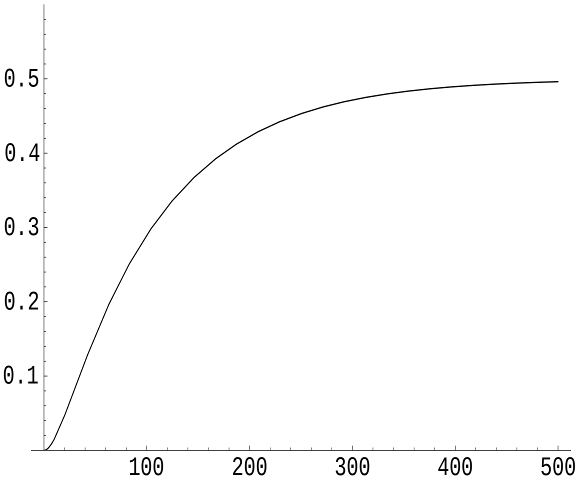

Fig 2.11 Approach to the steady state of the concentration c_2*.*

3.2 Second order linear equations

Second order equations occur very often in practice so that it is useful to specify the general theory of systems for this particular case.

$$\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = f(t) \tag{3.2.1}$$

where a_1, a_0 are real constants and f is a given continuous function. As before in what follows we shall abbreviate $d^2y/dt^2 = y''$ and $dy/dt = y'$.

As we mentioned earlier, (3.2.1) can be written as an equivalent system of 2 first order equations by introducing new variables $y_1 = y$, $y_2 = y' = y_1'$,

$$\begin{array}{rcl} y_1' & = & y_2, \\ \vdots & & \vdots \\ y_2' & = & -a_1y_2 - a_0y_1 + f(t). \end{array}$$

Note that if (3.2.1) was supplemented with the initial conditions $y(t_0) = y^0, y'(t_0) = y^1$, then these conditions will become natural initial conditions for the system as $y_1(t_0) = y^0, y_2(t_0) = y^1$.

Let us first recall the theory for first order linear equations, specified to the case of a constant coefficient a:

$$y' + ay = f(t). \tag{3.2.2}$$

By (2.3.14), the general solution to (3.2.2) is given by

$$y(t) = Ce^{-at} + e^{-at}\int e^{as}f(s)ds,$$

where the first term is the general solution of the homogeneous ($f \equiv 0$) version of (3.2.2) and the second is a particular solution to (3.2.2). This suggests that a sensible strategy for solving (3.2.1) is to look first for solutions to the associated homogeneous equation

$$\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = 0. \tag{3.2.3}$$

Let as denote by y_0 the general solution to (3.2.3), that is, y_0 is really a class of functions depending on two constants. Next, let y_p be a particular solution of (3.2.1) and consider $y(t) = y_p(t) + z(t)$. Then

$$\begin{array}{rcl} y'' + a_1y' + a_0y & = & y_p'' + a_1y_p' + a_0y_p + z'' + a_1z' + a_0z \\ & = & f(t) + z'' + a_1z' + a_0z, \end{array}$$

that is, y is a solution to (3.2.1) if and only if z is any solution to (3.2.3) or, in other words, if and only if z is the general solution to (3.2.3), $z = y_c$.

Accordingly, we shall first develop methods for finding general solutions to homogeneous equations.

3.2.1 Homogeneous equations

Let us consider the homogeneous equation (3.2.3)

$$\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0 y = 0. \tag{3.2.4}$$

Since the space of solutions of the corresponding 2×2 homogeneous system

$$\begin{array}{rcl} y_1' & = & y_2, \\ \vdots & & \vdots \\ y_2' & = & -a_1 y_2 - a_0 y_1. \end{array} \tag{3.2.5}$$

is two-dimensional, the space of solutions to (3.2.4) is also two-dimensional, that is, there are two independent solutions of (3.2.4) $y_1(t), y_2(t)$ such that any other solution is given by

$$y(t) = C_1 y_1(t) + C_2 y_2(t).$$

How can we recover the general solution to (3.2.4) from the solution $\mathbf{y}(t)$ of the system? A function $y(t)$ solves (3.2.4) if and only if $\mathbf{y}(t) = (y_1(t), y_2(t)) = (y(t), y'(t))$ solves the system (3.2.5), we see that the general solution to (3.2.4) can be obtained by taking first components of the solution of the associated system (3.2.5). We note once again that if $\mathbf{y^1}(t) = (y_1^1(t), y_2^1(t)) = (y^1(t), \frac{dy^1}{dt}(t))$ and $\mathbf{y^1}(t) = (y_1^2(t), y_2^2(t)) = (y^2(t), \frac{dy^2}{dt}(t))$ are two linearly independent solutions to (3.2.5), then $y^1(t)$ and $y^2(t)$ are linearly independent solutions to (3.2.4). In fact, otherwise we would have $y^1(t) = Cy^2(t)$ for some constant C and therefore also $\frac{dy^1}{dt}(t) = C\frac{dy^2}{dt}(t)$ so that the wronskian, having the second column as a scalar multiple of the first one, would be zero, contrary to the assumption that $\mathbf{y^1}(t)$ and $\mathbf{y^2}(t)$ are linearly independent.

Two find explicit formulae for two linearly independent particular solutions to (3.2.4) we write the equation for the characteristic polynomial of (3.2.5):

$$\begin{vmatrix} -\lambda & 1 \\ -a_0 & -a_1 - \lambda \end{vmatrix} = 0$$

that is

$$\lambda^2 + a_1\lambda + a_0 = 0,$$

which is also called the characteristic polynomial of (3.2.4). This is a quadratic equation in λ which is zero when $\lambda = \lambda_1$ or $\lambda = \lambda_2$ with

$$\lambda_{1,2} = \frac{-a_1 \pm \sqrt{\Delta}}{2}$$

where the discriminant $\Delta = a_1^2 - 4a_0$.

If $\Delta > 0$, then $\lambda_1 \neq \lambda_2$, and we obtain two different solutions $y_1 = e^{\lambda_1 t}$ and $y_2 = e^{\lambda_2 t}$. Thus

$$y(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$$

with two arbitrary constants is the sought general solution to (3.2.4). If $\Delta < 0$, then λ_1 and λ_2 are complex conjugates: $\lambda_1 = \xi + i\omega$, $\lambda_2 = \xi - i\omega$ with $\xi = -a_1/2$ and $\omega = -\sqrt{-\Delta}/2$. Since in many applications it is undesirable to work with complex functions, we shall express the solution in terms of real functions. Using the Euler formula for the complex exponential function, we obtain

$$\begin{aligned} y(t) &= C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} = C_1 e^{\xi t}(\cos \omega t + i \sin \omega t) + C_2 e^{\xi t}(\cos \omega t - i \sin \omega t) \\ &= (C_1 + C_2) e^{\xi t} \cos \omega t + i(C_1 - C_2) e^{\xi t} \sin \omega t. \end{aligned}$$

If as the constants C_1 and C_2 we take complex conjugates $C_1 = (A - iB)/2$ and $C_2 = (A + iB)/2$ with arbitrary real A and B, then we obtain y as a combination of two real functions with two arbitrary real coefficients

$$y(t) = Ae^{\xi t} \cos \omega t + Be^{\xi t} \sin \omega t.$$

We have left behind the case $\lambda_1 = \lambda_2$ (necessarily real). In this case we have only one function $e^{\lambda_1 t}$ with one arbitrary constant C_1 so that $y(t) = C_1 e^{\lambda_1 t}$ is not the general solution to (3.2.4). Using the theory for systems, we obtain the other solution in the form

$$y_2(t) = te^{\lambda_1 t}$$

with $\lambda_1 = -a_1/2$. Thus the general solution is given by

$$y(t) = (C_1 + C_2 t)e^{-ta_1/2}.$$

Summarizing, we have the following general solutions corresponding to various properties of the roots of the characteristic polynomial λ_1, λ_2.

$$\begin{array}{lll} y(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} & \text{if} & \lambda_1 \neq \lambda_2,\ \lambda_1, \lambda_2 \text{ real}, \\ y(t) = C_1 e^{\xi t} \cos \omega t + C_2 e^{\xi t} \sin \omega t & \text{if} & \lambda_{1,2} = \xi \pm i\omega, \\ y(t) = (C_1 + C_2 t)e^{-t\lambda} & \text{if} & \lambda_1 = \lambda_2 = \lambda. \end{array}$$

3.2.2 Nonhomogeneous equations

At the beginning of this section we have shown that to find the general solution to

$$\frac{d^2y}{dt^2}+a_1\frac{dy}{dt}+a_0y=f(t) \tag{3.2.6}$$

we have to find the general solution to the homogeneous version (3.2.4) and then just one particular solution to the full equation (3.2.6). In the previous subsection we have presented the complete theory for finding the general solution to homogeneous equations. Here we shall discuss two methods of finding solutions to nonhomogeneous equation. We start with the so-called variation of parameters method that is very general but sometimes rather cumbersome to apply. The second method, of judicious guessing, can be applied for special right-hand sides only, but then it gives the solution really quickly.

Variation of parameters The method of variations of parameters was introduced for systems of equations, specifying it for second order equations would be, however, quite cumbersome. Thus, we shall derive it from scratch. Let

$$y_0(t)=C_1y_1(t)+C_2y_2(t)$$

be the general solution to the homogeneous version of (3.2.6). We are looking for a solution to (3.2.6) in the form

$$y(t)=u(t)y_1(t)+v(t)y_2(t), \tag{3.2.7}$$

that is, we allow the arbitrary parameters C_1 and C_2 to depend on time. To determine $v(t)$ and $u(t)$ so that (3.2.7) is a solution to (3.2.6), we substitute $y(t)$ to the equation. Since there is only one equation, this will give one condition to determine two functions, giving some freedom to pick up the second condition is such a way that the resulting equation becomes the easiest. Let us work it out. Differentiating (3.2.7) we have

$$y'=uy_1'+vy_2'+u'y_1+v'y_2,$$

and

$$y''=u'y_1'+v'y_2'+uy_1''+vy_2''+u''y_1+v''y_2+u'y_1'+v'y_2'.$$

We see that there appear second order derivatives of the unknown functions and this is something we would like to avoid, as we are trying to simplify a second order equation. For the second order derivatives not to appear we simply require that the part of y' containing u' and v' to vanish, that is,

$$u'y_1+v'y_2=0.$$

With this, we obtain

$$\begin{aligned} y' &= uy_1' + vy_2', \\ y'' &= u'y_1' + v'y_2' + uy_1'' + vy_2''. \end{aligned}$$

Substituting these into (3.2.6) we obtain

$$\begin{aligned} &u'y_1' + v'y_2' + uy_1'' + vy_2'' + a_1(uy_1' + vy_2') + a_0(uy_1 + vy_2) \\ &= u(y_1'' + a_1y_1' + a_0y_1) + v(y_2'' + a_1y_2' + a_0y_2) + u'y_1' + v'y_2' \\ &= f(t). \end{aligned}$$

Since y_1 and y_2 are solutions of the homogeneous equation, first two terms in the second line vanish and for y to satisfy (3.2.6) we must have

$$u'y_1' + v'y_2' = f(t).$$

Summarizing, to find u and v such that (3.2.7) satisfies (3.2.6) we must solve the following system of equations

$$u'y_1 + v'y_2 = 0, \tag{3.2.8}$$

$$u'y_1' + v'y_2' = f(t) \tag{3.2.9}$$

System (3.2.9) is to be solved for u' and v' and the solution integrated to find u and v.

Remark 3.2.1. System (3.2.9) can be solved by determinants. The main determinant

$$W(t) = \begin{vmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{vmatrix} = y_1(t)y_2'(t) - y_2(t)y_1'(t) \tag{3.2.10}$$

is the *wronskian* and plays an important rôle in the general theory of differential equations. Here we shall only not that clearly for (3.2.9) to be solvable, $W(t) \neq 0$ for all t which is ensured by y_1 and y_2 being linearly independent which, as we know, must be the case if y_0 is the general solution to the homogeneous equation, see Remark 3.1.5.

Example 3.2.2. Find the solution to

$$y'' + y = \tan t$$

on the interval $-\pi/2 < t < \pi/2$ satisfying the initial conditions $y(0) = 1$ and $y'(0) = 1$.

Step 1.
General solution to the homogeneous equation

$$y'' + y = 0$$

is obtained by finding the roots of the characteristic equation

$$\lambda^2 + 1 = 0.$$

We have $\lambda_{1,2} = \pm i$ so that $\xi = 0$ and $\omega = 1$ and we obtain two independent solutions

$$y_1(t) = \cos t, \qquad y_2(t) = \sin t.$$

Step 2.
To find a solution to the nonhomogeneous equations we first calculate wronskian

$$W(t) = \begin{vmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{vmatrix} = 1.$$

Solving (3.2.9), we obtain

$$u'(t) = -\sin t \tan t, \qquad v'(t) = \cos t \tan t.$$

Then

$$\begin{aligned} u(t) &= -\int \sin t \tan t dt = -\int \frac{\sin^2}{\cos t} dt = -\int \frac{1-\cos^2}{\cos t} dt \\ &= \int \cos t dt - \int \frac{dt}{\cos t} = \sin t - \int \frac{dt}{\cos t} \\ &= \sin t - \ln|\sec t + \tan t| = \sin t - \ln(\sec t + \tan t), \end{aligned}$$

where the absolute value bars can be dropped as for $-\pi/2 < t < \pi t \sec t + \tan t > 0$. Integrating the equation for v we find

$$v(t) = -\cos t$$

and a particular solution the non-homogeneous equation can be taken to be

$$\begin{aligned} y_p(t) &= u(t)y_1(t) + v(t)y_2(t) = \cos t(\sin t - \ln(\sec t + \tan t)) + \sin t(-\cos t) \\ &= -\cos t \ln(\sec t + \tan t). \end{aligned}$$

Note that we have taken the constants of integration to be zero in each case. This is allowed as we are looking for particular integrals and we are free to pick up the simplest particular solution.

Thus, the general solution to the non-homogeneous equation is

$$y(t) = C_1 \cos t + C_2 \sin t - \cos t \ln(\sec t + \tan t).$$

Step 3.
To solve the initial value problem we must find the derivative of y:

$$y'(t) = -C_1 \sin t + C_2 \sin t + \sin t \ln(\sec t + \tan t) - 1$$

so that we obtain

$$1 = y(0) = C_1, \qquad 1 = y'(0) = C_2 - 1,$$

hence $C_1 = 1$ and $C_2 = 2$. Therefore

$$y(t) = \cos t + 2\sin t - \cos t \ln(\sec t + \tan t).$$

Judicious guessing

The method of judicious guessing, called also the method of undetermined coefficients, is based on the observation that for some functions the operations performed on the left-hand side of the differential equation, that is, taking derivatives, multiplying by constants and addition, does not change the form of the function. To wit, the derivative of a polynomial is a polynomial, the derivative of an exponential function is an exponential function and, in general the derivative of the product of an exponential function and a polynomial is again of the same form. Trigonometric functions $\sin t$ and $\cos t$ are included into this class by Euler's formulae $\sin t = \frac{e^{it}-e^{-it}}{2i}$ and $\cos t = \frac{e^{it}+e^{-it}}{2}$. Thus, if the right-hand side is of this form, then it makes sense to expect that the same of the solution. Let us test this hypothesis on the following example.

Example 3.2.3. Find a particular solution to

$$y'' - 2y' - 3y = 3t^2.$$

The right-hand side is a polynomial of the second degree so we will look for a solution amongst polynomials. To decide polynomial of what degree we should try we note that if we try polynomials of zero or first degree then the left-hand side will be at most of this degree, as the differentiation lowers the degree of a polynomial. Thus, the simplest candidate appears to be a polynomial of second degree

$$y(t) = At^2 + Bt + C,$$

where A, B, C are coefficients to be determined. Inserting this polynomial into the equation we get

$$y'' - 2y' - 3y = 2A - 2B - 3C - (4A + 3B)t - 3At^2 = 3t^2,$$

from which we obtain the system

$$\begin{array}{rcl} -3A & = & 3, \\ -4A - 3B & = & 0, \\ 2A - 2B - 3C & = & 0. \end{array}$$

Solving this system, we obtain $A = -1$, $B = 4/3$ and $C = -14/9$ so that the solution is

$$y(t) = -t^2 - \frac{4}{3}t - \frac{14}{9}.$$

Unfortunately, there are some pitfalls in this method, as shown in the following example.

Example 3.2.4. Find a particular solution to

$$y'' - 2y' - 3y = e^{-t}.$$

Using our method, we take $y(t) = Ae^{-t}$ but inserting it into the equation we find that

$$y'' - 2y' - 3y = Ae^{-t} + 2Ae^{-t} - 3Ae^{-t} = 0 \neq e^{-t},$$

so that no choice of the constant A can turn $y(t)$ into the solution of our equation. The reason for this is that e^{-t} is a solution to the homogeneous equation what could be ascertained directly by solving the characteristic equation $\lambda^2 - 2\lambda - 3 = (\lambda+1)(\lambda-3)$. A way of this trouble is to consider $y(t) = Ate^{-t}$ so that $y' = Ae^{-t} - Ate^{-t}$ and $y'' = -2Ae^{-t} + Ate^{-t}$ and

$$\begin{aligned} y'' - 2y' - 3y &= -2Ae^{-t} + Ate^{-t} - 2(Ae^{-t} - Ate^{-t}) - 3Ate^{-t} \\ &= -4e^{-t}, \end{aligned}$$

which agrees with e^{-t} if $A = -\frac{1}{4}$. Thus we have a particular solution

$$y(t) = -\frac{1}{4}te^{-t}.$$

In general, it can be proved that the following procedure always produces the solution to

$$y'' + a_1y' + a_0y = t^m e^{at} \tag{3.2.11}$$

where $a_0 \neq 0$ and m is a non-negative integer.

I. When a is not a root of the characteristic equation $\lambda^2 + a_1\lambda + a_0 = 0$, then we use

$$y(t) = e^{at}(A_m t^m + A_{m-1}t^{m-1} + \ldots + A_0); \tag{3.2.12}$$

II. If a is a single root of the characteristic equation, then use (3.2.12) multiplied by t and if a is a double root, then use (3.2.12) multiplied by t^2.

Remark 3.2.5. Note that if $a_0 = 0$, then (3.2.11) is reducible to a first order equation by methods of Subsection 2.3.4.

Also, equations with right-hand sides of the form

$$y'' + a_1y' + a_0y = f_1(t) + f_2(t) \ldots + f_n(t), \tag{3.2.13}$$

can be handled as if $y_i(t)$ is a particular solution to

$$y'' + a_1 y' + a_0 y = f_i(t), \qquad i = 1, \ldots, n,$$

then the sum $y_p(t) = y_1(t) + y_2(t) \ldots + y_n(t)$ is a particular solution to (3.2.13) as my be checked by direct substitution.

Example 3.2.6. Find a particular solution of

$$y'' + 4y = 32t \cos 2t - 8 \sin 2t.$$

Let us first find the characteristic roots. From the equation $\lambda^2 + 4 = 0$ we find $\lambda = \pm 2i$. Next we convert the RHS of the equation to the exponential form. Since $\cos 2t = (e^{i2t} + e^{-i2t})/2$ and $\sin 2t = (e^{i2t} - e^{-i2t})/2i$, we obtain

$$32t \cos 2t - 8 \sin 2t = (16t + 4i)e^{i2t} + (16t - 4i)e^{-i2t}.$$

In both cases we have the exponent being a single root of the characteristic equation so that we will be looking for solutions in the form $y_1(t) = t(At + B)e^{i2t}$ and $y_2(t) = t(Ct + D)e^{-i2t}$. For y_1 we obtain $y_1'(t) = (2At + B)e^{i2t} + 2it(At + B)e^{i2t}$ and $y_1''(t) = 2Ae^{i2t} + 4i(2At + B)e^{i2t} - 4t(At + B)e^{i2t}$ so that inserting these into the equation we obtain

$$2Ae^{i2t} + 4i(2At + B)e^{i2t} - 4(At^2 + Bt)e^{i2t} + 4t(At + B)e^{i2t} = (16t + 4i)e^{i2t}$$

which gives $2A + 4iB = 4i$ and $8iA = 16$. Thus $A = -2i$ and $B = 2$. Similarly, $C = 2i$ and $D = 2$ and we obtain the particular solution in the form

$$y(t) = t(-2it + 2)e^{i2t} + t(2it + 2)e^{-i2t} = 4t^2 \sin 2t + 4t \cos t,$$

where we used Euler's formula to convert exponential into trigonometric functions once again.

3.2.3 An application

Second order equations appear in many applications involving oscillations occurring due to the existence of an elastic force in the system. The reason for this is that the elastic force, at least for small displacements, is proportional to the displacement so that according to Newton's second law

$$my'' = -ky$$

where k is a constant. In general, there is a damping force (due to the resistance of the medium) and some external force, and then the full equation for oscillation reads

$$y'' + cy' + ky = F(t). \tag{3.2.14}$$

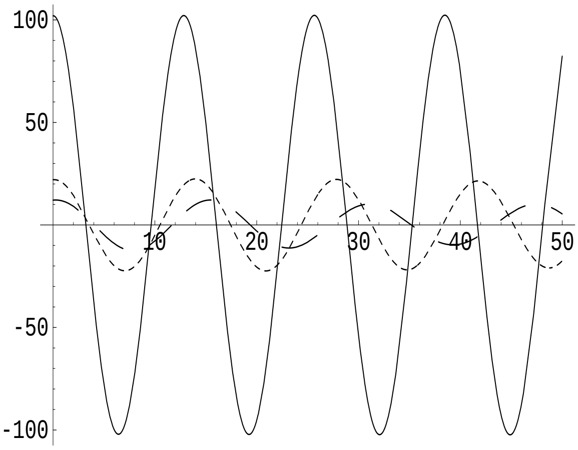

Fig 2.8 Forced free vibrations: non-resonance case with $\omega_0 = 0.5$ *and* $\omega = 0.4$ *(dashed line),* $\omega = 0.45$ *(dotted line) and* $\omega = 0.49$ *(solid line). Note the increase in amplitude of vibrations as the frequency of the external force approaches the natural frequency of the system*

We shall discuss in detail a particular example of this equation describing the so-called *forced free vibrations.* In this case we have

$$y'' + \omega_0^2 y = \frac{F_0}{m}\cos\omega t, \tag{3.2.15}$$

where we denoted $\omega_0^2 = k/m$ and introduced a special periodic force $F(t) = F_0 \cos\omega t$ with constant magnitude F_0 and period ω.

The characteristic equation is $\lambda^2 + \omega_0^2 = 0$ so that we have imaginary roots $\lambda_{1,2} = \pm i\omega_0$ and the general solution to the homogeneous equations is given by

$$y_0(t) = C_1 \cos\omega_0 t + C_2 \sin\omega_0 t.$$

The frequency ω_0 is called the natural frequency of the system. The case $\omega \neq \omega_0$ gives a particular solution in the form

$$y_p(t) = \frac{F_0}{m(\omega_0^2 - \omega^2)}\cos\omega t$$

so that the general solution is given by

$$y(t) = C_1 \cos\omega_0 t + C_2 \sin\omega_0 t + \frac{F_0}{m(\omega_0^2 - \omega^2)}\cos\omega t, \tag{3.2.16}$$

that is the solution is obtained as a sum of two periodic motions, as shown in Figure 2.8. Though there is nothing unusual here, we can sense that a trouble is brewing – if the the natural frequency of the system is close to

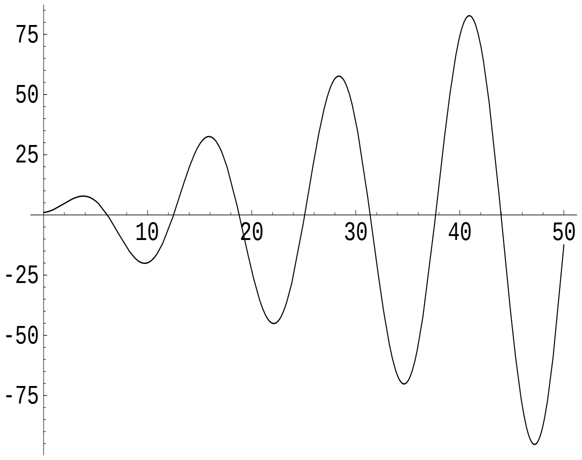

Fig 2.9 Forced free vibrations: the resonance case. The amplitude of vibrations increases to infinity.

the frequency of the external force, then the amplitude of vibrations can become very large, because the denominator in the last term in (3.2.16) is very small. Let us find out what happens if $\omega_0 = \omega$. In this case we convert $F(t) = F_0 \cos \omega t = F_0 \cos \omega_0 t = F_0(e^{i\omega_0 t} + e^{-i\omega_0 t})/2$ and look for the particular solution in the form $y_p(t) = t(Ae^{i\omega_0 t} + Be^{-i\omega_0 t})$. We obtain $y_p'(t) = Ae^{i\omega_0 t} + Be^{-i\omega_0 t} + ti\omega_0(Ae^{i\omega_0 t} - Be^{-i\omega_0 t})$ and

$$y_p''(t) = i2(Ae^{i\omega_0 t} - Be^{-i\omega_0 t}) - t\omega_0^2(Ae^{i\omega_0 t} + Be^{-i\omega_0 t}).$$

Inserting these into the equation and comparing coefficients we find that $A = B = -F_0/4i\omega_0$ so that

$$y_p(t) = \frac{F_0}{2\omega_0}\frac{e^{i\omega_0 t} - e^{-i\omega_0 t}}{2i} = \frac{F_0}{2\omega_0} t \sin \omega_0 t$$

and the general solution is given by

$$y(t) = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t + \frac{F_0}{2\omega_0} t \sin \omega_0 t.$$

A graph of such a function is shown in Figure 2.9. The important point of this example is that even small force can induce very large oscillations in the system if its frequency is equal or even only very close to the natural frequency of the system. This phenomenon is called the *resonance* and is responsible for a number of spectacular collapses of constructions, like the collapse of Tacoma Bridge in the USA (oscillations induced by wind) and Broughton suspension bridge in England (oscillations introduced by soldiers marching in cadence).

Chapter 4

Difference equations: models and solutions

4.1 First order difference equations

The general form of a first order difference equation is

$$x(n+1) = f(n, x(n)), \tag{4.1.1}$$

where f is any function of two variables defined on $\mathbb{N}_0 \times \mathbb{R}$, where $\mathbb{N}_0 = \{0, 1, 2 \ldots\}$ is the set of natural numbers enlarged by 0. In this chapter we will be concerned only with linear difference equations and systems of them, where $x(n)$ is a vector and f is a matrix.

4.1.1 Methods of solution

The simplest difference equations are these defining geometric and arithmetic progressions:

$$x(n+1) = ax(n),$$

and

$$y(n+1) = y(n) + a,$$

respectively, where a is a constant. The solutions of these equations are known to be

$$x(n) = a^n x(0),$$

and

$$y(n) = y(0) + na.$$

We shall consider the generalization of both these equations: the general first order difference equation,

$$x(n+1) = a(n)x(n) + g(n) \tag{4.1.2}$$

with the an initial condition $x(0) = x_0$. Calculating first few iterates, we obtain

$$\begin{aligned}
x(1) &= a(0)x(0) + g(0),\\
x(2) &= a(1)x(1) + g(1) = a(1)a(0)x(0) + a(1)g(0) + g(1),\\
x(3) &= a(2)x(2) + g(2) = a(2)a(1)a(0)x(0) + a(2)a(1)g(0) + a(2)g(1) + g(2),\\
x(4) &= a(3)x(3) + g(3)\\
&= a(3)a(2)a(1)a(0)x(0) + a(3)a(2)a(1)g(0) + a(3)a(2)g(1) + a(3)g(2) + g(3).
\end{aligned}$$

At this moment we have enough evidence to conjecture that the general form of the solution could be

$$x(n) = x(0)\prod_{k=0}^{n-1} a(k) + \sum_{k=0}^{n-1} g(k) \prod_{i=k+1}^{n-1} a(i) \tag{4.1.3}$$

where we adopted the convention that $\prod\limits_{n}^{n-1} = 1$. Similarly, to simplify notation, we agree to put $\sum\limits_{k=j+1}^{j} = 0$. To fully justify this formula, we shall use mathematical induction. Constructing (4.1.3) we have checked that the formula holds for a few initial values of the argument. Assume now that it is valid for n and consider

$$\begin{aligned}
x(n+1) &= a(n)x(n) + g(n)\\
&= a(n)\left(x(0)\prod_{k=0}^{n-1} a(k) + \sum_{k=0}^{n-1} g(k) \prod_{i=k+1}^{n-1} a(i)\right) + g(n)\\
&= x(0)\prod_{k=0}^{n} a(k) + a(n)\sum_{k=0}^{n-1} g(k) \prod_{i=k+1}^{n-1} a(i) + g(n)\\
&= x(0)\prod_{k=0}^{n} a(k) + \sum_{k=0}^{n-1} g(k) \prod_{i=k+1}^{n} a(i) + g(n) \prod_{i=n+1}^{n} a(i)\\
&= x(0)\prod_{k=0}^{n} a(k) + \sum_{k=0}^{n} g(k) \prod_{i=k+1}^{n} a(i)
\end{aligned}$$

which proves that (4.1.3) is valid for all $n \in \mathbb{N}$.

Two special cases

There are two special cases of (4.1.2) that appear in many applications. In the first, the equation is given by

$$x(n) = ax(n) + g(n), \tag{4.1.4}$$

with the value $x(0)$ given. In this case $\prod\limits_{k=k_1}^{k_2} a(k) = a^{k_2-k_1+1}$ and (4.1.3) takes the form

$$x(n) = a^n x(0) + \sum_{k=0}^{n-1} a^{n-k-1} g(k). \tag{4.1.5}$$

The second case is a simpler form of (4.1.4), given by

$$x(n) = ax(n) + g, \tag{4.1.6}$$

with g independent of n. In this case the sum in (4.1.5) can be evaluated in an explicit form giving

$$x(n) = \begin{cases} a^n x(0) + g\frac{a^n-1}{a-1} & \text{if} \quad a \neq 1, \\ x(0) + gn. \end{cases} \tag{4.1.7}$$

Example 4.1.1. Assume that a dose D_0 of a drug, that increases it's concentration in the patient's body by c_0, is administered at regular time intervals $t = 0, T, 2T, 3T \ldots$. Between the injections the concentration c of the drug decreases according to the differential equation $c' = -\gamma c$, where γ is a positive constant. It is convenient here to change slightly the notational convention and denote by c_n the concentration of the drug just after the nth injection, that is, c_0 is the concentration just after the initial (zeroth) injection, c_1 is the concentration just after the first injection, that is, at the time T, etc. We are to find formula for c_n and determine whether the concentration of the drug eventually stabilizes.

In this example we have a combination of two processes: continuous between the injections and discrete in injection times. Firstly, we observe that the process is discontinuous at injection times so we have two different values for $c(nT)$: just before the injection and just after the injection (assuming that the injection is done instantaneously). To avoid ambiguities, we denote by $c(nT)$ the concentration just before the nth injection and by c_n the concentration just after, in accordance with the notation introduced above. Thus, between the nth and $n+1$st injection the concentration changes according to the exponential law

$$c((n+1)T) = c_n e^{-\gamma T}$$

so that over each time interval between injection the concentration decreases by a constant fraction $a = e^{-\gamma T} < 1$. Thus, we are able to write down the difference equation for concentrations just after $n+1$st injection as

$$c_{n+1} = ac_n + c_0, \tag{4.1.8}$$

where c_0 is the dose of each injection. We can write down the solution using (4.1.7) as

$$c_n = c_0 a^n + c_0 \frac{a^n - 1}{a - 1} = -\frac{c_0}{1-a} a^{n+1} + \frac{c_0}{1-a}.$$

Fig 3.1 Long time behaviour of the concentration $c(t)$.

Since $a < 1$, we immediately obtain that $\bar{c} = \lim\limits_{n\to\infty} c_n = \frac{c_0}{1-a} = \frac{c_0}{1-e^{-\gamma T}}$.

Similarly, the concentration just before the nth injection is

$$\begin{aligned} c(nT) &= c_{n-1}e^{-\gamma T} = e^{-\gamma T}\left(\frac{c_0}{e^{-\gamma T}-1}e^{-\gamma T n} + \frac{c_0}{1-e^{-\gamma T}}\right) \\ &= \frac{c_0}{1-e^{\gamma T}}e^{-\gamma T n} + \frac{c_0}{e^{\gamma T}-1} \end{aligned}$$

and for the long run $\underline{c} = \lim\limits_{n\to\infty} c(nT) = \frac{c_0}{e^{\gamma T}-1}$.

For example, using $c_0 = 14$ mg/l, $\gamma = 1/6$ and $T = 6$ hours we obtain that after a long series of injections the maximal concentration, attained immediately after injections, will stabilize at around 22 mg/l. The minimal concentration, just before injection, will stabilize at around $\underline{c} = 14/e - 1 \approx 8.14$ mg/l. This effect is illustrated at Fig. 3.1.

Example 4.1.2. In Subsection 1.2.1 we discussed the difference equation governing long-term loan repayment:

$$D(k+1) = D(k) + \frac{\alpha p}{100}D(k) - R = D(k)\left(1 + \frac{\alpha p}{100}\right) - R, \tag{4.1.9}$$

where D_0 is the initial debt to be repaid, for each k, $D(k)$ is the outstanding debt after the kth repayment, the payment made after each conversion period is R, α% is the annual interest rate and p is the conversion period, that is, the number of payments in one year. To simplify notation we denote $r = \alpha p/100$

Using again (4.1.7) we obtain the solution

$$\begin{aligned} D(k) &= (1+r)^k D_0 - R\sum_{i=0}^{k-1}(1+r)^{k-i-1} \\ &= (1+r)^k D_0 - \left((1+r)^k - 1\right)\frac{R}{r} \end{aligned}$$

This equation gives answers to a number of questions relevant in taking a loan. For example, if we want to know what will be the monthly instalment on a loan of D_0 to be repaid in n payments, we observe that the loan is repaid in n instalments if $D(n) = 0$, thus we must solve

$$0 = (1+r)^n D_0 - ((1+r)^n - 1)\frac{R}{r}$$

in R, which gives

$$R = \frac{rD_0}{1-(1+r)^{-n}}.$$

For example, taking a mortgage of R200000 to be repaid over 20 years in monthly instalments at the annual interest rate of 13% we obtain $\alpha = 1/12$, hence $r = 0.0108$, and $n = 20 \times 12 = 240$. Therefore

$$R = \frac{0.0108 \cdot 200000}{1 - 1.0108^{-240}} \approx R2167.$$

4.2 Systems of difference equations and higher order equations

4.2.1 Homogeneous systems of difference equations

We start with the homogeneous system of difference equations

$$\begin{aligned} y_1(n+1) &= a_{11}y_1(n) + a_{12}y_2(n) + \ldots + a_{1k}y_k(n), \\ \vdots \;&\vdots\; \vdots, \\ y_k(n+1) &= a_{k1}y_1(n) + a_{k2}y_2(n) + \ldots + a_{kk}y_k(n), \end{aligned} \tag{4.2.1}$$

where, for $n \geq 0$, $y_1(n), \ldots, y_k(n)$ are unknown sequences, $a_{11}, \ldots, a_{kk}$ are constant coefficients and $g_1(n)\ldots, g_k(n)$ are known. As with systems of differential equations, we shall find it more convenient to use the matrix notation. Denoting $\mathbf{y} = (y_1, \ldots, y_k)$, $\mathcal{A} = \{a_{ij}\}_{1\leq i,j\leq k}$, that is,

$$\mathcal{A} = \begin{pmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{pmatrix},$$

(4.2.1) can be written as

$$\mathbf{y}(n+1) = \mathcal{A}\mathbf{y}(n). \tag{4.2.2}$$

Eq. (4.2.2) is usually supplemented by the initial condition $\mathbf{y}(0) = \mathbf{y^0}$. By induction, to see that the solution to (4.2.2) is given by

$$\mathbf{y}(n) = \mathcal{A}^n\mathbf{y^0}. \tag{4.2.3}$$

The problem with (4.2.3) is that it is rather difficult to give an explicit form of $\mathcal{A}^n$. We shall solve this problem in a way similar to Subsection 3.1.4, where we were to evaluate the exponential function of $\mathcal{A}$.

To proceed, we assume that the matrix $\mathcal{A}$ is nonsingular. This means, in particular, that if $\mathbf{v^1}, \ldots, \mathbf{v^k}$ are linearly independent vectors, then also $\mathcal{A}\mathbf{v^1}, \ldots, \mathcal{A}\mathbf{v^k}$ are linearly independent. Since $\mathbb{R}^k$ is k-dimensional, it is enough to find k linearly independent vectors $\mathbf{v^i}$, $i = 1, \ldots, k$ for which $\mathcal{A}^n\mathbf{v^i}$ can be easily evaluated. Assume for a moment that such vectors have been found. Then, for arbitrary $\mathbf{x^0} \in \mathbb{R}^k$ we can find constants $c_1, \ldots, c_k$ such that

$$\mathbf{x^0} = c_1\mathbf{v^1} + \ldots + c_2\mathbf{v^k}.$$

Precisely, let $\mathcal{V}$ be the matrix having vectors $\mathbf{v^i}$ as its columns, and let $\mathbf{c} = (c_1, \ldots, c_k)$, then

$$\mathbf{c} = \mathcal{V}^{-1}\mathbf{x^0}. \tag{4.2.4}$$

Note, that $\mathcal{V}$ is invertible as the vectors $\mathbf{v^i}$ are linearly independent.

Thus, for an arbitrary $\mathbf{x^0}$ we have

$$\mathcal{A}^n\mathbf{x^0} = \mathcal{A}^n(c_1\mathbf{v^1} + \ldots + c_2\mathbf{v^k}) = c_1\mathcal{A}^n\mathbf{v^1} + \ldots + c_k\mathcal{A}^n\mathbf{v^k}. \tag{4.2.5}$$

Now, if we denote by $\mathcal{A}_n$ the matrix whose columns are vectors $\mathcal{A}^n\mathbf{v^1}, \ldots, \mathcal{A}^n\mathbf{v^k}$, then we can write

$$\mathcal{A}^n = \mathcal{A}_n\mathcal{V}^{-1} \tag{4.2.6}$$

Hence, the problem is to find k linearly independent vectors $\mathbf{v^i}$, $i = 1, \ldots, k$, on which powers of $\mathcal{A}$ can be easily evaluated. As before, we use eigenvalues and eigenvectors for this purpose. Firstly, note that if $\mathbf{v^1}$ is an eigenvector of $\mathcal{A}$ corresponding to an eigenvalue λ_1, that is, $\mathcal{A}\mathbf{v^1} = \lambda_1\mathbf{v^1}$, then by induction

$$\mathcal{A}^n\mathbf{v^1} = \lambda_1^n\mathbf{v^1}.$$

Therefore, if we have k linearly independent eigenvectors $\mathbf{v^1}, \ldots, \mathbf{v^k}$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_k$ (not necessarily distinct), then from (4.2.5) we obtain

$$\mathcal{A}^n\mathbf{x^0} = c_1\lambda_1^n\mathbf{v^1} + \ldots + c_k\lambda_k^n\mathbf{v^k}.$$

with $c_1, \ldots, c_k$ given by (4.2.4). Note that, as for systems of differential equations, if λ is a complex eigenvalue with eigenvector $\mathbf{v}$, then both $\Re(\lambda^n\mathbf{v})$

and $\Im(\lambda^n\mathbf{v})$ are real valued solutions. To find explicit expressions for them we write $\lambda = re^{i\phi}$ where $r = |\lambda|$ and $\phi = Arg\lambda$. Then

$$\lambda^n = r^n e^{in\phi} = r^n(\cos n\phi + i\sin n\phi)$$

and

$$\begin{aligned}\Re(\lambda^n\mathbf{v}) &= r^n(\cos n\phi\Re\mathbf{v} - \sin n\phi\Im\mathbf{v}),\\ \Im(\lambda^n\mathbf{v}) &= r^n(\sin n\phi\Re\mathbf{v} + \cos n\phi\Im\mathbf{v}).\end{aligned}$$

Finally, if for some eigenvalue λ_i the number ν_i of linearly independent eigenvectors is smaller than its algebraic multiplicity n_i, then we follow the procedure described in Subsection 3.1.4, that is, we find all solutions to

$$(\mathcal{A} - \lambda_i\mathcal{I})^2\mathbf{v} = \mathbf{0}$$

that are not eigenvectors and, if we still do not have sufficiently many independent vectors, we continue solving

$$(\mathcal{A} - \lambda_i\mathcal{I})^j\mathbf{v} = \mathbf{0}$$

with $j \leq n_i$; it can be proven that in this way we find n_i linearly independent vectors. Let $\mathbf{v^j}$ is found as a solution to $(\mathcal{A}-\lambda_i\mathcal{I})^j\mathbf{v^j} = \mathbf{0}$ with $j \leq n_i$. Then, using the binomial expansion we find

$$\begin{aligned}\mathcal{A}^n\mathbf{v^j} &= (\lambda_i\mathcal{I} + \mathcal{A} - \lambda_i\mathcal{I})^n\mathbf{v^j} = \sum_{r=0}^{n}\lambda_i^{n-r}\begin{pmatrix} n\\ r\end{pmatrix}(\mathcal{A} - \lambda_i\mathcal{I})^r\mathbf{v^j}\\ &= \Big(\lambda_i^n\mathcal{I} + n\lambda_i^{n-1}(\mathcal{A} - \lambda_i\mathcal{I}) + \ldots\\ &\quad + \frac{n!}{(j-1)!(n-j+1)!}\lambda_i^{n-j+1}(\mathcal{A} - \lambda_i\mathcal{I})^{j-1}\Big)\mathbf{v^j}, \qquad (4.2.7)\end{aligned}$$

where

$$\begin{pmatrix} n\\ r\end{pmatrix} = \frac{n!}{r!(n-r)!}$$

is the Newton symbol. It is important to note that (4.2.7) is a finite sum for any n; it always terminates at most at the term $(\mathcal{A} - \lambda_1\mathcal{I})^{n_i-1}$ where n_i is the algebraic multiplicity of λ_i.

We shall illustrate these considerations by the following example.

Example 4.2.1. Find $\mathcal{A}^n$ for

$$\mathcal{A} = \begin{pmatrix} 4 & 1 & 2\\ 0 & 2 & -4\\ 0 & 1 & 6\end{pmatrix}.$$

We start with finding eigenvalues of $\mathcal{A}$:

$$p(\lambda)=\begin{vmatrix}4-\lambda & 1 & 2\\ 0 & 2-\lambda & -4\\ 0 & 1 & 6-\lambda\end{vmatrix}=(4-\lambda)(16-8\lambda+\lambda^2)=(4-\lambda)^3=0$$

gives the eigenvalue $\lambda=4$ of algebraic multiplicity 3. To find eigenvectors corresponding to $\lambda=3$, we solve

$$(\mathcal{A}-4\mathcal{I})\mathbf{v}=\begin{pmatrix}0 & 1 & 2\\ 0 & -2 & -4\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}.$$

Thus, v_1 is arbitrary and $v_2=-2v_3$ so that the eigenspace is two dimensional, spanned by

$$\mathbf{v^1}=\begin{pmatrix}1\\ 0\\ 0\end{pmatrix},\qquad \mathbf{v^2}=\begin{pmatrix}0\\ -2\\ 1\end{pmatrix}.$$

Therefore

$$\mathcal{A}^n\mathbf{v^1}=4^n\begin{pmatrix}1\\ 0\\ 0\end{pmatrix},\qquad \mathcal{A}^n\mathbf{v^2}=4^n\begin{pmatrix}0\\ -2\\ 1\end{pmatrix}.$$

To find the last vector we consider

$$\begin{aligned}(\mathcal{A}-4\mathcal{I})^2\mathbf{v} &= \begin{pmatrix}0 & 1 & 2\\ 0 & -2 & -4\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}0 & 1 & 2\\ 0 & -2 & -4\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix}\\ &= \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}.\end{aligned}$$

Any vector solves this equation so that we have to take a vector that is not an eigenvalue. Possibly the simplest choice is

$$\mathbf{v^3}=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}.$$

Thus, by (4.2.7)

$$\begin{aligned}\mathcal{A}^n\mathbf{v^3} &= \left(4^n\mathcal{I}+n4^{n-1}(\mathcal{A}-4\mathcal{I})\right)\mathbf{v^3}\\ &= \left(\begin{pmatrix}4^n & 0 & 0\\ 0 & 4^n & 0\\ 0 & 0 & 4^n\end{pmatrix}+n4^{n-1}\begin{pmatrix}0 & 1 & 2\\ 0 & -2 & -4\\ 0 & 1 & 2\end{pmatrix}\right)\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}\\ &= \begin{pmatrix}2n4^{n-1}\\ -n4^n\\ 4^n+2n4^{n-1}\end{pmatrix}.\end{aligned}$$

To find explicit expression for $\mathcal{A}^n$ we use (4.2.6). In our case

$$\mathcal{A}_n = \begin{pmatrix} 4^n & 0 & 2n4^{n-1} \\ 0 & -2\cdot 4^n & -n4^n \\ 0 & 4^n & 4^n + 2n4^{n-1} \end{pmatrix},$$

further

$$\mathcal{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

so that

$$\mathcal{V}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix}.$$

Therefore

$$\mathcal{A}^n = \mathcal{A}_n\mathcal{V}^{-1} = \begin{pmatrix} 4^n & n4^{n-1} & 2n4^{n-1} \\ 0 & 4^n - 2n4^{n-1} & -n4^n \\ 0 & n4^{n-1} & 4^n + 2n4^{n-1} \end{pmatrix}.$$

The next example shows how to deal with complex eigenvalues.

Example 4.2.2. Find $\mathcal{A}^n$ if

$$\mathcal{A} = \begin{pmatrix} 1 & -5 \\ 1 & -1 \end{pmatrix}.$$

We have

$$\begin{vmatrix} 1-\lambda & -5 \\ 1 & -1-\lambda \end{vmatrix} = \lambda^2 + 4$$

so that $\lambda_{1,2} = \pm 2i$. Taking $\lambda_1 = 2i$, we find the corresponding eigenvector by solving

$$\begin{pmatrix} 1-2i & -5 \\ 1 & -1-2i \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

we get

$$\mathbf{v^1} = \begin{pmatrix} 1+2i \\ 1 \end{pmatrix}$$

and

$$\mathbf{x}(n) = \mathcal{A}^n\mathbf{v^1} = (2i)^n\begin{pmatrix} 1+2i \\ 1 \end{pmatrix}.$$

To find real valued solutions, we have to take real and imaginary parts of $\mathbf{x}(n)$. Since $i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}$, we have by de Moivre's formula

$$(2i)^n = 2^n\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right)^n = 2^n\left(\cos\frac{n\pi}{2} + i\sin\frac{n\pi}{2}\right).$$

Therefore

$$\begin{aligned}\Re\mathbf{x}(n) &= 2^n\left(\cos\frac{n\pi}{2}\begin{pmatrix}1\\1\end{pmatrix} - \sin\frac{n\pi}{2}\begin{pmatrix}2\\0\end{pmatrix}\right)\\ \Im\mathbf{x}(n) &= 2^n\left(\cos\frac{n\pi}{2}\begin{pmatrix}2\\0\end{pmatrix} + \sin\frac{n\pi}{2}\begin{pmatrix}1\\1\end{pmatrix}\right).\end{aligned}$$

The initial values for $\Re\mathbf{x}(n)$ and $\Im\mathbf{x}(n)$ are, respectively, $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}2\\0\end{pmatrix}$. Since $\mathcal{A}^n$ is a real matrix, we have $\Re\mathcal{A}^n\mathbf{v^1} = \mathcal{A}^n\Re\mathbf{v^1}$ and $\Im\mathcal{A}^n\mathbf{v^1} = \mathcal{A}^n\Im\mathbf{v^1}$, thus

$$\mathcal{A}^n\begin{pmatrix}1\\1\end{pmatrix} = 2^n\left(\cos\frac{n\pi}{2}\begin{pmatrix}1\\1\end{pmatrix} - \sin\frac{n\pi}{2}\begin{pmatrix}2\\0\end{pmatrix}\right) = 2^n\begin{pmatrix}\cos\frac{n\pi}{2} - 2\sin\frac{n\pi}{2}\\ \cos\frac{n\pi}{2}\end{pmatrix}$$

and

$$\mathcal{A}^n\begin{pmatrix}2\\0\end{pmatrix} = 2^n\left(\cos\frac{n\pi}{2}\begin{pmatrix}2\\0\end{pmatrix} + \sin\frac{n\pi}{2}\begin{pmatrix}1\\1\end{pmatrix}\right) = 2^n\begin{pmatrix}2\cos\frac{n\pi}{2} + \sin\frac{n\pi}{2}\\ \sin\frac{n\pi}{2}\end{pmatrix}.$$

To find $\mathcal{A}^n$ we use again (4.2.6). In our case

$$\mathcal{A}_n = 2^n\begin{pmatrix}\cos\frac{n\pi}{2} - 2\sin\frac{n\pi}{2} & 2\cos\frac{n\pi}{2} + \sin\frac{n\pi}{2}\\ \cos\frac{n\pi}{2} & \sin\frac{n\pi}{2}\end{pmatrix},$$

further

$$\mathcal{V} = \begin{pmatrix}1 & 2\\ 1 & 0\end{pmatrix},$$

so that

$$\mathcal{V}^{-1} = -\frac{1}{2}\begin{pmatrix}0 & -2\\ -1 & 1\end{pmatrix}.$$

Therefore

$$\mathcal{A}^n = \mathcal{A}_n\mathcal{V}^{-1} = -2^{n-1}\begin{pmatrix}-2\cos\frac{n\pi}{2} - \sin\frac{n\pi}{2} & 5\sin\frac{n\pi}{2}\\ -\sin\frac{n\pi}{2} & -2\cos\frac{n\pi}{2} + \sin\frac{n\pi}{2}\end{pmatrix}.$$

4.2.2 Nonhomogeneous systems

Here we shall discuss solvability of the nonhomogeneous version of (4.2.1)

$$\begin{aligned}y_1(n+1) &= a_{11}y_1(n) + a_{12}y_2(n) + \ldots + a_{1k}y_k(n) + g_1(n),\\ \vdots \quad &\vdots \quad \vdots,\\ y_k(n+1) &= a_{k1}y_1(n) + a_{k2}y_2(n) + \ldots + a_{kk}y_k(n) + g_k(n),\end{aligned} \tag{4.2.8}$$

where, for $n \geq 0$, $y_1(n), \ldots, y_k(n)$ are unknown sequences, $a_{11}, \ldots a_{kk}$ are constant coefficients and $g_1(t) \ldots, g_k(n)$ are known. As before, we write it using the vector-matrix notation. Denoting $\mathbf{y} = (y_1, \ldots, y_k)$, $\mathbf{g} = (g_1, \ldots, g_k)$ and $\mathcal{A} = \{a_{ij}\}_{1\leq i,j\leq k}$, we have

$$\mathbf{y}(n+1) = \mathcal{A}\mathbf{y}(n) + \mathbf{g}(n). \tag{4.2.9}$$

Exactly as in Subsection 4.1.1 we find that the solution to (4.2.9) satisfying the initial condition $\mathbf{y}(0) = \mathbf{y}^0$ is given by the formula

$$\mathbf{y}(n) = \mathcal{A}^n\mathbf{y}^0 + \sum_{r=0}^{n-1} \mathcal{A}^{n-r-1}\mathbf{g}(r). \tag{4.2.10}$$

Example 4.2.3. Solve the system

$$\begin{array}{rcr} y_1(n+1) & = & 2y_1(n) + y_2(n) + n, \\ y_2(n+1) & = & 2y_2(n) + 1 \end{array}$$

with $y_1(0) = 1, y_2(0) = 0$. Here

$$\mathcal{A} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{g}(n) = \begin{pmatrix} n \\ 1 \end{pmatrix}, \qquad \mathbf{y}^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

We see that

$$p(\lambda) = \begin{vmatrix} 2-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (2-\lambda)^2,$$

so that we have double eigenvalue $\lambda = 2$. To find eigenvectors corresponding to this eigenvalue, we have to solve the system

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so that we have one-dimensional eigenspace spanned by $\mathbf{v}^1 = (1,0)$. To find the second linearly independent vector associated with $\lambda = 2$ we observe that

$$(\mathcal{A} - 2\mathcal{I})^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

so that we can take $\mathbf{v}^2 = (0,1)$. Thus, we obtain two independent solutions in the form

$$\mathbf{y}^1(n) = \mathcal{A}^n\mathbf{v}^1 = 2^n\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{y}^2(n) &= \mathcal{A}^n\mathbf{v}^2 = (2\mathcal{I} + (\mathcal{A} - 2\mathcal{I})^n\mathbf{v}^2 = \left(2^n\mathcal{I} + n2^{n-1}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right)\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} n2^{n-1} \\ 2^n \end{pmatrix}. \end{aligned}$$

Since $\mathbf{v}^1$ and $\mathbf{v}^2$ happen to be the canonical basis for $\mathbb{R}^2$, that is, $\mathbf{x}^0 = (x_1^0, x_2^0) = x_1^0\mathbf{v}^1 + x_2^0\mathbf{v}^1$, we obtain immediately

$$\mathcal{A}^n = \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix}.$$

To find the solution of the nonhomogeneous equation, we use formula (4.2.10). The first term is easily calculated as

$$\mathcal{A}^n\mathbf{x}^0 = \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2^n \\ 0 \end{pmatrix}.$$

Next,

$$\begin{aligned}
\sum_{r=0}^{n-1}\mathcal{A}^{n-r-1}\mathbf{g}(r) &= \sum_{r=0}^{n-1}\begin{pmatrix} 2^{n-r-1} & (n-r-1)2^{n-r-2} \\ 0 & 2^{n-r-1} \end{pmatrix}\begin{pmatrix} r \\ 1 \end{pmatrix} \\
&= \sum_{r=0}^{n-1}\begin{pmatrix} r2^{n-r-1} + (n-r-1)2^{n-r-2} \\ 2^{n-r-1} \end{pmatrix} \\
&= 2^n \begin{pmatrix} \frac{1}{4}\sum\limits_{r=1}^{n-1} r2^{-r} + \frac{n-1}{4}\sum\limits_{r=0}^{n-1} 2^{-r} \\ \frac{1}{2}\sum\limits_{r=0}^{n-1} 2^{-r} \end{pmatrix} \\
&= 2^n \begin{pmatrix} \frac{1}{2}\left(1-\left(\frac{1}{2}\right)^{n-1}\right) - (n-1)\left(\frac{1}{2}\right)^{n+1} + \frac{n-1}{2}\left(1-\left(\frac{1}{2}\right)^{n}\right) \\ 1-\left(\frac{1}{2}\right)^n \end{pmatrix} \\
&= 2^n \begin{pmatrix} -n\left(\frac{1}{2}\right)^n + \frac{n}{2} \\ 1-\left(\frac{1}{2}\right)^n \end{pmatrix} \\
&= \begin{pmatrix} -n + \frac{n2^n}{2} \\ 2^n - 1 \end{pmatrix}
\end{aligned}$$

Remark 4.2.4. Above we used the following calculations

$$\begin{aligned}
\sum_{r=1}^{n-1} ra^r &= a(1+a+\ldots+a^{n-2}) + a^2(1+a+\ldots+a^{n-3}) + \ldots + a^{n-1} \\
&= \frac{1}{1-a}\left(a(1-a^{n-1}) + a^2(1-a^{n-2}) + \ldots + a^{n-1}(1-a)\right) \\
&= \frac{1}{1-a}\left(a + a^2 + \ldots + a^{n-1} - (n-1)a^n\right) \\
&= \frac{a(1-a^{n-1}) - (n-1)a^n(1-a)}{(1-a)^2}
\end{aligned}$$

Thus, the solution is given by

$$\mathbf{y}(n) = \begin{pmatrix} 2^n - n + \frac{n2^n}{2} \\ 2^n - 1 \end{pmatrix}.$$

4.2.3 Higher order equations

Consider the linear difference equation of order k:

$$y(n+k) + a_1 y(n+k-1) + \ldots + a_k y(n) = g(n), \qquad n \geq 0 \tag{4.2.11}$$

where $a_1, \ldots, a_k$ are known numbers and $g(n)$ is a known sequence. This equation determines the values of $y(N)$, $N > k$ by k preceding values of $y(r)$. Thus, it is clear that to be able to solve this equation, that is, to start the recurrence procedure, we need k initial values $y(0), y(1), \ldots, y(k-1)$. Equation (4.2.11) can be written as a system of first order equations of dimension k. We let

$$\begin{array}{rcl} z_1(n) & = & y(n), \\ z_2(n) & = & y(n+1) = z_1(n+1), \\ z_3(n) & = & y(n+2) = z_2(n+1), \\ \vdots & \vdots & \vdots, \\ z_k(n) & = & y(n+k-1) = z_{k-1}(n-1), \end{array} \tag{4.2.12}$$

hence we obtain the system

$$\begin{array}{rcl} z_1(n+1) & = & z_2(n), \\ z_2(n+1) & = & z_3(n), \\ \vdots & \vdots & \vdots, \\ z_{k-1}(n+1) & = & z_k(n), \\ z_k(n+1) & = & -a_1 z_1(n) - a_2 z_2(n) \ldots - a_k z_k(n) + g(n), \end{array}$$

or, in matrix notation,

$$\mathbf{z}(n+1) = \mathcal{A}\mathbf{z}(n) + \mathbf{g}(n)$$

where $\mathbf{z} = (z_1, \ldots, z_k)$, $\mathbf{g}(n) = (0, 0, \ldots, g(n))$ and

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_k & -a_{k-1} & -a_{k-2} & \ldots & -a_1 \end{pmatrix}.$$

It is clear that the initial values $y(0), \ldots, y(k-1)$ give the initial vector $\mathbf{z}^0 = (y(0), \ldots, y(k-1))$. Next we observe that the eigenvalues of $\mathcal{A}$ can be obtained by solving the equation

$$\begin{vmatrix} -\lambda & 1 & 0 & \ldots & 0 \\ 0 & -\lambda & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_k & -a_{k-1} & -a_{k-2} & \ldots & -a_1 - \lambda \end{vmatrix}$$
$$= (-1)^k (\lambda^k + a_1 \lambda^{k-1} + \ldots + a_k) = 0,$$

that is, the eigenvalues can be obtained by finding roots of the characteristic polynomial. Consequently, solutions of higher order equations can be obtained by solving the associated first order systems but there is no need to repeat the whole procedure. In fact, to solve a $k \times k$ system we have to construct k linearly independent vectors $\mathbf{v}^1, \ldots, \mathbf{v}^k$ so that solutions are given by $\mathbf{z}^1(n) = \mathcal{A}^n\mathbf{v}^1, \ldots \mathbf{z}^k(n) = \mathcal{A}^n\mathbf{v}^k$ and coordinates of each $\mathbf{z}^\mathbf{i}$ are products of λ_i and polynomials in n of degree strictly smaller than the algebraic multiplicity of λ_i. To obtain n_i solutions of the higher order equation corresponding to the eigenvalue λ_i, by (4.2.12) we take only the first coordinates of all $\mathbf{z}^i(n)$ that correspond to λ_i. On the other hand, we must have here n_i linearly independent scalar solutions of this form and therefore we can use the set $\{\lambda_i^n, n\lambda_i^n, \ldots, n^{n_i-1}\lambda_i^n\}$ as a basis for the set of solutions corresponding to λ_i, and the union of such sets over all eigenvalues to obtain a basis for the set of all solutions.

Example 4.2.5. Consider the Fibonacci equation (1.2.6), written here as

$$y(n+2) = y(n+1) + y(n) \tag{4.2.13}$$

to be consistent with the notation of the present chapter. Introducing new variables $z_1(n) = y(n), z_2(n) = y(n+1) = z_1(n+1)$ so that $y(n+2) = z_2(n+1)$, we re-write the equation as the system

$$\begin{aligned} z_1(n+1) &= z_2(n), \\ z_2(n+1) &= z_1(n) + z_2(n). \end{aligned}$$

The eigenvalues of the matrix

$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

are obtained by solving the equation

$$\begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = \lambda^2 - \lambda - 1 = 0;$$

they are $\lambda_{1,2} = \frac{1\pm\sqrt{5}}{2}$. Since the eigenvalues are distinct, we immediately obtain that the general solution of (4.2.13) is given by

$$y(n) = c_1\left(\frac{1+\sqrt{5}}{2}\right)^n + c_2\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

To find the solution satisfying the initial conditions $y(0) = 1$, $y(1) = 2$ (corresponding to one pair of rabbits initially) we substitute these values and get the system of equations for c_1 and c_2

$$\begin{aligned} 1 &= c_1 + c_2, \\ 2 &= c_1\frac{1+\sqrt{5}}{2} + c_2\frac{1-\sqrt{5}}{2}, \end{aligned}$$

the solution of which is $c_1 = 1 + 3\sqrt{5}/5$ and $c_2 = -3\sqrt{5}/5$.

4.3 Miscellaneous applications

Gambler's ruin

A gambler plays a sequence of games against an adversary. The probability that the gambler wins R 1 in any given game is q and the probability of him losing R 1 is $1-q$. He quits the game if he either wins a prescribed amount of N rands, or loses all his money; in the latter case we say that he has been ruined. Let $p(n)$ denotes the probability that the gambler will be ruined if he starts gambling with n rands. We build the difference equation satisfied by $p(n)$ using the following argument. Firstly, note that we can start observation at any moment, that is, the probability of him being ruined with n rands at the start is the same as the probability of him being ruined if he acquires n rands at any moment during the game. If at some moment during the game he has n rands, he can be ruined in two ways: by winning the next game and ruined with $n+1$ rand, or by losing and then being ruined with $n-1$ rands. Thus

$$p(n) = qp(n+1) + (1-q)p(n-1). \tag{4.3.1}$$

Replacing n by $n+1$ and dividing by q, we obtain

$$p(n+2) - \frac{1}{q}p(n+1) + \frac{1-q}{q}p(n) = 0, \tag{4.3.2}$$

with $n = 0, 1 \ldots, N$. We supplement (4.3.2) with the (slightly untypical) side (boundary) conditions $p(0) = 1$ and $p(N) = 0$.

The characteristic equation is given by

$$\lambda^2 - \frac{1}{q}\lambda + \frac{1-q}{q} = 0$$

and the eigenvalues are $\lambda_1 = \frac{1-q}{q}$ and $\lambda_2 = 1$. Hence, if $q \neq 1/2$, then the general solution can be written as

$$p(n) = c_1 + c_2\left(\frac{1-q}{q}\right)^n$$

and if $q = 1/2$, then $\lambda_1 = \lambda_2 = 1$ and

$$p(n) = c_1 + c_2 n.$$

To find the solution for the given boundary conditions, we denote $Q = (1-q)/q$ so that for $q \neq 1/2$

$$\begin{aligned} 1 &= c_1 + c_2, \\ 0 &= c_1 + Q^N c_2, \end{aligned}$$

from where

$$c_2 = \frac{1}{1-Q^N}, \qquad c_1 = -\frac{Q^N}{1-Q^N}$$

and

$$p(n) = \frac{Q^n - Q^N}{1-Q^N}.$$

Analogous considerations for $q = 1/2$ yield

$$p(n) = 1 - \frac{n}{N}.$$

For example, if $q = 1/2$ and the gambler starts with $n = 20$ rands with the target $N = 1000$, then

$$p(20) = 1 - \frac{20}{1000} = 0,98,$$

that is, his ruin is almost certain.

In general, if the gambler plays a long series of games, which can be modelled here as taking $N \to \infty$, then he will be ruined almost certainly even if the game is fair ($q = \frac{1}{2}$).

Chapter 5

Qualitative theory of differential equations

5.1 Introduction

In this chapter we shall consider the system of differential equations

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}) \tag{5.1.1}$$

where, in general,

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix},$$

and

$$\mathbf{f}(t, \mathbf{x}) = \begin{pmatrix} f_1(t, x_1, \ldots, x_n) \\ \vdots \\ f_n(t, x_1, \ldots, x_n) \end{pmatrix}.$$

is a nonlinear function of $\mathbf{x}$. Our main focus will be on autonomous systems of two equations with two unknowns

$$\begin{aligned} x_1' &= f_1(x_1, x_2), \\ x_2' &= f_2(x_1, x_2). \end{aligned} \tag{5.1.2}$$

Unfortunately, even for such a simplified case there are no known methods of solving (5.1.2) in general form. Though it is, of course, disappointing, it turns out that knowing exact solution to (5.1.2) is not really necessary. For example, let $x_1(t)$ and $x_2(t)$ denote the populations, at time t, of two species competing amongst themselves for the limited food and living space in some region. Further, suppose that the rates of growth of $x_1(t)$ and $x_2(t)$

are governed by (5.1.2). In such a case, for most purposes it is irrelevant to know the population sizes at each time t but rather it is important to know some qualitative properties of them. Specifically, the most important questions biologists ask are:

1. Do there exist values ξ_1 and ξ_2 at which the two species coexist in a steady state? That is to say, are there numbers ξ_1 and ξ_2 such that $x_1(t) \equiv \xi_1$ and $x_2(t) \equiv \xi_2$ is a solution to (5.1.2)? Such values, if they exist, are called *equilibrium points* of (5.1.2).

2. Suppose that the two species are coexisting in equilibrium and suddenly a few members of one or both species are introduced to the environment. Will $x_1(t)$ and $x_2(t)$ remain close to their equilibrium values for all future times? Or may be these extra few members will give one of the species a large advantage so that it will proceed to annihilate the other species?

3. Suppose that x_1 and x_2 have arbitrary values at $t = 0$. What happens for large times? Will one species ultimately emerge victorious, or will the struggle for existence end in a draw?

Mathematically speaking, we are interested in determining the following properties of system (5.1.2).

Existence of equilibrium solutions. Do there exist constant vectors $\mathbf{x^0} = (x_1^0, x_2^0)$ for which $\mathbf{x}(t) \equiv \mathbf{x^0}$ is a solution of (5.1.2)?

Stability. Let $\mathbf{x}(t)$ and $\mathbf{y}(t)$ be two solutions of (5.1.2) with initial values $\mathbf{x}(0)$ and $\mathbf{y}(0)$ very close to each other. Will $\mathbf{x}(t)$ and $\mathbf{y}(t)$ remain close for all future times, or will $\mathbf{y}(t)$ eventually diverge from $\mathbf{x}(t)$?

Long time behaviour. What happens to solutions $\mathbf{x}(t)$ as t approaches infinity. Do all solutions approach equilibrium values? If they do not approach equilibrium, do they at least exhibit some regular behaviour, like e.g. periodicity, for large times.

The first question can be answered immediately. In fact, since $\mathbf{x}(t)$ is supposed to be constant, then $\mathbf{x}'(t) \equiv 0$ and therefore $\mathbf{x^0}$ is an equilibrium value of (5.1.2) if and only if

$$\mathbf{f}(\mathbf{x^0}) \equiv \mathbf{0}, \tag{5.1.3}$$

that is, finding equilibrium solutions is reduced to solving a system of algebraic equations.

Example 5.1.1. Find all equilibrium values of the system of differential equations

$$\begin{aligned} x_1' &= 1 - x_2, \\ x_2' &= x_1^3 + x_2. \end{aligned}$$

We have to solve the system of algebraic equations

$$\begin{aligned} 0 &= 1 - x_2, \\ 0 &= x_1^3 + x_2. \end{aligned}$$

From the first equation we find $x_2 = 1$ and therefore $x_1^3 = -1$ which gives $x_1 = -1$ and the only equilibrium solution is

$$\mathbf{x^0} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

5.2 The phase-plane and orbits

In this section we shall give rudiments of the "geometric" theory of differential equations. The aim of this theory is to obtain as complete a description as possible of all solutions of the system of differential equations (5.1.2)

$$\begin{aligned} x_1' &= f_1(x_1, x_2), \\ x_2' &= f_2(x_1, x_2), \end{aligned} \tag{5.2.1}$$

without solving it but by analysing geometric properties of its orbits. To explain the latter, we note that every solution $x_1(t), x_2(t)$ defines a curve in the three dimensional space (t, x_1, x_2).

Example 5.2.1. The solution $x_1(t) = \cos t$ and $x_2(t) = \sin t$ of the system

$$\begin{aligned} x_1' &= -x_2, \\ x_2' &= x_1 \end{aligned}$$

describes a helix in the (t, x_1, x_2) space.

The foundation of the geometric theory of differential equations is the observation that every solution $x_1(t), x_2(t)$, $t_0 \leq t \leq t_1$, of (5.2.1) also describes a curve in the $x_1 - x_2$ plane, that is, as t runs from t_0 to t_1, the points $(x_1(t), x_2(t)$ trace out a curve in the $x_1 - x_2$ plane. This curve is called the *orbit*, or the *trajectory*, of the solution $\mathbf{x}(t)$ and the $x_1 - x_2$ plane is called the *phase plane* of the solutions of (5.2.1). Note that the orbit of an equilibrium solution reduces to a point.

Example 5.2.2. The solution of the previous example, $x_1(t) = \cos t, x_2(t) = \sin t$ traces out the unit circle $x^2 + y^2 = 1$ when t runs from 0 to 2π, hence the unit circle is the orbit of this solution. If t runs from 0 to ∞, then the pair $(\cos t, \sin t)$ traces out this circle infinitely often.

Example 5.2.3. Functions $x_1(t) = e^{-t}\cos t$ and $x_2(t) = e^{-t}\sin t$, $-\infty < t < \infty$, are a solution of the system

$$\begin{aligned} x_1' &= -x_1 - x_2, \\ x_2' &= x_1 - x_2. \end{aligned}$$

Since $r^2(t) = x_1^2(t) + x_2^2(t) = e^{-2t}$, we see that the orbit of this solution is a spiral traced towards the origin as t runs towards ∞.

One of the advantages of considering the orbit of the solution rather than the solution itself is that it is often possible to find the orbit explicitly without prior knowledge of the solution. Let $x_1(t), x_2(t)$ be a solution of (5.2.1) defined in a neighbourhood of a point $\bar{t}$. If e.g. $x_1'(\bar{t}) \neq 0$, then we can solve $x_1 = x_1(t)$ getting $t = t(x_1)$ in some neighbourhood of $\bar{x} = x_1(\bar{t})$. Thus, for t near $\bar{t}$, the orbit of the solution $x_1(t), x_2(t)$ is given as the graph of $x_2 = x_2(t(x_1))$. Next, using the chain rule and the inverse function theorem

$$\frac{dx_2}{dx_1} = \frac{dx_2}{dt}\frac{dt}{dx_1} = \frac{x_2'}{x_1'} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)}.$$

Thus, the orbits of the solution $x_1 = x_1(t), x_2(t) = x_2(t)$ of (5.2.1) are the solution curves of the first order scalar equation

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)} \tag{5.2.2}$$

and therefore to find the orbit of a solution there is no need to solve (5.2.1); we have to solve only the single first-order scalar equation (5.2.2).

Example 5.2.4. The orbits of the system of differential equations

$$\begin{aligned} x_1' &= x_2^2, \\ x_2' &= x_1^2. \end{aligned} \tag{5.2.3}$$

are the solution curves of the scalar equation $dx_2/dx_1 = x_1^2/x_2^2$. This is a separable equation and it is easy to see that every solution is of the form $x_2 = (x_1^3 + c)^{1/3}$, c constant. Thus, the orbits are the curves $x_2 = (x_1^3 + c)^{1/3}$ whenever $x_2 = x_1 \neq 0$ as then $x_1' = x_2' \neq 0$ and the procedure described above can be applied, see the example below.

Example 5.2.5. A solution curve of (5.2.2) is an orbit of (5.2.1) if and only if $x_1' \neq 0$ and $x_2' \neq 0$ simultaneously along the solution. If a solution curve of (5.2.2) passes through an equilibrium point of (5.2.1), where $x_1'(\bar{t}) = 0$ and $x_2'(\bar{t}) = 0$ for some $\bar{t}$, then the entire solution curve is not an orbit but rather it is a union of several distinct orbits. For example, consider the system of differential equations

$$\begin{aligned} x_1' &= x_2(1 - x_1^2 - x_2^2), && (5.2.4)\\ x_2' &= -x_1(1 - x_1^2 - x_2^2). && (5.2.5) \end{aligned}$$

The solution curves of the scalar equation

$$\frac{dx_2}{dx_1} = -\frac{x_1}{x_2}$$

are the family of concentric circles $x_1^2 + x_2^2 = c^2$. Observe however that to get the latter equation we should have assumed $x_1^2 + x_2^2 = 1$ and that each point of this circle is an equilibrium point of (5.2.5). Thus, the orbits of (5.2.5) are the circles $x_1^2 + x_2^2 = c^2$ for $c \neq 1$ and each point of the unit circle.

Similarly, the full answer for the system (5.2.3) of the previous example is that $x_2 = (x_1^3 + c)^{1/3}$ are orbits for $c \neq 0$ as then neither solution curve passes through the only equilibrium point $(0,0)$. For $c = 0$ the solution curve $x_2 = x_1$ consists of the equilibrium point $(0,0)$ and two orbits $x_2 = x_1$ for $x_1 > 0$ and $x_1 < 0$.

Note that in general it is impossible to solve (5.2.2) explicitly. Hence, usually we cannot find the equation of orbits in a closed form. Nevertheless, it is still possible to obtain an accurate description of all orbits of (5.2.1). In fact, the system (5.2.1) provides us with an explicit information about how fast and in which direction solution is moving at each point of the trajectory. In fact, as the orbit of the solution $(x_1(t), x_2(t))$ is a curve of which $(x_1(t), x_2(t))$ is a parametric description, $(x_1'(t), x_2'(t)) = (f_1(x_1, x_2), f_2(x_1, x_2))$ is the tangent vector to the orbit at the point (x_1, x_2) showing, moreover, the direction at which the orbit is traversed. In particular, the orbit is vertical at each point (x_1, x_2) where $f_1(x_1, x_2) = 0$ and $f_2(x_1, x_2) \neq 0$ and it is horizontal at each point (x_1, x_2) where $f_1(x_1, x_2) \neq 0$ and $f_2(x_1, x_2) = 0$. As we noted earlier, each point (x_1, x_2) where $f_1(x_1, x_2) = 0$ and $f_2(x_1, x_2) = 0$ gives an equilibrium solution and the orbit reduces to this point.

5.3 Qualitative properties of orbits

Let us consider the initial value problem for the system (5.1.2):

$$\begin{aligned} x_1' &= f_1(x_1,x_2),\\ x_2' &= f_2(x_1,x_2)\\ x_1(t_0) &= x_1^0, \quad x_2(t_0)=x_2^0, \end{aligned} \tag{5.3.1}$$

As we have already mentioned in Subsection 3.1.3, Picard's theorem, Theorem 2.2.4, can be generalized to systems. Due to the importance of it for the analysis of orbits, we shall state it here in full.

Theorem 5.3.1. *If each of the functions $f_1(x_1,x_2)$ and $f_2(x_1,x_2)$ have continuous partial derivatives with respect to x_1 and x_2. Then the initial value problem (5.3.1) has one and only one solution $\mathbf{x}(t)=(x_1(t),x_2(t))$, for every $\mathbf{x^0}=(x_1^0,x_2^0)\in\mathbb{R}^2$ defined at least for t in some neighborhood of t_0.*

Firstly, we prove the following result.

Lemma 5.3.2. *If $\mathbf{x}(t)$ is a solution to*

$$\mathbf{x}'=\mathbf{f}(\mathbf{x}), \tag{5.3.2}$$

then for any c the function $\hat{\mathbf{x}}(t)=\mathbf{x}(t+c)$ also satisfies this equation.

Proof. Define $\tau=t+c$ and use the chain rule for $\hat{x}$. We get

$$\frac{d\hat{\mathbf{x}}(t)}{dt}=\frac{d\mathbf{x}(t+c)}{dt}=\frac{d\mathbf{x}(\tau)}{d\tau}\frac{d\tau}{dt}=\frac{d\mathbf{x}(\tau)}{d\tau}=\mathbf{f}(\mathbf{x}(\tau))=\mathbf{f}(\mathbf{x}(t+c))=\mathbf{f}(\hat{\mathbf{x}}(t)).$$

Example 5.3.3. For linear systems the result follows directly as $\mathbf{x}(t)=e^{t\mathcal{A}}\mathbf{v}$ for arbitrary vector $\mathbf{v}$, so that $\hat{\mathbf{x}}(t)=\mathbf{x}(t+c)=e^{(t+c)\mathcal{A}}\mathbf{v}=e^{t\mathcal{A}}e^{c\mathcal{A}}\mathbf{v}=e^{t\mathcal{A}}\mathbf{v}'$ for some other vector $\mathbf{v}'$ so that $\hat{\mathbf{x}}(t)$ is again a solution.

We shall now prove two properties of orbits that are crucial to analyzing system (5.1.2).

Theorem 5.3.4. *Assume that the assumptions of Theorem 5.3.1 are satisfied. Then*

(i) there exists one and only one orbit through every point $\mathbf{x^0}\in\mathbb{R}^2$. In particular, if the orbits of two solutions $\mathbf{x}(t)$ and $\mathbf{y}(t)$ have one point in common, then they must be identical.

(ii) Let $\mathbf{x}(t)$ *be a solution to (5.1.2). If for some* $T > 0$ *and some* t_0 *we have* $\mathbf{x}(t_0+T) = \mathbf{x}(t_0)$*, then* $\mathbf{x}(t+T) = \mathbf{x}(t)$ *for all* t*. In other words, if a solution* $\mathbf{x}(t)$ *returns to its starting value after a time* $T > 0$*, then it must be periodic (that is, it must repeat itself over every time interval of length* T*).*

Proof. ad (i) Let $\mathbf{x^0}$ be any point in $\mathbb{R}^2$. Then from Theorem 5.3.1 we know that there is a solution of the problem $\mathbf{x}' = \mathbf{f}(\mathbf{x}), \mathbf{x}(0) = \mathbf{x^0}$ and the orbit of this solution passes through $\mathbf{x^0}$ from the definition of the orbit. Assume now that there is another orbit passing through $\mathbf{x^0}$, that is, there is a solution $\mathbf{y}(t)$ satisfying $\mathbf{y}(t_0) = \mathbf{x^0}$ for some t_0. From Lemma 5.3.2 we know that $\hat{\mathbf{y}}(t) = \mathbf{y}(t+t_0)$ is also a solution. However, this solution satisfies $\hat{\mathbf{y}}(0) = \mathbf{y}(t_0) = \mathbf{x^0}$, that is, the same initial condition as $\mathbf{x}(t)$. By the uniqueness part of Theorem 5.3.1 we must then have $\mathbf{x}(t) = \hat{\mathbf{y}}(t) = \mathbf{y}(t+t_0)$ for all t for which the solutions are defined. This implies that the orbits are identical. In fact, if ξ is an element of the orbit of $\mathbf{x}$, then for some t' we have $\mathbf{x}(t') = \xi$. However, we have also $\xi = \mathbf{y}(t'+t_0)$ so that ξ belongs to the orbit of $\mathbf{y}(t)$. Conversely, if ξ belongs to the orbit of $\mathbf{y}$ so that $\xi = \mathbf{y}(t'')$ for some t'', then by $\xi = \mathbf{y}(t'') = \mathbf{x}(t''-t_0)$, we see that ξ belongs to the orbit of $\mathbf{x}$.

ad (ii) Assume that for some numbers t_0 and $T > 0$ we have $\mathbf{x}(t_0) = \mathbf{x}(t_0 + T)$. The function $\mathbf{y}(t) = \mathbf{x}(t+T)$ is again a solution satisfying $\mathbf{y}(t_0) = \mathbf{x}(t_0+T) = \mathbf{x}(t_0)$, thus from Theorem 5.3.1, $\mathbf{x}(t) = \mathbf{y}(t)$ for all t for which they are defined and therefore $\mathbf{x}(t) = \mathbf{x}(t+T)$ for all such t. ■

Example 5.3.5. A curve in the shape of a figure 8 cannot be an orbit. In fact, suppose that the solution passes through the intersection point at some time t_0, then completing the first loop, returns after time T, that is, we have $\mathbf{x}(t_0) = \mathbf{x}(t_0+T)$. From (ii) it follows then that this solution is periodic, that is, it must follow the same loop again and cannot switch to the other loop.

Corollary 5.3.6. *A solution* $\mathbf{y}(t)$ *of (5.1.2) is periodic if and only if its orbit is a closed curve in* $\mathbb{R}^2$*.*

Proof. Assume that $\mathbf{x}(t)$ is a periodic solution of (5.1.2) of period T, that is $\mathbf{x}(t) = \mathbf{x}(t+T)$. If we fix t_0, then, as t runs from t_0 to t_0+T, the point $\mathbf{x}(t) = (x_1(t), x_2(t))$ traces a curve, say C, from $\xi = \mathbf{x}(t)$ back to the same point ξ without intersections and, if t runs from $-\infty$ to ∞, the curve C is traced infinitely many times.

Conversely, suppose that the orbit is a closed curve (containing no equilibrium points). The point $\mathbf{x}(t)$ moves along this curve with a speed of magnitude $v(x_1,x_2) = \sqrt{f_1^2(x_1,x_2) + f_2^2(x_1,x_2)}$. The curve is closed and, since there is no equilibrium point on it, that is, f_1 and f_2 are not simultaneously zero at any point, the speed v has a non-zero minimum on it. Moreover, as the parametric description of this curve if differentiable, it has a finite length. Thus, the point $\mathbf{x}(t)$ starting from a point $\xi = \mathbf{x}(t_0)$ will traverse the whole curve in finite time, say T, that is $\mathbf{x}(t_0) = \mathbf{x}(t_0 + T)$ and the solution is periodic. ■

Example 5.3.7. Show that every solution $z(t)$ of the second order differential equation

$$z'' + z + z^3 = 0$$

is periodic. We convert this equation into a system: let $z = x_1$ so that

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -x_1 - x_1^3. \end{aligned}$$

The orbits are the solution curves of the equation

$$\frac{dx_2}{dx_1} = -\frac{x_1 + x_1^3}{x_2},$$

so that

$$\frac{x_2^2}{2} + \frac{x_1^2}{2} + \frac{x_1^4}{4} = c^2$$

is the equation of orbits. If $c \neq 0$, then none of them contains the unique equilibrium point $(0,0)$. By writing the above equation in the form

$$\frac{x_2^2}{2} + \left(\frac{x_1^2}{2} + \frac{1}{2}\right)^2 = c^2 + \frac{1}{4}$$

we see that for each $c \neq 0$ it describes a closed curve consisting of two branches $x_2 = \pm\frac{1}{\sqrt{2}}\sqrt{4c^2 + 1 - (x_1^2 + 1)^2}$ that stretch between $x_1 = \pm\sqrt{1 + \sqrt{4c^2 + 1}}$. Consequently, every solution is a periodic function.

5.4 An application

In this section we shall discuss the predator-prey model introduced in Section 1.3.5. It reads

$$\begin{aligned} \frac{dx_1}{dt} &= (r - f)x_1 - \alpha x_1 x_2, \\ \frac{dx_2}{dt} &= -(s + f)x_2 + \beta x_1 x_2 \end{aligned} \qquad (5.4.1)$$

where α, β, r, s, f are positive constants. In the predator-prey model x_1 is the density of the prey, x_2 is the density of the predators, r is the growth rate of the prey in the absence of predators, $-s$ is the growth rate of predators in the absence of prey (the population of predators dies out without the supply of the sole food source – prey). The quadratic terms account for predator–prey interaction and f represents indiscriminate killing of both prey and predators. The model was introduced in 1920s by Italian mathematician Vito Volterra to explain why, in the period of reduced (indiscriminate) fishing, the relative number predators (sharks) significantly increased.

Let us consider first the model without fishing

$$\begin{aligned}\frac{dx_1}{dt} &= rx_1 - \alpha x_1 x_2,\\ \frac{dx_2}{dt} &= -sx_2 + \beta x_1 x_2\end{aligned}\tag{5.4.2}$$

Observe that there are two equilibrium solutions $x_1(t) = 0, x_2(t) = 0$ and $x_1(t) = s/\beta, x_2(t) = r/\alpha$. The first solution is not interesting as it corresponds to the total extinction. We observe also that we have two other solutions $x_1(t) = c_1e^{rt}$, $x_2(t) = 0$ and $x_1(t) = 0, x_2(t) = c_2e^{-st}$ that correspond to the situation when one of the species is extinct. Thus, both positive x_1 and x_2 semi-axes are orbits and, by Theorem 5.3.4 (i), any orbit starting in the first quadrant will stay there or, in other words, any solution with positive initial data will remain strictly positive for all times.

The orbits of (5.4.2) are the solution curves of the first order separable equation

$$\frac{dx_2}{dx_1} = \frac{x_2(-s+\beta x_1)}{x_1(r-\alpha x_2)}\tag{5.4.3}$$

Separating variables and integrating we obtain

$$r\ln x_2 - \alpha x_2 + s\ln x_1 - \beta x_1 = k$$

which can be written as

$$\frac{x_2^r}{e^{\alpha x_2}}\frac{x_1^s}{e^{\beta x_1}} = K.\tag{5.4.4}$$

Next, we prove that the curves defined by (5.4.4) are closed. It is not an easy task. To accomplish this we shall show that for each x_1 from a certain open interval $(x_{1,m}, x_{1,M})$ we have exactly two solutions $x_{2,m}(x_1)$ and $x_{2,M}(x_1)$ and that these two solutions tend to common limits as x_1 approaches $x_{1,m}$ and $x_{1,M}$.

First, let as define $f(x_2) = x_2^re^{-\alpha x_2}$ and $g(x_1) = x_1^se^{-\beta x_1}$. We shall analyze only f as g is of the same form. Due to positivity of all the coefficients, we see that $f(0) = 0$, $\lim_{x_2\to\infty} f(x_2) = 0$ and also $f(x_2) > 0$ for $x_2 > 0$. Further

$$f'(x_2) = x_2^{r-1}e^{-\alpha x_2}(r-\alpha x_2),$$

so that f is increasing from 0 to $x_2 = r/\alpha$ where it attains global maximum, say M_2, and then starts to decrease monotonically to 0. Similarly, $g(0) = \lim_{x_1\to\infty} g(x_1) = 0$ and $g(x_1) > 0$ for $x_1 > 0$ and it attains global maximum M_1 at $x_1 = s/\beta$. We have to analyze solvability of

$$f(x_2)g(x_1) = K.$$

Firstly, there are no solutions if $K > M_1M_2$, and for $K = M_1M_2$ we have the equilibrium solution $x_1 = s/\beta, x_2 = r/\alpha$. Thus, we have to consider $K = \lambda M_2$ with $\lambda < 1$. Let us write this equation as

$$f(x_2) = \frac{\lambda}{g(x_1)}M_2. \tag{5.4.5}$$

From the shape of the graph g we find that the equation $g(x_1) = \lambda$ has no solution if $\lambda > M_1$ but then $\lambda/g(x_1) \geq \lambda/M_1 > 1$ so that (5.4.5) is not solvable. If $\lambda = M_1$, then we have again the equilibrium solution. Finally, for $\lambda < M_1$ there are two solutions $x_{1,m}$ and $x_{1,M}$ satisfying $x_{1,m} < s/\beta < x_{1,M}$. Now, for x_1 satisfying $x_{1,m} < x_1 < x_{1,M}$ we have $\lambda/g(x_1) < 1$ and therefore for such x_1 equation (5.4.5) has two solutions $x_{2,m}(x_1)$ and $x_{2,M}(x_1)$ satisfying $x_{2,m} < r/\alpha < x_{2,M}$, again on the basis of the shape of the graph of f. Moreover, if x_1 moves towards either $x_{1,m}$ or $x_{1,M}$, then both solutions $x_{2,m}$ and $x_{2,M}$ move towards r/α, that is the set of points satisfying (5.4.5) is a closed curve.

Summarizing, the orbits are closed curves encircling the equilibrium solution $(s/\beta, r/\alpha)$ and are traversed in the anticlockwise direction. Thus, the solutions are periodic in time. The evolution can be described as follows. Suppose that we start with initial values $x_1 > s/\beta, x_2 < r/\alpha$, that is, in the lower right quarter of the orbit. Then the solution will move right and up till the prey population reaches maximum $x_{1,M}$. Because there is a lot of prey, the number of predators will be still growing but then the number of prey will start decreasing, slowing down the growth of the predator's population. The decrease in the prey population will eventually bring the growth of predator's population to stop at the maximum $x_{2,M}$. From now on the number of predators will decrease but the depletion of the prey population from the previous period will continue to prevail till the population reaches the minimum $x_{1,m}$, when it will start to take advantage of the decreasing number of predators and will start to grow; this growth will, however, start to slow down when the population of predators will reach its minimum. However, then the number of prey will be increasing beyond the point when the number of predators is the least till the growing number of predators will eventually cause the prey population to decrease having reached its peak at $x_{1,M}$ and the cycle will repeat itself.

Now we are ready to provide the explanation of the observational data. Including fishing into the model, according to (5.4.1), amounts to changing parameters r and s to $r-f$ and $s+f$ but the structure of the system does not change, so that the equilibrium solution becomes

$$\left(\frac{s+f}{\beta}, \frac{r-f}{\alpha}\right). \tag{5.4.6}$$

Thus, with a moderate amount of fishing ($f<r$), in the equilibrium solution there is more fish and less sharks in comparison with no-fishing situation. Thus, if we reduce fishing, the equilibrium moves towards larger amount of sharks and lower amount of fish. Of course, this is true for equilibrium situation, which not necessarily corresponds to reality, but as the orbits are closed curves around the equilibrium solution, we can expect that the amounts of fish and sharks in a non-equilibrium situation will change in a similar pattern. We can confirm this hypothesis by comparing average numbers of sharks and fish over the full cycle. For any function f its average over an interval (a,b) is defined as

$$\bar{f} = \frac{1}{b-a}\int\limits_a^b f(t)dt,$$

so that the average numbers if fish and sharks over one cycle is given by

$$\overline{x_1} = \frac{1}{T}\int\limits_0^T x_1(t)dt, \qquad \overline{x_2} = \frac{1}{T}\int\limits_0^T x_2(t)dt.$$

It turns out that these averages can be calculated explicitly. Dividing the first equation of (5.4.2) by x_1 gives $x_1'/x_1 = r-\alpha x_2$. Integrating both sides, we get

$$\frac{1}{T}\int\limits_0^T \frac{x_1'(t)}{x_1(t)}dt = \frac{1}{T}\int\limits_0^T (r-\alpha x_2(t))dt.$$

The left-hand side can be evaluated as

$$\int\limits_0^T \frac{x_1'(t)}{x_1(t)}dt = \ln x_1(T) - \ln x_1(0) = 0$$

on account of the periodicity of x_1. Hence,

$$\frac{1}{T}\int\limits_0^T (r-\alpha x_2(t))dt = 0,$$

and

$$\overline{x_2} = \frac{r}{\alpha}. \tag{5.4.7}$$

In the same way,

$$\overline{x_1} = \frac{s}{\beta}, \tag{5.4.8}$$

so that the average values of x_1 and x_2 are exactly the equilibrium solutions. Thus, we can state that introducing fishing is more beneficial to prey than predators as the average numbers of prey increases while the average number of predators will decrease in accordance with (5.4.6), while reducing fishing will have the opposite effect of increasing the number of predators and decreasing the number of prey.