Methods of Mathematical Modelling 1

Björn Stinner
Mathematics Institute
University of Warwick

Last updated: Sat May 20 12:48:33 BST 2023

Acknowledgments

To a great extent, the notes for the module *MA146 Methods of Mathematical Modelling 1* have emerged from the predecessor module *MA133 Differential Equations*. Notes for that module had been provided by David Wood and David McCormick, which in turn are strongly linked with James Robinson's book on differential equations. Novel parts are centred around modelling and dimensional analysis, where some ideas have been discussed with Shreyas Mandre and Emma Davis. I'd also like to thank students, TA's and colleagues who helped to improve content and form of these notes.

Björn Stinner

Contents

Chapter 1

Mathematical modelling

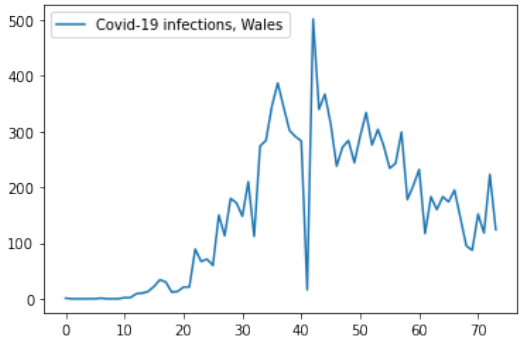
1.1 Motivation

What motivates us to undertake mathematical modelling? This motivation ultimately can be traced to the primal impulse in humans to use abstract concepts for explaining phenomena and processes, usually with the aim of making predictions about them and controlling them. An example is pre-school children realising that when two out of five apples are eaten, three remain, and then are able to transfer such understanding to using up construction bricks or hiding soft toys. More sophisticated examples are the concept of orthogonality, whether used to find minimal distances or for separating features, or computing with probabilities, whether to assess chances in a gamble or to inform decisions. In many cases, insight can simply be gained without much abstraction by 'trying out', i.e., performing suitable experiments. However, these might be too expensive, too dangerous, or simply not feasible, and then reusing abstract, theoretical ideas that turned out beneficial for describing similar phenomena may be the only way to gain insight.

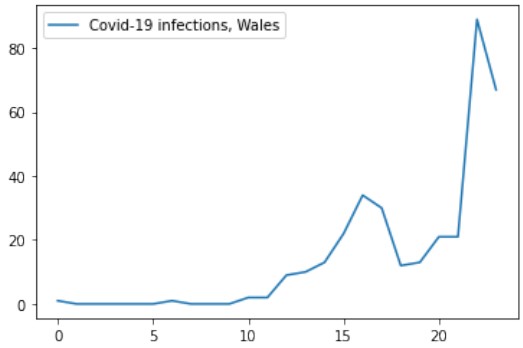
Generally, the purpose of mathematical modelling is gaining insight and expanding knowledge by describing behaviours and trends in real-world problems, and by making predictions about these. In the modelling process, assumptions are made that involve simplification and reduction to essential features, characteristics, or quantities of interest. Suitable state variables and parameters for these are identified and then related in mathematical equations. This process is guided by some principles and systematic procedures, and it hopefully leads to equations that can be solved and further analysed. In fact, the choices made to formulate the model are informed by available tools, methodology, and techniques for its study. Before we discuss the modelling process in more detail, let us look at an example.

The following graph displays the Covid-19 infections in Wales for the first 75 days of the pandemic, starting on 28/02/2020 (day 0 in the graph below):

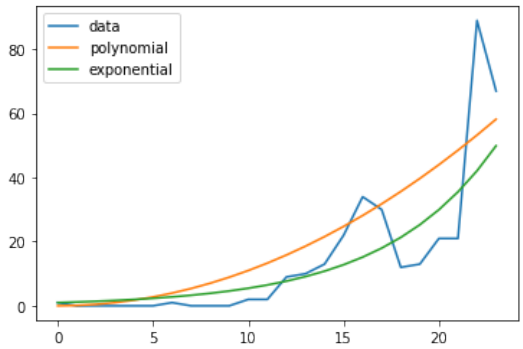
1.1. MOTIVATION



Suppose now we are 25 days into the outbreak and would like to make a prediction about the number of infected people. For this purpose, we simply 'fit' some curves to the data and use their extensions for a prediction. So we only use data processing tools and completely ignore background and context of the data. Let us first plot the data for the first 25 days to get an idea of possible curves.

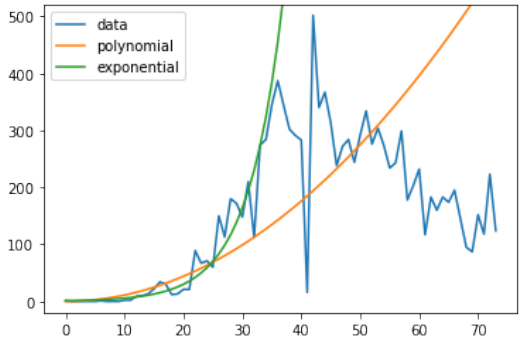


We observe that the numbers effectively increase, but there are many increasing functions that could describe the trend. Writing t for the time measured in days, for instance, a quadratic polynomial of the form $p(t) = mt^2$ with a suitable number $m > 0$ or an exponential function of the form $f(t) = se^{ct}$ with parameters $s, c > 0$ can effectively capture the growth. For some suitable parameters m, s, and c we indeed seem to be able to get good fits in that the curves 'nicely' run through the data points for the first 25 days:



Let us now use these curves for a prediction. We plot the data and the curves for 75 days.

1.1. MOTIVATION



This doesn't look very good. The polynomial fit quickly runs away from the data points. The exponential fit follows them at least for a while towards the spike. Both approaches are not able to describe the fact that the infection numbers reach a maximum before decaying again. Of course, for the fits we had chosen functions that increase, however this was inspired by the trend in the data for the first 25 days. In order to recover the observed behaviour of the infections numbers for the later days we need to combine the data with some additional knowledge about what they mean and describe.

Mathematical models of infectious disease epidemiology track the infected individuals in a population using assumptions about transmission and recovery of the disease. A classical approach is the *SEIR model*, on which many later, refined models are based. The population, denoted by N, is assumed constant in time and split into *susceptible* individuals $S(t)$, *exposed* (but not yet infectious) individuals $E(t)$, *infected* (and infectious) individuals $I(t)$, and *recovered*[1] individuals $R(t)$. As indicated, these population compartments depend on time t measured here in days, but they have to sum up to the overall population at any time,

$$S(t) + E(t) + I(t) + R(t) = N \quad \forall t.$$

Assumptions are then made about the changes of the compartments:

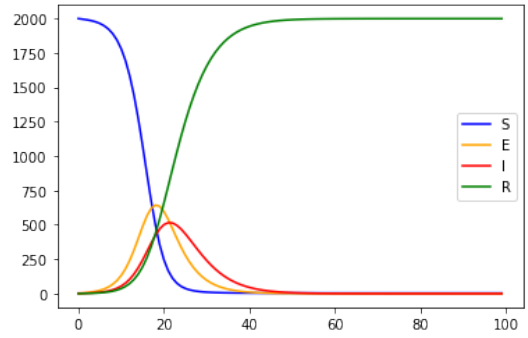
- Susceptible individuals get in contact with other people, and the chance that another person is infectious is $I(t)/N$. We assume that they become exposed from such a contact at a rate that is proportional to a factor $\beta > 0$.
- Exposed individuals become (ill and) infectious at a rate with a factor denoted by $\epsilon > 0$.
- Individuals from the compartment $I(t)$ recover at a rate with a factor denoted by $\gamma > 0$.

Denoting the change with a time derivative we obtain a set of differential equations, one equation for each compartment:

$$\begin{aligned}
\frac{dS}{dt}(t) &= -\beta \frac{I(t)}{N} S(t) \\
\frac{dE}{dt}(t) &= \beta \frac{I(t)}{N} S(t) - \epsilon E(t) \\
\frac{dI}{dt}(t) &= \epsilon E(t) - \gamma I(t) \\
\frac{dR}{dt}(t) &= \gamma I(t)
\end{aligned}$$

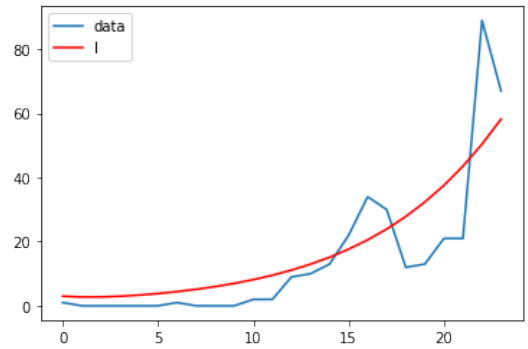
[1] At least in the best case. Somewhat cynically, the notation $R(t)$ also allows for the interpretation as a *removed* individual...

Here is a typical solution to the ODE system, where we start with a hign number $S(0)$, a small number $I(0)$, and $E(0) = R(0) = 0$:

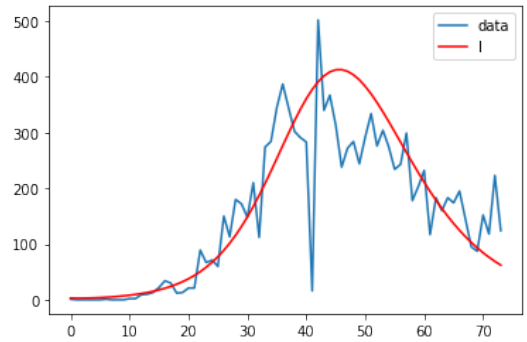


The curve for I looks very promising, this is effectively the behaviour that our numbers display and that we expect!

Getting back to our prediction problem, we choose parameters β, γ, ϵ and initial values such that fairly good fit of the $I(t)$ values with the infections in the first 25 days is obtained.



Let us now use these parameters or, respectively, the solution to the corresponding ODE system to make a prediction. Below we plot the infection data and the numerically obtained $I(t)$ values for the first 75 days.



This looks quite good, the solution captures the trend in the data points with the increase and the decay, even the maximum looks at a good position. So the mess of data is effectively fairly well described by a couple of equations with very few parameters. Now, can we further extend the prediction? Let us look into a year.

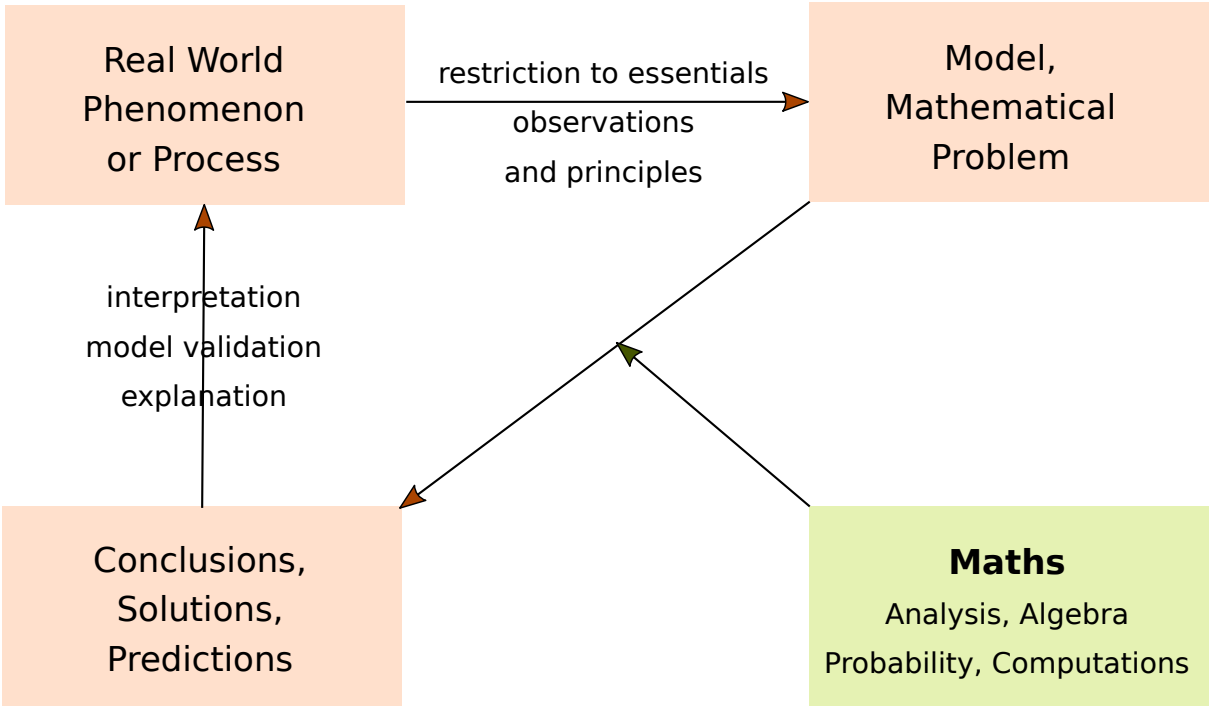
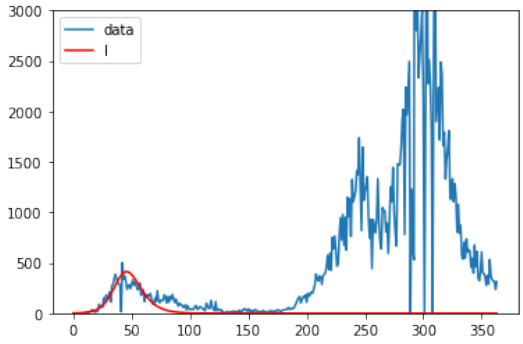


Figure 1.1: Illustration of the modelling cycle.



Whilst the model can describe the first wave, the computed values for $I(t)$ stay close to zero after and the model fails to pick up the later waves, particularly the one early in 2021 caused by a variant of the virus. For this purpose, more sophisticated models would be needed.

Improving the quality of models, such as by extending their range of applicability, is part of the overall modelling process. The process can be illustrated with a so-called modelling cycle, the aspects of which we study in the next section.

1.2 Modelling cycle

Figure 1.1 illustrates what we mean with a *modelling cycle*. Essentially this is a graph, here with three vertices that are connected with directed edges (arrows) so that a closed loop is obtained. These edges represent *modelling stages*. Before we discuss these let us briefly remark that the

modelling cycle in Figure 1.1 is fairly general but also basic. One can find many others in the literature with more detail, where some of its vertices and edges are broken down into sub-steps.

1.2.1 Model formulation

Mathematical models express knowledge, intuition, and assumptions about a real world phenomenon or process in a mathematical language. This involves identifying suitable *variables* that quantify aspects deemed of importance and of relevance in any description. These variables then are related to each other by means of equations. In this module, we will almost exclusively use *differential equations* for this purpose. In many applications, knowing the past and current *characteristics*, expressed in terms of the variables, of a phenomenon or process and the *laws* that govern their *changes* enables to make *predictions* about future characteristics, i.e., the values of the variables. Changes of the variables can be expressed in terms of their derivatives with respect to time, so when expressing the laws in mathematical form then we obtain equations that put the variables and their derivatives in a relation with each other. These ideas proved very successful as they often resulted in mathematical problems that could be solved, and moreover the solutions could be interpreted to yield insight into the phenomenon.

We have seen already that mathematical models always involve *restrictions* to aspects that are deemed of importance, which is the essential reason why we have a modelling cycle instead of something like a pipeline: when the model is analysed and we compare the results with observations and measurements then we might not be satisfied and need to re-think the choice of variables and their relations. In general, we are fairly free to define these relations and may simply express in mathematical terms what we *observe*.

Example 1.1 (Vicsek model for flocks and swarms)**.** *A simple model for flocks and swarms inspired by observations goes back to Tamas Vicsek and coworkers*[2]*. Assume a flock with $N \in \mathbb{N}$ members. Each member has a position denoted by $x_i(t)$, $i = 1, \dots, N$, in the square $\Omega = [0, L]^2$ with edge length $L > 0$. The members move with the same constant velocity $v_0 > 0$ but in an individual direction given by an individual angle θ_i relative to the first coordinate axis. We can express this by the differential equations*

$$\frac{d}{dt} x_i(t) = v_0 \begin{pmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{pmatrix}, \quad i = 1, \dots, N.$$

The original Vicsek model is discrete in time, however. Given a small time step size $\tau > 0$ we introduce discrete times $t^{(m)} = m\tau$ where $m \in \mathbb{N} \cup \{0\}$ at which we want to track the positions. We denote the position of member i at time $t^{(m)}$ with $x_i^{(m)}$ and the orientation with $\theta_i^{(m)}$. We then move the points representing the swarm members according to

$$x_i^{(m+1)} = x_i^{(m)} + \tau v_0 \begin{pmatrix} \cos(\theta_i^{(m)}) \\ \sin(\theta_i^{(m)}) \end{pmatrix}, \quad i = 1, \dots, N.$$

Observations showed that the swarm members adjust the orientation of their movement depending on their neighbours. The assumption was made that the new orientation in each time step is simply the average of the neighbours' orientation plus some random fluctuations (noise),

$$\theta_i^{(m)} = \left\langle \theta_j^{(m-1)} \right\rangle_r + \eta W_i.$$

[2] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet: *Novel Type of Phase Transition in a System of Self-Driven Particles.* Phys. Rev. Lett. 75, 1226 (1995).

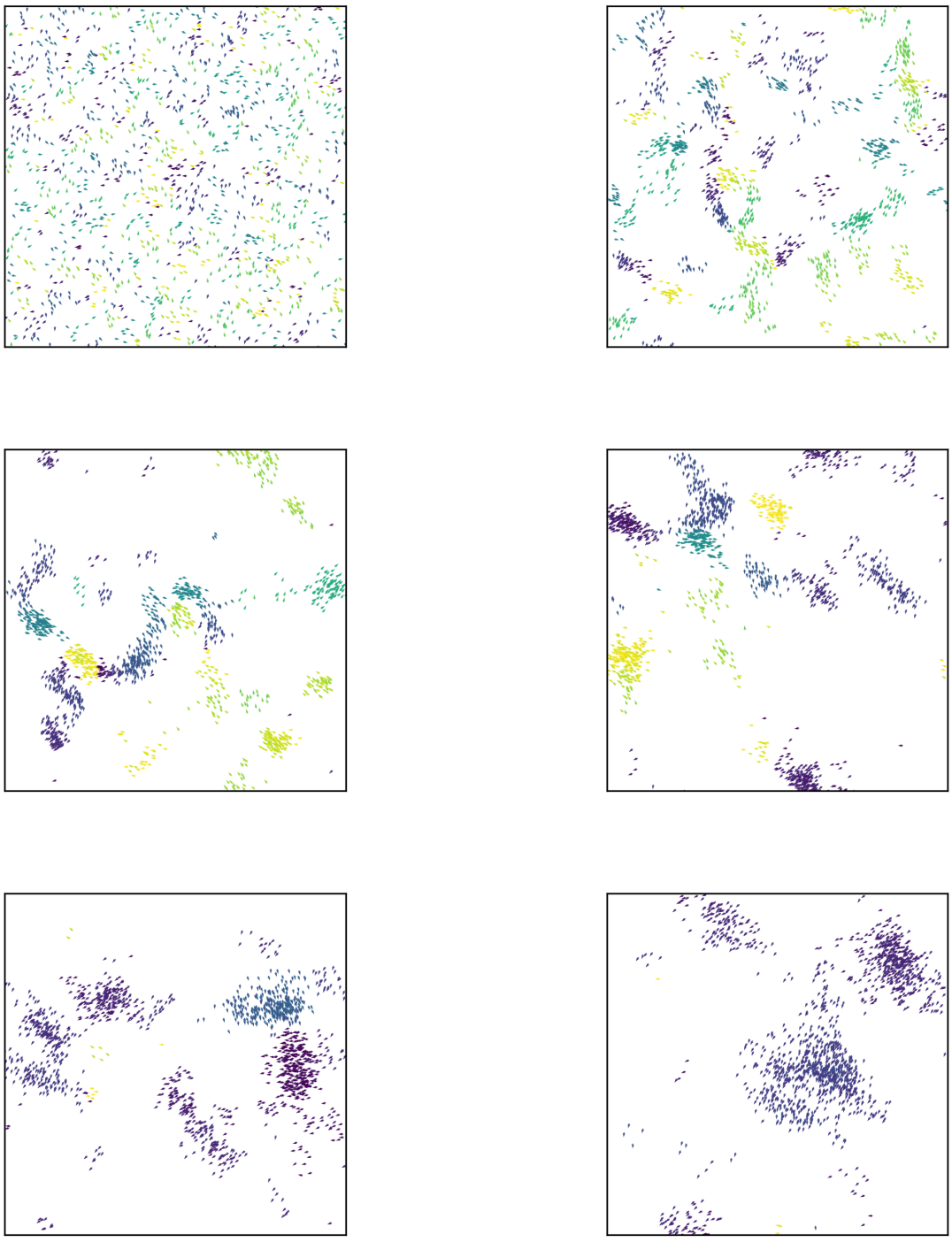


Figure 1.2: Swarming simulation with the Vicsek model in Example 1.1. A solution is displayed at the beginning and after 10, 30, 80, 160, 240 steps, starting at the upper left, from left to right and after from top to bottom.

Here, $\langle\cdot\rangle_r$ means the average of all swarm members in the ball of radius $r > 0$ around member i, $\eta > 0$ is a factor determining the strength of the fluctuations, and W_i is a normally distributed random variable.

Figure 1.2 gives an impression of a solution. Note that every solution is different as the fluctuations are random. Researchers are interested in the average behaviour *but also in the emergence of swarming behaviour. Indeed, we can see in the figure that the swarm members join up to form bigger and bigger groups, which then move together.*

In some applications, *principles* can guide us to come up with a model. These may restrict flexibility but maybe necessary to ensure that the model satisfies certain side conditions. They often provide structure that can be exploited when analysing the model. As an example, we look into chemical reactions, where a good model ensures that the total mass is conserved, as it can be observed experimentally.

Example 1.2 (Principle guided modelling, chemical reaction kinetics and the law of mass action)**.** *Chemical reaction are often expressed by graphs with vertices formed by (sums of) reactants and arrows as edges indicating possible reactions, possibly with rate constants. For instance,*

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \tag{1.1}$$

describes the reversible transformation of two reactants A and B into a reactant C.

The law of mass action *is a fundamental principle for translating chemical reactions into sets of differential equations, which are called* rate equations*. Consider a reactant R that is involved in j_R reactions and denote by $r(t)$ its density*[3]*. The rate equation for R is of the form*

$$\frac{d}{dt}r(t) = \sum_{k=1}^{j_R} (\textit{creation rate of } R \textit{ by reaction } k) - \sum_{k=1}^{j_R} (\textit{consumption rate of } R \textit{ by reaction } k).$$

For the reactants A, B, C in the above example (1.1), we obtain

$$\begin{aligned}
\frac{d}{dt}a(t) &= k_2 c(t) - k_1 a(t) b(t),\\
\frac{d}{dt}b(t) &= k_2 c(t) - k_1 a(t) b(t),\\
\frac{d}{dt}c(t) &= k_1 a(t) b(t) - k_2 c(t).
\end{aligned}$$

If several molecules of a reactant are involved in a reaction then we have to count each in the creation rate and the consumption rate. Let us consider a more complicated example and assume that n_A molecules of A and n_B molecules of B react to n_C molecules of C and n_D molecules of D,

$$n_A A + n_B B \overset{k}{\to} n_C C + n_D D.$$

The numbers n_A, n_B, n_C, n_D are called stoichiometries *and feature as prefactors of the rate in*

[3] In the literature, often, the word *concentration* is used, for instance, in [6].

the rate equations:

$$\begin{aligned}\frac{d}{dt}a(t) &= -n_A k a(t)^{n_A} b(t)^{n_B} \ ,\\ \frac{d}{dt}b(t) &= -n_B k a(t)^{n_A} b(t)^{n_B} \ ,\\ \frac{d}{dt}c(t) &= n_C k a(t)^{n_A} b(t)^{n_B} \ ,\\ \frac{d}{dt}d(t) &= n_D k a(t)^{n_A} b(t)^{n_B} \ .\end{aligned}$$

The actual rate *here is the same for all reactants and given by*

$$\frac{1}{n_A}\frac{d}{dt}a(t) = \frac{1}{n_B}\frac{d}{dt}b(t) = -\frac{1}{n_C}\frac{d}{dt}c(t) = -\frac{1}{n_D}\frac{d}{dt}d(t) = -k a(t)^{n_A} b(t)^{n_B}.$$

Note that then

$$\frac{d}{dt}\Big(\frac{1}{n_A}a(t) + \frac{1}{n_C}c(t)\Big) = 0, \tag{1.2}$$

and analogously for the time derivatives of the combinations $\frac{1}{n_A}a(t) + \frac{1}{n_D}d(t)$, $\frac{1}{n_B}b(t) + \frac{1}{n_C}c(t)$, *and* $\frac{1}{n_B}b(t) + \frac{1}{n_D}d(t)$. *These are identities that express* mass conservation, *namely, that all atoms (or, respectively, their masses) that are present before the reaction are also present after the reaction, just bound in different molecules.*

Mass conservation is a well established fact in many applications, and it is good if a model satisfies this observation. However, it can also prove useful otherwise. From (1.2) *we see that* $\frac{1}{n_A}a(t) + \frac{1}{n_C}c(t) = \mu$ *is a number that is independent of time* t, *and we may be able to work out what that constant is, for instance, because we may know the densities at the start of the reaction. If we then manage to work out* $a(t)$ *then* $c(t) = n_C\mu - \frac{n_C}{n_A}a(t)$ *is immediately determined, too.*

To be a bit more specific, let us consider an enzyme reaction. *These protein molecules* E *act on a substrate* S *to facilitate the generation of a product* P *but are not used up in this process. An example is the enzyme* reverse transcriptase *used by viruses such as* HIV *to generate DNA from their RNA. Leonor Michaelis and Maud Menten investigated such a the process and came up with the following diagram*

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\to} P + E.$$

By the law of mass action the corresponding differential equations read

$$\begin{aligned}\frac{d}{dt}s(t) &= -k_1 s(t)e(t) + k_{-1}c(t),\\ \frac{d}{dt}c(t) &= k_1 s(t)e(t) - \big(k_{-1} + k_2\big)c(t),\\ \frac{d}{dt}p(t) &= k_2 c(t),\\ \frac{d}{dt}e(t) &= -k_1 s(t)e(t) + \big(k_{-1} + k_2\big)c(t).\end{aligned}$$

At the beginning of the reaction (say, at time $t = 0$*) only the substrate and the enzyme are present,* $e(0) = e_0 \geq 0$ *and* $s(0) = s_0 \geq 0$ *whilst* $c(0) = p(0) = 0$. *Noting that* $\frac{d}{dt}(e(t) + c(t)) = 0$ *we see that* $e(t) + c(t) = e(0) + c(0) = e_0$ *at all times, hence*

$$e(t) = e_0 - c(t).$$

Moreover, if we know $c(t)$ then

$$p(t) = \int_0^t k_2 c(\tau) d\tau.$$

Effectively, the problem thus reduces to a system of two equations,

$$\frac{d}{dt}s(t) = -k_1 s(t) e_0 + \big(k_1 s(t) + k_{-1}\big)c(t),$$
$$\frac{d}{dt}c(t) = k_1 s(t) e_0 - \big(k_1 s(t) + k_{-1} + k_2\big)c(t),$$

with the additional conditions $s(0) = s_0$, $c(0) = 0$.

In this module we focus on modelling with differential equations. There are not the only way to formulate models, though. Arguably, all areas of mathematics provide tools to describe phenomena or processes. A few examples are:

- The n-gon is a geometric object that can be physically constructed. The dihedral group D_n is a model of some of its essential characteristics – the symmetries.
- Techniques from graph theory can be used to describe connections between participants in a social network. Such approaches can underpin models for the dynamics of opinions.
- Games and gambling involve random events. Methods from probability theory can be used to model these, usually with the aim of assessing and quantifying chances and risks.

1.2.2 Mathematical problems

The final outcome of the formulation stage is a *mathematical problem*, often also formulated in the form of a *mathematical question*. As such, it is located in the upper right box of the illustration of the modelling cycle in Figure 1.1. Whilst a mathematical model lists variables and relations between these, the problem states which variable are assume known and which variable one wishes to identify or find. These problems are the objects to which mathematical methods, tools, and techniques are applied with the aim of solving it, making predictions, and possibly also drawing other useful conclusions and thus getting insight.

There are three broad types of problems:

1. *Forward problems*, or *evaluation questions*. Given all needed information about a process or phenomenon, an we quantitatively predict its other properties and how it will function? Examples: What is the maximum attainable speed of this car? How quickly will this disease spread through the population of this city?

2. *Inverse problems*, or *detection questions*. If some information about a phenomenon is not directly available, can it be "reverse engineered" from observations?
   Examples: How can we use data from CT scans to estimate the location of a tumour? Can we determine the damping of an oscillating pendulum from the decay of its time series height data?

3. *Control problems*, or *design questions*: Can we create a solution that best meets a specified goal?
   Examples: What shape is best for a paper airplane to make it fly furthest? How should a pill be coated to release its drug at a constant rate over an entire day?

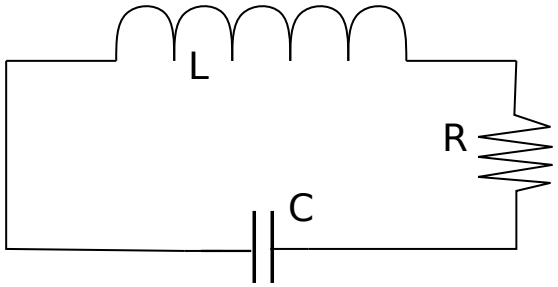


Figure 1.3: Sketch of an RLC circuit consisting of a capacitor (C), an inductor (L), and a resistor (R).

Example 1.3 (RLC circuit)**.** *Let us consider a serial electric circuit consisting of a capacitor, an inductor, and a resistor as sketched in Figure 1.3. According to Kirchhoff's law, the voltages across all elements in a closed circuit sum up to zero,*[4]

$$v_i(t) + v_r(t) + v_c(t) = 0.$$

Across the inductor, the voltage is given by $v_i(t) = L\frac{d}{dt}i(t)$*, where* $L > 0$ *is the inductance parameter and* $i(t) = \frac{d}{dt}q(t)$ *is the electric current arising from the capacitor charging or decharging. The voltage across the capacitor is given by* $v_c(t) = q(t)/C$*, where* $C > 0$ *is the capacity, a material parameter, and* $q(t)$ *is the electric charge. Finally, across the resistor, the voltage is* $v_r(t) = Ri(t)$ *with the electrical resistance* $R \geq 0$*. Substituting the formulas for the voltages we obtain the differential equation*

$$L\frac{d^2}{dt^2}q(t) + R\frac{d}{dt}q(t) + \frac{1}{C}q(t) = 0 \tag{1.3}$$

for the electric charge $q(t)$*. Let us now look at some mathematical problems centred around this differential equation and classify these.*

1. *Suppose that, initially, a charged capacitor is connected to the otherwise charge and current free elements. We want to know the charge of the capacitor at a later time.*
   This is a forward problem. Figure 1.4 displays some solutions for different regimes.

2. *Suppose we obtain a measurement such as in Figure 1.4. Knowing the capacity* C *of the capacitor, we are interested in the inductance* L *and the resistance* R*.*
   This is an inverse problem.

3. *Let us add a 'battery' to the circuit, where we assume that we can change the voltage over time. We aim for steering the circuit into a state such that the charge of the capacitor oscillates as in the undamped case (top in Figure 1.4).*
   This is a control problem.

1.2.3 Model validation

Generation of experimentally testable predictions is a hallmark of good mathematical modelling attempts. Testing against observations can be of *qualitative* nature, by which we mean that

[4]Here and in the rest of this example, all terms are considered nondimensional. We will discuss dimensions of this problem later on in Example 1.7.

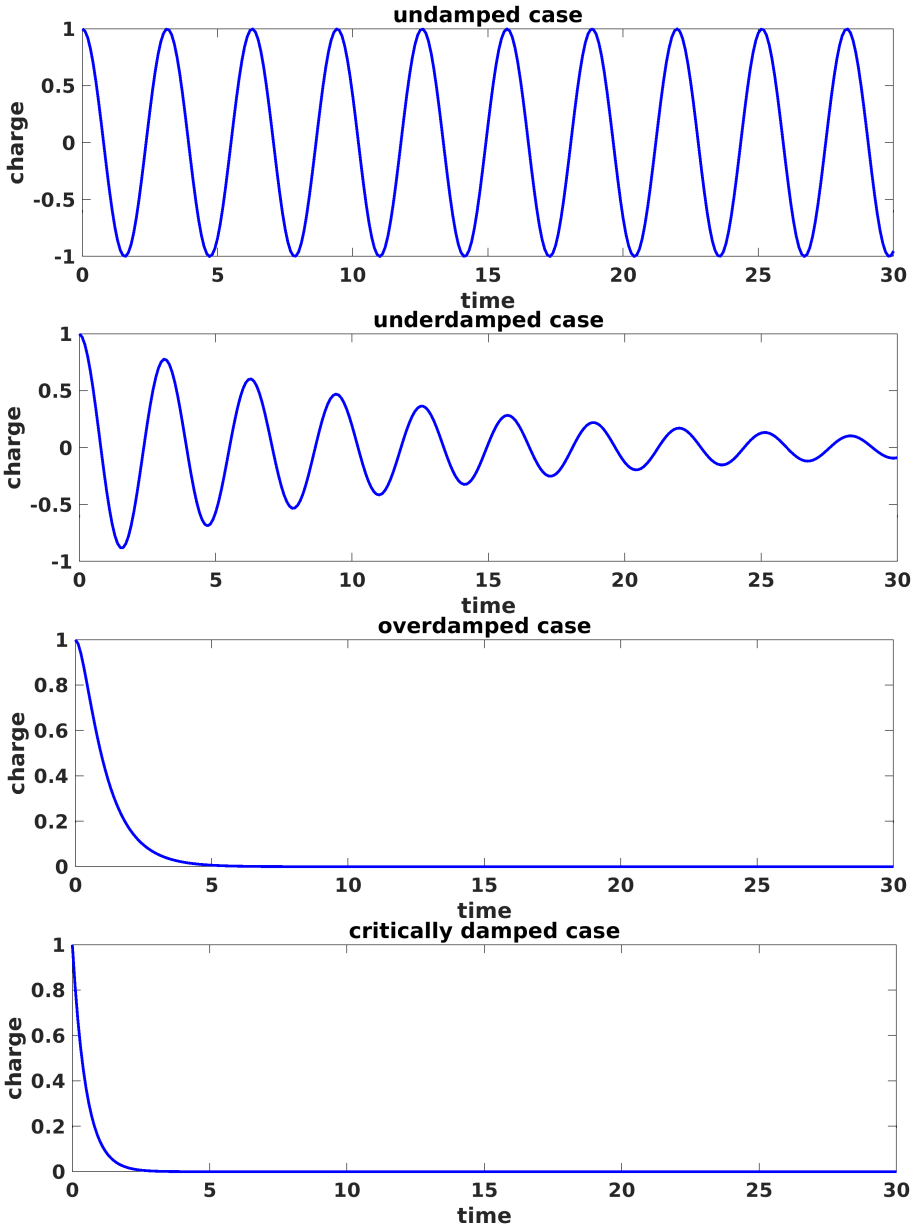


Figure 1.4: Solutions to (1.3) modelling the electric charge of the capacitor in an RLC circuit as in 1.3. The parameters are $q_0 = 1$, $C = 0.1$, $L = 2.5$. For R we chose from top to bottom $R = 0$ (undamped), $R = 0.4$ (underdamped), $R = 12$ (overdamped), and $R = 10$ (critically damped).

typical features and properties of the phenomenon are reproduced by the model. If detailed measurements and data are at hand then the model can be *quantitatively* validated by comparison with the value produced by quantitatively solving the associated problem.

Example 1.4 (Population growth)**.** *Around 1800 the English economist Thomas Malthus was interested in predicting the population and proposed that its growth is proportional to the present size. Denoting by $p(t)$ is the population as a function of time this means that $\frac{d}{dt}p(t) = kp(t)$ with some growth rate factor $k > 0$. The solutions to this differential equation are of the form $p(t) = ce^{kt}$ with some number $c > 0$, so the population grows exponentially fast without limits. Given that the resources on the planet are limited, this seemed a bit of a stretch, and indeed determining the growth rate factor k from early data seemed to consistently overestimate the population at later times.*

The Belgian mathematician Pierre Verhulst came up with the idea of limiting the population by assuming that the growth decays again then a maximal population is approached. More precisely, he proposed what is known today as the logistic equation *and reads*

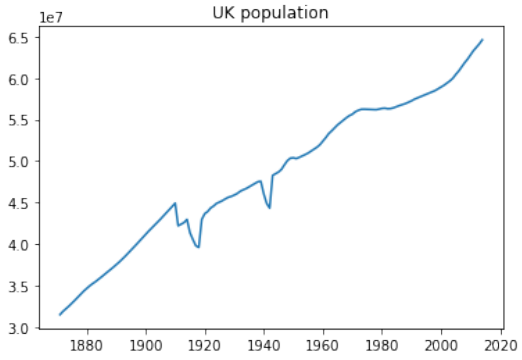
$$\frac{d}{dt}p(t) = kp(t)\Big(1 - \frac{p(t)}{p_m}\Big)$$

where $p_m > 0$ stands for the maximal population and $k > 0$ is the growth rate factor again. Note that when $p(t)$ is very small in comparison with p_m then the $1 - \frac{p}{p_m} \approx 1$ and we recover Malhus' model, the solution is then close to the exponential growth. However when $p(t)$ approaches p_m then $1 - \frac{p(t)}{p_m}$ becomes very small, so the growth $\frac{d}{dt}p(t)$ becomes small.

If we know the population at a specific time t_0, say $p(t_0) = p_0$ for a given number p_0, then there is a unique solution (we will later on see techniques to work it out):

$$p(t) = \frac{p_m p_0 e^{k(t-t_0)}}{(p_m - p_0) + p_0 e^{k(t-t_0)}}.$$

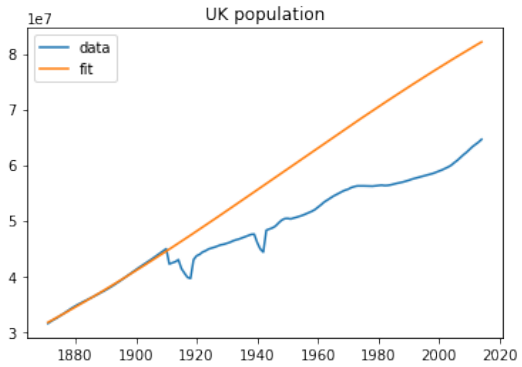
Let us now try to fit such a curve to the UK population and make a prediction. Census data from the Uk government[5] *look like this:*



We can see the impact of the two world wars but also notice a sharp drop shortly before the first one. The population of Ireland is no longer included then. This highlights a typical problem in dealing with data, but which we are not going to engage with in this module.

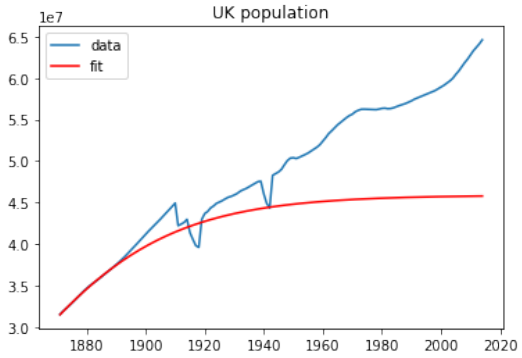
With fitting *we mean the (inverse problem of) identifying parameters k, p_0, and p_m such that the distance between the curve and the data points for specified times is minimal. If we use the data from 1871-1900 (then $t_0 = 1871$) then we obtain the following fit:*

[5] *Mid-1851 to Mid-2014 population estimates for the United Kingdom*, published by the Office for National Statistics in 2015.



This seems give a good fit until that drop due to neglecting the population of Ireland. From then the population is overestimated. However, even if we included the data of the Republic of Ireland (about 5 Mio in 2021) the fit would still significantly overestimate the population.

Interestingly, if we use the data from 1871-1890 for fitting then the population is significantly underestimated, an increase of the growth rate around 1900 when the society was prospering is not captured then:



To conclude, the model can give reasonable results for short time but a long time prediction is delicate.

Once a mathematical problem is set up, its mathematical investigation usually is an important step in understanding the original phenomenon. This is clearly true if the model does a good job at predicting the observed behaviours (a *positive outcome*). When the model does not work as expected (a *negative result*), it often still provides better insight and some understanding of which effects have significant influence on the phenomenon's behaviour, and possibly even how to further improve its accuracy of the model. In any case, however, the model is an abstraction and a restriction of reality, and it is important to keep in mind that:

Mathematical models have a limited range of validity!

Many scientists have expressed views about the limitations of models. Some notable examples are:

- Mark Kac explained *Models are, for the most part, caricatures of reality, but if they are good, they portray some features of the real world.*[6]

[6] M Kac: *Some mathematical models in science.* Science 166, 695–699 (1969).

- In one of his papers on developmental biology[7] Alan Turing wrote *This model will be a simplification and an idealisation, and consequently a falsification. It is to be hoped that the features retained for discussion are those of greatest importance ...*
- In short, *...all models are wrong, but some are useful.*[8]

1.3 Dimensional analysis

We have seen that modelling involves identifying and classifying quantities of interest. We distinguish *dependent variables*, denoted by $d = (d_1, \dots, d_N)$, and *independent variables* $i = (i_1, \dots i_M)$, which influence these. The outcome can be understood as the problem of having to find an unknown function u that tells us the values of the dependent variables for any values of the independent variables,

$$d = u(i) \quad \Leftrightarrow \quad \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix} = \begin{pmatrix} u_1(i_1, \dots i_M) \\ u_2(i_1, \dots i_M) \\ \vdots \\ u_N(i_1, \dots i_M) \end{pmatrix} \tag{1.4}$$

This function $u = (u_1, \dots u_N)$ might be the solution to a set of differential equations depending on time and some parameters, or it might be a function for which we have some data samples and further knowledge, and that we would like to learn more about. However, finding u may be very costly or complicated, particularly in the case of many independent variables.

Dimensional analysis is a technique to potentially significantly reduce the effort to solve the problem. It is based on the premises that

- all quantities and variables have *dimensions*, which we measure in (possibly various) *dimensional units*[9], and that
- the relations or laws between them do not alter when changing the units in which we measure them.

The model equations then are transformed into a *nondimensional form*, which is kind of a reference form. If the solution to the reference form is known then the solution to the original problem is simply obtained by adding the dimensions again (i.e., by transforming back). There are two essential benefits to this procedure:

1. The number of independent variables in the problem often is reduced as some variables become a number (usually one) in nondimensional form. This makes it often easier to find the solution.
2. Many specific problems reduce to the same nondimensional problem. From a mathematical point of view it therefore suffices to analyse the nondimensional problem in depth. The dimensional problems then automatically are covered, too.

Let us now explain these concepts and ideas and then apply them to some examples. For more detail we refer to the book by Barenblatt [1].

[7] A M Turing: *The chemical basis of morphogenesis.* Phil Trans. R. Soc. Lond., Series B – Biol. Sci. 237(641), 37–72 (1952).

[8] G E P Box, N R Draper: *Empirical Model-Building and Response Surfaces.* Wiley, New York (1987).

[9] Note that the distinction between dimension and dimensional units sometimes is not (or not properly) made in the literature, for instance, in [6].

1.3.1 Dimensions

Some fundamental dimensions that we will use in this module are contained in the following table:

notation	dimension	possible units
T	time	seconds, days
L	length	nanometers, feet, lightyears
M	mass	micrograms, how about elephants?
A	quantity	head count, mol
Θ	temperature	kelvin, degrees fahrenheit
Q	charge	coulomb

The dimension of a variable v is denoted by $[v]$. Dimensions of compound variables can be expressed in terms of the fundamental dimensions according to the rule that the dimension of a product equals the product of the dimensions of its factors, in mathematical terms $[v_1 v_2] = [v_1][v_2]$ for any two variables v_1, v_2. This then also helps to sort out the dimensions of derivatives of quantities by recalling that these emerge as the limit fractions of small differences. Similarly for integrals. Here are some examples:

- Consider a time dependent population $p(t)$ of bacteria. Then its derivative $\dot{p}(t)$ with respect to time has the dimension

$$[\dot{p}(t)] = \Big[\lim_{\Delta t \to 0} \frac{p(t+\Delta t) - p(t)}{\Delta t}\Big] = \lim_{\Delta t \to 0} \frac{[p(t+\Delta t) - p(t)]}{[\Delta t]} = \lim_{\Delta t \to 0} \frac{A}{T} = \frac{A}{T}.$$

- Similarly, a velocity w stands for a (possibly instantaneous) change of a position x (which is a length) in a given time interval. Hence

$$[w] = \frac{[\text{change of } x]}{[\text{change of } t]} = \frac{L}{T}.$$

- From physics you may remember that force equals mass times acceleration, $f = ma$. The latter is the derivative of the velocity with respect to time. Therefore $[a] = L/T^2$ and then $[f] = ML/T^2$.

- The change of the electric charge q of a capacitor over a time interval Δt equals the inflow, which is the time integral of the electroc current i,

$$q(t) = \int_{t-\Delta t}^{t} i(s)ds \approx i(t)\Delta t.$$

  This approximation might be crude but is sufficient for working out the dimensions of the electric current:

$$Q = [q(t)] = [i(t)\Delta t] = [i(t)][\Delta t] = [i]T \quad \Rightarrow \quad [i(t)] = \frac{Q}{T}.$$

The *principle of dimensional homogeneity* states that all terms that are summed to contribute to an equation in a model have to have the same dimension. When we have formulated a model this principle thus provides a first *sanity check* by making sure that all sums involve operands with the same dimension. A further consequence is that all arguments of more complicated

functions such as ε or sin need to be nondimensional. An easy way to see this, for instance for the exponential function, is to recall its series representation

$$\exp(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n.$$

Now $[x^n] = [x]^n$, so if x was not nondimensional then each term in the sum would have a different dimension.

1.3.2 Nondimensionalisation

Let us come back to the general problem $d = u(i)$ in (1.4). The fact that the both dependent and independent variables have dimensions restricts the possible function u. For instance, unless a variable i_m is nondimensional, a term of the form $\sin(i_m)$ doesn't make any sense. Similarly, the function $u_n(i_1, \dots, i_M)$ for d_n cannot contain the summand $i_1 i_2$ unless its dimension coincides with the dimension of d_n, i.e., unless $[i_1 i_2] = [d_n]$. Dimensional analysis is a technique to exploit this fact and to find the most general possible form for the function u, and to reduce its complexity (usually at least), i.e., the number of independent variables. It follows the following steps:

1. Write the dimensions of all variables in terms of fundamental dimensions.

2. Express all fundamental dimensions required for the dependent variables as a product of suitable independent variables. The choice made for each fundamental dimension is called a *scale*.
   Note that, often, there are several ways to do so. For finding the general form of the problem and reducing the complexity of the problem, the choice does not matter. However, when we want to analyse the reduced problem then some choices might turn out better than others. We will come back to this question in the context of *scaling*.

3. Nondimensionalise the variables.
   Consider any dependent variable d_n and let us write $D_n = [d_n]$ for its dimension. This dimension is made up of fundamental dimensions. We have expressed these in terms of the independent variables by defining suitable scales, hence $D_n = [p_n(i_1, \dots, i_M)]$ for some product p_n of some independent variables. This product is then called the *scale* for the variable d_n and usually denoted with a line of the letter in the following, $\bar{d}_n = p_n(i_1, \dots, i_M)$. Actually, it is then the scale for all variables with the same dimension. Let now $\tilde{d}_n = d_n/\bar{d}_n$. This variable has no dimension because

   $$[\tilde{d}_n] = \left[\frac{d_n}{\bar{d}_n}\right] = \frac{[d_n]}{[\bar{d}_n]} = 1.$$

   We proceed similarly with the independent variables to obtain nondimensional variables $\tilde{i}_m = i_m/\bar{i}_m$ with suitable products for the $\bar{i}_m$.

4. Write down the nondimensional problem and reduce it.
   The nondimensional problem is to find an unknown function $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_N)$ with

   $$\tilde{d} = \tilde{u}(\tilde{i}) \quad \Leftrightarrow \quad \begin{pmatrix} \tilde{d}_1 \\ \tilde{d}_2 \\ \vdots \\ \tilde{d}_N \end{pmatrix} = \begin{pmatrix} \tilde{u}_1(\tilde{i}_1, \dots \tilde{i}_M) \\ \tilde{u}_2(\tilde{i}_1, \dots \tilde{i}_M) \\ \vdots \\ \tilde{u}_N(\tilde{i}_1, \dots \tilde{i}_M) \end{pmatrix}. \tag{1.5}$$

Some nondimensional independent variables $\{\tilde{i}_1, \ldots, \tilde{i}_M\}$ in fact might not be independent any more, or they might have become a constant (typically one). We can then reduce the number of variables of $\tilde{u}$ until they are independent again.

5. Add the dimensions again by transforming the reduced equation back.
   This yields the form that the solution necessarily has.

We remark that the so-called *Buckingman-Pi Theorem* guarantees that every problem can be nondimensionalised provided that steps 3 works, i.e., all dimensions of the dependent variables can be expressed in terms of the dimensions of the independent variables[10]. Let us now study some examples.

Example 1.5 (Form of the solution to a differential equation)**.** *For a bacteria colony, the growth is proportional to the bacteria population until lack of nutrient restrict it. Denoting the population with p, time with t, and the initial population with $p_0 > 0$, we obtain*

$$\dot{p}(t) = \alpha p(t), \quad t > 0, \qquad p(0) = p_0 \tag{1.6}$$

where $\alpha > 0$ is a growth rate factor. We assume that p_0, and α are known and consider the forward problem of predicting the population p over time.
We now want to use dimensional analysis to learn something about the possible forms of the solution. The dependent variable here is p, and the independent variables t, α, and p_0, so we can write the problem abstractly in the form

$$p = u(t, \alpha, p_0).$$

Following the steps we first find the dimensions of the variables. We easily get that $[p] = [p_0] = A$ and $[t] = T$. For α we recall that $[\dot{p}] = A/T$, hence, using the differential equation,

$$\frac{A}{T} = [\dot{p}(t)] = [\alpha p(t)] = [\alpha][p(t)] = [\alpha]A \quad \Leftrightarrow \quad [\alpha] = \frac{1}{T}.$$

Next, we express the fundamental dimensions in terms of the independent variables. The initial populaton p_0 is the only independent variable that involves the dimension A, so we choose $\overline{p} = p_0$ as a scale for the fundamental dimension quantity, A. For the fundamental dimension time T we can choose t or $1/\alpha$. Let us choose $\overline{t} = 1/\alpha$, the other case is discussed below.

Let us proceed with the next step and introduce nondimensional variables. We have to divide p by a product of variable that yields the same dimension, which here is A. For this dimension we had chosen the scale $\overline{p} = p_0$. Similarly for the initial population, so we obtain the nondimensional population and the nondimensional initial population

$$\tilde{p} = \frac{p}{\overline{p}} = \frac{p}{p_0}, \quad \tilde{p}_0 = \frac{p_0}{\overline{p}} = \frac{p_0}{p_0} = 1.$$

Note that the nondimensional initial population is just one, so we will not need to consider it in the nondimensional problem. Regarding the other variables, recall that α has the dimension of inverse time, for which $1/\overline{t}$ then is a scale. So we obtain the nondimensional time and the nondimensional growth rate factor

$$\tilde{t} = \frac{t}{\overline{t}} = \frac{t}{1/\alpha} = \alpha t, \quad \tilde{\alpha} = \frac{\alpha}{1/\overline{t}} = \frac{\alpha}{\alpha} = 1.$$

[10] Proving it requires tools beyond the scope of this module.

The nondimensional problem therefore is of the form

$$\tilde{p} = \tilde{u}(\tilde{t}, \tilde{\alpha}, \tilde{p}_0),$$

however $\tilde{\alpha} = 1$ *and* $\tilde{p}_0 = 1$*, hence it reduces to*

$$\tilde{p} = \tilde{u}(\tilde{t}).$$

Transforming back by substituting the nondimensional variables with the dimensional ones again yields that

$$\frac{p}{p_0} = \tilde{p} = \tilde{u}(\tilde{t}) = \tilde{u}(\alpha t).$$

Altogether, the dimensional analysis yields that the solution has to be the form

$$p = p_0 \tilde{u}(\alpha t).$$

Or maybe not? We made a choice about how to express the fundamental dimensions in terms of the independent variables. For instance, we could have chosen $\bar{t} = t$ *instead of* $\bar{t} = 1/\alpha$*. In that case,* $\tilde{p} = p/p_0$ *and* $\tilde{p}_0 = 1$ *are as before, but we obtain*

$$\tilde{t} = \frac{t}{\bar{t}} = \frac{t}{t} = 1, \quad \tilde{\alpha} = \frac{\alpha}{1/\bar{t}} = \frac{\alpha}{1/t} = \alpha t$$

as $1/t$ *then has dimension* $1/T$ *as it is required for the nondimensional growth rate factor. We now obtain the nondimensional problem*

$$\tilde{p} = \tilde{u}(\tilde{t}, \tilde{\alpha}, \tilde{p}_0) = \tilde{u}(\tilde{\alpha}).$$

However, if we transform back then we obtain the same form for the solution as before:

$$\frac{p}{p_0} = \tilde{p} = \tilde{u}(\tilde{\alpha}) = \tilde{u}(\alpha t) \quad \Rightarrow \quad p = p_0 \tilde{u}(\alpha t).$$

Now, in this case we can fairly easily find the solution to the initial value problem (1.6)*. It is given by*

$$p(t) = p_0 e^{\alpha t}.$$

Observe that it is indeed of the required form that we have found. We see this by setting $\tilde{u}() = \exp()$*.*

Often, it is not possible to simulate or duplicate a phenomenon in a laboratory, for instance, because they are too expensive or too dangerous. Dimensional analysis can help to correctly rescale or otherwise transform the phenomenon, and thus to make it accessible to experimental measurements as the are required to develop and validate a model. Let us look at a specific example.

Example 1.6 (Similitude)**.** *In 2012, NASA's Mars rover Curiosity reached its destination. During the decent, a parachute was used to bring down its velocity. This was possible because Mars has an atmosphere, which is different from Earth's, however. Experiments and rehearsals on Mars were out of reach, so the engineers had to set them up on Earth. For this purpose they had to make sure that the experience and data they collected were meaningful to ensure that everything would work on Mars. As an example for one of the problems they had to solve we look at the size of the parachute.*

We are interested in the diameter d of the (spherical) parachute canopy and assume that it depends on the desired terminal velocity v of the probe, its mass m, the acceleration g by gravity, and the atmospheric density ρ. The abstract problem, to which we will apply dimensional analysis, reads

$$d = u(v, m, g, \rho).$$

The dimensions of the variable are

$$[d] = L, \quad [v] = \frac{L}{T}, \quad [m] = M, \quad [g] = \frac{L}{T^2}, \quad [\rho] = \frac{M}{L^3}.$$

The fundamental dimensions L, T, M appear and, using the independent variable v, m, g, and ρ can be expressed as

$$M = [m], \quad L = \Big[\Big(\frac{m}{\rho}\Big)^{1/3}\Big], \quad T = \Big[\frac{v}{g}\Big].$$

With other words, for these, we use the scales

$$\overline{m} = m, \quad \bar{l} = \Big(\frac{m}{\rho}\Big)^{1/3}, \quad \bar{t} = \frac{v}{g}.$$

Let us now use these combinations to nondimensionalise all variables. For d we obtain the scale $\overline{d} = \bar{l} = (m/\rho)^{1/3}$, for m we use $\overline{m} = m$, and for ρ we get $\overline{\rho} = \overline{m}/\bar{l}^3 = m/((m/\rho)^{1/3})^3 = \rho$. The corresponding nondimensional variables then are

$$\tilde{d} = \frac{d}{\overline{d}} = \frac{d}{(m/\rho)^{1/3}} = \Big(\frac{\rho}{m}\Big)^{1/3} d, \quad \tilde{m} = \frac{m}{\overline{m}} = \frac{m}{m} = 1, \quad \tilde{\rho} = \frac{\rho}{\overline{\rho}} = \frac{\rho}{\rho} = 1.$$

From the expressions for the fundamental units we also deduce the scales $\overline{v} = \bar{l}/\bar{t} = (\frac{m}{\rho})^{1/3}/(v/g)$ for the velocity and $\overline{g} = \bar{l}/\bar{t}^2 = (\frac{m}{\rho})^{1/3}/(v/g)^2$ for the acceleration. This yields that

$$\tilde{v} = \frac{v}{\overline{v}} = \frac{v}{(\frac{m}{\rho})^{1/3}/(v/g)} = \Big(\frac{\rho}{m}\Big)^{1/3}\frac{v^2}{g}, \quad \tilde{g} = \frac{g}{\overline{g}} = \frac{g}{(\frac{m}{\rho})^{1/3}/(v/g)^2} = \Big(\frac{\rho}{m}\Big)^{1/3}\frac{v^2}{g}.$$

Observe that these are the same and, thus, not independent! In the nondimensional problem we therefore need to account for this variable only once. Also dropping the nondimensional quantities that become one, the nondimensional problem reads and reduces to

$$\tilde{d} = \tilde{u}(\tilde{v}, \tilde{m}, \tilde{g}, \tilde{\rho}) = \tilde{u}(\tilde{v}).$$

This equation can now be used both on Earth and on Mars. For both planets we know g and ρ, and given m and v we can transform the nondimensional problem to

$$d = \Big(\frac{m}{\rho}\Big)^{1/3}\tilde{d} = \Big(\frac{m}{\rho}\Big)^{1/3}\tilde{u}(\tilde{v}) = \Big(\frac{m}{\rho}\Big)^{1/3}\tilde{u}\Big(\Big(\frac{\rho}{m}\Big)^{1/3}\frac{v^2}{g}\Big).$$

on each planet.

It remains to find the function $\tilde{u}$. For this purpose, we can now perform experiments and measurements on Earth. Having found the function $\tilde{u}$ we simply evaluate it with the atmospheric data for Mars and then get the required canopy diameter d.

1.3.3 Nondimensionalising differential equations

As we mainly work with differential equations in this module we have another look at their nondimensionalisation. Generally, the procedure is the same as for any model and as described in Section 1.3.2. We have also seen in Example 1.5 how the procedure is applied to a differential equation. Recall that the outcome, i.e., the nondimensional problem, is not unique, in general, because we can usually can choose different scales for the dimensions of relevance in the problem. In Example 1.5 we looked into two cases. First, we chose scales such that the growth rate factor became one, $\tilde{\alpha} = 1$. In the second scaling the time became one, $\tilde{t} = 1$.

In the context of differential equation we aim for deriving nondimensional differential equations of the same structure. We want to keep derivatives with respect to independent variables. We will therefore try to select scales with variables that act as parameters in the differential equation, rather than with dependent or independent variable that feature in derivatives. For instance, in the population growth model, the time appears in the derivative of the differential equation, and we are interested in deriving a nondimensional differential equation for a function of a nondimensional time. The first scaling discussed in Example 1.5 is therefore more natural as it leads to a differential equation in the nondimensional time $\tilde{t}$.

Two principles provide guidance to nondimensionalise differential equations:

1. Select scales such that as many parameters as possible are normalised to become one.
   With parameters we here mean independent variables that do not appear in any derivatives of the differential equation. These can be prefactors or also initial values if the full problem comes with information on an initial state.

2. Select scales such that all nondimensional parameters are 'small' and do not become infinity when certain variables become small or big.
   In the context of differential equations, the behaviour may be of interest as some parameters become small or big. The principle means that the scaling should be such that any nondimensional parameters become small in that case.

We will now study another example with a differential equation but focus on the first principle. The second one will be picked up later on.

Example 1.7 (Nondimensionaling the RLC circuit model)**.** *Recall the differential equation* (1.3) *modelling the voltages across the elements of an electric circuit from Example 1.3:*

$$\overline{L}\frac{d^2}{dt^2}q(t) + R\frac{d}{dt}q(t) + \frac{1}{C}q(t) = 0.$$

We have denoted the inductance with $\overline{L}$ to distinguish it from the dimension length, which we denoted with L.

Let us consider the problem centred around Figure 1.4: we connect a capacitor with an initial charge q_0 with the other elements and are interested in predicting its charge as a function of time. Our goal is now to derive a nondimensional differential equation for the (nondimensional) charge as a function of (nondimensional) time and then studying how changes of the (nondimensional) resistance affect the solution.

Following the nondimensionalisation procedures and the first guidance we would now (try to) write all fundamental dimensions in terms of the dimensions of the parameters $\overline{L}$, R, C, and q_0. In practice, one often proceeds differently for differential equations.

Let $\overline{q}$ and $\overline{t}$ denote some scales for charge and time. These have dimensions $[\overline{q}] = Q$ and $[\overline{t}] = T$ but are not determined yet in terms of the parameters, we will do that later. With these

scales, $\tilde{q} = q/\overline{q}$ and $\tilde{t} = t/\overline{t}$ are a nondimensional charge and time, respectively. We now try to derive a differential equation for $\tilde{q}$ as a function of the nondimensional time $\tilde{t}$. Starting with $\tilde{q}(\tilde{t}) = q(t)/\overline{q} = q(\widetilde{t}\overline{t})/\overline{q}$ we note that

$$\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) = \frac{1}{\overline{q}}\frac{d}{d\tilde{t}}q(\widetilde{t}\overline{t}) = \frac{1}{\overline{q}}q'(\widetilde{t}\overline{t})\overline{t} = \frac{\overline{t}}{\overline{q}}\frac{d}{dt}q(t)$$

where we applied the chain rule. Similarly, $\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) = \frac{\overline{t}^2}{\overline{q}}\frac{d^2}{dt^2}q(t)$. Substituting these into the differential equation we obtain that

$$\frac{\overline{L}\overline{q}}{\overline{t}^2}\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) + \frac{R\overline{q}}{\overline{t}}\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) + \frac{\overline{q}}{C}\tilde{q}(\tilde{t}) = 0. \tag{1.7}$$

Initially, we have that $\tilde{q}(\tilde{0}) = q(0)/\overline{q}$.

The initial value gives an idea how to choose the scale for the charge: $\overline{q} = q_0$. As we are not intending to vary this parameter, we thus ensure that is becomes one in the nondimensional problem. We initially now have that $\tilde{q}(\tilde{0}) = 1$. We still can choose the scale for the time and possibly 'kill' another parameter. We don't want to do so with the resistance (the factor before the first derivative) because we will want to vary it. However, we can try to make the factor before the second derivative or the factor of $\tilde{q}$ become one. Let us do the former and, for this purpose, divide the equation by $\frac{\overline{L}\overline{q}}{\overline{t}^2}$. This yields

$$\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) + \frac{R\overline{t}}{\overline{L}}\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) + \frac{\overline{t}^2}{\overline{L}C}\tilde{q}(\tilde{t}) = 0.$$

The factor in from of $\tilde{q}$ becomes one if we choose $\overline{t} = \sqrt{\overline{L}C}$, and then the nondimensional differential equation becomes

$$\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) + R\sqrt{\frac{C}{\overline{L}}}\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) + \tilde{q}(\tilde{t}) = 0.$$

A couple of remarks:

1. *A nondimensional resistance $\tilde{R} = R\sqrt{C/\overline{L}}$ indeed is present, and we can now try to solve the problem and study its dependence on $\tilde{R}$.*

2. *We could have divided by $\overline{q}/C$ in (1.7). We then would have got the prefactor $\overline{L}C/\overline{t}^2$ in front of $\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t})$. We see that it becomes one with the same choice of the time scale $\overline{t} = \sqrt{\overline{L}C}$.*

3. *Actually, is $\overline{t}$ a time scale? Let us briefly check.*
   Voltages are measured in volt = joule / coloumb and have the dimension of energy per charge, $ML^2/(T^2Q)$ in fundamental units. As $[\frac{d}{dt}q] = Q/T$ and $[\frac{d^2}{dt^2}q] = Q/T^2$ we see that the parameters have dimensions

   $$[\overline{L}] = \frac{ML^2}{Q^2}, \quad [R] = \frac{ML^2}{Q^2T}, \quad [C] = \frac{Q^2T^2}{ML^2}, \quad [q_0] = Q.$$

   We see that $[\overline{L}C] = T^2$ so that $\overline{t}$ indeed is a time scale, and one can also work out that $\tilde{R}$ indeed is nondimensional.

4. *We advocated choosing scales such that parameters become one in the nondimensional differential equation problem. This it no necessary, though, any constants are fine. Indeed, in Figure 1.4 the nondimensional values for the inductance and capacity are* $L = 2.5$ *and* $C = 0.1$*, respectively (note that dimensions are not discussed in Example 1.3, consider everything as nondimensional there).*

Chapter 2

First order equations

Because of their fundamental importance and success as a modelling approach in many fields, differential equations have become a field within the mathematical sciences and are an integral part of training in mathematics. One fundamental question is whether a differential equation has solutions and, if so, whether the solution is unique. We will see some counterexamples showing that this is not always the case but have to delay deeper results on this questions because they require sophisticated analytical tools that we don't have to our disposition yet.

If one can show that there is a solution, a natural question is whether we can find an explicit formula for it. We will study some techniques that enable us to do so for some simple differential equations. However, in general, this is not the case, and for this reason computational methods have been developed that approximate the solution. We will occasionally make use of the computer but the in-depth study of such methods is left to future modules.

Even if we cannot find an explicit solution formula, we still may be able to make some qualitative predictions about the behaviour of the solution. For time dependent problems this comprises the long-term behaviour, for instance, whether the solution grows, or remains bounded, or oscillates, or converges to an equilibrium state. We will also look a bit into so-called stability properties, i.e., how this qualitative behaviour changes if data or parameters in the differential equation change. Such questions will also be picked up in follow-up modules.

There are a couple of more aspects around differential equations that are not covered in this module but for which there are modules in later study years. Most notably, we will focus on forward problems and rarely consider inverse or control problems.

2.1 General concepts

General scalar differential equations can be written in the form

$$\mathcal{F}\big(t, x(t), x'(t), x''(t), \dots, x^{(k-1)}(t), x^{(k)}(t)\big) = 0 \tag{2.1}$$

for some given function $\mathcal{F} : (\alpha, \beta) \times \mathbb{R}^{k+1} \to \mathbb{R}$ with an interval $(\alpha, \beta) \subset \mathbb{R}$. We here use the notation

$$x^{(l)}(t) = \frac{d^l}{dt^l} x(t)$$

for the l-th derivative of a function $x : (\alpha, \beta) \to \mathbb{R}$. In the general form (2.1) we make the implicit assumption that the highest derivative $x^{(k)}$ occurs in the equation (otherwise we could consider a function F that does not depend on this argument). We will focus on equations where

we can the equation for $x^{(k)}$ and write it in the form

$$x^{(k)}(t) = \phi\big(t, x(t), \dots, x^{(k-1)}(t)\big) \tag{2.2}$$

for some function $\phi : (\alpha, \beta) \times \mathbb{R}^k \to \mathbb{R}$.

The equation is expressed in terms of the variables t (independent) and x (dependent). The variables are kind of placeholders and used to express the differential equation. We could have used other letters, for instance,

$$z^{(k)}(\theta) = \phi\big(\theta, z(\theta), \dots, z^{(k-1)}(\theta)\big) \quad \text{and} \quad y^{(k)}(x) = \phi\big(x, y(x), \dots, y^{(k-1)}(x)\big)$$

express the same differential equation (2.2). Note that x stands for the independent variable in the last reformulation.

In the literature, the variable t is often stands for time and x, y, z are spatial variables. Derivatives with respect to time are often denotes using dots,

$$\frac{d}{dt}z(t) = \dot{z}(t), \quad \frac{d^2}{dt^2}z(t) = \ddot{z}(t), \quad \dots$$

Spatial derivatives often use dashes,

$$\frac{d}{dx}y(x) = y'(x), \quad \frac{d^2}{dt^2}y(x) = y''(x), \quad \dots$$

Definition 2.1. *For a differential equation of the form (2.1) we define:*

1. *The* order *of a differential equation is the order of its highest derivative.*
2. *The differential equation is* autonomous *if $\mathcal{F}$ does not explicitly depend on the independent variable, so $\mathcal{F} : \mathbb{R}^{k+1} \to \mathbb{R}$, and the differential equation reads*
$$\mathcal{F}\big(x(t), x'(t), \dots, x^{(k)}(t)\big) = 0.$$
3. *The differential equation is* linear *if $\mathcal{F}$ can be written in the form*
$$\mathcal{F}\big(t, x(t), x'(t), \dots, x^{(k)}(t)\big) = s(t) + \sum_{i=0}^{k} a_i(t)x^{(i)}(t) \tag{2.3}$$
with some functions $s, a_i : (\alpha, \beta) \to \mathbb{R}$, $i = 0, \dots, k$.
4. *Let s be the function obtained if $x(t) = x'(t) = x^{(k)}(t) = 0$ in (2.1),*
$$s(t) = \mathcal{F}(t, 0, \dots, 0), \quad t \in (\alpha, \beta).$$
The equation (2.1) is called homogeneous *if $s = 0$.*
Otherwise, it is called inhomogeneous.

A couple of remarks on the above formal definitions:

- Note that the functions s defined in the last two items coincide if $\mathcal{F}$ is linear (which also explains the same notation).
- If the differential equation is autonomous then s cannot depend on the independent variable t but can be a constant, so the equation still can be inhomogeneous.

- (PDEs and ODEs) There are even more general differential equations, in particular for functions of several variables. For instance, the equation

$$\frac{\partial^2}{\partial t^2}u(y,t) - \frac{\partial^2}{\partial y^2}u(y,t) = 0$$

is a simple model for the propogation of waves. Such equations that involve partial derivatives with respect to multiple (more than one) independent variable are called *partial differential equations (PDE)*.

In this module, we focus on differential equations for functions of one independent variable only. These are called *ordinary differential equations (ODE)*.

Example. *Let us briefly classify some of the differential equations:*

1. *Recall* population growth *from the introduction:*

$$\dot{p}(t) = kp(t)\Big(1 - \frac{p(t)}{p_m}\Big).$$

We can write it in the form (2.1) *with*

$$\mathcal{F}(p(t), \dot{p}(t)) = \dot{p}(t) - kp(t)\Big(1 - \frac{p(t)}{p_m}\Big).$$

Its order is one as only the first derivative occurs. The independent variable is t, and the dependent variable p is the only function that depends on it. Therefore, the equation is autonomous. On the right-hand-side the quadratic term $p(t)^2$ occurs, so the equation is non-linear. It is homogeneous as there is no additive constant.

2. *In the introduction we have also seen an equation for an electric circuit* (1.3)*:*

$$L\frac{d^2}{dt^2}q(t) + R\frac{d}{dt}q(t) + \frac{1}{C}q(t) = S\sin(\omega t),$$

where we have added a term on the right-hand-side with a number $S \neq 0$ and a frequency $\omega \neq 0$. This additional term may model the effect of an oscillating voltage source. Writing it in the form (2.1) *is fairly straightforward using the function*

$$\mathcal{F}(t, q(t), \dot{q}(t), \ddot{q}(t)) = L\ddot{q}(t) + R\dot{q}(t) + \frac{1}{C}q(t) - S\sin(\omega t).$$

We see that the second derivative is the highest, so the equation is of order two. The independent variable is t and explicitly appears in the term on the right-hand-side. Therefore, the equation is non-autonomous. We can write

$$\mathcal{F}(t, q(t), \dot{q}(t), \ddot{q}(t)) = s(t) + \sum_{i=0}^{2} a_i(t)\frac{d^i}{dt^i}q(t)$$

with $a_2 = L$, $a_1 = R$, $a_0 = \frac{1}{C}$, and $s(t) = -S\sin(\omega t)$. The equation therefore is linear. As $s(t) = \mathcal{F}(t, 0, 0, 0) = -S\sin(\omega t)$ does not always vanish the equation is inhomogeneous.

Definition 2.2. *A function $x : (\alpha, \beta) \to \mathbb{R}$ is called a solution to the differential equation* (2.1) *if it satisfies the equation.*

- *The solution is called an* explicit solution*, if the dependent variable is given in terms of the independent variable as a combination of algebraic expressions or elementary functions.*
- *An* implicit form *of the solution is an equation that relates the dependent and independent variables and involves no derivatives.*

Example. *The function* $x(t) = 12e^{-2t}$ *is an explicit solution to the differential equation* $\frac{d}{dt}x(t) = -2x(t)$.
The expression

$$\log(x) - x + 4\log(t) - 2t + 4 = 0 \tag{2.4}$$

relates x *and* t *but cannot be algebraically resolved for* x *as a function of* t*. However, assuming that* $x = x(t) \neq 0$ *is a differentiable function, then differentiating the expression* (2.4) *with respect to* t *yields that*

$$\Big(\frac{1}{x(t)} - 1\Big)\frac{d}{dt}x(t) = 2 - 4\frac{1}{t} \quad \Rightarrow \quad \frac{d}{dt}x(t) = \frac{x(t)}{1 - x(t)}\Big(2 - 4\frac{1}{t}\Big),$$

and (2.4) *is an implicit solution to this differential equation.*

This distinction of a variable to formulate a differential equation from a solution might look a bit nit-picky but is similar to the careful distinction between a function, say g, and the expression $g(y)$, which is the value of g in a point y. In practice (and in the following), however, this distinction often is not made. In a slight abuse of notation, $x(t)$ sometimes is just a dependent variable used to formulate a differential equation such as (2.1), and sometimes it denotes a solution, depending on the context.

In the remainder of this chapter we study differential equations for scalar functions that involve first derivatives. The differential equations are of the form

$$\frac{d}{dt}x(t) = f\big(x(t), t\big), \quad t \in (\alpha, \beta), \tag{2.5}$$

where $f : \mathbb{R} \times (\alpha, \beta) \to \mathbb{R}$ with an interval $(\alpha, \beta) \subset \mathbb{R}$ is a given function.

2.2 Trivial differential equations

One can regard solving a differential equation of the form (2.5) as the opposite to differentiating a given function. In differentiation calculus, we are given the function and want to find the derivative, or slope of the graph of the function. In differential equations, we know these slopes and want to find (graphs of) functions that fit them. In the literature, solving differential equations is therefore often called *integrating differential equations.*

This connection is easiest seen when considering the case that the given function f in (2.5) is independent of $x(t)$ so that the differential equation reads

$$\frac{d}{dt}x(t) = f(t), \quad t \in (\alpha, \beta). \tag{2.6}$$

This case is called *trivial* for the following reason. Recall this fundamental result from analysis:

Theorem 2.3 (Fundamental Theorem of Calculus)**.** *Suppose* $g : [a, b] \to \mathbb{R}$ *is continuous and let*

$$G(x) := \int_a^x g(\hat{x})d\hat{x}, \quad x \in [a, b].$$

Then G is an anti-derivative *of g, i.e., it satisfies*

$$\frac{d}{dx}G(x) = g(x), \quad x \in (a,b).$$

Moreover, if $\tilde{G}$ is any other anti-derivative *of g then*

$$\int_a^b g(\hat{x})d\hat{x} = \tilde{G}(b) - \tilde{G}(a),$$

and the two anti-derivatives of g differ by a constant only, so for some $c \in \mathbb{R}$

$$\tilde{G}(x) = G(x) + c, \quad x \in [a,b].$$

We see that solving (2.6) amounts to finding any anti-derivative of f. Denoting if by F, the theorem says that then $F + c$ for any $c \in \mathbb{R}$ is an anti-derivative and, thus, a solution again.

Definition 2.4 (General solution, particular solution)**.** *The* general solution *to the trivial differential equation* (2.6) *is given by*

$$x(t) = F(t) + c, \quad t \in (\alpha, \beta)$$

where F is an anti-derivative of f and $c \in \mathbb{R}$ is any number.
For a specific number c, $x(t) = F(t) + c$ is a particular solution*.*

So the general solution in fact is a whole family of solutions parametrised by c whilst the particular solution is a unique function. To single out a solution (or to determine a number c), an additional condition is required. In evolution problems where t stands for time an *initial condition* can be imposed.

Definition 2.5 (Initial value problem)**.** *The* initial value problem *consists of finding a function $x : (\alpha, \beta) \to \mathbb{R}$ such that the differential equation* (2.5) *and the* initial condition

$$x(t_0) = x_0 \tag{2.7}$$

for some $t_0 \in [\alpha, \beta]$ and some given number $x_0 \in \mathbb{R}$ are satisfied.

For the differential equation (2.6) with such an initial condition (2.7) the anti-derivative

$$x(t) = x_0 + \int_{t_0}^t f(\tau)d\tau \tag{2.8}$$

then is the solution: indeed, as x_0 does not depend on t, it satisfies the differential equation because

$$\frac{d}{dt}x(t) = \frac{d}{dt}\Big(\int_{t_0}^t f(\tau)d\tau\Big) = f(t)$$

by the *Fundamental Theorem of Calculus*, and it also satisfies the initial condition because

$$x(t_0) = x_0 + \underbrace{\int_{t_0}^{t_0} f(\tau)d\tau}_{=0} = x_0.$$

Note that there is no unknown constant c any more in the solution, this is a particular solution.

Example (Spacecraft)**.** *For a spacecraft landing on a planet (or moon) Newton's equations of the two-body problem can be simplified by neglecting the motion of the planet. We only consider the spacecraft's height $h(t)$ and its vertical velocity $v(t) = \dot{h}(t)$, both depending on time t. We assume that the spacecraft is propelled by rockets, their thrust is another force acting on it. Newton's law that mass times acceleration equals the force then can be written in the form*

$$M\dot{v}(t) = -gM + u(t) \quad \Leftrightarrow \quad \dot{v}(t) = -g + \tfrac{1}{M}u(t)$$

where $M > 0$ is the mass (assumed constant here), $g > 0$ is the gravitation constant, and $u(t)$ is a given function modelling the thrust by the spacecraft's rockets. We here assume for simplicity that $u(t) = \bar{u} \in \mathbb{R}$ is constant.

Assume now that, initially at time $t = t_0 = 0$, the velocity is known and given by $v_0 < 0$. We want to find out when the spacecraft comes to rest. In mathematical terms, we want to find a time $t^* > 0$ such that the solution to the initial value problem satisfies $v(t^*) = 0$.

Using the above result (2.8), the solution is given by

$$v(t) = v_0 + \int_{t_0}^{t} \big(-g + \tfrac{1}{M}u(\tau)\big)d\tau = v_0 + \int_0^t \big(-g + \tfrac{1}{M}\bar{u}\big)d\tau = v_0 + \big(-g + \tfrac{1}{M}\bar{u}\big)t.$$

The spacecraft comes to rest if $0 = v(t^*) = v_0 + (-g + \frac{1}{M}\bar{u})t^*$, which is the case if and only if

$$t^* = \frac{v_0}{g - \frac{1}{M}\bar{u}}.$$

We want that t^* is positive. As $v_0 < 0$ this requires that $g - \frac{1}{M}\bar{u} < 0$, or the thrust $\bar{u} > Mg$ has to be big enough. Poor spacecraft otherwise...

We may ask at which height the spacecraft comes to rest (assuming it does so). To formulate this questions in mathematical terms, let us assume that the initial height at time $t = 0$ is $h(0) = h_0 > 0$, and we are interested in the value of $h(t^*)$ where h satisfies the differential equation $\dot{h}(t) = v(t)$. As we know v already this is again an initial value problem with a trivial differential equation. Its solution is given by

$$h(t) = h_0 + \int_{t_0}^{t} v(\tau)d\tau = h_0 + \int_0^t \Big(v_0 + \big(-g + \tfrac{1}{M}\bar{u}\big)\tau\Big)d\tau = h_0 + v_0 t + \frac{1}{2}\big(-g + \tfrac{1}{M}\bar{u}\big)t^2.$$

At time $t^* = \frac{v_0}{g - \frac{1}{M}\bar{u}}$ the height then is

$$h(t^*) = h_0 + \frac{v_0^2}{g - \frac{1}{M}\bar{u}} + \frac{1}{2}\frac{(-g + \frac{1}{M}\bar{u})v_0^2}{(g - \frac{1}{M}\bar{u})^2} = h_0 + \frac{1}{2}\frac{v_0^2}{g - \frac{1}{M}\bar{u}}.$$

We may also want that the spacecraft touches down at this time, i.e., $h(t^*) = 0$. We see that this is the case if and only if

$$2h_0 = \frac{v_0^2}{\frac{1}{M}\bar{u} - g} \quad \Leftrightarrow \quad \bar{u} = M\Big(g + \frac{v_0^2}{2h_0}\Big).$$

This is the thrust required for a smooth landing.

2.3 Autonomous equations

Let us now consider the special case that the given function f in (2.5) is independent of t so that the differential equation reads

$$\frac{d}{dt}x(t) = f(x(t)), \quad t \in (\alpha, \beta). \tag{2.9}$$

Note that this is now an autonomous equation. These equations can be very challenging to integrate, and no method might produce an explicit solution. However, there are tools that give us some insight into the qualitative behaviour of solutions.

2.3.1 Stationary points and stability

First, observe that zeros of f yield constant solutions to (2.9). Indeed, if $f(x^*) = 0$ then $x(t) = x^*$ for all t is a solution. Such solutions are called *stationary* or *equilibrium points*. An interesting question now is whether x^* is a *stable* point. For that the following investigation makes sense we assume that $\beta = \infty$.

Definition 2.6. *A stationary point x^* of an autonomous differential equation of the form (2.9) is called* stable *if nearby solutions remain close as the independent variable grows. In mathematical terms, the distance* $\operatorname{dist}(x(t), x^*)$[1] *remains bounded as $t \to \infty$.*

Consider first the case that $f(x) < 0$ if $x < x^*$ and that $f(x) > 0$ if $x > x^*$, where x are points near the x^*, say, $x \in [x^* - \varepsilon, x^* + \varepsilon]$ for some small $\varepsilon > 0$. Suppose now that x is a solution to (2.9) that starts at a point $x(t_0) < x^*$. Then $\dot{x}(t_0) = f(x(t_0)) < 0$, so the solution will decay and move further away from x^*. Using the same argument, this will continue for $t > t_0$ as long as $x(t) \in [x^* - \varepsilon, x]$, and ultimately $x(t)$ will leave the small interval. Similarly, if $x(t_0) > x^*$. Then $\dot{x}(t_0) = f(x(t_0)) > 0$, so the solution will increase and move away from x^* and finally end up outside of $[x^* - \varepsilon, x^* + \varepsilon]$. We call such a stationary point *unstable*.

Consider now the opposite case that $f(x) > 0$ if $x < x^*$ and that $f(x) < 0$ if $x > x^*$, $x \in [x^* - \varepsilon, x^* + \varepsilon]$ for some small $\varepsilon > 0$. Then solutions to (2.9) that start at $x(t_0) < x^*$ increase as t increases, whilst those that start at $x(t_0) > x^*$ will decay, so the solutions will move towards x^*. In this case, the stationary point x^* is called *stable*.

If the function f in (2.9) is differentiable then the sign of $f'(x^*)$ can be used as a criterion to decide on the stability of the point. If $f'(x^*) > 0$ then f is increasing locally around x^*, i.e., there is small $\varepsilon > 0$ such that $f(x) > f(x^*)$ for all $x \in (x^*, x^* + \varepsilon]$ and $f(x) < f(x^*)$ for all $x \in [x^* - \varepsilon, x^*))$. As $f(x^*) = 0$ the conditions for instability are satisfied. Similarly, if $f'(x^*) < 0$ then f is decreasing locally around x^* (there is a small $\varepsilon > 0$ such that $f(x) < f(x^*)$ for all $x \in (x^*, x^* + \varepsilon]$ and $f(x) > f(x^*)$ for all $x \in [x^* - \varepsilon, x^*))$, and we see that x^* then is stable. In summary:

$$\begin{aligned} f(x^*) = 0,\ f'(x^*) > 0: &\quad x^* \text{ is unstable,} \\ f(x^*) = 0,\ f'(x^*) < 0: &\quad x^* \text{ is stable.} \end{aligned}$$

The connection with the derivative will later on be studied again in greater detail in the context of linearisation around stationary points of systems of differential equations (see Section 4.2.3).

A *phase line* is a popular way to illustrate the dynamics of an automonous differential equation such as (2.9). This is a (usually horizontal) line for the points $x \in \mathbb{R}$ with a couple of arrows to

[1] There are multiple ways to measure the distance, for instance, we can take the absolute value of the difference, $|x(t) - x^*|$.

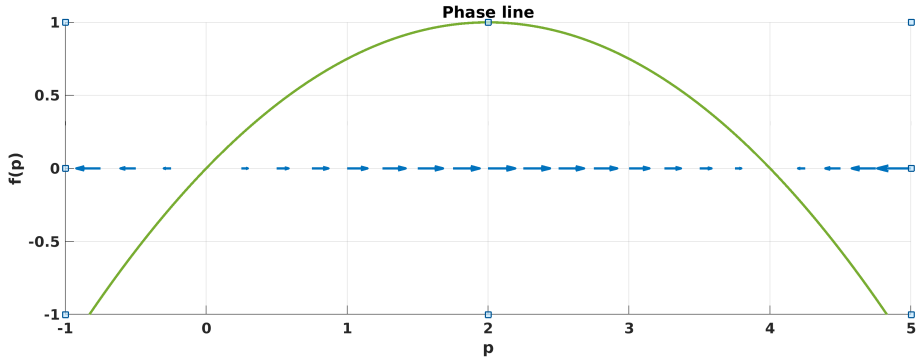


Figure 2.1: Phase line for the population growth equation in Example 2.7. The parameters are $k = 1$ and $p_m = 4$, and the blue arrows indicate the sign and magnitude of $f(p)$. A graph of the function $f(p)$ is included, too.

the left or right that indicate the direction of the evolution and are proportional to the value of $f(x)$. Figure 2.1 gives an impression of a phase line for the following example. The phase line is a topic that we will discuss later on in the context of systems of differential equations (then called *phase diagram* or *phase portrait*, see Section 4.2.2).

Example 2.7 (Population growth model)**.** *Recall the equation*

$$\dot{p}(t) = kp(t)\Big(1 - \frac{p(t)}{p_m}\Big)$$

from the introduction. We are typically interested in the dynamics with initial populations $p_0 \in (0, p_m)$ at any time $t_0 \in \mathbb{R}$ (for simplicity, assume that $t_0 = 0$).
Here, $f(p) = kp(1 - p/p_m)$ has two zeros at $p_1^* = 0$ and $p_2^* = p_m$. For all these initial values $p_0 \in (p_1^*, p_2^*)$ we have that $f(p_0) > 0$, so solutions will increase and move towards p_2^*. We conclude that p_2^* is stable whilst p_1^* is unstable. Figure 2.1 displays the phase line with arrows for the direction of the dynamics.
Note that f is differentiable with $f'(p) = k(1 - 2p/p_m)$. Using the criterion involving the derivative we can also conclude that

$$f(p_1^*) = 0,\ f'(p_1^*) = k > 0 \quad \Rightarrow \quad p_1^* \text{ is unstable,}$$

and

$$f(p_2^*) = 0,\ f'(p_2^*) = -k < 0 \quad \Rightarrow \quad p_2^* \text{ is stable.}$$

In this case one can actually explicitly work out the solution (we will see suitable techniques later on in Section 2.6.1). It is given by

$$p(t) = \frac{p_0 p_m}{(p_m - p_0)e^{-kt} + p_0}.$$

The function indeed increases and converges to $p_2^* = p_m$ as $t \to \infty$, so this is the stable point that attracts nearby solutions.

2.4 Well-posedness

A natural question is as to whether any differential equation of the form (2.5) has a unique solution. We have seen already in the case of a trivial differential equation that we need to impose an additional condition to single out a particular solution. However, this is not sufficient. Consider the following examples:

Example (non-existence)**.** *Consider the initial value problem*

$$\frac{d}{dt}x(t) = \frac{-x(t)^2}{t^2}, \quad x(0) = x_0 \in \mathbb{R}.$$

The right-hand-side function $f(x,t) = -x^2/t^2$ is not defined at $t = 0$. Intuitively, imposing a condition at time $t = 0$ seems problematic. Indeed, if we rewrite the differential equation in the form $x(t)^2 + t^2 \frac{d}{dt}x(t) = 0$ then evaluating at $t = 0$ yields that $x(0)^2 = 0$. The condition $x(0) = x_0 \in \mathbb{R}$ therefore cannot be satisfied if $x_0 \neq 0$, and the initial value problem has no solution then.

Example (non-uniqueness)**.** *Consider the initial value problem*

$$\frac{d}{dt}x(t) = \sqrt{x(t)}, \quad x(0) = 0.$$

The function $x(t) = 0$ for all t obviously is a solution. However, it is not the only one. For any $c > 0$ the function

$$x_c(t) = \begin{cases} 0 & \text{if } t \leq c, \\ \frac{(t-c)^2}{4} & \text{otherwise} \end{cases}$$

also is a solution. Indeed, if $t > c$ then

$$x_c'(t) = \frac{d}{dt}\Big(\frac{(t-c)^2}{4}\Big) = \frac{t-c}{2} = \sqrt{\frac{(t-c)^2}{4}} = \sqrt{x_c(t)},$$

if $t < c$ then

$$x_c'(t) = 0 = \sqrt{x_c(t)},$$

and we note that in the point $t = c$ the limits of the derivatives from both sides exist and coincide (both are zero), so that the derivative exists, too.

Whilst these results are discouraging, one actually can ensure that there is a unique solution to (2.5) with (2.7) if the function f is sufficiently smooth, or 'nice' enough. We will now specify here what this means as we are lacking the tools to prove these results as this stage. This problem is picked up and addressed in detail in later modules.

Assumption 2.8. *For the remainder of this chapter,*

$$f \text{ is assumed to be 'nice' unless stated otherwise.}$$

As a final remark, a *problem is well-posed* if is has a unique solution that 'nicely' depends on the *data*, more precisely

$$\text{well-posedness} \;=\; \text{existence} \;+\; \text{uniqueness} \;+\; \text{stability}.$$

With *data* we mean parameters in f such as the growth factors in the infectious disease model in the introduction but also initial values such as x_0 in (2.7), and *stability* means that solutions for different data are not too far away from each other if the data points are close. Investigating stability requires the ability to measure distances between functions (the solutions). Suitable concepts will be introduced in later modules in great depth.

2.5 Solution techniques for linear equations

Let us now start to explicitly solve some non-trivial cases where the function f in (2.5) depends on x. Probably the simplest case is if f is linear in x,

$$f(x,t) = r(t)x + s(t)$$

with some functions $r, s : (\alpha, \beta) \to \mathbb{R}$. This is consistent with the defnition or linearity in (2.3) in Definition 2.1 with $a_1 = 1$ and $a_0 = r(t)$, noting that the independent variable is denoted by t. Note also that $f(x,t)$ is 'nice' in the sense of Assumption 2.8 if r and s are 'nice'.

2.5.1 Homogeneous equations

We first consider the case $s = 0$ of a homogeneous differential equation. The initial value problem then reads

$$\frac{d}{dt}x(t) = r(t)x(t), \quad t \in (\alpha, \beta), \qquad x(t_0) = x_0. \tag{2.10}$$

If $r(t) = \bar{r}$ is constant then this means we are searching for a function such that its derivative is proportional to itself. We know of such functions: the exponential functions satisfy this property. Substituting thus informed guess (also called an *ansatz*) $x(t) = e^{\lambda t}$ we obtain

$$0 = \frac{d}{dt}x(t) - \bar{r}x(t) = \lambda e^{\lambda t} - \bar{r}e^{\lambda t},$$

and we see that $x(t) = e^{\bar{r}t}$ is a solution. Actually, any multiple

$$x_c(t) = ce^{\bar{r}t}, \quad c \in \mathbb{R},$$

is a solution, too, so we have found a general solution again. The initial condition singles out a particular solution by fixing c. The function

$$x(t) = x_0 e^{\bar{r}(t-t_0)} \tag{2.11}$$

satisfies $x(t_0) = x_0$ and thus the initial value problem (2.10) with $r(t) = \bar{r}$, and the specific parameter is $c = x_0 e^{-\bar{r}t_0}$ in the general solution.

Consider now the case that r depends on the independent variable t. Let us try a similar ansatz,

$$x(t) = ce^{R(t)}$$

with $c \in \mathbb{R}$ and with a suitable function $R : (\alpha, \beta) \to \mathbb{R}$. Substituting this approach in the differential equation we see that R has to be such that

$$\dot{R}(t)ce^{R(t)} = \frac{d}{dt}x(t) = r(t)x(t) = r(t)ce^{R(t)}.$$

We see that R then has to satisfy $\dot{R} = r$, so R has to be an anti-derivative of r. If we choose

$$R(t) = \int_{t_0}^{t} r(\tau)d\tau$$

then $R(t_0) = 0$. The initial condition then implies that

$$x_0 = x(t_0) = ce^{R(t_0)} = c,$$

so the solution of (2.10) is given by

$$x(t) = x_0 e^{R(t)} \quad \text{with } R(t) = \int_{t_0}^{t} r(\tau)d\tau. \tag{2.12}$$

All the previous operations seem fine from a mathematical points of view. However, we have to bear in mind that, in general, all variables have dimensions. Recall that the principle of dimensional homogeneity implies that functions such as $e^{(\cdot)}$ can take nondimensional arguments. Obviously, if the differential equation have been nondimensionalised then also $R(t)$ is nondimensional. However, even in the dimensional case there is no problem in (2.12):

Lemma 2.9. *The initial value problem (2.10) satisfies the principle of dimensional homogeneity if and only if $[r(t)] = 1/[t]$ and $[x_0] = [x(t)]$.*
If so, then $R(t) = \int_{t_0}^{t} r(\tau)d\tau$ is nondimensional. In particular, $e^{R(t)}$ is well-defined.

Proof. For the differential equation, the principle of dimensional homogeneity is satisfied if and only if both terms $\frac{d}{dt}x(t)$ and $r(t)x(t)$ have the same dimension. Now

$$\frac{[x(t)]}{[t]} = \left[\frac{d}{dt}x(t)\right] = [r(t)x(t)] = [r(t)][x(t)] \quad \Leftrightarrow \quad [r(t)] = \frac{1}{[t]},$$

so the differential equation satisfies the principle if and only if $[r(t)] = 1/[t]$. Similarly, the initial condition satisfied the principle if and only if $[x(t)] = [x(t_0)] = [x_0]$.
If this is the case then indeed

$$[R(t)] = \left[\int_{t_0}^{t} r(\tau)d\tau\right] = [r(t)][t] = 1.$$

□

Example. *Let us use the above ideas to solve the (nondimensional) initial value problem*

$$\frac{d}{dt}x(t) + 2tx(t) = 0, \quad x(10) = 3.$$

So here $t_0 = 10$ and $x_0 = 3$. Note that $r(t) = -2t$. According to (2.12) we need to compute

$$R(t) = \int_{t_0}^{t} r(\tau)d\tau = \int_{10}^{t} -2\tau d\tau = -\tau^2\Big|_{10}^{t} = -t^2 + 100.$$

Then the solution is given by

$$x(t) = x_0 e^{R(t)} = 3e^{100-t^2}.$$

Example (Mixing problem)**.** *Consider a tank containing $v_0 = 1000l$ of a mixture of water and chlorine. To reduce the amount of chlorine, fresh water is pumped into the tank at a rate of $i = 6l/s$. The mixture is well-stirred and pumped out at a rate of $\omega = 8l/s$. The initial concentration of chlorine is $c_0 = 0.02g/l$.*
We want to find the total amount (=mass) of chlorine in the tank as a function of time.

Let us denote the chlorine mass by $q(t)$ as a function of time t, and the volume in the tank by $v(t)$. The change of the chlorine mass is equal to minus the outflow times ω the concentration $q(t)/v(t)$, hence

$$\frac{d}{dt}q(t) = -\frac{\omega}{v(t)}q(t)$$

which comes with the initial condition $q(0) = c_0 v_0$. The volume of the water / chlorine mixture changes with the net outflow $i - \omega$, i.e.,

$$\frac{d}{dt}v(t) = i - \omega \quad \Rightarrow \quad v(t) = v_0 + (i-\omega)t$$

where we used the initial condition $v(0) = v_0$. Note that this is only true as long as the tank it not empty. With $i - \omega = -2l/s$ and $v_0 = 1000l$ this is the case if $t = 500s$ as then $v(500s) = v_0 - (i-\omega)t = 1000l - 2\frac{l}{s} \times 500s = 0l$. Altogether we obtain the initial value problem

$$\frac{d}{dt}q(t) = -\frac{\omega}{v_0 + (i-\omega)t}q(t), \quad t \in (0s, 500s), \qquad q(0) = c_0 v_0 = 0.02\frac{g}{l} \times 1000l = 20g.$$

The differential equation is of the form $\dot{q}(t) = r(t)q(t)$ with $r(t) = -\omega/(v_0 + (i-\omega)t)$. Before we proceed, let us check whether the principle of dimensional homogeneity is satisfied. We have that $[\omega] = [8l/s] = L^3/T$ and similarly $[i] = L^3/T$. For the volume we have that $[v_0] = L^3$. Therefore

$$[r(t)] = \Big[-\frac{\omega}{v_0 + (i-\omega)t}\Big] = \frac{[\omega]}{[v_0] + [i-\omega][t]} = \frac{L^3/T}{L^3 + (L^3/T)T} = \frac{1}{T} = \frac{1}{[t]}.$$

Moreover, $[c_0 v_0] = [c_0][v_0] = (M/L^3)L^3 = M = [q(t)]$. By Lemma 2.9 the initial value problem thus satisfies the principle. With this and $t_0 = 0s$

$$\begin{aligned} R(t) &= \int_{t_0}^{t} r(\tau)d\tau \\ &= \int_0^t -\frac{\omega}{v_0 + (i-\omega)\tau}d\tau \\ &= -\frac{\omega}{v_0}\int_0^t \frac{1}{1 + ((i-\omega)/v_0)\tau}d\tau \\ &= -\frac{\omega}{v_0}\frac{v_0}{i-\omega}\big(\log(1 + ((i-\omega)/v_0)t) - \log(1)\big) \\ &= \frac{\omega}{\omega - i}\log\Big(\frac{v_0 + (i-\omega)t}{v_0}\Big), \end{aligned}$$

where we note that the logarithm's arguments always are nondimensional. Finally,

$$q(t) = q(t_0)e^{R(t)} = c_0 v_0 \Big(\frac{v_0 + (i-\omega)t}{v_0}\Big)^{\frac{\omega}{\omega - i}}.$$

Substituting the value of the parameters we get

$$q(t) = 20g\Big(\frac{1000l + (6l/s - 8l/s)t}{1000l}\Big)^{\frac{8l/s}{8l/s - 6l/s}} = 20g\Big(\frac{1000l - 2l/s \times t}{1000l}\Big)^4.$$

2.5.2 Inhomogeneous equations and integrating factor

Consider now the case $s \neq 0$. The differential equation then reads

$$\frac{d}{dt}x(t) = r(t)x(t) + s(t), \quad t \in (\alpha, \beta). \tag{2.13}$$

The exponential function from the previous section will prove quite useful in this case. Recall that, by Lemma 2.9, it is well-defined subject to the assumption that the differential equation satisfies the principle of dimensional homogeneity.

Definition 2.10. *The function*

$$I(t) = e^{-R(t)} \tag{2.14}$$

where $R(t) = \int r(\tau)d\tau$ is an anti-derivative of r, is called an integrating factor *for the differential equation (2.13). If an initial condition of the form $x(t_0) = x_0$ as in (2.7) is imposed then a specific anti-derivative is chosen, namely*

$$R(t) = \int_{t_0}^{t} r(\tau)d\tau.$$

Note that

$$\frac{d}{dt}I(t) = \frac{d}{dt}e^{-R(t)} = \Big(-\frac{d}{dt}R(t)\Big)e^{-R(t)} = -r(t)I(t) \tag{2.15}$$

Let us multiply the differential equation (2.13) with $I(t)$ so that

$$I(t)\,\frac{d}{dt}x(t) - I(t)r(t)x(t) = I(t)s(t).$$

Then, using (2.15),

$$I(t)\,\frac{d}{dt}x(t) - I(t)r(t)x(t) = I(t)\,\frac{d}{dt}x(t) + x(t)\frac{d}{dt}I(t) = \frac{d}{dt}\Big(I(t)x(t)\Big)$$

and we obtain

$$\frac{d}{dt}\Big(I(t)x(t)\Big) = I(t)s(t). \tag{2.16}$$

This is a trivial differential equation for $I(t)x(t)$, and such a transformation is exactly behind the idea of the integrating factor[2].

The initial condition for (2.16) reads $I(t_0)x(t_0) = I(t_0)x_0$. Recalling the solution formula for such problems (2.8) (or just integrating with respect to t), we obtain that

$$I(t)x(t) = I(t_0)x_0 + \int_{t_0}^{t} I(\tau)s(\tau)d\tau.$$

Substituting the integrating factor (2.14) we end up with the following result:

$$x(t) = \frac{1}{I(t)}x_0 + \int_{t_0}^{t}\frac{I(\tau)}{I(t)}s(\tau)d\tau = e^{R(t)}x_0 + \int_{t_0}^{t} e^{R(t)-R(\tau)}s(\tau)d\tau.$$

In summary:

Lemma 2.11. *The solution to the initial value problem (2.13), (2.7) is given by*

$$x(t) = e^{R(t)}x_0 + \int_{t_0}^{t} e^{R(t)-R(\tau)}s(\tau)d\tau \quad \text{with } R(t) = \int_{t_0}^{t} r(\tau)d\tau. \tag{2.17}$$

Recipe 2.12. *A 'recipe' to solve (2.13), (2.7) with an integrating factor is as follows: first, find an anti-derivative of r,*

$$R(t) = \int_{t_0}^{t} r(\tau)d\tau.$$

[2]More generally, so-called *exact equations* are obtained after transformation, but which are beyond the scope of this module. See [4], Sec 10 for more detail.

Then, work out the integral

$$x_s(t) = \int_{t_0}^{t} e^{R(t)-R(\tau)} s(\tau) d\tau.$$

The solution is then given by (2.17),

$$x(t) = e^{R(t)} x_0 + x_s(t).$$

Observe that the integrating factor (2.14) *featured as a tool in the derivation but does not have to be explicitly computed.*

Remark 2.13. *This recipe is aligned with the solution formula* (2.17). *There are other recipes to produce the solution with an integrating factor. These are fine when we have to solve a single problem.*
The significant advantage of solution formulas such as (2.17) *is of analytical nature. The formula tells us, how variable and parameter enter the solution and influence it. We can therefore analyse, how small perturbations such a due to measurement uncertainty can affect the solution. In other words, we can nicely investigate the stability properties of the problem.*
As an example, let $x_{01} \neq x_{02}$ *be two different initial values and let* $x_1(t)$ *and* $x_2(t)$ *denote the respective solutions, i.e.,* x_1 *and* x_2 *solve the differential equation* (2.13) *with initial conditions* $x_1(t_0) = x_{01}$ *and* $x_2(t_0) = x_{20}$ *(the other data* r, s, *and* t_0 *are assumed to be the same). Assume furthermore that* r *is nonnegative and bounded so that there is a number* $b.0$ *such that* $0 \leq r(t) \leq B$ *for all* t. *Note that then* $0 \leq R(t) = \int_{t_0}^{t} r(\tau)d\tau \leq \int_{t_0}^{t} B d\tau = B(t - t_0)$. *Using the solution formula* (2.17) *we obtain that*

$$\begin{aligned} |x_1(t) - x_2(t)| &= \left| e^{R(t)} x_{01} + \int_{t_0}^{t} e^{R(t)-R(\tau)} s(\tau) d\tau - e^{R(t)} x_{02} - \int_{t_0}^{t} e^{R(t)-R(\tau)} s(\tau) d\tau \right| \\ &= \left| e^{R(t)} (x_{01} - x_{02}) \right| \leq e^{B(t-t_0)} |x_{01} - x_{02}|. \end{aligned}$$

This is true for all t. *We thus have estimated the distance between the solutions at any time* t *(here measure with the absolute value of the difference) in terms of the distance between the initial data points. Such type of estimates are called* stability estimates.

Example. *In forensic science, Newton's law of cooling can be used to determine the time of a body's death. Newton's law states that the temperature change of the body is porportional to the difference between its temperature and the surrounding temperature. Assuming the latter to be constant and denoting it by* $\bar{T}$, *a mathematical formulation of the law reads*

$$\frac{d}{dt} T(t) = -k\big(T(t) - \bar{T}\big)$$

for a constant $k > 0$. *Note that the negative sign in front of* k *ensures the right direction of the evolution. For instance,* $\frac{d}{dt} T(t) < 0$ *if* $T(t) > \bar{T}$, *so the body's temperature decays if it is warmer than the surroundings, and the other way round. If we know the solution then we can trace it backward in time and find the time* t^* *at which it was the value of a living body,* $T^* = 37°C$. *In practice,* k *is not known, however. The idea to identify it is to take temperature measurements at two different times.*

The differential equation is of the form (2.13) *with both* $r(t) = -k$ *and* $s(t) = k\bar{T}$ *constant. Following the recipe we first compute*

$$R(t) = \int_{t_0}^{t} r(\tau) d\tau = \int_{t_0}^{t} (-k) d\tau = -k(t - t_0).$$

Then

$$\begin{aligned} x_s(t) &= \int_{t_0}^{t} e^{R(t)-R(\tau)} s(\tau) d\tau \\ &= \int_{t_0}^{t} e^{-k(t-t_0)+k(\tau-t_0)} k\bar{T} d\tau \\ &= \bar{T} \int_{t_0}^{t} k e^{k(\tau-t)} d\tau \\ &= \bar{T}\big(e^{k(\tau-t)}\big)\big|_{\tau=t_0}^{\tau=t} \\ &= \bar{T}\big(1 - e^{k(t_0-t)}\big). \end{aligned}$$

Finally, with $x_0 = T(t_0)$,

$$T(t) = e^{R(t)}x_0 + x_s(t) = e^{-k(t-t_0)}T(t_0) + \bar{T}\big(1 - e^{k(t_0-t)}\big) = \big(T(t_0) - \bar{T}\big)e^{-k(t-t_0)} + \bar{T}$$

Suppose now that we have measured the temperature at time t_0*, denoting it by* $T_0 = T(t_0)$*, and that there is a second measurement* $T_1 = T(t_1)$ *at a later time* $t_1 > t_0$*. Using the solution formula yields that*

$$T_1 = T(t_1) = \big(T(t_0) - \bar{T}\big)e^{-k(t_1-t_0)} + \bar{T} = (T_0 - \bar{T})e^{-k(t_1-t_0)} + \bar{T}$$

so that

$$e^{-k(t_1-t_0)} = \frac{T_1 - \bar{T}}{T_0 - \bar{T}} \quad \Leftrightarrow \quad k = -\log\Big(\frac{T_1 - \bar{T}}{T_0 - \bar{T}}\Big)\Big/(t_1 - t_0).$$

With the thus computed number k *we can now try to find* t^**. It is such that*

$$\begin{aligned} & T^* = (T_0 - \bar{T})e^{-k(t^*-t_0)} + \bar{T} \\ \Leftrightarrow \quad & e^{-k(t^*-t_0)} = \frac{T^* - \bar{T}}{T_0 - \bar{T}} \\ \Leftrightarrow \quad & -k(t^* - t_0) = \log\Big(\frac{T^* - \bar{T}}{T_0 - \bar{T}}\Big) \\ \Leftrightarrow \quad & t^* = t_0 - \frac{1}{k}\log\Big(\frac{T^* - \bar{T}}{T_0 - \bar{T}}\Big) \\ & \quad = t_0 + (t_1 - t_0)\log\Big(\frac{T^* - \bar{T}}{T_0 - \bar{T}}\Big)\Big/\log\Big(\frac{T_1 - \bar{T}}{T_0 - \bar{T}}\Big). \end{aligned}$$

Let us consider some specific values. Assume that at $t_0 = 8am$ *we measured* $T_0 = 28°C$ *and at* $t_1 = 9am$ *we measured* $T_1 = 27°C$*, and that the ambient room temperature is* $\bar{T} = 22°C$*. Then the temperature decay factor is* $k = -\log(\frac{5}{6})/(9-8) = \log(\frac{6}{5})$*, and the (likely) time of death is*

$$t^* = 8 + (9-8)\log\big(\tfrac{15}{6}\big)\big/\log\big(\tfrac{5}{6}\big) \approx 2.97 \quad (\textit{which is around 3am}).$$

Suppose now that we made a small error when measuring at 9am and obtained a values of $T_1 = 27.2°C$*. Then the result is* $t^* \approx 1.59$*, which is a round half past 1am and thus significantly earlier. So the solutions sensitively depends on the data in the sense that small errors in the data lead to significant deviations in the solution.*
In practice, this issue is dealt with by taking some more measurements and fitting a solution curve to these. Suppose there is another measurements at time $t_2 = 10am$ *of* $T_2 = 26.5 \circ C$*.*

The solution for the initial conditions $T(t_i) = T_i$ with some given time t_i but some unknown temperature T_i reads

$$T(t, k, T_i) = \big(T_i - \bar{T}\big)e^{-k(t-t_0)} + \bar{T}.$$

Note that we have written it as a function of t, k, T_i as k and T_i here also are treated as unknown parameters (we assume that we are confident about the values of $\bar{T}$, though). Fitting this curve to the measurements now means to minimise a function that penalises the distance $|T(t_m, k, T_i) - T_m|$ between the function at the times t_m and the corresponding measured temperature values T_m. For instance,

$$\mathcal{L}(k, T_i) = \sum_{m=0}^{2} \big(T(t_m, k, T_i) - T_m\big)^2$$

is a so-called least squares *function that models such a penalisation*[3].
If we do this with the original values $T_0 = 28°C$, $T_1 = 27°C$, $T_2 = 26.5°C$ and then proceed as before to determine the time of death then we obtain that $t^* \approx 1.73$. If we do the same with $T_1 = 27.2°C$ then $t^* \approx 1.63$. We notice that not only is this result fairly close to the previous value of 1.59 obtained before from two data points, but also the error in T_1 has a much less impact on the solution.

2.6 Further solution techniques

2.6.1 Separable equations, separation of variables

Let us now investigate the case where the function f in (2.5) is a product such that the dependent and independent variables are separated:

$$f(x, t) = \xi(x)\tau(t)$$

with some functions $\xi : \mathbb{R} \to \mathbb{R}$ and $\tau : (\alpha, \beta) \to \mathbb{R}$. We will assume these to be 'nice' so that the initial values problem

$$\frac{d}{dt}x(t) = \xi(x(t))\tau(t), \quad t \in (\alpha, \beta), \qquad x(t_0) = x_0, \tag{2.18}$$

has a unique solution.

Lemma 2.14. *If x_0 is a zero of ξ then the constant function $x(t) = x_0$, $t \in (\alpha, \beta)$, is the solution to (2.18).*
If $\xi(x_0) \neq 0$ then the solution x to (2.18) is such that $\xi(x(t)) \neq 0$ for all $t \in (\alpha, \beta)$.

Proof. If x_0 is a zero of ξ then the constant function $x(t) = x_0$ satisfies both $\frac{d}{dt}x(t) = 0$ and $\xi(x(t))\tau(t) = \xi(x_0)\tau(t) = 0$ for all t, and thus the initial value problem (2.18).

Consider now the case $\xi(x_0) \neq 0$ and let x denote the corresponding solution to (2.18). Assume that there is a time $t_1 \in (\alpha, \beta)$ such that $\xi(x(t_1)) = 0$. Let y now denote the solution to (2.18) but with the initial condition $y(t_1) = x(t_1)$ at the point t_1. Recall that the initial point in (2.18) is arbitrary, and that the solution is unique. As x satisfies the differential equation and the initial condition we conclude that $y = x$. However, as $\xi(y(t_1)) = \xi(x(t_1)) = 0$, our previous result yields that the unique solution is also given by the constant function $y(t) = x(t_1)$, $t \in (\alpha, \beta)$. But then $\xi(y(t_0)) = 0 \neq \xi(x(t_0))$, which contradicts $y = x$. Therefore, $\xi(x(t)) \neq 0$ at all times $t \in (\alpha, \beta)$. □

[3] We will not discuss how to minimise functions of several variables here, such problems will be addressed in follow-up modules

Let us now study the second case in more detail. If $\xi(x(t)) \neq 0$ for all t we can divide the differential equation by $\xi(t)$ and obtain

$$\frac{1}{\xi(x(t))}\frac{d}{dt}x(t) = \tau(t).$$

Suppose now that $H(x)$ is an anti-derivative of $1/\xi(x)$, so that $H'(x) = \frac{1}{\xi(x)}$. By the chain rule

$$\frac{d}{dt}H(x(t)) = H'(x(t))\frac{d}{dt}x(t) = \frac{1}{\xi(x(t))}\frac{d}{dt}x(t).$$

Substituting this we get that

$$\frac{d}{dt}H(x(t)) = \tau(t).$$

We see that we have transformed the differential equation into a trivial differential equation for $H(x)$. Integrating with respect to t and using the initial condition yields that

$$H(x(t)) - H(x_0) = \int_{t_0}^{t} \tau(\hat{t})d\hat{t}. \tag{2.19}$$

This is an implicit solution. If we can resolve fo $x(t)$ then we can even obtain an explicit solution.

Recipe 2.15. *A popular* recipe *to solve* (2.18) *consists of 'multiplying the differential equation with dt' so that*

$$dx = \xi(x)\tau(t)dt,$$

then to divide by ξ*,*

$$\frac{1}{\xi(x)}dx = \tau(t)dt,$$

and finally to integrate,

$$\int_{x_0}^{x(t)} \frac{1}{\xi(\hat{x})}d\hat{x} = \int_{t_0}^{t} \tau(\hat{t})d\hat{t},$$

which is (2.19) *as* $H' = 1/\xi$*.*

Example. *Let us use separation of variables to solve the initial value problem*

$$\frac{d}{dt}x(t) = 3t^2e^{-x(t)}, \quad x(0) = 1.$$

Here $\xi(x) = e^x$ *and* $\tau(t) = 3t^2$*. Using Recipe 2.15 first obtain*

$$e^x dx = 3t^2 dt$$

and then, with $t_0 = 0$ *and* $x_0 = 1$*,*

$$\int_1^{x(t)} e^{\hat{x}}d\hat{x} = \int_0^t 3\hat{t}^2 d\hat{t}.$$

Therefore

$$e^{x(t)} - e^1 = t^3 \quad \Rightarrow \quad x(t) = \log\left(t^3 + e\right).$$

2.6.2 Substitution methods

Occasionally, differential equations can be simplified by formulating them in a different dependent variable. Writing t and $x = x(t)$ for the independent and the dependent variable, respectively, a new dependent variable $u(t, x(t))$ may be considered. The challenge is to identify a suitable variable. There is no general 'recipe' other than studying the equation to potentially 'see' something, or trying with approaches that proved beneficial for similar equations. Below we consider two examples.

Example 2.16 (Bernoulli equations)**.** *Differential equations of the form*

$$\frac{d}{dt}x(t) = -p(t)x(t) + q(t)x^n(t), \quad t \in (\alpha, \beta) \subset \mathbb{R},$$

with $n \in \mathbb{N}\backslash\{0, 1\}$ and p, q given functions are named after Bernoulli *and, for instance, feature in mechanical problems where linear drag forces are insufficient.*

Let us consider the variable $u = x^{1-n}$. If $x(t)$ now denotes a solution then

$$\begin{aligned}\frac{d}{dt}u(t) &= (1-n)x^{-n}(t)\frac{d}{dt}x(t)\\ &= (1-n)x^{-n}(t)\big(-p(t)x(t) + q(t)x^n(t)\big)\\ &= (1-n)(-p(t))x^{1-n}(t) + (1-n)q(t)\\ &= (n-1)p(t)u(t) + (1-n)q(t),\end{aligned}$$

so u solves a linear, inhomogeneous equation. This can be solved with an integrating factor and then can help us to work out x.

Example 2.17 (Another substitution)**.** *Consider now the differential equation*

$$\frac{d}{dt}x(t) = 1 + \frac{x(t)}{t} + \frac{x(t)^2}{t^2}.$$

Here, we may consider the new variable $u = \frac{x}{t}$. Assuming again that $x(t)$ is a solution we then have that $x(t) = tu(t)$, and differentiating yields that

$$\frac{d}{dt}x(t) = u(t) + t\frac{d}{dt}u(t).$$

The differential equation therefore becomes

$$u(t) + t\frac{d}{dt}u(t) = 1 + u(t) + u(t)^2 \quad \Rightarrow \quad \frac{d}{dt}u(t) = \frac{1}{t}\Big(1 + u(t)^2\Big).$$

This is a separable equation. With an initial condition of the form $u(t_0) = u_0$ (equivalent to $x(t_0) = u_0 t_0$) we may follow the recipe in Remark 2.15 to obtain that

$$\int_{u_0}^{u(t)} \frac{1}{1+\tilde{u}^2}d\tilde{u} = \int_{t_0}^{t} \frac{1}{\hat{t}}d\hat{t}$$

and then (assuming $t > t_0$ for simplicity, the other case is left as an exercise)

$$\operatorname{atan}(u(t)) - \operatorname{atan}(u_0) = \log(t/t_0).$$

Reverting the substitution yields that

$$x(t) = t\tan\big(\log(t/t_0) + \operatorname{atan}(x(t_0)/t_0)\big)$$

Chapter 3

Higher order equations

We first look into linear differential equations with $\mathcal{F}$ of the form (2.3), however we confine ourselves mostly to second order equations. The essential difficulties and ideas to overcome them can be explained for second order equations already, though there are some technical subtleties if the order is higher. Moreover, second order equations are of importance in many applications and can describe oscillatory behaviour. We briefly also study a nonlinear second order equation, which turns out to be a perturbed linear equation, and we discuss corrections to the solution of the linearised equations.

3.1 Linear second order equations

In this section we consider differential equations of the form

$$a(t)\frac{d^2}{dt^2}x(t) + b(t)\frac{d}{dt}x(t) + c(t)x(t) = s(t) \tag{3.1}$$

with given functions $a(t) \neq 0$, $b(t)$, $c(t)$, and $s(t)$, $t \in (\alpha, \beta)$. First, a brief observation. A special case is ($a(t) = 1$, $b(t) = c(t) = s(t) = 0$)

$$\frac{d^2}{dt^2}x(t) = 0.$$

Integrating twice we obtain that

$$x(t) = C_1 t + C_0$$

with two integration constants $C_1, C_0 \in \mathbb{R}$. We see that, in order to identify a specific solution we need to impose two conditions. This motivates the following initial value problem:

Definition 3.1. *The* initial value problem for linear second order equations *consists of finding a function* $x : (\alpha, \beta) \to \mathbb{R}$ *such that the differential equation* (3.1) *and the* initial conditions

$$x(t_0) = x_0, \quad \frac{d}{dt}x(t_0) = v_0 \tag{3.2}$$

for some $t_0 \in [\alpha, \beta]$ *and some given numbers* $x_0, v_0 \in \mathbb{R}$ *are satisfied.*

Assumption 3.2. *Throughout this chapter and unless stated otherwise we assume that the data* a, b, c, *and* s *are 'nice' enough so that the initial value problem in Definition 3.1 has a unique solution.*

There are abstract results that address this problem. However, these results usually do not produce explicit solutions. Our strategy for this purpose is to exploit the linearity as follows:

1. We first study the homogeneous equation ($s = 0$ in (3.1)). Thanks to linearity this will require solving two problems. The outcome will be a general solution called *complementary function* with two parameters. These enable us to account for two initial conditions (3.2).

2. The inhomogeneous case then requires finding a specific solution called *particular integral* of (3.1). Thanks to linearity again, we can then use the results for the homogeneous case to account for the initial condition.

3.1.1 Homogeneous equation, complementary function

We consider now the case of a homogeneous equation with $s = 0$ in (3.1), i.e.,

$$a(t)\frac{d^2}{dt^2}x(t) + b(t)\frac{d}{dt}x(t) + c(t)x(t) = 0, \tag{3.3}$$

and aim for describing the set of solutions. The following lemma can be generalised to higher order in a straightforward way.

Lemma 3.3. *Assume that x_1 and x_2 are two solutions of (3.3). Then so is* linear combination

$$x(t) = l_1x_1(t) + l_2x_2(t), \quad t \in (\alpha, \beta),$$

for all $l_1, l_2 \in \mathbb{R}$.

Proof. Substituting x in (3.3) we see that indeed

$$\begin{aligned}
&a(t)\frac{d^2}{dt^2}x(t) + b(t)\frac{d}{dt}x(t) + c(t)x(t)\\
=&\,a(t)\frac{d^2}{dt^2}\big(l_1x_1(t) + l_2x_2(t)\big) + b(t)\frac{d}{dt}\big(l_1x_1(t) + l_2x_2(t)\big) + c(t)\big(l_1x_1(t) + l_2x_2(t)\big)\\
=&\,l_1a(t)\frac{d^2}{dt^2}x_1(t) + l_2a(t)\frac{d^2}{dt^2}x_2(t) + l_1b(t)\frac{d}{dt}x_1(t) + l_2b(t)\frac{d}{dt}x_2(t) + l_1c(t)x_1(t) + l_2c(t)x_2(t)\\
=&\,l_1\big(a(t)\frac{d^2}{dt^2}x_1(t) + b(t)\frac{d}{dt}x_1(t) + c(t)x_1(t)\big) + l_2\big(a(t)\frac{d^2}{dt^2}x_2(t) + b(t)\frac{d}{dt}x_2(t) + c(t)x_2(t)\big)\\
=&\,l_1 \times 0 + l_2 \times 0 = 0.
\end{aligned}$$

□

Definition 3.4 (Linearly independent functions)**.** *Two functions x_1, x_2 on an interval (α, β) are* linearly independent[1] *if the only solution to the problem of finding $l_1, l_2 \in \mathbb{R}$ such that*

$$l_1x_1(t) + l_2x_2(t) = 0 \text{ for all } t \in (\alpha, \beta)$$

is given by $l_1 = l_2 = 0$.

Note that linear independence implies that $x_1 \neq 0$ and $x_2 \neq 0$ (as whole functions, at some points t their value is permitted to be zero).

[1]The concept of linear independence might be known from linear algebra. Here, the vector space is that of functions, which can be pointwise added and multiplied with a scalar.

Theorem 3.5. *Given any two linearly independent solutions x_1, x_2, all solutions of (3.3) are of the form*

$$x(t) = l_1 x_1(t) + l_2 x_2(t), \quad t \in (\alpha, \beta),$$

for some $l_1, l_2 \in \mathbb{R}$.

Proof. In the following, we write $\dot{y} = \frac{d}{dt}y(t)$ for the derivative to keep the presentation a bit neater.
Assume that x is a solution to (3.3) and, for some $t_0 \in (\alpha, \beta)$ let $x_0 = x(t_0)$ and $v_0 = \dot{x}(t_0)$. If we can write x in the form $l_1 x_1 + l_2 x_2$ then then the coefficients l_1 and l_2 have to be such that

$$\begin{aligned} x_0 &= l_1 x_1(t_0) + l_2 x_2(t_0), \\ v_0 &= l_1 \dot{x}_1(t_0) + l_2 \dot{x}_2(t_0). \end{aligned} \tag{3.4}$$

This is a 2×2 system of linear equations. Suppose that the matrix is singular or, equivalently, that the determinant[2] is zero. This means that

$$0 = \det \begin{pmatrix} x_1(t_0) & x_2(t_0) \\ \dot{x}_1(t_0) & \dot{x}_2(t_0) \end{pmatrix} = x_1(t_0)\dot{x}_2(t_0) - x_2(t_0)\dot{x}_1(t_0).$$

Assume for simplicity that $x_2(t_0) \neq 0$ and $\dot{x}_2(t_0) \neq 0$ (we will omit the discussion of the other cases, which involves changing the roles of x_1 and x_2 in the following but don't require any new ideas). We then obtain that

$$\frac{x_1(t_0)}{x_2(t_0)} = \frac{\dot{x}_1(t_0)}{\dot{x}_2(t_0)} =: \bar{c} \in \mathbb{R}.$$

Therefore

$$x_1(t_0) = \bar{c} x_2(t_0), \quad \dot{x}_1(t_0) = \bar{c}\dot{x}_2(t_0).$$

Now, by linearity (i.e., Lemma 3.3) the function $z(t) = \bar{c}x_2(t)$ is a solution of (3.3). As $z(t_0) = \bar{c}x_2(t_0) = x_1(t_0)$ and $\dot{z}(t_0) = \bar{c}\dot{x}_2(t_0) = \dot{x}_1(t_0)$ it satisfies the same initial conditions as x_1. The solution to the initial value problem is unique by Assumption 3.2. Therefore, the two functions have to coincide. This means that $x_1(t) = z(t) = \bar{c}x_2(t)$ for all t, so the functions x_1 and x_2 are not linearly independent as assumed.

Consequently, the matrix in the system (3.5) is regular, and the coefficients l_1 and l_2 can be uniquely determined. We know from Lemma 3.3 that $l_1 x_1 + l_2 x_2$ is a solution to the differential equation (3.3) for any l_1, l_2 and therefore also solves the initial value problem from Def 3.1 (the differential equation (3.3) is a special case of (3.1)). But x also solves this problem. Hence, by Assumption 3.2 on the uniqueness, the two functions have to coincide, $x(t) = l_1 x_1(t) + l_2 x_2(t)$ for all $t \in (\alpha, \beta)$, and we see that x indeed can be written as claimed. □

Definition 3.6 (Complementary function)**.** *Given any two linearly independent solutions x_1, x_2 of the linear, homogeneous, second order differential equation (3.3), the* general solution *(or set of solutions)*

$$\left\{ l_1 x_1(t) + l_2 x_2(t) \,\middle|\, t \in (\alpha, \beta),\, l_1, l_2 \in \mathbb{R} \right\}$$

is called complementary function.

Corollary 3.7. *Assume that x_1 and x_2 are two linearly independent solutions of (3.10). Then the solution to the initial value problem (3.10) with (3.2) is given by*

$$x(t) = l_1 x_1(t) + l_2 x_2(t), \quad t \in (\alpha, \beta),$$

[2] This determinant is known as the *Wronskian*.

where l_1, l_2 are such that

$$\begin{aligned} x_0 &= l_1 x_1(t_0) + l_2 x_2(t_0), \\ v_0 &= l_1 \frac{d}{dt} x_1(t_0) + l_2 \frac{d}{dt} x_2(t_0). \end{aligned} \tag{3.5}$$

Proof. This follows directly from Theorem 3.5 and the discussion around (3.4) regarding the regularity of the system matrix. □

We now derive a specific set of solutions to (3.3) that makes it convenient to account for initial conditions of the form (3.2). Let $x_d(t)$ denote the solution with initial conditions

$$x_d(t_0) = 1, \quad \dot{x}_d(t_0) = 0, \tag{3.6}$$

and $x_n(t)$ the solution with initial condition

$$x_n(t_0) = 0, \quad \dot{x}_n(t_0) = 1. \tag{3.7}$$

By the Assumption 3.2 these solutions exist and are unique.

Lemma 3.8. *For the solutions x_d and x_n we have the following results:*

1. *The functions x_d and x_n are linearly independent.*
2. *The solution to (3.3) and (3.2) is given by*

$$x(t) = x_0 x_d(t) + v_0 x_n(t).$$

Proof. 1. If this was not the case then one function would be the multiple of another function. Now, x_d cannot be a multiple of x_n because $x_n(t_0) = 0$ but $x_d(t_0) = 1$. In turn, if x_n was a multiple of x_d then, again because of their values at t_0 we see that $x_n = 0 \times x_d$. But then $x_n = 0$ everywhere, and could not satisfy $\dot{x}_n(t_0) = 1$.

2. By Assumption 3.2 the solution is unique, so we only have to show that x solves the initial value problem. By Lemma 3.3, x is a solution to the differential equation (3.1) (with $s = 0$). Moreover, using the properties of x_d and x_n,

$$x(t_0) = x_0 \underbrace{x_d(t_0)}_{=1} + v_0 \underbrace{x_n(t_0)}_{=0} = x_0$$

and

$$\dot{x}(t_0) = x_0 \underbrace{\dot{x}_d(t_0)}_{=0} + v_0 \underbrace{\dot{x}_n(t_0)}_{=1} = v_0,$$

so x also satisfies the initial conditions.

□

3.1.2 Inhomogeneous equation, particular integral

Let us now assume that $s \neq 0$ in (3.1). A usual objective then is to find a solution x_p that satisfies

$$x_p(t_0) = 0, \quad \frac{d}{dt}x_p(t_0) = 0. \tag{3.8}$$

Such functions are called *particular integrals*, and they are convenient for the following reason:

Lemma 3.9. *With x_d, x_n from the previous section, the function*

$$x(t) = x_p(t) + x_0 x_d(t) + v_0 x_n(t), \quad t \in (\alpha, \beta),$$

is the solution to (3.1) with initial condition (3.2).

Proof. By the Assumption 3.2 we know that the solution is unique and thus only have to show that x solves the equation and the conditions. Regarding the initial conditions,

$$x(t_0) = \underbrace{x_p(t_0)}_{=0} + x_0 \underbrace{x_d(t_0)}_{=1} + v_0 \underbrace{x_n(t_0)}_{=0} = x_0$$

and

$$\dot{x}(t_0) = \underbrace{\dot{x}_p(t_0)}_{=0} + x_0 \underbrace{\dot{x}_d(t_0)}_{=0} + v_0 \underbrace{\dot{x}_n(t_0)}_{=1} = v_0.$$

Furthermore, also the inhomogeneous differential equation is satisfied:

$$\begin{aligned}
&a(t)\frac{d^2}{dt^2}x(t) + b(t)\frac{d}{dt}x(t) + c(t)x(t) \\
&= a(t)\frac{d^2}{dt^2}\big(x_p(t) + x_0 x_d(t) + v_0 x_n(t)\big) \\
&\quad + b(t)\frac{d}{dt}\big(x_p(t) + x_0 x_d(t) + v_0 x_n(t)\big) \\
&\quad + c(t)\big(x_p(t) + x_0 x_d(t) + v_0 x_n(t)\big) \\
&= a(t)\frac{d^2}{dt^2}x_p(t) + b(t)\frac{d}{dt}x_p(t) + c(t)x_p(t) \\
&\quad + x_0\Big(a(t)\frac{d^2}{dt^2}x_d(t) + b(t)\frac{d}{dt}x_d(t) + c(t)x_d(t)\Big) \\
&\quad + v_0\Big(a(t)\frac{d^2}{dt^2}x_n(t) + b(t)\frac{d}{dt}x_n(t) + c(t)x_n(t)\Big) \\
&= s(t) + x_0 \times 0 + v_0 \times 0 = s(t),
\end{aligned}$$

where we used that x_d and x_n solve the homogeneous equation. □

The idea of building solutions as linear combinations of elementary solutions is known as *superposition principle*. Here, we build it from a particular integral (satisfying homogeneous initial conditions (3.8)) and from solutions to the homogeneous differential equation that enable us to account for arbitrary initial values.

Remark 3.10. *In practice, finding x_p can be challenging. But note that the initial conditions (3.8) do not have to be satisfied. In fact, if $x_p(t_0) \neq 0$ and $\dot{x}_p(t_0) \neq 0$ then it is easy to show that the solution to (3.1), (3.2) is given by*

$$x(t) = x_p(t) + \big(x_0 - x_p(t_0)\big)x_d(t) + \big(v_0 - \dot{x}_p(t_0)\big)x_n(t).$$

In the next section, we look into applying these ideas around the superposition principle to the specific case of constant coefficients.

3.2 Linear and second order with constant coefficients

To illustrate how the procedure outlined in the previous section works let us investigate the case of an equation (3.1) with constant coefficients,

$$a\frac{d^2}{dt^2}x(t) + b\frac{d}{dt}x(t) + cx(t) = s(t), \tag{3.9}$$

where we assume that $a \neq 0$ (otherwise the order of the differential equation would be less than two).

We first need to find the solutions to the homogeneous equation, so assume that $s = 0$ and study

$$a\frac{d^2}{dt^2}x(t) + b\frac{d}{dt}x(t) + cx(t) = 0, \tag{3.10}$$

We have seen that function of the form $e^{\lambda t}$ solved the linear first order equations with constant coefficients, so let us try this form. Substituting it in the differential equation (3.10) yields that

$$0 = a\lambda^2 e^{\lambda t} + b\lambda e^{\lambda t} + ce^{\lambda t} = \big(a\lambda^2 + b\lambda + c\big)e^{\lambda t}.$$

Definition 3.11 (Auxiliary equation)**.** *The equation*

$$a\lambda^2 + b\lambda + c = 0$$

is called the auxiliary equation *or* characteristic equation *of the differential equation* (3.10).

So if λ is a solution to the auxiliary equation then $e^{\lambda t}$ solves the differential equation. The solutions to this quadratic equation are given by

$$\lambda_{1,2} = -\frac{b}{2a} \pm \frac{1}{2a}\sqrt{b^2 - 4ac}.$$

We have to distinguish several cases, depending on the sign of $b^2 - 4ac$.

3.2.1 Auxiliary equation with two real roots

If $b^2 - 4ac > 0$ then $\lambda_1 \neq \lambda_2$ are two real roots of the auxiliary equation. The functions

$$x_1(t) = e^{\lambda_1 t}, \quad x_2(t) = e^{\lambda_2 t}$$

then are two linearly independent solutions of (3.10). Let us now find the solutions $x_d(t)$ and $x_n(t)$. For simplicity, let us assume that

$$t_0 = 0.$$

Otherwise replace t with $t - t_0$ in the following computations.
We proceed as explained around (3.5). Starting with

$$x_d(t) = l_{1d}x_1(t) + l_{2d}x_2(t) = l_{1d}e^{\lambda_1 t} + l_{2d}e^{\lambda_2 t},$$

the initial conditions (3.6) yield that

$$\begin{aligned} 1 &= x_d(0) = l_{1d}x_1(0) + l_{2d}x_2(0) = l_{1d} + l_{2d}, \\ 0 &= \dot{x}_d(0) = l_{1d}\dot{x}_1(0) + l_{2d}\dot{x}_2(0) = l_{1d}\lambda_1 + l_{2d}\lambda_2. \end{aligned}$$

Solving this system (eg by substituting $l_{2d} = 1 - l_{1d}$ in the second equation and resolving for l_{1d}) yields that

$$l_{1d} = \frac{\lambda_2}{\lambda_2 - \lambda_1}, \quad l_{2d} = \frac{\lambda_1}{\lambda_1 - \lambda_2}$$

and thus the solution

$$x_d(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2} e^{\lambda_2 t}.$$

For x_n we start with

$$x_n(t) = l_{1n} x_1(t) + l_{2n} x_2(t) = l_{1n} e^{\lambda_1 t} + l_{2n} e^{\lambda_2 t}.$$

The initial conditions (3.7) yield that

$$\begin{aligned} 0 &= x_n(0) = l_{1n} x_1(0) + l_{2n} x_2(0) = l_{1n} + l_{2n}, \\ 1 &= \dot{x}_n(0) = l_{1n} \dot{x}_1(0) + l_{2n} \dot{x}_2(0) = l_{1n} \lambda_1 + l_{2n} \lambda_2. \end{aligned}$$

The solution is given by

$$l_{1n} = \frac{1}{\lambda_1 - \lambda_2}, \quad l_{2n} = \frac{1}{\lambda_2 - \lambda_1}$$

and then

$$x_n(t) = \frac{1}{\lambda_1 - \lambda_2} e^{\lambda_1 t} + \frac{1}{\lambda_2 - \lambda_1} e^{\lambda_2 t}.$$

Example 3.12. *Consider the initial value problem*

$$\frac{d^2}{dt^2} x(t) + \frac{d}{dt} x(t) - 6x(t) = 0, \qquad x(0) = 5, \quad \frac{d}{dt} x(0) = 0.$$

We know from Lemma 3.8 that the solution then is given by

$$x(t) = x_0 x_d(t) + v_0 x_n(t) = 5 x_d(t).$$

The auxiliary equation reads

$$\lambda^2 + \lambda - 6 = 0$$

and has the solutions

$$\lambda_{1,2} = -\frac{1}{2} \pm \frac{1}{2}\sqrt{1 + 24} = \frac{1}{2}\big(1 \pm 5\big), \quad \Rightarrow \quad \lambda_1 = 2,\ \lambda_2 = -3,$$

so $\lambda_1 - \lambda_2 = 5$ *and*

$$x(t) = 5\Big(\frac{\lambda_2}{\lambda_2 - \lambda_1} e^{\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2} e^{\lambda_2 t}\Big) = 3e^{2t} + 2e^{-3t}.$$

In the example, one could alternatively proceed directly as outlined around (3.5), starting with

$$x(t) = l_1 x_1(t) + l_2 x_2(t)$$

and identifying l_1, l_2 by solving

$$\begin{aligned} 5 &= x(0) = l_1 x_1(0) + l_2 x_2(0) = l_1 + l_2, \\ 0 &= \dot{x}(0) = l_1 \dot{x}_1(0) + l_2 \dot{x}_2(0) = l_1 \lambda_1 + l_2 \lambda_2. \end{aligned}$$

This procedure usually is faster than computing x_d and x_n first. However, the latter is useful if solutions for different initial conditions have to be computed.

3.2.2 Auxiliary equation with complex roots

If $b^2 - 4ac < 0$ then

$$\lambda_{1,2} = p \pm \mathbf{i}q \quad \text{where } p = -\frac{b}{2a},\ q = \frac{\sqrt{4ac - b^2}}{2a}.$$

The functions

$$\tilde{x}_1(t) = e^{(p+\mathbf{i}q)t}, \quad \tilde{x}_2(t) = e^{(p-\mathbf{i}q)t}$$

are linearly independent but complex-valued solutions to the differential equation (3.10). We are interested in two linearly independent real-valued solutions. For this purpose, we first note that the arguments in Lemma 3.3 still go through if $l_1, l_2 \in \mathbb{C}$. Second, by Euler's formula

$$e^{\pm \mathbf{i}qt} = \cos(qt) \pm \mathbf{i}\sin(qt).$$

The functions

$$x_1(t) = \frac{1}{2}\big(\tilde{x}_1(t) + \tilde{x}_2(t)\big) = e^{pt}\cos(qt),$$
$$x_2(t) = \frac{1}{2\mathbf{i}}\big(\tilde{x}_1(t) - \tilde{x}_2(t)\big) = e^{pt}\sin(qt),$$

therefore are solutions to (3.10), are real-valued, and it is not difficult to see that they are linearly independent.

Let us again assume that $t_0 = 0$ but skip the computation of x_d and x_n but solve an example directly, following the lines around (3.5)

Example 3.13. *Consider the initial value problem*

$$\frac{d^2}{dt^2}x(t) - 2\frac{d}{dt}x(t) + 26x(t) = 0, \qquad x(0) = 3, \quad \frac{d}{dt}x(0) = -12.$$

Then $a = 1$, $b = -2$, $c = 26$, *and then*

$$p = -\frac{b}{2a} = 1, \quad q = \frac{\sqrt{4ca - b^2}}{2a} = \frac{\sqrt{104 - 4}}{2} = 5,$$

so we obtain that

$$x_1(t) = e^t\cos(5t), \quad x_2(t) = e^t\sin(5t).$$

The general solution thus is

$$x(t) = l_1x_1(t) + l_2x_2(t) = l_1e^t\cos(5t) + l_2e^t\sin(5t), \quad l_1, l_2 \in \mathbb{R}.$$

Note that

$$\frac{d}{dt}x(t) = l_1\big(e^t\cos(5t) - 5e^t\sin(5t)\big) + l_2\big(e^t\sin(5t) + 5e^t\cos(5t)\big).$$

To satisfy the initial conditions we have to solve

$$3 = l_1x_1(0) + l_2x_2(0) = l_1,$$
$$-12 = l_1\dot{x}_1(0) + l_2\dot{x}_2(0) = l_1 + 5l_2.$$

Substituting the result for l_1 *into the second equation yields* $l_2 = -3$. *Hence,*

$$x(t) = 3e^t\big(\cos(5t) - \sin(5t)\big).$$

3.2.3 Auxiliary equation with one repeated root

In the case $b^2 - 4ac = 0$ the auxiliary equation has only one root

$$\lambda := \lambda_{1,2} = -\frac{b}{2a}.$$

Then $\tilde{x}_1(t) = e^{\lambda t}$ is only one solution. A convenient second, linearly independent solution is given by[3]

$$\tilde{x}_2(t) = te^{\lambda t}.$$

To see that this indeed is a solution, note that

$$\frac{b}{a} = -2\lambda$$

and

$$b^2 - 4ac = 0 \quad \Rightarrow \quad c = \frac{b^2}{4a} \quad \Rightarrow \quad \frac{c}{a} = \frac{1}{4}\frac{b^2}{a^2} = \lambda^2.$$

Now

$$\frac{d}{dt}\tilde{x}_2(t) = e^{\lambda t} + \lambda te^{\lambda t},$$
$$\frac{d^2}{dt^2}\tilde{x}_2(t) = 2\lambda e^{\lambda t} + \lambda^2 te^{\lambda t}.$$

Altogether,

$$\frac{d^2}{dt^2}\tilde{x}_2(t) + \frac{b}{a}\frac{d}{dt}\tilde{x}_2(t) + \frac{c}{a}\tilde{x}_2(t)$$
$$= 2\lambda e^{\lambda t} + \lambda^2 te^{\lambda t} - 2\lambda\big(e^{\lambda t} + \lambda te^{\lambda t}\big) + \lambda^2 te^{\lambda t} = e^{\lambda t}\big(2\lambda + \lambda^2 t - 2\lambda - 2\lambda^2 t + \lambda^2 t\big) = 0,$$

so $\tilde{x}_2(t)$ indeed solves (3.10).

To account for initial data at any time $t_0 \neq 0$ let us 'shift in time' and consider

$$x_1(t) = e^{\lambda(t-t_0)}, \quad x_2(t) = (t - t_0)e^{\lambda(t-t_0)}.$$

Note that $x_1(t) = \tilde{x}_1(t)e^{-\lambda t_0}$ is just a multiple of $\tilde{x}_1$ and thus a solution again, and

$$x_2(t) = \tilde{x}_2(t)e^{-\lambda t_0} - t_0 x_1(t)$$

is a linear combination of solutions and thus solves (3.10), too.

To find the solution satisfying (3.6) we start as usual with a linear combination

$$x_d(t) = l_{1d}x_1(t) + l_{2d}x_2(t) = l_{1d}e^{\lambda(t-t_0)} + l_{2d}(t - t_0)e^{\lambda(t-t_0)}, \quad l_{1d}, l_{2d} \in \mathbb{R}.$$

Noting that

$$\frac{d}{dt}x_d(t) = \lambda l_{1d}e^{\lambda(t-t_0)} + l_{2d}\big(1 + \lambda(t - t_0)\big)e^{\lambda(t-t_0)}$$

[3]This second solution drops out of the blue here. It is fairly natural, though, but deriving it requires some tools from Linear Algebra, namely the *Jordan Canonical Form* and the *Matrix Exponential*. Details are left for future modules, interested readers may resort to [5], Sec 3.3.

we see that $x_1(t_0) = 1$, $x_2(t_0) = 0$, $\dot{x}_1(t_0) = \lambda$, and $\dot{x}_2(t_0) = 1$ we obtain the system

$$\begin{aligned} 1 &= x_d(t_0) = l_{1d}x_1(t_0) + l_{2d}x_2(t_0) = l_{1d}, \\ 0 &= \dot{x}_d(t_0) = l_{1d}\dot{x}_1(t_0) + l_{2d}\dot{x}_2(t_0) = l_{1d}\lambda + l_{2d} \end{aligned}$$

for the coefficients and then the solution

$$x_d(t) = x_1(t) - \lambda x_2(t) = e^{\lambda(t-t_0)} - \lambda(t-t_0)e^{\lambda(t-t_0)}.$$

Regarding the solution satisfying (3.7), we have already discovered and used that $x_2(t_0) = 0$ and $\dot{x}_2(t_0) = 1$, so we have that

$$x_n(t) = x_2(t) = (t-t_0)e^{\lambda(t-t_0)}.$$

Example 3.14. *Consider the initial value problem*

$$\frac{d^2}{dt^2}x(t) + 6\frac{d}{dt}x(t) + 9x(t) = 0, \qquad x(-1) = 2, \quad \frac{d}{dt}x(-1) = -5.$$

Then $a = 1$, $b = 6$, $c = 9$, and indeed $b^2 - 4ac = 0$. Noting that $\lambda = -b/2a = -3$, $t_0 = -1$, $x_0 = 2$, and $v_0 = -5$ we obtain from Lemma 3.8 that the solution is given by

$$x(t) = x_0x_d(t) + v_0x_n(t) = 2\big(e^{-3(t+1)} + 3(t+1)e^{-3(t+1)}\big) - 5(t+1)e^{-3(t+1)} = \big(3+t\big)e^{-3(t+1)}.$$

3.2.4 Inhomogeneous equation

Let us now assume that $s \neq 0$ and investigate the inhomogeneous equation (3.9). As discussed around Lemma 3.9 we may try to find a *particular integral* x_p. However, there is no general technique to do so, other than by the method of 'inspired guesswork'. Let us consider two examples:

Example 3.15. *Let us try to find the solutions to*

$$\frac{d^2}{dt^2}x(t) + \frac{d}{dt}x(t) - 2x(t) = t^2.$$

The solutions to the homogeneous equation can be found using the methods in the previous sections and are given in terms of

$$x_d(t) = \frac{2}{3}e^{t} + \frac{1}{3}e^{-2t}, \quad x_n(t) = \frac{1}{3}e^{t} - \frac{1}{3}e^{-2t}.$$

As $s(t)$ is a polynomial we may try to find a particular integral in the form of a polynomial, too. Choosing the same degree let us try the ansatz

$$x_p(t) = p_2t^2 + p_1t + p_0, \quad \text{with some } p_i \in \mathbb{R}.$$

Substituting this and its derivatives in the differential equation yields that

$$t^2 = \frac{d^2}{dt^2}x_p(t) + \frac{d}{dt}x_p(t) - 2x_p(t) = 2p_2 + (2p_2t + p_1) - 2(p_2t^2 + p_1t + p_0).$$

Equating coefficients of the various powers of t yields the equations

$$-2p_2 = 1, \quad 2p_2 - 2p_1 = 0, \quad 2p_2 + p_1 - 2p_0 = 0.$$

We see that $p_2 = -1/2$, *then* $p_1 = p_2 = -1/2$ *and finally* $p_0 = p_2 + p_1/2 = -3/4$ *so that*

$$x_p(t) = -\frac{1}{2}t^2 - \frac{1}{2}t - \frac{3}{4}.$$

This function does not satisfy the initial conditions (3.8), *so to find the solution for any initial condition of the form* $x(0) = x_0$, $\dot{x}(0) = v_0$ *we use Remark 3.10. With* $x_p(0) = -3/4$ *and* $\dot{x}_p(0) = -1/2$ *we obtain the solution*

$$\begin{aligned} x(t) &= x_p(t) + \big(x_0 - x_p(t_0)\big)x_d(t) + \big(v_0 - \dot{x}_p(t_0)\big)x_n(t) \\ &= -\frac{1}{2}t^2 - \frac{1}{2}t - \frac{3}{4} + \big(x_0 + \frac{3}{4}\big)\Big(\frac{2}{3}e^{t} + \frac{1}{3}e^{-2t}\Big) + \big(v_0 + \frac{1}{2}\big)\Big(\frac{1}{3}e^{t} - \frac{1}{3}e^{-2t}\Big). \end{aligned}$$

Example 3.16. *Let us investigate the same equation but with* $s(t) = e^{-t}$,

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

In this case we may try with an exponential function for the particular integral, too. Substituting

$$x_p(t) = ke^{-t}, \quad k \in \mathbb{R},$$

into the differential equation yields

$$e^{-t} = ke^{-t} - ke^{-t} - 2ke^{-t} = -2ke^{-t},$$

and we see that $k = -1/2$ *yields that particular integral*

$$x_p(t) = -\frac{1}{2}e^{-t}.$$

Again, the initial conditions (3.8) *are not satisfied but Remark 3.10 tells us how to incorporate any initial conditions.*
Note that $x_1(t) = e^{t}$ *and* $x_2(t) = e^{-2t}$ *are solutions to the homogeneous equation. So if, say,*

$$s(t) = e^{-2t}$$

then the ansatz $x_p(t) = ke^{-2t}$ *will not work. Remembering what we did in the case of repeated roots of the auxiliary equation we may try with* $x_p(t) = p(t)e^{-2t}$ *where* $p(t)$ *is a polynomial. In this specific case,* $x_p = -\frac{1}{3}te^{-2t}$ *turns out to work.*

3.2.5 Example: RLC circuit

Recall from Example 1.3 in the introduction the model for an electric circuit displayed in Figure 1.3,

$$L\frac{d^2}{dt^2}Q(t) + R\frac{d}{dt}Q(t) + \frac{1}{C}Q(t) = 0,$$

where $L > 0$ is the inductance of an inductor, $R > 0$ is the resistance or a resistor, and $C > 0$ is the capacity of a capacitor. If initially a charged capacitor with charge q_0 is connected to the otherwise charge and current free elements then we can express this with the initial conditions

$$q(0) = q_0, \quad \frac{d}{dt}q(0) = 0. \tag{3.11}$$

In Example 1.7 the equation was nondimensionalised, which lead to the nondimensional equation

$$\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) + \tilde{R}\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) + \tilde{q}(\tilde{t}) = 0, \tag{3.12}$$

where $\tilde{R} = R\sqrt{\frac{C}{L}}$, with nondimensional initial conditions

$$\tilde{q}(0) = \tilde{q}_0 = 1, \quad \frac{d}{d\tilde{t}}\tilde{q}(0) = 0. \tag{3.13}$$

The above differential equation is of the form (3.10) with $a = 1$, $b = \tilde{R}$, and $c = 1$. The auxiliary equation then reads

$$\lambda^2 + \tilde{R}\lambda + 1 = 0$$

and has the solutions

$$\lambda_{1,2} = -\frac{\tilde{R}}{2} \pm \frac{1}{2}\sqrt{\tilde{R}^2 - 4}.$$

Depending on the parameters we have the following solution regimes:

- *underdamped case*:
  If $\tilde{R}^2 - 4 < 0$ then Section 3.2.2 gives us the linearly independent solutions

  $$\tilde{q}_1(\tilde{t}) = e^{p\tilde{t}}\cos(\omega\tilde{t}), \quad \tilde{q}_2(\tilde{t}) = e^{p\tilde{t}}\sin(\omega\tilde{t}),$$

  of the differential equation, where $p = -\tilde{R}/2$ and $\omega = \frac{1}{2}\sqrt{4 - \tilde{R}^2}$. Starting with $\tilde{q}(\tilde{t}) = l_1\tilde{q}_1(\tilde{t}) + l_2\tilde{q}_2(\tilde{t})$, $l_1, l_2 \in \mathbb{R}$, and accounting for the initial conditions (3.13) we obtain the solution

  $$\tilde{q}(\tilde{t}) = \tilde{q}_0 e^{p\tilde{t}}\big(\cos(\omega\tilde{t}) - \frac{p}{\omega}\sin(\omega\tilde{t})\big).$$

  We expect an oscillatory behaviour in time with exponentially fast decay of the extrema. Note that if $\tilde{R} = 0$ then $p = 0$ and $\omega = 1$, so we then have perfect oscillations. Figure 3.1 confirms these expectations.

- *overdamped case*:
  If $\tilde{R}^2 - 4 > 0$ then Section 3.2.1 yields the linearly independent solutions

  $$\tilde{q}_d(\tilde{t}) = \frac{\lambda_2}{\lambda_2 - \lambda_1}e^{\lambda_1\tilde{t}} + \frac{\lambda_1}{\lambda_1 - \lambda_2}e^{\lambda_2\tilde{t}}, \quad \tilde{q}_n(\tilde{t}) = \frac{1}{\lambda_1 - \lambda_2}e^{\lambda_1\tilde{t}} + \frac{1}{\lambda_2 - \lambda_1}e^{\lambda_2\tilde{t}},$$

  which satisfy

  $$\tilde{q}_d(0) = 1, \quad \frac{d}{d\tilde{t}}\tilde{q}_d(0) = 0, \qquad \tilde{q}_n(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{q}_n(0) = 1.$$

  Accounting for the initial conditions the solution of initial value problem (3.12), (3.13) is given by

  $$\tilde{q}(\tilde{t}) = \tilde{q}_0\tilde{q}_d(\tilde{t}) = \frac{\lambda_2}{\lambda_2 - \lambda_1}e^{\lambda_1\tilde{t}} + \frac{\lambda_1}{\lambda_1 - \lambda_2}e^{\lambda_2\tilde{t}}.$$

  Note that $0 > \lambda_1 > \lambda_2$ because $\frac{1}{2}\sqrt{\tilde{R}^2 - 4} < \tilde{R}/2$, so $\tilde{q}(\tilde{t})$ decays exponentially fast in time. This can indeed be observed in Figure 3.1.

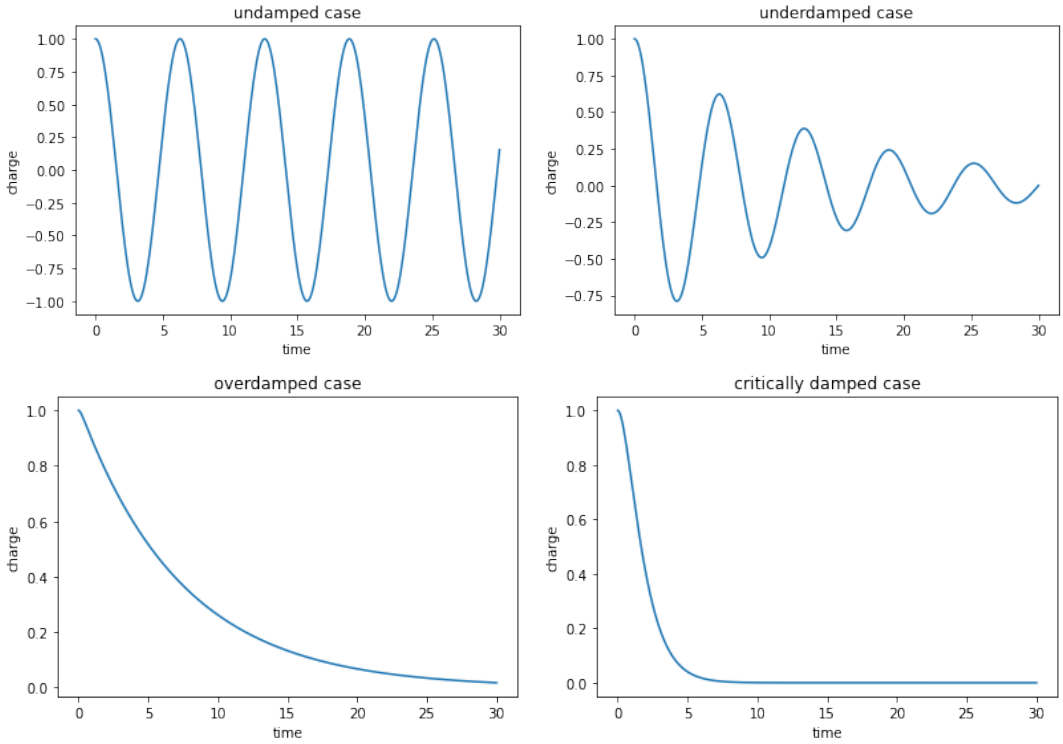


Figure 3.1: Solutions to the initial value problem (3.12), (3.13) modelling the electric charge in an RLC circuit as in Figure 1.3. For the nondimensional resistance $\tilde{R}$ we chose $\tilde{R} = 0$ (undamped), $\tilde{R} = 0.15$ (underdamped), $\tilde{R} = 7.5$ (overdamped), and $\tilde{R} = 2$ (critically damped).

- *critically damped case*:
  If $\tilde{R}^2 - 4 = 0$ then $\tilde{R} = 2$ and we have repeated eigenvalues $\lambda = -\tilde{R}/2 = -1$. Following the arguments in Section 3.2.3 one can fairly easily work out the solution to the initial value problem (3.12), (3.13):
  $$\tilde{q}(\tilde{t}) = (1 + \tilde{t})e^{-\tilde{t}}.$$
  Let us briefly check this: We clearly have that $\tilde{q}(0) = 1$. With
  $$\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) = e^{-\tilde{t}} - (1 + \tilde{t})e^{-\tilde{t}} = -\tilde{t}e^{-\tilde{t}}$$
  we also obtain that $\tilde{q}'(0) = 0$. Finally, regarding the differential equation, with
  $$\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) = -e^{-\tilde{t}} + \tilde{t}e^{-\tilde{t}}$$
  we see that indeed
  $$\frac{d^2}{d\tilde{t}^2}\tilde{q}(\tilde{t}) + \tilde{R}\,\frac{d}{d\tilde{t}}\tilde{q}(\tilde{t}) + \tilde{q}(\tilde{t}) = -e^{-\tilde{t}} + \tilde{t}e^{-\tilde{t}} - 2\tilde{t}e^{-\tilde{t}} + (1 + \tilde{t})e^{-\tilde{t}} = 0.$$
  The function is illustrated in Figure 3.1. It looks similar to the overdamped case. However, the tiniest deviations in the data such that $\tilde{R}^2 - 4$ becomes negative will lead to oscillations.

3.3 Perturbed linear second order equations

We consider now a case of a nonlinear differential equation that is 'close' to a linear equation in the sense that the term that makes it nonlinear scales with a small parameter. The idea is to expand the problem and the solution as a series in terms of the small parameter, which then leads to subsequent linear problems. The presentation is centred around a specific example, however the concepts and ideas have turned out very successful in many other applications.

The specific example is that of a projectile that is launched vertically from the Earth with initial speed v. Newton's law of gravitation yields that the height $y(t)$ then satisfies (until it drops to the ground again) the differential equation

$$\frac{d^2}{dt^2}y(t) = -\frac{g}{(1+y(t)/R)^2}$$

where $g \approx 9.81 m/s^2$ is the gravity constant and $R \approx 6.4 \times 10^6 m$ is the Earth's radius. At launch y satisfies the initial conditions

$$y(0) = 0 \quad \text{and} \quad \frac{d}{dt}y(0) = v.$$

3.3.1 Nondimensionalisation

There are several possibilities to choose the scales for a nondimensionalisation. Before discussing these we establish some identities that are useful for all choices. Let $\tilde{t} = t/\bar{t}$ denote a nondimensional time for some time scale $\bar{t}$, and $\tilde{y}(\tilde{t}) = y(t)/\bar{l}$ a nondimensional height with some scale $\bar{l}$. Then by the chain rule (see Example 1.7 around equation (1.7) for a similar calculation)

$$\frac{d}{d\tilde{t}}\tilde{y}(\tilde{t}) = \frac{\bar{t}}{\bar{l}}\frac{d}{dt}y(t), \quad \frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = \frac{\bar{t}^2}{\bar{l}}\frac{d^2}{dt^2}y(t).$$

The differential equation then becomes $(\bar{l}/\bar{t}^2)\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = g/(1+\tilde{y}(\tilde{t})\,\bar{l}/R)^2$ and the initial condition for the velocity $(\bar{l}/\bar{t})\frac{d}{d\tilde{t}}\tilde{y}(\tilde{t}) = v$. Rewriting these we obtain that

$$\frac{\bar{l}}{g\bar{t}^2}\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -\frac{1}{(1+\tilde{y}(\tilde{t})\,\bar{l}/R)^2} = 0, \quad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(\tilde{t}) = \frac{\bar{t}}{\bar{l}}v.$$

We note that the dimensions of the parameters are $[v] = L/T$, $[g] = L/T^2$, $[R] = L$. Therefore, the parameters in the nondimensional equation

$$\frac{\bar{l}}{g\bar{t}^2}, \quad \frac{\bar{l}}{R}, \quad \frac{\bar{t}v}{\bar{l}}$$

indeed are nondimensional.

With two scales $\bar{t}$ and $\bar{l}$ at our disposal we can ensure that two parameters become one, and there are three options on how to do so. We use the nondimensional variable $\varepsilon = v^2/(gR)$ to express the results in the following.

1. $\bar{l}/(g\bar{t}^2) = 1$ and $\bar{l}/R = 1$ is achieved when $\bar{l} = R$ and $\bar{t} = \sqrt{R/g}$. In this case $\bar{t}v/\bar{l} = \sqrt{R/g}v/R = \sqrt{\varepsilon}$ and the initial value problem becomes

$$\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -\frac{1}{(1+\tilde{y}(\tilde{t}))^2}, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = \sqrt{\varepsilon}.$$

2. $\bar{l}/R = 1$ and $\bar{t}v/\bar{l} = 1$ is achieved when $\bar{l} = R$ and $\bar{t} = R/v$. Then $\bar{l}/(g\bar{t}^2) = \varepsilon$, and we obtain the problem

$$\varepsilon \frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -\frac{1}{(1+\tilde{y}(\tilde{t}))^2}, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = 1.$$

3. $\bar{l}/(g\bar{t}^2) = 1$ and $\bar{t}v/\bar{l} = 1$ is achieved if $\bar{l} = v^2/g$ and $\bar{t} = v/g$. Then $\bar{l}/R = \varepsilon$ and the initial value problem reads

$$\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -\frac{1}{(1+\varepsilon\tilde{y}(\tilde{t}))^2}, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = 1.$$

Which nondimensionalisation is a sensible one? When throwing something the velocity may be in the range of $v = 10m/s$. Then $\varepsilon = v^2/(gR) \approx 1.6 \times 10^{-6}$ is very small. Any terms that comes with ε as a factor in the initial value problem therefore barely have any influence. This motivates to set $\varepsilon = 0$ and study the thus reduced problems in the above three cases.

1. If $\varepsilon = 0$ then we obtain

$$\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -\frac{1}{(1+\tilde{y}(\tilde{t}))^2}, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = 0.$$

   Without any initial velocity, the projectile remains on Earth (recall that the differential equation is only valid as long as the projectile is in the air). In others words, nothing happens. To interpret this result, note that $\bar{l} = R = 6.4 \times 10^6 m = 6400km$. Choosing such a length scale is sensible if displacements in the same range are of interest, but throws of a couple of meters are barely noticeable. Similarly, the time scale $\bar{t} = \sqrt{R/g} \approx 807.7s$ is not quite appropriate for events taking place a few seconds only. So choosing these scales is not very suitable for the problem at hand.

2. In this case we obtain the initial value problem

$$0 = -\frac{1}{(1+\tilde{y}(\tilde{t}))^2}, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = 1.$$

   This problem is not well-posed, there is no value $\tilde{y}(\tilde{t})$ to satisfy the first equation (let alone also the initial conditions). The scales are $\bar{l} = 6.4 \times 10^6 m$ again for the length and $\bar{t} = 6.4 \times 10^5 s$ for the time. As in the case before, these scales are not appropriate for the problem data.

3. If we here set $\varepsilon = 0$ then we get

$$\frac{d^2}{d\tilde{t}^2}\tilde{y}(\tilde{t}) = -1, \qquad \tilde{y}(0) = 0, \quad \frac{d}{d\tilde{t}}\tilde{y}(0) = 1.$$

   This is a nice problem that has the well-known parabola $\tilde{y}(\tilde{t}) = t - \frac{1}{2}t^2$ for a solution. The scales here are $\bar{l} \approx 10.19m$ and $\bar{t} \approx 1.019s$, which indeed seem very reasonable for this problem.

Recall the objectives when nondimensionalising differential equations:

1. Ensuring that as many parameters as possible are one.

2. Any remaining parameters should be 'small'.

Whilst the first objective simply serves to reduce the number of parameters in the system, the second can help to identify a good choice of scales if several options are available. If a well-posed mathematical problem is obtained when setting the small parameter to zero then the choice probably is good. There still is a need to validate the solution, though, by comparing with what can be observed in experiments (or simply with what one would expect to happen).

3.3.2 Asymptotic expansions

Upon choosing suitable scales, the nondimensionalisation has lead to the initial value problem (for simplicity, we drop the tildas $\tilde{\cdot}$)

$$\frac{d^2}{dt^2}y(t) = -\frac{1}{(1+\varepsilon y(t))^2}, \qquad y(0) = 0, \quad \frac{d}{dt}y(0) = 1. \tag{3.14}$$

When setting $\varepsilon = 0$ then a linear differential equation is obtained. The above differential equation therefore can be considered as a *perturbed* linear differential equation.

The solution for $\varepsilon = 0$ (let us denote it with $y_0(t)$) can give a good idea of the effective dynamics for the perturbed problem. However, if ε is not that small then the solution $y_\varepsilon(t)$ to the perturbed problem can significantly deviate from $y_0(t)$. In this case, we may want to find a *first order correction* $y_1(t)$ such that

$$y_\varepsilon(t) \approx y_0(t) + \varepsilon y_1(t) \tag{3.15}$$

yields a better approximation. The question is then what is a good problem for $y_1(t)$.

First, note that then

$$\frac{d^2}{dt^2}y_\varepsilon(t) \approx \frac{d^2}{dt^2}y_0(t) + \varepsilon\frac{d^2}{dt^2}y_1(t).$$

For the right-hand-side of the equation we use the following approximation: for any differentiable function $f(x)$ and a small h

$$f'(x) \approx \frac{f(x+h)-f(x)}{h} \quad \Rightarrow \quad f(x+h) \approx f(x) + hf'(x).$$

Applying this to $f(x) = -1/(1+x)^2$, which as the derivative $f'(x) = 2/(1+x)^3$ with $x = 1$ and $h = \varepsilon y_\varepsilon(t)$ yields that

$$-\frac{1}{(1+\varepsilon y_\varepsilon(t))^2} \approx -1 + 2\varepsilon y_\varepsilon(t) \approx -1 + 2\varepsilon y_0(t) + 2\varepsilon^2 y_1(t).$$

Substituting this and the approximation $y \approx y_0(t) + \varepsilon y_1(t)$ in the initial value problem (3.14) we obtain that

$$\frac{d^2}{dt^2}y_0(t) + \varepsilon\frac{d^2}{dt^2}y_1(t) \approx -1 + 2\varepsilon y_0(t) + 2\varepsilon^2 y_1(t),$$

with the initial conditions

$$y_0(0) + \varepsilon y_1(0) \approx 0, \quad \frac{d}{dt}y_0(0) + \frac{d}{dt}y_1(0) \approx 1.$$

The idea is now to equate all terms that have a factor of the same power in ε. For a given power p, these terms then are called to *scale with* ε^p. The terms scaling with ε^0 yield that

$$\frac{d^2}{dt^2}y_0(t) = -1, \qquad y_0(0) = 0, \quad \frac{d}{dt}y_0(0) = 1.$$

This is the problem that we had obtained from (3.14) by simply setting ε to zero. The solution $y_0(t) = t - \frac{1}{2}t^2$ is called *leading order solution* or *zeroth order solution*. The terms scaling with ε^1 yield a problem for the first order correction, or *first order solution*:

$$\frac{d^2}{dt^2}y_1(t) = 2y_0(t) = 2t - t^2, \qquad y_1(0) = 0, \quad \frac{d}{dt}y_1(0) = 0.$$

As y_0 is known we can easily compute the solution to this trivial differential equation, which is

$$y_1(t) = \frac{1}{3}t^3 - \frac{1}{12}t^4.$$

There is even a term scaling ε^2, which yields $0 = 2y_1(t)$. This contradicts what we had obtained for y_1 before. For the moment, let s simply get rid of this issue with the argument that ε^2 is much smaller than ε, and we may therefore drop all terms scaling with ε^2. Altogether, we then obtain

$$y_\varepsilon(t) \approx t - \frac{1}{2}t^2 + \varepsilon\big(\frac{1}{3}t^3 - \frac{1}{12}t^4\big).$$

In Figure 3.2 the solution y of (3.14) for $\varepsilon = 0.2$[4] together with the approximations to zeroth order y_0 as well as to first order $y_0 + \varepsilon y_1$ are displayed. We notice that y_0 is quite a bit way from y, however with the first order correction a much better approximation is obtained.

Finally, let us briefly come back to the issue of the ε^2 term. We could augment (3.15) with terms of higher powers in ε,

$$y_\varepsilon(t) \approx y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \varepsilon^3 y_3(t) + \dots$$

The right-hand-side function $-1/((1+\varepsilon y)^2)$ then also should yield an expansion as an ε series. For this purpose, a Taylor expansions at $\varepsilon = 0$ is appropriate, however such tools as the question of convergence of these expansions is beyond the scope of this module.

[4]This value corresponds to a projectile velocity about $3544 m/s$. High-performance guns used for assessing shields of satellites against micrometeorites and space debris achieve such velocities.

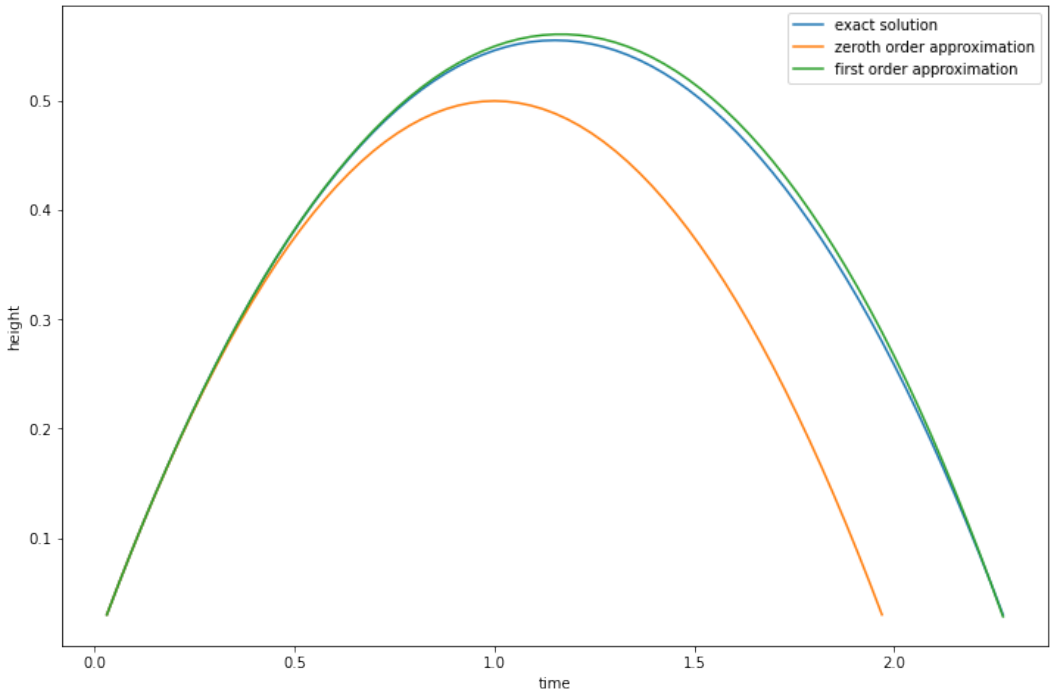


Figure 3.2: For $\varepsilon = 0.2$, solution $y(t)$ to (3.14) (blue, numerically approximated) and approximations $y_0(t)$ to zeroth order (orange) and $y_0(t) + \varepsilon y_1(t)$ to first order (green). See Section 3.3.2 for more detail.

Chapter 4

Systems

So far, we have studied *one-dimensional*, or *scalar* differential equations for one dependent variable. Many applications involve several species or objects that motivate models using several dependent variables. For instance, the infectious disease dynamics and the two-body problem that were mentioned in the introduction. In this section we will study such systems involving differential equations of first order. We will focus on 2×2 systems and note that many ideas and concepts generalise in a fairly generic way to $n \times n$ systems, $n \in \mathbb{N}$.

4.1 Introduction to systems

4.1.1 Notation

In this chapter we study 2×2 systems of first-order equations of the form

$$\begin{aligned} \frac{d}{dt}x_1(t) &= f_1\big(x_1(t), x_2(t), t\big), \\ \frac{d}{dt}x_2(t) &= f_2\big(x_1(t), x_2(t), t\big), \end{aligned} \qquad t \in (\alpha, \beta) \tag{4.1}$$

with some given functions $f_i : \mathbb{R}^2 \times (\alpha, \beta) \to \mathbb{R}$, $i = 1, 2$. So the 2×2 stands for two equations for two independent variables $x_1(t), x_2(t)$. We can write (4.1) in *vector form*

$$\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), t\big) \tag{4.2}$$

where $\boldsymbol{f} = (f_1, f_2) : \mathbb{R}^2 \times (\alpha, \beta) \to \mathbb{R}^2$.

The most *general form* of systems of first order equations is

$$\boldsymbol{\mathcal{F}}\Big(t, \boldsymbol{x}(t), \frac{d}{dt}\boldsymbol{x}(t)\Big) = 0^{1} \tag{4.3}$$

for some known function $\boldsymbol{\mathcal{F}} : (\alpha, \beta) \times (\mathbb{R}^2)^2 \to \mathbb{R}^2$, however we restrict our attention mostly to the specific form (4.1) or (4.2), which is obtained if we can resolve the equations in (4.3) for $\frac{d}{dt}\boldsymbol{x}(t)$.

[1] Properly, we should write $(0, 0)$ instead of 0 as this is a system of two equations. However, the notation 0 is common in the literature if it is clear from the context whether the scalar zero or a vector containing zeros is meant.

The *initial value problem* the consists of finding a solution to the differential equation together with an initial condition imposed at some time $t_0 \in [\alpha, \beta]$, which is of the form

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \quad \text{for some given } \boldsymbol{x}_0 \in \mathbb{R}^2. \tag{4.4}$$

Note that for the differential equation we assume that t belongs to the open interval (α, β). If $t_0 = \alpha$ or $t_0 = \beta$ then we implicitly assume that we can extend the solution to these boundary points.

Subject to suitable assumptions on the right-hand-side functions f_i one can prove that there is a unique solution to the initial value problem[2]. For this module we make the following assumption:

Assumption 4.1. *Unless stated otherwise, we assume that the functions f_i, $i = 1, 2$, in (4.1) are sufficiently 'nice' such that there is a unique solution of the system of differential equations (4.1) (or, equivalently, (4.2)) with an initial condition of the form (4.4).*

4.1.2 Relation to higher order equations

Lemma 4.2. *Any (scalar) second order differential equation can be written as a 2×2 system of first order equations.*

Proof. Recalling the definition (2.1) of a general scalar differential equation, let

$$\mathcal{F}(t, x(t), x'(t), x''(t)) = 0$$

be any second order equation. Let now $\eta(t) = x'(t)$. Then $\eta'(t) = x''(t)$ and we can reformulate the equation in the form

$$\begin{aligned} x'(t) - \eta(t) &= 0, \\ \mathcal{F}(t, x(t), \eta(t), \eta'(t)) &= 0. \end{aligned}$$

This is a 2×2 system of the most general form (4.3). □

Remark. *The above concepts for 2×2 systems can be generalised to $n \times n$ systems for $n \in \mathbb{N}$ in a straightforward way, including rewriting higher order equations of nth order as a system.*

Example. *Consider the second order equation*

$$\frac{d^2}{dt^2}x + \cos(t)\frac{d}{dt}x - x^2 = 0.$$

Writing $x_1(t) = x(t)$ and $x_2(t) = \frac{d}{dt}x_1(t)$ we obtain the system

$$\frac{d}{dt}x_1 - x_2 = 0, \quad \frac{d}{dt}x_2 + \cos(t)x_2 - x_1^2 = 0.$$

As we can rewrite the second order equation as a system this raises the question whether we can do so the other way round as well. In general, this is not possible, however sometimes we can do so.

[2] Well-posedness of initial value problems is a key topic of second year modules on differential equations.

Example 4.3. *Consider the system*

$$\begin{aligned}\frac{d}{dt}x_1 &= x_1 + x_2,\\ \frac{d}{dt}x_2 &= 4x_1 - 2x_2 + 4e^{-2t}.\end{aligned}$$

Rearranging the first equation we see that

$$x_2 = \frac{d}{dt}x_1 - x_1 \quad \Rightarrow \quad \frac{d}{dt}x_2 = \frac{d^2}{dt^2}x_1 - \frac{d}{dt}x_1.$$

Therefore, using the second equation,

$$\frac{d^2}{dt^2}x_1 - \frac{d}{dt}x_1 = \frac{d}{dt}x_2 = 4x_1 - 2x_2 + 4e^{-2t} = 4x_1 - 2\Big(\frac{d}{dt}x_1 - x_1\Big) + 4e^{-2t}.$$

Rearranging yields the second order equation

$$\frac{d^2}{dt^2}x_1 + \frac{d}{dt}x_1 - 6x_1 = 4e^{-2t}.$$

4.2 Autonomous systems

4.2.1 Notation

In this section we present some ideas to establish some qualitative results for autonomous systems, specifically related to their long-time behaviour. These are of the form

$$\begin{aligned}\frac{d}{dt}x_1(t) &= f_1\big(x_1(t), x_2(t)\big),\\ \frac{d}{dt}x_2(t) &= f_2\big(x_1(t), x_2(t)\big),\end{aligned} \qquad t \in (\alpha, \beta) \tag{4.5}$$

with some given functions $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$, $i = 1, 2$, so in contrast to (4.1) these functions f_i do not depend on the independent variable t.

Of particular interest are *stationary solutions* and their *stability*. Suppose $(\bar{x}_1, \bar{x}_2) \in \mathbb{R}^2$ is a point such that $f_1(\bar{x}_1, \bar{x}_2) = 0$ and $f_2(\bar{x}_1, \bar{x}_2) = 0$. Then we easily see that $(x_1(t), x_2(t)) = (\bar{x}_1, \bar{x}_2)$ is a stationary solution of (4.5). The stationary solution (or point) of (4.5) is called *stable* if solutions $(x_1(t), x_2(t))$ that start nearby remain close as time t increases.[3]

4.2.2 Phase portraits

A popular method to visualise solutions, to qualitatively predict their behaviour as time t increases, and to get an idea whether a stationary point is stable is the combine direction fields with phase portraits.

Definition 4.4. *1. A* direction field *for* (4.5) *is the attachment of an arrow proportional to* $(f_1(x_1, x_2), f_2(x_1, x_2))$ *at each point* $(x_1, x_2) \in \mathbb{R}^2$.

2. A phase portrait *(also called* phase diagram*) of* (4.5) *is the set of curves in the* x_1-x_2*-plane that are parametrised by solutions to* (4.5). *These curves traced out by solutions are so-called* trajectories.

[3] This is rather a notion. A precise definition of stability will be given in follow-up modules once measuring distances and limits will have been discussed in greater depth.

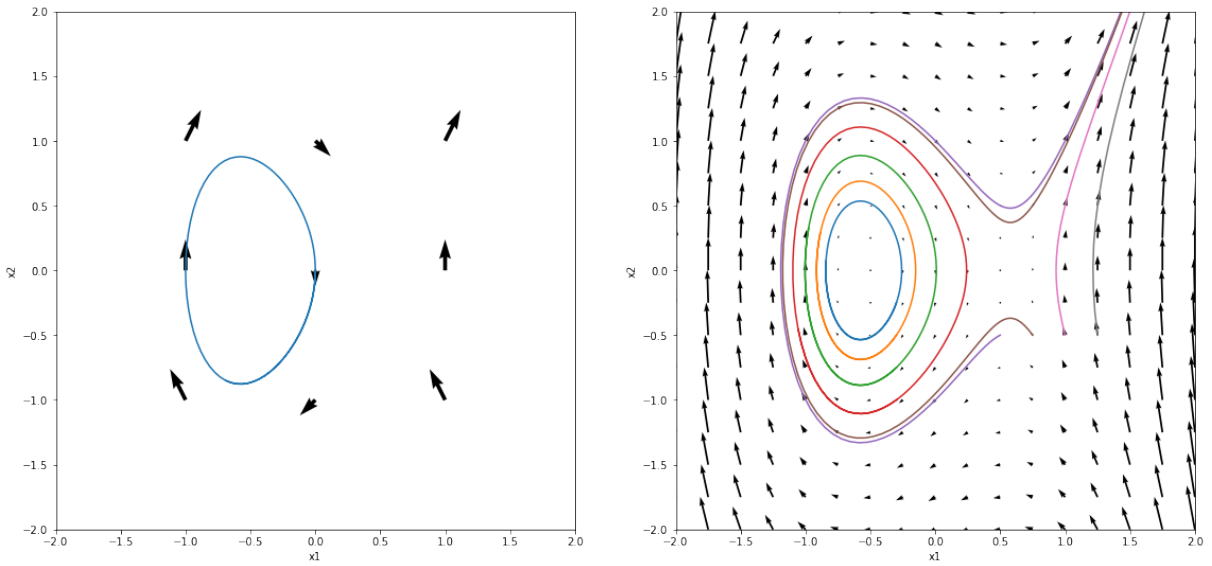


Figure 4.1: Direction field and phase portrait for the system (4.6).

In practice, a finite set of points in a bounded part of the plane is selected to illustrate a direction field. Similarly, a few solutions, usually around the stationary points, often suffice to give an idea of the phase portrait. If direction field and phase portrait are displayed together then the solution curves they are 'aligned' with the direction field in the sense that the arrows are tangential to the curves. The arrows then also indicate the direction of evolution as the independent variable increases.

Example. *Consider the system*

$$\begin{aligned} \frac{d}{dt}x_1 &= f_1(x_1, x_2) = x_2, \\ \frac{d}{dt}x_2 &= f_2(x_1, x_2) = 3x_1^2 - 1. \end{aligned} \tag{4.6}$$

Stationary points $(\bar{x}_2, \bar{x}_1)$ *have to satisfy* $0 = f_1(\bar{x}_1.\bar{x}_2) = \bar{x}_2$ *and* $0 = f_2(\bar{x}_1, \bar{x}_2) = 3\bar{x}_1^2 - 1$, *so the set of stationary points if given by* $\{(\sqrt{1/3}, 0), (-\sqrt{1/3}, 0)\}$.

A couple of function values for $\boldsymbol{f}(x_1, x_2) = (f_1(x_1, x_2), f_2(x_1, x_2))$ *are given by*

1	$(1,2)$	$(1-1)$	(12)
0	$(0,2)$	$(0,-1)$	$(0,2)$
-1	$(-1,2)$	$(-1,-1)$	$(-1,2)$
x_2/x_1	-1	0	1

In Figure 4.1 on the left, arrows proportional to $\boldsymbol{f}(x_1, x_2)$ *are displayed in these points. In addition, the trajectory starting at* $\boldsymbol{x}(0) = (0, 0)$ *is displayed. Is also goes throug the point* $(-1, 0)$. *Both in this point and the initial point we see that the arrow of the direction field is tangential to the curve. On the right, Figure 4.1 gives a more complete impression of the phase portrait.*

We notice that the solutions around the stationary point $(-\sqrt{1/3}, 0)$ *form closed curves. In particular, they stay nearby. We conclude that this stationary point is stable. In turn, solutions*

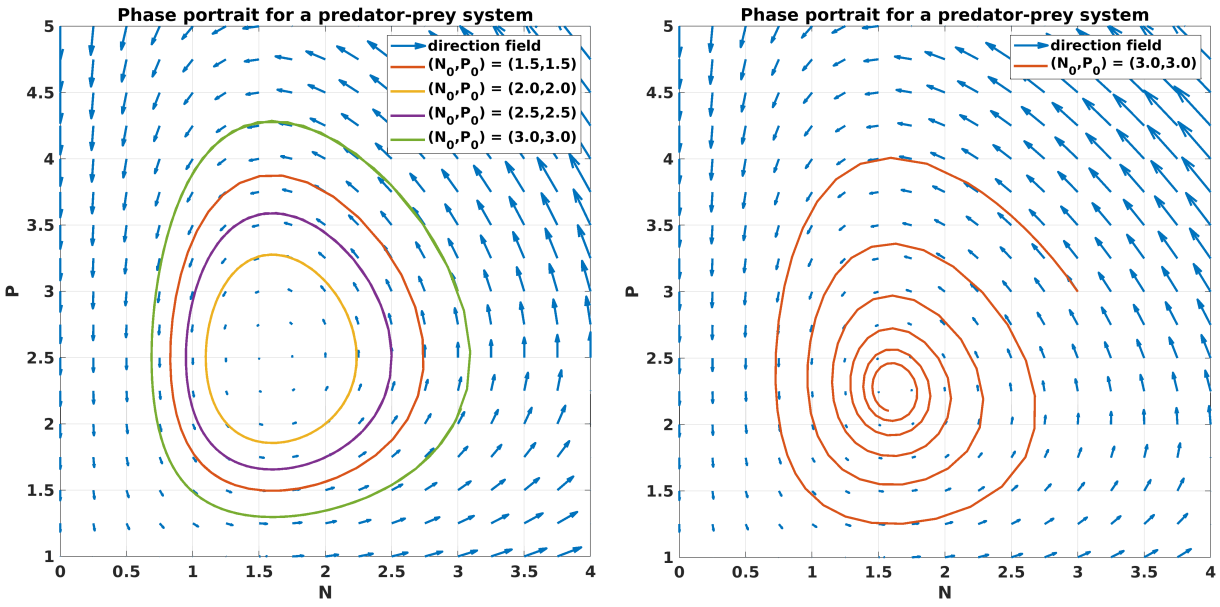


Figure 4.2: Phase portraits for the predator-prey model (4.7) and solutions for several initial points $(N(0), P(0)) = (N_0, P_0)$. The parameters are $a = 2.5$, $b = 1$, $c = 1$, $d = 1.6$. On the left, $e = 0.0$ (a stationary point is $(\bar{N}, \bar{P}) = (1.6, 2.5)$), on the right, $e = 0.15$ (a stationary point is $(\bar{N}, \bar{P}) = (1.6, 2.26)$).

starting to the right of the other stationary point $(\sqrt{1/3}, 0)$ *drift off to the upper right and don't stay nearby, so this point is unstable.*

Example 4.5. *Let us consider a* Lotka-Volterra *system modelling* predator-prey *interaction of the form*

$$\begin{aligned} \frac{d}{dt}N(t) &= N(t)\big(a - (bP(t) + eN(t))\big), \\ \frac{d}{dt}P(t) &= P(t)\big(cN(t) - d\big), \end{aligned} \qquad t \in \mathbb{R}, \tag{4.7}$$

with given parameters $a, b, c, d > 0$ *and* $e \geq 0$. *Here,* N *stands for a prey population and* P *for a predator population. Note that if* $b = 0$ *was permitted then the equation for* N *would be the population growth equation seen in the introduction. If predators are present then there is a negative impact on the growth given by* $-bN(t)P(t)$. *In turn, the predator population can only grow if prey is present, which is modelled by the term* $cN(t)P(t)$. *Otherwise, they die with a rate of* $-d$. *All variables are assumed nondimensional already.*

Stationary points $(\bar{N}, \bar{P})$ *have to satisfy*

$$\bar{N}\big(a - (b\bar{P} + e\bar{N})\big) = 0, \quad \bar{P}\big(c\bar{N} - d\big) = 0.$$

The second equation is satisfied if $\bar{P} = 0$ *or* $\bar{N} = d/c$. *If* $\bar{P} = 0$ *then the first equation is satisfied if* $\bar{N} = 0$ *or* $\bar{N} = a/e$ *(if* $e \neq 0$*). If* $\bar{N} = d/c$ *then the first equation is satisfied if* $\bar{P} = (ca - ed)/(cb)$. *Altogether, the solutions to these equations are (the last only if* $e \neq 0$*)*

$$(\bar{N}, \bar{P}) \in \Big\{(0,0), \Big(\frac{d}{c}, \frac{ca - ed}{cb}\Big), \Big(\frac{a}{e}, 0\Big)\Big\}.$$

Figure 4.2 displays two phase portraits, one with $e = 0$ and one with $e = 0.15$ (the other parameters are the same, see the figure caption for further detail). Some solutions are also displayed. The stationary points with $\bar{P} \neq 0$ are $(1.6, 2.5)$ if $e = 0$ and $(1.6, 2.26)$ if $e = 0.15$.

In the case $e = 0$ we can derive implicit solutions. Assuming that $N(t), P(t) > 0$ we note that

$$-\frac{cN(t)-d}{N(t)}\frac{d}{dt}N(t)+\frac{a-bP(t)}{P(t)}\frac{d}{dt}P(t) = -(cN(t)-d)(a-bP(t))+(a-bP(t))(cN(t)-d) = 0.$$

Re-writing the terms on the left-hand-side, this is equivalent to

$$\frac{d}{dt}\Big(-cN(t)+d\log(N(t))+a\log(P(t))-bP(t)\Big) = 0.$$

Integrating with respect to time we obtain that

$$-cN(t)+d\log(N(t))+a\log(P(t))-bP(t) = constant.$$

Therefore, the closed solutions displayed in Figure 4.2 on the left are level sets of the function

$$\mathcal{L}(N,P) = -cN+d\log(N)+a\log(P)-bP.$$

If $e > 0$ then we observe that the solution spirals inwards towards the stationary point $(\bar{N}, \bar{P}) = (1.6, 2.26)$ (recall from the definition that the arrows indicate the direction of the evolution). The stationary point therefore seems stable.

4.2.3 Linearisation and linear stability

Let $(\bar{x}_1, \bar{x}_2)$ denote a stationary solution (or point) of (4.5). In order to get some more insight into whether $(\bar{x}_1, \bar{x}_2)$ is stable we study solutions of the form

$$(x_1(t), x_2(t)) = (\bar{x}_1 + \xi_1(t), \bar{x}_2 + \xi_2(t))$$

where $(\xi_1(t), \xi_2(t))$ is a 'small' deviation from the stationary point. First, note that then $\frac{d}{dt}(x_1(t), x_2(t)) = \frac{d}{dt}(\xi_1(t), \xi_2(t))$ so that the system for $(\xi_1(t), \xi_2(t))$ becomes

$$\begin{aligned}\frac{d}{dt}\xi_1(t) &= f_1\big(\bar{x}_1+\xi_1(t), \bar{x}_2+\xi_2(t)\big),\\ \frac{d}{dt}\xi_2(t) &= f_2\big(\bar{x}_1+\xi_1(t), \bar{x}_2+\xi_2(t)\big),\end{aligned} \qquad t \in (\alpha, \beta).$$

Now, we approximate the right-hand-side by[4]

$$\begin{aligned}f_1\big(\bar{x}_1+\xi_1(t), \bar{x}_2+\xi_2(t)\big) &\approx \frac{\partial}{\partial x_1}f_1(\bar{x}_1,\bar{x}_2)\xi_1(t)+\frac{\partial}{\partial x_2}f_1(\bar{x}_1,\bar{x}_2)\xi_2(t),\\ f_2\big(\bar{x}_1+\xi_1(t), \bar{x}_2+\xi_2(t)\big) &\approx \frac{\partial}{\partial x_1}f_2(\bar{x}_1,\bar{x}_2)\xi_1(t)+\frac{\partial}{\partial x_2}f_2(\bar{x}_1,\bar{x}_2)\xi_2(t).\end{aligned} \tag{4.8}$$

Here, the *partial derivatives* $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$ with respect to the first and second argument of the functions f_1 and f_2 appear. There are defined by differentiating with respect to one argument whilst keeping the other fixed, for instance,

$$\frac{\partial}{\partial x_2}f_1(x_1,x_2) = \lim_{\Delta x_2 \to 0}\frac{f_1(x_1, x_2+\Delta x_2)-f_1(x_1,x_2)}{\Delta x_2}.$$

[4]Those who know about Taylor-expansions will recognise these approximations as the first-order expansions.

The idea in the above approximation (4.8) simply is to drop the limit and keep the difference on the left-hand-side in the definition of the partial derivative. This is justified by the assumption that $\xi_1(t)$ and $\xi_2(t)$ are 'small' deviations. With this approximation we obtain the following system for $(\xi_1(t), \xi_2(t))$:

$$\begin{aligned} \frac{d}{dt}\xi_1(t) &= \frac{\partial}{\partial x_1} f_1(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_1(\bar{x}_1, \bar{x}_2)\xi_2(t), \\ \frac{d}{dt}\xi_2(t) &= \frac{\partial}{\partial x_1} f_2(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_2(\bar{x}_1, \bar{x}_2)\xi_2(t), \end{aligned} \qquad t \in (\alpha, \beta). \tag{4.9}$$

This system is *linear* in $(\xi_1(t), \xi_2(t))$. We therefore talk about a *linearisation* of (4.5) around the stationary solution $(\bar{x}_1, \bar{x}_2)$. In particular, the coefficients $\frac{\partial}{\partial x_2} f_1(\bar{x}_1, \bar{x}_2)$ etc are constant, and in that case we can find explicit solutions.

The *principle of linearised stability* now relates the stability properties of the original system (4.5) around the stationary point $(\bar{x}_1, \bar{x}_2)$ to the stability properties of the linearised system (4.9) around $(0, 0)$. Roughly, the stationary point is stable if and only if the linearised systems is stable around the origin. We will not precisely formulate the principle in this module. It is known as *Hartman-Grobman* theorem in the literature, see [4], Sec 32, for more detail and further references. We use it here as a motivation to study linear systems of differential equations. Before doing so, let us look at some examples of linearisations.

Example. *Recall the system* (4.6),

$$\begin{aligned} \frac{d}{dt}x_1 &= f_1(x_1, x_2) = x_2, \\ \frac{d}{dt}x_2 &= f_2(x_1, x_2) = 3x_1^2 - 1 \end{aligned}$$

which we had shown to have the stationary points $\{(\sqrt{1/3}, 0), (-\sqrt{1/3}, 0)\}$. *We have that*

$$\begin{aligned} \frac{\partial}{\partial x_1} f_1(x_1, x_2) &= 0, \quad & \frac{\partial}{\partial x_2} f_1(x_1, x_2) &= 1, \\ \frac{\partial}{\partial x_1} f_2(x_1, x_2) &= 6x_1, \quad & \frac{\partial}{\partial x_2} f_1(x_1, x_2) &= 0. \end{aligned}$$

The linearised system around the point $(\bar{x}_1, \bar{x}_2) = (\sqrt{1/3}, 0)$ *therefore reads*

$$\begin{aligned} \frac{d}{dt}\xi_1(t) &= \frac{\partial}{\partial x_1} f_1(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_1(\bar{x}_1, \bar{x}_2)\xi_2(t) = \xi_2(t), \\ \frac{d}{dt}\xi_2(t) &= \frac{\partial}{\partial x_1} f_2(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_2(\bar{x}_1, \bar{x}_2)\xi_2(t) = \sqrt{12}\xi_1(t), \end{aligned}$$

and around the point $(\bar{x}_1, \bar{x}_2) = (-\sqrt{1/3}, 0)$ *we obtain*

$$\begin{aligned} \frac{d}{dt}\xi_1(t) &= \frac{\partial}{\partial x_1} f_1(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_1(\bar{x}_1, \bar{x}_2)\xi_2(t) = \xi_2(t), \\ \frac{d}{dt}\xi_2(t) &= \frac{\partial}{\partial x_1} f_2(\bar{x}_1, \bar{x}_2)\xi_1(t) + \frac{\partial}{\partial x_2} f_2(\bar{x}_1, \bar{x}_2)\xi_2(t) = -\sqrt{12}\xi_1(t). \end{aligned}$$

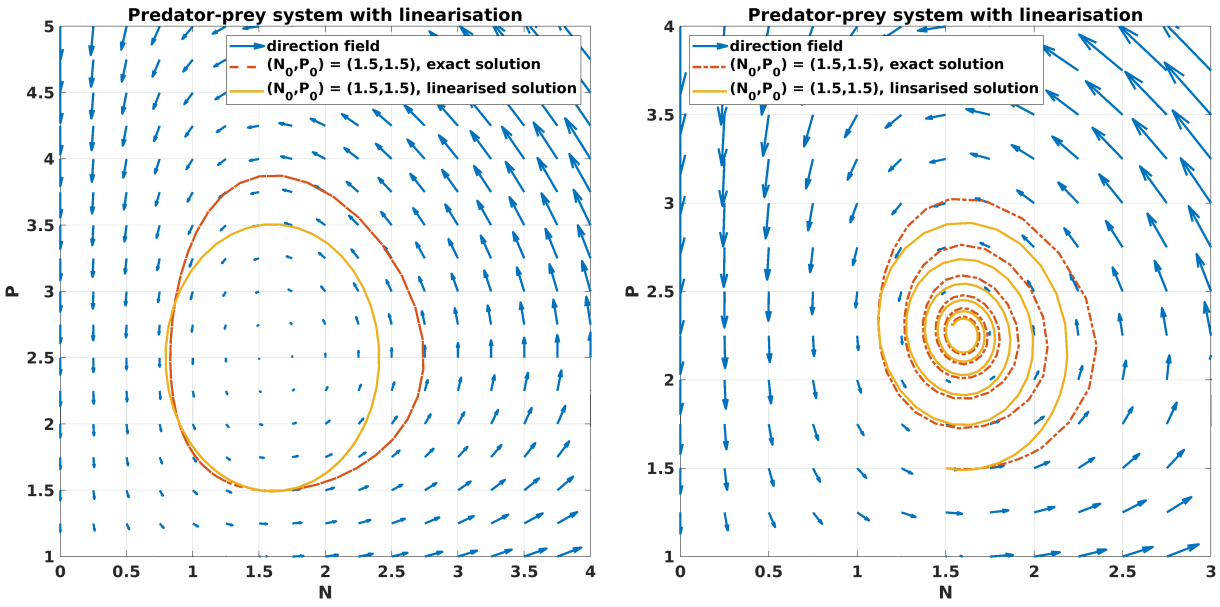


Figure 4.3: A solution $(N(t), P(t)$ of (4.7) together with the linearised solution $(\bar{N}+n(t), \bar{P}+p(t))$ starting in the same initial point $(N_0, P_0) = (1.5, 1.5)$. The parameters are $a = 2.5$, $b = 1$, $c = 1$, $d = 1.6$. On the left, $e = 0.0$, on the right, $e = 0.15$.

Example 4.6. *Let us linearise the predator-prey model* (4.7) *around the stationary point* $(\bar{N}, \bar{P}) = (\frac{d}{c}, \frac{ca-ed}{cb})$. *We have that*

$$f_N(N,P) = N(a-(bP+eN)) \quad \Rightarrow \quad \frac{\partial}{\partial N} f_N(N,P) = a - bP - 2eN, \quad \frac{\partial}{\partial P} f_N(N,P) = -bN,$$
$$f_P(N,P) = P(cN-d) \quad \Rightarrow \quad \frac{\partial}{\partial N} f_P(N,P) = cP, \quad \frac{\partial}{\partial P} f_P(N,P) = cN - d.$$

In the point $(\bar{N}, \bar{P})$ *we then obtain*

$$\frac{\partial}{\partial N} f_N(\bar{N},\bar{P}) = a - b\tfrac{ca-ed}{cb} - 2e\tfrac{d}{c} = -\tfrac{ed}{c}, \qquad \frac{\partial}{\partial P} f_N(\bar{N},\bar{P}) = -\tfrac{bd}{c},$$
$$\frac{\partial}{\partial N} f_P(\bar{N},\bar{P}) = \tfrac{ca-ed}{b}, \qquad \frac{\partial}{\partial P} f_P(\bar{N},\bar{P}) = c\tfrac{d}{c} - d = 0.$$

Let us denote small deviations by $(n(t), p(t))$. *The linearised system corresponding to* (4.9) *now reads*

$$\frac{d}{dt} n(t) = \frac{\partial}{\partial N} f_N(\bar{N},\bar{P}) n(t) + \frac{\partial}{\partial P} f_N(\bar{N},\bar{P}) p(t) = -\tfrac{ed}{c} n(t) - \tfrac{bd}{c} p(t),$$
$$\frac{d}{dt} p(t) = \frac{\partial}{\partial N} f_P(\bar{N},\bar{P}) n(t) + \frac{\partial}{\partial P} f_P(\bar{N},\bar{P}) p(t) = \tfrac{ca-ed}{b} n(t).$$

We can write it in matrix-vector form

$$\frac{d}{dt}\begin{pmatrix} n(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} -\frac{ed}{c} & -\frac{bd}{c} \\ \frac{ca-ed}{b} & 0 \end{pmatrix} \begin{pmatrix} n(t) \\ p(t) \end{pmatrix}.$$

Figure 4.3 displays solutions of the original system $(N(t), P(t))$ *and the linearised system* $(\bar{N} + n(t), \bar{P} + p(t))$. *Recall that the case* $e = 0$ *features solutions in the form of closed curves.*

This is also the case for the linearised solutions, but which here are ellipses, and the shapes differ. Similarly, in the case $e = 0.15$ both solutions show the same qualitative behaviour of spiralling into the stationary point, however the shapes are different.

4.3 Homogeneous linear systems with constant coefficients

The linearisation motivates to study linear systems of differential equations with constant coefficients, so let us look into linear systems of the form

$$\begin{aligned} \frac{d}{dt}x_1(t) &= a_{1,1}x_1(t) + a_{1,2}x_2(t), \\ \frac{d}{dt}x_2(t) &= a_{2,1}x_1(t) + a_{2,2}x_2(t), \end{aligned} \qquad t \in \mathbb{R},$$

where the $a_{i,j}$ now are independent of t. These can be formulated in the matrix-vector form

$$\frac{d}{dt}\boldsymbol{x}(t) = A\boldsymbol{x}(t), \qquad A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad \boldsymbol{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}. \tag{4.10}$$

We found that exponential functions have been working well with the linear equations that we saw so far, so let us try to find a solution in the form

$$\boldsymbol{x}(t) = e^{\lambda t}\boldsymbol{v} \quad \text{with } \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2 \text{ and } \lambda \in \mathbb{R}.$$

Substituting this approach in the system yields

$$\lambda e^{\lambda t}\boldsymbol{v} = \frac{d}{dt}\boldsymbol{x}(t) = A\boldsymbol{x}(t) = e^{\lambda t}A\boldsymbol{v}.$$

Using that $e^{\lambda t} \neq 0$ we obtain that

$$\lambda\boldsymbol{v} = A\boldsymbol{v}. \tag{4.11}$$

So $\boldsymbol{x}(t) = e^{\lambda t}\boldsymbol{v}$ is a solution if $\boldsymbol{v}$ is an eigenvector of A with corresponding eigenvalue λ. These eigenvalues might be complex, though, so we should extend the above approach to the case that

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{C}^2 \text{ and } \lambda \in \mathbb{C},$$

and investigate whether we can find derive real-valued solutions.

In general, the 2×2 matrix A has two distinct eigenvalues. Assuming this is the case we denote them by λ_1, λ_2 and corresponding eigenvectors by $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}$. Note that these eigenvectors are linearly independent[5]. The functions $\boldsymbol{x}^{(1)}(t) = e^{\lambda_1 t}\boldsymbol{v}^{(1)}$ and $\boldsymbol{x}^{(2)}(t) = e^{\lambda_1 t}\boldsymbol{v}^{(2)}$ then are solutions of (4.10) and, moreover, are linearly independent (easy to see, if we evaluate them at $t = 0$ we obtain the $\boldsymbol{v}^{(i)}$, which are linearly independent). One can show that[6] all solutions of (4.10) are of the form of the complementary function,

$$\boldsymbol{x}(t) = l_1\boldsymbol{x}^{(1)}(t) + l_2\boldsymbol{x}^{(2)}(t) = l_1 e^{\lambda_1 t}\boldsymbol{v}^{(1)} + l_2 e^{\lambda_2 t}\boldsymbol{v}^{(2)}, \quad l_1, l_2 \in \mathbb{C},$$

and the unique solution to the initial value problem consisting of (4.10) and (4.4) is obtained by solving the 2×2 system

$$\boldsymbol{x}_0 = l_1\boldsymbol{x}^{(1)}(t_0) + l_2\boldsymbol{x}^{(2)}(t_0).$$

Some natural questions are:

[5] this will be proved in Linear Algebra

[6] The proof is similar to the one for second order linear equations.

- Can we guarantee real-valued solutions?
- Is the stationary solution $(0,0)$ stable?

Before we answer these questions for some (arguably, most) cases we recall some properties of eigenvalues and eigenvectors.

4.3.1 Eigenvalues and eigenvectors

The eigenvalues of a matrix $A \in \mathbb{R}^{2\times 2}$ are given as the roots to its *characteristic polynomial*. So with I denoting the identity matrix they satisfy the equation

$$\begin{aligned} 0 &= \det(\lambda I - A) \\ &= \det \begin{pmatrix} \lambda - a_{1,1} & -a_{1,2} \\ -a_{2,1} & \lambda - a_{2,2} \end{pmatrix} \\ &= (\lambda - a_{1,1})(\lambda - a_{2,2}) - a_{2,1}a_{1,2} \\ &= \lambda^2 - (a_{1,1} + a_{2,2})\lambda + a_{1,1}a_{2,2} - a_{2,1}a_{1,2} \\ &= \lambda^2 - \operatorname{trace}(A)\lambda + \det(A). \end{aligned}$$

Note that, equivalently, we may consider the equation $0 = \det(A - \lambda I)$. Let us denote the solutions by

$$\begin{aligned} \lambda_{1,2} &= \frac{\operatorname{trace}(A)}{2} \pm \frac{1}{2}\sqrt{(\operatorname{trace}(A))^2 - 4\det(A)} \\ &= \frac{a_{1,1} + a_{2,2}}{2} \pm \frac{1}{2}\sqrt{(a_{1,1} + a_{2,2})^2 - 4(a_{1,1}a_{2,2} - a_{2,1}a_{1,2})} \\ &= \frac{a_{1,1} + a_{2,2}}{2} \pm \frac{1}{2}\sqrt{(a_{1,1} - a_{2,2})^2 + 4a_{2,1}a_{1,2}}. \end{aligned}$$

Corresponding eigenvectors then are the solutions $\boldsymbol{v}^{(i)} \in \mathbb{C}^2$, $\boldsymbol{v}^{(i)} \neq (0,0)$, of

$$(\lambda_i I - A)\boldsymbol{v}^{(i)} = 0, \quad i = 1, 2.$$

There are two equations, and if $\boldsymbol{v}^{(i)}$ is an eigevector then also every multiple $c\boldsymbol{v}^{(i)}$, $c \in \mathbb{C}$, is an eigenvector again. In practice, it is sufficient to consider only one of the two equations.

Example 4.7. *Let us consider some examples and compute the eigenvalues and some corresponding eigenvectors of the following matrices.*

1. *Consider the matrix*

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}.$$

In this case $\operatorname{trace}(A) = 1$ *and* $\det(A) = -2$ *so that*

$$\lambda_{1,2} = \frac{\operatorname{trace}(A)}{2} \pm \frac{1}{2}\sqrt{(\operatorname{trace}(A))^2 - 4\det(A)} = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - 4 * (-2)} = \frac{1}{2} \pm \frac{3}{2} = \{2, -1\}.$$

An eigenvector $\boldsymbol{v}^{(1)} = (v_1^{(1)}, v_2^{(1)})$ *corresponding to* λ_1 *satisfies*

$$0 = (\lambda_1 I - A)\boldsymbol{v} = \begin{pmatrix} 2-1 & -2 \\ -1 & 2-0 \end{pmatrix} \begin{pmatrix} v_1^{(1)} \\ v_2^{(1)} \end{pmatrix} = \begin{pmatrix} v_1^{(1)} - 2v_2^{(1)} \\ -v_1^{(1)} + 2v_2^{(1)} \end{pmatrix}.$$

We see that a solution is given by $\boldsymbol{v}^{(1)} = (2,1)$. *Let us briefly check that this is an eigenvalue for* λ_1*:*

$$A\boldsymbol{v}^{(1)} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \lambda_1 \boldsymbol{v}^{(1)}.$$

Regarding λ_2*, a corresponding eigenvector* $\boldsymbol{v}^{(2)} = (v_1^{(2)}, v_2^{(2)})$ *satisfies*

$$0 = (\lambda_2 I - A)\boldsymbol{v} = \begin{pmatrix} -1-1 & -2 \\ -1 & -1-0 \end{pmatrix} \begin{pmatrix} v_1^{(2)} \\ v_2^{(2)} \end{pmatrix} = \begin{pmatrix} -2v_1^{(2)} - 2v_2^{(2)} \\ -v_1^{(2)} - v_2^{(2)} \end{pmatrix}.$$

A solution to these equations is $\boldsymbol{v}^{(2)} = (1,-1)$*, and it can be checked similarly.*

2. *Let now*

$$A = \begin{pmatrix} 2 & 5 \\ -2 & 0 \end{pmatrix}.$$

In this case $\operatorname{trace}(A) = 2$ *and* $\det(A) = 10$ *so that*

$$\lambda_{1,2} = \frac{\operatorname{trace}(A)}{2} \pm \frac{1}{2}\sqrt{(\operatorname{trace}(A))^2 - 4\det(A)} = 1 \pm \frac{1}{2}\sqrt{4 - 4 * 10} = 1 \pm \mathbf{i}3.$$

Starting with $\lambda_1 = 1 + \mathbf{i}3$*, the system for a corresponding eigenvector* $\boldsymbol{v}^{(1)} = (v_1^{(1)}, v_2^{(1)})$ *reads*

$$0 = (\lambda_1 I - A)\boldsymbol{v}^{(1)} = \begin{pmatrix} 1+\mathbf{i}3-2 & -5 \\ 2 & 1+\mathbf{i}3-0 \end{pmatrix} \begin{pmatrix} v_1^{(1)} \\ v_2^{(1)} \end{pmatrix} = \begin{pmatrix} (-1+\mathbf{i}3)v_1^{(1)} - 5v_2^{(1)} \\ 2v_1^{(1)} + (1+\mathbf{i}3)v_2^{(1)} \end{pmatrix}.$$

If we choose $\boldsymbol{v}^{(1)} = (5, -1+\mathbf{i}3)$ *then clearly the first equation is satisfied. The second one as well because it is a multiple of the first one, but let us briefly check as this is may note be immediately clear here:*

$$2v_1^{(1)} + (1+\mathbf{i}3)v_2^{(1)} = 2*5 + (1+\mathbf{i}3)*(-1+\mathbf{i}3) = 10 + (-1+\mathbf{i}3-\mathbf{i}3+\mathbf{i}^2 9) = 10 + (-1-9) = 0.$$

Note that $\lambda_2 = 1 - \mathbf{i}3 = \overline{\lambda}_1$ *are complex conjugates of each other. Because the matrix* A *has real entries this is also true for the corresponding eigenvectors. So here* $\boldsymbol{v}^{(2)} = \overline{\boldsymbol{v}^{(1)}} = (5, -1-\mathbf{i}3)$ *is an eigenvector of* A *with eigenvalue* λ_2*. Let us briefly check this:*

$$\begin{aligned} A\boldsymbol{v}^{(2)} = \begin{pmatrix} 2 & 5 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} 5 \\ -1-\mathbf{i}3 \end{pmatrix} &= \begin{pmatrix} 10 - 5 - \mathbf{i}15 \\ -10 \end{pmatrix} \\ &= \begin{pmatrix} (1-\mathbf{i}3)*5 \\ (1-\mathbf{i}3)*(-1-\mathbf{i}3) \end{pmatrix} = (1-\mathbf{i}3)\begin{pmatrix} 5 \\ -1-\mathbf{i}3 \end{pmatrix} = \lambda_2 \boldsymbol{v}^{(2)}, \end{aligned}$$

where we used that

$$(1-\mathbf{i}3)*(-1-\mathbf{i}3) = -1 - \mathbf{i}3 + \mathbf{i}3 + \mathbf{i}^2 9 = -1 - 9 = -10.$$

3. *Finally, let us consider*

$$A = \begin{pmatrix} 5 & -4 \\ 1 & 1 \end{pmatrix}.$$

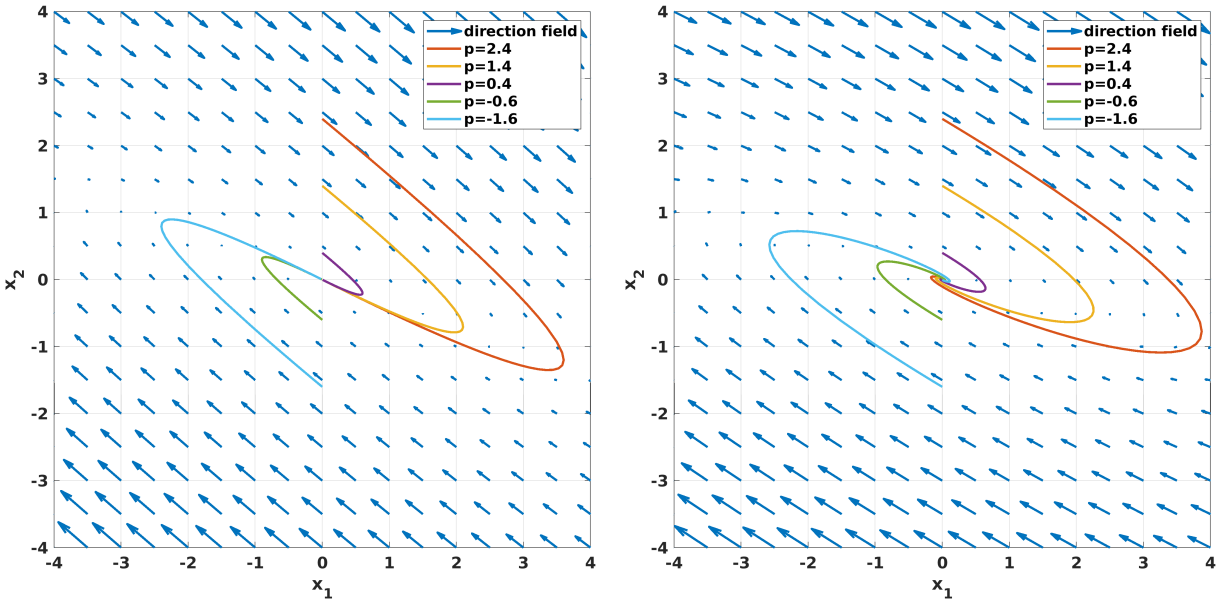


Figure 4.4: Direction fields and phase portraits for systems discussed in Example 4.8 (left) and Example 4.9 (right). Solutions start at points $(0, p)$ where $p = 2.4, 1.4, 0.4, -0.6, -1.6$.

In this case, $\operatorname{trace}(A) = 6$ *and* $\det(A) = 9$ *are such that* $(\operatorname{trace}(A))^2 - 4\det(A) = 36 - 4*9 = 0$ *so that the matrix only has one eigenvalue,*

$$\lambda_{1,2} = \frac{\operatorname{trace}(A)}{2} \pm \frac{1}{2}\sqrt{(\operatorname{trace}(A))^2 - 4\det(A)} = 3 \pm 0.$$

We can only expect to find one corresponding eigenvector (or, rather, eigendirection, recall that multiples of eigenvectors are eigenvectors again), and a short calculation shows that $\boldsymbol{v} = (2, 1)$ *is such an eigenvector. Finding two linearly independent solutions* $\boldsymbol{x}^{(1)}(t), \boldsymbol{x}^{(2)}(t)$ *of* (4.10) *is a bit more involved.*

4.3.2 Distinct real eigenvalues

If $(a_{1,1} - a_{2,2})^2 + 4a_{2,1}a_{1,2} > 0$ or, equivalently, $(\operatorname{trace}(A))^2 - 4\det(A) > 0$ then the two eigenvalues are real and distinct. The corresponding eigenspaces then are distinct and we find linearly independent corresponding eigenvectors $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)} \in \mathbb{R}^2$. The functions

$$\boldsymbol{x}^{(1)}(t) = e^{\lambda_1 t}\boldsymbol{v}^{(1)}, \quad \boldsymbol{x}^{(2)}(t) = e^{\lambda_2 t}\boldsymbol{v}^{(2)}$$

are then are linearly independent solutions, and at time t_0 the equations

$$\boldsymbol{x}_0 = \boldsymbol{x}(t_0) = l_1 e^{\lambda_1 t_0}\boldsymbol{v}^{(1)} + l_2 e^{\lambda_2 t_0}\boldsymbol{v}^{(2)}$$

have a unique solution that determine $l_1, l_2 \in \mathbb{R}$.

Note that $\boldsymbol{x}^{(i)}(t)$ remains bounded if $\lambda_i \leq 0$, $i = 1, 2$, because the prefactors $e^{\lambda_i t}$ either are 1 (if $\lambda_i = 0$) or decrease to zero (if $\lambda_i < 0$). So if both eigenvalues are non-positive then the stationary solution $(0, 0)$ is stable. In turn, if one of the eigenvalues is positive, say λ_1, then $e^{\lambda_1 t}$ keeps on increasing as t increases. The corresponding solution $\boldsymbol{x}^{(1)}(t) = e^{\lambda_1 t}\boldsymbol{v}^{(1)}$ then moves away from $(0, 0)$, so the stationary points is unstable.

Example 4.8. *Consider (4.10) with an initial condition of the form (4.4) where $t_0 = 0$ and*

$$A = \begin{pmatrix} 2 & 6 \\ -2 & -5 \end{pmatrix}, \quad \boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

As $\operatorname{trace}(A) = -3$ *and* $\det(A) = 2$ *the eigenvalues of A satisfy $\lambda^2 + 3\lambda + 2 = 0$ and are*

$$\lambda_{1,2} = -\frac{3}{2} \pm \frac{1}{2}\sqrt{3^2 - 4 * 2} = \{-1, -2\}.$$

An eigenvector $\boldsymbol{v}^{(1)}$ has to satisfy

$$0 = (\lambda_1 I - A)\boldsymbol{v}^{(1)} = \begin{pmatrix} -3 & -6 \\ 2 & 4 \end{pmatrix} \boldsymbol{v}^{(1)}.$$

A possible choice is $\boldsymbol{v}^{(1)} = (-2, 1)$. Proceeding similarly for λ_2 we find that a corresponding eigenvector is given by $\boldsymbol{v}^{(2)} = (-3, 2)$. A general solution to the differential equations is thus given by

$$\boldsymbol{x}(t) = l_1 e^{\lambda_1 t}\boldsymbol{v}^{(1)} + l_2 e^{\lambda_2 t}\boldsymbol{v}^{(2)} = l_1 e^{-t}\begin{pmatrix} -2 \\ 1 \end{pmatrix} + l_2 e^{-2t}\begin{pmatrix} -3 \\ 2 \end{pmatrix}, \quad l_1, l_2 \in \mathbb{R}.$$

The initial condition is satisfied if

$$\boldsymbol{x}_0 = \boldsymbol{x}(0) = l_1\boldsymbol{v}^{(1)} + l_2\boldsymbol{v}^{(2)} \quad \Leftrightarrow \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} = l_1\begin{pmatrix} -2 \\ 1 \end{pmatrix} + l_2\begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

The solution to this 2×2 systems is $l_1 = -2$, $l_2 = 1$, so the unique solution of (4.10), (4.4) with the above data is

$$\boldsymbol{x}(t) = -2e^{-t}\begin{pmatrix} -2 \\ 1 \end{pmatrix} + e^{-2t}\begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

As $\lambda_2 < \lambda_1 < 0$ the solution converges to $(0, 0)$ as $t \to \infty$. In particular, it remains close to the stationary solution. This is actually true for any initial point, so the stationary point is stable. Figure 4.4 (left) give an impression of the direction field and the behaviour of solutions starting at different points. In particular, the convergence to $(0, 0)$ is visible.

4.3.3 Distinct complex eigenvalues

If $(a_{1,1}-a_{2,2})^2+4a_{2,1}a_{1,2} < 0$ or, equivalently, $(\operatorname{trace}(A))^2-4\det(A) > 0$ then the two eigenvalues are complex and satisfy

$$\lambda_2 = \overline{\lambda_1} = \frac{a_{1,1} + a_{2,2}}{2} - \frac{\mathbf{i}}{2}\sqrt{4a_{2,1}a_{1,2} - (a_{1,1} - a_{2,2})^2} = \frac{\operatorname{trace}(A)}{2} - \frac{\mathbf{i}}{2}\sqrt{4\det(A) - (\operatorname{trace}(A))^2}.$$

Moreover $\boldsymbol{v}^{(1)} \in \mathbb{C}^2$ is an eigenvector corresponding to λ_1 if and only if its conjugate $\overline{\boldsymbol{v}^{(1)}}$ is an eigenvector corresponding to $\lambda_2 = \overline{\lambda_1}$. Writing $\boldsymbol{v}^{(2)} = \overline{\boldsymbol{v}^{(1)}}$ we then also obtain that

$$\overline{e^{\lambda_1 t}\boldsymbol{v}^{(1)}} = e^{\lambda_2 t}\boldsymbol{v}^{(2)}.$$

Recall that any linear combinations

$$l_1 e^{\lambda_1 t}\boldsymbol{v}^{(1)} + l_2 e^{\lambda_2 t}\boldsymbol{v}^{(2)}, \quad l_1, l_2 \in \mathbb{C},$$

are solutions to the system of differential equations. So let us consider the real and the imaginary parts of $e^{\lambda_1 t}\boldsymbol{v}^{(1)}$, which are such linear combinations with suitable l_i:

$$\boldsymbol{x}^{(1)}(t) = \mathfrak{Re}\big[e^{\lambda_1 t}\boldsymbol{v}^{(1)}\big] = \frac{1}{2}\big(e^{\lambda_1 t}\boldsymbol{v}^{(1)} + \overline{e^{\lambda_1 t}\boldsymbol{v}^{(1)}}\big) = \frac{1}{2}e^{\lambda_1 t}\boldsymbol{v}^{(1)} + \frac{1}{2}e^{\lambda_2 t}\boldsymbol{v}^{(2)} = \mathfrak{Re}\big[e^{\lambda_2 t}\boldsymbol{v}^{(2)}\big],$$
$$\boldsymbol{x}^{(2)}(t) = \mathfrak{Im}\big[e^{\lambda_1 t}\boldsymbol{v}^{(1)}\big] = \frac{1}{2\mathbf{i}}\big(e^{\lambda_1 t}\boldsymbol{v}^{(1)} - \overline{e^{\lambda_1 t}\boldsymbol{v}^{(1)}}\big) = \frac{1}{2\mathbf{i}}e^{\lambda_1 t}\boldsymbol{v}^{(1)} - \frac{1}{2\mathbf{i}}e^{\lambda_2 t}\boldsymbol{v}^{(2)} = -\mathfrak{Im}\big[e^{\lambda_2 t}\boldsymbol{v}^{(2)}\big].$$

To identify and further work out real and imaginary parts let us write

$$\lambda_1 = p + \mathbf{i}\omega \text{ with } p, \omega \in \mathbb{R}, \quad \text{and } \boldsymbol{v}^{(1)} = \boldsymbol{r} + \mathbf{i}\boldsymbol{b} \text{ with } \boldsymbol{r}, \boldsymbol{b} \in \mathbb{R}^2.$$

Note that $\omega = \frac{1}{2}\sqrt{4\det(A) - (\operatorname{trace}(A))^2} \neq 0$, and one can show that $\boldsymbol{r} \neq (0,0)$ and $\boldsymbol{b} \neq (0,0)$. Then

$$e^{\lambda_1 t} = e^{pt}(\cos(\omega t) + \mathbf{i}\sin(\omega t)),$$

and consequently

$$\begin{aligned} e^{\lambda_1 t}\boldsymbol{v}^{(1)} &= e^{pt}\big(\cos(\omega t) + \mathbf{i}\sin(\omega t)\big)(\boldsymbol{r} + \mathbf{i}\boldsymbol{b}) \\ &= e^{pt}\big(\cos(\omega t)\boldsymbol{r} - \sin(\omega t)\boldsymbol{b}\big) + \mathbf{i}e^{pt}\big(\sin(\omega t)\boldsymbol{r} + \cos(\omega t)\boldsymbol{b}\big). \end{aligned}$$

Two linearly independent real-valued solutions therefore are

$$\boldsymbol{x}^{(1)}(t) = e^{pt}\big(\cos(\omega t)\boldsymbol{r} - \sin(\omega t)\boldsymbol{b}\big), \quad \boldsymbol{x}^{(2)}(t) = e^{pt}\big(\sin(\omega t)\boldsymbol{r} + \cos(\omega t)\boldsymbol{b}\big).$$

We obtain a general (real-valued) solution of the form

$$\begin{aligned} \boldsymbol{x}(t) &= l_1\boldsymbol{x}^{(1)}(t) + l_2\boldsymbol{x}^{(2)}(t) \\ &= l_1 e^{pt}\big(\cos(\omega t)\boldsymbol{r} - \sin(\omega t)\boldsymbol{b}\big) + l_2 e^{pt}\big(\sin(\omega t)\boldsymbol{r} + \cos(\omega t)\boldsymbol{b}\big), \quad l_1, l_2 \in \mathbb{R}. \end{aligned} \tag{4.12}$$

Any initial condition of the form (4.4) with $t_0 = 0$ (this is for simplicity, otherwise, we 'shift in time' by replacing t with $t - t_0$ in the general solution) then becomes

$$\boldsymbol{x}_0 = \boldsymbol{x}(0) = l_1\boldsymbol{r} + l_2\boldsymbol{b}.$$

If $p = \mathfrak{Re}(\lambda_1) = \mathfrak{Re}(\lambda_2) = \operatorname{trace}(A)/2 < 0$ then e^{pt} decreases to zero as t increases, and if $p = 0$ then $e^{pt} = 1$ is bounded. We see that if $p \leq 0$ then also $\boldsymbol{x}(t)$ is bounded or even approaches $(0,0)$, hence, the stationary point $(0,0)$ is stable then. In turn, if $p > 0$ then e^{pt} increases as t increases, and solutions move away from $(0,0)$. The stationary point then is unstable.

Example 4.9. *Consider* (4.10) *with an initial condition of the form* (4.4) *where* $t_0 = 0$ *and*

$$A = \begin{pmatrix} 1 & 5 \\ -1 & -3 \end{pmatrix}, \quad \boldsymbol{x}_0 = \begin{pmatrix} 20 \\ 0 \end{pmatrix}.$$

As $\operatorname{trace}(A) = -2$ *and* $\det(A) = 2$ *the eigenvalues of* A *satisfy* $\lambda^2 + 2\lambda + 2 = 0$ *and are*

$$\lambda_{1,2} = -\frac{2}{2} \pm \frac{1}{2}\sqrt{2^2 - 4 * 2} = -1 \pm \mathbf{i}, \quad \textit{so } p = -1,\ \omega = 1.$$

An eigenvector corresponding to λ_1 *can be worked out to be*

$$\boldsymbol{v}^{(1)} = \begin{pmatrix} 5 \\ -2 + \mathbf{i} \end{pmatrix}, \quad \textit{so } \boldsymbol{r} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

A general solution to the differential equations is thus given by

$$\begin{aligned}\boldsymbol{x}(t) &= l_1 e^{pt}\big(\cos(\omega t)\boldsymbol{r} - \sin(\omega t)\boldsymbol{b}\big) + l_2 e^{pt}\big(\sin(\omega t)\boldsymbol{r} + \cos(\omega t)\boldsymbol{b}\big)\\ &= l_1 e^{-t}\cos(t)\begin{pmatrix}5\\-2\end{pmatrix} - l_1 e^{-t}\sin(t)\begin{pmatrix}0\\1\end{pmatrix} + l_2 e^{-t}\sin(t)\begin{pmatrix}5\\-2\end{pmatrix} + l_2 e^{-t}\cos(t)\begin{pmatrix}0\\1\end{pmatrix}\\ &= l_1 e^{-t}\begin{pmatrix}5\cos(t)\\-2\cos(t)-\sin(t)\end{pmatrix} + l_2 e^{-t}\begin{pmatrix}5\sin(t)\\-2\sin(t)+\cos(t)\end{pmatrix}.\end{aligned}$$

The initial condition is satisfied if

$$\boldsymbol{x}_0 = l_1\boldsymbol{r} + l_2\boldsymbol{b} \quad \Leftrightarrow \quad \begin{pmatrix}20\\0\end{pmatrix} = l_1\begin{pmatrix}5\\-2\end{pmatrix} + l_2\begin{pmatrix}0\\1\end{pmatrix}.$$

Hence $l_1 = 4$ and $l_2 = 8$, and the unique solution of (4.10), (4.4) with the above data is

$$\boldsymbol{x}(t) = e^{-t}\begin{pmatrix}20\cos(t)\\-8\cos(t)-4\sin(t)\end{pmatrix} + e^{-t}\begin{pmatrix}40\sin(t)\\-16\sin(t)+8\cos(t)\end{pmatrix} = e^{-t}\begin{pmatrix}20\cos(t)+40\sin(t)\\-20\sin(t)\end{pmatrix}.$$

As $e^{-t} \to 0$ as $t \to \infty$ the solution converges to $(0,0)$ as $t \to \infty$. In particular, it remains close to the stationary solution. This is actually true for any initial point, so the stationary point is stable. Figure 4.4 (right) give an impression of the direction field and the behaviour of solutions starting at different points. In particular, the convergence to $(0,0)$ is visible. Notice the difference to the other example (left figure): the solutions 'spiral' into the origin (right) rather than 'hitting' it directly (left).

4.3.4 Other cases and summary

It can happen that the two eigenvalues of A coincide, namely if $(a_{1,1} - a_{2,2})^2 + 4a_{2,1}a_{1,2} = 0$. We will not discuss this 'critical' case here. For a complete characterisation of solutions to homogeneous 2×2 systems with constants coefficients including sketches of phase diagrams wee Chap 31 of [4].

Bibliography

[1] Grigory I Barenblatt, *Scaling, self-similarity, and intermediate asymptotics: dimensional analysis and intermediate asymptotics*. No. 14, Cambridge University Press, 1996.

[2] Mark H Holmes, *Introduction to the foundations of applied mathematics*. Springer, Texts in Applied Mathematics 56, 2019.

[3] David Logan, *A first course in differential equations*. Springer, 2017.

[4] James C Robinson, *An introduction to ordinary differential equations*. Cambridge University Press, 2004.

[5] Gerald Teschl, *Ordinary differential equations and dynamical systems*. Vol. 140. American Mathematical Soc., 2012.

[6] Thomas Witelski and Mark Bowen, *Methods of Mathematical Modelling: Continuous Systems and Differential Equations*. Springer Undergraduate Mathematics Series, 2015.