Introduction to Differential Equations

(For smart kids)

Andrew D. Lewis

This version: 2017/07/17

Preface

This book is intended to be suggest a revision of the way in which the first course in differential equations is delivered to students, normally in their second year of university. This course has traditionally grown as an offshoot of the calculus courses taught in the first year, where students often learn some techniques and tricks for solving specific problems, e.g., for computing specific derivatives and integrals. This is not an entirely unreasonable thing to do, since it is difficult to imagine being able to practice mathematics without being able to handle calculus with some of the basic special functions one encounters in a first course.[1] Moreover, a first calculus course often comes complete with many insights into the meaning of, and uses of, differentiation and integration. However, this "techniques and tricks" method becomes less valuable for ordinary differential equations. The fact is that there are very few differential equations that can be solved, and those that can be solved only succumb after quite a lot of work.[2] Thus, while I do believe it is essential to be able to solve a number of differential equations "by inspection"—and I expect students taking the course for which this is the text to be able to do this—the proliferation of computer packages to carry out efficiently and effectively the tedious computations typically learnt in a differential equations course makes one reconsider why we teach students multiple ways to solve the same small set of differential equations. This text is the result of my own reconsideration of the traditional first course in differential equations.

As an instructor, the question becomes, "If I do not teach all of the usual techniques and tricks for solving differential equations, what do I replace it with?" My choices for answers to this question are the following.

1. *Make sure students know that differential equations arise naturally in a wide variety of fields, including the sciences, engineering, and the social sciences.* This is done by starting the text with a long list of examples of models involving differential equations. For some of these, we are able to provide a pretty complete rationale for where the equations come from, and in some cases we can make predictions based on our human experience about how solutions of these differential equations may behave. However, in some cases we are merely able to describe in plain English what the equations represent, and then just write them down. But,

[1] That being said, the following conversation happens in most courses I teach:

Me: Oh no, an integral! Class, how do we solve integrals?
Class: Google! Wolfram Alpha!
Me: Correct!

[2] And so we have. . .

Me: Oh no, a differential equation! Class, how do we solve differential equations?
Class: Google! Wolfram Alpha!
Me: Correct!

at the end of the day, we produce a long list of differential equations of various kinds. Thus we do not restrict ourselves to modelling involving simple differential equations, but also provide some impossibly complicated differential equations so that the subject is not oversimplified.

And this leads to the following choice.

2. *Make sure students know what a differential equation* is. A differential equation is normally written as just that: an equation. The problem with this is that equations are not really mathematically useful. When one writes down an equation with an unknown, this is something to be solved, not something to be understood. Thus we demur from just writing differential equations, and define them initially as maps whose properties can be enumerated and understood. In treating differential equations in this way, it is seen that there is a common starting point for *all* differential equations, and that the ones that we learn how to solve are very special and degenerate in some way.
3. *Appreciate how to use a computer when working with differential equations.* Because there are so few differential equations that can be solved analytically, and also because the analytical solution procedures are often extremely tedious to apply, one may wish to have at hand computer methods for working with differential equations. Computer packages come in two basic flavours, which give, along with some examples of these.
   I. *Computer algebra systems:* A computer algebra system can typically find analytic solutions to differential equations, when these can be easily found. For example, any decent computer algebra system can solve any differential equation we solve using the methods in this book. Some examples of commonly-used computer algebra systems are:
      (a) Maple®: http://www.maplesoft.com/
      (b) Mathematica®: http://www.wolfram.com/mathematica/
      (c) Maxima®: http://maxima.sourceforge.net/
      (d) SageMath®: http://www.sagemath.org/
   II. *Numerical computation packages:* Even if one cannot use a computer algebra system to obtain analytic solutions to differential equations, one can often use algorithms that approximate differential equations and produce numerical solutions. This is very often the only thing one is interested in in hardcore applications of differential equations, even in cases where analytical solutions are possible. Some examples of commonly used numerical computation packages are:
      (a) Maple®: http://www.maplesoft.com/
      (b) Mathematica®: http://www.wolfram.com/mathematica/
      (c) Matlab®: http://www.mathworks.com/
      (d) Octave®: https://www.gnu.org/software/octave/
      (e) Scilab®: http://www.scilab.org/

As can be seen, it is often (but not always) the case that a computer algebra system offers the facility to do numerical computations with differential equations, along with that for doing symbolic computations.

The above list is by no means an exhaustive accounting of what is available, and for a more complete (but still not complete) list, please visit the appropriate WIKIPEDIA® pages:

https://en.wikipedia.org/wiki/List_of_computer_algebra_systems
https://en.wikipedia.org/wiki/List_of_numerical_analysis_software

We wish to emphasise that we do not go deeply *at all* into numerical analysis in this text. Indeed, we use computer packages as a tool, and one must be aware—as when using any tool—of its limitations. However, from a pragmatic point of view, computer packages in the present day are so sophisticated that one typically must go very deeply to see why they might break, and to do so is well beyond our present scope.

4. *Understand the character of solutions, rather than just producing their closed-form expressions.* While there is something gratifying in being able to go through a long involved process, and arrive at a correct solution to a differential equation, it is far more interesting and useful to be able to understand (a) why the solution process works and (b) what is the character of the solution one obtained. Thus, while we do consider some of the standard methods for solving differential equations, we do not either start or stop there.
5. *Introduce transform methods for differential equations, since these are very powerful.* However, we do not wish to introduce transform methods as providing an algorithmic procedure for solving (in practice, only very simple) differential equations. What we wish to do is illustrate, in as general a way as possible in an introductory text, the *raison d'être* for transform methods, which is that they turn differential equations into algebraic equations, maybe only partially so. Thus we introduce a variety of transforms used in a variety of problems.

This is Version 1 of these notes, so please indicate errors or suggestions for improvements.

Andrew D. Lewis *Kingston, Ontario, Canada*

Table of Contents

Chapter 1

What are differential equations?

In this chapter we provide what we hope is a substantial backdrop and motivation for the study of differential equations. We do this first, in Section 1.1, by considering an array of physical systems that are modelled by differential equations. While we provide a diverse collection of such motivating examples, the fact of the matter is that this is a pitifully small sampling of the ways in which differential equations arise in modelling. Nonetheless, we hope that we can justify a broad and vague assertion like, "Differential equations are endemic in mathematical models of the physical world."

As we shall see in Section 1.3, differential equations come in various flavours. The two main branches in the classification tree for differential equations are "ordinary differential equations" and "partial differential equations." In these notes we will primarily consider short twigs coming off each of these two branches, corresponding to "linear" differential equations. However, it is extremely important to realise that, while linear differential equations are of fundamental importance in the general theory, many, many differential equations encountered in practice are *not* linear. Thus a good understanding of just what a linear differential equation *is* is essential.

In the final section of this chapter, we consider two important questions: (1) do differential equations possess solutions? (2) if a differential equation possesses a solution, is it unique? We shall see that the answer to both questions is generally, "No," but that under very weak hypotheses this answer is, "Yes," at least for ordinary differential equations. (These questions for partial differential equations are far more sinister, at least if one wishes to pursue any degree of generality.)

We also provide, in Section 1.2, an overview of the mathematical background we will use. Most of this will have been acquired by a typical second-year student, although perhaps not with the degree of notational precision we require. There is also likely to be a few topics that are not a part of the standard background of a second-year student, and so some parts of this section may be new.

Contents

Section 1.1

How do differential equations arise in mathematical modelling?

It is possible to approach the subject of differential equations from a purely mathematical point of view. And, indeed, even if one is interested in only applying the theory of differential equations in specific areas, a good knowledge of this mathematical subject is necessary. However, a primary reason for the importance of differential equations in mathematics is that they arise so naturally and broadly in areas of application, ranging from engineering, physics, economics, and biology, to name a few. Indeed, it may not be inaccurate to say that differential equations provide the most important (but definitely not the only) conduit from developments in mathematics to applications. In this section, we illustrate this with an array of examples.

Caveat We mainly shall not be precise in this section with things like whether functions are continuous, differentiable, etc.. In the remainder of the text we shall be more careful about these things. •

1.1.1 Mass-spring-damper systems

Let us start by considering a single mass connected to the ground by a spring and a damper, as in Figure 1.1. The mass has mass m, the spring is a linear spring with

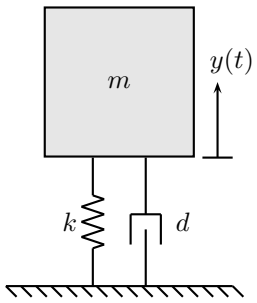


Figure 1.1 A simplified model of a car suspension

a restoring force proportional to the change in length from its equilibrium—i.e., the spring force is $-k\Delta$, $k \geq 0$, where Δ is the change in length— and the damper is also linear with a restoring force proportional of the velocity at which the damper is contracted–i.e., the damper force is $-d\dot{\Delta}$, $d \geq 0$, where "$\dot{}$" means "derivative with respect to time. This may be thought of as a simple model for a car suspension.

We shall derive an equation that governs the vertical motion of the mass as a function of time. We let $y(t)$ be the vertical displacement of the mass, with the assumption that $y = 0$ corresponds to the undeflected position of the spring. We

suppose that we have a gravitational force acting "downwards" in the diagram and with a gravitational constant a_g. One then performs a force balance, setting vertical forces equal to the mass times the acceleration:

$$-d\dot{y}(t) - ky(t) - ma_g = m\ddot{y}(t) \iff m\ddot{y}(t) + d\dot{y}(t) + ky(t) = -ma_g. \tag{1.1}$$

Note that this is an equation with single independent variable t (time) and single dependent variable y (vertical displacement). Moreover, the equation us *not* an algebraic equation for y as a function of t, since derivatives of y with respect to t arise.

During the course of these notes, we shall learn how to exactly solve a differential equation like this. But before we do so, let us see if we can, based on our common sense, deduce what sort of behaviour a system like this should exhibit. First let's determine the equilibrium of the system, since it is *not* when $y = 0$, because of the gravitational force. Indeed, as equilibrium the mass should not be in motion and so we ought to have $\dot{y} = 0$ and $\ddot{y} = 0$. In this case, $y = -\frac{ma_g}{k}$. Now let's think about what happens when $d = 0$. What we expect here is that the mass will oscillate in the vertical direction around the equilibrium. Moreover, we may expect that as k becomes relatively larger, the frequency of oscillations will increase. Now, adding the damping constant $d > 0$, perhaps our intuition is not quite so reliable a means of deducing what is going on here. But what happens is this: the damper dissipates energy. This causes the oscillations to decay to zero as $t \to \infty$. Moreover, if d gets relatively large, it actually happens that the oscillations do not occur, and the mass just moves towards its equilibrium. These are things we will investigate systematically.

Next let us complicate matters a little, and consider two interconnected masses as in Figure 1.2. In this case, to simplify things we interconnect the masses only

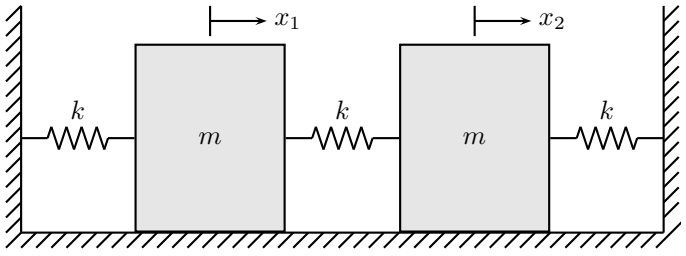


Figure 1.2 Interconnected masses

with springs. As in the figure, we let x_1 and x_2 denote the positions of the masses, assuming that all springs are uncompressed with $x_1 = x_2 = 0$. In this case, the force balance equations for the two masses give the equations

$$\begin{aligned} -kx_1(t) - k(x_1(t) - x_2(t)) = m\ddot{x}_1(t), &\iff m\ddot{x}_1(t) + 2kx_1(t) - x_2(t) = 0, \\ -kx_2(t) - k(x_2(t) - x_1(t)) = m\ddot{x}_2(t), &\iff m\ddot{x}_2(t) + 2kx_2(t) - x_1(t) = 0. \end{aligned}$$

Let us express this using matrix/vector notation:

$$m\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\ddot{x}_1(t)\\ \ddot{x}_2(t)\end{bmatrix} + k\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}.$$

If we introduce the notation

$$\boldsymbol{M} = m\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad \boldsymbol{K} = k\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}, \quad \boldsymbol{x}(t) = \begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix},$$

then we can further write this as

$$\boldsymbol{M}\ddot{\boldsymbol{x}}(t) + \boldsymbol{K}\boldsymbol{x}(t) = \boldsymbol{0}. \tag{1.2}$$

Note that this is an equation with single independent variable t (time) and two dependent variables x_1 and x_2, or equivalently a vector dependent variable $(x_1, x_2) \in \mathbb{R}^2$ (horizontal displacements). As was the case with the single mass, the key point is that the equation involves derivatives of the dependent variables with respect to the independent variable.

In the text, we will see how to analyse such equations as this. Let us say a few words about the most interesting features of how this system behaves. There are two interesting classes of behaviours, one occurring when $x_1(t) = x_2(t)$ (the masses move together) and one occurring when $x_1(t) = -x_2(t)$ (the masses move exactly opposite one another). These "modes" of the system are important, as we shall see that every solution is a linear combination of these two. This has to do with fundamental properties of systems of this general type.

1.1.2 The motion of a simple pendulum

Let us consider the motion of a pendulum as depicted in Figure 1.3. We suppose

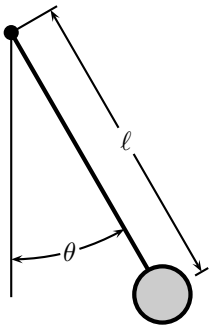


Figure 1.3 A simple pendulum

that we have a mass m attached to a rod of length ℓ whose mass we consider to be negligible compared to m. We have a gravitational force with gravitational

constant a_g that acts downward in the figure. Summing moments about the pivot point gives

$$-ma_g\ell \sin\theta(t) = m\ell^2\ddot{\theta}(t) \iff \ddot{\theta}(t) + \frac{a_g}{\ell}\sin\theta(t) = 0. \tag{1.3}$$

This is an equation in a single independent variable t (time) and a single dependent variable θ (pendulum angle), and again is an equation in derivatives of the dependent variable with respect to the independent variable.

We shall *not* learn how to solve this equation in this text, although a "closed-form solution" is possible with a suitably flexible notion of "closed-form." However, problems such as this one call into question the value of having a closed-form solution. What is, perhaps, a more useful way to understand the behaviour of a simple pendulum is to try some sort of approximation. We shall make an approximation near the two equilibria of the pendulum, corresponding to $\theta = 0$ (the "down" equilibrium) and $\theta = \pi$ (the "up" equilibrium). To make the approximation, we note that, for ϕ near zero,

$$\begin{aligned}&\sin\phi \approx \phi,\\ &\sin(\pi+\phi) = \sin\pi\cos\phi + \cos\pi\sin\phi \approx -\phi.\end{aligned}$$

Therefore, the equation governing the behaviour of the simple pendulum are approximated near $\theta = 0$ (say $\theta = 0 + \phi$) by

$$\ddot{\phi}(t) + \frac{a_g}{\ell}\phi(t) = 0.$$

We shall see during the course of our studies that a general solution to these equations takes the form

$$\phi(t) = \phi(0)\cos(\omega\phi(t)) + \frac{\dot{\phi}(0)}{\omega}\sin(\omega\phi(t)),$$

where $\omega = \sqrt{a_g/\ell}$. Thus, if the approximation is valid, this suggests that the motion of the simple pendulum, for small angles, consists of periodic motions with frequency ω. It turns out that this behaviour is indeed close to that of the genuine pendulum equations. To be precise, the motion of the pendulum for small angles is indeed periodic, and as the angle gets smaller, the frequency approaches ω. However, the motion is *not* sinusoidal. Moreover, the period gets larger for larger amplitude motions.

A very large amplitude motion would be when θ starts at π. If we take $\theta = \pi+\phi$ then the governing equation is approximately

$$\ddot{\phi}(t) - \frac{a_g}{\ell}\phi(t) = 0.$$

We shall see that a general solution to these equations takes the form

$$\phi(t) = \phi(0)\cosh(\omega\phi(t)) + \frac{\dot{\phi}(0)}{\omega}\sinh(\omega\phi(t)), \tag{1.4}$$

where $\omega = \sqrt{a_g/\ell}$. (Here cosh and sinh are the hyperbolic cosine and sine functions, defined by

$$\cosh(x) = \tfrac{1}{2}(e^x + e^{-x}), \quad \sinh(x) = \tfrac{1}{2}(e^x - e^{-x}).)$$

For most values of $\dot{\phi}(0)$ and $\dot{\phi}(0)$, the solutions of this equation diverge to ∞ as $t \to \infty$. Of course, as ϕ gets large, this approximation becomes unreliable. Nonetheless, the behaviour observed for small times agrees with what we think the dynamics ought to be: since the "up" equilibrium is unstable, trajectories generally move away from this equilibrium. Note, however, that there are a small number of the solutions (1.4) that do not diverge to ∞, but approach $\phi = 0$ as $t \to \infty$, namely those for which $\phi(0) = -\frac{\dot{\phi}(0)}{\omega}$. In terms of the physics of the pendulum, these solutions correspond to the motions of the pendulum where the pendulum swings with just enough energy to approach the upright equilibrium as $t \to \infty$.

1.1.3 Bessel's equation

We shall not motivate here precisely how the equation we consider in this section arises in practice. We shall be content with the following description: If one tries to solve the potential equation (1.19) in two-dimensions and in polar coordinates, then one arrives at the equation

$$r^2\frac{\partial^2 y}{\partial r^2} + r\frac{\partial y}{\partial r} + (r^2 - \alpha^2)y = 0, \tag{1.5}$$

for $\alpha \in \mathbb{R}$ (actually, in the particular case of the potential equation, α is a nonnegative integer). This equation, for example, describes the radial displacement in a drum when it has been struck. The equation is known as ***Bessel's equation***.

We note that Bessel's equation has one independent variable r, one dependent variable y, and is an equation in the derivatives of the dependent variable with respect to the independent variable.

1.1.4 RLC circuits

Next let us consider differential equations such as arise in circuits comprised of ideal resistors, inductors, and capacitors. Let us define these terms. We will use "E," "I," and "q" to denote voltage, current, and charge, respectively.

1. A ***resistor*** is a device across which the voltage drop is proportional to the current through the device. The constant of proportionality is the ***resistance*** R: $E = RI$.
2. An ***inductor*** is a device across which the voltage drop is proportional to the time rate of change of current through the device. The constant of proportionality is the ***inductance*** L: $E = L\frac{\mathrm{d}I}{\mathrm{d}t}$.

3. A ***capacitor*** is a device across which the voltage drop is proportional to the charge in the device. The constant of proportionality is the $\frac{1}{C}$ with C being the ***capacitance***: $E = \frac{1}{C}q$.

The three devices are typically given the symbols as in Figure 1.4. The physical

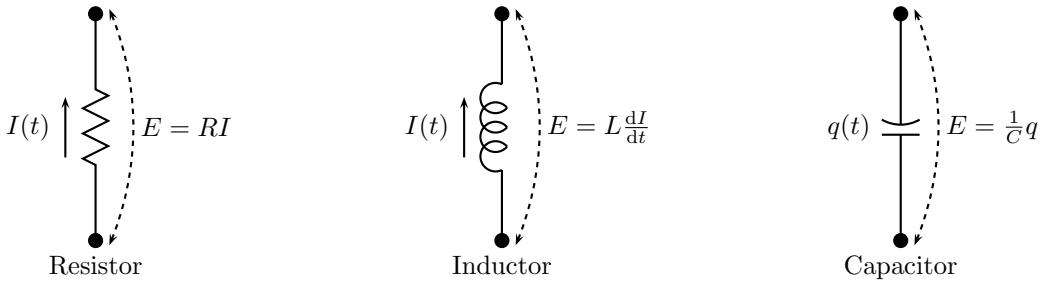


Figure 1.4 Electrical devices

laws governing the behaviour of ideal circuits are:

1. the current I is related to the charge q by $I = \frac{dq}{dt}$;
2. ***Kirchhoff's voltage law*** states that the sum of voltage drops around a closed loop must be zero;
3. ***Kirchhoff's current law*** states that the sum of the currents entering a node must be zero.

Given a collection of such devices arranged in some way—i.e., a "circuit"—along with voltage and/or current sources, we can imagine that governing equations for the behaviour of the circuit can be deduced. In Figure 1.5 we have a particularly

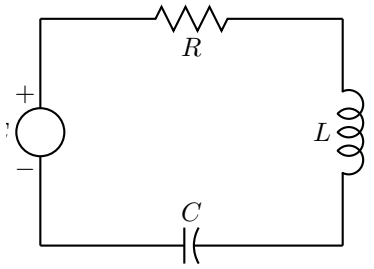


Figure 1.5 A series *RLC* circuit

simple configuration. The voltage drop around the circuit must be zero which gives the governing equations

$$E(t) = RI(t) + L\dot{I}(t) + \tfrac{1}{C}q(t) \qquad \Longrightarrow \qquad L\ddot{q}(t) + R\dot{q}(t) + \tfrac{1}{C}q(t) = E(t)$$

where $E(t)$ is an external voltage source. This may also be written as a current equation by merely differentiating:

$$L\ddot{I}(t) + R\dot{I}(t) + \tfrac{1}{C}I(t) = \dot{E}(t). \tag{1.6}$$

In either case, we have an equation in a single independent variable (time) and a single dependent variable (charge or current). The equations involve, of course, derivatives of the dependent variable with respect to the dependent variable.

We comment here on similarity with the equation (1.6) with the equation (1.1) describing the motion of a damped mass/spring system are worth remarking upon. The capacitor plays the rôle of a spring (stores energy), the resistor plays the rôle of a damper (dissipates energy), and the inductor plays the rôle of a mass (it energy is obtained from "motion" in the circuit). This gives rise to an important "electro-mechanical analogy" in the modelling of physical systems.

1.1.5 Tank systems

Here we consider two tanks with fluid in a configuration shown in Figure 1.6. Here are the variables and parameters:

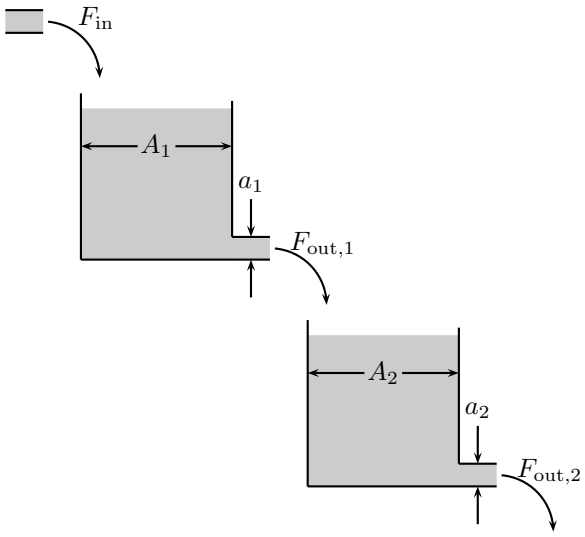


Figure 1.6 Mass balance in coupled tank flow

F_{in}	volume flow into tank 1
$F_{\text{out,j}}$	volume flow out of tank j, $j \in \{1, 2\}$
A_j	cross-sectional area of tank j, $j \in \{1, 2\}$
a_j	cross-sectional area of orifice j, $j \in \{1, 2\}$
h_j	height of water in tank j, $j \in \{1, 2\}$

Let us state the rules we shall use to deduce the behaviour of the system, assuming that the fluid is "incompressible" so the mass of a given volume of fluid will be constant:

1. according to ***Bernoulli's Law***, the velocity of the fluid exiting a small orifice at the bottom of a tank with level h is $\sqrt{2a_g h}$, where a_g is the acceleration due to gravity;
2. the volume of rate of fluid flow passing through an orifice with constant cross-sectional area A with velocity v (assumed to be constant across the cross-section) is Av;
3. the rate of change of volume in a tank with constant cross-sectional area A and fluid height h is $A\frac{\mathrm{d}h}{\mathrm{d}t}$.

We can thus form the balance equations for each tank by setting the rate of change of volume in the tank equal to the volume flow in minus the volume flow out:

$$\begin{aligned} A_1\dot{h}_1(t) &= F_{\text{in}}(t) - F_{\text{out},1} = F_{\text{in}}(t) - \sqrt{2a_1h_1(t)}, \\ A_2\dot{h}_2(t) &= F_{\text{out},1}(t) - F_{\text{out},2} = \sqrt{2a_1h_1(t)} - \sqrt{2a_2h_2(t)}. \end{aligned} \tag{1.7}$$

The equations governing the behaviour of the system have one independent variable t (time) and two dependent variables h_1 and h_2, or a single vector variable $(h_1, h_2) \in \mathbb{R}^2$ (the heights of fluid in the tanks). As with all of our examples, the equations involve the derivatives of the dependent variables with respect to the independent variable.

1.1.6 Population models

An important area of application of differential equations is in biological sciences, in areas such as epidemiology and population dynamics. We shall consider here two simple models of population dynamics as an illustration.

First let us consider a population that we model as a scalar variable $p \in \mathbb{R}$. First we consider a situation where the rate of population growth is proportional to p for small values of p, but then diminishes as we approach some "limiting population size, p_0, representing the fact that there may be limited resources. This can be represented by a model like

$$\dot{p}(t) = kp(t)\left(1 - \frac{p(t)}{p_0}\right). \tag{1.8}$$

This is often referred to as the ***logistical model*** of population dynamics. This is an equation with a single independent variable t (time) and a single dependent variable p (population).

While we will not explicitly examine this equation in this text, the reader may relatively easily verify the following behaviour, under the natural assumption that $k > 0$.

1. There is an equilibrium at $p = 0$ that is not stable. That is, for small positive populations, the rate of population change is positive.
2. There is an equilibrium at $p = p_0$ that is stable. That is, for populations less than the limiting population p_0, the rate of population change is positive.

Let us now consider two populations a and b, with a representing the population of a prey species and b representing the population of a predator species. The following assumptions are made:

1. prey population increases exponentially in the absence of predation;
2. predators die off exponentially in the absence of predation;
3. predator growth and prey death due to predation is proportional to the rate of predation;
4. the rate of predation is proportional to the encounters between predators and prey, and encounters themselves are proportional to the populations.

Putting all of this together, the behaviour of the prey population a can be modelled by

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

and the behaviour of the predator population can be modelled by

$$\dot{b}(t) = \delta a(t)b(t) - \gamma b(t).$$

We should combine these equations:

$$\begin{aligned} \dot{a}(t) &= \alpha a(t) - \beta a(t)b(t), \\ \dot{b}(t) &= \delta a(t)b(t) - \gamma b(t). \end{aligned} \tag{1.9}$$

These equations have a single independent variable t (time) and two dependent variables a and b, or equivalently a single vector variable $(a, b) \in \mathbb{R}^2$. This model is called the ***Lotka–Volterra predator-prey model***.

We shall not in this text undertake a detailed analysis of this equation. However, a motivated reader can easily find many sources where this model is discussed in great depth and detail.

1.1.7 Economics models

Another area where differential equations are useful is in social sciences, and especially economics. We consider an example of this, known as the ***Rapoport production and exchange model***.

The setup is this. Individuals A and B produce goods that we measure by scalar variables $a, b \in \mathbb{R}$. The individuals A and B trade, each trying to maximise their "happiness," typically referred to as "utility."[1] We denote by p the proportion of

[1] In philosophy, the notion of "utility" as a measure of general happiness dates, in its most explicit form, to Thomas Hobbes (1588–1679) and John Locke (1632–1704). While early versions of utilitarianism were based in religion, John Stuart Mill (1806–1873) developed a powerful secular utilitarian ethic, which itself led to the secular philosophy of Jeremy Bentham (1748–1832).

goods produced and retained, and by q the proportion of goods produced and traded: thus $p + q = 1$. The assumptions made by Rapoport are these:

1. people are lazy, so the act of production is a loss of utility;
2. people are gauche, so possessing something produced is a gain in utility;
3. the loss of utility due to the agonies of production are proportional to the amount produced;
4. while there is no cap in a person's desire to acquire crap, the utility they gain from acquiring crap diminishes, the more crap they have;
5. the rate at which A or B makes product a and b is proportional to the rate at which utility increases with respect to a and b.

With all this as backdrop, let us introduce something meaningful. First of all, let us give the utility functions for A and B:

$$U_A(a, b) = \log(1 + pa + qb) - r_A a, \quad U_B(a, b) = \log(1 + qa + pb) - r_B b.$$

If one examines these expressions, one can see that they capture in form and shape the characteristics of individuals A and B described above. Of course, many other forms are also viable candidates.

Now, according to condition 5, the equations that govern the amounts a and b are:

$$\begin{aligned} \dot{a}(t) &= c_A \left(\frac{p}{1 + pa(t) + qb(t)} - r_A a(t) \right), \\ \dot{b}(t) &= c_B \left(\frac{p}{1 + pa(t) + qb(t)} - r_B b(t) \right). \end{aligned} \tag{1.10}$$

These equations have a single independent variable t (time), and two dependent variables a and b, or equivalently one vector variable $(a, b) \in \mathbb{R}^2$ (production). The equation is one that involves the derivatives of the dependent variables with respect to the independent variable.

An indepth analysis of these equations is not something we will undertake here.

1.1.8 Euler–Lagrange equations

We consider here the following problem. Suppose we are given $y_1, y_2 \in \mathbb{R}$ and $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$. Denote by

$$\begin{aligned} &\Gamma(y_1, y_2, x_1, x_2) \\ &= \{\gamma \colon [x_1, x_2] \to \mathbb{R} \mid \ \gamma \text{ is twice continuously differentiable } \gamma(x_1) = y_1,\ \gamma(x_2) = y_2\} \end{aligned}$$

the set of all twice continuously differentiable functions with value y_1 at the left endpoint and y_2 at the right endpoint, as in Figure 1.7. Suppose that we have a function $L \colon [x_1, x_2] \times \mathbb{R}^2 \to \mathbb{R}$ that we call the ***Lagrangian***. Associated to this Lagrangian and a function $\gamma \in \Gamma(y_1, y_2, x_1, x_2)$ we have an associated ***cost***

$$C_L(\gamma) = \int_{x_1}^{x_2} L(x, \gamma(x), \gamma'(x)) \, \mathrm{d}x.$$

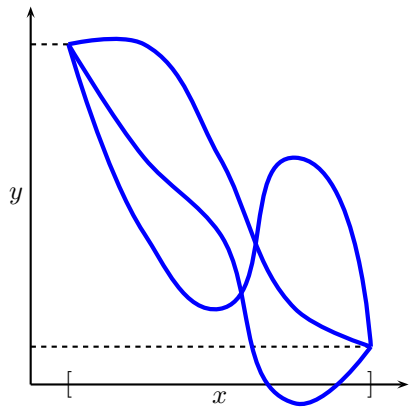


Figure 1.7 Candidate curves in an optimisation problem

The objective is to find γ that minimises $C_L(\gamma)$. That is, we seek $\gamma_* \in \Gamma(y_1, y_2, x_1, x_2)$ such that

$$C_L(\gamma_*) \leq C_L(\gamma), \qquad \gamma \in \Gamma(y_1, y_2, x_1, x_2).$$

Such a function γ_* is a ***minimiser*** for the Lagrangian L. One can show, without much difficulty, but using methods from the calculus of variations that are a little far afield for us at the moment, that if γ_* is given by $\gamma_*(x) = y(x)$ is a minimiser for L, then it necessarily satisfies the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial y'}\right) - \frac{\partial L}{\partial y} = 0,$$

which are the ***Euler–Lagrange equations*** for this problem. We give the equations in their traditional form, although this form is genuinely confusing. Let us be a little more explicit about what the equations mean. By an application of the Chain Rule, the Euler–Lagrange equations can be written as

$$\frac{\partial^2 L}{\partial y' \partial y'} y''(x) + \frac{\partial^2 L}{\partial y' \partial y} y'(x) - \frac{\partial L}{\partial y} = 0. \tag{1.11}$$

Note that this is an equation in the single independent variable x and the single dependent variable y. Again, it is an equation involving derivatives of the dependent variable with respect to the independent variable. However, this equation has, in general, an important difference with some of the other equations we have seen. To illustrate this, let us consider the Lagrangians $L(x, y, y') = y'$. In this case

$$\frac{\partial^2 L}{\partial y' \partial y'} y''(x) + \frac{\partial^2 L}{\partial y' \partial y} y'(x) - \frac{\partial L}{\partial y}$$

is identically zero: a circumstance unlike the equations we have encountered before.

The Euler–Lagrange equations are important equations in physics and optimisation, but to study them in any depth is not something we will be able to undertake in this text.

1.1.9 Maxwell's equations

Maxwell's equations are famously important equations governing the behaviour of electromagnetic phenomenon. Let us introduce the physical variables of Maxwell's equations:

$\boldsymbol{E}$	electric field
$\boldsymbol{B}$	magnetic field
$\boldsymbol{J}$	current density
ρ	charge density

The first three of these quantities are vector fields on the physical space $\mathbb{R}^3$. Thus we should think of each of these physical quantities as defining a direction in $\mathbb{R}^3$ and a length at each point in $\mathbb{R}^3$, i.e., an arrow. The charge density ρ is a scalar-valued function on $\mathbb{R}^3$. Let us say a word or two about how we should interpret these quantities. First of all, the charge density ρ is relatively easy to understand: it prescribes the density of charge provided by subatomic particles per unit volume as we move through physical space. The electric field indicates how charge moves through space; at each point (x_1, x_2, x_3) in space, it moves in the direction of $\boldsymbol{E}(x_1, x_2, x_3)$. Thus $\boldsymbol{E}(x_1, x_2, x_3)$ can be thought of as a "force" acting on a charge at the point (x_1, x_2, x_3). The magnetic field $\boldsymbol{B}$[2] acts for magnetic field lines rather like the electric field acts from the flow of charge: it indicates the direction of magnetic force applied to a moving charge. The current density $\boldsymbol{J}$ gives the current, as a vector quantity, rather in the manner of a fluid flow.

There are also some physical constants in the equations of electromagnetism. These are the following:

ϵ_0	permittivity of free space
μ_0	permeability of free space

These constants are proportionality constants, rather in the manner of the acceleration due to gravity, which we have been denoting by a_g.

With this preparation, we shall produce ***Maxwell's equations*** which indicate

[2]There is another quantity $\boldsymbol{H}$ that also represents the magnetic field, and is proportional to $\boldsymbol{B}$ in a vacuum, but has a more complicated relationship within a magnetic material. Very often $\boldsymbol{H}$ is referred to as the magnetic field, and $\boldsymbol{B}$ is called something different. But often the name "magnetic field" is applied to $\boldsymbol{B}$

how these quantities interact with one another:

$$\begin{aligned} \epsilon_0 \nabla \cdot \boldsymbol{E} &= \rho, \\ \nabla \cdot \boldsymbol{B} &= \mathbf{0}, \\ \nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t}, \\ \nabla \times \boldsymbol{B} &= \mu_0 \boldsymbol{J} + \mu_0 \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}. \end{aligned} \tag{1.12}$$

Let us first describe the mathematical symbols "$\nabla\cdot$" and "$\nabla\times$" that you will learn about in a course on vector calculus. The operator $\nabla\cdot$ is the ***divergence*** and acts on a vector field $\boldsymbol{X} = (X_1, X_2, X_3)$, giving a function according to the definition

$$\nabla \cdot \boldsymbol{X} = \frac{\partial X_1}{\partial x_1} + \frac{\partial X_2}{\partial x_2} + \frac{\partial X_3}{\partial x_3}.$$

The precise meaning of the divergence of a vector field requires a few ways of thinking about things that are not part of ones makeup prior to a course like this, but basically vanishing divergence corresponds to "volume preserving." The operator $\nabla\times$ is ***curl*** and again acts on a vector field $\boldsymbol{X} = (X_1, X_2, X_3)$ giving another vector field according to the definition

$$\nabla \times \boldsymbol{X} = \left(\frac{\partial X_2}{\partial x_3} - \frac{\partial X_3}{\partial x_2}, \frac{\partial X_1}{\partial x_3} - \frac{\partial X_3}{\partial x_1}, \frac{\partial X_2}{\partial x_1} - \frac{\partial X_1}{\partial x_2} \right).$$

As with divergence, a really good understanding of curl of a bit beyond us at this point. Let us say two things: (1) $\nabla \times \boldsymbol{X}$ measures the "rotationality" of a vector field $\boldsymbol{X}$, so its vanishing somehow means it is not rotational; (2) if $\nabla \times \boldsymbol{X} = \mathbf{0}$, then there exists a function f such that $\boldsymbol{X} = \nabla f$, with ∇f being the ***gradient*** of f:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right).$$

What we can now see is that there are four independent variables (x_1, x_2, x_3, t) in Maxwell's equations, representing spacetime, and $3 + 3 + 3 + 1 = 10$ dependent variables $\boldsymbol{E}$, $\boldsymbol{B}$, $\boldsymbol{J}$, and ρ. The equations involve the partial derivatives of the dependent variables with respect to the independent variable.

Now we can say a few words about the meaning of Maxwell's equations. The first equation, called ***Gauss's law for electricity***, says that the "expansiveness" of the electric field is proportional to the charge density. The second equation, called ***Gauss's law for magnetism***, says that the expansiveness of the magnetic field is zero. The third equation, called ***Faraday's law of induction***, tells us that a time-varying magnetic field gives rise to an electric field. Finally, the fourth equation, called ***Ampére's law***, says that both a time-varying electric field and a current density field give rise to magnetic field.

Of course, any systematic investigation of Maxwell's equations is not something we can undertake here, and indeed in complete generality is not possible, by any reasonable meaning of "systematic investigation."

1.1.10 The Navier–Stokes equations

The Navier–Stokes equations deal with the motion of a Newtonian, viscous, and compressible fluid. This means (1) there are viscous, i.e., friction, effects that are accounted for, (2) the viscous stresses arise as a consequence of temporal deformation of the fluid, (3) and the mass of fluid in a given volume is allowed to vary. The motion of the fluid we represent by a mapping $\boldsymbol{\phi}\colon \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, so that $\boldsymbol{\phi}(t, \boldsymbol{x})$ indicates where the fluid particle at $\boldsymbol{x} \in \mathbb{R}^3$ at time 0 resides at time t. We shall abbreviate $\boldsymbol{\phi}_t\colon \mathbb{R}^3 \to \mathbb{R}^3$ the mapping $\boldsymbol{\phi}_x(\boldsymbol{x}) = \boldsymbol{\phi}(t, \boldsymbol{x})$. We shall not deal directly with this mapping $\boldsymbol{\phi}$, but rather with its associated velocity field, by which we mean the mapping $\boldsymbol{u}\colon \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\boldsymbol{u}(t, \boldsymbol{\phi}_t(\boldsymbol{x})) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\phi}_t(\boldsymbol{x}).$$

Thus $\boldsymbol{u}(t, \boldsymbol{x})$ is the velocity of the fluid particle initially at position $\boldsymbol{x}$ at time t. In Figure 1.8 we illustrate how one can think of the velocity field by depicting the

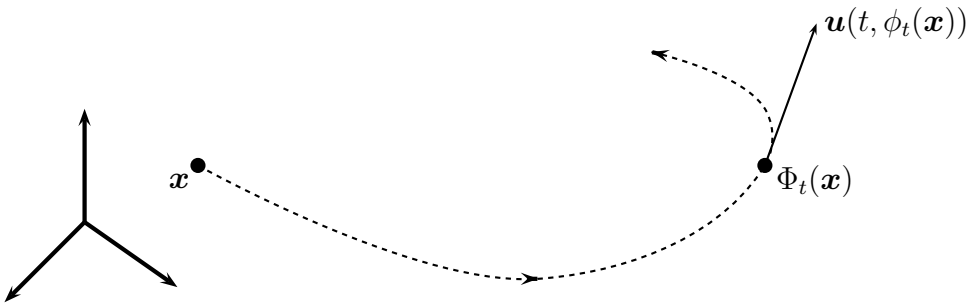


Figure 1.8 The velocity field for a fluid motion

trajectory followed by a single particle, along with the velocity of that particle at time t.

The Navier–Stokes equations are equations for the velocity field $\boldsymbol{u}$. The first part of these equations is the ***continuity equation***, which represents the law of conservation of mass:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0. \tag{1.13}$$

Here ρ is a scalar-valued function on $\mathbb{R}^3$ giving the mass density of the fluid as a function on physical space. The operator "$\nabla\cdot$" is the divergence which we encountered in our discussion of Maxwell's equations above. Note that when ρ

is constant—which corresponds to incompressible flow—the continuity equation reads

$$\nabla \cdot \boldsymbol{u} = 0,$$

meaning that the velocity field preserves volume. Along with the mass conservation equation, we have a force/momentum balance equation that we will not provide any details for:

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}\right) = -\nabla p + \nabla \cdot (\mu(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T) - \tfrac{2}{3}\mu(\nabla \cdot \boldsymbol{u})\boldsymbol{I}) + \boldsymbol{f}. \tag{1.14}$$

These are the ***Navier–Stokes equations***.

Let us first define all of the mathematical components of this equation, at least so one can imagine writing these equations down in explicit form. The term $\nabla \boldsymbol{u}$ is the ***gradient*** or ***Jacobian*** of the velocity field, which is a 3×3-matrix:

$$\nabla \boldsymbol{u} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{bmatrix}.$$

The second term in the Navier–Stokes equations is the vector obtained by multiplying this matrix on the left by the vector $\boldsymbol{u}$. The variable p is the ***pressure field*** which is a scalar function, and ∇p represents the gradient of the pressure field, i.e., the vector field $\operatorname{grad} p = (\frac{\partial p}{\partial x_1}, \frac{\partial p}{\partial x_2}, \frac{\partial p}{\partial x_3})$. The variable μ is the ***viscosity***, and represents the internal forces in the fluid due to friction causes when creating strain gradients. Of course, $\boldsymbol{I}$ is the 3×3 identity matrix. Note that the second term on the right-hand side has the form $\nabla \cdot \boldsymbol{M}$ for a matrix function $\boldsymbol{M}$. This is a vector field, called the ***divergence*** of M. It is given explicitly by

$$\nabla \cdot \boldsymbol{M} = \left(\sum_{j=1}^{3} \frac{\partial M_{1j}}{\partial x_j}, \sum_{j=1}^{3} \frac{\partial M_{2j}}{\partial x_j}, \sum_{j=1}^{3} \frac{\partial M_{3j}}{\partial x_j}\right).$$

Finally, $\boldsymbol{f}$ are ***body forces***, e.g., gravitational effects.

The Navier–Stokes equations have four independent variables (x_1, x_2, x_3, t) and five dependent variables, ρ, p, and (u_1, u_2, u_3). It is, of course, an equation in the derivatives of the dependent variables with respect to the independent variables.

1.1.11 Heat flow due to temperature gradients

Our next modelling task is that of heat flow in a homogeneous medium. Let us specify the physical assumptions we make.

1. For simplicity we work with a one-dimensional medium, i.e., a rod.

2. We assume a homogeneous medium, i.e., its characteristics are constant as we move throughout. We assume the rod to have a constant cross-sectional area A.
3. Thermal energy is given by $Q = c\rho Vu$, where ρ is the mass density, V is the volume, u is temperature, and c is the specific heat of the medium. We assume ρ and c to be constant throughout the material.
4. We assume that rate of heat transfer from one region to another through a slice of the rod is proportional to the temperature gradient:

$$q = -K\frac{\partial u}{\partial x},$$

where q is the heat flow per unit area and x measures the distance along the rod. This is ***Fourier's law***.

5. Thermal energy is conserved in each chunk of the rod.

Let us use these assumptions to derive an equation governing the temperature distribution in a rod. Consider a chunk of the rod as shown in Figure 1.9. In the

Figure 1.9 A chunk of rod used in the derivation of the heat equation

figure, the rod chunk is shown at a fixed time. The quantity $q(a)$ denotes the rate of heat flow at the position $x = a$ on the rod, and $q(b)$ denotes the rate of heat flow at the position $x = b$ on the rod. In terms of the quantities in Figure 1.9, Fourier's law reads

$$q(a) = -K\frac{\partial u}{\partial x}\Big|_a, \quad q(b) = K\frac{\partial u}{\partial x}\Big|_b$$

for some constant $c > 0$. The signs result from the fact that heat will flow in a direction opposite the temperature gradient. If we assume that no heat escapes from the upper and lower boundaries of the rod, then the net change in heat in the rod chunk in a time Δt will be

$$\Delta Q = KA\Delta t\left(\frac{\partial u}{\partial x}\Big|_b - \frac{\partial u}{\partial x}\Big|_a\right), \tag{1.15}$$

With the assumptions we have made, the net change in heat in the chunk over a time Δt is given by

$$\Delta Q = c\rho A(b-a)\Delta t \frac{\partial u}{\partial t}, \tag{1.16}$$

where $\frac{\partial u}{\partial t}$ is the average of the time rate of change of temperature throughout the chunk and ρ is the mass density of the material. By making $(b-a)$ and Δt sufficiently small, one may ensure that $\frac{\partial u}{\partial t}$ does not vary much through the chunk. Equating (1.15) and (1.16) we get

$$c\rho A(b-a)\Delta t \frac{\partial u}{\partial t} = KA\Delta t \left(\frac{\partial u}{\partial x}\Big|_b - \frac{\partial u}{\partial x}\Big|_a \right)$$

Now, dividing by $\mu\Delta t(b-a)$ and taking the limit as $b-a$ goes to zero we get the ***heat equation***:

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}, \tag{1.17}$$

where $k = \frac{K}{c\rho} > 0$ is the ***diffusion constant***.

The heat equation has two independent variables x and t and a single dependent variable u. It is an equation in the derivatives of the dependent variable with respect to the independent variables. A multidimensional (in space) analogue of the heat equation is imaginable, and takes the form

$$\frac{\partial u}{\partial t} = k\left(\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} \right).$$

The operator in the right-hand side is of independent interest, and is known as the ***Laplacian*** of u and given by

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}.$$

With this bit of notation, the heat equation can be written as

$$\frac{\partial u}{\partial t} = k\Delta u.$$

We shall subsequently look at the heat equation in some detail, and shall say some things about the behaviour of its solutions at that time.

1.1.12 Waves in a taut string

Next we consider the small transverse vibrations of a taut string when it is plucked, e.g., a guitar string. To derive the equations governing these transverse vibrations, we use simple force balance on a short segment of the string. In Figure 1.10 we depict a little segment of a string with its transverse displacement

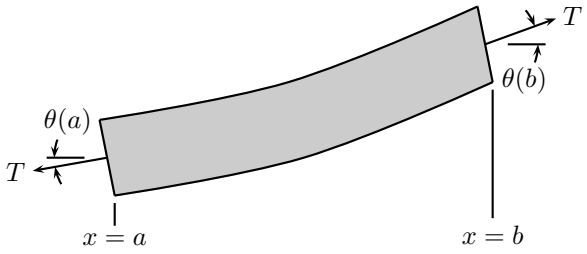


Figure 1.10 A segment of string used in the derivation of the wave equation

denoted u. It is assumed that the tension T in the string is independent of x and t. This is acceptable for small string deflections. The vertical component of the force on the string is given by

$$F_y = -T\sin(\theta(a)) + T\sin(\theta(b)).$$

Let us manipulate this until it looks like something we want. We denote the vertical deflection of the string by u. We then have

$$\tan\theta(a) = \frac{\partial u}{\partial x}\Big|_a, \quad \tan\theta(b) = \frac{\partial u}{\partial x}\Big|_b.$$

Now recall that for small angles θ we have $\sin\theta \approx \tan\theta$. This then gives

$$F_y \approx T\left(\frac{\partial u}{\partial x}\Big|_b - \frac{\partial u}{\partial x}\Big|_a\right).$$

Now the mass of the segment of string is $\rho(b-a)$ with ρ the length mass density of the string, which we assume to be constant. The vertical acceleration is then $\frac{\partial^2 u}{\partial t^2}$, which we suppose to be constant in the segment. By making the length of the segment sufficiently small, this becomes closer to being true. An application of force balance now gives

$$\rho(b-a)\frac{\partial^2 u}{\partial t^2} \approx T\left(\frac{\partial u}{\partial x}\Big|_b - \frac{\partial u}{\partial x}\Big|_a\right).$$

Dividing by $\rho(b-a)$ and letting $b-a$ go to zero, we have the ***wave equation***:

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2}, \tag{1.18}$$

where $c = \sqrt{\frac{T}{\rho}} > 0$ is the ***wave speed*** for the problem.

There are two independent variables x and t for the wave equation, and a single dependent variable u. The equation itself is one involving derivatives of the

dependent variable with respect to the independent variables. As with the heat equation, a multidimensional (in space) analogue of the heat equation is possible, and takes the form

$$\frac{\partial^2 u}{\partial t^2} = c^2 \left(\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} \right).$$

The operator in the right-hand side is the Laplacian which we saw with the heat equation:

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}.$$

The wave equation can be thus written as

$$\frac{\partial^2 u}{\partial t^2} = k \Delta u.$$

In the text we shall examine the wave equation in a little detail, and say some things about the behaviour of its solutions.

1.1.13 The potential equation in electromagnetism and fluid mechanics

In this section we shall see how the Laplacian, introduced in our discussion of the wave equation, arises in special cases of Maxwell's and Navier–Stokes' equations.

We first consider Maxwell's equations of electromagnetism. We make a few assumptions about the physics that will allow us to simplify the complicated Maxwell's equations.

1. We assume we are in steady-state, so the dependent variable do not depend on time.
2. We assume that the electric field $\boldsymbol{E}$ is a potential field. This means that there exists a function V, called the ***electric potential***, such that $\boldsymbol{E} = \nabla V = (\frac{\partial V}{\partial x_1}, \frac{\partial V}{\partial x_2}, \frac{\partial V}{\partial x_3})$.
3. We assume that we are in free space so the charge density is zero.

The equations for the potential function are determined by Gauss's law:

$$\nabla \cdot \boldsymbol{E} = 0 \quad \Longrightarrow \quad \nabla \cdot \nabla V = 0.$$

A direct computation gives

$$\nabla \cdot \nabla V = \Delta V = \frac{\partial^2 V}{\partial x_1^2} + \frac{\partial^2 V}{\partial x_2^2} + \frac{\partial^2 V}{\partial x_3^2}. \tag{1.19}$$

This is the ***potential equation*** in $\mathbb{R}^3$.

Next we turn to a special case of the Navier–Stokes equations, making the following physical assumptions.

1. The flow in inviscid, so the viscosity μ vanishes.

2. The flow in incompressible, so the divergence of the fluid velocity vanishes.
3. We assume the fluid velocity is derived from a velocity potential: $\boldsymbol{u} = -\nabla\phi$.
4. We suppose that body forces are potential forces, i.e., $\boldsymbol{f} = -\nabla V$, e.g., gravitational forces.

In this case, the assumptions of incompressibility and the existence of a velocity potential give the following form of the equation of continuity:

$$\nabla \cdot \boldsymbol{u} = 0 \quad \Longrightarrow \quad \Delta\phi = 0.$$

Let us investigate the impact of this, along with the other physical assumptions, in describing properties of the fluid flow. First of all, a direct computation gives

$$\boldsymbol{u} \cdot \nabla \boldsymbol{u} = (\nabla \times \boldsymbol{u}) \times \boldsymbol{u} + \operatorname{grad}(\tfrac{1}{2}\boldsymbol{u} \cdot \boldsymbol{u}),$$

where $\boldsymbol{a} \times \boldsymbol{b}$ denotes the vector cross-product and $\boldsymbol{a} \cdot \boldsymbol{b}$ denotes the Euclidean inner product of $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$. Since $\boldsymbol{u} = -\nabla\phi$, we calculate that $\nabla \times \boldsymbol{u} = \boldsymbol{0}$, and so the Navier—Stokes equations read

$$\nabla\left(\frac{\partial\phi}{\partial t} + \frac{1}{2}(\boldsymbol{u} \cdot \boldsymbol{u}) + \frac{p}{\rho} + V\right) = 0.$$

This implies that

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}(\boldsymbol{u} \cdot \boldsymbol{u}) + \frac{p}{\rho} + V$$

depends only on t. This is known as ***Bernoulli's principle***.

Let us indicate another way in which the Laplacian arises in fluid flow problems, in this case with planar flow problems, i.e., that $u_3 = 0$. We assume that the fluid velocity $(u_1, u_2, 0)$ has the special form

$$u_1 = \frac{\partial\psi}{\partial x_2}, \quad u_2 = -\frac{\partial\psi}{\partial x_1}$$

for a function ψ of (x_1, x_2) called the ***stream function***. Note that the resulting fluid velocity automatically satisfies the incompressible continuity equation:

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} = \frac{\partial^2\psi}{\partial x_1 \partial x_2} - \frac{\partial^2\psi}{\partial x_2 \partial x_1} = 0.$$

If we additionally require that $\Delta\psi = 0$, then $\nabla \times \boldsymbol{u} = \boldsymbol{0}$. In this case, we recall from vector calculus that $\boldsymbol{u} = -\operatorname{grad}\phi$, i.e., the flow is a potential flow.

1.1.14 Einstein's field equations

In Einstein's theory of general relativity, a ***spacetime*** is a four-dimensional "differentiable manifold." This means that around every point in spacetime there is a parameterisation by $\mathbb{R}^4$. To keep things simple (and still representative), we just assume that our spacetime is equal to $\mathbb{R}^4$. There are two physical objects defined on spacetime, and Einstein's field equations relate these. The first is the ***stress-energy tensor*** $\boldsymbol{T}$ which is a symmetric 4×4 matrix function. This encodes the properties of spacetime like mass and electromagnetic fields. The other object defined on spacetime of interest is the ***metric tensor*** $\boldsymbol{g}$, which is another symmetric 4×4 matrix function, this one having the property that it has one negative and three positive eigenvalues. Physically, $\boldsymbol{g}$ determines the gravitational properties of spacetime, as well as the space and time structure.

We definitely will not derive Einstein's field equations, but will simply produce them. First of all, we denote the coordinates for spacetime by (x^1, x^2, x^3, x^4); the use of superscripts as indices is traditional in general relativity. The components of the matrices $\boldsymbol{T}$ and $\boldsymbol{g}$ we denote by T^{jk} and g_{jk}, $j, k \in \{1, 2, 3, 4\}$. First we define the ***Christoffel symbols*** associated with $\boldsymbol{g}$:

$$\gamma^j_{kl} = \frac{1}{2} \sum_{m=1}^{4} g^{jm} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right),$$

where g^{jk}, $j, k \in \{1, 2, 3, 4\}$, are the components of $\boldsymbol{g}^{-1}$. Next, the ***curvature tensor*** is then defined by

$$R^j_{klm} = \frac{\partial \Gamma^j_{lm}}{\partial x^k} - \frac{\partial \Gamma^j_{km}}{\partial x^l} + \Gamma^j_{km} \Gamma^m_{lm} - \Gamma^j_{lm} \Gamma^m_{km},$$

the ***Ricci tensor*** is the 4×4-symmetric matrix function **Ric** defined by

$$\mathrm{Ric}_{jk} = \sum_{l=1}^{4} R^l_{ljk}, \qquad j, k \in \{1, 2, 3, 4\},$$

and the ***scalar curvature*** is function defined by

$$\rho = \sum_{j,k=1}^{4} g^{jk} \, \mathrm{Ric}_{jk} \, .$$

Finally, we define the contravariant form of the stress-energy tensor, which is the symmetric 4×4-matrix function $\overline{\boldsymbol{T}}$ with components

$$\overline{T}_{jk} = \sum_{l,m=1}^{4} g_{jl} g_{km} T^{lm}, \qquad j, k \in \{1, 2, 3, 4\}.$$

With all of this data, we can now write the ***Einstein field equations***:

$$\mathbf{Ric} - \tfrac{1}{2}\rho \boldsymbol{g} + \Lambda \boldsymbol{g} = \frac{8\pi G}{c^4}\overline{\boldsymbol{T}}, \tag{1.20}$$

where Λ is the ***cosmological constant***, G is the ***gravitational constant***, and c is the speed of light in a vacuum.

There are four independent variables in Einstein's field equations, the coordinates (x^1, x^2, x^3, x^4) for spacetime. There are nominally ten dependent variables (the sixteen components of $\boldsymbol{g}$ taking into account symmetry). The equations are complicated equations in the derivatives of dependent variables with respect to the independent variables.

Of course, we will not say anything about the nature of the solutions to Einstein's field equations. This is the subject of deep work by many smart people.

1.1.15 The Schrödinger equation

In quantum mechanics, the Schrödinger equation governs the behaviour of a function known as the ***wave function***. The wave function encodes the state of a quantum system in the form of a "probability amplitude." These are typically complex-valued as they come equipped with, not just an amplitude, but a phase. This phase allows for the wave part of the particle/wave duality seen in the behaviour of subatomic particles. We shall not delve into the quantum mechanical machinations required to understand where the equation comes from, but shall merely produce the Schrödinger equation for the wave function ψ of a single particle moving in $\mathbb{R}^3$ in an electric field with electric potential function V:

$$\mathrm{i}\hbar\frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2\mu}\Delta\psi + V\psi, \tag{1.21}$$

where $\mathrm{i} = \sqrt{-1}$, $\hbar$ is Planck's constant, and μ is the effective mass of the particle.

Note that the Schrödinger equation is an equation with four independent variables, (x_1, x_2, x_3) and t, and a single complex-valued dependent variable ψ, or equivalently, regarding a complex number as determined by its real and imaginary parts, two real dependent variables. Of course, the equation is one involving the derivatives of the dependent variable with respect to the independent variables.

1.1.16 The Black–Scholes equation

The model we arrive at in this section is widely used in options trading, and has garnered a Nobel Prize in Economics for its developers. It is also true that the widespread misuse of this model, and models like it, combined with greed and governments divesting themselves of regulatory responsibilities, has led to the ruination of millions of lives. So mathematics *can* make a difference in peoples lives!

The equation we give provides the price V of an option as a function of stock price S and time t. It also has the following parameters:

r risk-free compound interest rate
σ standard deviation of stock's returns

We shall not describe the "derivation" of the model, but simply state the ***Black–Scholes equation***:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0.$$

For this equation, there are two independent variables (t, S) and a single independent variable V. The equation involves derivatives of the independent variable with respect to the dependent variables.

Now you can go off into a room and run Black–Scholes simulations, and make yourself rich!

1.1.17 Summary

In this section we have presented myriad illustrations of how equations involving various numbers of independent and dependent variables, along with derivatives of these, may arise in applications. The subject of this text is how to solve some such equations, and how to look for the essential attributes of equations such as these. This is the subject of "differential equations." It is a subject that is impossible to comprehend fully in any sort of generality, which is not unreasonable since differential equations describe physical phenomenon that we do not expect to be able to understand fully. Thus the subject of differential equations is a combination of looking deeply at certain special cases (particularly linear equations) and working hard to determine characteristic behaviour of general classes of systems.

1.1.18 Notes

[Brown 2007, page 68]

Exercises

1.1.1 Think of, or Google, three models (not included in the text) where differential equations arise in practice. In each case, do the following:
   (a) indicate the independent and dependent variables;
   (b) give some meaning to these variable in terms of the particular application;
   (c) provide a tiny bit of background about where the equations come from.

Section 1.2

The mathematical background and notation required to read this text

One of the attributes of the course for which these notes are developed is a slightly higher level of mathematical rigour. In this section, therefore, we overview our expectations of the background of students. We also introduce some concepts, terminology, and notation that will arise frequently in the course of our presentation.

1.2.1 Elementary mathematical notation

We will use standard mathematical notation which we overview here for reference.

Given a set S, if x is an element of S we shall write $x \in S$. If A is a subset of S, we shall write $A \subseteq S$. This allows for the possibility that $A = S$. If we wish to exclude this possibility, then A is a ***strict*** subset of S and we will write $A \subset S$. If S is a set and $A \subseteq S$, then $S \setminus A$ is the set of elements of S that are not in A. For sets S and T, $S \cup T$ denotes the ***union*** of S and T, i.e., the set whose elements are from either S or T. By $S \cap T$ we denote the ***intersection*** of S and T, i.e., the set whose elements are in both of S and T. For sets S and T, we denote the Cartesian product of S and T by $S \times T$, noting that elements of $S \times T$ take the form (x, y) for $x \in S$ and $y \in T$. Of course, we can talk about arbitrary finite unions, intersections, and products in the same way. Thus if $S_1, \dots, S_k$ are sets an element of the Cartesian product $S_1 \times \cdots \times S_k$ takes the form $(x_1, \dots, x_k)$, where $x_j \in S_j$ for $j \in \{1, \dots, k\}$. The empty set, i.e., the set with no elements is denoted by $\emptyset$.

Sets will frequently be prescribed by placing restrictions on elements of another set. Let S be a set and let P be a predicate in S. Thus P is a rule for assigning a value of True or False to each element of S. Thus we can regard P as a map $P\colon S \to \{\text{True}, \text{False}\}$. We can then define a subset of S to be the set of elements of S to which P assigns a value True. The notation we use for this is

$$\{x \in S \mid P(x) = \text{True}\}.$$

For sets S and T, a ***function*** or ***map*** from S to T is a rule that assigns to each point in S a unique point in T. The rule is typically given a name like "f," and so, given $x \in S$, $f(x) \in T$ is the element assigned to x by the map f. To signify that f is a map from S to T, we shall write $f\colon S \to T$. We call S the ***domain*** of f and T the ***codomain*** of f. Thus we shall frequently write things like, "Consider a map $f\colon S \to T$." If $f\colon R \to S$ and $g\colon S \to T$ are maps, the ***composition*** of f and g is the map $g \circ f\colon R \to T$ defined by $g \circ f(x) = g(f(x))$ for $x \in R$. If $f\colon S \to T$ and if $A \subseteq S$, we denote by $f|A$ the ***restriction*** of f to A, which is the same map as f, but

only taking inputs as points in A. We say that $f\colon S \to T$ is ***injective*** if $f(x_1) = f(x_2)$ for $x_1, x_2 \in S$, then this implies that $x_1 = x_2$. We say that $f\colon S \to T$ is ***surjective*** if, given $y \in T$, there exists $x \in S$ such that $f(x) = y$. We say that $f\colon S \to T$ is ***bijective*** if it is both injective and surjective. For a set S, the ***identity map*** on S is the map $\mathrm{id}_S\colon S \to S$ defined by id_S.

By $\mathbb{Z}$ we denote the set of integers, with $\mathbb{Z}_{\geq 0}$ denoting the nonnegative integers and $\mathbb{Z}_{>0}$ denoting the positive integers. The real numbers we denote by $\mathbb{R}$, with $\mathbb{R}_{\geq 0}$ denoting the nonnegative real numbers and $\mathbb{R}_{>0}$ denoting the positive real numbers. By $\mathbb{R}^n$ we denote the n-fold Cartesian product of $\mathbb{R}$ with itself. Thus an element of $\mathbb{R}^n$ has the form $(x_1, \dots, x_n)$ for $x_j \in \mathbb{R}$, $j \in \{1, \dots, n\}$. We will denote this with a bold font:

$$\boldsymbol{x} = (x_1, \dots, x_n).$$

We shall frequently consider subsets of $\mathbb{R}$ known as intervals. An ***interval*** is a subset of one of the following nine forms:

$$\begin{aligned}
(a, b) &= \{x \in \mathbb{R} \mid a < x < b\},\\
(a, b] &= \{x \in \mathbb{R} \mid a < x \leq b\},\\
[a, b) &= \{x \in \mathbb{R} \mid a \leq x < b\},\\
[a, b] &= \{x \in \mathbb{R} \mid a \leq x \leq b\},\\
(a, \infty) &= \{x \in \mathbb{R} \mid a < x < \infty\},\\
[a, \infty) &= \{x \in \mathbb{R} \mid a \leq x < b\},\\
(-\infty, b) &= \{x \in \mathbb{R} \mid -\infty < x < b\},\\
(-\infty, b] &= \{x \in \mathbb{R} \mid -\infty < x \leq b\},\\
(-\infty, \infty) &= \mathbb{R}.
\end{aligned} \tag{1.22}$$

Note that, for our purposes, things like $[a, \infty]$ do not make sense: ∞ is not an element of $\mathbb{R}$.

1.2.2 Complex numbers

We will work with complex numbers, and we suppose that the reader has a passing familiarity with these. Here we shall provide the few facts that we shall require.

1.2.2.1 Complex arithmetic We define the set $\mathbb{C}$ of ***complex numbers*** to be $\mathbb{C} = \mathbb{R}^2$, equipped with the following operations:

1. ***Addition:*** $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$;
2. ***Multiplication:*** $(x_1, y_1) \cdot (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1)$;
3. ***Inversion:*** $(x, y)^{-1} = (\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2})$ when $x^2 + y^2 \neq 0$.

One can readily verify that these definitions of arithmetic have the familiar commutativity, associativity, and distributivity properties of arithmetic with real numbers.

In practice, one almost never writes a complex number as (x, y). Instead, one denotes $\mathrm{i} = (0, 1)$ and notes that

$$(x, 0) + \mathrm{i}(0, y) = (x, 0) + (0, 1)(0, y) = (x, 0) + (0, y) = (x, y).$$

Thus it is entirely reasonable to write $(x, y) = x + \mathrm{i}y$, and this is the almost universally used notation for writing a complex number. If we wish to abbreviate, a typical complex number is often written as $z = x + \mathrm{i}y$ for $x, y \in \mathbb{R}$. We call x the ***real part*** of z, denoted by $x = \mathrm{Re}(z)$, and y the ***imaginary part*** of z, denoted by $y = \mathrm{Im}(z)$. Note that $\mathrm{i} \cdot \mathrm{i} = -1 + \mathrm{i}0$, and so we have $\mathrm{i} = \sqrt{-1}$. The fact that -1 does not have a real square root is, in some sense, the whole point of using complex numbers. If $z = x + \mathrm{i}y \in \mathbb{C}$, the ***complex conjugate*** of z is $\overline{z} = x - \mathrm{i}y$: $\overline{z}$ is the reflection of z about the real axis as in Figure 1.11.

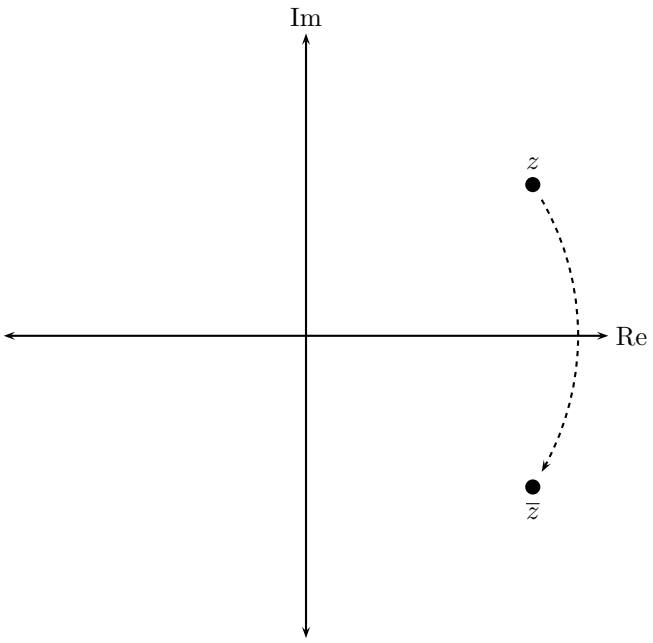


Figure 1.11 Complex conjugation

1.2.2.2 Polar representation There is an alternative means of representing a complex number, namely the ***polar representation***. This works as follows. Let $z = x + \mathrm{i}y \in \mathbb{C} \setminus \{0\}$. Then there exists a unique $r \in \mathbb{R}_{\geq 0}$ and $\theta \in (-\pi, \pi]$ such that

$$z = r(\cos\theta + \mathrm{i}\sin\theta),$$

see Figure 1.12. Specifically, $r = \sqrt{x^2 + y^2}$ and $\theta = \mathrm{atan}(x, y)$, where $\mathrm{atan}\colon \mathbb{R}^2 \setminus \{(0, 0)\} \to (-\pi, \pi]$ is the "smart" arctangent function, see Figure 1.13. For the polar

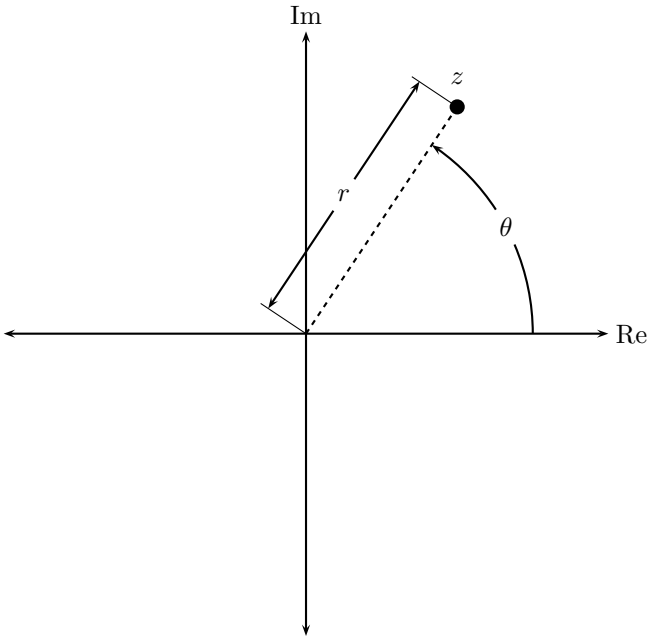


Figure 1.12 The polar representation of a complex number

representation, it is common to use ***Euler's formula*** which is

$$e^{i\theta} = \cos\theta + i\sin\theta$$

for $\theta \in \mathbb{R}$.[3]

For our purposes, we shall just use this as a short form, but when you learn about functions of a complex variable, you will learn about the exponential function, and

[3]Here's a justification of Euler's formula. A reader likely knows the Taylor series formula for the exponential function:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This formula is valid for complex numbers, and defines the ***complex exponential function***:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Let us work this out for $z = i\theta$:

$$\begin{aligned} e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} = \sum_{n=0}^{\infty} \frac{(i\theta)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(i\theta)^{2n+1}}{(2n+1)!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n\theta^{2n}}{(2n)!} + i\sum_{n=0}^{\infty} \frac{(-1)^n\theta^{2n+1}}{(2n+1)!} = \cos\theta + i\sin\theta. \end{aligned}$$

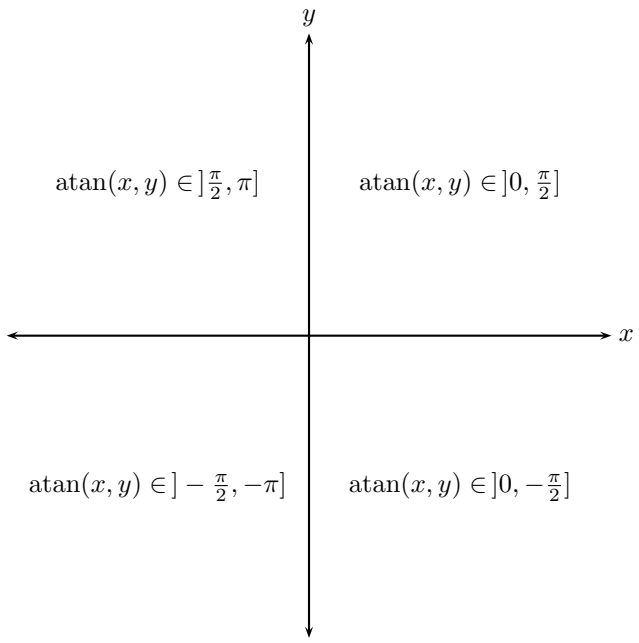


Figure 1.13 The "smart" arctangent function

then see that this is a relationship between three functions, the exponential, cosine, and sine functions. Note that, with Euler's formula, we write $z = re^{i\theta}$. One can verify that this representation interacts in predictable ways with multiplication. Thus, if we write $z_1, z_2 \in \mathbb{C} \setminus \{0\}$ as

$$z_1 = r_1 e^{i\theta_1}, \;\; z_2 = r_2 e^{i\theta_2},$$

then

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}.$$

In particular, we have the useful formula

$$z = re^{i\theta} \quad \Longrightarrow \quad z^n = r^n e^{in\theta}.$$

1.2.2.3 Roots of complex numbers Complex numbers are, in some way of thinking about them, conjured especially because of their nice properties upon taking roots. For example, not all real numbers have real square roots, e.g., all negative real numbers do not have real square roots. However, given any $w \in \mathbb{C}$, the equation $z^n = w$ can be solved for z, and indeed has exactly n solutions. Let us explore this.

We write w and z using polar representations:

$$w = \rho e^{i\phi}, \quad z = re^{i\theta}.$$

The equation $z^n = w$ then reads $r^n e^{in\theta} = \rho e^{i\phi}$. This has the one "obvious" solution given by $r = \sqrt[n]{\rho}$ and $\theta = \phi/n$. However, there are other solutions. Indeed, note that, for any $k \in \mathbb{Z}$, we have $w = e^{i\phi+2k\pi}$. Thus we have solutions

$$z_k = \sqrt[n]{\rho} e^{i(\phi+2k\pi)/n}, \quad k \in \mathbb{Z}.$$

This makes it seem like there are then infinitely many solutions for $z^n = w$. However, note that

$$e^{i(\phi+2(k+n)\pi)/n} = e^{i(\phi+2k\pi)/n+i(2\pi)} = e^{i(\phi+2k\pi)/n}, \qquad k \in \mathbb{Z}.$$

Therefore, there are, in fact, n solutions to the equation $z^n = w$, and these are

$$z_k = \sqrt[n]{\rho} e^{i(\phi+2k\pi)/n}, \qquad k \in \{0, 1, \dots, n\}.$$

1.2.3 Polynomials

We shall suppose that the reader knows what a polynomial is, and how to find the roots of a degree 2 polynomial using the quadratic formula. Here we shall say a few more things about this, since part of some algorithms for solving differential equations involves first finding the roots of a polynomial. Moreover, the nature of the solution of the differential equation depends on the particularities of the roots. We shall be primarily interested in polynomials with real coefficients, although we shall see that complex numbers inevitably enter the frame, even in this case.

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A ***polynomial over*** $\mathbb{F}$ is a linear combination

$$a_k X^k + \cdots + a_1 X + a_0$$

of powers of an ***indeterminate*** X,[4] with ***coefficients*** $a_0, a_0, \dots, a_k \in \mathbb{F}$. The ***degree*** of a polynomial is the largest $k \in \mathbb{Z}_{\geq 0}$ for which $a_k \neq 0$. A polynomial of degree k is ***monic*** if $a_k = 1$. We denote by $\mathbb{F}[X]$ the set of polynomials over $\mathbb{F}$. We shall use symbols like "P" to denote polynomials. We assume the reader knows how to multiply and add polynomials.

For us, the main feature of a polynomials is its roots. First we need to evaluate polynomials. Let $P \in \mathbb{F}[X]$ be a degree k polynomial given by

$$P = a_k X^k + \cdots + a_a X + a_0.$$

Associated to P is a function $\widehat{P}\colon \mathbb{F} \to \mathbb{F}$ defined by, of course,

$$\widehat{P}(x) = a_k x^k + \cdots + a_1 x + a_0.$$

A ***root*** of P is then $\lambda \in \mathbb{F}$ such that $\widehat{P}(\lambda) = 0$. If $\lambda \in \mathbb{F}$ is a root of P, then we can write

$$P = (X - \lambda)P_1,$$

[4] We shall not be very precise about just what an indeterminate is; you can think of it as being a variable.

where P_1 is a polynomial of degree $k-1$. It may happen that λ is a root of P_1, in which case the same argument gives

$$P = (X-\lambda)^2 P_2$$

for a polynomial P_2 of degree $k-2$. We can continue in this way until we arrive at the largest $m \in \{1,\dots,k\}$ for which

$$P = (X-\lambda)^m P_m$$

for a polynomial P_m of degree $k-m$ for which $\widehat{P}_m(\lambda) \neq 0$. The number m is the ***multiplicity*** of the root λ, and we denote this by $m(\lambda, P)$.

Let us consider the nature of roots of polynomials.

1. The Fundamental Theorem of Algebra says that, if $P \in \mathbb{C}[X]$, then P has a root. If P has degree k, the number of roots can be any number in $\{1,\dots,k\}$. For example, if $\lambda \in \mathbb{C}$, the polynomial

   $$P = (X-\lambda)^k$$

   has only one root, while, if $\lambda_1,\dots,\lambda_k \in \mathbb{C}$ are distinct, then the polynomial

   $$P = (X-\lambda_1)\cdots(X-\lambda_k)$$

   has k roots.
2. If $P \in \mathbb{R}[X]$, then it is possible that P has no roots, e.g., $P = X^1 + 1$. However, if $P \in \mathbb{R}[X]$ there is the naturally associated $\overline{P} \in \mathbb{C}[X]$ with the same coefficients, keeping in mind that $\mathbb{R} \subseteq \mathbb{C}$. This polynomials will have roots, as in 1. In this case, we say that the real polynomial P has ***complex roots***.
3. Because of the realness of the coefficients of $P \in \mathbb{R}[X]$, the arrangement of the roots is not arbitrary, however. For example, if $\lambda \in \mathbb{C}$ is a root of P, possibly complex, then $\overline{\lambda}$ is also a root.[5] Thus, if one lays down points in the complex plane corresponding to the complex roots of a real polynomial, the configuration will be symmetric about the real axis.

[5] This is easy to see. If λ is a root, then

$$\begin{aligned}
& a_k\lambda^k + \cdots + a_1\lambda + a_0 = 0 \\
\implies & \overline{a_k\lambda^k + \cdots + a_1\lambda + a_0} = 0 \\
\implies & \overline{a_k\lambda^k} + \cdots + \overline{a_1\lambda} + \overline{a_0} = \overline{0} \\
\implies & \overline{a_k}\overline{\lambda}^k + \cdots + \overline{a_1}\overline{\lambda} + \overline{a_0} = 0 \\
\implies & a_k\overline{\lambda}^k + \cdots + a_1\overline{\lambda} + a_0 = 0.
\end{aligned}$$

4. One can compute the roots of a degree 2 polynomial using the quadratic formula. There are similar formulae for polynomials of degree 3 and 4. Things change with degree 5, however. The ***Abel–Ruffini Theorem*** tells us that there is no formula for the roots of a degree 5 polynomial that involves addition, subtraction, multiplication, division, and rational powers in the coefficients.

The reader is invited to enumerate the possible root configurations of a degree 5 real polynomial in Exercise 1.2.1.

1.2.4 Linear algebra

We shall make use of linear algebra in not completely trivial ways, particularly when dealing with systems of linear ordinary differential equations. In this section we overview the required ideas.

One of the complications of dealing with linear differential equations is that, even if one works solely with real equations, one must work with complex solutions. Thus, to have at hand a useful theory, we will need to work with vector spaces over both the real and complex numbers. To facilitate this, we shall use the symbol $\mathbb{F}$ to represent either the real or complex numbers; we shall do that by writing things like, "let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$."

1.2.4.1 Vector spaces and subspaces Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. We start by recalling that a ***$\mathbb{F}$-vector space*** is a set V equipped with two operations, vector addition and scalar multiplication, wherein one adds elements of V and multiplies an element of V by a scalar from $\mathbb{F}$, respectively. One also posits the existence of a zero vector, typically denote by 0, and an additive inverse $-v$ for every $v \in \mathsf{V}$. These operations are required to satisfy a list of commutativity, associativity, and distributivity axioms, and there are a host of other properties one derives from these. We suppose the reader to have seen this sort of thing in their first course on linear algebra. A subset $\mathsf{U} \subseteq \mathsf{V}$ is a ***subspace*** if $u_1 + u_2, au \in \mathsf{U}$ for every $u, u_1, u_2 \in \mathsf{U}$ and every $a \in \mathbb{F}$, i.e., if U is "closed under vector addition and scalar multiplication."

While we suppose the reader to have seen such example previously, let us give two examples of vector spaces that represent the sorts of vector spaces we shall use in this text.

1.2.1 Examples (Vector spaces)

1. The set $\mathbb{R}^n$ is a vector space with the vector space operations

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$
$$a(x_1, \dots, x_n) = (ax_1, \dots, ax_n).$$

As example of a subspace of $\mathbb{R}^n$ is the subset

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_{k+1} = \cdots = x_n = 0\}$$

for some $k \in \{1, \dots, n\}$.

2. Let I be an interval, any one of the nine intervals defined in (1.22), and let $\mathbb{F} \in \{\mathbb{R}; \mathbb{C}\}$. For $k \in \mathbb{Z}_{\geq 0}$, we denote by $\mathsf{C}^k(I; \mathbb{F})$ the set of k-times continuously differentiable functions from I to $\mathbb{F}$. If $k = 0$ we adopt the convention that this means continuous functions. We define a $\mathbb{F}$-vector space structure on $\mathsf{C}^k(I; \mathbb{F})$ by
$$(f + g)(x) = f(x) + g(x), \quad (af)(x) = a(f(x)).$$
Note that, because the derivative of a sum of functions is the sum of the derivatives of the functions, and because the derivative of a scalar times a function is the scalar times the derivative of the function, it follows that $\mathsf{C}^k(I; \mathbb{F})$ is indeed a well-defined $\mathbb{F}$-vector space. Note that, if $k < l$, then $\mathsf{C}^l(I; \mathbb{F}) \subseteq \mathsf{C}^k(I; \mathbb{F})$ is a subspace. •

1.2.4.2 Linear independence and bases Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. Let us merely enumerate the notions that will be of interest to us here.

1. A set $\{v_1, \ldots, v_k\} \subseteq \mathsf{V}$ of vectors is ***linearly independent*** if, for $c_1, \ldots, c_k \in \mathbb{F}$ satisfying
$$c_1 v_1 + \cdots + c_k v_k = 0,$$
it must be the case that $c_1 = \cdots = c_k = 0$. We suppose the reader to be intimately familiar with the notion of linear independence.
2. For a subset $\{v_1, \ldots, v_k\} \subseteq \mathsf{V}$, we denote
$$\operatorname{span}_{\mathbb{F}}(v_1, \ldots, v_k) = \{c_1 v_1 + \cdots + c_k v_k \mid c_1, \ldots, c_k \in \mathbb{F}\},$$
which is the ***span*** of $\{v_1, \ldots, v_k\}$. Then this is the set of all linear combinations of the vectors, and is the smallest subspace that contains all of the vectors.
3. A ***basis*** for V is a set $\{e_1, \ldots, e_n\} \subseteq \mathsf{V}$ that is (a) linearly independent and for which (b) $\operatorname{span}_{\mathbb{F}}(e_1, \ldots, e_n) = \mathsf{V}$. It follows, for example, that if $\{e_1, \ldots, e_n\}$ is a basis for V, then, for every $v \in \mathsf{V}$, there exist *unique* $c_1, \ldots, c_n \in \mathbb{F}$ such that
$$v = c_1 e_1 + \cdots + c_n e_n.$$
These are the ***components*** of v relative to this basis. The number n, if it exists, is the same for any basis, and is the ***dimension*** of V, denoted by $\dim_{\mathbb{F}}(\mathsf{V})$. A vector space possessing a basis in the sense we define here is ***finite-dimensional***.[6]

There is a change of basis formula for the components of a vector. To set this up, let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $v \in \mathsf{V}$. Suppose that we are given bases $\mathscr{E} = \{e_1, \ldots, e_n\}$ and $\mathscr{E}' = \{e'_1, \ldots, e'_n\}$. We can then write
$$e'_j = \sum_{k=1}^{n} P_{kj} e_k, \qquad j \in \{1, \ldots, n\}, \tag{1.23}$$

[6] All vector spaces possess a basis in a more general sense that we will not discuss here. In this more general sense, a vector space with a finite basis is finite-dimensional, and one with a basis that is not finite is ***infinite-dimensional***.

for some (necessarily invertible) matrix

$$\boldsymbol{P} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{1n} & P_{2n} & \cdots & P_{nn} \end{bmatrix},$$

called the ***change of basis matrix***. If we write

$$v = c_1 e_1 + \cdots + c_n e_n = c'_1 e'_1 + \cdots + c'_n e'_n$$

where $(c_1, \ldots, c_n)$ and $(c'_1, \ldots, c'_n)$ are the components of v relative to the bases $\mathscr{E}$ and $\mathscr{E}'$, then one readily determines that

$$c'_j = \sum_{k=1}^{n} Q_{jk} c_k,$$

where

$$\boldsymbol{P}^{-1} = \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{1n} & Q_{2n} & \cdots & Q_{nn} \end{bmatrix}.$$

In matrix form

$$\boldsymbol{c}' = \boldsymbol{P}^{-1} \boldsymbol{c}, \tag{1.24}$$

and this is the ***change of basis formula***.

1.2.4.3 Linear maps Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let U and V be $\mathbb{F}$-vector spaces. We recall that a ***linear map*** from U to V is a map $L \colon \mathsf{U} \to \mathsf{V}$ for which

$$L(u_1 + u_2) = L(u_1) + L(u_2), \quad L(au) = aL(u)$$

for every $u, u_1, u_2 \in \mathsf{U}$ and $a \in \mathbb{F}$. In case $\mathsf{U} = \mathsf{V}$ and so $L \colon \mathsf{V} \to \mathsf{V}$, then we may refer to L as a ***linear transformation***. We denote by

$$\operatorname{image}(L) = \{L(u) \in \mathsf{V} \mid \ u \in \mathsf{U}\}, \quad \ker(L) = \{u \in \mathsf{U} \mid \ L(u) = 0\}$$

the ***image*** and ***kernel*** of L. By $\mathrm{L}(\mathsf{U}; \mathsf{V})$ we denote the set of linear maps from U to V. The set of linear maps is itself an $\mathbb{F}$-vector space, with the vector space operations defined by

$$(L_1 + L_2)(u) = L_1(u) + L_2(u), \quad (aL)(u) = a(L(u)), \qquad u \in \mathsf{U},$$

for $L, L_1, L_2 \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ and $a \in \mathbb{F}$.

For the specific case of the vector spaces $\mathsf{U} = \mathbb{R}^m$ and $\mathsf{V} = \mathbb{R}^n$, linear maps are identified with matrices in a natural way. Given a matrix

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix},$$

the associated linear map is given by matrix/vector multiplication:

$$\boldsymbol{A}(\boldsymbol{x}) = \left(\sum_{a=1}^{m} A_{1a} x_a, \ldots, \sum_{a=1}^{m} A_{na} x_a \right).$$

Thus we can think of, and will think of, the set of $n \times m$ matrices as being $\mathrm{L}(\mathbb{R}^m; \mathbb{R}^n)$, not using any special notation for these being matrices. We do have some special notation for matrices. The $n \times n$ ***identity matrix***, i.e., corresponding to the identity map on $\mathbb{F}^n$, is

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The ***transpose*** of the matrix $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^m; \mathbb{R}^n)$ is the matrix $\boldsymbol{A}^T \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ given by

$$\boldsymbol{A}^T = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1m} & A_{2m} & \cdots & A_{nm} \end{bmatrix},$$

i.e., $\boldsymbol{A}^T$ is $\boldsymbol{A}$ with the columns turned into rows. A matrix $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ is ***symmetric*** if $\boldsymbol{A}^T = \boldsymbol{A}$ and ***skew-symmetric*** if $\boldsymbol{A}^T = -\boldsymbol{A}$.

Linear maps can be composed in the obvious way. If $L \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ and $M \in \mathrm{L}(\mathsf{V}; \mathsf{W})$, then the composition $M \circ L$ is an element of $\mathrm{L}(\mathsf{U}; \mathsf{W})$. If $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, then we denote by L^k the k-fold composition of L with itself:

$$L^k = \underbrace{L \circ \cdots \circ L}_{k \text{ times}} \in \mathrm{L}(\mathsf{V}; \mathsf{V}).$$

There is a notion regarding linear maps and subspaces that perhaps are unfamiliar to readers, but which are elementary and will be useful for us.

1.2.2 Definition Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. A subspace $\mathsf{U} \subseteq \mathsf{V}$ is **L-*invariant*** if $L(u) \in \mathsf{U}$ for every $u \in \mathsf{U}$. •

Note that the notion of invariant subspaces generally only makes sense for linear transformations.

We assume the reader is familiar with matrix representations of linear maps in bases. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let U and V be finite-dimensional $\mathbb{F}$-vector spaces, and let $L \in \mathrm{L}(\mathsf{U}; \mathsf{V})$. Suppose that we have bases $\mathscr{F} = \{f_1, \dots, f_m\}$ for U and $\mathscr{E} = \{e_1, \dots, e_n\}$ for V. Then there exist unique $L_a^j \in \mathbb{F}$, $a \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$, for which

$$L(f_a) = \sum_{j=1}^{n} L_{ja} v_j,$$

merely by properties of bases. We then define the $n \times m$ matrix

$$[L]_{\mathscr{F}}^{\mathscr{E}} = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1m} \\ L_{21} & L_{22} & \ddots & L_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \cdots & L_{nm} \end{bmatrix},$$

which is the ***matrix representative*** of L with respect to the bases $\mathscr{F}$ and $\mathscr{E}$.

As with the components of vectors, there is a change of basis formula for matrix representatives of linear maps. To formulate this, we suppose that we are given two bases $\mathscr{F} = \{f_1, \dots, f_m\}$ and $\mathscr{F}' = \{f'_1, \dots, f'_m\}$ for U, and two bases $\mathscr{E} = \{e_1, \dots, e_m\}$ and $\mathscr{E}' = \{e'_1, \dots, e'_n\}$ for V. We then have two change of basis matrices $\boldsymbol{Q}$ and $\boldsymbol{P}$ that are defined by

$$f'_a = \sum_{b=1}^{m} Q_{ba} f_b, \quad e'_j = \sum_{k=1}^{m} P_{kj} e_k,$$

cf. (1.23). We then determine that

$$[L]_{\mathscr{F}'}^{\mathscr{E}'} = \boldsymbol{P}^{-1} [L]_{\mathscr{F}}^{\mathscr{E}} \boldsymbol{Q}, \tag{1.25}$$

which is the ***change of basis formula*** for matrix representations. We shall have occasion to make use of this formula in the case when $\mathsf{U} = \mathsf{V}$ and $\mathscr{F} = \mathscr{E}$ and $\mathscr{F}' = \mathscr{E}'$. In this case the change of basis formula reads

$$[L]_{\mathscr{E}'}^{\mathscr{E}'} = \boldsymbol{P}^{-1} [L]_{\mathscr{E}}^{\mathscr{E}} \boldsymbol{P} \tag{1.26}$$

for a linear transformation L.

The fact that we can represent a linear transformation by a matrix in a basis means that we can define the determinant of a linear map by using its matrix representative. (We assume the reader is familiar with row or column expansions

for computing determinants.) That is to say, if $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$ for a finite-dimensional vector space V, we define the ***determinant*** of L to be

$$\det L = \det[L]^{\mathscr{E}}_{\mathscr{E}}$$

for some basis $\mathscr{E} = \{e_1, \dots, e_n\}$. The formula (1.26) and familiar properties of determinants (that we assume known) shows that the definition of determinant is independent of basis.

There is another scalar one can associate to a linear map $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$ on a finite-dimensional vector space, and this is as follows. Again, we work with matrix representations. The ***trace*** of an $n \times n$ matrix $\boldsymbol{A}$ is the sum of its diagonal elements:

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{j=1}^{n} A_{jj}.$$

Now, given $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$, we define the ***trace*** of L by

$$\mathrm{tr}(L) = \mathrm{tr}([L]^{\mathscr{E}}_{\mathscr{E}})$$

for some basis $\mathscr{E} = \{e_1, \dots, e_n\}$ for V. As with the determinant, the change of basis formula (1.26) allows one to show that this definition of trace does not depend on the basis $\mathscr{E}$.

1.2.4.4 Affine maps and inhomogeneous linear equations

Closely related to the notion of a linear map is the following.

1.2.3 Definition (Affine map) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let U and V be $\mathbb{F}$-vector spaces. A map $A \colon \mathsf{U} \to \mathsf{V}$ is ***affine*** if it is given by $A(u) = L(u) + v_0$ for $L \in \mathrm{L}(\mathsf{U};\mathsf{V})$ and $v_0 \in \mathsf{V}$. •

Our primary interest in affine maps will come from how they arise in the theory of systems of linear algebraic equations. It will be beneficial to recall in some detail the structure associated with such equations, as this structure repeats itself in the theory of linear differential equations. We thus let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let U and V be $\mathbb{F}$-vector spaces. We let $L \in \mathrm{L}(\mathsf{U};\mathsf{V})$ and let $v_0 \in \mathsf{V}$. A ***linear algebraic equation*** is then the equation

$$L(u) = v_0$$

which is to be solved for $u \in \mathsf{U}$. Of course,

$$L(u) = v_0 \iff L(u) - v_0 = 0,$$

i.e., u solves the linear algebraic equation if and only if it is in the "kernel" of the affine map $A(u) = L(u) - v_0$. We denote by $\mathrm{Sol}(L, v_0) \subseteq \mathsf{U}$ the set of all solutions to this equation. There are various possible scenarios that arise in the attempt to find solutions to this equation, and we outline these in the following proposition.

1.2.4 Proposition (Solutions to linear algebraic equations) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* U *and* V *be* $\mathbb{F}$*-vector spaces. For* $\mathrm{L} \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ *and* $\mathrm{v}_0 \in \mathsf{V}$*, denote by*

$$\mathrm{Sol}(\mathrm{L}, \mathrm{v}_0) = \{\mathrm{u} \in \mathsf{U} \mid \mathrm{L}(\mathrm{u}) = \mathrm{v}_0\}$$

the set of solutions to the linear algebraic equation. Then the following statements hold:

(i) (existence of solutions) $\mathrm{Sol}(\mathrm{L}, \mathrm{v}_0) \neq \emptyset$ *if and only if* $\mathrm{v}_0 \in \mathrm{image}(\mathrm{L})$*;*

(ii) (characterisation of all solutions when one exists) if $\mathrm{v}_0 \in \mathrm{image}(\mathrm{L})$*, let* $\mathrm{u}_0 \in \mathrm{Sol}(\mathrm{L}, \mathrm{v}_0)$*, and then*

$$\mathrm{Sol}(\mathrm{L}, \mathrm{v}_0) = \{\mathrm{u}_0 + \mathrm{u} \mid \mathrm{u} \in \ker(\mathrm{L})\};$$

(iii) (uniqueness of solutions) if $\mathrm{v}_0 \in \mathrm{image}(\mathrm{L})$*, then* $\mathrm{Sol}(\mathrm{L}, \mathrm{v}_0) = \{\mathrm{u}_0\}$ *(i.e., there is only one solution) if and only if* $\ker(\mathrm{L}) = 0$ *(i.e.,* L *is injective).*

Proof (i) This follows by definition of image(L).

(ii) First, since $v_0 \in \mathrm{image}(L)$, there exists $u_0 \in \mathrm{Sol}(L, v_0)$ by part (i).

Now let $u \in \mathrm{Sol}(L, v_0)$ so that $L(u) = v_0$. Then, since $L(u_0) = v_0$ and since L is linear:

$$L(u) - L(u_0) = v_0 - v_0 = 0 \implies L(u - u_0) = 0 \implies u - u_0 \in \ker(L) \implies u = u_0 + \underbrace{(u - u_0)}_{\in \ker(L)}$$

which shows that

$$u \in \{u_0 + u' \mid u' \in \ker(L)\}.$$

Next suppose that

$$u \in \{u_0 + u' \mid u' \in \ker(L)\}.$$

Then, if $u' = u - u_0 \in \ker(L)$, we have

$$L(u) = L(u_0 + u') = L(u_0) + L(u') = L(u_0) = v_0,$$

by linearity of L.

(iii) This follows immediately by part (ii). ■

We suppose that the reader has seen various ways of determining the existence and uniqueness properties of a linear equation using matrix representatives and row reduction, although this is a skill we will not make use of here.

1.2.4.5 Eigenvalues and eigenvectors There are special sorts of invariant subspaces that arise for linear transformations, and these are one-dimensional subspaces on which the transformation acts by multiplication. To be precise, let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ be a linear transformation. An ***eigenvalue*** for L is $\lambda \in \mathbb{F}$ such that $L(v) = \lambda v$ for some nonzero $v \in \mathsf{V}$. We define the ***eigenspace*** associated to an eigenvalue λ by

$$\mathsf{W}(\lambda, L) = \{v \in \mathsf{V} \mid L(v) = \lambda v\}.$$

Nonzero vectors in $\mathsf{W}(\lambda, L)$ are ***eigenvectors*** for λ. The ***geometric multiplicity*** of λ is $\dim_{\mathbb{F}}(\mathsf{W}(\lambda, L))$, and is denote by $m_{\mathrm{g}}(\lambda, L)$.

We suppose that the reader knows that eigenvalues of a linear map are exactly the roots of the ***characteristic polynomial*** which is

$$P_L = \det(X \operatorname{id}_{\mathsf{V}} - L) \in \mathbb{F}[X].$$

This is a monic polynomial of degree $n = \dim(\mathsf{V})$. If λ is an eigenvalue, i.e., a root of P_L, then the ***algebraic multiplicity*** of λ is the multiplicity of λ as a root of P_L, and is denoted by $m_{\mathrm{a}}(\lambda, L)$. We assume it known—or just simply assume it to be—that $m_g(\lambda) \leq m_a(\lambda)$. Matters such as this will be of great concern to us when we discuss systems of linear ordinary differential equations.

1.2.4.6 Internal and external direct sums The notion we discuss in this section is quite simple, but may not be a part of the linear algebra background of a student using this text.

If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if V_j, $j \in \{1, \dots, k\}$, are $\mathbb{F}$-vector spaces, then we can put an $\mathbb{F}$-vector space structure on the product $\mathsf{V}_1 \times \cdots \times \mathsf{V}_k$ in a more or less obvious way:

$$(u_1, \dots, u_k) + (v_1, \dots, v_k) = (u_1 + v_1, \dots, u_k + v_k), \quad a(v_1, \dots, v_k) = (av_1, \dots, av_k),$$

where $(u_1, \dots, u_k), (v_1, \dots, v_k) \in \mathsf{V}$ and $a \in \mathbb{F}$. The resulting $\mathbb{F}$-vector space is called the ***direct sum***, more specifically the ***external direct sum***, of $\mathsf{V}_1, \dots, \mathsf{V}_k$, and is denoted by $\mathsf{V}_1 \oplus \cdots \oplus \mathsf{V}_k$. Note that, as a set, this is simply the product $\mathsf{V}_1 \times \cdots \times \mathsf{V}_k$, but with the prescribed vector space structure.

A similar construction can be made with subspaces of a vector space. Thus we let V be a $\mathbb{F}$-vector space and let $\mathsf{U}_1, \dots, \mathsf{U}_k$ be subspaces of V. We say that V is the ***direct sum***, more specifically the ***internal direct sum***, of $\mathsf{U}_1, \dots, \mathsf{U}_k$ is either of the following two equivalent properties hold:

1. $\mathsf{U}_1 \cap \cdots \cap \mathsf{U}_k = \{0\}$ and, for each $v \in \mathsf{V}$, there exists $u_j \in \mathsf{U}_j$, $j \in \{1, \dots, k\}$, such that
$$v = u_1 + \cdots + u_k;$$
2. for each $v \in \mathsf{V}$, there exist unique $u_j \in \mathsf{U}_j$, $j \in \{1, \dots, k\}$, such that
$$v = u_1 + \cdots + u_k.$$

We shall seldom, perhaps never, distinguish between external and internal direct sums, the intended usage being clear from context.

1.2.4.7 Complexification One of the complications in linear algebra that arises naturally when working with linear ordinary differential equations is the unavoidability of complex numbers, even when all data in the equation are real. In this section we shall see systematically how to handle the need to use complex numbers for $\mathbb{R}$-vector spaces.

We begin with a definition.

1.2.5 Definition (Complexification of a $\mathbb{R}$-vector space) Let V be a $\mathbb{R}$-vector space. The ***complexification*** of V is the set $\mathsf{V} \times \mathsf{V}$ with the following structure as a $\mathbb{C}$-vector space:

(i) ***Vector addition:***
$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2);$$

(ii) ***Scalar multiplication:***
$$(a + \mathrm{i}b) \cdot (u, v) = (au - bv, av + bu);$$

(iii) ***additive inverse:***
$$-(u, v) = (-u, -v);$$

(iv) ***zero vector:***
$$0 = (0, 0).$$

The complexification of V, with the above $\mathbb{C}$-vector space structure, we denote by $\mathsf{V}^{\mathbb{C}}$. •

The only slightly subtle thing here is scalar multiplication, and for this the reader should compare the definition we give to the definition of multiplication of complex numbers above.

The notion of complexification also extends to linear maps.

1.2.6 Definition Suppose that U and V are $\mathbb{R}$-vector spaces and that $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. The ***complexification*** of L is the linear map $L^{\mathbb{C}} \colon \mathsf{V}^{\mathbb{C}} \to \mathsf{V}^{\mathbb{C}}$ defined by
$$L^{\mathbb{C}}(u, v) = (L(u), L(v)).$$ •

1.2.4.8 Multilinear maps Another topic that is probably new to most readers is that of a multilinear map. This is, however, a straightforward generalisation of a linear map.

1.2.7 Definition (Multilinear map, symmetric multilinear map) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let U and V be $\mathbb{F}$-vector spaces. For $k \in \mathbb{Z}_{>0}$, a **k-*multilinear map*** from U to V is a map
$$T \colon \underbrace{\mathsf{U} \times \cdots \times \mathsf{U}}_{k \text{ times}} \to \mathsf{V}$$

with the property that, for each $j \in \{1, \dots, k\}$ and each $u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_k \in \mathsf{U}$, the mapping

$$u \mapsto T(u_1, \dots, u_{j-1}, u, u_{j+1}, \dots, u_k)$$

is linear. A k-multilinear map T is ***symmetric*** if, for each $j_1, j_2 \in \{1, \dots, k\}$ with $j_1 < j_2$, we have

$$T(u_1, \dots, u_{j_1}, \dots, u_{j_2}, \dots, u_k) = T(u_1, \dots, u_{j_2}, \dots, u_{j_1}, \dots, u_k)$$

for every $u_1, \dots, u_k \in \mathsf{U}$.

By $\mathrm{L}^k(\mathsf{U}; \mathsf{V})$ we denote the set of k-multilinear maps from U to V, and by $\mathrm{L}^k_{\mathrm{sym}}(\mathsf{U}; \mathsf{V})$ we denote the set of k-multilinear maps from U to V. •

Thus k-multilinear maps take k arguments, and are linear in each argument if the remaining arguments are fixed. We shall use the special terminology ***bilinear map*** for 2-multilinear maps. Note that a 1-multilinear map is nothing but a linear map.

1.2.5 Calculus

It goes without saying that a basic course in differential and integral calculus is essential background for any study of differential equations. We shall assume readers to be completely familiar with continuous functions, limits, ordinary derivatives, partial derivatives, and integration with respect to a single variable.

What we consider in this section is some particular notation that we shall use. First of all, we wish to talk about derivatives of functions of multiple variables, as we saw in many of our examples in Section 1.1. To talk in a precise way about such derivatives, we need to say a few words about the topology of Euclidean space $\mathbb{R}^n$. First of all, we define the ***Euclidean norm*** by

$$\|\boldsymbol{x}\| = \left(\sum_{j=1}^{n} x_j^2 \right)^{1/2} \tag{1.27}$$

for $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. This gives the usual length of a vector familiar from 1-, 2-, and 3-dimensions to arbitrary numbers of dimensions. We then define the ***open ball*** of radius r at $\boldsymbol{x}_0 \in \mathbb{R}^n$ by

$$\mathsf{B}(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}_0\| < r\}.$$

In the case $n = 1$, the open ball is a line segment (not containing its endpoints) of length $2r$ centred at x_0. In the case of $n = 2$, the open ball is a disk (not containing its boundary) of radius r centred at $\boldsymbol{x}_0$. In the case of $n = 3$, the open ball is a ball in the colloquial sense (not including its boundary) of radius r centred at $\boldsymbol{x}_0$. We then say that a subset $U \subseteq \mathbb{R}^n$ is ***open*** if, for each $\boldsymbol{x} \in U$, there exists $r \in \mathbb{R}_{>0}$ such that $\mathsf{B}(r, \boldsymbol{x}) \subseteq U$. For our purposes, the important attribute of an open set is that one can approach any point $\boldsymbol{x}$ in U from any direction, remaining in U.

Open sets are the natural domain of differentiable functions. We will not carefully provide all the subtleties concerning differentiation of functions of multiple variables, since the subject is one with which students are supposed to have a nodding acquaintance, and this nodding acquaintance is really enough for the material we present in this text. We let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open sets and let $\boldsymbol{f}\colon U \to V$. Note that we can write

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \dots, f_m(\boldsymbol{x})),$$

so a V-valued function can also be thought of as m $\mathbb{R}$-valued functions (with appropriate restrictions so they lie in V). Thus, when we speak of the continuity or differentiability of $\boldsymbol{f}$, we mean continuity or differentiability of each of these m functions.

A function $\boldsymbol{f}\colon U \to V$ is ***continuous*** precisely when each of the functions $f_1, \dots, f_m\colon U \to \mathbb{R}$ are continuous. We shall say that $\boldsymbol{f}$ is of ***class*** $\mathbf{C^k}$, $k \in \mathbb{Z}_{>0}$, if all partial derivatives of $f_1, \dots, f_m$ of degree k exist and are continuous. Note that, if $\boldsymbol{f}$ is of class C^k, then it is **k-*times continuously differentiable***, or of ***class*** $\mathbf{C^l}$ for every $l \in \{1, \dots, k\}$. Let us organise what derivatives *are*, rather than just how to compute them, which is what students likely learned in their past.

We start with the first derivative of $\boldsymbol{f}\colon U \to V$. In this case, being of class C^1 if the partial derivatives

$$\frac{\partial f_a}{\partial x_j}, \qquad a \in \{1, \dots, m\},\ j \in \{1, \dots, n\},$$

exist and are continuous functions of $\boldsymbol{x}$. We organise these derivatives into an $m \times n$ matrix

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

that we denote by $\boldsymbol{Df}$, noting that this is a matrix-valued function of $\boldsymbol{x}$. This is called the ***Jacobian*** of $\boldsymbol{f}$, but we can just think of it as being the ***derivative*** of $\boldsymbol{f}$, since it encodes all partial derivatives. Being an $m \times n$ matrix, we can think of it as being a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e., an element of $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$. In summary, the first derivative is a map

$$\boldsymbol{Df}\colon U \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m).$$

Now what about the second derivative? In this case, $\boldsymbol{f}$ is of class C^2 when the partial derivatives

$$\frac{\partial^2 f_a}{\partial x_j \partial x_k}, \qquad a \in \{1, \dots, m\},\ j, k \in \{1, \dots, n\},$$

exist and are continuous. There are $m \times n^2$ possible partial derivatives, so it is more difficult to arrange them on the page than it was for the derivative. We can nonetheless do so by writing m matrices of second partial derivatives:

$$\begin{bmatrix} \dfrac{\partial^2 f_1}{\partial x_1^2} & \dfrac{\partial^2 f_1}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f_1}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f_1}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f_1}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f_1}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f_1}{\partial x_n \partial x_1} & \dfrac{\partial^2 f_1}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f_1}{\partial x_n^2} \end{bmatrix}, \cdots, \begin{bmatrix} \dfrac{\partial^2 f_m}{\partial x_1^2} & \dfrac{\partial^2 f_m}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f_m}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f_m}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f_m}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f_m}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f_m}{\partial x_n \partial x_1} & \dfrac{\partial^2 f_m}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f_m}{\partial x_n^2} \end{bmatrix}.$$

Conglomerated, these form the ***second derivative*** of $\boldsymbol{f}$. The question is, "While the first derivative is an element of $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$, where does the second derivative live?" The answer is that, just as the first derivative is a *linear* map from $\mathbb{R}^n$ to $\mathbb{R}^m$, the second derivative is to be regarded as a *bilinear map* from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^m$. Specifically, the bilinear map is this one:

$$(\boldsymbol{v}_1, \boldsymbol{v}_2) \mapsto \left(\sum_{j,k=1}^{n} \frac{\partial^2 f_1}{\partial x_j \partial x_k} v_{1,j} v_{2,k}, \ldots, \sum_{j,k=1}^{n} \frac{\partial^2 f_m}{\partial x_j \partial x_k} v_{1,j} v_{2,k} \right),$$

noting that we write

$$\boldsymbol{v}_1 = (v_{1,1}, \ldots, v_{1,n}), \quad \boldsymbol{v}_2 = (v_{2,1}, \ldots, v_{2,n}).$$

Note that for $\boldsymbol{f}$ of class C^2, mixed partial commute:

$$\frac{\partial^2 f_a}{\partial x_j \partial x_k} = \frac{\partial^2 f_a}{\partial x_k \partial x_j}, \qquad a \in \{1, \ldots, m\},\ j, k \in \{1, \ldots, n\}.$$

Thus the second derivative is a *symmetric* bilinear map. We denote the second derivative by $\boldsymbol{D}^2\boldsymbol{f}$, noting that this is a $\mathrm{L}^2_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m)$-valued function of $\boldsymbol{x}$. In summary, the second derivative is a mapping

$$\boldsymbol{D}^2\boldsymbol{f} \colon U \to \mathrm{L}^2_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m).$$

The situation with higher-order derivatives is similar, but the difference is that trying to tabulate these derivatives on the page is difficult, and in any case pointless. Instead, we jump right to the multilinear map version of things. Let us see how this goes. For $k \in \mathbb{Z}_{>0}$, $\boldsymbol{f} \colon U \to V$ is of class C^k if and only all partial derivatives

$$\frac{\partial f_a}{\partial x_{j_1} \cdots \partial x_{j_k}}, \qquad a \in \{1, \ldots, m\},\ j_1, \ldots, j_k \in \{1, \ldots, n\},$$

exist and are continuous. We represent the kth derivative as a k-multilinear map from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ to $\mathbb{R}^m$ defined by

$$(v_1, \dots, v_k) \mapsto \left(\sum_{j_1,\dots,j_k=1}^{n} \frac{\partial f_1}{\partial x_{j_1} \cdots \partial x_{j_k}} v_{1,j_1} \cdots v_{k,j_k}, \dots, \sum_{j_1,\dots,j_k=1}^{n} \frac{\partial f_m}{\partial x_{j_1} \cdots \partial x_{j_k}} v_{1,j_1} \cdots v_{k,j_k} \right).$$

Again, since mixed partials commute, this is a symmetric multilinear map. Thus the kth derivative, which we denote by $\boldsymbol{D}^k \boldsymbol{f}$, is a map

$$\boldsymbol{D}^k \boldsymbol{f} \colon U \to \mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m).$$

Finally, we need some notation for all derivatives of $\boldsymbol{f}$ up to order k. For this we first denote

$$\mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) = \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m) \oplus \mathrm{L}^2_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \oplus \cdots \oplus \mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m).$$

This is the space where all derivatives up to order k of maps $\boldsymbol{f} \colon U \to V$ take their values. Then we define the map

$$\begin{aligned} \boldsymbol{D}^{\leq k} \boldsymbol{f} \colon U &\to V \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \\ \boldsymbol{x} &\mapsto (\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{D}^2 \boldsymbol{f}(\boldsymbol{x}), \dots, \boldsymbol{D}^k \boldsymbol{f}(\boldsymbol{x})). \end{aligned}$$

Thus $\boldsymbol{D}^{\leq k} \boldsymbol{f}$ encodes all partial derivatives of $\boldsymbol{f}$ of all orders (including the zeroth-order) up to k. This is all just organising things that are already known. However, this organisation will be useful in Section 1.3 when we classify the various types of differential equations.

1.2.6 Real analysis

For the most part, a background in real analysis is not required to (1) understand the definitions and the statements of the results in the text or (2) to apply the methods described in the text to solve particular problems. However, it is very often (but not always) the case that proofs of results require some background in analysis. A reader who wishes to understand these proofs should expect to have/acquire this background in a suitable course, or do a substantial amount of independent study. In this short section we merely list the concepts that must be learnt to understand some of the proofs.

1. *Supremum and infimum.* Given a subset $A \subseteq \mathbb{R}$, an ***upper bound*** for A is a number $u \in \mathbb{R}$ such that $x < u$ for every $x \in A$. A ***lower bound*** for A is a number $l \in A$ such that $l < x$ for every $x \in A$. The ***supremum*** of A, denoted $\sup A$, is the least upper bound for A. The ***infimum*** of A, denoted $\inf A$, is the greatest lower bound for A.
2. *Open set.* We defined the notion of an open set on Page 43 above.
3. *Closed set.* A subset $C \subseteq \mathbb{R}^n$ is ***closed*** if its complement is open.

4. *Bounded set.* A subset $B \subseteq \mathbb{R}^n$ is ***bounded*** if there exists $R \in \mathbb{R}_{>0}$ such that $B \subseteq \mathsf{B}(R, \mathbf{0})$.
5. *Compact set.* A subset $K \subseteq \mathbb{R}^n$ is ***compact*** if it is closed and bounded.[7]
6. *Interior, closure, boundary.* Let $A \subseteq \mathbb{R}^n$. A point $\boldsymbol{x} \in \mathbb{R}^n$ is an ***interior point*** for A if there exists $r \in \mathbb{R}_{>0}$ such that $\mathsf{B}(r, \boldsymbol{x}) \subseteq U$. A point $\boldsymbol{x} \in \mathbb{R}^n$ is a ***limit point*** for A if, for any $r \in \mathbb{R}_{>0}$, $\mathsf{B}(r, \boldsymbol{x}) \cap A \neq \emptyset$ and $\mathsf{B}(r, \boldsymbol{x}) \cap (\mathbb{R}^n \setminus A) \neq \emptyset$. A point $\boldsymbol{x} \in \mathbb{R}^n$ is a ***boundary point*** for A if, for any $r \in \mathbb{R}_{>0}$, $\mathsf{B}(r, \boldsymbol{x}) \cap (\mathbb{R}^n \setminus A) \neq \emptyset$. The ***interior*** of A is the set of all interior points for A. The ***closure*** of A is the set of all limit points for A. The ***boundary*** of A is the set of all boundary points for A.
7. *Bounded function.* Let $A \subseteq \mathbb{R}^n$. A function $\boldsymbol{f} \colon A \to \mathbb{R}^m$ is ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{f}(\boldsymbol{x})\| \leq M$ for every $\boldsymbol{x} \in A$.
8. *Continuous function.* Let $A \subseteq \mathbb{R}^n$. A function $\boldsymbol{f} \colon A \to \mathbb{R}^m$ is ***continuous at $\boldsymbol{x}_0 \in \mathbf{A}$*** if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in A$ satisfies $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, then $\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0)\| < \epsilon$. If $\boldsymbol{f}$ is continuous at every $\boldsymbol{x}_0 \in A$, then we say $\boldsymbol{f}$ is ***continuous***.
9. *Convergence of sequences in $\mathbb{R}^n$.* Let $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n$. The sequence ***converges*** to $\boldsymbol{x}_0$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\boldsymbol{x}_j - \boldsymbol{x}_0\| < \epsilon$ for every $j \geq N$.
10. *Pointwise and uniform convergence of sequences of functions.* Let $A \subseteq \mathbb{R}^n$, let $(\boldsymbol{f}_j)A\mathbb{R}^m$ be a sequence of functions, and let $\boldsymbol{f} \colon A \to \mathbb{R}^m$.
    (a) The sequence $(\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ ***converges pointwise*** to $\boldsymbol{f}$ if, for every $\epsilon \in \mathbb{R}_{>0}$ and every $\boldsymbol{x} \in A$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\boldsymbol{f}_j(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})\| < \epsilon$ for $j \geq N$.
    (b) The sequence $(\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ ***converges uniformly*** to $\boldsymbol{f}$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\boldsymbol{f}_j(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})\| < \epsilon$ for $\boldsymbol{x} \in A$ and $j \geq N$.
11. *Results on interchanging operations.* If a sequence $(\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ of functions converges (in some way) to a function $\boldsymbol{f}$, one would like to know what attributes of the functions $\boldsymbol{f}_j$, $j \in \mathbb{Z}_{>0}$, are inherited by $\boldsymbol{f}$. Here are some facts:
    (a) if each of the functions $\boldsymbol{f}_j$, $j \in \mathbb{Z}_{>0}$, is continuous and if $(\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ converges uniformly to $\boldsymbol{f}$, then $\boldsymbol{f}$ is continuous;
    (b) if each of the functions $\boldsymbol{f}_j$, $j \in \mathbb{Z}_{>0}$, is continuously differentiable, if $(\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ converges uniformly to $\boldsymbol{f}$, and if $(\boldsymbol{D}\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ converges uniformly to $\boldsymbol{D}\boldsymbol{f}$, then
    $$\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}) = \lim_{j \to \infty} \boldsymbol{D}\boldsymbol{f}_j(\boldsymbol{x});$$
    (c) if $m = n = 1$, if $A = [a, b]$ is an interval, if the functions f_j, $j \in \mathbb{Z}_{>0}$, are continuous, and if $(f_j)_{j \in \mathbb{Z}_{>0}}$ converges uniformly to f¡ then
    $$\int_a^b f(x)\,\mathrm{d}x = \lim_{j \to \infty} \int_a^b f_j(x)\,\mathrm{d}x.$$

[7] This definition of compactness is particular to $\mathbb{R}^n$; there is an alternative definition of compactness that can and should be used in more general situations. It is equivalent to the one we give here for subsets of $\mathbb{R}^n$, but is generally inequivalent to "closed and bounded."

In all cases, the conditions given are sufficient but not necessary.

Exercises

1.2.1 Let $P \in \mathbb{R}[X]$ have degree 5. List the possible configurations of roots of P, treating differing multiplicities as different cases.
Hint: *You should get 12 cases.*

1.2.2 For the given sets of numbers, find the unique monic polynomial having these as its roots:

(a) $\{-1, 2\}$;
(b) $\{2 + 2\mathrm{i}, 2 - 2\mathrm{i}, -2\}$;
(c) $\{-\frac{1}{\tau}\}$, $\tau \in \mathbb{R} \setminus \{0\}$;
(d) $\{-a, -a, 2\}$, $a \in \mathbb{R}$;
(e) $\{\omega_0(-\zeta + \mathrm{i}\sqrt{1-\zeta^2}), \omega_0(-\zeta - \mathrm{i}\sqrt{1-\zeta^2})\}$, $\omega_0, \zeta \in \mathbb{R}$, $\omega_0 \neq 0$, $|\zeta| \leq 1$;
(f) $\{\sigma + \mathrm{i}\omega, \sigma - \mathrm{i}\omega\}$, $\sigma, \omega \in \mathbb{R}$, $\omega \neq 0$.

1.2.3 Let $L \in \mathrm{L}(\mathbb{R}^2; \mathbb{R}^2)$ be defined by the 2×2 matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Show that L has no eigenvalues.

Section 1.3

Classification of differential equations

In Section 1.1 we saw many examples of differential equations, and there were many different types of differential equations represented in these examples. In this section we provide some procedures for separating differential equations into classes that are special. Such a process cannot be exhaustive, especially at the level which we are able to treat the subject. Nonetheless, the classifications we provide here give important first steps in any classification procedure, and allow us to clearly distinguish the very few differential equations that we can treat in detail by pointing these the special attributes of these equations.

1.3.1 Variables in differential equations

In all of the examples in Section 1.1 we pointed out the independent and dependent variables. In this section we chat about this in a general sort of way.

The independent variables for a differential equation typically reside in an open subset $D \subseteq \mathbb{R}^n$ for some $n \in \mathbb{Z}_{>0}$. These are the variables upon which our objects of interest depend. In the case of $n = 1$, this variable is often thought of as time, although it is also common for this single variable to be a spatial variable.

The dependent variables in a differential equation represent the quantities whose behaviour, as functions of the independent variable, one wishes to understand. We typically regard dependent variables as being in an open subset $U \subseteq \mathbb{R}^m$ for some $m \in \mathbb{Z}_{>0}$. Very often, when one wishes to understand the behaviour of a solution of a differential equation, one plots graphs of the dependent variables as functions of the independent variables. For large numbers of variables, such graphical representations become difficult, and one is forced to think abstractly to understand the behaviour of solutions.

In cases where the number of independent variables is 1, as we mention above this variable typically represents time or space. We shall assume, in general situations, that this variable represents time which we denote by "t." In such cases we represent derivatives of the dependent variables with a dot, e.g., $\dot{x}$ for the first derivative, $\ddot{x}$ for the second derivative, and so on. Thus

$$\dot{x} = \frac{\mathrm{d}x}{\mathrm{d}t}, \quad \ddot{x} = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}.$$

In the case of a single independent variable which is regarded as a spatial variable, we denote this spatial variable by "x." Derivatives of this spatial variable we denote by a prime, e.g., y' is the first derivative and y'' is the second derivative. Thus

$$y' = \frac{\mathrm{d}y}{\mathrm{d}x}, \quad y'' = \frac{\mathrm{d}^2y}{\mathrm{d}x^2}.$$

When there is more than one independent variable, we will not use this notation, and indeed it is faulty to do so; stick to the partial derivative notation in this case. Some commonly encountered notation in this case is to use subscripts to connote the variable with which differentiation is occurring. For example, one sees

$$\frac{\partial^2 u}{\partial x^2} = u_{xx}, \quad \frac{\partial u}{\partial t} = u_t, \quad \frac{\partial^2 u}{\partial x \partial t} = u_{xt}.$$

Note that this notation is *never* to be used when dealing specifically with a single independent variable.

Let us adapt this subscript notation to give a general notation for derivatives. Let $D \subseteq \mathbb{R}^n$ be open and denote coordinates for D by $(x_1, \dots, x_n)$. As we have seen, the kth-order partial derivatives for a function $\boldsymbol{u}\colon D \to U$ are those partial derivatives

$$\frac{\partial u_a}{\partial x_{j_1} \cdots \partial x_{j_k}}, \qquad a \in \{1, \dots, m\},\ j_1, \dots, j_k \in \{1, \dots, n\}.$$

We can use this to motivate notation for coordinates for $\mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m)$. Indeed, we shall use

$$u^a_{j_1 \cdots j_k}, \qquad a \in \{1, \dots, m\},\ j_1, \dots, j_k \in \{1, \dots, n\}, \tag{1.28}$$

for coordinates. Thus a k-multilinear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be denoted by

$$(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) \mapsto \left(\sum_{j_1, \dots, j_k = 1}^{n} u^1_{j_1 \cdots j_k} v_{1, j_1} \cdots v_{k, j_k}, \dots, \sum_{j_1, \dots, j_k = 1}^{n} u^m_{j_1 \cdots j_k} v_{1, j_1} \cdots v_{k, j_k} \right).$$

1.3.2 Differential equations and solutions

In this section we give a *very* general definition of what is meant by a differential equation. While the definition we give is well suited to the objectives of classification in this section, we will not work deeply with this definition outside this section.

First let us give this definition.

1.3.1 Definition (Differential equation) A ***differential equation*** consists of a mapping

$$\boldsymbol{F}\colon D \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}^l,$$

where $k, l, m, n \in \mathbb{Z}_{>0}$, and $D \subseteq \mathbb{R}^n$ and $U \subseteq \mathbb{R}^m$ are open. We also have the following terminology:

(i) n is the number of ***independent variables***;

(ii) m is the number of ***unknowns*** or ***states***;

(iii) k is the ***order***;

(iv) l is the number of ***equations***;

(v) $D \subseteq \mathbb{R}^n$ is the ***domain*** for the differential equation;
(vi) $U \subseteq \mathbb{R}^m$ is the ***state space*** for the differential equation. •

To get an understanding of why the preceding definition might encode the notion of a differential equation, let us define what we mean by a solution to a differential equation.

1.3.2 Definition (Solution to a differential equation) Let

$$\boldsymbol{F} \colon D \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}^l,$$

be a differential equation. A ***solution*** to the differential equation is a function $\boldsymbol{u} \colon D' \to U$ of class C^k defined on an open subset $D' \subseteq D$ such that

$$F(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}), \boldsymbol{D}\boldsymbol{u}(\boldsymbol{x}), \dots, \boldsymbol{D}^k\boldsymbol{u}(\boldsymbol{x})) = \boldsymbol{0}, \qquad \boldsymbol{x} \in D'. \quad \bullet$$

This definitions seem quite abstract at this point, so let us illustrate how this works in all of our examples from Section 1.1. In doing this, we shall use the notation (1.28) to denote coordinates for derivatives. Some of the examples are a little tedious to write out in full detail, so we do not do so. However, we encourage the interested reader to undertake to carry out the procedure we describe for any of their favourite equations that we do not work out. For example, Star Wars nerds will probably *need* to work out how to write Einstein's field equations as a formal differential equation in the sense of Definition 1.3.1.

1.3.3 Examples (Differential equations and solutions)

1. For the mass-spring-damper equation we derived in (1.1), we have $n = 1$, $m = 1$, $l = 1$, and $k = 2$. We take $D = \mathbb{R}$ and $U = \mathbb{R}$ for concreteness. Thus we consider all possible times and vertical displacements in the description of the system; this is something that one generally chooses with the specific instantiation of the problem. We use the coordinate t for independent variable time, y for the unknown vertical displacement. Then we have coordinates y_t and y_{tt} for derivatives. We then have

$$F \colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

defined by

$$F(t, y, y_t, y_{tt}) = m y_{tt} + d y_t + k y + m a_g.$$

A solution to this equation is then a mapping $y \colon \mathbb{T} \to \mathbb{R}$ defined on some interval $\mathbb{T}' \subseteq \mathbb{R}$ that satisfies

$$F\left(t, y(t), \frac{\mathrm{d}y}{\mathrm{d}t}(t), \frac{\mathrm{d}^2 y}{\mathrm{d}t^2}(t)\right) = m \frac{\mathrm{d}^2 y}{\mathrm{d}t^2}(t) + d \frac{\mathrm{d}y}{\mathrm{d}t}(t) + k y(t) + m a_g = 0.$$

This, of course, is exactly the equation (1.1).

2. For the coupled mass-spring-damper equation of (1.2), we have $n = 1$, $m = 2$, $k = 2$, and $l = 2$. We again take $D = \mathbb{R}$ and $U = \mathbb{R}$ for concreteness, and we use t as the independent variable time, x_1 and x_2 as the states, the displacements of the masses, and we denote the coordinates for the derivatives by

$$x_{1,t},\ x_{2,t},\ x_{1,tt},\ x_{2,tt}.$$

The map

$$\boldsymbol{F}\colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}^2; \mathbb{R}^2) \to \mathbb{R}^2$$

for this differential equation is then

$$\boldsymbol{F}(t, x_1, x_2, x_{1,t}, x_{2,t}, x_{1,tt}, x_{2,tt}) = (mx_{1,tt} + 2kx_1 - kx_2, mx_{2,tt} - kx_1 + 2kx_2),$$

and a solution $\boldsymbol{x}\colon \mathbb{T} \to \mathbb{R}^2$ satisfies the equation

$$\begin{aligned}\boldsymbol{F}\left(t, x_1(t), x_2(t), \frac{\mathrm{d}x_1}{\mathrm{d}t}(t), \frac{\mathrm{d}x_2}{\mathrm{d}t}(t), \frac{\mathrm{d}^2x_1}{\mathrm{d}t^2}(t), \frac{\mathrm{d}^2x_2}{\mathrm{d}t^2}(t)\right)\\ = \left(m\frac{\mathrm{d}^2x_1}{\mathrm{d}t^2}(t) + 2kx_1(t) - kx_2(t), m\frac{\mathrm{d}^2x_2}{\mathrm{d}t^2}(t) - kx_1(t) + 2kx_2(t)\right) = (0, 0).\end{aligned}$$

These equations are, of course, simply the equations (1.2) written in a different form. We can unify the two forms of the equations a little more by writing

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}_t, \boldsymbol{x}_{tt}) = \boldsymbol{M}\boldsymbol{x}_{tt} + \boldsymbol{K}\boldsymbol{x},$$

where $\boldsymbol{x}_t = (x_{1,t}, x_{2,t})$ and $\boldsymbol{x}_{tt} = (x_{1,tt}, x_{2,tt})$.

3. For the simple pendulum equation of (1.3), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.1.
4. For Bessel's equation (1.5), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.2.
5. For the equation (1.6) governing the current in a series RLC circuit, we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.3.
6. For the tank equations of (1.7), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.4.
7. For the logistical model (1.8) of a population, we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.5.
8. For the Lotka–Volterra predator prey model of (1.9), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.6.
9. For the Rapoport production and exchange model of (1.10), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.7.

10. The Euler–Lagrange equations of (1.11) have $n = 1$, $m = 1$, $k = 2$, and $l = 1$. We take $D = [x_1, x_2]$ (let's overlook, for the moment, the fact that this D is not open) and $U = \mathbb{R}$, and use x as the independent variable, y as the unknown, and y_x and y_{xx} as variables for the required derivatives. The Lagrangian L is then a function of x, y, and y_x. The differential equation is then prescribed by the map

$$F\colon [x_1, x_2] \times \mathbb{R} \times \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

given by

$$F(x, y, y_x, y_{xx}) = \frac{\partial^2 L}{\partial y_x^2} y_{xx} + \frac{\partial^2 L}{\partial y_x \partial y} y_x - \frac{\partial L}{\partial y}.$$

A solution to these equations is then a function $y\colon [x_1, x_2] \to \mathbb{R}$ satisfying

$$F\left(x, y(x), \frac{\mathrm{d}y}{\mathrm{d}x}(x), \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}(x)\right) = \frac{\partial^2 L}{\partial y_x^2} \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}(x) + \frac{\partial^2 L}{\partial y_x \partial y} \frac{\mathrm{d}y}{\mathrm{d}x}(x) - \frac{\partial L}{\partial y} = 0,$$

which is exactly the Euler–Lagrange equation.

11. In Maxwell's equations (1.12), we have $n = 4$, $m = 10$, $k = 1$, and $l = 1+1+3+3 = 8$. To write the function $\boldsymbol{F}$ defining Maxwell's equations is tedious because of the largish number of variables. For example, if we include all required derivatives, the number of arguments for $\boldsymbol{F}$ in this case is $4 + 10 + 40 = 54$.

12. For the Navier–Stokes equations (1.14), along with the equations of continuity (1.13), we have $n = 4$, $m = 5$, $k = 1$, and $l = 3 + 1 = 4$. In this case, the number of variables is manageable, but the equations themselves are quite lengthy and complicated. Thus we do not go through the details of writing down $\boldsymbol{F}$ in this case.

13. For the heat equation (1.17), we have $n = 2$, $m = 1$, $k = 2$, and $l = 1$. For the domain D, we will suppose that we are working with a rod of length ℓ and that we consider positive times. Thus we take $D = [0, \ell] \times \mathbb{R}_{\geq 0}$ (sweeping under the rug the fact that D is not open). We also take $U = \mathbb{R}$. We denote the independent time/space variables as (x, t), the unknown temperature as u, and the required derivatives are

$$u_x,\ u_t,\ u_{xx},\ u_{xt},\ u_{tt},$$

keeping in mind that $u_{tx} = u_{xt}$ by symmetry of derivatives. The map

$$F\colon [0, \ell] \times \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}^2; \mathbb{R}) \to \mathbb{R}$$

is given by

$$F(x, t, u, u_x, u_t, u_{xx}, u_{xt}, u_{xx}) = u_t - k u_{xx}.$$

A solution is then a function $u\colon [0, \ell] \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ satisfying

$$F\left(x, t, u(x,t), \frac{\partial u}{\partial x}(x,t), \frac{\partial u}{\partial t}(x,t), \frac{\partial^2 u}{\partial x^2}(x,t), \frac{\partial^2 u}{\partial x \partial t}(x,t), \frac{\partial^2 u}{\partial t^2}(x,t)\right)$$
$$= \frac{\partial u}{\partial t}(x,t) - k \frac{\partial^2 u}{\partial x^2}(x,t) = 0,$$

which is just the heat equation, of course.

14. For the wave equation (1.18), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.8.

15. For the potential equation (1.19), we leave the working out of this as a differential equation and the conditions for a solution as Exercise 1.3.9.

16. For the Einstein field equations (1.20), we have $n = 4$, $m = 10$, $k = 2$ (can you work out why?), and $l = 10$. These equations are extremely complicated to write as a differential equation as per Definition 1.3.1, and so we do not do this here. For example, the number of arguments of $\boldsymbol{F}$ in this case would be $4 + 10 + 40 + 100 = 154$!

17. Finally, we consider the Schrödinger equation (1.21). For this equation we have $n = 4$, $m = 2$, $k = 2$, and $l = 2$. Here, for simplicity, we take $D = \mathbb{R}^4$ and $U = \mathbb{C} \simeq \mathbb{R}^2$. We use coordinates (x_1, x_2, x_3, t) the independent variables, (ψ_1, ψ_2) for the unknown real and imaginary parts of the wave function, and the required derivatives are

$$\begin{aligned}&\psi_{1,x_1},\ \psi_{1,x_2},\ \psi_{1,x_3},\ \psi_{1,t},\ \psi_{2,x_1},\ \psi_{2,x_2},\ \psi_{2,x_3},\ \psi_{2,t},\\ &\quad \psi_{1,x_1x_1},\ \psi_{1,x_1x_2},\ \psi_{1,x_1x_3},\ \psi_{1,x_1t},\ \psi_{1,x_2x_2},\ \psi_{1,x_2x_3},\ \psi_{1,x_2t},\ \psi_{1,x_3x_3},\psi_{1,x_3t},\ \psi_{1,tt},\\ &\quad\quad \psi_{2,x_1x_1},\ \psi_{2,x_1x_2},\ \psi_{2,x_1x_3},\ \psi_{2,x_1t},\ \psi_{2,x_2x_2},\ \psi_{2,x_2x_3},\ \psi_{2,x_2t},\ \psi_{2,x_3x_3},\psi_{2,x_3t},\ \psi_{2,tt}.\end{aligned}$$

The map

$$\boldsymbol{F}\colon \mathbb{R}^4 \times \mathbb{R}^2 \times \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}^4; \mathbb{R}^2) \to \mathbb{R}$$

defining the Schrödinger equation is

$$\begin{aligned}&\boldsymbol{F}(x_1, x_2, x_3, t, \psi_1, \psi_2, \psi_{1,x_1},\ \psi_{1,x_2},\ \psi_{1,x_3},\ \psi_{1,t},\ \psi_{2,x_1},\ \psi_{2,x_2},\ \psi_{2,x_3},\ \psi_{2,t},\\ &\quad \psi_{1,x_1x_1},\ \psi_{1,x_1x_2},\ \psi_{1,x_1x_3},\ \psi_{1,x_1t},\ \psi_{1,x_2x_2},\ \psi_{1,x_2x_3},\ \psi_{1,x_2t},\ \psi_{1,x_3x_3},\psi_{1,x_3t},\ \psi_{1,tt},\\ &\quad \psi_{2,x_1x_1},\ \psi_{2,x_1x_2},\ \psi_{2,x_1x_3},\ \psi_{2,x_1t},\ \psi_{2,x_2x_2},\ \psi_{2,x_2x_3},\ \psi_{2,x_2t},\ \psi_{2,x_3x_3},\psi_{2,x_3t},\ \psi_{2,tt})\\ &= (\hbar\psi_{2,t}+\tfrac{\hbar^2}{2\mu}(\psi_{1,x_1x_1}+\psi_{1,x_2x_2}+\psi_{1,x_3,x_3})-V\psi_1, -\hbar\psi_{1,t}+\tfrac{\hbar^2}{2\mu}(\psi_{2,x_1x_1}+\psi_{2,x_2x_2}+\psi_{2,x_3,x_3})-V\psi_2).\end{aligned}$$

A solution is then a map $\boldsymbol{\psi}\colon D' \to \mathbb{R}^2$ defined on some open set $D' \subseteq \mathbb{R}^4$ that satisfies the equation (with the tedious arguments abbreviated)

$$\begin{aligned}\boldsymbol{F}\left(\boldsymbol{x}, t, \boldsymbol{\psi}(\boldsymbol{x}), \frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{x}}, \frac{\partial^2\boldsymbol{\psi}}{\partial\boldsymbol{x}^2}\right) = \Bigg(&\hbar\frac{\partial\psi_2}{\partial t} + \tfrac{\hbar^2}{2\mu}\left(\frac{\partial^2\psi_1}{\partial x_1^2} + \frac{\partial^2\psi_1}{\partial x_2^2} + \frac{\partial^2\psi_1}{\partial x_3^2}\right) - V\psi_1,\\ &-\hbar\frac{\partial\psi_1}{\partial t} + \tfrac{\hbar^2}{2\mu}\left(\frac{\partial^2\psi_2}{\partial x_1^2} + \frac{\partial^2\psi_2}{\partial x_2^2} + \frac{\partial^2\psi_2}{\partial x_3^2}\right) - V\psi_2\Bigg).\end{aligned}$$

One can check that, indeed, these are the Schrödinger equations, broken into their real and imaginary parts. •

Since this is likely to be a student's first encounter with the subject of differential equations, the preceding way of doing things may seem excessively complicated. Indeed, we went through a lot of trouble to just write down equations that were comparatively easy to write down in our modelling exercises of Section 1.1. The benefits of our work will now be seen. Since we know what a differential equation *is* (it is the map $\boldsymbol{F}$), we can speak intelligently about its attributes. And it is this that we now do.

1.3.3 Ordinary differential equations

We begin with a consideration of differential equations with a single independent variable, which we will think of as representing time. The states or unknowns we will represent by $\boldsymbol{x} \in U \subseteq \mathbb{R}^m$. Because of the simplicity of the single independent variable, we can make a more concrete representation for the derivatives. Specifically, we will denote the coordinates for the derivatives up to order k by

$$(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) \in \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^n).$$

Thus $\boldsymbol{x}^{(j)}$ represents the jth derivative with respect to time (this is not uncommon notation, the only difference here is we are thinking of this as being a coordinate rather than an actual derivative).

1.3.4 Remark (Simplification of derivatives with one independent variable) Now, we make a few observations to make things even more concrete:

1. because the domain is 1-dimensional, every multilinear map from $\mathbb{R}$ to $\mathbb{R}^m$ is symmetric;
2. we have a natural isomorphism of the vector spaces $\mathrm{L}^k(\mathbb{R}; \mathbb{R}^m)$ with $\mathbb{R}^m$ by assigning to the k-multilinear map $\boldsymbol{T} \in \mathrm{L}^k(\mathbb{R}; \mathbb{R}^m)$ the element $\boldsymbol{v}_{\boldsymbol{T}} \in \mathbb{R}^m$ given by

$$\boldsymbol{v}_{\boldsymbol{T}} = \boldsymbol{T}(1, \dots, 1).$$

The punchline of the preceding is that we can think of

$$\mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \simeq \mathbb{R}^m \implies \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \simeq \underbrace{\mathbb{R}^m \oplus \cdots \oplus \mathbb{R}^m}_{k+1 \text{ times}}.$$

While we will continue to write things using the notation on the left of these isomorphisms, we shall, when convenient, use the isomorphisms to simplify things. •

1.3.3.1 General ordinary differential equations

With the preceding notation, we have the following definition.

1.3.5 Definition (Ordinary differential equation) An ***ordinary differential equation*** is a differential equation $\boldsymbol{F}$ subject to the following conditions:

(i) there is one independent variable, i.e., $n = 1$;

(ii) the independent variable takes values in an interval $\mathbb{T} \subseteq \mathbb{R}$ called the ***time-domain***;

(iii) the ***state space*** is an open subset $U \subseteq \mathbb{R}^m$;

(iv) there are the same number of equations as states, i.e., $l = m$;

(v) if the order of the differential equation is k, for each

$$(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}) \in \mathbb{R} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m)$$

the equation

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}, \boldsymbol{x}^{(k)}) = \boldsymbol{0}$$

can be uniquely solved to give

$$\boldsymbol{x}^{(k)} = \widehat{\boldsymbol{F}}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}).$$

We call $\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$ the ***right-hand side*** for the ordinary differential equation. •

We can give an alternative characterisation for solutions for ordinary differential equations.

1.3.6 Proposition (Solutions to ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with time-domain* $\mathbb{T}$*, state space* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{m}}$*, and right-hand side* $\widehat{\mathbf{F}}$*. Then the following statements are equivalent for a* C^{k} *map* $\xi \colon \mathbb{T}' \to \mathrm{U}$ *defined on a subinterval* $\mathbb{T}' \subseteq \mathbb{T}$*:*

(i) ξ *is a solution for* $\mathbf{F}$*;*

(ii) ξ *satisfies the equation*

$$\frac{\mathrm{d}^{\mathrm{k}}\xi}{\mathrm{dt}^{\mathrm{k}}}(\mathrm{t}) = \widehat{\mathbf{F}}\left(\mathrm{t}, \xi(\mathrm{t}), \frac{\mathrm{d}\xi}{\mathrm{dt}}(\mathrm{t}), \dots, \frac{\mathrm{d}^{\mathrm{k}-1}\xi}{\mathrm{dt}^{\mathrm{k}}}(\mathrm{t})\right).$$

Proof First suppose that ξ is a solution for $\boldsymbol{F}$. Then

$$\boldsymbol{F}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t)\right) = \boldsymbol{0}.$$

The property (v) of Definition 1.3.5, we immediately have

$$\frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t) = \widehat{\boldsymbol{F}}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^k}(t)\right).$$

Next suppose that ξ satisfies the preceding equation. Fix $t \in \mathbb{T}$ and consider the equation

$$\boldsymbol{F}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^k}(t), \boldsymbol{x}^{(k)}\right) = \boldsymbol{0}.$$

By property (v) of Definition 1.3.5, there exists a unique $\boldsymbol{x}^{(k)} \in \mathbb{R}^m$ that solves this equation and, moreover,

$$\boldsymbol{x}^{(k)} = \widehat{\boldsymbol{F}}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^k}(t)\right).$$

This means, however, that

$$\boldsymbol{x}^{(k)} = \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t).$$

Thus

$$\boldsymbol{F}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^k}(t), \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t)\right) = \boldsymbol{0},$$

i.e., ξ is a solution for $\boldsymbol{F}$. ■

This last condition in Definition 1.3.5 is one that very often arises naturally when looking at specific differential equations. To see how this arises, let us consider the examples of Section 1.1 with one independent variable, and see how their right-hand sides are naturally defined.

1.3.7 Examples (Ordinary differential equations)

1. For the mass-spring-damper equation we derived in (1.1), we can use our ordinary differential equation specific notation to write

$$F(t, y, y^{(1)}, y^{(2)}) = my^{(2)} + dy^{(1)} + ky + ma_g.$$

Note that this is indeed an ordinary differential equation since (1) $n = 1$, (2) $l = m = 1$, and (3) we can solve the equation

$$F(t, y, y^{(1)}, y^{(2)}) = 0$$

for $y^{(2)}$ as

$$y^{(2)} = \tfrac{1}{m}(-dy^{(1)} - ky - ma_g).$$

Thus the right-hand side is

$$\widehat{F}(t, y, y^{(1)}) = \tfrac{1}{m}(-dy^{(1)} - ky - ma_g).$$

As per Proposition 1.3.6, a solution to the differential equation then satisfies

$$\ddot{y}(t) = \tfrac{1}{m}(-d\dot{y}(t) - ky(t) - ma_g),$$

as expected.

2. For the coupled mass-spring-damper equation of (1.2), the differential equation can be conveniently expressed as

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \boldsymbol{M}\boldsymbol{x}^{(2)} + \boldsymbol{K}\boldsymbol{x}.$$

This is an ordinary differential equation since (1) $n = 1$, (2) $l = m = 2$, and (3) we can solve the equation

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \boldsymbol{0}$$

for $\boldsymbol{x}^{(2)}$ as

$$\boldsymbol{x}^{(2)} = -\boldsymbol{M}^{-1}\boldsymbol{K}\boldsymbol{x}.$$

Thus the right-hand side of this ordinary differential equation is

$$\widehat{\boldsymbol{F}}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}) = -\boldsymbol{M}^{-1}\boldsymbol{K}\boldsymbol{x}.$$

As per Proposition 1.3.6, a solution satisfies

$$\ddot{\boldsymbol{x}}(t) = -\boldsymbol{M}^{-1}\boldsymbol{K}\boldsymbol{x}(t),$$

which is simply our original equation, rewritten.

3. For the simple pendulum equation of (1.3), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.10.
4. For Bessel's equation (1.5), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.11.
5. For the current in a series RLC circuit of (1.6), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.12.
6. For the tank flow model of (1.7), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.13.
7. For the logistical model population of (1.8), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.14.
8. For the Lotka–Volterra predator prey model of (1.9), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.15.
9. For the Rapoport production and exchange model of (1.10), we leave the working out of the right-hand side and corresponding conditions for solutions as Exercise 1.3.16.
10. Our final example, that of the Euler–Lagrange equations, shows that one must sometimes take care with what is and is not an ordinary differential equation. We let x denote the single independent variable, y the unknown, and we follow our ordinary differential equation notation and denote derivatives by $y^{(1)}$ and $y^{(2)}$. The Lagrangian is then a function of x, y, and $y^{(1)}$, and the Euler–Lagrange equations are differential equations prescribed by

$$F(x, y, y^{(1)}, y^{(2)}) = \frac{\partial^2 L}{\partial y^{(1)} \partial y^{(1)}} y^{(2)} + \frac{\partial^2 L}{\partial y^{(1)} \partial y} y^{(1)} - \frac{\partial L}{\partial y}.$$

This differential equation is an ordinary differential equation if and only if

$$\frac{\partial^2 L}{\partial y^{(1)} \partial y^{(1)}}$$

is non-zero for every $(x, y, y^{(1)})$. This is true, for example, if

$$L(x, y, y^{(1)}) = (y^{(1)})^2.$$

It is not true, for example, when

$$L(x, y, y^{(1)}) = f(x, y)$$

for any function of (x, y) or when

$$L(x, y, y^{(1)}) = y^{(1)}.$$

Thus we cannot say that the Euler–Lagrange equations are ordinary differential equations, in general, but must examine particular Lagrangians. •

Note that an ordinary differential equation $\boldsymbol{F}$ determines uniquely its right-hand side $\widehat{\boldsymbol{F}}$, but that it is possible that two different ordinary differential equations can give rise to the same right-hand side. To resolve this ambiguity, we make the following definition.

1.3.8 Definition (Normalised ordinary differential equation) An ordinary differential equation

$$\boldsymbol{F} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

with right-hand side

$$\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

is ***normalised*** if

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) = \boldsymbol{x}^{(k)} - \widehat{\boldsymbol{F}}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)})$$

for all

$$(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) \in \mathbb{T} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m).$$ •

If $\boldsymbol{F}$ is an ordinary differential equation that is *not* normalised, we can alway replace it with an ordinary differential equation $\boldsymbol{F}^*$ that *is* normalised, according to the formula

$$\boldsymbol{F}^*(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) = \boldsymbol{x}^{(k)} - \widehat{\boldsymbol{F}}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}).$$

Moreover, by Proposition 1.3.6, $t \mapsto \xi(t)$ is a solution for $\boldsymbol{F}$ if and only if it is a solution for $\boldsymbol{F}^*$. In short, we can without loss of generality assume that an ordinary differential equation is normalised. That being said, we will only rarely make this assumption.

Now that we have defined what we mean, in general terms, by an ordinary differential equation, let us examine certain special kinds of such equations.

We begin with a general and common sort of simplification that can be made with the general definition.

1.3.9 Definition (Autonomous ordinary differential equation) An ordinary differential equation

$$\boldsymbol{F}\colon \mathbb{T} \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

is ***autonomous*** if there exists $\boldsymbol{F}_0\colon U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$ so that

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) = \boldsymbol{F}_0(\boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)})$$

for every $(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) \in \mathbb{T} \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m)$. An ordinary differential equation that is not autonomous is ***nonautonomous***. •

Simply put, an autonomous ordinary differential equation is independent of time.

One can equivalently characterise the notion of autonomous in terms of right-hand sides.

1.3.10 Proposition (Right-hand sides of autonomous ordinary differential equations) *If an ordinary differential equation*

$$\mathbf{F}\colon \mathbb{T} \times \mathrm{U} \times \mathrm{L}^{\le \mathrm{k}}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{m}}$$

with right-hand side

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \times \mathrm{L}^{\le \mathrm{k}-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{m}}$$

is autonomous, then there exists

$$\widehat{\mathbf{F}}_0\colon \mathrm{U} \times \mathrm{L}^{\le \mathrm{k}-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{m}}$$

such that

$$\widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mathrm{k}-1)}) = \widehat{\mathbf{F}}_0(\mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mathrm{k}-1)}).$$

for every $(\mathrm{t}, \mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mathrm{k}-1)}) \in \mathbb{T} \times \mathrm{U} \times \mathrm{L}^{\le \mathrm{k}-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}})$.

Proof Suppose that $\boldsymbol{F}$ is autonomous. Let

$$(\boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}) \in U \times \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m)$$

and let $t_1, t_2 \in \mathbb{T}$. Then there exists a unique $\boldsymbol{x}^{(k)}_1, \boldsymbol{x}^{(k)}_2 \in \mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m)$ such that

$$\boldsymbol{F}(t_a, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}, \boldsymbol{x}^{(k)}_a) = \boldsymbol{0}.$$

Moreover, since $\boldsymbol{F}$ is autonomous, we conclude that $\boldsymbol{x}^{(k)}_1 = \boldsymbol{x}^{(k)}_2$. We also have

$$\boldsymbol{x}^{(k)}_a = \widehat{\boldsymbol{F}}(t_a, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}), \qquad a \in \{1, 2\},$$

and so

$$\widehat{\boldsymbol{F}}(t_1, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}) = \widehat{\boldsymbol{F}}(t_2, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}).$$

Thus $\widehat{\boldsymbol{F}}$ is independent of t, which is the assertion of the proposition. ■

It is easy to see that the converse of the preceding proposition is not generally true. This is because, while a differential equation uniquely determines its right-hand side, a right-hand side does not uniquely determine a differential equation. This is pursued in Exercise 1.3.20.

1.3.3.2 Linear ordinary differential equations Next we turn to a very important class of ordinary differential equations, namely those that are linear.

1.3.11 Definition (Linear ordinary differential equation) Let

$$\boldsymbol{F}\colon \mathbb{T} \times \mathbb{R}^m \oplus \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

be an ordinary differential equation with state space $U = \mathbb{R}^m$. The ordinary differential equation $\boldsymbol{F}$ is:

(i) ***linear*** if, for each $t \in \mathbb{T}$, the map

$$\begin{aligned} \boldsymbol{F}_t\colon \mathbb{R}^m \oplus \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) &\to \mathbb{R}^m \\ (\boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) &\mapsto \boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) \end{aligned}$$

is affine;

(ii) ***linear homogeneous*** if, for each $t \in \mathbb{T}$, the map $\boldsymbol{F}_t$ is linear;

(iii) ***linear inhomogeneous*** if it is linear but not linear homogeneous. •

Before we get to examples, let us characterise linearity in terms of the right-hand side of the ordinary differential equation.

1.3.12 Proposition (Right-hand sides of linear ordinary differential equations) *Let*

$$\mathbf{F}\colon \mathbb{T} \times \mathbb{R}^{\mathrm{m}} \oplus \mathrm{L}^{\leq \mathrm{k}}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{m}}$$

be an ordinary differential equation with right-hand side

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathbb{R}^{\mathrm{m}} \oplus \mathrm{L}^{\leq \mathrm{k}-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{m}}$$

The following statements hold:

(i) if $\mathbf{F}$ *is linear, then, for each* $\mathrm{t} \in \mathbb{T}$*, the map*

$$\begin{aligned} \widehat{\mathbf{F}}_{\mathrm{t}}\colon \mathbb{R}^{\mathrm{m}} \oplus \mathrm{L}^{\leq \mathrm{k}-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^{\mathrm{m}}) &\to \mathbb{R}^{\mathrm{m}} \\ (\mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mathrm{k}-1)}) &\mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mathrm{k}-1)}) \end{aligned}$$

is affine;

(ii) if $\mathbf{F}$ *is linear homogeneous, then, for each* $\mathrm{t} \in \mathbb{T}$*, the map* $\widehat{\mathbf{F}}_{\mathrm{t}}$ *is linear;*

(iii) if $\mathbf{F}$ *is linear inhomogeneous, then, for each* $\mathrm{t} \in \mathbb{T}$*, the map* $\widehat{\mathbf{F}}_{\mathrm{t}}$ *is affine but not linear.*

Proof (i) Fix $t \in \mathbb{T}$. Since $\boldsymbol{F}_t$ is affine, there exists $\boldsymbol{L}_{0,t} \in \mathrm{L}(\mathbb{R}^m; \mathbb{R}^m)$,

$$\boldsymbol{L}_{j,t} \in \mathrm{L}(\mathrm{L}^{j}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m); \mathbb{R}^m), \qquad j \in \{1, \dots, k\},$$

and $\boldsymbol{b}_t \in \mathbb{R}^m$ such that

$$\boldsymbol{F}_t(\boldsymbol{x}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}) = \boldsymbol{L}_{k,t}(\boldsymbol{x}^{(k)}) + \cdots + \boldsymbol{L}_{1,t}(\boldsymbol{x}^{(1)}) + \boldsymbol{L}_{0,t}(\boldsymbol{x}) + \boldsymbol{b}_t. \tag{1.29}$$

Keeping in mind Remark 1.3.4, we have

$$\mathrm{L}^j_{\mathrm{sym}}(\mathbb{R};\mathbb{R}^m) \simeq \mathrm{L}(\mathbb{R}^m;\mathbb{R}^m), \qquad j \in \{1,\dots,m\},$$

and so we can use this identification to think of $x^{(j)}$, $j \in \{1,\dots,m\}$, as being in $\mathbb{R}^m$ and the linear maps $\boldsymbol{L}_{j,t}$ as being elements of $\mathrm{L}(\mathbb{R}^m;\mathbb{R}^m)$. We will denote by $\boldsymbol{A}_{j,t} \in \mathrm{L}(\mathbb{R}^m;\mathbb{R}^m)$ the corresponding linear maps, so equation (1.29) reads

$$\boldsymbol{F}_t(\boldsymbol{x},\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(k)}) = \boldsymbol{A}_{k,t}(\boldsymbol{x}^{(k)}) + \cdots + \boldsymbol{A}_{1,t}(\boldsymbol{x}^{(1)}) + \boldsymbol{A}_{0,t}(\boldsymbol{x}) + \boldsymbol{b}_t.$$

Since $\boldsymbol{F}$ is an ordinary differential equation, $\boldsymbol{A}_{k,t}$ must be invertible, and we must also have

$$\widehat{\boldsymbol{F}}_t(\boldsymbol{x},\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(k-1)}) = -\boldsymbol{A}_{k,t}^{-1} \circ \boldsymbol{A}_{0,t}(\boldsymbol{x}) - \boldsymbol{A}_{k,t}^{-1} \circ \boldsymbol{A}_{1,t}(\boldsymbol{x}^{(1)}) - \cdots - \boldsymbol{A}_{k,t}^{-1} \circ \boldsymbol{A}_{k-1,t}(\boldsymbol{x}^{(k-1)}) - \boldsymbol{A}_{k,t}^{-1}(\boldsymbol{b}_t).$$

This gives the desired conclusion that $\widehat{\boldsymbol{F}}_t$ is affine.

(ii) This follows from the calculations of part (i), but with $\boldsymbol{b}_t = \boldsymbol{0}$.

(iii) This follows from parts (i) and (ii). ■

As with Proposition 1.3.10, the converses to the statements in the preceding result are generally false, and the reader can explore this in Exercise 1.3.21.

The proof of the proposition reveals the form for linear ordinary differential equations, and we reproduce this here outside the proof for emphasis. To wit, a differential equation

$$\boldsymbol{F}\colon \mathbb{T} \times \mathbb{R}^m \oplus \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R};\mathbb{R}^m) \to \mathbb{R}^m$$

is linear if and only if there exist maps

$$\boldsymbol{A}_j\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^m;\mathbb{R}^m), \qquad j \in \{0,1,\dots,k\},$$

and $\boldsymbol{b}\colon \mathbb{T} \to \mathbb{R}^m$ such that

$$\boldsymbol{F}(t,\boldsymbol{x},\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(k)}) = \boldsymbol{A}_k(t)(\boldsymbol{x}^{(k)}) + \cdots + \boldsymbol{A}_1(t)(\boldsymbol{x}^{(1)}) + \boldsymbol{A}_0(t)(\boldsymbol{x}) + \boldsymbol{b}(t). \tag{1.30}$$

The right-hand side is then

$$-\boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_0(t)(\boldsymbol{x}) - \boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_1(t)(\boldsymbol{x}^{(1)}) - \cdots - \boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_{k-1}(t)(\boldsymbol{x}^{(k-1)}) - \boldsymbol{A}_k^{-1}(t)(\boldsymbol{b}(t)).$$

Solutions to this ordinary differential equation are then functions $t \mapsto \boldsymbol{x}(t)$ satisfying

$$\frac{\mathrm{d}^k\boldsymbol{x}}{\mathrm{d}t^k}(t) = -\boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_0(t)(\boldsymbol{x}(t)) - \boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_1(t)\left(\frac{\mathrm{d}^{k-1}\boldsymbol{x}}{\mathrm{d}t}(t)\right) - \dots - \boldsymbol{A}_k^{-1}(t) \circ \boldsymbol{A}_{k-1}(t)\left(\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}(t)\right) - \boldsymbol{A}_k^{-1}(t)(\boldsymbol{b}(t)).$$

We shall study equations like this in great detail subsequently, particularly in the case when the linear maps $\boldsymbol{A}_0, \boldsymbol{A}_1, \dots, \boldsymbol{A}_k$ are independent of t. Indeed, equations like this have a particular name.

1.3.13 Definition (Constant coefficient linear ordinary differential equation) A linear ordinary differential equation given by (1.30) is a ***constant coefficient linear ordinary differential equation*** if the functions $A_0, A_1, \dots, A_k$ are independent of t. •

Let us consider the examples of Section 1.1 in terms of their linearity.

1.3.14 Examples (Linear ordinary differential equations (or not))

1. The mass-spring-damper equation we derived in (1.1) is an autonomous linear constant coefficient inhomogeneous ordinary differential equation. According to the notation of (1.30), we have
$$A_2 = m,\ A_1 = d,\ A_0 = k,\ b = -ma_g.$$
2. The coupled mass-spring-damper equation of (1.2) is an autonomous linear constant coefficient homogeneous ordinary differential equations. According to the notation of (1.30), we have
$$\boldsymbol{A}_2 = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix},\ \boldsymbol{A}_1 = \boldsymbol{0},\ \boldsymbol{A}_0 = \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix},\ \boldsymbol{b} = \boldsymbol{0}.$$
3. For the simple pendulum equation of (1.3), we leave the working out of its attributes as Exercise 1.3.22.
4. For Bessel's equation (1.3), we leave the working out of its attributes as Exercise 1.3.22.
5. For the current in a series RLC circuit of simple pendulum equation of (1.6), we leave the working out of its attributes as Exercise 1.3.22.
6. For the tank flow model of (1.7), we leave the working out of its attributes as Exercise 1.3.22.
7. For the logistical population model of (1.8), we leave the working out of its attributes as Exercise 1.3.22.
8. For the Lotka–Volterra predator prey model of (1.9), we leave the working out of its attributes as Exercise 1.3.22.
9. For the Rapoport production and exchange model of (1.10), we leave the working out of its attributes as Exercise 1.3.22. •

1.3.4 Partial differential equations

In the preceding section we called differential equations with one independent variable, and satisfying a certain nondegeneracy condition, "ordinary differential equations." The other kind of differential equations are what we define next.

To do so, we introduce some useful general notation for the various variables and for the derivative coordinates. Independent variables will be denoted by $\boldsymbol{x}$ and states or unknowns by $\boldsymbol{u}$. Then the list of the coordinates representing the

derivatives up to order k of the dependent variables with respect to the independent variables will be denoted by

$$(\boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k)}) \in U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m).$$

Note that, in the general case when $n > 1$, the simplifications of Remark 1.3.4 do not apply, and each of the derivative variables lives in a different space.

1.3.4.1 General partial differential equations

We begin with the definition.

1.3.15 Definition (Partial differential equation) A ***partial differential equation*** is a differential equation

$$\boldsymbol{F} \colon D \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}^l$$

with the following properties:

(i) $n > 1$;

(ii) there exists $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k-1)}) \in D \times U \times \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m)$ such that the function

$$\boldsymbol{u}^{(k)} \mapsto \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k-1)}, \boldsymbol{u}^{(k)})$$

is not constant. •

The second condition merits explanation. It serves a similar function to the nondegeneracy condition (v) of Definition 1.3.5 for ordinary differential equation. In the case of ordinary differential equations, we wished to be able to solve for the highest-order derivative. For partial differential equations, this is asking too much as it is typically *not* the case that the entire highest-order derivative can be solved for. However, the condition we give is that $\boldsymbol{F}$ should not be everywhere independent of the highest-order derivative. This is a condition that, while technically required for a sensible notion of order for a partial differential equation, is always met in practice.

There is not much to say about general partial differential equations. All of the examples of Section 1.1 that have more than one independent variable are partial differential equations as per Definition 1.3.15. The dichotomy into autonomous and nonautonomous equations is not so interesting for partial differential equations, so we do not give the definition here, although it is possible to do so. We also comment that there is no natural notion of a right-hand side for a partial differential equation as there is for an ordinary differential equation.

Thus we begin our specialisation of partial differential equations with various flavours of linearity.

1.3.4.2 Linear and quasilinear partial differential equations

Let us provide the appropriate definitions of linearity for partial differential equations.

1.3.16 Definition (Linear partial differential equation) Let

$$\boldsymbol{F}\colon D \times \mathbb{R}^m \oplus \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}^l$$

be a partial differential equation with state space $U = \mathbb{R}^m$. The partial differential equation $\boldsymbol{F}$ is:

(i) ***linear*** if, for each $\boldsymbol{x} \in D$, the map

$$\begin{aligned}\boldsymbol{F}_{\boldsymbol{x}}\colon \mathbb{R}^m \oplus \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) &\to \mathbb{R}^l\\ (\boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k)}) &\mapsto \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k)})\end{aligned}$$

is affine;

(ii) ***linear homogeneous*** if, for each $\boldsymbol{x} \in D$, the map $\boldsymbol{F}_{\boldsymbol{x}}$ is linear;

(iii) ***linear inhomogeneous*** if it is linear but not linear homogeneous. •

1.3.17 Definition (Quasilinear partial differential equation) A partial differential equation

$$\boldsymbol{F}\colon D \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}^l$$

is ***quasilinear*** if, for each

$$(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k-1)}) \in D \times U \times \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}^n; \mathbb{R}^m),$$

the map

$$\boldsymbol{u}^{(k)} \mapsto \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(k)})$$

is affine. •

We can immediately deduce from the definitions the following forms for the various flavours of linear and quasilinear partial differential equations.

1.3.18 Proposition (Linear partial differential equations) *Let*

$$\mathbf{F}\colon \mathrm{D} \times \mathbb{R}^{\mathrm{m}} \oplus \mathrm{L}^{\le \mathrm{k}}_{\mathrm{sym}}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}}) \to \mathbb{R}^{\mathrm{l}}$$

be a partial differential equation with state space $\mathrm{U} = \mathbb{R}^{\mathrm{m}}$. *Then the following statements hold:*

(i) $\mathbf{F}$ *is linear if and only if there exist maps*

$$\mathbf{A}_{\mathrm{j}}\colon \mathrm{D} \to \mathrm{L}(\mathrm{L}^{\mathrm{j}}_{\mathrm{sym}}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}}); \mathbb{R}^{\mathrm{l}}), \qquad \mathrm{j} \in \{0, 1, \dots, \mathrm{k}\},$$

and $\mathbf{b}\colon \mathrm{D} \to \mathbb{R}^{\mathrm{l}}$, *with* $\mathbf{A}_{\mathrm{k}}$ *not identically zero, such that*

$$\mathbf{F}(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(\mathrm{k})}) = \mathbf{A}_{\mathrm{k}}(\mathbf{x})(\mathbf{u}^{(\mathrm{k})}) + \cdots + \mathbf{A}_1(\mathbf{x})(\mathbf{u}^{(1)}) + \mathbf{A}_0(\mathbf{x})(\mathbf{u}) + \mathbf{b}(\mathbf{x}); \quad (1.31)$$

(ii) $\mathbf{F}$ *is linear homogeneous if and only if it has the form from part (i) with* $\mathbf{b}(\mathbf{x}) = \mathbf{0}$ *for every* $\mathbf{x} \in \mathrm{D}$;

(iii) $\mathbf{F}$ *is linear inhomogeneous if and only if it has the form from part (i) with* $\mathbf{b}(\mathbf{x}) \neq \mathbf{0}$ *for some* $\mathbf{x} \in \mathrm{D}$.

1.3.19 Proposition (Quasilinear partial differential equations) *A partial differential equation*

$$\mathbf{F}\colon \mathrm{D} \times \mathrm{U} \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}^l$$

is quasilinear if and only if there exist maps

$$\mathbf{A}_1\colon \mathrm{D} \times \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}^n;\mathbb{R}^m) \to \mathrm{L}(\mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}^n;\mathbb{R}^m);\mathbb{R}^l), \quad \mathbf{A}_0\colon \mathrm{D} \times \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}^l,$$

with $\mathbf{A}_1$ *not identically zero, such that*

$$\mathrm{F}(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}) = \mathbf{A}_1(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k-1)})(\mathbf{u}^{(k)}) + \mathbf{A}_0(\mathbf{x}, \mathbf{u}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k-1)}).$$

The notion of having constant coefficients that we encountered for ordinary differential equations also makes sense for partial differential equations.

1.3.20 Definition (Constant coefficient linear partial differential equation) A linear partial differential equation given by (1.31) is a ***constant coefficient linear partial differential equation*** if the functions $A_0, A_1, \dots, A_k$ are constant. •

We leave to the reader in Exercise 1.3.24 the pleasure of classifying the example partial differential equations of Section 1.1.

1.3.4.3 Elliptic, hyperbolic, and parabolic second-order linear partial differential equations Many of the partial differential equations that arise from physics are linear second-order equations with a single unknown, and there are various classifications that can be applied to such equations that bear on the attributes of the solutions to these equations.

Let us write the general form of such a differential equation. In doing so, let us remind ourselves what our derivative notation means in this case. We will deal with derivatives of a single variable of at most second-order, so the first derivative $u^{(1)}$ represents a vector of partial derivatives

$$u^{(1)} = (u_{x_1}, \dots, u_{x_n})$$

and $u^{(2)}$ represents a matrix of partial derivatives

$$u^{(2)} = \begin{bmatrix} u_{x_1x_1} & u_{x_1x_2} & \cdots & u_{x_1x_n} \\ u_{x_2x_1} & u_{x_2x_2} & \cdots & u_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{x_nx_1} & u_{x_nx_2} & \cdots & u_{x_nx_n} \end{bmatrix},$$

keeping in mind that this matrix will be symmetric. With this in mind, a general linear second-order partial differential equation will have the form

$$F(\boldsymbol{x}, u, u^{(1)}, u^{(2)}) = \sum_{j,k=1}^{n} A_{jk}(\boldsymbol{x}) u_{x_jx_k} + \sum_{j=1}^{n} a_j(\boldsymbol{x}) u_{x_j} + b(\boldsymbol{x}) \tag{1.32}$$

for functions

$$\boldsymbol{A}\colon D \to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n),\quad \boldsymbol{a}\colon D \to \mathbb{R}^n,\quad b\colon D \to \mathbb{R}.$$

We can, without loss of generality, suppose that $\boldsymbol{A}(\boldsymbol{x})$ is a symmetric matrix for all $\boldsymbol{x} \in D$.[8] In this case, we know that the eigenvalues of $\boldsymbol{A}$ are real, allowing the following definition.

1.3.21 Definition (Elliptic, hyperbolic, parabolic) Let

$$F\colon D \times \mathbb{R} \oplus \mathrm{L}^{\le 2}_{\mathrm{sym}}(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$$

be a second-order linear partial differential equation, and so given by (1.32). Then $\boldsymbol{F}$ is:

(i) ***elliptic*** at $\boldsymbol{x} \in D$ if all eigenvalues of $\boldsymbol{A}(\boldsymbol{x})$ are positive;

(ii) ***hyperbolic*** at $\boldsymbol{x} \in D$ if all eigenvalues of $\boldsymbol{A}(\boldsymbol{x})$ are nonzero;

(iii) ***parabolic*** at $\boldsymbol{x} \in D$ if all eigenvalues of $\boldsymbol{A}(\boldsymbol{x})$ are nonnegative, and at least one of them is zero. •

Note that if $\boldsymbol{F}$ has constant coefficients, then the notion of being in one of the three cases of elliptic, hyperbolic, or parabolic does not depend on $\boldsymbol{x} \in D$. Generally, however, it will. Thus the notions are most frequently applied in the constant coefficient case. Let us consider examples that we have seen thus far, and see where they sit relative to the elliptic/hyperbolic/parabolic classification.

1.3.22 Examples (Elliptic, hyperbolic, and parabolic partial differential equations)

1. The standard example of an elliptic partial differential equation is the ***potential equation***, or ***Laplace's equation***. The domain $D \subseteq \mathbb{R}^n$ is normally thought of as

[8] Indeed, suppose that $\boldsymbol{A}$ is not symmetric. Then write $\boldsymbol{A}$ as a sum of a symmetric and skew-symmetric matrix:

$$\boldsymbol{A} = \underbrace{\tfrac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T)}_{\boldsymbol{A}^+} + \underbrace{\tfrac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)}_{\boldsymbol{A}^-},$$

with $\boldsymbol{A}^+$ being symmetric and $\boldsymbol{A}^-$ being skew-symmetric. Then we have

$$\sum_{j,k=1}^{n} A^-_{jk} u_{x_j x_k} = -\sum_{j,k=1}^{n} A^-_{kj} u_{x_j x_k} = -\sum_{j,k=1}^{n} A^-_{kj} u_{x_k x_j} = -\sum_{j,k=1}^{n} A^-_{jk} u_{x_j x_k},$$

and so we conclude that

$$\sum_{j,k=1}^{n} A^-_{jk} u_{x_j x_k} = 0,$$

and so

$$\sum_{j,k=1}^{n} A_{jk} u_{x_j x_k} = \sum_{j,k=1}^{n} A^+_{jk} u_{x_j x_k},$$

giving our claim that we can assume that $\boldsymbol{A}$ is symmetric.

being "space" in this case, so we denote coordinates for D by $(x_1, \dots, x_n)$. Then the differential equation is given by

$$\boldsymbol{F}(\boldsymbol{x}, u, u^{(1)}, u^{(2)}) = u_{x_1x_1} + \cdots + u_{x_nx_n}.$$

Thus $u\colon D' \to \mathbb{R}$ is a solution if it satisfies

$$\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} = 0.$$

We saw examples of how this equation arises in applications in Section 1.1.13. Note that, in this case,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

so all eigenvalues are 1, i.e., are positive. This ensures that $\boldsymbol{F}$ in this case is indeed elliptic.

2. The standard example of an hyperbolic partial differential equation is the ***wave equation***. In this case, the domain D is normally thought of as encoding time and space, and so we denote coordinates by $(x_1, \dots, x_n, t)$. The differential equation is given by

$$F((t, \boldsymbol{x}), u, u^{(1)}, u^{(2)}) = -u_{tt} + u_{x_1x_1} + \cdots + u_{x_nx_n}.$$

Solutions u thus satisfy the equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}.$$

We saw that in Section 1.1.12 that the wave equation arises in the model of the transverse vibrations of a taut string. In this case we have

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

and so the eigenvalues are $-1, 1, \dots, 1$, showing that this is indeed an hyperbolic equation.

3. The usual example of a parabolic equation is the ***heat equation***, which we saw modelled the temperature distribution in a rod in Section 1.1.11. In this case, like the wave equation, the domain D is coordinatised by time and space: $(1, x_1, \dots, x_n)$. The differential equation is

$$F((t, \boldsymbol{x}), u, u^{(1)}, u^{(2)}) = -u_t + u_{x_1x_1} + \cdots + u_{x_2x_2}.$$

Solutions $u\colon D' \to \mathbb{R}$ satisfy

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}.$$

In this case

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

and so the eigenvalues are $0, 1, \ldots, 1$, showing that this is indeed a parabolic equation. •

In the text we shall come back to these three equations, and when we are able to solve them shall comment on their general characteristics.

1.3.5 How to think about differential equations

A reader having read and understood the content of this section will have an excellent understanding of what a differential equation is, and some of the special classes of differential equations. The reader will now embark on actually *solving* some differential equations (after a brief diversion in Section 1.4 on the important matter of existence and uniqueness of solutions). Before doing so, it is worth putting this process of solving differential equations into a general context.

First of all, let us state very clearly: *if you reach into the bag of differential equations and pull one out, it is extremely unlikely you will be able to solve it*. This is rather like what a student has already encountered in their study of differentiation and integration; one has at hand a small but important collection of functions that one can actually differentiate or integrate, and these are to be regarded as isolated and valuable gems. But this does raise the question of what one can *do* with a differential equation pulled at random from the bag of differential equations.

Let us explore this a little.

1. *Analysis:* Even if one cannot explicitly solve a given differential equation, there are still sometimes things that can be done to get some insight into its behaviour. Let us consider some of the things one might try to do.
   (a) *Understand steady-state behaviour:* In some equations one has time t as the, or one of the, independent variables. In such cases, it is often of interest to understand the behaviour of solutions as $t \to \infty$. This behaviour is known as ***steady-state*** behaviour. Sometimes the steady-state behaviour is not interesting, as in "blows up to infinity." But sometimes this behaviour is all one really wants, and sometimes it can even be determined. We shall see some instances of this sort of investigation in the text.

(b) *Approximating solutions:* Sometimes in a differential equation there are effects that are dominant, and the remaining effects can be regarded as "perturbations" of these dominant effects. If the dominant part of the equations are something that one can understand, one can hope (pray, really) that the perturbations do not materially affect the dominant behaviour. In practice, methods like this should be used with great care, since the "perturbations," while small, may have significant impact on the character of solutions, particularly for long times in cases where time is one of the independent variables. However, there are cases where "perturbation theory" can be applied to give useful conclusions. However, this is not something we will get deeply into in any sort of general way.

(c) *Equilibria and their stability:* A special case of the preceding idea of approximation involves the study of equilibria. This is most easily discussed by reference to ordinary differential equations, but the basic ideas can be adapted by a flexible mind to partial differential equations. Suppose that we have an ordinary differential equation

$$\boldsymbol{F}\colon \mathbb{T} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m.$$

An ***equilibrium*** is a point $\boldsymbol{x}_0 \in U$ for which

$$\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{0}, \dots, \boldsymbol{0}) = \boldsymbol{0}, \qquad t \in \mathbb{T}.$$

Note that the constant function $t \mapsto \boldsymbol{x}_0$ is then a solution of this differential equation. The fact that it is constant is what leads to its being called an "equilibrium." One can then consider the ***stability*** of this equilibrium, which loosely means the matter of whether solutions starting near $\boldsymbol{x}_0$ (i) remain near $\boldsymbol{x}_0$, (ii) approach $\boldsymbol{x}_0$ as $t \to \infty$, or (iii) diverge away from $\boldsymbol{x}_0$. We shall be precise about this in the text in various situations.

2. *Numerical solution:* One can attempt to use a computer to solve the differential equation. For most ordinary differential equations, there are reliable methods for solving them numerically. The situation with partial differential equations is quite different, and significant science has been, is, and will be dedicated to numerical techniques for solving partial differential equations. In the text we will talk a little about using numerical methods to solve ordinary differential equations, and will give the reader some opportunity to use the standard package MATLAB® for plotting numerical solutions to differential equations.

A matter related to what one can *do* with a differential is the manner in which one can think of a solution, since it is solutions in which we are interested. No matter what else you do, here is how you should *not* think about solutions:

> ***Be a grown up about what a solution is:*** *A solution to a differential equation, or any equation for that matter, is not a formula that you write on the page as the byproduct of some algorithmic procedure. This way of*

thinking about "solution" should remain in high school, which is where it was unfortunately taught to you.

So. . . how *should* you think about what a solution is?

For ordinary differential equations, a profitable way to think about it is to think about curves, since a solution is indeed a curve $t \mapsto \boldsymbol{x}(t)$. Let us focus on first-order ordinary differential equations.[9] In this case, $\dot{\boldsymbol{x}}(t)$ is the tangent vector to this curve, and so the equation

$$\dot{\boldsymbol{x}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{x}(t))$$

should be thought of as prescribing the tangent vectors to solution curves. What becomes important, then is the vector $\widehat{\boldsymbol{F}}(t, \boldsymbol{x})$ one assigns to the point $(t, \boldsymbol{x})$.

Let us be explicit about this in an example.

1.3.23 Example (Differential equations and vector fields) We consider the autonomous first-order ordinary differential equation in two unknowns defined by

$$\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = \widehat{\boldsymbol{F}}_0(x_1, x_2) = (x_2, -x_1 + \tfrac{1}{2}x_2(1 - x_1^2)).$$

Thus solutions are defined by the equations

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= -x_1(t) + \tfrac{1}{2}x_2(t)(1 - x_1(t)^2). \end{aligned}$$

In Figure 1.14 we plot the vector field. Thus, at each point $(x_1, x_2) \in \mathbb{R}^2$ we draw an arrow in the direction of

$$\boldsymbol{F}_0(x_1, x_2) = (x_2, -x_1 + \tfrac{1}{2}x_2(1 - x_1^2)).$$

A solution to the differential equation will then be a curve $t \mapsto (x_1(t), x_2(t))$ whose tangent vector at $(x_1(t), x_2(t))$ points in the direction of $\boldsymbol{F}_0(x_1(t), x_2(t))$. In Figure 1.15 we show a few such solution curves; these are known in the business as ***integral curves***.

It is also not uncommon to look at plots of $x_1(t)$ and $x_2(t)$ as functions of t. In Figure 1.16 we show such plots starting at a fixed point $(x_1(0), x_2(0))$ at $t = 0$.[10]

We hope that a reader will find looking at pictures like this, particularly Figure 1.15, more insightful than looking at some formula for the solution, produced as a byproduct of some algorithmic procedure. Also, for this equation, there is no algorithmic procedure for determining the solutions. . . but the pictures can still be produced and offer insight. •

[9] We shall see that a kth-order ordinary differential equation can always be converted into a first-order ordinary differential equation, so the assumption of the equation being first-order is made without loss of generality.

[10] As one varies $(x_1(0), x_2(0))$, one also varies these plots, and this is something we will consider in Section 1.4.

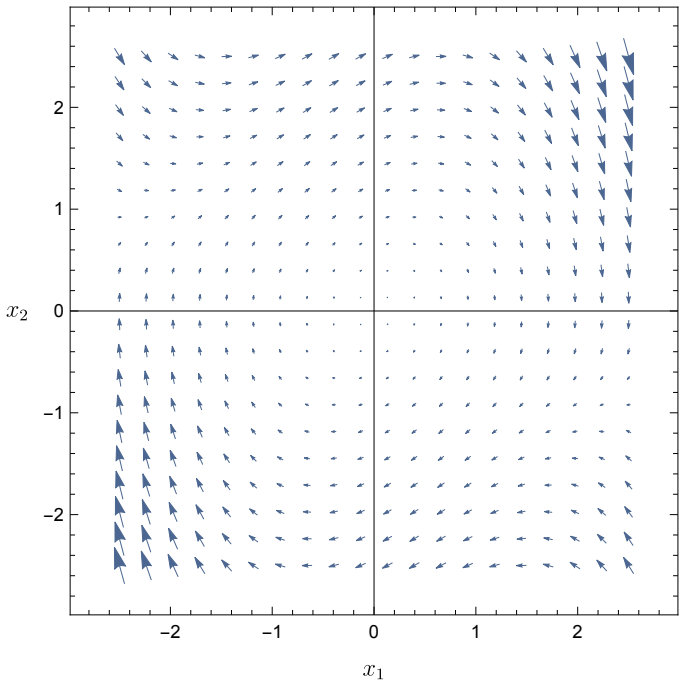


Figure 1.14 A vector field in $\mathbb{R}^2$

For partial differential equations, solutions are no longer curves, i.e., vector functions of a single independent variable, but it is still worthwhile to think about, and represent where possible, a solution as a graph of a function of the independent variables.

Exercises

1.3.1 Work out Example 1.3.3–3. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.2 Work out Example 1.3.3–4. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

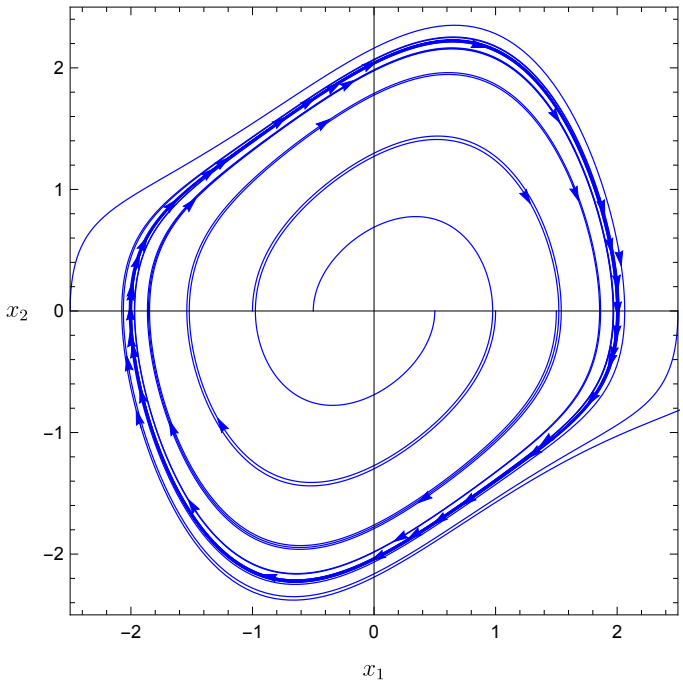


Figure 1.15 A few solution curves for the vector field of Figure 1.14

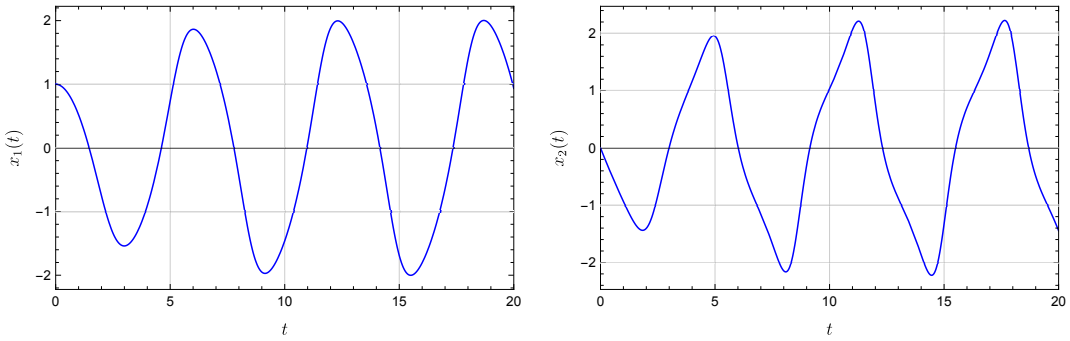


Figure 1.16 Plots of the solutions as functions of time

1.3.3 Work out Example 1.3.3–5. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.4 Work out Example 1.3.3–6. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.5 Work out Example 1.3.3–7. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.6 Work out Example 1.3.3–8. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.7 Work out Example 1.3.3–9. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.8 Work out Example 1.3.3–14. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.9 Work out Example 1.3.3–15. Thus:

(a) identify n, m, k, and l;
(b) name the independent variables;
(c) name the states;
(d) write $\boldsymbol{F}$ as a map, explicitly denoting its domain and codomain;
(e) write the equation that must be satisfied by a solution.

1.3.10 Work out Example 1.3.7–3. Thus:

(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;

(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.11 Work out Example 1.3.7–4. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.12 Work out Example 1.3.7–5. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.13 Work out Example 1.3.7–6. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.14 Work out Example 1.3.7–7. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.15 Work out Example 1.3.7–8. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.16 Work out Example 1.3.7–9. Thus:
(a) write $\boldsymbol{F}$ using the ordinary differential equation notation for derivatives;
(b) show that $\boldsymbol{F}$ is an ordinary differential equation;
(c) write down the right-hand side;
(d) write the condition for a solution using Proposition 1.3.6.

1.3.17 For each of the following ordinary differential equations $\boldsymbol{F}$, determine their right-hand sides:
(a) $F(t, x, x^{(1)}, x^{(2)}) = 3(1 + t^2)x^{(2)}$;
(b) $\boldsymbol{F}(t, (x_1, x_2), (x_1^{(1)}, x_2^{(1)})) = (x_2^{(1)} + 2x_1 - x_2, -x_1^{(1)} - x_1^2)$;
(c) $F(t, x, x^{(1)}, x^{(2)}, x^{(3)}) = -x^{(3)} + t(x^{(1)})^2 + \sin(x)$;
(d) $\boldsymbol{F}(t, (x_1, x_2), (x_1^{(1)}, x_2^{(1)})) = (-x_1^{(1)} + x_2^{(1)} + x_1^2 - x_2, 2x_1^{(1)} + 2x_2^{(1)} + \cos(x_2) - x_1)$;

(e) $F(t, x, x^{(1)}) = x^{(1)} + a(t)x$.

1.3.18 For each of the following right-hand sides $\widehat{F}$, determine the associated normalised ordinary differential equation $\boldsymbol{F}$:

(a) $\widehat{F}(t, x, x^{(1)}) = 0$;

(b) $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (-x_1^2, -2x_2 + x_2)$;

(c) $\widehat{F}(t, x, x^{(1)}, x^{(2)}) = t(x^{(1)})^2 + \sin(x)$;

(d) $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (\frac{1}{4}(x_1 + 2x_1^2 - 2x_2 - \cos(x_2)), \frac{1}{4}(x_1 - 2x_1^2 + 2x_2 - \cos(x_2)))$;

(e) $\widehat{F}(t, x) = -a(t)x$.

In the next exercise we shall show how autonomous ordinary differential equations are special in terms of their solutions. In order for the exercise to make sense, we require the existence and uniqueness theorem we state below, Theorem 1.4.8.

1.3.19 Let

$$\boldsymbol{F} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

be an autonomous ordinary differential equation satisfying the conditions of Theorem 1.4.8(ii), let

$$(\boldsymbol{x}_0, \boldsymbol{x}_0^{(1)}, \dots, \boldsymbol{x}_0^{(k-1)}) \in U \times \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m),$$

and let $t_1, t_2 \in \mathbb{T}$. Let $\xi_1 \colon \mathbb{T} \to U$ and $\xi_2 \colon \mathbb{T} \to U$ be solutions for $\boldsymbol{F}$ satisfying

$$\xi_1(t_1) = \xi_2(t_2) = \boldsymbol{x}_0, \quad \frac{\mathrm{d}^j \xi_1}{\mathrm{d}t^j}(t_1) = \frac{\mathrm{d}^j \xi_2}{\mathrm{d}t^j}(t_2) = \boldsymbol{x}_0^{(j)}, \qquad j \in \{1, \dots, k-1\}.$$

Answer the following questions.

(a) Show that $\xi_2(t) = \xi_1(t + t_1 - t_2)$ for all $t \in \mathbb{T}$ for which $\xi(t)$ is defined and for which $t + t_1 - t_2 \in \mathbb{T}$.

(b) Assuming that $\mathbb{T} = \mathbb{R}$ and that all solutions are defined for all time for simplicity, express your conclusion from part (a) as a condition on the flow $\Phi^{\boldsymbol{F}}$.

1.3.20 Let us consider the following two differential equations:

$$F_1 \colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\leq 1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R} \qquad F_2 \colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\leq 1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$
$$(t, x, x^{(1)}) \mapsto x^{(1)}, \qquad (t, x, x^{(1)}) \mapsto (1 + t^2)x^{(1)}.$$

Answer the following questions.

(a) Show that both F_1 and F_1 are ordinary differential equations, and determine the right-hand sides $\widehat{F}_1$ and $\widehat{F}_2$.

(b) Show that both $\widehat{F}_1$ and $\widehat{F}_2$ are independent of t.

(c) Which of F_1 and F_2 is autonomous?

1.3.21 Let us consider the following two differential equations:

$$F_1 \colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\le 1}_{\mathrm{sym}}(\mathbb{R};\mathbb{R}) \to \mathbb{R} \qquad (t, x, x^{(1)}) \mapsto x^{(1)},$$

$$F_2 \colon \mathbb{R} \times \mathbb{R} \times \mathrm{L}^{\le 1}_{\mathrm{sym}}(\mathbb{R};\mathbb{R}) \to \mathbb{R} \qquad (t, x, x^{(1)}) \mapsto (1 + x^2)x^{(1)}.$$

Answer the following questions.

(a) Show that both F_1 and F_1 are ordinary differential equations, and determine the right-hand sides $\widehat{F_1}$ and $\widehat{F_2}$.

(b) Show that both $\widehat{F_1}$ and $\widehat{F_2}$ are linear.

(c) Which of F_1 and F_2 is linear?

1.3.22 Consider the ordinary differential equations of Examples 1.3.3–3 to 9.

(a) Which of the equations is autonomous?

(b) Which of the equations is linear?

(c) Which of the equations is linear and homogeneous?

(d) Which of the equations is linear and inhomogeneous?

(e) Which of the equations is a linear constant coefficient equation?

1.3.23 Let

$$\boldsymbol{F} \colon \mathbb{T} \times U \times \mathrm{L}^{\le k}_{\mathrm{sym}}(\mathbb{R};\mathbb{R}^n) \to \mathbb{R}^m$$

be an ordinary differential equation with right-hand side $\widehat{\boldsymbol{F}}$. As usual, let t be the independent variable and $\boldsymbol{x}$ the state, with $\boldsymbol{x}^{(j)} \in \mathrm{L}^j_{\mathrm{sym}}(\mathbb{R};\mathbb{R}^m)$ being the coordinate for the jth derivative. As per Remark 1.3.4, we can think of $\boldsymbol{x}^{(j)}$ as being an element of $\mathbb{R}^m$.

We will associate to $\boldsymbol{F}$ a first-order ordinary differential equation $\boldsymbol{F}_1$ with time domain $\mathbb{T}$ and state space

$$U_1 = U \times \underbrace{\mathbb{R}^m \times \cdots \times \mathbb{R}^m}_{k-1 \text{ times}}.$$

To do so, answer the following questions.

(a) Denote coordinates for the state space U_1 by $\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_{k-1}$, and relate these to $(\boldsymbol{x}, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k-1)})$ by

$$\boldsymbol{y}_0 = \boldsymbol{x}, \ \boldsymbol{y}_j = \boldsymbol{x}^{(j)}, \qquad j \in \{1, \ldots, k-1\}.$$

If $t \mapsto \boldsymbol{x}(t)$ is a solution for $\boldsymbol{F}$, write down the corresponding differential equations that must be satisfied by $(\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1})$.
Hint: *For each* $\mathrm{j} \in \{0, 1, \ldots, \mathrm{k}-1\}$, *write down* $\dot{\mathbf{y}}_{\mathrm{j}}(\mathrm{t})$, *and express the result in terms of the coordinates for* U_1.

(b) What is the right-hand side $\widehat{\boldsymbol{F}_1}$ corresponding to the equations you derived in part (a)?

(c) Write down a first-order ordinary differential equation $\boldsymbol{F}_1$ with time domain $\mathbb{T}$ and state space U_1 whose right-hand side is the function $\widehat{\boldsymbol{F}}_1$ you determined in part (b).

(d) State *precisely* the relationship between solutions for $\boldsymbol{F}$ and solutions for $\boldsymbol{F}_1$, and show that if solutions for $\boldsymbol{F}_1$ are of class C^1, then solutions for $\boldsymbol{F}$ are of class C^k.

(e) Show that $\boldsymbol{F}_1$ is can be taken to be linear if $\boldsymbol{F}$ is linear, and show that $\boldsymbol{F}_1$ is homogeneous if and only if $\boldsymbol{F}$ is, in this case.

1.3.24 For the partial differential equations of Examples 1.3.3–11 to 17, determine whether they are (a) linear homogeneous, (b) linear inhomogeneous, (c) quasilinear, and/or (d) has constant coefficients.

The next exercise concerns itself with the so-called method of characteristics for simple second-order linear partial differential equations. Although the presentation is for a simple class of equations, the language and methodology we introduce is readily generalised. The class of differential equations we consider are given by

$$\begin{aligned} F\colon D \times \mathbb{R} \oplus \mathrm{L}^{\leq 2}_{\mathrm{sym}}(\mathbb{R}^2; \mathbb{R}) &\to \mathbb{R} \\ (x, y, u, u^{(1)}, u^{(2)}) &\mapsto a u_{xx} + 2b u_{x,y} + d u_{yy} + d u_x + e u_y + f u \end{aligned} \tag{1.33}$$

for functions $a, b, c, d, e, f, g\colon D \to \mathbb{R}$ defined on an open subset D of $\mathbb{R}^2$. The ***symbol*** for the equation is the $\mathbb{C}$-valued function

$$\begin{aligned} \sigma(F)\colon D \times \mathbb{R}^2 &\to \mathbb{C} \\ (x, y, \xi, \eta) &\mapsto -a\xi^2 - 2b\xi\eta - c\eta^2 + \mathrm{i}d\xi + \mathrm{i}e\eta + f, \end{aligned}$$

defined by "substituting" $\mathrm{i}\xi$ for $\frac{\partial u}{\partial x}$ and $\mathrm{i}\eta$ for $\frac{\partial u}{\partial y}$. The ***principal symbol*** $\sigma_0(F)$ is the quadratic part of the symbol

$$\sigma_0(F)(x, y, \xi, \eta) = -a\xi^2 - 2b\xi\eta - c\eta^2.$$

Consider a curve in D defined by $\phi(x, y) = 0$. The curve is a ***characteristic*** if

$$\sigma_0(F)\left(x, y, \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}\right) = 0.$$

It turns out that it is possible for a solution of a partial differential equation to have points of discontinuity, but one may determine that these are necessarily located along characteristic curves.

The above development outlines why the symmetric matrix

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is useful in determining some properties of a partial differential equation of the form (1.33).

1.3.25 In the preceding, suppose that a, b, and c are constant, and define the function $f_{a,b,c}\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f_{a,b,c}(\xi, \eta) = a\xi^2 + 2b\xi\eta + c\eta^2,$$

and answer the following questions.

(a) Show that when $b^2 - ac = 0$ the following statements hold:
   (a) the curve $f_{a,b,c}(x, y) = 1$ is a parabola for $a > 0$;
   (b) through each point in $\mathbb{R}^2$ there passes a single characteristic for (1.33).
   Show that the heat equation falls into this category.

(c) Show that when $b^2 - ac > 0$ the following statements hold:
   (a) the curve $f_{a,b,c}(x, y) = 1$ is an hyperbola for $a > 0$;
   (b) through each point in $\mathbb{R}^2$ there passes two characteristics for (1.33).
   Show that the wave equation falls into this category.

(c) Show that when $b^2 - ac < 0$ the following statements hold:
   (a) the curve $f_{a,b,c}(x, y) = 1$ is an ellipse for $a > 0$;
   (b) the differential equation (1.33) possesses no characteristics curves.
   Show that the potential equation falls into this category.

Section 1.4

The question of existence and uniqueness of solutions

This chapter, up to this point, has been a bit chatty, with some nice examples, some arcane definitions and some quite obvious results. In this section we produce a few important results, especially for ordinary differential equations. The results are concerned with two important questions: (1) does a given differential equation possess solutions; (2) how many solutions does a differential equation possess? In mathematics, questions like this are known as questions of "existence and uniqueness" (think about similar sorts of questions for linear algebraic equations, as discussed in Section 1.2.4.)

1.4.1 Existence and uniqueness of solutions for ordinary differential equations

We begin our discussion by looking at the situation for ordinary differential equations, where a fairly complete story can be told. We shall begin by framing the sort of questions and answers we might expect by looking at some examples. Then we state the principal existence and uniqueness theorems for solutions of ordinary differential equations. We close the section by considering how all solutions of an ordinary differential equation "fit together."

1.4.1.1 Examples motivating existence and uniqueness of solutions for ordinary differential equations Our first three examples make use of the fact that, when a differential has a right-hand side that is independent of the unknown, then solutions are obtained by integration.

1.4.1 Example (An ordinary differential equation with no solutions (sometimes)) We consider the scalar nonautonomous first-order differential equation with time-domain $\mathbb{R}$ and with right-hand side

$$\widehat{F}(t,x) = \begin{cases} t^{-1}, & t \neq 0, \\ 0, & t = 0. \end{cases}$$

A solution to this differential equation satisfies

$$\dot{x}(t) = f(t),$$

where

$$f(t) = \begin{cases} t^{-1}, & t \neq 0, \\ 0, & t = 0. \end{cases}$$

Since we ask that a solution be a C^1-function, the Fundamental Theorem of Calculus gives that a solution should satisfy

$$x(t) = x(t_0) + \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau.$$

We claim that, if $t_0 t \leq 0$ and if $t \neq t_0$, then the integral does not exist. Indeed, if $t_0 t \leq 0$, then one of the following four instances must hold: (1) $t = 0$; (2) $t_0 = 0$; (3) $t < 0 < t_0$; (4) $t_0 < 0 < t$. In all four of these instances, the integral will not exist since the function $f(t) = t^{-1}$ is not integrable about 0. Thus this differential equation only can be solved when t and t_0 are both on the same side of 0. •

1.4.2 Example (An ordinary differential equation with no solutions (all the time)) This example is beyond the abilities of a typical student taking a first course in differential equations, but we present it because it shows something interesting.

We let $f\colon \mathbb{R} \to \mathbb{R}$ be a function with the properties that (1) f takes values in $[0, 1]$ and (2) the integral of the restriction of f to any interval does not exist. Such a function is not likely to come readily to hand, but they do exist; this is the part of this example that is beyond most students using this as a course text.

In any case, given such an f, we define a scalar autonomous ordinary differential equation with right-hand side $\widehat{F}(t, x) = f(t)$. As in Example 1.4.1, a solution of this differential equation is given by

$$x(t) = x(t_0) + \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau.$$

In this case, because no matter how we choose t and t_0, the integral of $f|[t_0, t]$ (or $f|[t, t_0]$ if $t < t_0$) does not exist, and so a solution cannot exist for any choice of t and t_0. •

1.4.3 Example (Uniqueness of solutions is not the right thing to ask for) Let us now let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function, which implies that the integral of $f|[a, b]$ exists for any $a < b$. As in our preceding two examples, we consider a differential equation with right-hand side $F(t, x) = f(t)$. And, as with the preceding two examples, solutions to this differential equation satisfy

$$x(t) = x(t_0) + \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau.$$

In this case, the integral exists for any t_0 and t, and this shows that this differential equation has *many* solutions. But what we notice is that, once we fix an initial time t_0 and an initial value $x(t_0)$ at this time, then the solution does become unique. •

1.4.4 Example (Solutions, when they exist, may have a limited domain of definition) The next example we consider shows that, even for seemingly well-behaved right-hand sides, solutions to differential equations will not be defined for all time. We consider a scalar autonomous ordinary differential equation with right-hand side $\widehat{F}(t, x) = x^2$. Thus solutions satisfy the equation

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

This equation can be easily solved (we shall see how to solve a class of equations including this one in Section 2.1) to give

$$x(t) = \begin{cases} 0, & x(t_0) = 0, \\ \frac{x(t_0)}{x(t_0)(t_0 - t) + 1}, & x(t_0) \neq 0. \end{cases}$$

(Alternatively, one can just verify by substitution that this is a solution of the differential equation and satisfies "$x(t_0) = x(t_0)$.") Let us assume that $x(t_0) \neq 0$. One can see that the solution in this case is only defined for

$$x(t_0)(t_0 - t) + 1 \neq 0 \iff t \neq t_0 + \frac{1}{x(t_0)} \triangleq t_*.$$

From this we conclude the following about solutions:

1. if $x(t_0) > 0$, then $\lim_{t \downarrow -\infty} x(t) = 0$ and $\lim_{t \uparrow t_*} x(t) = \infty$;
2. if $x(t_0) < 0$, then $\lim_{t \downarrow t_*} x(t) = -\infty$ and $\lim_{t \uparrow \infty} x(t) = 0$.

The essential point is that although (1) solutions exist for any initial time t_0 and any initial value $x(t_0)$ at that time and (2) the differential equation is defined for all times (and indeed is independent of time), solutions with initial values different from 0 will not exist for all times. •

1.4.5 Example (Solutions may not be unique even when things seem nice) We consider the scalar autonomous differential equation with right-hand side $\widehat{F}(t, x) = x^{1/3}$. We will show that there are infinitely many solutions $t \mapsto x(t)$ satisfying the equation

$$\dot{x}(t) = x(t)^{1/3}$$

with $x(0) = 0$. One can use the techniques of Section 2.1 to obtain the solution $t \mapsto x_0(t)$ given by

$$x_0(t) = \left(\frac{2}{3}t\right)^{3/2}.$$

However, $x_1(t) = 0$ is also clearly a solution. Indeed, there is a family of solutions of the form

$$x(t) = \begin{cases} x_0(t + t_-), & t \in (-\infty, -t_-], \\ 0, & t \in (-t_-, t_+), \\ x_0(t - t_+), & t \in [t_+, \infty), \end{cases}$$

where $t_-, t_+ \in \mathbb{R}_{>0}$. •

From the preceding examples, we draw the following conclusions about the questions of existence and uniqueness of solutions to ordinary differential equations.

1. From Examples 1.4.1 and 1.4.2 we conclude that we must prescribe some conditions on the right-hand side of $\widehat{F}$ of an ordinary differential equation if we are to expect solutions to exist. This is hardly a surprise, of course. However, just what are the right conditions is something that took smart people some time to figure out, cf. the proof of Theorem 1.4.8 below.
2. Example 1.4.3 shows that in the case when we have solutions, we will have lots of them, so ordinary differential equations should not be expected to have unique solutions. However, in the example we saw that perhaps the matter of uniqueness can be resolved by asking that the unknown x take on a prescribed value at a prescribed time t_0. This is altogether akin to constants of integration disappearing when fixed upper and lower limits for the integral are chosen.
3. Example 1.4.4 shows that, even when solutions exist for all initial times and values of the unknown at these times, and even when the differential equation is autonomous, it can arise that solutions only exist locally in time, i.e., solutions cannot be defined for all times. It turns out that this is just a fact of life when dealing with differential equations.
4. Finally, Example 1.4.5 shows that, even when the differential equation is autonomous with a continuous right-hand side, it can happen that multiple, indeed infinitely many, solutions pass through the same initial value for the unknown at the same time. This is a quite undesirable state of affairs, and can be hypothesised away easily by conditions that are nearly always met in practice.

With an excellent understanding of the context of the existence and uniqueness problem bestowed upon us by these motivational examples, we can now state precisely with the problem is, and provide some notation for stating the main theorem.

Let us first state precisely the problem for whose solutions we consider existence and uniqueness.

1.4.6 Definition (Initial value problem) Let

$$\boldsymbol{F}\colon \mathbb{T} \times U \to \mathbb{R}^m$$

be an ordinary differential equation with right-hand side $\widehat{\boldsymbol{F}}$. Let $t_0 \in \mathbb{T}$ and $\boldsymbol{x}_0 \in U$. A map $\xi\colon \mathbb{T}' \to U$ is a ***solution*** for $\boldsymbol{F}$ with ***initial value*** $\boldsymbol{x}_0$ at t_0 if it satisfies the following conditions:

(i) $\mathbb{T}' \subseteq \mathbb{T}$ is an interval;

(ii) ξ is of class C^1;

(iii) $\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t))$ for all $t \in \mathbb{T}'$;

(iv) $\xi(t_0) = x_0$.

In this case, we say that ξ is a solution to the ***initial value problem***

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x_0. \qquad \bullet$$

1.4.1.2 Principal existence and uniqueness theorems for ordinary differential equations In order to state an appropriate existence and uniqueness theorem, we need to define the following attribute of map between Euclidean spaces.

1.4.7 Definition (Lipschitz map) Let $U \subseteq \mathbb{R}^n$ be an open set.

(i) A map $f\colon U \to \mathbb{R}^m$ is ***Lipschitz*** if there exists $L \in \mathbb{R}_{>0}$ such that

$$\|f(x) - f(y)\| \leq L\|x - y\|, \qquad x, y \in U.$$

(ii) A map $f\colon U \to \mathbb{R}^m$ is ***locally Lipschitz*** if, for each $x \in U$, there exists $r \in \mathbb{R}_{>0}$ such that $f|\mathsf{B}(r, x)$ is Lipschitz. $\bullet$

One can show that if a map $f\colon U \to \mathbb{R}^m$ is differentiable, then it is locally Lipschitz, so this provides for us a wealth of functions that are locally Lipschitz.

Finally, we can state the main existence and uniqueness theorem for solutions to initial value problems. Because of Exercise 1.3.23, it is sufficient to consider first-order ordinary differential equations.

1.4.8 Theorem (Existence and uniqueness of solutions for ordinary differential equations) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{m}}$ *be open, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, and let* $\mathbf{F}$ *be a first-order ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{m}}.$$

We have the following two statements.

(i) Existence for continuous ordinary differential equations. *Suppose that* $\mathbf{F}$ *satisfies the following conditions:*

(a) the map $\mathrm{t} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})$ *is continuous for each* $\mathbf{x} \in \mathrm{U}$*;*

(b) the map $\mathbf{x} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})$ *is continuous for each* $\mathrm{t} \in \mathbb{T}$*;*

(c) for each $\mathbf{x} \in \mathrm{U}$*, there exists* $\mathrm{r} \in \mathbb{R}_{>0}$ *and a continuous function* $\mathrm{g}\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ *such that*

$$\|\widehat{\mathbf{F}}(\mathrm{t}, \mathbf{y})\| \leq \mathrm{g}(\mathrm{t}), \qquad (\mathrm{t}, \mathbf{y}) \in \mathbb{T} \times \mathsf{B}(\mathrm{r}, \mathbf{x}).$$

Then, for each $(\mathrm{t}_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$*, there exists a subinterval* $\mathbb{T}' \subseteq \mathbb{T}$*, relatively open in* $\mathbb{T}$ *and with* $\mathrm{t}_0 \in \mathrm{int}_{\mathbb{T}}(\mathbb{T}')$*, and a solution* $\xi\colon \mathbb{T}' \to \mathrm{U}$ *for* $\mathbf{F}$ *such that* $\xi(\mathrm{t}_0) = \mathbf{x}_0$*.*

(ii) Uniqueness for Lipschitz ordinary differential equations. *Suppose that* $\mathbf{F}$ *satisfies the following conditions:*

(a) the map $\mathrm{t} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})$ *is continuous for each* $\mathbf{x} \in \mathrm{U}$*;*

(b) *the map* $\mathbf{x} \mapsto \widehat{\mathbf{F}}(t, \mathbf{x})$ *is locally Lipschitz for each* $t \in \mathbb{T}$;

(c) *for each* $\mathbf{x} \in U$, *there exist* $r \in \mathbb{R}_{>0}$ *and continuous functions* $g, L \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ *such that*

$$\|\widehat{\mathbf{F}}(t, \mathbf{y})\| \leq g(t), \qquad (t, \mathbf{y}) \in \mathbb{T} \times B(r, \mathbf{x}), \tag{1.34}$$

and

$$\|\widehat{\mathbf{F}}(t, \mathbf{y}_1) - \widehat{\mathbf{F}}(t, \mathbf{y}_2)\| \leq L(t)\|\mathbf{y}_1 - \mathbf{y}_2\|, \qquad t \in \mathbb{T},\ \mathbf{y}_1, \mathbf{y}_2 \in B(r, \mathbf{x}). \tag{1.35}$$

Then, for each $(t_0, \mathbf{x}_0) \in \mathbb{T} \times U$, *there exists a subinterval* $\mathbb{T}' \subseteq \mathbb{T}$, *relatively open in* $\mathbb{T}$ *and with* $t_0 \in \operatorname{int}_{\mathbb{T}}(\mathbb{T}')$, *and a solution* $\xi \colon \mathbb{T}' \to U$ *for* $\mathbf{F}$ *such that* $\xi(t_0) = \mathbf{x}_0$. *Moreover, if* $\mathbb{T}''$ *is another such interval and* $\eta \colon \mathbb{T}'' \to U$ *is another such solution, then* $\eta(t) = \xi(t)$ *for all* $t \in \mathbb{T}'' \cap \mathbb{T}'$.

Before we embark on a proof of this theorem, let us make a few comments. The nature of our assumptions concerning the explicit dependence of a differential equation on time is far stronger than is required. Indeed, in many applications, the assumption of continuous dependence on time does not hold, e.g., when there is "switching." With this in mind, we will provide two versions of proofs of the theorem. First we will sketch a proof of part (ii) of the theorem, since the main ideas can be mainly understood by a typical student using this as a course text. (There is no such "simple" proof of part (i).) Then we provide a full proof of the theorem, but in a context more general than the theorem statement (the precise hypotheses are given in the proof). We make no attempt in the proof to develop the machinery of the proof. We do this not so it can be learnt as part of an introductory course, but rather so ambitious students can see what is involved and get an appreciation for how much they have yet to learn at this point in their lives.

A sketch of a proof of part (ii) The basic idea of the proof is that, by the Fundamental Theorem of Calculus, the function $t \mapsto \xi(t)$ satisfies the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \qquad \xi(t_0) = x_0$$

if and only if

$$\xi(t) = x_0 + \int_{t_0}^{t} \widehat{F}(s, \xi(s))\, \mathrm{d}s. \tag{1.36}$$

We iteratively construct a solution to this last equation. Thus we construct a sequence of functions $(\xi_j)_{j \in \mathbb{Z}_{>0}}$ that converges, in some sense, to a limit function, and the limit function satisfies the equation (1.36).

We first define $\xi_1(t) = x_0$, i.e., ξ_1 is a constant function. Then we define

$$\xi_2(t) = x_0 + \int_{t_0}^{t} \widehat{F}(s, \xi_0(s))\, \mathrm{d}s.$$

We then proceed recursively. Thus, if we have defined $\xi_1, \dots, \xi_j$, we define ξ_{j+1} by

$$\xi_{j+1}(t) = x_0 + \int_{t_0}^{t} \widehat{F}(s, \xi_j(s))\, \mathrm{d}s.$$

The objective is to show that this sequence converges. We do this by showing that, for t's sufficiently near t_0, the points $\xi_j(t)$ get closer and closer together as $j \to \infty$. To see how this might work, let us estimate the distance between $\xi_j(t)$ and $\xi_{j+1}(t)$. We have

$$\begin{aligned} \|\xi_{j+1}(t) - \xi(t)\| &= \left\| x_0 + \int_{t_0}^{t} \widehat{F}(s, \xi_j(s))\, \mathrm{d}s - \left(x_0 + \int_{t_0}^{t} \widehat{F}(s, \xi_{j-1}(s))\, \mathrm{d}s \right) \right\| \\ &\leq \int_{t_0}^{t} \|\widehat{F}(s, \xi_j(s)) - \widehat{F}(s, \xi_{j-1}(s))\|\, \mathrm{d}s. \end{aligned} \tag{1.37}$$

We can see how assumption (1.35) now becomes useful, since it is easy to see that this assumption gives (skipping some easy details)

$$\|\xi_{j+1}(t) - \xi(t)\| \leq \int_{t_0}^{t} L(s)\|\xi_j(s) - \xi_{j-1}(s)\|\, \mathrm{d}s$$

for some continuous function $L \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$, provided that $\xi_j(t)$ can be made to be close enough to x_0 for all $j \in \mathbb{Z}_{>0}$ and all t. The way we do this is as follows. We first choose $r \in \mathbb{R}_{>0}$ such that $\overline{\mathsf{B}}(r, x_0) \subseteq U$. We then choose some $\lambda \in (0, 1)$ (matters not which). Then we choose $T > t_0$ small enough that

$$\int_{t_0}^{T} g(s) < r, \quad \int_{t_0}^{T} L(s)\, \mathrm{d}s < \frac{\lambda}{2r}.$$

Such a T exists since the functions

$$t \mapsto \int_{t_0}^{t} g(s)\, \mathrm{d}s, \quad t \mapsto \int_{t_0}^{t} L(s)\, \mathrm{d}s$$

are continuous (in fact, differentiable) and by our assumption (1.34). It is then straightforward (using the triangle inequality) to show that

$$\left\| \int_{t_0}^{t} \widehat{F}(s, \xi_j(s))\, \mathrm{d}s - x_0 \right\| < r$$

and

$$\int_{t_0}^{t} \|\widehat{F}(s, \xi_j(s)) - \widehat{F}(s, \xi_{j-1}(s))\|\, \mathrm{d}s < \lambda$$

for all $t \in [t_0, T]$ and $j \in \mathbb{Z}_{>0}$. This shows two things:

1. as long as we choose r and T as above, the functions ξ_j take values in $\mathsf{B}(r, x_0)$ for all $t \in [t_0, T]$ (i.e., these functions do not blow up as either t or j increase);
2. given a $\lambda \in (0, 1)$, as long as we choose r and T as above, by our calculation (1.37), we have that, for each $j \in \mathbb{Z}_{>0}$ and each $t \in [t_0, T]$,

$$\|\xi_{j+1}(t) - \xi_j(t)\| < \lambda \sup\{\|\xi_j(s) - \xi_{j-1}(s)\| \mid s \in [t_0, T]\}.$$

This last observation can be applied recursively as follows:

$$\begin{aligned}
& \|\xi_2(t) - \xi_1(t)\| < \lambda \sup\{\|\xi_1(s) - \xi_0(s)\| \mid s \in [t_0, T]\}, \\
\implies \quad & \|\xi_3(t) - \xi_2(t)\| < \lambda \sup\{\|\xi_2(s) - \xi_1(s)\| \mid s \in [t_0, T]\} \\
& \qquad < \lambda^2 \sup\{\|\xi_1(s) - \xi_0(s)\| \mid s \in [t_0, T]\}, \\
& \qquad \vdots \\
\implies \quad & \|\xi_{j+1}(t) - \xi_j(t)\| < \lambda^j \sup\{\|\xi_1(s) - \xi_0(s)\| \mid s \in [t_0, T]\} \\
& \qquad \vdots
\end{aligned}$$

Therefore, by making j large, $\|\xi_{j+1}(t) - \xi_j(t)\|$ can be made small, uniformly in $t \in [t_0, T]$. One can readily make oneself believe that this means that the sequence (ξ_j) converges to a function $\xi \colon [t_0, T] \to \mathsf{B}(r, x_0)$. It does, but more is true. One can also fairly easily show that the resulting limit function satisfies (1.36), and so solves the initial value problem. ■

Complete proof of Theorem 1.4.8 As mentioned above, we prove a result with weaker hypotheses on time-dependence than those of the theorem statement, and with correspondingly modified conclusions. The precise hypotheses we use are:

Part (i) we suppose that (a) $t \mapsto \widehat{F}(t, x)$ is locally integrable for every $x \in U$, (b) $x \mapsto \widehat{F}(t, x)$ is continuous for every $t \in \mathbb{T}$, and (c) for every compact set $K \subseteq U$, there exists a locally integrable function $g \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{F}(t, x)\| \leq g(t), \qquad (t, x) \in \mathbb{T} \times K;$$

Part (ii) we suppose that (a) $t \mapsto \widehat{F}(t, x)$ is locally integrable for every $x \in U$, (b) $x \mapsto \widehat{F}(t, x)$ is locally Lipschitz for every $t \in \mathbb{T}$, and (c) for every compact set $K \subseteq U$, there exists locally integrable functions $g, L \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{F}(t, x)\| \leq g(t), \qquad (t, x) \in \mathbb{T} \times K;$$

and

$$\|\widehat{F}(t, x) - \widehat{F}(t, y)\| \leq L(t)\|x - y\|, \qquad t \in \mathbb{T},\ x, y \in K.$$

The conclusions must then be slightly modified, since a solution will no longer be of class C^1. Instead a solution will have a property known as "absolute continuity." This is precisely the property that a function be the indefinite integral of an integrable function. Such solutions will have the property that the equation

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t))$$

does not hold for every t, but for "almost every" t, this being made precise using Lebesgue measure, which we will not talk about.

(i) Let us first prove a lemma.

1 Lemma *For a continuous map $\xi\colon \mathbb{T} \to \mathrm{U}$, the function $\mathrm{t} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t}))$ is locally integrable.*

Proof First of all, let us show that $t \mapsto \widehat{\boldsymbol{F}}(t, \xi(t))$ is measurable. It suffices to prove this when $\mathbb{T}$ is compact, so we make this assumption. Since ξ is continuous, there exists a sequence $(\xi_j)_{j\in\mathbb{Z}_{>0}}$ of piecewise constant functions converging uniformly to ξ.***missing stuff*** That is, for each $j \in \mathbb{Z}_{>0}$ there exists a partition $(\mathbb{T}_{j,1}, \dots, \mathbb{T}_{j,k_j})$ of $\mathbb{T}$ such that $\xi_j(t) = \boldsymbol{x}_{j,l}$ for some $\boldsymbol{x}_{j,l} \in \mathbb{R}^m$ when $t \in \mathbb{T}_{j,l}$ for $l \in \{1, \dots, k_j\}$. Then

$$\widehat{\boldsymbol{F}}(t\, \xi_j(t)) = \sum_{l=1}^{k_j} \widehat{\boldsymbol{F}}(t, \boldsymbol{x}_{j,l}) \chi_{\mathbb{T}_{j,l}},$$

where χ_A denotes the characteristic function of a subset A of a set S, and so $t \mapsto \widehat{\boldsymbol{F}}(t, \xi_j(t))$ is measurable. Now, by continuity of $\boldsymbol{x} \mapsto \widehat{\boldsymbol{F}}(t, \boldsymbol{x})$,

$$\lim_{j\to\infty} \widehat{\boldsymbol{F}}(t, \xi_j(t)) = \widehat{\boldsymbol{F}}(t, \xi(t))$$

and measurability of $t \mapsto \widehat{\boldsymbol{F}}(t, \xi(t))$ follows since the pointwise limit of measurable functions is measurable***missing stuff***.

Now let $t, t_0 \in \mathbb{T}$ and suppose that $t > t_0$. Then, by continuity of ξ, there exists a compact set $K \subseteq U$ such that $\xi(s) \in K$ for every $s \in [t_0, t_0 + t]$. By assumption, there exists a locally integrable function $g\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that $\|\widehat{\boldsymbol{F}}(s, \boldsymbol{x})\| \leq g(s)$ for every $(s, \boldsymbol{x}) \in \mathbb{T} \times K$. Therefore,

$$\int_{t_0}^{t} \|\widehat{\boldsymbol{F}}(s, \xi(s))\|\, \mathrm{d}s \leq \int_{t_0}^{t} g(s)\, \mathrm{d}s < \infty.$$

The same statement holds if $t < t_0$, flipping the limits of integration, and this gives the desired local integrability. ▼

Let $r \in \mathbb{R}_{>0}$ be chosen so that $\overline{\mathsf{B}}(r, \boldsymbol{x}_0) \subseteq U$. By assumption, there exists a locally integrable $g\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that $\|\widehat{\boldsymbol{F}}(t, \boldsymbol{x})\| \leq g(t)$ for every $(t, \boldsymbol{x}) \in \mathbb{T} \times \overline{\mathsf{B}}(r, \boldsymbol{x}_0)$. Then, since g is locally integrable, the function $G_+\colon [t_0, \infty) \cap \mathbb{T} \to \mathbb{R}$ defined by

$$G_+(t) = \int_{t_0}^{t} g(s)\, \mathrm{d}s \tag{1.38}$$

is continuous.

Let us suppose that $t_0 \neq \sup \mathbb{T}$ so that there exists $b \in \mathbb{R}_{>0}$ such that $[t_0, t_0 + b] \subseteq \mathbb{T}$. Thus, since g is nonnegative, there exists $T_+ \in \mathbb{R}_{>0}$ such that $[t_0, t_0 + T_+] \subseteq \mathbb{T}$ and such that

$$G_+(t) = \int_{t_0}^{t} g(s)\, \mathrm{d}s < r, \qquad t \in [t_0, t_0 + T_+].$$

For the remainder of the proof, we consider r and T_+ to be chosen as above.

Let $\mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$ be the Banach space of continuous $\mathbb{R}^m$-valued functions on $[t_0, t_0 + T_+]$ equipped with the norm

$$\|\xi\|_\infty = \sup\{\|\xi(t)\| \mid t \in [t_0, t_0 + T_+]\}$$

(see ***missing stuff***). Let $\xi_0 \in \mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$ be defined by $\xi_0(t) = x_0$. Let $\overline{\mathsf{B}}_+(r, \xi_0)$ be the closed ball of radius r and centre ξ_0 in $\mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$. For $\alpha \in (0, T_+]$, let us define $\xi_\alpha \in \overline{\mathsf{B}}_+(r, \xi_0)$ by

$$\xi_\alpha(t) = \begin{cases} x_0, & t \in [t_0, t_0 + \alpha], \\ x_0 + \int_{t_0}^t \widehat{F}(s, \xi_\alpha(s - \alpha)) \, \mathrm{d}s, & t \in (t_0 + \alpha, t_0 + T_+]. \end{cases}$$

It is not clear that this definition makes sense, so let us verify how it does. We fix $\alpha \in (0, T_+]$. If $t \in [t_0, t_0 + \alpha]$, then the meaning of $\xi_\alpha(t)$ is unambiguous. If $t \in (t_0 + \alpha, t_0 + 2\alpha] \cap [t_0, t_0 + T_+]$, then $\xi_\alpha(t)$ is determined from the already known value of ξ_α on $[t_0, t_0 + \alpha]$. Similarly, if $t \in (t_0 + 2\alpha, t_0 + 3\alpha] \cap [t_0, t_0 + T_+]$, then $\xi_\alpha(t)$ is determined from the already known value of ξ_α on $[t_0, t_0 + 2\alpha]$. In a finite number of such steps, one determines ξ_α on $[t_0, t_0 + T_+]$. We now show that $\xi_\alpha \in \overline{\mathsf{B}}_+(r, \xi_0)$. If $t \in [t_0, t_0 + \alpha]$, then $\|\xi_\alpha(t) - x_0\| = 0$. If $t \in (t_0 + \alpha, t_0 + 2\alpha]$, then

$$\begin{aligned} \|\xi_\alpha(t) - x_0\| &= \left\| \int_{t_0}^{t_0 + \alpha} \mathbf{0} \, \mathrm{d}s + \int_{t_0 + \alpha}^t \widehat{F}(s, x_0) \, \mathrm{d}s \right\| \\ &\leq \int_{t_0}^{t_0 + \alpha} \mathbf{0} \, \mathrm{d}s + \int_{t_0 + \alpha}^t \|\widehat{F}(s, x_0)\| \, \mathrm{d}s \leq \int_{t_0}^t g(s) \, \mathrm{d}s < r. \end{aligned}$$

By induction, if $t \in (t_0 + (k-1)\alpha, t_0 + k\alpha]$, then

$$\|\xi_\alpha(t) - x_0\| \leq \sum_{j=0}^{k-2} \int_{t_0 + j\alpha}^{t_0 + (j+1)\alpha} g(s) \, \mathrm{d}s + \int_{t_0 + (k-1)\alpha}^t g(s) \, \mathrm{d}s \leq r,$$

giving $\xi_\alpha \in \overline{\mathsf{B}}_+(r, \xi_0)$, as desired.

We claim that the family $(\xi_\alpha)_{\alpha \in (0, T_+]}$ is equicontinuous, i.e., for each $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that

$$|t_1 - t_2| < \delta \quad \Longrightarrow \quad \|\xi_\alpha(t_1) - \xi_\alpha(t_2)\| < \epsilon$$

for all $\alpha \in (0, T_+]$. So let $\epsilon \in \mathbb{R}_{>0}$ and note that the function $G_+ \colon [t_0, t_0 + T_+] \to \mathbb{R}$ defined by (1.38) is continuous, and so uniformly continuous, its domain being compact. Therefore, there exists $\delta \in \mathbb{R}_{>0}$ such that

$$|t_1 - t_2| < \delta \quad \Longrightarrow \quad |G_+(t_1) - G_+(t_2)| < \epsilon.$$

Let δ be so chosen. Then, if $|t_1 - t_2| < \delta$ with $t_1 > t_2$,

$$\begin{aligned}\|\xi_\alpha(t_1) - \xi_\alpha(t_2)\| &= \left\| \int_{t_0}^{t_1} \widehat{\boldsymbol{F}}(s, \xi_\alpha(t-\alpha))\,\mathrm{d}s - \int_{t_0}^{t_2} \widehat{\boldsymbol{F}}(s, \xi_\alpha(t-\alpha))\,\mathrm{d}s \right\| \\ &\leq \int_{t_2}^{t_1} \|\widehat{\boldsymbol{F}}(s, \xi_\alpha(t-\alpha))\|\,\mathrm{d}s \leq \int_{t_2}^{t_1} g(s)\,\mathrm{d}s = G_+(t_1) - G_+(t_2) < \epsilon,\end{aligned}$$

as desired.

Thus we have an equicontinuous family $(\xi_\alpha)_{\alpha\in(0,T_+]}$ contained in the bounded set $\overline{\mathsf{B}}_+(r, \xi_0)$. Consider then the sequence $(\xi_{T_+/j})_{j\in\mathbb{Z}_{>0}}$ contained in this family. By the Arzelà–Ascoli Theorem***missing stuff*** and the Bolzano–Weierstrass Theorem***missing stuff*** there exists an increasing sequence $(j_k)_{k\in\mathbb{Z}_{>0}}$ such that the sequence $(\xi_{T_+/j_k})_{k\in\mathbb{Z}_{>0}}$ converges in $\mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$, i.e., converges uniformly. Let us denote the limit by $\xi_+ \in \overline{\mathsf{B}}_+(r, \xi_0)$. It remains to show that the ξ_+ is a solution for $\boldsymbol{F}$ satisfying $\xi_+(t_0) = \boldsymbol{x}_0$. For this, an application of the Dominated Convergence Theorem***missing stuff***, continuity of $\widehat{\boldsymbol{F}}$ in the second argument, and equicontinuity of $(\xi_\alpha)_{\alpha\in(0,T_+]}$ gives

$$\begin{aligned}\xi_+(t) &= \lim_{k\to\infty} \xi_{T_+/j_k}(t) = \boldsymbol{x}_0 + \lim_{j_k\to\infty} \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \xi_{T_+/j_k}(s - T_+/j_k))\,\mathrm{d}s \\ &= \boldsymbol{x}_0 + \int_{t_0} \widehat{\boldsymbol{F}}(s, \lim_{\alpha\to 0} \xi_\alpha(s-\alpha))\,\mathrm{d}s = \boldsymbol{x}_0 + \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \xi_+(s))\,\mathrm{d}s.\end{aligned}$$

Therefore, by the lemma above, ξ_+ is absolutely continuous and

$$\dot{\xi}_+(t) = \widehat{\boldsymbol{F}}(t, \xi_+(t))$$

for almost every $t \in [t_0, t_0 + T_+]$. Thus ξ_+ is a solution for $\boldsymbol{F}$. Obviously $\xi_+(t_0) = \boldsymbol{x}_0$.

Next suppose that $t_0 \neq \inf \mathbb{T}$. Then there exists $a \in \mathbb{R}_{>0}$ such that $[t_0 - a, t_0] \subseteq \mathbb{T}$. As above, we let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{B}}(r, \boldsymbol{x}_0) \subseteq U$. Define $G_-\colon (-\infty, t_0] \cap \mathbb{T} \to \mathbb{R}$ by

$$G_-(t) = \int_t^{t_0} g(s)\,\mathrm{d}s$$

so that G_- is continuous. Since g is nonnegative, there exists $T_- \in \mathbb{R}_{>0}$ such that $[t_0, t_0 - T_-] \subseteq \mathbb{T}$ and such that

$$G_-(t) = \int_t^{t_0} g(s)\,\mathrm{d}s < r, \qquad t \in [t_0 - T_-, t_0].$$

Now, with r and T_- thusly defined, we can proceed as above to show the existence of a solution $\xi_-\colon [t_0 - T_-, t_0] \to U$ for $\boldsymbol{F}$ such that $\xi_-(t_0) = \boldsymbol{x}_0$.

The proof of this part of the theorem is complete if we define $\mathbb{T}'$ and ξ as follows.

1. $\operatorname{int}(\mathbb{T}) = \emptyset$: The interval $\mathbb{T}' = \{t_0\}$ and the trivial solution $\xi_0(t) = x_0$ satisfies the conclusions of the theorem.
2. $t_0 \neq \sup \mathbb{T}$ and $t_0 = \inf \mathbb{T}$: The interval $\mathbb{T}' = [t_0, t_0 + T_+)$ and the solution $\xi = \xi_+$ as defined above satisfy the conclusions of the theorem.
3. $t_0 = \sup \mathbb{T}$ and $t_0 \neq \inf \mathbb{T}$: The interval $\mathbb{T}' = [t_0 - T_-, t_0)$ and the solution $\xi = \xi_-$ as defined above satisfy the conclusions of the theorem.
4. $t_0 \neq \sup \mathbb{T}$ and $t_0 \neq \inf \mathbb{T}$: The interval $\mathbb{T}' = (t_0 - T_-, t_0 + T_+)$ and the solution

$$\xi(t) = \begin{cases} \xi_-(t), & t \in (t_0 - T_-, t_0], \\ \xi_+(t), & t \in (t_0, t_0 + T_+] \end{cases}$$

satisfy the conclusions of the theorem.

(ii) Note that the existence statement follows from part (i) since the hypotheses of part (ii) imply those of part (i). However, we shall reprove this via an argument that also ensures uniqueness.

Let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{B}}(r, x_0) \subseteq U$. As in the proof of part (i), there exists a locally integrable $g\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, x)\| \leq g(t), \qquad (t, x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0).$$

By hypothesis, there exists a locally integrable $L\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, x) - \widehat{\boldsymbol{F}}(t, y)\| \leq L(t)\|x - y\|$$

for all $t \in \mathbb{T}$ and $x, y \in \overline{\mathsf{B}}(r, x_0)$. Let us choose $\lambda \in (0, 1)$.

We first consider the case where $t_0 \neq \sup \mathbb{T}$ so that there exists $b \in \mathbb{R}_{>0}$ such that $[t_0, t_0 + b] \subseteq \mathbb{T}$. Define $G_+, \ell_+\colon [t_0, \infty) \cap \mathbb{T} \to \mathbb{R}$ by

$$G_+(t) = \int_{t_0}^{t} g(s)\,\mathrm{d}s, \quad \ell_+(t) = \int_{t_0}^{t} L(s)\,\mathrm{d}s.$$

Since g and L are nonnegative, we can choose $T_+ \in \mathbb{R}_{>0}$ such that

$$G_+(t) = \int_{t_0}^{t} g(s)\,\mathrm{d}s \leq r, \quad \ell_+(t) = \int_{t_0}^{t} L(s)\,\mathrm{d}s < \lambda$$

for $t \in [t_0, t_0 + T_+]$.

As in the proof of part (i), let ξ_0 be the trivial function $t \mapsto x_0$, $t \in [t_0, t_0 + T_+]$, and let $\overline{\mathsf{B}}_+(r, \xi_0)$ be the ball of radius r and centre ξ_0 in $\mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$. Define $F_+\colon \overline{\mathsf{B}}_+(r, \xi_0) \to \mathsf{C}^0([t_0, t_0 + T_+]; \mathbb{R}^m)$ by

$$F_+(\xi)(t) = x_0 + \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \xi(s))\,\mathrm{d}s.$$

By the lemma from the proof of part (i), $s \mapsto \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s))$ is locally integrable, showing that F_+ is well-defined and that $F_+(\boldsymbol{\xi})$ is absolutely continuous.

We claim that $F_+(\overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)) \subseteq \overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$. Suppose that $\boldsymbol{\xi} \in \overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$ so that

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq r, \qquad t \in [t_0, t_0 + T_+].$$

Then, for $t \in [t_0, t_0 + T_+]$,

$$\|F_+(\boldsymbol{\xi})(t) - \boldsymbol{x}_0\| = \left\| \int_{t_0}^t \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) \, \mathrm{d}s \right\| \leq \int_{t_0}^t \|\widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s))\| \, \mathrm{d}s \leq \int_{t_0}^t g(s) \, \mathrm{d}s \leq r,$$

as desired.

We claim that $F_+|\overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$ is a contraction mapping. That is, we claim that there exists $\rho \in [0, 1)$ such that

$$\|F_+(\boldsymbol{\xi}) - F_+(\boldsymbol{\eta})\|_\infty \leq \rho \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty$$

for every $\boldsymbol{\xi}, \boldsymbol{\eta} \in \overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$. Indeed, let $\boldsymbol{\xi}, \boldsymbol{\eta} \in \overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$ and compute, for $t \in [t_0, t_0 + T_+]$,

$$\begin{aligned} \|F_+(\boldsymbol{\xi})(t) - F_+(\boldsymbol{\eta})(t)\| &= \left\| \int_{t_0}^t \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) \, \mathrm{d}s - \int_{t_0}^t \widehat{\boldsymbol{F}}(s, \boldsymbol{\eta}(s)) \right\| \\ &\leq \int_{t_0}^t \|\widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) - \widehat{\boldsymbol{F}}(s, \boldsymbol{\eta}(s))\| \, \mathrm{d}s \\ &\leq \int_{t_0}^t L(s) \|\boldsymbol{\xi}(s) - \boldsymbol{\eta}(s)\| \, \mathrm{d}s \leq \ell_+(t) \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty \leq \lambda \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty, \end{aligned} \tag{1.39}$$

since $\boldsymbol{\xi}(s), \boldsymbol{\eta}(s) \in \mathsf{B}(r, \boldsymbol{x}_0)$ for every $s \in [t_0, t_0 + T_+]$. This proves that $F_+|\overline{\mathsf{B}}_+(r, \boldsymbol{\xi}_0)$ is a contraction mapping.

By the Contraction Mapping Theorem***missing stuff*** there exists a unique fixed point for F_+ which we denote by $\boldsymbol{\xi}_+$. Thus

$$\boldsymbol{\xi}_+(t) = F_+(\boldsymbol{\xi}_+)(t) = \boldsymbol{x}_0 + \int_{t_0}^t \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}_+(s)) \, \mathrm{d}s.$$

Differentiating the first and last expressions with respect to t shows that $\boldsymbol{\xi}_+$ is a solution for $\boldsymbol{F}$.

Now we consider the case when $t_0 \neq \inf \mathbb{T}$ so there exists $a \in \mathbb{R}_{>0}$ such that $[t_0 - a, t_0] \subseteq \mathbb{T}$. We proceed as above, cf. the corresponding part of the proof of part (i), to provide $T_- \in \mathbb{R}_{>0}$ such that

$$G_-(t) \triangleq \int_t^{t_0} g(s) \, \mathrm{d}s < r, \quad \ell_-(t) \triangleq \int_t^{t_0} L(s) \, \mathrm{d}s < \lambda$$

for $t \in [t_0 - T_-, t_0]$. We then define $\overline{\mathsf{B}}_-(r, \xi_0)$ as the ball of radius r and centre ξ_0 in $\mathsf{C}^0([t_0 - T_-, t_0]; \mathbb{R}^m)$ and define $F_- \colon \overline{\mathsf{B}}_-(r, x_0) \to \mathsf{C}^0([t_0 - T_-, t_0]; \mathbb{R}^m)$ by

$$F_-(\xi)(t) = x_0 + \int_{t_0}^t \widehat{F}(s, \xi(s))\, \mathrm{d}s.$$

We show, as above, that $F_-(\overline{\mathsf{B}}_-(r, \xi_0)) \subseteq \overline{\mathsf{B}}_-(r, \xi_0)$ and that $F_-|\overline{\mathsf{B}}_-(r, \xi_0)$ is a contraction mapping, so possessing a unique fixed point ξ_-. This fixed point is a solution for F, as above.

We can then define an interval $\mathbb{T}'$ and a solution ξ for F as at the end of the proof of part (i). We now prove uniqueness of this solution on $\mathbb{T}'$. Suppose that $\eta \colon \mathbb{T}' \to U$ is another solution satisfying $\eta(t_0) = x_0$. Then

$$\dot{\eta}(t) = \widehat{F}(t, \eta(t)), \qquad t \in \mathbb{T}'.$$

Therefore, by the Fundamental Theorem of Calculus,

$$\eta(t) = \eta(t_0) + \int_{t_0}^t \dot{\eta}(s)\, \mathrm{d}s = x_0 + \int_{t_0}^t \widehat{F}(s, \eta(s))$$

for $t \geq t_0$ and

$$\eta(t) = \eta(t_0) + \int_{t_0}^t \dot{\eta}(s)\, \mathrm{d}s = x_0 + \int_{t_0}^t \widehat{F}(s, \eta(s))$$

for $t \leq t_0$. It then follows that $\eta|[t_0, \infty) \cap \mathbb{T}'$ is a fixed point for F_+ and $\eta|(-\infty, t_0] \cap \mathbb{T}'$ is a fixed point for F_-. Therefore, η agrees with ξ on $\mathbb{T}'$ by the uniqueness part of the Contraction Mapping Theorem.

Now suppose that $\mathbb{T}'' \subseteq \mathbb{R}$ is some other interval containing t_0 and that $\eta \colon \mathbb{T}'' \to U$ is a solution for F satisfying $\eta(t_0) = x_0$. Suppose that $\xi(t) \neq \eta(t)$ for some $t \in \mathbb{T}'' \cap \mathbb{T}'$. Suppose that $t < t_0$. Let

$$t_1 = \inf\{t \in [t_0, \infty) \cap \mathbb{T}'' \cap \mathbb{T}' \mid \xi(t) \neq \eta(t)\}.$$

Then $\xi(t) = \eta(t)$ for $t \in [t_0, t_1)$. Continuity of solutions implies that $\xi(t_1) = \eta(t_1)$. Denote $x_1 = \xi(t_1)$. Note that both ξ and η are solutions for F satisfying $\xi(t_1) = \eta(t_1) = x_1$. By our above arguments for existence and uniqueness, there exists $T_+ \in \mathbb{R}_{>0}$ and a unique solution ζ on $[t_1, t_1 + T_+]$ satisfying $\zeta(t_1) = x_1$. Thus ξ and η must agree with ζ on $[t_1, t_1 + T_+]$ contradicting the definition of t_1. A similar argument leads to a similar contradiction when we assume that ξ and η disagree at some $t \in \mathbb{T}'' \cap \mathbb{T}'$ with $t < t_0$. ■

The matter of checking the conditions of Theorem 1.4.8 is normally quite straightforward, particularly since if we know that a function is differentiable, then it is locally Lipschitz. Indeed, let us encode in the following result a situation where the hypotheses of Theorem 1.4.8 are easily verified.

1.4.9 Corollary (An existence and uniqueness result that is easy to apply) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{m}}$ *be open, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, and let* $\mathbf{F}$ *be a first-order ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}} \colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{m}}.$$

If $\widehat{\mathbf{F}}$ *is of class* C^1 *on* $\mathbb{T} \times \mathrm{U}$*, then, for each* $(\mathrm{t}_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$*, there exists a subinterval* $\mathbb{T}' \subseteq \mathbb{T}$*, relatively open in* $\mathbb{T}$ *and with* $\mathrm{t}_0 \in \mathrm{int}_{\mathbb{T}}(\mathbb{T}')$*, and a solution* $\xi \colon \mathbb{T}' \to \mathrm{U}$ *for* $\mathbf{F}$ *such that* $\xi(\mathrm{t}_0) = \mathbf{x}_0$*. Moreover, if* $\mathbb{T}''$ *is another such interval and* $\eta \colon \mathbb{T}'' \to \mathrm{U}$ *is another such solution, then* $\eta(\mathrm{t}) = \xi(\mathrm{t})$ *for all* $\mathrm{t} \in \mathbb{T}'' \cap \mathbb{T}'$*.*

We ask the reader to check that the hypotheses of Theorem 1.4.8 are satisfied for the examples of Section 1.1 as Exercise 1.4.3. In Exercise 1.4.4 we ask the reader to show which hypotheses of Theorem 1.4.8 are violated for the examples we gave at the beginning of this section.

1.4.1.3 Flows for ordinary differential equations With the above notions of existence and uniqueness of solutions for initial value problems, in this section we give some notation that ties together *all* solutions to *all* initial value problems. In doing this, we naturally run up against the question of how solutions to initial value problems depend on initial conditions. We shall at various points in the text run into situations where this sort of dependence is important, so the results in this section, while a bit technical, are certainly an essential part of any deep understanding of ordinary differential equations.

First we introduce the notation.

1.4.10 Definition (Interval of existence, domain of solutions) Let F be an ordinary differential equation with right-hand side

$$\widehat{F} \colon \mathbb{T} \times U \to \mathbb{R}^n,$$

and assume that F satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems.

(i) For $(t_0, x_0) \in \mathbb{T} \times U$, denote

$$J_F(t_0, x_0) = \cup\{J \subseteq \mathbb{T} \mid \ J \text{ is an interval and there is a solution } \xi \colon J \to U \text{ for } F \text{ satisfying } \xi(t_0) = x_0\}.$$

The interval $J_F(t_0, x_0)$ is the ***interval of existence*** for the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x_0.$$

(ii) The ***domain of solutions*** for F is

$$D_F = \{(t, t_0, x_0) \in \mathbb{T} \times \mathbb{T} \times U \mid \ t \in J_F(t_0, x_0)\}. \qquad \bullet$$

We shall carefully enumerate various properties of intervals of existence and domains of solutions, but to do this let us first introduce a very useful bit of notation.

1.4.11 Definition (Flow of an ordinary differential equation) Let F be an ordinary differential equation with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

and assume that F satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems. The ***flow*** of F is the map $\Phi^F\colon D_F \to U$ defined by asking that $\Phi^F(t, t_0, x_0)$ is the solution, evaluated at t, of the initial value problem

$$\frac{\mathrm{d}\xi}{\mathrm{d}\tau}(\tau) = \widehat{F}(\tau, \xi(\tau)), \quad \xi(t_0) = x_0. \qquad \bullet$$

The definition, phrased differently, says that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^F(t, t_0, x_0) = \widehat{F}(t, t_0, x_0), \quad \Phi^F(t_0, t_0, x_0) = x_0.$$

For $t, t_0 \in \mathbb{T}$, it is sometimes convenient to denote

$$D_F(t, t_0) = \{x \in U \mid (t, t_0, x) \in D_F\},$$

and then

$$\begin{aligned}\Phi^F_{t,t_0}\colon D_F(t, t_0) &\to U\\ x &\mapsto \Phi^F(t, t_0, x).\end{aligned}$$

Along similar lines, for $t_0 \in \mathbb{T}$, we denote

$$D_F(t_0) = \{(t, x) \in \mathbb{T} \times U \mid (t, t_0, x) \in D_F\},$$

and then

$$\begin{aligned}\Phi^F(t_0)\colon D_F(t_0) &\to U\\ (t, x) &\mapsto \Phi^F(t, t_0, x).\end{aligned}$$

Let us enumerate some of the more elementary properties of the flow.

1.4.12 Proposition (Elementary properties of flow) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}},$$

and assume that $\mathbf{F}$ *satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems. Then the following statements hold:*

(i) for each $(\mathrm{t}_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$, $(\mathrm{t}_0, \mathrm{t}_0, \mathbf{x}_0) \in \mathrm{D}_{\mathbf{F}}$ *and* $\Phi^{\mathbf{F}}(\mathrm{t}_0, \mathrm{t}_0, \mathbf{x}_0) = \mathbf{x}_0$*;*

(ii) if $(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x}) \in \mathrm{D}_{\mathbf{F}}$*, then* $(\mathrm{t}_3, \mathrm{t}_2, \Phi^{\mathbf{F}}(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x})) \in \mathrm{D}_{\mathbf{F}}$ *if and only if* $(\mathrm{t}_3, \mathrm{t}_1, \mathbf{x}) \in \mathrm{D}_{\mathbf{F}}$ *and, if this holds, then*

$$\Phi^{\mathbf{F}}(\mathrm{t}_3, \mathrm{t}_1, \mathbf{x}) = \Phi^{\mathbf{F}}(\mathrm{t}_3, \mathrm{t}_2, \Phi^{\mathbf{F}}(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x})).$$

(iii) if $(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x}) \in \mathrm{D}_{\mathbf{F}}$*, then* $(\mathrm{t}_1, \mathrm{t}_2, \Phi^{\mathbf{F}}(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x})) \in \mathrm{D}_{\mathbf{F}}$ *and* $\Phi^{\mathbf{F}}(\mathrm{t}_1, \mathrm{t}_2, \Phi^{\mathbf{F}}(\mathrm{t}_2, \mathrm{t}_1, \mathbf{x})) = \mathbf{x}$.

Proof (i) This is part of the definition of the flow.

(ii) Suppose that $t_2 \geq t_1$ and $t_3 \geq t_2$.

First suppose that $(t_2, t_1, \boldsymbol{x}) \in D_F$ and $(t_3, t_2, \Phi^F(t_2, t_1, \boldsymbol{x})) \in D_F$. We then have solutions $\xi_1 \colon [t_1, t_2] \to U$ and $\xi_2 \colon [t_2, t_3] \to U$ to the initial value problems

$$\dot{\xi}_1(t) = \widehat{\boldsymbol{F}}(t, \xi_1(t)), \quad \xi_1(t_1) = \boldsymbol{x},$$

and

$$\dot{\xi}_2(t) = \widehat{\boldsymbol{F}}(t, \xi_2(t)), \quad \xi_2(t_2) = \Phi^F(t_2, t_1, \boldsymbol{x}),$$

respectively. Then define $\xi \colon [t_1, t_3] \to U$ by

$$\xi(t) = \begin{cases} \xi_1(t), & t \in [t_1, t_2], \\ \xi_2(t), & t \in [t_2, t_3]. \end{cases}$$

It is clear, then, that

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_1) = \boldsymbol{x}.$$

It is then also clear that

$$\xi(t_3) = \Phi^F(t_3, t_2, \Phi^F(t_2, t_1, \boldsymbol{x}))$$

and that $\xi(t_3) = \Phi^F(t_3, t_1, \boldsymbol{x})$. This gives $(t_3, t_1, \boldsymbol{x}) \in D_F$.

Now suppose that $(t_2, t_1, \boldsymbol{x}) \in D_F$ and $(t_3, t_1, \boldsymbol{x}) \in D_F$. Let $\xi_1 \colon [t_1, t_2] \to U$ and $\xi_3 \colon [t_1, t_3] \to U$ be the solutions to the initial value problems

$$\dot{\xi}_1(t) = \widehat{\boldsymbol{F}}(t, \xi_1(t)), \quad \xi_1(t_1) = \boldsymbol{x},$$

and

$$\dot{\xi}_3(t) = \widehat{\boldsymbol{F}}(t, \xi_3(t)), \quad \xi_3(t_1) = \boldsymbol{x},$$

respectively. Then, by uniqueness of solutions, the curve $\xi_2 \colon [t_2, t_3] \to U$ give by

$$\xi_2(t) = \xi_1(t) = \xi_3(t)$$

satisfies the initial value problem

$$\dot{\xi}_2(t) = \widehat{\boldsymbol{F}}(t, \xi_2(t)), \quad \xi_2(t_2) = \xi_1(t_2) = \Phi^F(t_2, t_1, \boldsymbol{x}),$$

and so $(t_3, t_2, \Phi^F(t_2, t_1, \boldsymbol{x})) \in D_F$.

The assertion that

$$\Phi^F(t_3, t_1, \boldsymbol{x}) = \Phi^F(t_3, t_2, \Phi^F(t_2, t_1, \boldsymbol{x}))$$

follows from uniqueness of solutions.

In the cases that (1) $t_1 \geq t_2$ and $t_3 \leq t_2$, (2) $t_2 \leq t_1$ and $t_3 \geq t_2$, and (3) $t_3 \leq t_2$ and $t_2 \leq t_1$, similarly styled arguments can be made, appropriately fussing with going in "different directions" in cases (1) and (2).

(iii) This is a special case of (ii), using (i). ■

Useful mnemonics associated with parts (i)–(iii) are:

$$\Phi^F_{t_0,t_0} = \mathrm{id}_U, \quad (\Phi^F_{t_2,t_1})^{-1} = \Phi^F_{t_1,t_2}, \quad \Phi^F_{t_3,t_2} \circ \Phi^F_{t_2,t_1} = \Phi^F_{t_3,t_1}.$$

However, these really are just mnemonics, since they do not account carefully for the domains of the mappings being used.

1.4.13 Theorem (Properties of flow) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}} \colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^n,$$

and assume that $\mathbf{F}$ *satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems. Then the following statements hold:*

(i) *for* $(t_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$, $\mathrm{J}_{\mathbf{F}}(t_0, \mathbf{x}_0)$ *is an interval that is a relatively open subset of* $\mathbb{T}$*;*

(ii) *for* $(t_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$, *the curve*

$$\begin{aligned} \gamma_{(t_0,\mathbf{x}_0)} \colon \mathrm{J}_{\mathbf{F}}(t_0, \mathbf{x}_0) &\to \mathrm{U} \\ t &\mapsto \Phi^{\mathbf{F}}(t, t_0, \mathbf{x}_0) \end{aligned}$$

is well-defined and continuously differentiable;

(iii) *for* $t, t_0 \in \mathbb{T}$, $\mathrm{D}_{\mathbf{F}}(t, t_0) \neq \emptyset$, $\mathrm{D}_{\mathbf{F}}(t, t_0)$ *is open in* U*;*

(iv) *for* $t, t_0 \in \mathbb{T}$ *for which* $\mathrm{D}_{\mathbf{F}}(t, t_0) \neq \emptyset$, $\mathrm{D}_{\mathbf{F}}(t, t_0)$ *is open and* $\Phi^{\mathbf{F}}_{t,t_0}$ *is a locally bi-Lipschitz homeomorphism onto its image.*

(v) *for* $t_0 \in \mathbb{T}$, $\mathrm{D}_{\mathbf{F}}(t_0)$ *is relatively open in* $\mathbb{T} \times \mathrm{U}$*;*

(vi) *for* $t_0 \in \mathbb{T}$, *the map*

$$\begin{aligned} \Phi^{\mathbf{F}}(t_0) \colon \mathrm{D}_{\mathbf{F}}(t_0) &\to \mathrm{U} \\ (t, \mathbf{x}) &\mapsto \Phi^{\mathbf{F}}(t, t_0, \mathbf{x}) \end{aligned}$$

is well-defined and continuous;

(vii) $\mathrm{D}_{\mathbf{F}}$ *is relatively open in* $\mathbb{T} \times \mathbb{T} \times \mathrm{U}$*;*

(viii) *the map*

$$\Phi^{\mathbf{F}} \colon \mathrm{D}_{\mathbf{F}} \to \mathrm{U}$$

is continuous;

(ix) *for* $(t_0, \mathbf{x}_0) \in \mathbb{T} \times \mathrm{U}$ *and for* $\epsilon \in \mathbb{R}_{>0}$, *there exists* $r, \alpha \in \mathbb{R}_{>0}$ *such that*

$$\sup \mathrm{J}_{\mathbf{F}}(t, \mathbf{x}) > \sup \mathrm{J}_{\mathbf{F}}(t_0, \mathbf{x}_0) - \epsilon, \qquad \inf \mathrm{J}_{\mathbf{F}}(t, \mathbf{x}) < \inf \mathrm{J}_{\mathbf{F}}(t_0, \mathbf{x}_0) + \epsilon$$

for all $(t, \mathbf{x}) \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}\mathsf{B}(r, \mathbf{x}_0)$.

Proof As with our proof of Theorem 1.4.8(ii), we prove a result with weaker hypotheses on time-dependence than those of the theorem statement. The precise hypotheses we use are: (a) $t \mapsto \widehat{\boldsymbol{F}}(t, \boldsymbol{x})$ is locally integrable for every $\boldsymbol{x} \in U$, (b) $\boldsymbol{x} \mapsto$

$\widehat{F}(t,x)$ is locally Lipschitz for every $t \in \mathbb{T}$, and (c) for every compact set $K \subseteq U$, there exists locally integrable functions $g, L \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{F}(t,x)\| \leq g(t), \qquad (t,x) \in \mathbb{T} \times K;$$

and

$$\|\widehat{F}(t,x) - \widehat{F}(t,y)\| \leq L(t)\|x - y\|, \qquad t \in \mathbb{T},\ x, y \in K.$$

Note that the conclusion of part (ii) must then be modified to assert that $\gamma_{(t_0,x_0)}$ is locally absolutely continuous.

(i) Since $J_F(t_0,x_0)$ is a union of intervals, each of which contains t_0, it follows that it is itself an interval. To show that it is an open subset of $\mathbb{T}$, we show that, if $t \in J_F(t_0,x_0)$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$(-\epsilon, \epsilon) \cap \mathbb{T} \subseteq J_F(t_0,x_0).$$

First let us consider the case when t is not an endpoint of $\mathbb{T}$, in the event that $\mathbb{T}$ contains one or both of its endpoints. In this case, by definition of $J_F(t_0,x_0)$, there is an open interval $J \subseteq \mathbb{T}$ containing t_0 and t, and a solution $\xi \colon J \to U$ of the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x_0.$$

In particular, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $(-\epsilon, \epsilon) \subseteq J \subseteq J_F(t_0,x_0)$, which gives the desired conclusion in this case.

Next suppose that t is the right endpoint of $\mathbb{T}$, which we assume is contained in $\mathbb{T}$, of course. In this case, by definition of $J_F(t_0,x_0)$, there is an interval $J \subseteq \mathbb{T}$ containing t_0 and t, and a solution $\xi \colon J \to U$ of the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x_0.$$

In this case, there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$(-\epsilon, \epsilon) \cap \mathbb{T} = (-\epsilon, t] \subseteq J_F(t_0,x_0),$$

which gives the desired conclusion in this case.

A similar argument gives the desired conclusion when t is the left endpoint of $\mathbb{T}$.

(ii) That $\gamma_{(t_0,x_0)}$ is defined in $J_F(t_0,x_0)$ was proved as part of the preceding part of the proof. The assertion that $\gamma_{(t_0,x_0)}$ is locally absolutely continuous follows from Theorem 1.4.8(ii).

We shall prove the assertions (iii)–(vi) of the theorem together, first by proving that these conditions hold locally, and then giving an extension argument to give the global version of the statement.

Let us first prove a few technical lemmata that will be useful to us.

1 Lemma *Let $\mathbb{T}$ be an interval, and let $\alpha, \beta, \xi \colon \mathbb{T} \to \mathbb{R}$, and $t_0 \in \mathbb{T}$ be such that*

(i) *α is continuous,*

(ii) *β is nonnegative-valued and locally integrable,*

(iii) *ξ is nonnegative-valued and continuous, and*

(iv) $\xi(t) \leq \alpha(t) + \int_{t_0}^{t} \beta(s)\xi(s)\,ds$ *for all* $t \in \mathbb{T} \cap [t_0, \infty)$.

Then

$$\xi(t) \leq \alpha(t) + \int_{t_0}^{t} \alpha(s)\beta(s)e^{\int_s^t \beta(\tau)\,d\tau}\,ds, \qquad t \in \mathbb{T} \cap [t_0, \infty).$$

Moreover, if α is additionally nondecreasing, then we have

$$\xi(t) \leq \alpha(t)e^{\int_{t_0}^{t} \beta(s)\,ds}, \qquad t \in \mathbb{T} \cap [t_0, \infty).$$

Proof Define

$$\eta(s) = e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau} \int_{t_0}^{s} \beta(\tau)\xi(\tau)\,d\tau$$

and calculate, for almost every $s \in [t_0, t]$,

$$\begin{aligned}
\frac{d\eta}{ds}(s) &= -\beta(s)e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau} \int_{t_0}^{s} \beta(\tau)\xi(\tau)\,d\tau + \beta(s)\xi(s)e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau} \\
&= \beta(s)e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau}\left(\xi(s) - \int_{t_0}^{s} \beta(\tau)\xi(\tau)\,d\tau\right) \\
&\leq \alpha(s)\beta(s)e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau},
\end{aligned}$$

using the hypotheses of the lemma. Therefore,

$$\eta(t) \leq \int_{t_0}^{t} \alpha(s)\beta(s)e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau}ds.$$

Using the definition of η we then have

$$\begin{aligned}
\int_{t_0}^{t} \beta(s)\xi(s)\,ds &\leq \int_{t_0}^{t} \alpha(s)\beta(s)e^{\int_{t_0}^{t} \beta(s)\,ds}e^{-\int_{t_0}^{s} \beta(\tau)\,d\tau}\,ds \\
&= \int_{t_0}^{t} \alpha(s)\beta(s)e^{\int_{s}^{t} \beta(\tau)\,d\tau}ds,
\end{aligned}$$

which immediately gives the first conclusion of the lemma.

For the second, we first note that, for almost every $s \in [t_0, t]$,

$$\frac{d}{ds}e^{\int_s^t \beta(\tau)\,d\tau} = -\beta(s)e^{\int_s^t \beta(\tau)\,d\tau}.$$

Then

$$\int_{t_0}^{t} \beta(s) e^{\int_s^t \beta(\tau)\,d\tau} ds = -e^{\int_s^t \beta(\tau)\,d\tau}\Big|_{s=t_0}^{s=t} = e^{\int_{t_0}^t \beta(\tau)\,d\tau} - 1.$$

Then we use the first part of the lemma and the additional assumption on α:

$$\begin{aligned}\xi(t) &\le \alpha(t) + \int_{t_0}^{t} \alpha(s)\beta(s) e^{\int_s^t \beta(\tau)\,d\tau} ds \\ &\le \alpha(t) + \alpha(t)\left(\int_{t_0}^{t} e^{\beta(s)ds} - 1\right),\end{aligned}$$

and the lemma follows. ▼

Now we give the initial part of the local version of the theorem.

2 Lemma *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathsf{U} \to \mathbb{R}^n,$$

and assume that $\mathbf{F}$ *satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems. Then, for each* $(t_0, \mathbf{x}_0) \in \mathbb{T} \times \mathsf{U}$*, there exists* $r, \alpha \in \mathbb{R}_{>0}$ *such that* $(t, t_0, \mathbf{x}) \in \mathsf{D}_{\mathbf{F}}$ *for each* $\mathbf{x} \in \mathsf{B}(r, \mathbf{x}_0)$ *and* $t \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}$*. Moreover,*

(i) the map

$$\mathsf{B}(r, \mathbf{x}_0) \ni \mathbf{x} \mapsto \Phi^{\mathbf{F}}_{t,t_0}(\mathbf{x}) \in \mathbb{R}^m$$

is Lipschitz for every $t \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}$*;*

(ii) the map

$$(t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T} \times \mathsf{B}(r, \mathbf{x}_0) \ni (t, \mathbf{x}) \mapsto \Phi^{\mathbf{F}}(t, t_0, \mathbf{x})$$

is continuous.

Proof First let $r' \in \mathbb{R}_{>0}$ be such that $\mathsf{B}(r', x_0) \subseteq U$ and let $r = \frac{r'}{2}$. As in the proof of Theorem 1.4.8(ii), there exist locally integrable $g, L\colon \mathbb{T} \to \mathbb{R}_{\ge 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{x})\| \le g(t), \qquad (t, \boldsymbol{x}) \in \mathbb{T} \times \overline{\mathsf{B}}(r', \boldsymbol{x}_0).$$

and

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{x}_1) - \widehat{\boldsymbol{F}}(t, \boldsymbol{x}_2)\| \le L(t)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$$

for all $t \in \mathbb{T}$ and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \overline{\mathsf{B}}(r', \boldsymbol{x}_0)$. Let us choose $\lambda \in (0, 1)$. As in the proof of Theorem 1.4.8(ii), there exists $\alpha \in \mathbb{R}_{>0}$ such that

$$\left|\int_{t_0}^{t} g(s)\,ds\right| \le r, \quad \left|\int_{t_0}^{t} L(s)\,ds\right| < \lambda, \qquad t \in [t_0 - \alpha, t_0 + \alpha].$$

If $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$, then $\mathsf{B}(r, \boldsymbol{x}) \subseteq \mathsf{B}(r', \boldsymbol{x}_0)$. Therefore,

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{y})\| \le g(t), \qquad (t, \boldsymbol{y}) \in \mathbb{T} \times \overline{\mathsf{B}}(r, \boldsymbol{x}).$$

and

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{y}_1) - \widehat{\boldsymbol{F}}(t, \boldsymbol{y}_2)\| \leq L(t)\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|$$

for all $t \in \mathbb{T}$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \overline{\mathsf{B}}(r, \boldsymbol{x})$. If $\xi_1 \in \mathsf{C}^0([t_0 - \alpha, t_0 + \alpha]; \mathbb{R}^m)$ is the constant function $\xi_0(t) = \boldsymbol{x}_0$, then the arguments from the proof of Theorem 1.4.8(ii) allow us to conclude that there is a solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

in $\overline{\mathsf{B}}(r, \xi_0) \subseteq \mathsf{C}^0([t_0 - \alpha, t_0 + \alpha]; \mathbb{R}^m)$. This is the existence assertion of the lemma.

(i) Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathsf{B}(r, \boldsymbol{x}_0)$ and let $t \in [t_0 - \alpha, t_0 + \alpha]$. Then

$$\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_1) = \boldsymbol{x}_1 + \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_1))\,\mathrm{d}s, \quad \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_2) = \boldsymbol{x}_2 + \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\,\mathrm{d}s,$$

for all $t \in [t_0 - \alpha, t_0 + \alpha]$. Therefore,

$$\begin{aligned}
\|\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_1) - \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_2)\| &\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + \int_{t_0}^{t} \|\widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_1)) - \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\|\,\mathrm{d}s \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + \int_{t_0}^{t} L(s)\|\Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_1) - \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2)\|\,\mathrm{d}s \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \mathrm{e}^{\int_{t_0}^{t} L(s)\,\mathrm{d}s} \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \mathrm{e}^{\lambda}.
\end{aligned}$$

This shows that $\Phi^{\boldsymbol{F}}_{t,t_0}|\mathsf{B}(r, \boldsymbol{x}_0)$ is Lipschitz, as claimed, when $t \geq t_0$. A similar computation gives the analogous conclusion when $t \leq t_0$.

(ii) Let $t_1, t_2 \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}$ be such that $t_1 \leq t_2$. Just as above, we have

$$\begin{aligned}
\|\Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_1) - \Phi^{\boldsymbol{F}}(t_2, t_0, \boldsymbol{x})\| &\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \\
&\quad + \int_{t_0}^{t_1} \|\widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_1)) - \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\|\,\mathrm{d}s + \int_{t_1}^{t_2} \|\widehat{\boldsymbol{F}}(s, t_0, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\|\,\mathrm{d}s.
\end{aligned}$$

Let $\epsilon \in \mathbb{R}_{>0}$. By Lemma 1 from the proof of Theorem 1.4.8, there exists $\delta_1 \in \mathbb{R}_{>0}$ sufficiently small that, if $|t_2 - t_1| < \delta_1$, then

$$\int_{t_1}^{t_2} \|\widehat{\boldsymbol{F}}(s, t_0, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\|\,\mathrm{d}s < \frac{\epsilon}{2}.$$

Since $\Phi^{\boldsymbol{F}}_{t_1,t_0}$ is continuous, let $\delta_2 \in \mathbb{R}_{>0}$ be sufficiently small that, if $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < \delta_1$, then

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + \int_{t_0}^{t_1} \|\widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_1)) - \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}_2))\|\,\mathrm{d}s < \frac{\epsilon}{2}.$$

Then, if $|t_1 - t_2| < \delta_1$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < \delta_2$,

$$\|\Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_1) - \Phi^{\boldsymbol{F}}(t_2, t_0, \boldsymbol{x}_2)\| < \epsilon,$$

giving the desired conclusion. ▼

The next lemma is a refinement of the preceding one, giving the local version of the theorem statement.

3 Lemma *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}^{\mathrm{n}},$$

and assume that $\mathbf{F}$ *satisfies the conditions of Theorem 1.4.8(ii) for existence and uniqueness of solutions for initial value problems. Then, for each* $(\mathrm{t}_0, \mathbf{x}_0) \in \mathbb{T}\times\mathrm{U}$*, there exists* $\mathrm{r}, \alpha \in \mathbb{R}_{>0}$ *such that*

(i) $(\mathrm{t}, \mathrm{t}_0, \mathbf{x}) \in \mathrm{D}_{\mathbf{F}}$ *for each* $\mathbf{x} \in \mathsf{B}(\mathrm{r}, \mathbf{x}_0)$ *and* $\mathrm{t} \in (\mathrm{t}_0 - \alpha, \mathrm{t}_0 + \alpha) \cap \mathbb{T}$*,*

(ii) the map

$$(\mathrm{t}_0 - \alpha, \mathrm{t}_0 + \alpha) \times \mathsf{B}(\mathrm{r}, \mathbf{x}_0) \ni (\mathrm{t}, \mathbf{x}) \mapsto \Phi^{\mathbf{F}}(\mathrm{t}, \mathrm{t}_0, \mathbf{x})$$

is continuous, and

(iii) the map

$$\mathsf{B}(\mathrm{r}, \mathbf{x}_0) \ni \mathbf{x} \mapsto \Phi^{\mathbf{F}}_{\mathrm{t},\mathrm{t}_0}(\mathbf{x}) \in \mathbb{R}^{\mathrm{m}}$$

is a bi-Lipschitz homeomorphism onto its image for every $\mathrm{t} \in (\mathrm{t}_0 - \alpha, \mathrm{t}_0 + \alpha) \cap \mathbb{T}$*.*

Proof Let r', α' be as in Lemma 2 and let $r \in (0, r']$ and $\alpha \in (0, \alpha']$ be such that

$$\Phi^F_{t,t_0}(\boldsymbol{x}) \in \mathsf{B}(r', \boldsymbol{x}_0), \qquad \boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0),\ t \in [t_0 - \alpha, t_0 + \alpha],$$

this being possible by Lemma 2(ii). Let $t \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}$ and denote

$$V = \Phi^F_{t,t_0}(\mathsf{B}(r, \boldsymbol{x}_0)) \subseteq \mathsf{B}(r', \boldsymbol{x}_0).$$

Let $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$. Since $\boldsymbol{y} \triangleq \Phi^F_{t,t_0}(\boldsymbol{x}) \in \mathsf{B}(r', \boldsymbol{x}_0)$ and $t \in [t_0 - \alpha', t_0 + \alpha'] \cap \mathbb{T}$, there exists $\rho \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{y}' \in \mathsf{B}(\rho, \boldsymbol{y})$, then $(t_0, t, \boldsymbol{y}') \in D_F$. Moreover, since $\Phi^F_{t_0,t}$ is continuous (indeed, Lipschitz, with Lipschitz constant e^{λ}, with λ as in the proof of Lemma 2) and $\Phi^F_{t_0,t}(\boldsymbol{y}) = \boldsymbol{x}$, we may choose ρ sufficiently small that $\Phi^F_{t_0,t}(\boldsymbol{y}') \in \mathsf{B}(r, \boldsymbol{x}_0)$ if $\boldsymbol{y}' \in \mathsf{B}(\rho, \boldsymbol{y})$. By Lemma 2, $\Phi^F_{t_0,t}|\mathsf{B}(\rho, \boldsymbol{y})$ is Lipschitz with Lipschitz constant e^{λ}. Thus there is a neighbourhood of $\boldsymbol{x}$ on which the restriction of $\Phi^F_t t_0$ is invertible, Lipschitz, and with a Lipschitz inverse. ▼

We now need to show that the theorem holds globally. To this end, let $(t_0, \boldsymbol{x}_0) \in \mathbb{T} \times U$ and denote by $J_+(t_0, \boldsymbol{x}_0) \subseteq \mathbb{T}$ the set of $b > t_0$ such that, for each $b' \in [t_0, b)$, there exists a relatively open interval $J \subseteq \mathbb{T}$ and a $r \in \mathbb{R}_{>0}$ such that

1. $b' \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r, \boldsymbol{x}_0) \subseteq D_F$,
3. $J \times \mathsf{B}(r, \boldsymbol{x}_0) \ni (t, \boldsymbol{x}) \mapsto \Phi^F(t, t_0, \boldsymbol{x}) \in U$ is continuous, and
4. for each $t \in J$, $\mathsf{B}(r, \boldsymbol{x}_0) \ni \boldsymbol{x} \mapsto \Phi^F(t, t_0, \boldsymbol{x})$ is a locally bi-Lipschitz homeomorphism onto its image.

By Lemma 3, $J_+(t_0, \boldsymbol{x}_0) \neq \emptyset$. We then consider two cases.

The first case is $J_+(t_0, \boldsymbol{x}_0) \cap [t_0, \infty) = \mathbb{T} \cap [t_0, \infty)$. In this case, for each $t \in \mathbb{T}$ with $t \geq t_0$, there exists a relatively open interval $J \subseteq \mathbb{T}$ and $r \in \mathbb{R}_{>0}$ such that

1. $t \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r, x_0) \subseteq D_F$,
3. $J \times \mathsf{B}(r, x_0) \ni (\tau, x) \mapsto \Phi^F(\tau, t_0, x) \in U$ is continuous, and
4. for each $\tau \in J$, $\mathsf{B}(r, x_0) \ni x \mapsto \Phi^F(\tau, t_0, x)$ is a locally bi-Lipschitz homeomorphism onto its image.

The second case is $J_+(t_0, x_0) \cap [t_0, \infty) \subset \mathbb{T} \cap [t_0, \infty)$. In this case we let $t_1 = \sup J_+(t_0, x_0)$ and note that $t_1 \neq \sup \mathbb{T}$. We claim that $t_1 \in J_F(t_0, x_0)$. Were this not the case, then we must have $b \triangleq \sup J_F(t_0, x_0) < t_1$. Since $b \in J_+(t_0, x_0)$, there must be a relatively open interval $J \subseteq \mathbb{T}$ containing b such that $t \in J_F(t_0, x_0)$ for all $t \in J$. But, since there are t's in J larger than b, this contradicts the definition of $J_F(t_0, x_0)$, and so we conclude that $t_1 \in J_F(t_0, x_0)$. Let us denote $x_1 = \Phi^F(t_1, t_0, x_0)$. By Lemma 3, there exists $\alpha_1, r_1 \in \mathbb{R}_{>0}$ such that $(t, t_1, x) \in D_F$ for every $t \in (t_1 - \alpha_1, t_1 + \alpha_1)$ and $x \in \mathsf{B}(r_1, x_1)$, and such that the map

$$(t_1 - \alpha_1, t_1 + \alpha_1) \times \mathsf{B}(r_1, x_1) \ni (t, x) \mapsto \Phi^F(t, t_1, x)$$

is continuous, and the map

$$\mathsf{B}(r_1, x_1) \ni x \mapsto \Phi^F(t, t_1, x)$$

is a locally bi-Lipschitz homeomorphism onto its image for every $t \in (t_1 - \alpha_1, t_1 + \alpha_1)$. Since $t \mapsto \Phi^F(t, t_0, x_0)$ is continuous and $\Phi^F(t_1, t_0, x_0) = x_1$, let $\delta \in \mathbb{R}_{>0}$ be such that $\delta < \frac{\alpha_1}{2}$ and $\Phi^F(t, t_0, x_0) \in \mathsf{B}(r_1/4, x_1)$ for $t \in (t_1 - \delta, t_1)$. Now let $\tau_1 \in (t_1 - \delta, t_1)$ and, by our hypotheses on t_1, there exists an open interval J and $r_1' \in \mathbb{R}_{>0}$ such that

1. $\tau_1 \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r_1', x_0) \subseteq D_F$,
3. $J \times \mathsf{B}(r_1', x_0) \ni (\tau, x) \mapsto \Phi^F(\tau, t_0, x) \in U$ is continuous, and
4. for each $\tau \in J$, $\mathsf{B}(r_1', x_0) \ni x \mapsto \Phi^F(\tau, t_0, x)$ is a locally bi-Lipschitz homeomorphism onto its image.

We also choose J and r_1' sufficiently small that

$$\{\Phi^F(t, t_0, x) \mid t \in J,\ x \in \mathsf{B}(r_1', x_0)\} \subseteq \mathsf{B}(r_1/2, x_1).$$

Now we claim that

$$(\tau_1 - \alpha_1, \tau_1 + \alpha_1) \times \{t_0\} \times \mathsf{B}(r_1', x_0) \subseteq D_F.$$

We first show that

$$[\tau_1, \tau_1 + \alpha_1) \times \{t_0\} \times \mathsf{B}(r_1', x_0) \subseteq D_F. \tag{1.40}$$

Indeed, we have $(\tau_1, t_0, x) \in D_F$ for every $x \in \mathsf{B}(r_1', x_0)$ since $\tau_1 \in J$. By definition of J, $\Phi^F(\tau_1, t_0, x) \in \mathsf{B}(r_1/2, x_1)$. By definition of τ_1, $t_1 - \tau_1 < \delta < \frac{\alpha_1}{2}$. Then, by definition of α_1 and r_1,

$$(t_1, \tau_1, \Phi^F(\tau_1, t_0, x)) \in D_F$$

for every $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. From this we conclude that $(t_1, t_0, \boldsymbol{x}) \in D_F$ for every $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. Now, since

$$t \in [\tau_1, \tau_1 + \alpha_1) \quad \Longrightarrow \quad t \in (t_1 - \alpha_1, t_1 + \alpha_1),$$

we have $(t, t_1, \Phi^F(t, t_1, \boldsymbol{x})) \in D_F$ for every $t \in (\tau_1 - \alpha_1, \tau_1 + \alpha_1)$ and $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. Since

$$\Phi^F(t, t_1, \Phi^F(t_1, t_0, \boldsymbol{x})) = \Phi^F(t, t_0, \boldsymbol{x}),$$

we conclude (1.40). A similar but less complicated argument gives

$$(\tau_1 - \alpha_1, \tau_1) \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_F.$$

Now we claim that the map

$$(\tau_1 - \alpha_1, \tau_1 + \alpha_1) \times \mathsf{B}(r_1', \boldsymbol{x}_0) \ni (t, \boldsymbol{x}) \mapsto \Phi^F(t, t_0, \boldsymbol{x})$$

is continuous. This map is continuous at

$$(t, \boldsymbol{x}) \in (\tau_1 - \alpha_1, \tau_1] \times \mathsf{B}(r_1', \boldsymbol{x}_0)$$

by definition of τ_1. For $t \in (\tau_1, \tau_1 + \alpha_1)$ we have

$$\Phi^F(t, t_0, \boldsymbol{x}) = \Phi^F(t, \tau_1, \Phi^F(\tau_1, t_0, \boldsymbol{x})),$$

and continuity in this case follows since compositions of continuous maps are continuous.

Next we claim that the map

$$\mathsf{B}(r_1', \boldsymbol{x}_0) \ni \boldsymbol{x} \mapsto \Phi^F(t, t_0, \boldsymbol{x})$$

is a locally bi-Lipschitz homeomorphism onto its image for every $t \in (\tau_1 - \alpha_1, \tau_1 + \alpha_1)$. By definition of τ_1, the map

$$\Phi^F_{t,t_0} \colon \mathsf{B}(r_1', \boldsymbol{x}_0) \to \mathsf{B}(r_1/2, \boldsymbol{x}_1)$$

is a locally bi-Lipschitz homeomorphism onto its image for $t \in (\tau_1 - \alpha_1, \tau_1]$. We also have that

$$\Phi^F_{t,\tau_1} \colon \mathsf{B}(r_1, \boldsymbol{x}_1) \to U$$

is a locally bi-Lipschitz homeomorphism onto its image for $t \in (\tau_1, \tau_1 + \alpha_1)$. Since the composition of locally bi-Lipschitz homeomorphisms onto their image is a locally bi-Lipschitz homeomorphisms onto its image, our assertion follows.

By our above arguments, we have an open interval J' and $r_1' \in \mathbb{R}_{>0}$ such that

1. $t_1 \in J'$,
2. $J' \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_F$,
3. $J' \times \mathsf{B}(r_1', \boldsymbol{x}_0) \ni (t, \boldsymbol{x}) \mapsto \Phi^F(t, t_0, \boldsymbol{x}) \in U$ is continuous, and

4. for each $t \in J'$, $\mathsf{B}(r_1', x_0) \ni x \mapsto \Phi^F(t, t_0, x)$ is a locally bi-Lipschitz homeomorphism onto its image.

This contradicts the fact that $t_1 = \sup J_+(t_0, x_0)$ and so the condition

$$J_+(t_0, x_0) \cap [t_0, \infty) \subset \mathbb{T} \cap [t_0, \infty)$$

cannot obtain.

One similarly shows that it must be the case that $J_-(t_0, x_0) \cap (-\infty, t_0] = \mathbb{T} \cap (-\infty, t_0]$; where $J_-(t_0, x_0)$ has the obvious definition.

Finally, we note that Φ^F_{t,t_0} injective by uniqueness of solutions for $\boldsymbol{F}$. Now, assertions (i)–(vi) of the theorem now follow since the notions of "continuous" and "locally bi-Lipschitz homeomorphism" can be tested locally, i.e., in a neighbourhood of any point.

We shall prove assertions (vii) and (viii) together. We let $(t_1, t_0, x_0) \in D_{\boldsymbol{F}}$. As above, there exists $r_1, \alpha_1 \in \mathbb{R}_{>0}$ such that

$$(t_1 - \alpha_1, t_1 + \alpha_1) \cap \mathbb{T} \times \{t_0\} \times \mathsf{B}(r_1, x_0) \subseteq D_{\boldsymbol{F}},$$

and the map $(t, x) \mapsto \Phi^F(t, t_0, x_0)$ is continuous on this domain. We claim that the map

$$(t, x) \mapsto \Phi^F(t_0, t, x) \tag{1.41}$$

is continuous for (t, x) nearby (t_0, x_0). To see this, we proceed rather as in the proof of Theorem 1.4.8, using the Contraction Mapping Theorem.

Let $r \in \mathbb{R}_{>0}$ be such that there exists a locally integrable $g \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, x)\| \leq g(t), \qquad (t, x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0),$$

and also there exists a locally integrable $L \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, x) - \widehat{\boldsymbol{F}}(t, y)\| \leq L(t)\|x - y\|$$

for all $t \in \mathbb{T}$ and $x, y \in \overline{\mathsf{B}}(r, x_0)$. Let us choose $\lambda \in (0, 1)$. Let us suppose that $t \leq t_0$. Define $G_-, \ell_- \colon (-\infty, t_0] \cap \mathbb{T} \to \mathbb{R}$ by

$$G_-(t) = \int_t^{t_0} g(s)\,\mathrm{d}s, \quad \ell_+(t) = \int_t^{t_0} L(s)\,\mathrm{d}s.$$

Since g and L are nonnegative, we can choose $T_- \in \mathbb{R}_{>0}$ such that

$$G_-(t) = \int_t^{t_0} g(s)\,\mathrm{d}s \leq \frac{r}{2}, \quad \ell_-(t) = \int_t^{t_0} L(s)\,\mathrm{d}s < \lambda$$

for $t \in [t_0 - T_-, t_0]$. For $x \in \mathsf{B}(r/2, x_0)$, let ξ_0 be the trivial function $t \mapsto x, t \in [t_0 - T_-, t_0]$, and let $\overline{\mathsf{B}}_-(r, \xi_0)$ be the ball of radius r and centre ξ_0 in $\mathsf{C}^0([t_0 - T_-, t_0]; \mathbb{R}^m)$. Define $F_- \colon \overline{\mathsf{B}}_-(r, \xi_0) \to \mathsf{C}^0([t_0 - T_-, t_0]; \mathbb{R}^m)$ by

$$F_-(\xi)(t) = x + \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \xi(s))\,\mathrm{d}s.$$

By the lemma from the proof of Theorem 1.4.8, $s \mapsto \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s))$ is locally integrable, showing that F_- is well-defined and that $F_-(\boldsymbol{\xi})$ is absolutely continuous.

We claim that $F_-(\overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)) \subseteq \overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$. Suppose that $\boldsymbol{\xi} \in \overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$ so that

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq r, \qquad t \in [t_0 - T_-, t_0].$$

Then, for $t \in [t_0 - T_-, t_0]$,

$$\begin{aligned}\|F_-(\boldsymbol{\xi})(t) - \boldsymbol{x}_0\| &\leq \|\boldsymbol{x} - \boldsymbol{x}_0\| + \left\| \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) \, \mathrm{d}s \right\| \\ &\leq \frac{r}{2} + \int_t^{t_0} \|\widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s))\| \, \mathrm{d}s \leq \frac{r}{2} + \int_t^{t_0} g(s) \, \mathrm{d}s \leq r,\end{aligned}$$

as desired.

We claim that $F_-|\overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$ is a contraction mapping. That is, we claim that there exists $\rho \in [0, 1)$ such that

$$\|F_-(\boldsymbol{\xi}) - F_-(\boldsymbol{\eta})\|_\infty \leq \rho \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty$$

for every $\boldsymbol{\xi}, \boldsymbol{\eta} \in \overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$. Indeed, let $\boldsymbol{\xi}, \boldsymbol{\eta} \in \overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$ and compute, for $t \in [t_0 - T_-, t_0]$,

$$\begin{aligned}\|F_-(\boldsymbol{\xi})(t) - F_-(\boldsymbol{\eta})(t)\| &= \left\| \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) \, \mathrm{d}s - \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \boldsymbol{\eta}(s)) \right\| \\ &\leq \int_t^{t_0} \|\widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}(s)) - \widehat{\boldsymbol{F}}(s, \boldsymbol{\eta}(s))\| \, \mathrm{d}s \\ &\leq \int_t^{t_0} L(s) \|\boldsymbol{\xi}(s) - \boldsymbol{\eta}(s)\| \, \mathrm{d}s \leq \ell_-(t) \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty \leq \lambda \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty, \quad (1.42)\end{aligned}$$

since $\boldsymbol{\xi}(s), \boldsymbol{\eta}(s) \in \mathsf{B}(r, \boldsymbol{x}_0)$ for every $s \in [t_0, t_0 + T_+]$. This proves that $F_-|\overline{\mathsf{B}}_-(r, \boldsymbol{\xi}_0)$ is a contraction mapping.

By the Contraction Mapping Theorem***missing stuff*** there exists a unique fixed point for F_- which we denote by $\boldsymbol{\xi}_-$. Thus

$$\boldsymbol{\xi}_-(t) = F_-(\boldsymbol{\xi}_+)(t) = \boldsymbol{x} + \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \boldsymbol{\xi}_-(s)) \, \mathrm{d}s.$$

Differentiating the first and last expressions with respect to t shows that $\boldsymbol{\xi}_+$ is a solution for $\boldsymbol{F}$, and we moreover have $\boldsymbol{\xi}(t_0) = \boldsymbol{x}$. This show that, if $\boldsymbol{x} \in \mathsf{B}(r/2, \boldsymbol{x}_0)$ and $t \in [t_0 - T_-, t_0]$, then we have $\Phi^{\boldsymbol{F}}(t_0, t, \boldsymbol{x}) \in \mathsf{B}(r, \boldsymbol{x}_0)$ and

$$\Phi^{\boldsymbol{F}}(t_0, t, \boldsymbol{x}) = \boldsymbol{x} + \int_t^{t_0} \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(t_0, s, \boldsymbol{x}) \, \mathrm{d}s.$$

A similar argument, of course, can be fabricated for $t \geq t_0$, and we conclude that there exists $\alpha_0 \in \mathbb{R}_{>0}$ and $r_0 \in \mathbb{R}_{>0}$ such that

$$\Phi^F(t_0, t, \boldsymbol{x}) \in \mathsf{B}(r_1, \boldsymbol{x}_0), \qquad (t, \boldsymbol{x}) \in (t_0 - \alpha_0, t_0 + \alpha_0) \cap \mathbb{T} \times \mathsf{B}(r_0, \boldsymbol{x}_0).$$

Finally, we show that the map (1.41) is continuous on $(t_0-\alpha_0, t_0+\alpha_0)\cap\mathbb{T}\times\mathsf{B}(r_0, \boldsymbol{x}_0)$. Note that, as in the proof of Lemma 2 above and assuming that $\tau_2 \geq \tau_1$,

$$\begin{aligned}\|\Phi^F(t_0, \tau_1, \boldsymbol{x}_1) - \Phi^F(t_0, \tau_2, \boldsymbol{x}_2)\| &\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \\ &+ \int_{\tau_2}^{t_0} \|\widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_1)) - \widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_2))\| \, \mathrm{d}s + \int_{\tau_1}^{\tau_2} \|\widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_1))\| \, \mathrm{d}s.\end{aligned}$$

Let $\epsilon \in \mathbb{R}_{>0}$. By Lemma 1 from the proof of Theorem 1.4.8, there exists $\delta_1 \in \mathbb{R}_{>0}$ sufficiently small that, if $|\tau_2 - \tau_1| < \delta_1$, then

$$\int_{\tau_1}^{\tau_2} \|\widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_2))\| \, \mathrm{d}s < \frac{\epsilon}{2}.$$

Since $\Phi^F_{t_0,\tau_2}$ is continuous, let $\delta_2 \in \mathbb{R}_{>0}$ be sufficiently small that, if $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < \delta_2$, then

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + \int_{\tau_2}^{t_0} \|\widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_1)) - \widehat{\boldsymbol{F}}(s, \Phi^F(t_0, s, \boldsymbol{x}_2))\| \, \mathrm{d}s < \frac{\epsilon}{2}.$$

Then, if $|t_1 - t_2| < \delta_1$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < \delta_2$,

$$\|\Phi^F(t_0, \tau_1, \boldsymbol{x}_1) - \Phi^F(t_0, \tau_2, \boldsymbol{x}_2)\| < \epsilon,$$

given the desired continuity.

Finally, if $(t', t_0', \boldsymbol{x}) \in (t - \alpha, t + \alpha) \cap \mathbb{T} \times (t_0 - \alpha_0, t_0 + \alpha_0) \cap \mathbb{T} \times \mathsf{B}(r_0, \boldsymbol{x}_0)$, then

$$\Phi^F(t', t_0, \Phi^F(t_0, t_0', \boldsymbol{x})) = \Phi^F(t', t_0', \boldsymbol{x}),$$

which shows both that D_F is open and that Φ^F is continuous, since the composition of continuous mappings is continuous.

(ix) Let $T_+ = \sup J_F(t_0, \boldsymbol{x}_0)$. Then $(T_+ - \epsilon, t_0, \boldsymbol{x}_0) \in D_F$. Since D_F is open, there exists $r \in \mathbb{R}_{>0}$ such that

$$\{T_+ - \tfrac{\epsilon}{2}\} \times (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T} \times \mathsf{B}(r, \boldsymbol{x}_0) \subseteq D_F.$$

In other words, $[t_0, T_+ - \frac{\epsilon}{2}] \subseteq J_F(t, \boldsymbol{x})$ for every $(t, \boldsymbol{x}) \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T} \times \mathsf{B}(r, \boldsymbol{x}_0)$. Thus, for such $(t, \boldsymbol{x})$,

$$\sup J_F(t, \boldsymbol{x}) \geq T_- - \tfrac{\epsilon}{2} > T_- - \epsilon = \sup J_F(t_0, \boldsymbol{x}_0) - \epsilon,$$

as claimed. A similar argument holds for the left endpoint of intervals of convergence. ■

1.4.2 Existence and uniqueness of solutions for partial differential equations. . . NOT!!

The questions of existence and uniqueness of solutions for partial differential equations is far more difficult than for ordinary differential equations. Situations range from relatively simple cases where one can prove existence and uniqueness directly by writing down solutions, to equations where proving an existence and uniqueness result becomes a triumph of analysis, resulting in a paper in the *Annals of Mathematics*. Thus it is not possible to have a comprehensive discussion of a theory of existence and uniqueness for general partial differential equations. Instead we content ourselves with some mostly vague observations about the nature of the problem.

First we note that all of the examples of Section 1.4.1 can be turned into partial differential equations is an entirely artificial way, merely by artificially adding an extra independent variable. This is not an interesting thing to do, except that it ensures that all of the conclusions 1–4 enumerated after these examples equally apply to partial differential equations.

Let us list some of the difficulties that arise in arriving at existence and uniqueness theorems for partial differential equations.

1. For ordinary differential equations, we saw that appropriate combinations of continuity, boundedness, and Lipschitz hypotheses ensured existence, and often uniqueness, of solutions. For partial differential equations, this is no longer true. A partial differential equation with lots of nice properties can fail to have any solutions. Moreover, this failure of solutions to exist can arise in various ways. So any attempt at a general theorem is dead from the start, and one must make assumptions on the sort of partial differential equation for solutions to even exist, cf. the discussion of elliptic, hyperbolic, and parabolic equations in Section 1.3.4.
2. For ordinary differential equations, we saw that to uniquely prescribe a solution one must specify an initial value of the state at some time to arrive at an initial value problem. For partial differential equations, this process is more difficult. Typically one must specify values of the solution along some surface or some such thing. This is known as prescribing "Cauchy data." However, the type of Cauchy data that is to be specified is not as easy a matter to understand as for ordinary differential equations. For many problems arising from physics, e.g., the heat, wave, and potential equations, the "natural" prescriptions of values for the solution and/or its derivatives at "boundaries" of the domain is often correct, as we shall see when we study these problems subsequently.***missing stuff*** However, these partial differential equations are "nice." In general, finding the analogue of initial conditions for ordinary differential equations is quite hard for partial differential equations.
3. The properties of a solution of an ordinary differential equation as it depends on the independent variable are quite easy: it is of class C^1. For partial differential

equations, finding the right attributes for a solution beforehand is often crucial to proving existence and uniqueness theorems for an equation.

We shall say nothing more about the subject of existence and uniqueness theorems for partial differential equations, except to say this:

Go to

http://www.claymath.org/millennium-problems/navier-stokes-equation

to win $1,000,000! •

Exercises

1.4.1 Which of the following maps are locally Lipschitz?

(a) $f\colon \mathbb{R} \to \mathbb{R}$
$x \mapsto \sqrt{|x|}$;

(b) $f\colon \mathbb{R}_{>0} \to \mathbb{R}$
$x \mapsto \sqrt{|x|}$;

(c) $f\colon \mathbb{R} \to \mathbb{R}$
$x \mapsto |x|$;

(d) $f\colon [0,\pi] \to \mathbb{R}$
$x \mapsto \sin(x)$;

(e) $f\colon \mathbb{R}_{>0} \to \mathbb{R}$
$x \mapsto x^{-1}$.

1.4.2 For the ordinary differential equations F with right-hand sides $\widehat{F}$ as given, determine which, if either, of the parts of Theorem 1.4.8 apply, and indicate what conclusions, if any, you can make about existence and uniqueness of solutions for F. Here are the right-hand sides:

(a) $\widehat{F}\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$
$(t,x) \mapsto \sqrt{t}x$;

(b) $\widehat{F}\colon \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$
$(t,x) \mapsto \dfrac{x}{t}$;

(c) $\widehat{F}\colon \mathbb{R} \times [0,1] \to \mathbb{R}$
$(t,x) \mapsto \begin{cases} 1, & x \in [0,\frac{1}{2}], \\ -1, & x \in (\frac{1}{2},1]; \end{cases}$

(d) $\widehat{F}\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$
$(t,x) \mapsto |xt|$;

(e) $\widehat{F}\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$
$(t,x) \mapsto x^2$.

1.4.3 For the ordinary differential equations of Examples 1.3.3–1 to 9, show that the hypotheses of Theorem 1.4.8 hold, and so these equations possess unique solutions, at least for small times around any initial time.

1.4.4 In each of Examples 1.4.1–1.4.5, state the hypotheses of Theorem 1.4.8 that are violated by the example.

1.4.5 Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n$$

and suppose that, for each $\boldsymbol{x}_0 \in U$, there exist $M, r \in \mathbb{R}_{>0}$ such that

$$\left|\frac{\partial \widehat{F}_j}{\partial x_k}(t, \boldsymbol{x})\right| \leq M, \qquad j, k \in \{1, \dots, n\},\ (t, \boldsymbol{x}) \in \mathbb{T} \times \mathsf{B}(r, \boldsymbol{x}_0).$$

Show that

$$pderivt\Phi^{\boldsymbol{F}}(t_0, t, \boldsymbol{x}) + \sum_{j=1}^{n} \widehat{F}_j(t, \boldsymbol{x}) \frac{\partial}{\partial x_j} \Phi^{\boldsymbol{F}}(t_0, t, \boldsymbol{x}) = 0.$$

Chapter 2

Scalar ordinary differential equations

In this chapter, we begin our studies in earnest, doing what one does with differential equations: where possible, solve them and/or understand the nature of their solutions or sets of solutions. We shall study ordinary differential equations with a single state and arbitrary order. Thus, in the notation of Section 1.3.3, we consider an ordinary differential equation with time domain $\mathbb{T} \subseteq \mathbb{R}$, state space $U \subseteq \mathbb{R}$, and with right-hand side

$$\widehat{F} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

that gives an equation

$$\frac{\mathrm{d}^k \xi}{\mathrm{d}t^k}(t) = \widehat{F}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t)\right)$$

that must be satisfied by solutions $t \mapsto \xi(t)$.

There is not much one can say in any generality about such an equation, except to say that we can use Theorem 1.4.8 to assert the existence and uniqueness of solutions, at least for small times (making use of Exercise 1.3.23). Thus we focus in this chapter on special equations for which one *can* say something useful. In Section 2.1 we consider a very special class of first-order equations which can, in some sense, be solved. In Sections 2.2 and 2.3 we consider linear differential equations, first homogeneous equations then inhomogeneous equations.

Contents

Section 2.1

Separable first-order scalar equations

In this short section we consider a very special class of first-order scalar differential equation, one that can sometimes be solved explicitly. The following definition encodes what we are after.

2.1.1 Definition (Separable scalar differential equation) A differential equation $F\colon \mathbb{T}\times U\times \mathrm{L}^1_{\mathrm{sym}}(\mathbb{R};\mathbb{R})\to\mathbb{R}$ is ***separable*** if it has the form

$$\boldsymbol{F}(t,x,x^{(1)}) = f_1(x)x^{(1)} - f_0(t). \qquad \bullet$$

We note that a separable differential equation is an ordinary differential equation if and only if $f_1(x)$ is nonzero for every $x\in U$, because in this case we can solve for $x^{(1)}$ for a given $(t,x)\in\mathbb{T}\times U$ by

$$x^{(1)} = \frac{f_0(t)}{f_1(x)} = \widehat{F}(t,x).$$

Note that $t\mapsto x(t)$ is a solution to a separable differential equation if

$$f_1(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}(t) = f_0(t), \qquad x(t_0)=t_0$$

for some $(t_0,x_0)\in\mathbb{T}\times U$. There is a naïve way to "solve" such as equation. First do some (a priori meaningless) manipulations:

$$f_1(x)\frac{\mathrm{d}x}{\mathrm{d}t} = f_0(t) \quad\implies\quad \int_{x_0}^{x(t)} f_1(\xi)\,\mathrm{d}\xi = \int_{t_0}^{t} f_0(\tau)\,\mathrm{d}\tau.$$

If F_1 and F_0 are antiderivatives of f_1 and f_0, respectively, we have

$$F_1(x(t)) - F_1(x_0) = F_0(t) - F_0(t_0).$$

This is an equation that you pray you can solve for $x(t)$.

This naïve procedure does, in fact, work, as the following result indicates.

2.1.2 Proposition (Solutions for separable differential equations) *Let $\mathbb{T}\subseteq\mathbb{R}$ be a time-domain, let $\mathrm{U}\subseteq\mathbb{R}$ be an open set, let $\mathrm{f}_0\colon\mathbb{T}\to\mathbb{R}$ and $\mathrm{f}_1\colon\mathrm{U}\to\mathbb{R}$ be continuous functions for which $\mathrm{f}_1(\mathrm{x})\neq 0$ for every $\mathrm{x}\in\mathrm{U}$. Let F_0 and F_1 be antiderivatives of f_0 and f_1, respectively. Let $(\mathrm{t}_0,\mathrm{x}_0)\in\mathbb{T}\times\mathrm{U}$. Then the following statements hold:*

(i) if $\mathbb{T}'\subseteq\mathbb{T}$ is a subinterval containing t_0 and if a class C^1-function $\xi\colon\mathbb{T}'\to\mathrm{U}$ satisfies

$$\mathrm{F}_1(\xi(\mathrm{t})) - \mathrm{F}_1(\mathrm{x}_0) = \mathrm{F}_0(\mathrm{t}) - \mathrm{F}_0(\mathrm{t}_0), \qquad \mathrm{t}\in\mathbb{T}',$$

then ξ is a solution to the separable ordinary differential equation

$$F(t, x, x^{(1)}) = f_1(x)x^{(1)} - f_0(t)$$

satisfying the initial condition $\xi(t_0) = x_0$;

(ii) if there exists a subinterval $\mathbb{T}' \subseteq \mathbb{T}$ and a solution $\xi\colon \mathbb{T}' \to U$ to F satisfying $\xi(t_0) = x_0$, then

$$F_1(\xi(t)) - F_1(x_0) = F_0(t) - F_0(t_0), \qquad t \in \mathbb{T}'.$$

Proof (i) Let us define

$$G\colon \mathbb{T} \times U \to \mathbb{R}$$

by

$$G(t, x) = F_1(x) - F_1(x_0) - F_0(t) + F_0(t_0),$$

noting that $G(t, \xi(t)) = 0$. Note that G is of class C^1 and that

$$\frac{\partial G}{\partial x}(t, \xi(t)) \neq 0, \qquad t \in \mathbb{T}'$$

Thus, by the Implicit Function Theorem,***missing stuff*** there exists a relatively open interval $\mathbb{T}'_t \subseteq \mathbb{T}'$ containing t and a unique map $\xi_t\colon \mathbb{T} \to U$ of class C^1 such that $\xi_t(t) = \xi(t)$ and that $G(\tau, \xi_t(\tau)) = 0$ for all $\tau \in \mathbb{T}'_t$. Therefore, by the Chain Rule,

$$0 = \frac{\mathrm{d}}{\mathrm{d}\tau} G(\tau, \xi_t(\tau)) = \frac{\mathrm{d}}{\mathrm{d}\tau}(F_1(\xi_t(\tau)) - F_1(x_0) - F_0(\tau) + F_0(t_0)) = f_1(\xi_t(\tau))\dot{\xi}_t(\tau) - f_0(\tau),$$

giving ξ_t as a solution to $\boldsymbol{F}$.

It remains to show that $\xi(\tau) = \xi_t(\tau)$ for every $t \in \mathbb{T}'$ and every $\tau \in \mathbb{T}'_t$. Let $\mathbb{T}'' \subset \mathbb{T}'$ be the largest subinterval such that $\xi(\tau) = \xi_t(\tau)$ for every $t \in \mathbb{T}''$ and every $\tau \in \mathbb{T}'_t$. We claim that $\mathbb{T}'' = \mathbb{T}'$. We need only show that $\mathbb{T}' \subseteq \mathbb{T}''$. Let $t \in \mathbb{T}'$. By construction, we have $\xi_t(t) = \xi(t)$. Note that, for every $\tau \in \mathbb{T}'_t$ we have $G(\tau, \xi(\tau)) = 0$. Moreover, $\xi|\mathbb{T}'_t$ is of class C^1. Thus the uniqueness part of the Implicit Function Theorem gives $\xi_t(\tau) = \xi(\tau)$ for all $\tau \in \mathbb{T}'_t$. Therefore, $t \in \mathbb{T}''$. From this we conclude that, indeed $\xi(\tau) = \xi_t(\tau)$ for every $\tau \in \mathbb{T}'_t$, and this shows that ξ is a solution for F, since ξ_t is a solution for F.

(ii) We have, for all $t \in \mathbb{T}'$,

$$\begin{aligned} & f_1(\xi(t))\dot{\xi}(t) - f_0(t) = 0 \\ \implies \quad & \frac{\mathrm{d}}{\mathrm{d}t}(F_1(\xi(t)) - F_0(t)) = 0 \\ \implies \quad & F_1(\xi(t)) - F_1(x_0) - F_0(t) + F_0(t_0) \end{aligned}$$

since ξ is continuous, and using the Fundamental Theorem of Calculus. ■

Now let us look at some examples.

2.1.3 Examples (Separable ordinary differential equations)

1. Consider the ordinary differential equation

$$F(t, x, x^{(1)}) = x^{(1)} - ax$$

for $a \in \mathbb{R}$, which is defined for $(t, x) \in \mathbb{R} \times \mathbb{R}$, i.e., $\mathbb{T} = \mathbb{R}$ and $U = \mathbb{R}$. Solutions of this differential equation satisfy

$$\dot{x}(t) = ax(t).$$

This is not immediately in the form of a separable equation, but it can be converted into the separable equation

$$\tilde{F}(t, x, x^{(1)}) = \frac{x^{(1)}}{x} - a,$$

but only at the cost of limiting the state space to be $\tilde{U} = \mathbb{R} \setminus \{0\}$. But let us do this and see what happens. We have $f_1(x) = x^{-1}$ and $f_0(t) = a$ and so $F_1(x) = \ln(|x|)$ and $F_0(t) = at$. Thus, by Proposition 2.1.2, a solution $t \mapsto \xi(t)$ with values in $\tilde{U}$ will satisfy

$$\begin{aligned}
&\ln(|\xi(t)|) - \ln(|\xi(t_0)|) = a(t - t_0)\\
\iff\ &\ln\left(\left|\frac{\xi(t)}{\xi(t_0)}\right|\right) = a(t - t_0)\\
\iff\ &\left|\frac{\xi(t)}{\xi(t_0)}\right| = \mathrm{e}^{a(t-t_0)}\\
\iff\ &|\xi(t)| = |\xi(t_0)|\mathrm{e}^{a(t-t_0)}.
\end{aligned}$$

Now, since ξ must be of class C^1, in particular continuous, it follows that the sign of $\xi(t)$ must be the same as that of $\xi(t_0)$, and so we have

$$\xi(t) = \xi(t_0)\mathrm{e}^{a(t-t_0)}.$$

Note that this only applies when $\xi(t_0) \neq 0$. However, if $\xi(t_0) = 0$ then we immediately have the solution as $\xi(t) = 0$ for all t.
We will encounter this differential as a special case of various other sorts of differential equations in the sequel.

2. Next we consider the differential equation

$$F(t, x, x^{(1)}) = x^{(1)} - x^2$$

with $(t, x) \in \mathbb{R} \times \mathbb{R}$ that we initially investigated in Example 1.4.4. Again, this equation is not in the form of a separable ordinary differential equation, but can be converted into the separable equation

$$\tilde{F}(t, x, x^{(1)}) = \frac{x^{(1)}}{x^2} - 1$$

with $f_0(x) = x^{-2}$ and $f_1(t) = 1$. Again, in making this conversion, we must restrict our state to be in $\tilde{U} = \mathbb{R} \setminus \{0\}$. We then have

$$F_1(x) = -x^{-1}, \quad F_0(t) = t.$$

Therefore, skipping the details, a solution $t \mapsto \xi(t)$ satisfies

$$-\frac{1}{\xi(t)} + \frac{1}{\xi(t_0)} = t - t_0 \quad \Longrightarrow \quad \xi(t) = \frac{\xi(0)}{\xi(t_0)(t_0 - t) + 1},$$

just as in Example 1.4.4. As we saw in this previous example, the solution cannot be defined on the entire time interval $\mathbb{R}$. Also, we can recover the solution with the initial condition $\xi(t_0) = 0$ by noting that, in this case, the solution is $\xi(t) = 0$.

3. Here we consider the differential equation

$$F(t, x, x^{(1)}) = x^{(1)} - x^{1/3}$$

first considered in Example 1.4.5. As with our other examples, this one is not separable by can be converted to a separable equation on the reduced state space $U' = \mathbb{R} \setminus \{0\}$:

$$\tilde{F}(t, x, x^{(1)}) = \frac{x^{(1)}}{x^{1/3}} - 1.$$

We then have

$$F_1(x) = \frac{3x^{2/3}}{2}, \quad F_0(t) = t$$

and so solutions $t \mapsto \xi(t)$ are determined by

$$\frac{3\xi(t)^{2/3}}{2} - \frac{3\xi(t_0)^{2/3}}{2} = t - t_0 \quad \Longrightarrow \quad \xi(t) = \frac{(2t - 2t_0 + 3\xi(t_0)^{2/3})^{3/2}}{3\sqrt{3}}.$$

Again, if we include the possibility that $\xi(t_0) = 0$, we arrive at the situation described in Example 1.4.5.

4. Finally, we consider the separable ordinary differential equation

$$F(t, x, x^{(1)}) = (x^4 + x^2 + 1)x^{(1)} - \mathrm{e}^{-t^2}$$

with $f_1(x) = x^4 + x^2 + 1$ and $f_0(t) = \mathrm{e}^{-t^2}$ with $(t, x) \in \mathbb{R} \times \mathbb{R}$. Here we have

$$F_1(x) = \frac{x^5}{5} + \frac{x^3}{3} + x, \quad F_0(t) = \frac{\sqrt{\pi}}{2} \operatorname{erf}(t),$$

where erf is the ***error function*** defined by

$$\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \mathrm{e}^{-\tau^2} \, \mathrm{d}\tau.$$

Thus a solution $t \mapsto \xi(t)$ satisfies

$$\frac{\xi(t)^5}{5} + \frac{\xi(t)^3}{3} + \xi(t) - \frac{\xi(t_0)^5}{5} - \frac{\xi(t_0)^3}{3} - \xi(t_0) = \frac{\sqrt{\pi}}{2}(\operatorname{erf}(t) - \operatorname{erf}(t_0)).$$

This is an implicit equation that will be unpleasant to solve. Note that one might have five possible solutions for $\xi(t)$ at a given time, since we have the solution as the root of a fifth-order polynomial. •

Exercises

2.1.1 Solve the following initial value problems, taking care to provide the domain of definition for the solution:

(a) $t\dot{\xi}(t) = 2(\xi(t) - 4)$, $\xi(1) = 5$;

(b) $(t^2 + 1)\dot{\xi}(t) = t\xi(t)$, $\xi(0) = 1$;

(c) $\dot{\xi}(t) = \xi(t)\tan(t)$, $\xi(0) = 1$;

(d) $\dot{\xi}(t) = t\xi(t) + 2t + \xi(t) + 2$, $\xi(0) = -1$.

2.1.2 Solve the following initial value problems, taking care to provide the domain of definition for the solution:

(a) $\dot{\xi}(t) + t\xi(t) = t$, $\xi(1) = 5$;

(b) $t\dot{\xi}(t) + \xi(t) = t + 1$, $\xi(1) = 0$;

(c) $\dot{\xi}(t) + \mathrm{e}^t\xi(t) = \mathrm{e}^t$, $\xi(0) = x_0$;

(d) $(1 + t)\dot{\xi}(t) + \tan(t)\xi(t) = \sec(t)$, $\xi(\frac{\pi}{4}) = 0$.

Section 2.2

Scalar linear homogeneous ordinary differential equations

Now we turn to scalar linear equations, looking first in this section at the homogeneous case. That is to say, we consider differential equations with $\mathbb{T} \subseteq \mathbb{R}$ an interval, the state space $U = \mathbb{R}$, and right-hand sides of the form

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}(t)x^{(k-1)} - \cdots - a_1(t)x^{(1)} - a_0(t)x \tag{2.1}$$

for functions $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$. Thus solutions $t \mapsto \xi(t)$ satisfies

$$\frac{\mathrm{d}^k \xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = 0.$$

In this section we shall (1) investigate the character of the solutions, (2) investigate the set of all solutions in the general case, and (3) provide a procedure for, in principle, solving the equations in the constant coefficient case.

2.2.1 Equations with time-varying coefficients

We start by working with the general situation where the coefficients $a_0, a_1, \dots, a_{k-1}$ depend on time. In this case, we will study the properties of solutions and sets of solutions, and as well introduce an important tool, the "Wronskian," for dealing with linear ordinary differential equations.

2.2.1.1 Solutions and their properties We begin by listing the general properties of solutions. First let us be sure that the equations with which we are dealing possess solutions.

2.2.1 Proposition (Local existence and uniqueness of solutions for scalar linear homogeneous ordinary differential equations) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $\mathrm{a}_0, \mathrm{a}_1, \dots, \mathrm{a}_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let*

$$(\mathrm{t}_0, \mathrm{x}_0, \mathrm{x}_0^{(1)}, \dots, \mathrm{x}_{k-1}^{(0)}) \in \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}).$$

Then there exists an interval $\mathbb{T}' \subseteq \mathbb{T}$ *and a map* $\xi \colon \mathbb{T}' \to \mathbb{R}$ *of class* C^k *that is a solution for* F *and which satisfies*

$$\xi(\mathrm{t}_0) = \mathrm{x}_0, \frac{\mathrm{d}\xi}{\mathrm{dt}}(\mathrm{t}_0) = \mathrm{x}_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{dt}^{k-1}}(\mathrm{t})(\mathrm{t}_0) = \mathrm{x}_0^{(k-1)}.$$

Moreover, if $\tilde{\mathbb{T}}' \subseteq \mathbb{T}$ *is another subinterval and if* $\tilde{\xi} \colon \tilde{\mathbb{T}}' \to \mathbb{R}$ *is another* C^k*-solution for* F *satisfying*

$$\tilde{\xi}(\mathrm{t}_0) = \mathrm{x}_0, \frac{\mathrm{d}\tilde{\xi}}{\mathrm{dt}}(\mathrm{t}_0) = \mathrm{x}_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\tilde{\xi}}{\mathrm{dt}^{k-1}}(\mathrm{t})(\mathrm{t}_0) = \mathrm{x}_0^{(k-1)},$$

then $\tilde{\xi}(t) = \xi(t)$ *for every* $t \in \tilde{\mathbb{T}}' \cap \mathbb{T}'$.

Proof This is Exercise 2.2.1. ■

As we have seen in Example 1.4.4, a solution to a general ordinary differential equation will not be defined for all times in $\mathbb{T}$, even for seemingly "nice" differential equations. One might then wonder whether linear ordinary differential equations are sufficiently nice to permit solutions defined for all time. This is, indeed, the case.

2.2.2 Proposition (Global existence of solutions for scalar linear homogeneous ordinary differential equations) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $a_0, a_1, \dots, a_{k-1} : \mathbb{T} \to \mathbb{R}$ *are continuous. If* $\xi : \mathbb{T}' \to \mathbb{R}$ *is a solution for* F*, then there exists a solution* $\overline{\xi} : \mathbb{T} \to \mathbb{R}$ *for which* $\overline{\xi}|\mathbb{T}' = \xi$.

Proof Note that, as per Exercise 1.3.23, we can convert the differential equation F into a first-order differential equation linear homogeneous differential equation with states $(x, x^{(1)}, \dots, x^{(k-1)})$. Thus the result will follow from the analogous result for first-order systems of equations, and this is stated and proved as Proposition 3.2.5. ■

Now that we know the domain of definition of a scalar linear homogeneous ordinary differential equation, we can talk in a reasonable manner about the set of *all* solutions of such equations, as the structure of these is what is most interesting about the equations. Thus we consider a scalar linear homogeneous ordinary differential equation

$$F : \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

with right-hand side

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}(t)x^{(k-1)} - \cdots - a_1(t)x^{(1)} - a_0(t)x,$$

where $a_0, a_1, \dots, a_{k-1} : \mathbb{T} \to \mathbb{R}$ are continuous. Let us denote by

$$\mathrm{Sol}(F) = \Big\{ \xi \in \mathrm{C}^k(\mathbb{T}; \mathbb{R}) \Big| \ \xi \text{ satisfies } \frac{\mathrm{d}^k \xi}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = 0 \Big\}$$

the set of solutions for F. The following result is then the main structural result for the class of differential equations we are considering in this section. We recall from Example 1.2.1–1 that $\mathrm{C}^k(\mathbb{T}; \mathbb{R})$ is a $\mathbb{R}$-vector space.

2.2.3 Theorem (Vector space structure of sets of solutions) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Then* Sol(F) *is a* k*-dimensional subspace of* $C^k(\mathbb{T}; \mathbb{R})$.

Proof We first show that Sol(F) is a subspace. Let $\xi, \xi_1.\xi_2 \in \mathrm{Sol}(F)$ and $\alpha \in \mathbb{R}$. Then we immediately have

$$\begin{aligned}\frac{\mathrm{d}^k(\xi_1+\xi_2)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1})\xi_1+\xi_2}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}(\xi_1+\xi_2)}{\mathrm{d}t}(t) + a_0(t)(\xi_1+\xi_2)(t)\\ = \frac{\mathrm{d}^k\xi_1}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) + a_0(t)\xi_1(t)\\ + \frac{\mathrm{d}^k\xi_2}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t) + a_0(t)\xi_2(t) = 0+0=0\end{aligned}$$

and

$$\begin{aligned}\frac{\mathrm{d}^k(\alpha\xi)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}(\alpha\xi)}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}(\alpha\xi)}{\mathrm{d}t}(t) + a_0(t)(\alpha\xi)(t)\\ = \alpha\left(\frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t)\right) = 0,\end{aligned}$$

using linearity of differentiation.

Next we prove that the dimension of Sol(F) is k. We shall do this by showing that, for a given $t_0 \in \mathbb{T}$, the map

$$\begin{aligned}\sigma_{t_0} \colon \mathrm{Sol}(F) &\to \mathbb{R}^k\\ \xi &\mapsto \left(\xi(t_0), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t_0)\right)\end{aligned}$$

is an isomorphism of $\mathbb{R}$-vector spaces. Since the map is surjective by the existence part of Proposition 2.2.1, it suffices to show that it is an injective linear map. Linearity of σ_{t_0} is immediate since the identities

$$\begin{aligned}\left((\xi_1+\xi_2)(t_0), \frac{\mathrm{d}(\xi_1+\xi_2)}{\mathrm{d}t}(t_0), \dots, \frac{\mathrm{d}^{k-1}(\xi_1+\xi_2)}{\mathrm{d}t^{k-1}}(t_0)\right)\\ = \left(\xi_1(t_0), \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t_0), \dots, \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t_0)\right) + \left(\xi_2(t_0), \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t_0), \dots, \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t_0)\right),\end{aligned}$$

by definition of the vector space structure for Sol(F). To show that σ_{t_0} is injective, it suffices so show that, if $\sigma_{t_0}(\xi) = \mathbf{0}$, then ξ is the zero vector in Sol(F),[1] i.e., that

[1] This relies on the fact, presumably familiar to you from your first linear algebra course, that a linear map is injective L if and only if $\ker(L) = \{0\}$.

$\xi(t) = 0$ for all $t \in \mathbb{T}$. So, suppose that $\sigma_{t_0}(\xi) = \mathbf{0}$. Then

$$\xi(t_0) = 0, \ \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) = 0, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t_0) = 0.$$

Consider the function $\zeta \colon \mathbb{T} \to \mathbb{R}$ given by $\zeta(t) = 0$ for all $t \in \mathbb{T}$. Then $\zeta \in \mathrm{Sol}(F)$ and

$$\zeta(t_0) = 0, \ \frac{\mathrm{d}\zeta}{\mathrm{d}t}(t_0) = 0, \dots, \frac{\mathrm{d}^{k-1}\zeta}{\mathrm{d}t^{k-1}}(t_0) = 0.$$

Therefore, by Proposition 2.2.1, $\xi = \zeta$, giving the theorem. ■

Being a finite-dimensional $\mathbb{R}$-vector space, the set $\mathrm{Sol}(F)$ of solutions to the scalar linear homogeneous differential equation F is capable of possessing a basis. One has a special name for a basis of $\mathrm{Sol}(F)$, i.e., a set of k linearly independent solutions for F.

2.2.4 Definition (Fundamental set of solutions) Consider the linear homogeneous ordinary differential equation F with right-hand side (2.1) and suppose that the functions $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$ are continuous. A set $\{\xi_1, \dots, \xi_k\}$ of linearly independent elements of $\mathrm{Sol}(F)$ is a ***fundamental set of solutions*** for F. •

There is not much more one can say easily, in general, about scalar linear homogeneous ordinary differential equations with coefficients that depend on time. There is, however, one case where they can be solved "explicitly," and this is when $k = 1$.

2.2.5 Example (Degree one scalar linear homogeneous equations) The differential equation we consider here is given by

$$\begin{aligned} F \colon \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^1_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) &\to \mathbb{R} \\ (t, x, x^{(1)}) &\mapsto x^{(1)} + a(t)x, \end{aligned}$$

for a continuous function $a \colon \mathbb{T} \to \mathbb{R}$. Thus a solution $t \mapsto \xi(t)$ satisfies

$$\dot{\xi}(t) = -a(t)\xi(t).$$

Note that F is equivalent to the separable equation

$$\tilde{F}(t, x, x^{(1)}) = \frac{x^{(1)}}{x} + a(t)$$

with $f_1(x) = x^{-1}$ and $f_0(t) = -a(t)$. Thus we can apply the methods of Section 2.1 to solve this equation; indeed, note that Example 2.1.3–1 is a special case that we have already treated in this manner. Let $t_0 \in \mathbb{T}$ and $x_0 \in \mathbb{R}$. We have the antiderivatives

$$F_1(x) = \ln(|x|) - \ln(|x_0|), \quad F_0(t) = -\int_{t_0}^{t} a(\tau)\,\mathrm{d}\tau.$$

In the same manner as Example 2.1.3–1, we conclude that

$$\xi(t) = \xi(t_0)\mathrm{e}^{-\int_{t_0}^{t} a(\tau)\,\mathrm{d}\tau}.$$

Note that this solution is also valid when $\xi(t_0) = 0$, although this is not covered by this solution method, since we had to eliminate 0 from the state space to make the equation a separable equation. •

2.2.1.2 The Wronskian, and its properties and uses In this section we present a fairly simple construction that turns out to have great importance in the treatment of linear differential equations. We first make a simple general definition that seems to not be *a priori* relating to differential equations.

2.2.6 Definition (Wronskian) Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval and let $f_1, \dots, f_k \in \mathsf{C}^{k-1}(\mathbb{T}; \mathbb{R})$ for $k \in \mathbb{Z}_{>0}$. The ***Wronskian*** for the functions $f_1, \dots, f_k$ is the function $\mathrm{W}(f_1, \dots, f_k) \colon \mathbb{T} \to \mathbb{R}$ defined by

$$\mathrm{W}(f_1, \dots, f_k)(t) = \det \begin{bmatrix} f_1(t) & f_2(t) & \cdots & f_k(t) \\ \frac{\mathrm{d}f_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}f_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}f_k}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}f_1}{\mathrm{d}t^{k-1}}(t) & \frac{\mathrm{d}^{k-1}f_2}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}f_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix}$$ •

An essential feature of the Wronskian is that it gives a sufficient condition for measuring the linear independence of finite sets of functions in the space of functions. More precisely, we have the following result, which again is not *a priori* related to differential equations.

2.2.7 Proposition (The Wronskian and linear independence) *Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval and let $\mathrm{f}_1, \dots, \mathrm{f}_k \in \mathsf{C}^{k-1}(\mathbb{T}; \mathbb{R})$ for $\mathrm{k} \in \mathbb{Z}_{>0}$. If $\mathrm{W}(\mathrm{f}_1, \dots, \mathrm{f}_k)(\mathrm{t}) \neq 0$ for some $\mathrm{t} \in \mathbb{T}$, then the set $\{\mathrm{f}_1, \dots, \mathrm{f}_k\}$ is linearly independent in $\mathsf{C}^{k-1}(\mathbb{T}; \mathbb{R})$.*

Proof We prove the contrapositive, i.e., that, if the functions $\{f_1, \dots, f_k\}$ are linearly dependent, then $\mathrm{W}(f_1, \dots, f_k)(t) = 0$ for all $t \in \mathbb{T}$.

So suppose that $\{f_1, \dots, f_k\}$ is linearly dependent, and let $c_1, \dots, c_k \in \mathbb{R}$, not all zero, be such that

$$c_1 f_1 + \cdots + c_k f_k = 0.$$

Then, for any $j \in \{1, \dots, k-1\}$,

$$c_1 \frac{\mathrm{d}^j f_1}{\mathrm{d}t^j} + \cdots + c_n \frac{\mathrm{d}^j f_n}{\mathrm{d}t^j} = 0.$$

Assembling these relationships for $j \in \{0, 1, \dots, k-1\}$ gives the single equation

$$\begin{bmatrix} f_1(t) & f_2(t) & \cdots & f_k(t) \\ \frac{\mathrm{d}f_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}f_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}f_k}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}f_1}{\mathrm{d}t^{k-1}}(t) & \frac{\mathrm{d}^{k-1}f_2}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}f_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This means that the matrix on the left has a nontrivial kernel (since this kernel contains $(c_1, \dots, c_k)$) and so must have zero determinant. ■

Note that the converse of the preceding result is not generally true, as demonstrated by the following example.

2.2.8 Example (The Wronskian is not adequate to characterise linear independence) Let $\mathbb{T} = [-1, 1]$ and consider the two functions $f_1, f_2 \colon [-1, 1] \to \mathbb{R}$ of class C^1 defined by

$$f_1(t) = t^2, \quad f_2(t) = t|t|.$$

We have

$$\frac{\mathrm{d}f_1}{\mathrm{d}t}(t) = 2t, \quad \frac{\mathrm{d}f_2}{\mathrm{d}t} = 2|t|$$

We thus have

$$\mathrm{W}(f_1, f_2) = \det \begin{bmatrix} t^2 & t|t| \\ 2t & 2|t| \end{bmatrix} = 2t^2|t| - 2t^2|t| = 0.$$

However, the set $\{f_1, f_2\}$ *is* linearly independent. Indeed, suppose that $c_1, c_2 \in \mathbb{R}$ satisfy

$$c_1 f_1(t) + c_2 f_2(t) = 0, \qquad t \in [-1, 1].$$

Then, taking $t = -1$, we get $c_1 - c_2 = 0$ and taking $t = 1$ we get $c_1 + c_2 = 0$. The only way both of these equations can be satisfied is when $c_1 = c_2 = 0$. •

Thus the Wronskian is not quite the thing for precisely characterising the linear independence of general sets of functions. However, it is just the thing when the set of functions under consideration are solutions to a scalar linear homogeneous ordinary differential equation.

2.2.9 Proposition (Wronskians and linear independence in Sol(F)) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $\mathrm{a}_0, \mathrm{a}_1, \dots, \mathrm{a}_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Then the following statements are equivalent for* $\xi_1, \dots, \xi_k \in \mathrm{Sol}(\mathrm{F})$*:*

(i) $\{\xi_1, \dots, \xi_k\}$ *is linearly independent;*

(ii) $\mathrm{W}(\xi_1, \dots, \xi_k)(\mathrm{t}) \neq 0$ *for some* $\mathrm{t} \in \mathbb{T}$*;*

(iii) $\mathrm{W}(\xi_1, \dots, \xi_k)(\mathrm{t}) \neq 0$ *for all* $\mathrm{t} \in \mathbb{T}$*.*

Proof (i) $\implies$ (ii) We prove the contrapositive, i.e., we prove that, if $W(\xi_1, \dots, \xi_k)(t) = 0$ for all $t \in \mathbb{T}$, then $\{\xi_1, \dots, \xi_k\}$ is linearly dependent.

So suppose that $W(\xi_1, \dots, \xi_k)(t) = 0$ for all $t \in \mathbb{T}$, which means that there exists $c_1, \dots, c_k \in \mathbb{R}$, not all zero, such that

$$\begin{bmatrix} \xi_1(t) & \xi_2(t) & \cdots & \xi_k(t) \\ \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}\xi_k}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t) & \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

for all $t \in \mathbb{T}$. If we simply expand this out, we see that it is equivalent to

$$c_1\sigma_t(\xi_1) + \cdots + c_k\sigma_t(\xi_k) = 0$$

for all $t \in \mathbb{T}$, recalling the isomorphism $\sigma_t \colon \mathrm{Sol}(F) \to \mathbb{R}^k$, defined for some $t \in \mathbb{T}$, from the proof of Theorem 2.2.3. Since σ_t is linear, this gives

$$\sigma_t(c_1\xi_1 + \cdots + c_k\xi_k) = 0, \qquad t \in \mathbb{T}.$$

Injectivity of σ_t then gives

$$c_1\xi_1 + \cdots + c_k\xi_k = 0,$$

showing linear dependence of $\{\xi_1, \dots, \xi_k\}$.

(ii) $\implies$ (iii) From Proposition 2.2.7, noting that $\xi_1, \dots, \xi_k$ are of class C^k, and so of class C^{k-1}, the assumption of (i) implies that $\{\xi_1, \dots, \xi_k\}$ is linearly independent. Suppose now that there exists $t' \in \mathbb{T}$ such that $W(\xi_1, \dots, \xi_k)(t') = 0$. Then there exists $c_1, \dots, c_k \in \mathbb{R}$, not all zero, such that

$$\begin{bmatrix} \xi_1(t') & \xi_2(t') & \cdots & \xi_k(t') \\ \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t') & \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t') & \cdots & \frac{\mathrm{d}\xi_k}{\mathrm{d}t}(t') \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t') & \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t') & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t') \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.2}$$

Now, define $\xi \colon \mathbb{T} \to \mathbb{R}$ by

$$\xi = c_1\xi_1 + \cdots + c_kxi_k.$$

By Theorem 2.2.3, $\xi \in \mathrm{Sol}(F)$. Moreover, the equation (2.2) gives

$$\xi(t') = 0, \ \frac{\mathrm{d}\xi}{\mathrm{d}t}(t') = 0, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^k}(t') = 0.$$

By Proposition 2.2.1 we conclude that $\xi(t) = 0$ for all $t \in \mathbb{T}$. This contradicts the linear independence of $\{\xi_1, \dots, \xi_k\}$.

(iii) $\implies$ (i) This follows from Proposition 2.2.7, noting that $\xi_1, \dots, \xi_k$ are of class C^k, and so of class C^{k-1}. ■

The following result gives an interesting characterisation of the Wronskian, further illustrating the fact that, when applied to solutions of scalar linear homogeneous ordinary differential equations, it serves to characterise linear independence of sets of solutions.

2.2.10 Proposition (Liouville's formula) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. If* $\{\xi_1, \dots, \xi_k\}$ *are linearly independent, then, for any* $t_0, t \in \mathbb{T}$,

$$W(\xi_1, \dots, \xi_k)(t) = W(\xi_1, \dots, \xi_k)(t_0) e^{-\int_{t_0}^{t} a_{k-1}(\tau)\, d\tau}.$$

Proof This is Exercise 3.2.6, which can be proved using some attributes of systems of linear ordinary differential equations in Section 3.2. ■

One of the sort of peculiar features of the Wronskian is that it can be used to actually write down a differential equation. While it seems, at this point, to be just a mere trick, the next result will be important when we consider inhomogeneous equations in Section 2.3.

2.2.11 Proposition (A Wronskian representation of a differential equation) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side (2.1) and suppose that the functions* $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let* $\{\xi_1, \dots, \xi_k\}$ *be a fundamental set of solutions for* F. *Then, for* $\xi \in C^k(\mathbb{T}; \mathbb{R})$,

$$\frac{d^k \xi}{dt^k}(t) + a_{k-1}(t) \frac{d^{k-1} \xi}{dt^{k-1}}(t) + \cdots + a_1(t) \frac{d\xi}{dt}(t) + a_0 \xi(t) = \frac{W(\xi_1, \dots, \xi_k, \xi)(t)}{W(\xi_1, \dots, \xi_k)(t)}.$$

In particular,

$$\operatorname{Sol}(F) = \left\{ \xi \in C^k(\mathbb{T}; \mathbb{R}) \;\middle|\; \frac{W(\xi_1, \dots, \xi_k, \xi)(t)}{W(\xi_1, \dots, \xi_k)(t)} = 0 \right\}.$$

Proof First of all, note by Proposition 2.2.9 that $W(\xi_1, \dots, \xi_k)(t)$ is never zero, so this is valid to appear in denominators, as in the statement of the proposition.

We shall prove the last assertion first. First suppose that $\xi \in \operatorname{Sol}(F)$, then

$$\xi = c_1 \xi_1 + \cdots + c_k \xi_k$$

for some (unique) constants $c_1, \dots, c_k \in \mathbb{R}$. Therefore, the functions $\{\xi, \xi_1, \dots, \xi_k\}$ are linearly dependent, cf.

$$-c_1 \xi_1 - \cdots - c_k \xi_k + 1\xi = 0.$$

Therefore, differentiating this equation k-times gives

$$\begin{bmatrix} \xi_1(t) & \xi_2(t) & \cdots & \xi_k(t) & \xi(t) \\ \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}\xi_k}{\mathrm{d}t}(t) & \frac{\mathrm{d}\xi}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t') & \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t') & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t') & \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) \\ \frac{\mathrm{d}^{k}\xi_1}{\mathrm{d}t^{k}}(t') & \frac{\mathrm{d}^{k}\xi_2}{\mathrm{d}t^{k}}(t') & \cdots & \frac{\mathrm{d}^{k}\xi_k}{\mathrm{d}t^{k}}(t') & \frac{\mathrm{d}^{k}\xi}{\mathrm{d}t^{k}}(t) \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ \vdots \\ -c_k \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

for all $t \in \mathbb{T}$. From this we immediately conclude that $\mathrm{W}(\xi_1, \dots, \xi_k, \xi)(t) = 0$ for all $t \in \mathbb{T}$, and so

$$\xi \in \left\{ \xi \in \mathsf{C}^k(\mathbb{T}; \mathbb{R}) \;\middle|\; \frac{\mathrm{W}(\xi_1, \dots, \xi_k, \xi)(t)}{\mathrm{W}(\xi_1, \dots, \xi_k)(t)} = 0 \right\}.$$

Now note that, if we expand the determinant $\mathrm{W}(\xi_1, \dots, \xi_k, \xi)$ about the last column, we get an expression of the form

$$\begin{aligned} &\mathrm{W}(\xi_1, \dots, \xi_k, \xi)(t) \\ &\qquad = \mathrm{W}(\xi_1, \dots, \xi_k)(t) \frac{\mathrm{d}^k \xi}{\mathrm{d}t^k}(t) + b_{k-1}(t) \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + b_1(t) \frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) \end{aligned}$$

for some continuous functions $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$. By Proposition 2.2.9 it follows that

$$\left\{ \xi \in \mathsf{C}^k(\mathbb{T}; \mathbb{R}) \;\middle|\; \frac{\mathrm{W}(\xi_1, \dots, \xi_k, \xi)(t)}{\mathrm{W}(\xi_1, \dots, \xi_k)(t)} = 0 \right\}$$

is the set of solutions to a kth-order scalar linear homogeneous ordinary differential equation. Moreover, since we clearly have $\mathrm{W}(\xi_1, \dots, \xi_k, \xi_j) = 0$ for every $j \in \{1, \dots, k\}$, (it is the determinant of a $(k+1) \times (k+1)$ matrix with two equal columns), it follows that $\{\xi_1, \dots, \xi_k\}$ is a fundamental set of solutions for this differential equation. Thus we have shown that

$$\mathrm{Sol}(F) = \left\{ \xi \in \mathsf{C}^k(\mathbb{T}; \mathbb{R}) \;\middle|\; \frac{\mathrm{W}(\xi_1, \dots, \xi_k, \xi)(t)}{\mathrm{W}(\xi_1, \dots, \xi_k)(t)} = 0 \right\}.$$

To prove the first assertion, we shall show that the set of solutions for a kth-order scalar linear homogeneous ordinary differential equation uniquely determines its coefficients. That is, we show that if two such equations F and G with right-hand sides

$$\begin{aligned} \widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) &= -a_{k-1}(t)x^{(k-1)} - \cdots - a_1(t)x^{(1)} - a_0(t)x, \\ \widehat{G}(t, x, x^{(1)}, \dots, x^{(k-1)}) &= -b_{k-1}(t)x^{(k-1)} - \cdots - b_1(t)x^{(1)} - b_0(t)x \end{aligned}$$

satisfy $\mathrm{Sol}(F) = \mathrm{Sol}(G)$, then $a_j = b_j$, $j \in \{0, 1, \dots, k-1\}$. Let us consider the differential equation

$$H(t, x, x^{(1)}, \dots, x^{(k-1)}) = F(t, x, x^{(1)}, \dots, x^{(k-1)}) - G(t, x, x^{(1)}, \dots, x^{(k-1)}).$$

Note that this is not necessarily a $(k-1)$st-order ordinary differential equation, since we may have $a_{k-1} = b_{k-1}$. However, suppose that $\widehat{F} \neq \widehat{G}$ and let j be the largest element of $\{0, 1, \dots, k-1\}$ such that $a_j \neq b_j$. Thus there exists $t_0 \in \mathbb{T}$ so that $a_j(t_0) \neq b_j(t_0)$. Since a_j and b_j are continuous, there is an interval $\mathbb{T}' \subseteq \mathbb{T}$ around t_0 such that $a_j(t) \neq b_j(t)$ for all $t \in \mathbb{T}'$. We then define an ordinary differential equation H' with right-hand side

$$\begin{aligned}\widehat{H}' \colon \mathbb{T}' \times \mathbb{R} \oplus \mathrm{L}^{lej-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) &\to \mathbb{R}\\ (t, x, x^{(1)}, \dots, x^{(j-1)}) &\mapsto -\frac{a_{j-1}(t) - b_{j-1}(t)}{a_j(t) - b_j(t)} x^{(j-1)} - \cdots - \frac{a_1(t) - b_1(t)}{a_j(t) - b_j(t)} x^{(1)}\\ &\quad - \frac{a_0(t) - b_0(t)}{a_j(t) - b_j(t)} x.\end{aligned}$$

This jth-order ordinary differential equation has $\xi_1, \dots, \xi_k$ as linearly independent solutions, and this is in contradiction with Theorem 2.2.3. Thus we must have $\widehat{F} = \widehat{G}$, as claimed. ■

2.2.2 Equations with constant coefficients

Having said about as much as one can say, in general, about the situation with time-varying coefficients, we now turn to the case of constant coefficient scalar linear homogeneous ordinary differential equations. If

$$F \colon \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

is such an equation, then its right-hand side must be given by

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k)}) = -a_{k-1} x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x \tag{2.3}$$

for $a_0, a_1, \dots, a_{k-1} \in \mathbb{R}$. Thus a solution $t \mapsto \xi(t)$ satisfies the equation

$$\frac{\mathrm{d}^k \xi(t)}{\mathrm{d}t^k}(t) + a_{k-1} \frac{\mathrm{d}^{k-1} \xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0 \xi(t) = 0. \tag{2.4}$$

These equations are, of course, a special case of the equations considered in Section 2.2.1, and so all statements made about the general case of time-varying coefficients hold in the special case of constant coefficients. In particular, Propositions 2.2.1 and 2.2.2, and Theorem 2.2.3 hold for equations of the form (2.4). However, for these constant coefficient equations, it is possible to explicitly describe the character of the solutions, and this is what we undertake to do.

The trick, motivated to some extent by Example 2.1.3–1, is to *assume* a solution of the form $\xi(t) = ae^{rt}$ for $a, r \in \mathbb{R}$, and see what happens. A direct substitution into the equation (2.4) shows that, with ξ in this assumed form,

$$\frac{\mathrm{d}^k(ae^{rt})}{\mathrm{d}t^k} + a_{k-1}\frac{\mathrm{d}^{k-1}(ae^{rt})}{\mathrm{d}t^{k-1}} + \cdots + a_1\frac{\mathrm{d}(ae^{rt})}{\mathrm{d}t} + a_0(ae^{rt}) = ae^{rt}(r^k + a_{k-1}r^{k-1} + \cdots + a_1 r + a_0).$$

Since we are looking for nontrivial solutions, we suppose that $a \neq 0$, in which case $\xi(t) = ae^{rt}$ is a solution for F if and only if

$$r^k + a_{k-1}r^{k-1} + \cdots + a_1 r + a_0.$$

With this as backdrop, we make the following definition.

2.2.12 Definition (Characteristic polynomial of a scalar linear homogeneous differential equation with constant coefficients) Consider the linear homogeneous ordinary differential equation F with constant coefficients and with right-hand side (2.3). The ***characteristic polynomial*** of F is

$$P_F = X^k + a_{k-1}X^{k-1} + \cdots + a_1 X + a_0 \in \mathbb{R}[X]. \qquad \bullet$$

Now we systematically develop the methodology for solving scalar linear homogeneous ordinary differential equations with constant coefficients.

2.2.2.1 Complexification of scalar linear ordinary differential equations It turns out that to solve constant coefficient linear ordinary differential equations, one needs to work with complex numbers. To do this systematically, we introduce the notion of "complexification," by which a real equation is converted into a complex one. This is rather elementary in this setting, but will be less elementary in Section 3.2.3. Thus it will do not harm, and maybe do some good, to treat this systematically here.

First let us understand the notation for derivatives of $\mathbb{C}$-valued functions of a single real variable, i.e., functions of time. Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval and suppose that we have a mapping $\zeta\colon \mathbb{T} \to \mathbb{C}$. Since we have $\mathbb{C} \simeq \mathbb{R}^2$, it makes sense to say that ζ is of class C^k for any $k \in \mathbb{Z}_{\geq 0}$: it is of class C^k if and only if both its real and imaginary parts are of class C^k. Moreover, if we write ζ as a sum of its real and imaginary parts, $\zeta(t) = \xi(t) + \mathrm{i}\eta(t)$, then we have

$$\frac{\mathrm{d}^k\zeta}{\mathrm{d}t^k} = \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k} + \mathrm{i}\frac{\mathrm{d}^k\eta}{\mathrm{d}t^k}.$$

Thus derivatives of order k are just $\mathbb{C}$-valued functions of t. Thus we can follow the same line of reasoning as Remark 1.3.4 and make the identification $\mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}; \mathbb{C}) \simeq \mathbb{C}$.

Here is the basic and quite elementary construction.

2.2.13 Definition (Complexification of scalar linear ordinary differential equation) Consider the linear homogeneous ordinary differential equation F with constant coefficients and with right-hand side (2.3). The ***complexification*** of F is the mapping

$$\begin{aligned} F^{\mathbb{C}} \colon \mathbb{T} \times \mathbb{C} \oplus \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{C}) &\to \mathbb{C} \\ (t, z, z^{(1)}, \dots, z^{(k)}) &\mapsto z^{(k)} + a_{k-1} z^{(k-1)} + \cdots + a_1 z^{(1)} + a_0 z. \end{aligned}$$

A ***solution*** for $F^{\mathbb{C}}$ is a C^k-function $\zeta \colon \mathbb{T} \to \mathbb{C}$ that satisfies

$$\frac{\mathrm{d}^k \zeta(t)}{\mathrm{d}t^k}(t) + a_{k-1} \frac{\mathrm{d}^{k-1} \zeta}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\zeta}{\mathrm{d}t}(t) + a_0 \zeta(t) = 0.$$

By $\mathrm{Sol}(F^{\mathbb{C}})$ we denote the set of solutions for $F^{\mathbb{C}}$. •

Everything we said in Section 2.2.1 about scalar linear homogeneous ordinary differential equations holds in the case of the complex differential equation $F^{\mathbb{C}}$, even when the coefficients are not constant. In particular, Propositions 2.2.1 and 2.2.2, and Theorem 2.2.3 hold in this case to give us the basic attributes of the complex differential equation, merely by replacing the appropriate occurrences of the symbol "$\mathbb{R}$" with the symbol "$\mathbb{C}$." In particular, $\mathrm{Sol}(F^{\mathbb{C}})$ is a k-dimensional $\mathbb{C}$-vector space if F has order k.

An essential result for returning to "reality" after complexification is the following simple result.

2.2.14 Lemma (Real and imaginary parts of complex solutions are solutions) *Consider the linear homogeneous ordinary differential equation* F *with constant coefficients, with right-hand side (2.3) and with complexification* $\mathrm{F}^{\mathbb{C}}$. *If* $\zeta \colon \mathbb{T} \to \mathbb{C}$ *is a solution for* $\mathrm{F}^{\mathbb{C}}$, *then* $\mathrm{Re}(\zeta)$ *and* $\mathrm{Im}(\zeta)$ *are solutions for* F.

Proof Since ζ is a solution for $F^{\mathbb{C}}$, we have

$$\frac{\mathrm{d}^k \zeta}{\mathrm{d}t^k}(t) + a_{k-1} \frac{\mathrm{d}^{k-1} \zeta}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\zeta}{\mathrm{d}t}(t) + a_0 \zeta(t) = 0.$$

Now we note that $\mathrm{Re} \colon \mathbb{C} \to \mathbb{R}$ and $\mathrm{Im} \colon \mathbb{C} \to \mathbb{R}$ are $\mathbb{R}$-linear maps. Since the coefficients $a_0, a_1, \dots, a_{k-1}$ are real, this gives

$$\begin{aligned} 0 &= \mathrm{Re}\left(\frac{\mathrm{d}^k \zeta}{\mathrm{d}t^k}(t) + a_{k-1} \frac{\mathrm{d}^{k-1} \zeta}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\zeta}{\mathrm{d}t}(t) + a_0 \zeta(t) \right) \\ &= \frac{\mathrm{d}^k \mathrm{Re}(\zeta)}{\mathrm{d}t^k}(t) + a_{k-1} \frac{\mathrm{d}^{k-1} \mathrm{Re}(\zeta)}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\,\mathrm{Re}(\zeta)}{\mathrm{d}t}(t) + a_0 \,\mathrm{Re}(\zeta)(t), \end{aligned}$$

showing that $\mathrm{Re}(\zeta)$ is a solution for F. In like manner, of course, $\mathrm{Im}(\zeta)$ is also a solution for F. ■

2.2.2.2 Differential operator calculus We introduce a simple object that will be say a few simple things about.

2.2.15 Definition (Scalar differential operator with constant coefficients) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let $k \in \mathbb{Z}_{\geq 0}$. A ***k*th-order scalar differential operator with constant coefficients in** $\mathbb{F}$ is a mapping

$$D \colon \mathsf{C}^\infty(\mathbb{T}; \mathbb{F}) \to \mathsf{C}^\infty(\mathbb{T}; \mathbb{F})$$

of the form

$$D(f) = d_k \frac{\mathrm{d}^k f}{\mathrm{d}t}(t) + d_{k-1} \frac{\mathrm{d}^{k-1} f}{\mathrm{d}t^{k-1}}(t) + \cdots + d_1 \frac{\mathrm{d}f}{\mathrm{d}t}(t) + d_0 f(t)$$

for $d_0, d_1, \ldots, d_k \in \mathbb{F}$ with $d_k \neq 0$. The ***symbol*** for such an object is

$$\sigma(D) = d_k X^k + d_{k-1} X^{k-1} + \cdots + d_1 X + d_0 \in \mathbb{F}[X]. \qquad \bullet$$

Note that, while the domain and range of D in the preceding definition is the set of infinitely differentiable functions, clearly the definition makes sense when applied to functions that are at least k-times continuously differentiable. Indeed, we can think of D as a mapping from $\mathsf{C}^{k+m}(\mathbb{T}; \mathbb{F})$ to $\mathsf{C}^m(\mathbb{T}; \mathbb{F})$ for any $m \in \mathbb{Z}_{\geq 0}$. The definition as stated just allows us to not fuss about this sort of thing for the purposes of our discussion.

Note that differential operators of the sort we are talking about have a product given by composition. Thus, if D_1 and D_1 are k_1th- and k_2th-order scalar differential operators with constant coefficients, then we define a $(k_1 + k_2)$th-order scalar differential operator $D_1 D_2$ with constant coefficients by $D_1 D_2(f) = D_1(D_2(f))$.

A simplifying observation about scalar differential operators with constant coefficients is the following.

2.2.16 Proposition (The symbol of a product is the product of the symbols) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, let* $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{Z}_{\geq 0}$*. If* D_1 *and* D_1 *are* k_1*th- and* k_2*th-order scalar differential operators with constant coefficients, then* $\sigma(\mathrm{D}_1\mathrm{D}_2) = \sigma(\mathrm{D}_1)\sigma(\mathrm{D}_2)$*.*

Proof Let us write

$$\sigma(D_1) = \sum_{j=0}^{k_1} d_{1,j} X^j, \quad \sigma(D_2) = \sum_{j=0}^{k_2} d_{2,j} X^j.$$

Then, for $f \in \mathsf{C}^\infty(\mathbb{T}; \mathbb{F})$,

$$D_1 D_2(f) = \sum_{j=0}^{k_1} d_{1,j} \frac{\mathrm{d}^j}{\mathrm{d}t^j} \left(\sum_{l=0}^{k_2} d_{2,l} \frac{\mathrm{d}^l f}{\mathrm{d}t^l} \right) = \sum_{k=0}^{k_1+k_2} \sum_{j=0}^{k} d_{1,j} d_{2,k-j} \frac{\mathrm{d}^k f}{\mathrm{d}t^k}.$$

Since

$$\sigma(D_1)\sigma(D_2) = \sum_{k=0}^{k_1+k_2} \sum_{j=0}^{k} d_{1,j} d_{2,k-j} X^k,$$

the result follows. ■

2.2.17 Corollary (The product for differential operators is commutative) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, let* $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{Z}_{\geq 0}$*. If* D_1 *and* D_1 *are* k_1*th- and* k_2*th-order scalar differential operators with constant coefficients, then* $\mathrm{D}_1\mathrm{D}_2 = \mathrm{D}_2\mathrm{D}_1$*.*

Proof This follows from the following facts: (1) polynomial multiplication is commutative; (2) the mapping that assigns $\sigma(D)$ to D is injective. ■

2.2.2.3 Bases of solutions Now we construct a family of solutions for a scalar linear homogeneous ordinary differential equation. We do this via a procedure.

2.2.18 Procedure (Basis of solutions for scalar linear homogeneous ordinary differential equations with constant coefficients) Given a scalar linear homogeneous ordinary differential equation

$$F\colon \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

with right-hand side

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x,$$

do the following.

1. Let $F^{\mathbb{C}}$ be the complexification of F,
2. Consider the kth-order scalar differential operator D_F with constant coefficients in $\mathbb{C}$ defined by
$$\sigma(D_{F^{\mathbb{C}}}) = X^k + a_{k-1}X^k + \cdots + a_1 X + a_0.$$
3. Let $r_1, \dots, r_s$ be the distinct roots of $\sigma(D_F)$ and let $m(r_j)$, $j \in \{1, \dots, s\}$, be the multiplicity of the root r_j. Thus
$$\sigma(D_{F^{\mathbb{C}}}) = (X - r_1)^{m(r_1)} \cdots (X - r_s)^{m(r_s)}.$$
4. Fix $j \in \{1, \dots, s\}$ and consider the following cases.
   (a) $r_j \in \mathbb{R}$: Define functions $\xi_{r_j,l}\colon \mathbb{T} \to \mathbb{R}$, $l \in \{1, \dots, m(r_j)\}$, by
   $$\xi_{r_j,l}(t) = t^l \mathrm{e}^{r_j t}, \qquad l \in \{0, 1, \dots, m(r_j) - 1\}.$$
   (b) $r_j \in \mathbb{C} \setminus \mathbb{R}$: Note that, since r_j is complex and not real, $\bar{r}_j$ is also a root of $\sigma(D_{F^{\mathbb{C}}})$. We will work only with one of these roots, so we write $r_j = \sigma_j + \mathrm{i}\omega_j$ with $\omega_j > 0$. Define functions $\mu_{r_j,l}, \nu_{r_j,l}\colon \mathbb{T} \to \mathbb{R}$ by
   $$\mu_{r_j,l}(t) = t^l \mathrm{e}^{\sigma_j t} \cos(\omega_j t), \ \ \nu_{r_j,l}(t) = t^l \mathrm{e}^{\sigma_j t} \sin(\omega_j t), \qquad l \in \{0, 1, \dots, m(r_j) - 1\}.$$
5. Note that the result of the above steps is k functions. We will show that these functions form a basis for $\mathrm{Sol}(F)$. •

2.2.19 Theorem (Basis of solutions for scalar linear homogeneous ordinary differential equations with constant coefficients) *Given a scalar linear homogeneous ordinary differential equation*

$$\mathrm{F}\colon \mathbb{T}\times\mathbb{R}\oplus \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R};\mathbb{R}) \to \mathbb{R}$$

with right-hand side

$$\widehat{\mathrm{F}}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x,$$

define k *functions as in Procedure 2.2.18. Then these functions form a basis for* Sol(F).

Proof First we show that each of the functions defined in Procedure 2.2.18 is a solution for F.

First we consider the functions $\xi_{r_j,l}(t) = t^l e^{r_j t}$, $l \in \{0, 1, \dots, m(r_j) - 1\}$, associated with a real root r_j of the characteristic polynomial for F. Since

$$\sigma(D_{F^{\mathbb{C}}}) = (X - r_1)^{m(r_1)} \cdots (X - r_s)^{m(r_s)},$$

by Corollary 2.2.17 we can write

$$\sigma(D_{F^{\mathbb{C}}}) = P(X - r_j)^{m(r_j)}$$

for some $P \in \mathbb{C}[X]$. Therefore, it suffices to show that, for $r \in \mathbb{R}$ and for $m, l \in \mathbb{Z}_{\ge 0}$ with $m \ge 1$ and $l < m$, we have

$$\left(\frac{d}{dt} - r\right)^m P(t)e^{rt} = 0, \tag{2.5}$$

where P is any polynomial function of degree $l \in \{0, 1, \dots, m-1\}$. To prove (2.5), we first prove a simple lemma.

1 Lemma *Let* $m \in \mathbb{Z}_{>0}$ *and* $r \in \mathbb{C}$. *If* $\xi\colon \mathbb{T} \to \mathbb{C}$ *is of class* C^m *then*

$$\left(\frac{d}{dt} - r\right)^m (\xi(t)e^{rt}) = e^{rt}\frac{d^m\xi}{dt^m}(t).$$

Proof We prove this by induction on m. For $m = 1$ we have

$$\left(\frac{d}{dt} - r\right)(\xi(t)e^{rt}) = \frac{d\xi}{dt}(t)e^{rt} + r\xi(t)e^{rt} - r\xi(t)e^{rt} = e^{rt}\frac{d\xi}{dt}(t),$$

giving the lemma when $m = 1$. Now suppose that the lemma holds when $m = k$. Then

$$\begin{aligned}\left(\frac{d}{dt} - r\right)^{k+1} (\xi(t)e^{rt}) &= \left(\frac{d}{dt} - r\right)\left(\frac{d}{dt} - r\right)^k (\xi(t)e^{rt}) \\ &= \left(\frac{d}{dt} - r\right)e^{rt}\frac{d^k\xi}{dt^k}(t) \\ &= re^{rt}\frac{d^k\xi}{dt^k}(t) + e^{rt}\frac{d^{k+1}\xi}{dt^{k+1}}(t) - r\frac{d^k\xi}{dt^k}(t) \\ &= e^{rt}\frac{d^{k+1}\xi}{dt^{k+1}}(t),\end{aligned}$$

as desired. ▼

Now, if P is a polynomial function of degree $l \in \{0, 1, \dots, m\}$, by the Lemma 1 we have

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - r\right)^m P(t)\mathrm{e}^{rt} = \mathrm{e}^{rt}\frac{\mathrm{d}^m P}{\mathrm{d}t^m}(t) = 0.$$

Thus shows that the functions $\xi_{r_j,l}(t) = t^l \mathrm{e}^{r_j t}$, $l \in \{0, 1, \dots, m(r_j) - 1\}$, are solutions for F.

Next we consider the functions

$$\mu_{r_j,l} = t^l \mathrm{e}^{\sigma_j t}\cos(\omega_j t), \quad \nu_{r_j,l} = t^l \mathrm{e}^{\sigma_j t}\sin(\omega_j t), \qquad l \in \{0, 1, \dots, m(r_j) - 1\},$$

corresponding to a complex root $r_j = \sigma_j + \mathrm{i}\omega_j$, $\omega_j > 0$, of the characteristic polynomial of F. In this case, we argue, exactly as in the case of a real root above, that the $\mathbb{C}$-valued functions $\zeta_{r_j,l}(t) = t^l \mathrm{e}^{r_j t}$, $l \in \{0, 1, \dots, m(r_j) - 1\}$, are solutions for $F^{\mathbb{C}}$. Then, by Lemma 2.2.14, we have that

$$\begin{aligned}\mu_{r_j,l}(t) &= t^l \mathrm{e}^{\sigma_j t}\cos(\omega_j t)\\ &= \mathrm{Re}(t^l \mathrm{e}^{\sigma_j t}(\cos(\omega_j t) + \mathrm{i}\sin(\omega_j t))\\ &= \mathrm{Re}(t^l \mathrm{e}^{\sigma_j t}\mathrm{e}^{\mathrm{i}\omega_j t}) = \mathrm{Re}(\zeta_{r_j,l}(t))\end{aligned}$$

and, similarly,

$$\nu_{r_j,l} = t^l \mathrm{e}^{\sigma_j t}\sin(\omega_j t) = \mathrm{Im}(\zeta_{r_j,l}(t))$$

are solutions for F for $l \in \{0, 1, \dots, m(r_j) - 1\}$.

Our above arguments show that the functions produced in Procedure 2.2.18 are solutions. Moreover, since Procedure 2.2.18 produces k solutions for F, by Theorem 2.2.3 it suffices to show that these solutions are linearly independent to show that they form a basis for $\mathrm{Sol}(F)$. We achieve this with the aid of the following lemma.

2 Lemma *Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval containing more than one point. Let $\mathrm{r}_1, \dots, \mathrm{r}_s \in \mathbb{R}$ be distinct and let $\mathrm{P}_1, \dots, \mathrm{P}_s$ be $\mathbb{C}$-valued polynomial functions on $\mathbb{T}$. If*

$$\mathrm{P}_1(t)\mathrm{e}^{\mathrm{r}_1 t} + \cdots + \mathrm{P}_s(t)\mathrm{e}^{\mathrm{r}_s t} = 0, \qquad t \in \mathbb{T},$$

then $\mathrm{P}_j(t) = 0$ for all $j \in \{1, \dots, s\}$ and $t \in \mathbb{T}$.

Proof We prove the lemma by induction on s. For $s = 1$ we have, for $r_1 \in \mathbb{R}$ and a polynomial function P_1,

$$\begin{aligned}& P_1(s)\mathrm{e}^{r_1 t} = 0, \qquad t \in \mathbb{T},\\ \Longrightarrow\ & P_1(t) = 0, \qquad t \in \mathbb{T},\end{aligned}$$

giving the result in this case. Now suppose that the lemma is true for $s = k$ and suppose that

$$P_1(t)\mathrm{e}^{r_1 t} + \cdots + P_k(t)\mathrm{e}^{r_k t} + P_{k+1}(t)\mathrm{e}^{r_{k+1} t} = 0, \qquad t \in \mathbb{T},$$

for distinct $r_1, \dots, r_k, r_{k+1} \in \mathbb{R}$ and for polynomial functions $P_1, \dots, P_k, P_{k+1}$. Then

$$P_1(t)\mathrm{e}^{(r_1-r_{k+1})t} + \cdots + P_k(t)\mathrm{e}^{(r_k-r_{k+1})t} + P_{k+1}(t) = 0, \qquad t \in \mathbb{T}. \tag{2.6}$$

Now let us differentiate this expression m times with respect to t, using the Leibniz Rule for higher-order derivatives, which reads

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}(fg) = \sum_{l=1}^{m} \binom{m}{l} \frac{\mathrm{d}^{m-l} f}{\mathrm{d}t^{m-l}} \frac{\mathrm{d}^l g}{\mathrm{d}t^l}.$$

After m differentiations we get

$$P_1^m(t)\mathrm{e}^{(r_1-r_{k+1})t} + \cdots + P_k^m(t)\mathrm{e}^{(r_k-r_{k+1})t} + \frac{\mathrm{d}^m P_{k+1}}{\mathrm{d}t^m}(t) = 0, \qquad t \in \mathbb{T},$$

where

$$P_j^m(t) = \sum_{l=0}^{m} (r_j - r_{k+1})^l \binom{m}{l} \frac{\mathrm{d}^{m-l} P_j}{\mathrm{d}t^{m-l}}(t). \tag{2.7}$$

Since $r_j - r_{k+1} \neq 0$, P_j^m is a polynomial function whose degree is the same as the degree of P_j. Now, for m sufficiently large (larger than the degree of P_{k+1}, to be precise), $\frac{\mathrm{d}^m P_{k+1}}{\mathrm{d}t^m} = 0$. With m so chosen, we have

$$P_1^m(t)\mathrm{e}^{(r_1-r_{k+1})t} + \cdots + P_k^m(t)\mathrm{e}^{(r_k-r_{k+1})t} = 0, \qquad t \in \mathbb{T}.$$

By the induction hypothesis, $P_j^m(t) = 0$ for $j \in \{1, \dots, k\}$ and $t \in \mathbb{T}$. Now, in the expression (2.7) for P_j^m, note that the highest polynomial degree term in t in the sum occurs when $m = 0$, and this term is $(r_j - r_{k+1})^m P_j(t)$. For the polynomial P_j^m to vanish, this term in the sum must vanish, i.e., $P_j(t) = 0$ for every $j \in \{1, \dots, k\}$ and $t \in \mathbb{T}$. Finally, (2.6) then gives $P_{k+1}(t) = 0$ for all $t \in \mathbb{T}$, giving the result. ▼

Now we can show that the solutions produced by Procedure 2.2.18 are linearly independent. Suppose that there are s_1 distinct real roots, $r_1, \dots, r_{s_1}$, and s_2 distinct complex roots,

$$\rho_j = \sigma_1 + \mathrm{i}\omega_j, \dots, \rho_{s_2} = \sigma_{s_2} + \mathrm{i}\omega_{s_2},$$

with $\omega_1, \dots, \omega_{s_2} > 0$, for the characteristic polynomial of F. Thus $s_1 + 2s_2 = k$. Suppose that we have k scalars

$$c_{j,l}, \qquad j \in \{1, \dots, s_1\},\ l \in \{0, 1, \dots, m(r_j) - 1\}, \tag{2.8}$$

and

$$a_{j,l},\ b_{j,l}, \qquad j \in \{1, \dots, s_2\},\ l \in \{0, 1, \dots, m(\rho_j) - 1\}, \tag{2.9}$$

satisfying

$$\begin{aligned}(c_{1,0} + c_{1,1}t + \cdots + c_{1,m(r_1)-1}t^{m(r_1)-1})\mathrm{e}^{r_1 t} + \ldots \\ + (c_{s_1,0} + c_{s_1,1}t + \cdots + c_{s_1,m(r_{s_1})-1}t^{m(r_{s_1})-1})\mathrm{e}^{r_1 t} \\ + (a_{1,0} + a_{1,1}t + \cdots + a_{1,m(\rho_1)-1}t^{m(\rho_1)-1})\cos(\omega_1 t) \\ + (b_{1,0} + b_{1,1}t + \cdots + b_{1,m(\rho_1)-1}t^{m(\rho_1)-1})\sin(\omega_1 t) + \ldots \\ + (a_{s_2,0} + a_{s_2,1}t + \cdots + a_{s_2,m(\rho_{s_2})-1}t^{m(\rho_{s_2})-1})\cos(\omega_{s_2} t) \\ + (b_{s_2,0} + b_{s_2,1}t + \cdots + b_{s_2,m(\rho_{s_2})-1}t^{m(\rho_{s_2})-1})\sin(\omega_{s_2} t) = 0, \qquad t \in \mathbb{T}.\end{aligned}$$

By Lemma 1, the polynomial functions

$$\begin{aligned}c_{1,0} + c_{1,1}t + \cdots + c_{1,m(r_1)-1}t^{m(r_1)-1}, \ldots, \\ c_{s_1,0} + c_{s_1,1}t + \cdots + c_{s_1,m(r_{s_1})-1}t^{m(r_{s_1})-1}, \\ a_{1,0} + a_{1,1}t + \cdots + a_{1,m(\rho_1)-1}t^{m(\rho_1)-1}, \\ b_{1,0} + b_{1,1}t + \cdots + b_{1,m(\rho_1)-1}t^{m(\rho_1)-1}, \ldots, \\ a_{s_2,0} + a_{s_2,1}t + \cdots + a_{s_2,m(\rho_{s_2})-1}t^{m(\rho_{s_2})-1}, \\ b_{s_2,0} + b_{s_2,1}t + \cdots + b_{s_2,m(\rho_{s_2})-1}t^{m(\rho_{s_2})-1}\end{aligned}$$

must all vanish. But this implies that the scalars (2.8) and (2.9) must all vanish. This gives the desired linear independence. ■

2.2.2.4 Some examples As concerns the general theory of scalar linear homogeneous ordinary differential equations, the matter is settled pretty much by Theorem 2.2.19. It remains to consider a few examples.

We first consider an "academic" example, one that illustrates Procedure 2.2.18, but which has no particular deep meaning.

2.2.20 Example ("Academic" example) We consider the 4th-order scalar linear homogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}, x^{(2)}, x^{(3)}) = -5x + 8x^{(1)} - 2x^{(2)}.$$

Thus solutions $t \mapsto \xi(t)$ to this equation satisfy

$$\frac{\mathrm{d}^4\xi}{\mathrm{d}t^4}(t) + 2\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(t) - 8\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + 5\xi(t) = 0.$$

The characteristic polynomial is

$$P_F = X^4 + 2X^2 - 8X + 5$$

which can be verified to have roots and multiplicities

$$r_1 = 1,\ m(r_1) = 2,\ \ \rho_1 = -1 + 2\mathrm{i},\ m(\rho_1) = 1.$$

Of course, we also have the root $\overline{\rho}_1 = -1 - 2\mathrm{i}$, but the bookkeeping for this is dealt with when we produce two solutions corresponding to ρ_1. According to Procedure 2.2.18 the 4 solutions that form a basis for $\mathrm{Sol}(F)$ are then

$$\xi_{r_1,0}(t) = \mathrm{e}^t,\ \xi_{r_1,1}(t) = t\mathrm{e}^t,\ \mu_{\rho_1,0}(t) = \mathrm{e}^{-t}\cos(2t),\ \nu_{\rho_1,0}(t) = \mathrm{e}^{-t}\sin(2t).$$

Thus *any* solution for F can be written as

$$\xi(t) = c_1\mathrm{e}^t + c_2 t\mathrm{e}^t + c_3\mathrm{e}^{-t}\cos(2t) + c_4\mathrm{e}^{-t}\sin(2t).$$

To prescribe a *specific* solution, according to Proposition 2.2.1, we specify initial conditions. For simplicity, let us do this at $t = 0$:

$$\xi(0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(0) = x + 0^{(1)},\ \frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(0) = x_0^{(2)},\ \frac{\mathrm{d}^3\xi}{\mathrm{d}t}(0) = x_0^{(3)}. \tag{2.10}$$

To use these conditions to determine c_1, c_2, c_3, c_4 is a tedious problem in linear algebra. We compute

$$\begin{aligned}
\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) &= c_1\mathrm{e}^t + c_2(\mathrm{e}^t + t\mathrm{e}^t) + c_3(-\mathrm{e}^{-t}\cos(2t) - 2\mathrm{e}^{-t}\sin(2t))\\
&\quad + c_4(2\mathrm{e}^{-t}\cos(2t) - \mathrm{e}^{-t}\sin(2t)),\\
\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(t) &= c_1\mathrm{e}^t + c_2(2\mathrm{e}^t + t\mathrm{e}^t) + c_3(-3\mathrm{e}^{-t}\cos(2t) + 4\mathrm{e}^{-t}\sin(2t))\\
&\quad + c_4(-4\mathrm{e}^{-t}\cos(2t) - 3\mathrm{e}^{-t}\sin(2t)),\\
\frac{\mathrm{d}^3\xi}{\mathrm{d}t^3}(t) &= c_1\mathrm{e}^t + c_2(3\mathrm{e}^t + t\mathrm{e}^t) + c_3(11\mathrm{e}^{-t}\cos(2t) + 2\mathrm{e}^{-t}\sin(2t))\\
&\quad + c_4(-2\mathrm{e}^{-t}\cos(2t) + 11\mathrm{e}^{-t}\sin(2t)).
\end{aligned}$$

Evaluating these at $t = 0$ gives the equations

$$\begin{aligned}
c_1 + c_3 &= x_0,\\
c_1 + c_2 - c_3 + 2c_4 &= x_0^{(1)},\\
c_1 + 2c_2 - 3c_3 - 4c_4 &= x_0^{(2)},\\
c_1 + 3c_2 + 11c_3 - 2c_4 &= x_0^{(3)}.
\end{aligned}$$

These are four linear equations in four unknowns that, because of Proposition 2.2.1, we know possesses unique solutions. These can be solved to give

$$\begin{aligned}
c_1 &= \tfrac{1}{16}(15x_0 + x_0^{(1)} + x_0^{(2)} - x_0^{(3)}),\\
c_2 &= \tfrac{1}{8}(-5x_0 + 3x_0^{(1)} + x_0^{(2)} + x_0^{(3)}),\\
c_3 &= \tfrac{1}{16}(x_0 - x_0^{(1)} - x_0^{(2)} + x_0^{(3)}),\\
c_4 &= \tfrac{1}{8}(-x_0 + 2x_0^{(1)} - x_0^{(2)}).
\end{aligned}$$

Go ahead and plug numbers into this bad boy, if this is your thing. •

The next two examples are intended to illustrate the how the behaviour of the roots of the characteristic polynomial affect the behaviour of solutions.

2.2.21 Example (First-order system behaviour) Here we consider a general 1st-order scalar linear homogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x) = -\frac{x}{\tau}$$

for $\tau \in \mathbb{R}$. Solutions $t \mapsto \xi(t)$ satisfy

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} + \tau^{-1}\xi(t) = 0.$$

This is an easy equation to solve. It characteristic polynomial is $P_F = X + \tau^{-1}$ which has the single real root $r_1 = -\tau^{-1}$. Thus, by Procedure 2.2.18, *any* solution has the form $\xi(t) = ce^{-t/\tau}$. To determine c, we use initial conditions as in Proposition 2.2.1. We take a general initial time t_0 and prescribe $\xi(t_0) = x_0$. Thus

$$\xi(t_0) = ce^{-t_0/\tau} \quad \Longrightarrow \quad c = \xi(t_0)e^{t_0/\tau},$$

and so $\xi(t) = \xi(0) = e^{-(t-t_0)/\tau}$.

Let us think about this solution for a moment. When $\tau > 0$, this is ***exponential decay*** and when $\tau < 0$ it is ***exponential growth***. In Figure 2.1 we graph $\xi(t)$ as a

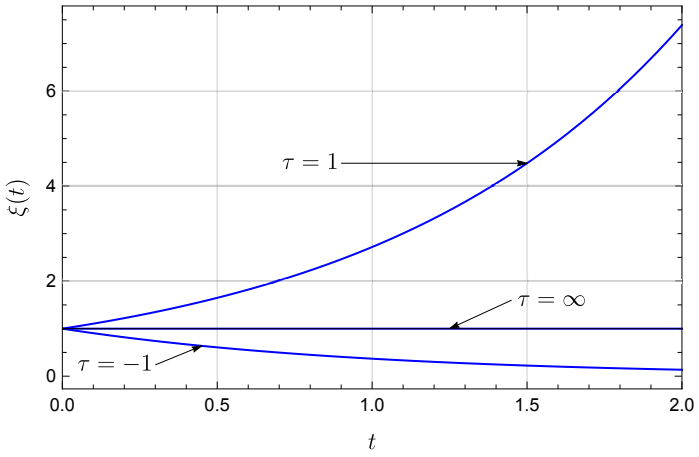


Figure 2.1 Solutions of a first-order scalar linear homogeneous ordinary differential equation with $\xi(0) = 1$

function of t for a few different τ's. Note that τ is not the rate of growth or decay, but the inverse of this. This is sometimes known as the ***time constant*** for the equation, since the units for τ are time. We can see that small (in magnitude) τ's give rise to relatively faster growth or decay. When $\tau = \infty$ (whatever that means), the decay or growth is infinitely slow, i.e., solutions neither grow nor decay. •

2.2.22 Example (Second-order system behaviour) We next consider a certain form of 2nd-order scalar linear homogeneous ordinary differential equation, namely such an equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}) = -\omega_0^2 x - 2\zeta\omega_0 x^{(1)}$$

for $\omega \in \mathbb{R}_{>0}$ and $\zeta \in \mathbb{R}$. The equations (1.1) for a mass-spring-damper and (1.6) for the current in a series RLC circuit are of this general form. A solution $t \mapsto \xi(t)$ satisfies

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(t) + 2\zeta\omega_0\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + \omega_0^2\xi(t) = 0.$$

The characteristic polynomial is

$$P_F = X^2 + 2\zeta\omega_0 X + \omega_0^2.$$

The roots of this equation are found using the quadratic formula, and their character depends on discriminant which is $\Delta = 2\omega_0^2(\zeta^2 - 1)$. When $\Delta > 0$ the roots are real and when $\Delta < 0$ the roots are complex. To be precise, the roots are the following:

1. $\zeta^2 > 1$: two distinct real roots

$$r_1 = \omega_0(-\zeta + \sqrt{\zeta^2 - 1}),\ m(r_1) = 1,\ \ r_2 = \omega_0(-\zeta - \sqrt{\zeta^2 - 1}),\ m(r_2) = 1;$$

2. $\zeta = 1$: one repeated real root

$$r_1 = -\omega_0\zeta,\ m(r_2) = 2;$$

3. $\zeta^2 < 1$: a complex conjugate pair of roots with

$$\rho_1 = \omega_0(-\zeta + \mathrm{i}\sqrt{1 - \zeta^2}),\ m(\rho_1) = 1.$$

This then gives rise, according to Procedure 2.2.18, to the following solutions of the differential equation:

1. $\zeta^2 > 1$: $\xi(t) = c_1\mathrm{e}^{\omega_0(-\zeta+\sqrt{\zeta^2-1})t} + c_2\mathrm{e}^{\omega_0(-\zeta-\sqrt{\zeta^2-1})t}$;
2. $\zeta^2 = 1$: $\xi(t) = c_1\mathrm{e}^{-\omega_0\zeta t} + c_2 t\mathrm{e}^{-\omega_0\zeta t}$;
3. $\zeta^2 < 1$: $\xi(t) = c_1\mathrm{e}^{-\omega_0\zeta t}\cos(\omega_0\sqrt{1-\zeta^2}t) + c_2\mathrm{e}^{-\omega_0\zeta t}\sin(\omega_0\sqrt{1-\zeta^2}t)$.

To determine the constants c_1 and c_2, one applies initial conditions. Let us keep things simple and prescribe initial conditions

$$\xi(0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(0) = x_0^{(1)}.$$

Skipping the tedious manipulations...

1. $\zeta^2 > 1$:

$$c_1 = \frac{\omega_0(\zeta + \sqrt{\zeta^2 - 1})x_0 + x_0^{(1)}}{2\omega_0\sqrt{\zeta^2 - 1}},$$
$$c_2 = \frac{-\omega_0(\zeta - \sqrt{\zeta^2 - 1})x_0 - x_0^{(1)}}{2\omega_0\sqrt{\zeta^2 - 1}};$$

2. $\zeta^2 = 1$:

$$c_1 = x_0,$$
$$c_2 = \omega_0\zeta x_0 + x_0^{(1)};$$

3. $\zeta^2 < 1$:

$$c_1 = x_0,$$
$$c_2 = \frac{\omega_0\zeta x_0 + x_0^{(1)}}{\omega_0\sqrt{1 - \zeta^2}}.$$

In Figure 2.2 we graph solutions for fixed ω_0 and varying ζ. We $\zeta > 0$ we say the

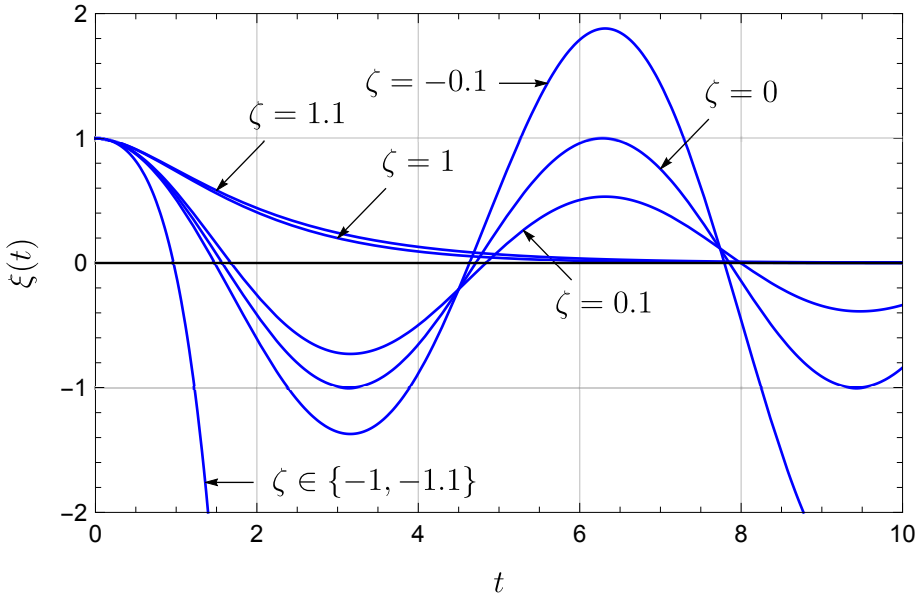


Figure 2.2 Solutions of a second-order scalar linear homogeneous ordinary differential equation with $\omega_0 = 1$, $\xi(0) = 1$, and $\frac{d\xi}{dt}(0) = 0$

equation is ***positively damped***, when $\zeta = 0$ we say the equation is ***undamped***, and when $\zeta < 0$ we say the equation is ***negatively damped***. In practice, systems are

positively damped, or possibly undamped. So let us focus on this situation for a moment. Here we break things down into $\zeta < 1$, which is called ***underdamped***, $\zeta = 1$ which is called ***critically damped***, and $\zeta > 1$ which is called ***overdamped***. The underdamped case is distinguished by there being oscillations in the motion, corresponding to the imaginary part of the roots. In the critical and overdamped cases, we do not get this oscillation. •

Exercises

2.2.1 Consider the ordinary differential equation F with right-hand side given by (2.1).

(a) Convert this to a first-order equation with k states, following Exercise 1.3.23.

(b) Show that, if the functions $a_0, a_1, \dots, a_k$ are continuous, then the resulting first-order equation satisfies the conditions of Theorem 1.4.8 for existence of a unique solution $t \mapsto \xi(t)$ satisfying the initial conditions

$$\xi(t_0) = x_0, \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) = x_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t)(t_0) = x_0^{(k-1)}$$

at time $t_0 \in \mathbb{T}$.

2.2.2 Let $a, b, c, \omega, \phi \in \mathbb{R}$ and define

$$\xi_1(t) = a\cos(\omega t + \phi), \quad \xi_2(t) = b\cos(\omega t) + c\sin(\omega t).$$

Show that $\xi_1, \xi_2 \in \mathrm{Sol}(F)$ where F is the second-order scalar linear homogeneous ordinary differential equation with constant coefficients whose right-hand side is

$$\widehat{F}(t, x, x^{(1)}) = -\omega^2 x.$$

Explain in at least two ways why this is not a violation of Proposition 2.2.1 concerning uniqueness of solutions.

2.2.3 In each of the following cases, show that the functions given are a basis for $\mathrm{Sol}(F)$ with F as given:

(a) take

$$F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - x$$

and

$$\xi_1(t) = \mathrm{e}^t, \quad \xi_2(t) = \mathrm{e}^{-t};$$

(b) take

$$F(t, x, x^{(1)}, x^{(2)}, x^{(3)}) = x^{(3)} + 4x^{(2)} + 4x^{(1)}$$

and

$$\xi_1(t) = 1, \quad \xi_2(t) = \mathrm{e}^{-2t}, \quad \xi_3(t) = t\mathrm{e}^{-2t}.$$

(c) take
$$F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + \omega^2 x$$
and
$$\xi_1(t) = \cos(\omega t), \quad \xi_2(t) = \sin(\omega t).$$

(d) take
$$F(t, x, x^{(1)}, x^{(2)}) = t^2 x^{(2)} + t x^{(1)} - x$$
and
$$\xi_1(t) = t, \quad \xi_2(t) = t^{-1}$$
(here the time-domain must be an interval not containing 0).

2.2.4 For each of the ordinary differential equations F of Exercise 2.2.3, give the general form of a solution of the differential equation, i.e., the general form of $t \mapsto \xi(t)$ satisfying
$$F\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t)\right) = 0.$$

2.2.5 For each of the ordinary differential equations F of Exercise 2.2.3 for which you found a general form of their solution in Exercise 2.2.4, give the solution satisfying the given initial conditions:
(a) $\xi(0) = 1$ and $\dot{\xi}(0) = 1$;
(b) $\xi(0) = 1$, $\dot{\xi}(0) = 1$, and $\ddot{\xi}(0) = 1$;
(c) $\xi(0) = 1$ and $\dot{\xi}(0) = 0$;
(d) $\xi(1) = 1$ and $\dot{\xi}(1) = 1$.

2.2.6 If possible, find the characteristic polynomial for the following scalar ordinary differential equations:
(a) $F(t, x, x^{(1)}) = x^{(1)} + tx$;
(b) $F(t, x, x^{(1)}) = x^{(1)} + 3x$;
(c) $F(t, x, x^{(1)}, x^{(2)}) = 2x^{(2)} - x^{(1)} + 8x$;
(d) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + \frac{a_g}{\ell}\sin(x)$;
(e) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + \omega^2 x$;
(f) $F(t, x, x^{(1)}, \dots, x^{(k)}) = a_k x^{(k)} + \cdots + a_1 x^{(1)} + a_0 x$.

2.2.7 Find the unique lowest degree normalised scalar linear homogeneous ordinary differential equation with constant coefficients whose characteristic polynomial has the following roots:
(a) $\{-1, 2\}$;
(b) $\{2 + 2\mathrm{i}, 2 - 2\mathrm{i}, -2\}$;
(c) $\{-\frac{1}{\tau}\}$, $\tau \in \mathbb{R} \setminus \{0\}$;
(d) $\{-a, -a, 2\}$, $a \in \mathbb{R}$;
(e) $\{\omega_0(-\zeta + \mathrm{i}\sqrt{1-\zeta^2}), \omega_0(-\zeta - \mathrm{i}\sqrt{1-\zeta^2})\}$, $\omega_0, \zeta \in \mathbb{R}$, $\omega_0 \neq 0$, $|\zeta| \leq 1$;

(f) $\{\sigma + \mathrm{i}\omega, \sigma - \mathrm{i}\omega\}$, $\sigma, \omega \in \mathbb{R}$, $\omega \neq 0$.

2.2.8 Find the unique lowest degree normalised scalar linear homogeneous ordinary differential equation with the following functions as a fundamental set of solutions:

(a) $\xi_1(t) = \mathrm{e}^{-t}$ and $\xi_2(t) = \mathrm{e}^{2t}$;
(b) $\xi_1(t) = \mathrm{e}^{2t}\cos(2t)$, $\xi_2(t) = \mathrm{e}^{2t}\sin(2t)$, $\xi_3(t) = \mathrm{e}^{-2t}$;
(c) $\xi_1(t) = \mathrm{e}^{-t/\tau}$, $\tau \in \mathbb{R} \setminus \{0\}$;
(d) $\xi_1(t) = \mathrm{e}^{-at}$, $\xi_2(t) = t\mathrm{e}^{-at}$, $a \in \mathbb{R}$, and $\xi_3(t) = \mathrm{e}^2 t$;
(e) $\xi_1(t) = \mathrm{e}^{-\omega_0\zeta t}\cos(\omega_0\sqrt{1-\zeta^2}t)$ and $\xi_2(t) = \mathrm{e}^{-\omega_0\zeta t}\sin(\omega_0\sqrt{1-\zeta^2}t)$, $\omega_0, \zeta \in \mathbb{R}$, $\omega_0 \neq 0$, $|\zeta| \leq 1$;
(f) $\xi_1(t) = \mathrm{e}^{\sigma t}\cos(\omega t)$ and $\xi_2(t) = \mathrm{e}^{\sigma t}\sin(\omega t)$, $\sigma, \omega \in \mathbb{R}$, $\omega \neq 0$.

2.2.9 In Proposition 2.2.11 it is proved that the set of solutions for a scalar linear inhomogeneous ordinary differential uniquely determines the differential equation. Show how you would, given a fundamental set of solutions to a homogeneous such equation, with constant coefficients, recover the coefficients in the differential equation.

2.2.10 Solve the following initial value problems:

(a) $\dot{\xi}(t) + 3\xi(t) = 0$, $\xi(0) = 4$;
(b) $\ddot{\xi}(t) - 4\dot{\xi}(t) + 4\xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 1$;
(c) $\ddot{\xi}(t) - 4\dot{\xi}(t) - 4\xi(t) = 0$, $\xi(0) = 1$, $\dot{\xi}(0) = 1$;
(d) $\dddot{\xi}(t) - 7\ddot{\xi}(t) + 15\dot{\xi}(t) - 9\xi(t) = 0$, $\xi(0) = 1$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 1$;
(e) $\dddot{\xi}(t) + 3\ddot{\xi}(t) + 4\dot{\xi}(t) + 2\xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 2$;
(f) $\ddddot{\xi}(t) + \dddot{\xi}(t) + \ddot{\xi}(t) + \dot{\xi}(t) + \xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 0$, $\ddot{\xi}(0) = 0$, $\dddot{\xi}(0) = 0$.

Section 2.3

Scalar linear inhomogeneous ordinary differential equations

In this section we still consider scalar linear ordinary differential equations, but now we consider the inhomogeneous case. We still, this have a time-domain $\mathbb{T}$ and the state space $U = \mathbb{R}$, but now we have a right-hand side of the form

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}(t)x^{(k-1)} - \cdots - a_1(t)x^{(1)} - a_0 x + b(t) \tag{2.11}$$

for functions $a_0, a_1, a_{k-1}, b \colon \mathbb{T} \to \mathbb{R}$. Thus solutions $t \mapsto \xi(t)$ satisfy

$$\frac{\mathrm{d}^k \xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t).$$

We shall proceed in this section much as in the preceding section, first saying some things about the general case, and then focussing on the case where F has constant coefficients, as in this case there is more that can be said.

2.3.1 Equations with time-varying coefficients

We begin by stating some general properties of general scalar linear inhomogeneous ordinary differential equations.

2.3.1.1 Solutions and their properties First we state the local existence and uniqueness result that one needs to get off the ground for any class of differential equations.

2.3.1 Proposition (Local existence and uniqueness of solutions for scalar linear homogeneous ordinary differential equations) *Consider the linear inhomogeneous ordinary differential equation* F *with right-hand side equation (2.11) and suppose that the functions* $\mathrm{a}_0, \mathrm{a}_1, \dots, \mathrm{a}_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let*

$$(\mathrm{t}_0, \mathrm{x}_0, \mathrm{x}_0^{(1)}, \dots, \mathrm{x}_{k-1}^{(0)}) \in \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\le k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}).$$

Then there exists an interval $\mathbb{T}' \subseteq \mathbb{T}$ *and a map* $\xi \colon \mathbb{T}' \to \mathbb{R}$ *of class* C^1 *that is a solution for* F *and which satisfies*

$$\xi(\mathrm{t}_0) = \mathrm{x}_0, \frac{\mathrm{d}\xi}{\mathrm{d}t}(\mathrm{t}_0) = \mathrm{x}_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(\mathrm{t})(\mathrm{t}_0) = \mathrm{x}_0^{(k-1)}.$$

Moreover, if $\tilde{\mathbb{T}}' \subseteq \mathbb{T}$ *is another subinterval and if* $\tilde{\xi} \colon \tilde{\mathbb{T}}' \to \mathbb{R}$ *is another* C^k*-solution for* F *satisfying*

$$\tilde{\xi}(\mathrm{t}_0) = \mathrm{x}_0, \frac{\mathrm{d}\tilde{\xi}}{\mathrm{d}t}(\mathrm{t}_0) = \mathrm{x}_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\tilde{\xi}}{\mathrm{d}t^{k-1}}(\mathrm{t})(\mathrm{t}_0) = \mathrm{x}_0^{(k-1)},$$

then $\tilde{\xi}(\mathrm{t}) = \xi(\mathrm{t})$ *for every* $\mathrm{t} \in \tilde{\mathbb{T}}' \cap \mathbb{T}'$.

Proof This is Exercise 2.3.1. ■

As with homogeneous equations, for the scalar linear inhomogeneous ordinary differential equations we can show that solutions exist for all times.

2.3.2 Proposition (Global existence of solutions for scalar linear inhomogeneous ordinary differential equations) *Consider the linear in homogeneous ordinary differential equation* F *with right-hand side equation (2.11) and suppose that the functions* $a_0, a_1, \dots, a_{k-1}, b\colon \mathbb{T} \to \mathbb{R}$ *are continuous. If* $\xi\colon \mathbb{T}' \to \mathbb{R}$ *is a solution for* F*, then there exists a solution* $\overline{\xi}\colon \mathbb{T} \to \mathbb{R}$ *for which* $\overline{\xi}|\mathbb{T}' = \xi$.

Proof Note that, as per Exercise 1.3.23, we can convert the differential equation F into a first-order differential equation linear homogeneous differential equation with states $(x, x^{(1)}, \dots, x^{(k-1)})$. Thus the result will follow from the analogous result for first-order systems of equations, and this is stated and proved as Proposition 3.3.2. ■

As in the homogeneous case, we can now talk sensibly about the set of *all* solutions for F. Thus we can define

$$\operatorname{Sol}(F) = \Big\{\xi\colon \mathbb{T} \to \mathbb{R} \Big| \; \xi \text{ satisfies } \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t)\Big\},$$

which is exactly this set of all solutions for F. While $\operatorname{Sol}(F)$ was a vector space in the homogeneous case, in the inhomogeneous case this is no longer the case. However, the set of all solutions for the homogeneous case plays an important rôle, even in the homogeneous case. To organise this discussion, we let F_{h} be the "homogeneous part" of F. Thus the right-hand side of F_{h} is

$$\widehat{F_{\mathrm{h}}}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}(t)x^{(k-1)} - \cdots - a_1(t)x^{(1)} - a_0(t)x.$$

As in Theorem 2.2.3, $\operatorname{Sol}(F_{\mathrm{h}})$ is a $\mathbb{R}$-vector space of dimension k. We can now state the character of $\operatorname{Sol}(F)$.

2.3.3 Theorem (Affine space structure of sets of solutions) *Consider the linear inhomogeneous ordinary differential equation* F *with right-hand side equation (2.11) and suppose that the functions* $a_0, a_1, \dots, a_{k-1}, b\colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let* $\xi_{\mathrm{p}} \in \operatorname{Sol}(\mathrm{F})$*. Then*

$$\operatorname{Sol}(\mathrm{F}) = \{\xi + \xi_{\mathrm{p}} \mid \xi \in \operatorname{Sol}(\mathrm{F_h})\}.$$

Proof First note that, by Theorem 2.2.3, $\operatorname{Sol}(F) \neq \emptyset$ and so there exists some $\xi_{\mathrm{p}} \in \operatorname{Sol}(F)$. We have, of course,

$$\frac{\mathrm{d}^k\xi_{\mathrm{p}}}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_{\mathrm{p}}}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t). \tag{2.12}$$

Next let $\xi \in \mathrm{Sol}(F)$ so that

$$\frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t). \tag{2.13}$$

Subtracting (2.12) from (2.13) we get

$$\frac{\mathrm{d}^k(\xi - \xi_{\mathrm{p}})}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}(\xi - \xi_{\mathrm{p}})}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}(\xi - \xi_{\mathrm{p}})}{\mathrm{d}t}(t) + a_0(t)(\xi - \xi_{\mathrm{p}})(t) = 0,$$

i.e., $\xi - \xi_{\mathrm{p}} \in \mathrm{Sol}(F_{\mathrm{h}})$. In other words, $\xi = \tilde{\xi} + \xi_{\mathrm{p}}$ for $\tilde{\xi} \in \mathrm{Sol}(F_{\mathrm{h}})$.

Conversely, suppose that $\xi = \tilde{\xi} + \xi_{\mathrm{p}}$ for $\tilde{\xi} \in \mathrm{Sol}(F_{\mathrm{h}})$. Then

$$\frac{\mathrm{d}^k\tilde{\xi}}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\tilde{\xi}}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\tilde{\xi}}{\mathrm{d}t}(t) + a_0(t)\tilde{\xi}(t) = 0. \tag{2.14}$$

Adding (2.12) and (2.14) we get

$$\frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t),$$

i.e., $\xi \in \mathrm{Sol}(F)$. ■

2.3.4 Remark (Comparison of Theorem 2.3.3 with systems of linear algebraic equations) The reader should compare here the result of Theorem 2.3.3 with the situation concerning linear algebraic equations of the form $L(u) = v_0$, for vector spaces U and V, a linear map $L \in \mathrm{L}(\mathsf{U}; \mathsf{V})$, and a fixed $v_0 \in \mathsf{V}$. In particular, we can make reference to Proposition 1.2.4. In the setting of scalar linear inhomogeneous ordinary differential equations, we have

$$\begin{aligned}
\mathsf{U} &= \mathsf{C}^k(\mathbb{T}; \mathbb{R}),\\
\mathsf{V} &= \mathsf{C}^0(\mathbb{T}; \mathbb{R}),\\
L(f)(t) &= \frac{\mathrm{d}^k f}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1} f}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}f}{\mathrm{d}t}(t) + a_0(t)f(t),\\
v_0 &= b.
\end{aligned}$$

Then Propositions 2.3.1 and 2.3.2 tell us that L is surjective, and so $v_0 \in \mathrm{image}(L)$. Thus we are in case (ii) of Proposition 1.2.4, which exactly the statement of Theorem 2.3.3. Note that L is not injective, since Theorem 2.2.3 tells us that $\dim_{\mathbb{R}}(\ker(L)) = k$. •

Note that Theorem 2.3.3 tells us that, to solve a scalar linear inhomogeneous ordinary differential equation, we must do two things: (1) find *some* solution for the equation; (2) find *all* solutions for the homogeneous part. Then we know our solution will be found amongst the set of sums of these. Generally, both

of these things is impossible, in any general way. We do know, however, that Procedure 2.2.18 can be used, in principle, to find all solutions for the homogeneous part. Thus one need only find some solution of the equation in this case. Upon finding such a solution, one calls it a ***particular solution***. *Note that there are many particular solutions.* Indeed, Proposition 2.2.1 tells us that there is one solution for every set of initial conditions. So one should always speak of *a* particular solution, not *the* particular solution.

2.3.1.2 Finding a particular solution using the Wronskian So... how do we find a particular solution? In this section we outline a general (and not very efficient) way of arriving at some such solution, using the Wronskian of Definition 2.2.6. To state the result, suppose that we have a fundamental set of solutions $\{\xi_1, \dots, \xi_k\}$ for F_{h}, where F has right-hand side (2.11), and denote

$$\mathrm{W}_{b,j}(\xi_1, \dots, \xi_k)(t) = \det \begin{bmatrix} \xi_1(t) & \cdots & \xi_{j-1}(t) & 0 & \xi_{j+1}(t) & \cdots & \xi_k(t) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-2}\xi_1}{\mathrm{d}t^{k-2}}(t) & \cdots & \frac{\mathrm{d}^{k-2}\xi_{j-1}}{\mathrm{d}t^{k-2}}(t) & 0 & \frac{\mathrm{d}^{k-2}\xi_{j+1}}{\mathrm{d}t^{k-2}}(t) & \cdots & \frac{\mathrm{d}^{k-2}\xi_k}{\mathrm{d}t^{k-2}}(t) \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}\xi_{j-1}}{\mathrm{d}t^{k-1}}(t) & b(t) & \frac{\mathrm{d}^{k-1}\xi_{j+1}}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix},$$

for $j \in \{1, \dots, k\}$, i.e., $\mathrm{W}_{b,j}(\xi_1, \dots, \xi_k)(t)$ is the determinant of the matrix used to compute the Wronskian, but with the jth column replaced by $(0, \dots, 0, b(t))$.

We then have the following result.

2.3.5 Proposition (A particular solution using Wronskians) *Consider the linear in homogeneous ordinary differential equation* F *with right-hand side equation (2.11) and suppose that the functions* $\mathrm{a}_0, \mathrm{a}_1, \dots, \mathrm{a}_{k-1}, \mathrm{b} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let* $\{\xi_1, \dots, \xi_k\}$ *be a fundamental set of solutions for* F_{h} *and let* $\mathrm{t}_0 \in \mathbb{T}$*. Then the function* $\xi_{\mathrm{p}} \colon \mathbb{T} \to \mathbb{R}$ *defined by*

$$\xi_{\mathrm{p}}(\mathrm{t}) = \sum_{\mathrm{j}=1}^{\mathrm{k}} \xi_{\mathrm{j}}(\mathrm{t}) \int_{\mathrm{t}_0}^{\mathrm{t}} \frac{\mathrm{W}_{\mathrm{b,j}}(\xi_1, \dots, \xi_k)(\tau)}{\mathrm{W}(\xi_1, \dots, \xi_k)(\tau)} \, \mathrm{d}\tau$$

is a particular solution for F.

Proof Let us define

$$c_j(t) = \int_{t_0}^{t} \frac{\mathrm{W}_{b,j}(\xi_1, \dots, \xi_k)(\tau)}{\mathrm{W}(\xi_1, \dots, \xi_k)(\tau)} \, \mathrm{d}\tau, \qquad j \in \{1, \dots, k\},\ t \in \mathbb{T},$$

so that

$$\frac{\mathrm{d}c_j}{\mathrm{d}t}(t) = \frac{\mathrm{W}_{b,j}(\xi_1, \dots, \xi_k)(t)}{\mathrm{W}(\xi_1, \dots, \xi_k)(t)}, \qquad j \in \{1, \dots, k\},\ t \in \mathbb{T}.$$

Note that this is equivalent, by Cramer's Rule for linear systems of algebraic equations, to the set of equations

$$\begin{bmatrix} \xi_1(t) & \xi_2(t) & \cdots & \xi_k(t) \\ \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}\xi_k}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t) & \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}c_1}{\mathrm{d}t}(t) \\ \frac{\mathrm{d}c_2}{\mathrm{d}t}(t) \\ \vdots \\ \frac{\mathrm{d}c_k}{\mathrm{d}t}(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ b(t) \end{bmatrix}, \qquad t \in \mathbb{T}. \tag{2.15}$$

Note that the proposition is then that

$$\xi_{\mathrm{p}}(t) = \sum_{j=1}^{k} c_j(t)\xi_j(t), \qquad t \in \mathbb{T},$$

defines a particular solution for F. This we shall prove by direct computation.

We compute

$$\frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t) = \sum_{j=1}^{k} \frac{\mathrm{d}c_j}{\mathrm{d}t}(t)\xi_j(t) + \sum_{j=1}^{k} c_j(t)\frac{\mathrm{d}\xi_j}{\mathrm{d}t}(t) = \sum_{j=1}^{k} c_j(t)\frac{\mathrm{d}\xi_j}{\mathrm{d}t}(t)$$

for $t \in \mathbb{T}$, using the first of equations (2.15). Repeatedly differentiating and using successive equations from (2.15), we deduce that

$$\frac{\mathrm{d}^j\xi_{\mathrm{p}}}{\mathrm{d}t^j}(t) = \sum_{j=1}^{k} c_j(t)\frac{\mathrm{d}^j\xi}{\mathrm{d}t^j}(t), \qquad j \in \{0, 1, \ldots, k-1\},\ t \in \mathbb{T}.$$

We also have, using the last of equations (2.15),

$$\frac{\mathrm{d}^k\xi_{\mathrm{p}}}{\mathrm{d}t^k}(t) = \sum_{j=1}^{k} \frac{\mathrm{d}c_j(t)}{\mathrm{d}t}(t)\frac{\mathrm{d}^{k-1}\xi_j}{\mathrm{d}t^{k-1}}(t) + \sum_{j=1}^{k} c_j(t)\frac{\mathrm{d}^k\xi_j}{\mathrm{d}t^k}(t) = \sum_{j=1}^{k} c_j(t)\frac{\mathrm{d}^k\xi_j}{\mathrm{d}t^k}(t) + b(t).$$

Therefore, combining these calculations,

$$\begin{aligned} &\frac{\mathrm{d}^k\xi_{\mathrm{p}}(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_{\mathrm{p}}}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t) + a_0(t)\xi(t) \\ &= \sum_{j=1}^{k} c_j(t)\left(\frac{\mathrm{d}^k\xi_j(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_j}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_j}{\mathrm{d}t}(t) + a_0(t)\xi_j(t)\right) + b(t) = b(t), \end{aligned}$$

using the fact that $\xi_1, \ldots, \xi_k$ are solutions for F_{h}. Thus ξ_{p} is indeed a particular solution. ■

Let us illustrate this result on an example.

2.3.6 Example (First-order scalar linear inhomogeneous ordinary differential equations) We consider here the first-order equation F with right-hand side

$$\widehat{F}(t,x) = -a(t)x + b(t)$$

for continuous functions $a, b\colon \mathbb{T} \to \mathbb{R}$. We have seen in Example 2.2.5 that a fundamental set of solutions is given by $\{\xi_1(t)\}$, with

$$\xi_1(t) = \mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau}$$

for some $t_0 \in \mathbb{T}$. Therefore,

$$\mathrm{W}(\xi_1)(t) = \det\Big[\xi_1(t)\Big] = \xi_1(t), \quad \mathrm{W}(\xi_1)_{b,1} = \det\Big[b(t)\Big] = b(t).$$

Thus

$$\begin{aligned}\xi_{\mathrm{p}}(t) &= \xi_1(t)\left(\int_{t_0}^t \frac{b(\tau)}{\xi_1(\tau)}\,\mathrm{d}\tau\right)\\ &= \mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau}\int_{t_0}^t b(\tau)\mathrm{e}^{\int_{t_0}^\tau a(\sigma)\,\mathrm{d}\sigma}\mathrm{d}\tau\end{aligned}$$

defines a particular solution for F. Thus, as in Theorem 2.3.3, *any* solution for F has the form

$$\xi(t) = C\mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau} + \mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau}\int_{t_0}^t b(\tau)\mathrm{e}^{\int_{t_0}^\tau a(\sigma)\,\mathrm{d}\sigma}\mathrm{d}\tau$$

for some $C \in \mathbb{R}$. In we apply an initial condition $\xi(t_0) = x_0$, then we see that $C = x_0$. Therefore, finally, we have the solution

$$\xi(t) = x_0\mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau} + \mathrm{e}^{-\int_{t_0}^t a(\tau)\,\mathrm{d}\tau}\int_{t_0}^t b(\tau)\mathrm{e}^{\int_{t_0}^\tau a(\sigma)\,\mathrm{d}\sigma}\mathrm{d}\tau$$

to the initial value problem

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a(t)\xi(t) = b(t), \quad \xi(t_0) = x_0.$$

Because we have expressed the solution of a differential equation as an integral, we declare victory![2] •

2.3.1.3 The Green's function In this section we describe another means of determining a particular solution. In this case, what we arrive at is a description of a particular solution that allows for the inhomogeneous term "b" to be plugged into an integral. We shall see a close variant of this in Section 3.3 when we discuss linear inhomogeneous *systems* of equations.

The result is the following.

[2] Because victories are few and far between in the business of solving differential equations.

2.3.7 Theorem (Existence of, and properties of, the Green's function) *Consider the linear homogeneous ordinary differential equation* F *with right-hand side equation (2.1) and suppose that the functions* $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$ *are continuous. Let* $t_0 \in \mathbb{T}$ *and denote* $\mathbb{T}_{t_0} = \mathbb{T} \cap [t_0, \infty)$.

$$F_{t_0} \colon \mathbb{T}_{t_0} \times \mathbb{R} \oplus L^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

be defined by $F_{t_0}(t, x, x^{(1)}, \dots, x^{(k)}) = F(t, x, x^{(1)}, \dots, x^{(k)})$. *If* $b \in C^0(\mathbb{T}; \mathbb{R})$ *then define the inhomogeneous ordinary differential equation*

$$\begin{aligned} F_{t_0,b} \colon \mathbb{T}_{t_0} \times \mathbb{R} \oplus L^{\le k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) &\to \mathbb{R} \\ (t, x, x^{(1)}, \dots, x^{(k)}) &\mapsto F_{t_0}(t, x, x^{(1)}, \dots, x^{(k)}) - b(t). \end{aligned}$$

Then there exists

$$G_{F,t_0} \colon \mathbb{T}_{t_0} \times \mathbb{T}_{t_0} \to \mathbb{R}$$

with the following properties:

(i) $\dfrac{\partial^l G_{F,t_0}}{\partial t^l}$ *is continuous for* $l \in \{0, 1, \dots, k-2\}$;

(ii) $\dfrac{\partial^l G_{F,t_0}}{\partial t^l}$ *is continuous on*

$$\{(t, \tau) \in \mathbb{T}_{t_0} \times \mathbb{T}_{t_0} \mid t \neq \tau\}$$

for $l \in \{k-1, k\}$;

(iii) for $\tau \in \mathbb{T}_{t_0}$, *we have*

$$\lim_{t \uparrow \tau} \frac{\partial^l G_{F,t_0}}{\partial t^l}(t, \tau) = 0, \quad \lim_{t \downarrow \tau} \frac{\partial^l G_{F,t_0}}{\partial t^l}(t, \tau) = 0, \qquad l \in \{0, 1, \dots, k-2\},$$

and

$$\lim_{t \uparrow \tau} \frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(t, \tau) = 0, \quad \lim_{t \downarrow \tau} \frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(t, \tau) = 1;$$

(iv) $\dfrac{\partial^l G_{F,t_0}}{\partial t^l}(t_0, \tau) = 0, l \in \{0, 1, \dots, k-1\}$;

(v) for $t \in \mathbb{T}_{t_0} \setminus \{\tau\}$ *we have*

$$\frac{\partial^k G_{F,t_0}}{\partial t^k}(t, \tau) + a_{k-1}(t) \frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(t, \tau) + \cdots + a_1(t) \frac{\partial G_{F,t_0}}{\partial t}(t, \tau) + a_0(t) G_{F,t_0}(t, \tau) = 0;$$

(vi) if $b \in C^o(\mathbb{T}; \mathbb{R})$ *and if* $\xi_{p,b} \colon \mathbb{T}_{t_0} \to \mathbb{R}$ *is given by*

$$\xi_{p,b}(t) = \int_{t_0}^{t} G_{F,t_0}(t, \tau) b(\tau) \, d\tau,$$

then $\xi_{p,b} \in \mathrm{Sol}(F_{t_0,b})$.

Moreover, there is only one such function satisfying all of the above properties.

Proof Let $\{\xi_1, \dots, \xi_k\}$ be a fundamental set of solutions and define

$$g_j(s) = (-1)^{k+j} \det \begin{bmatrix} \xi_1(s) & \cdots & \xi_{j-1}(s) & \xi_{j+1}(s) & \cdots & \xi_k(s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\mathrm{d}^{k-2}\xi_1}{\mathrm{d}t^{k-2}}(s) & \cdots & \dfrac{\mathrm{d}^{k-2}\xi_{j-1}}{\mathrm{d}t^{k-2}}(s) & \dfrac{\mathrm{d}^{k-2}\xi_{j+1}}{\mathrm{d}t^{k-2}}(s) & \cdots & \dfrac{\mathrm{d}^{k-2}\xi_k}{\mathrm{d}t^{k-2}}(s) \end{bmatrix},$$

for $j \in \{1, \dots, k\}$, and observe that we also have

$$g_j(s) = \det \begin{bmatrix} \xi_1(s) & \cdots & \xi_{j-1}(s) & 0 & \xi_{j+1}(s) & \cdots & \xi_k(s) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\mathrm{d}^{k-2}\xi_1}{\mathrm{d}t^{k-2}}(s) & \cdots & \dfrac{\mathrm{d}^{k-2}\xi_{j-1}}{\mathrm{d}t^{k-2}}(s) & 0 & \dfrac{\mathrm{d}^{k-2}\xi_{j+1}}{\mathrm{d}t^{k-2}}(s) & \cdots & \dfrac{\mathrm{d}^{k-2}\xi_k}{\mathrm{d}t^{k-2}}(s) \\ \dfrac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(s) & \cdots & \dfrac{\mathrm{d}^{k-1}\xi_{j-1}}{\mathrm{d}t^{k-1}}(s) & 1 & \dfrac{\mathrm{d}^{k-1}\xi_{j+1}}{\mathrm{d}t^{k-1}}(s) & \cdots & \dfrac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(s) \end{bmatrix}, \tag{2.16}$$

for $j \in \{1, \dots, k\}$. Note, then, that $\xi_{\mathrm{p}} \colon \mathbb{T}_{t_0} \to \mathbb{R}$ defined by

$$\xi_{\mathrm{p}}(t) = \sum_{j=1}^{k} \xi_j(t) \int_{t_0}^{t} \frac{g_j(s)}{\mathrm{W}(\xi_1, \dots, \xi_k)(s)} \, \mathrm{d}s$$

is the particular solution with "$b(t) = 1$" from Proposition 2.3.5. Now define $c_j \colon \mathbb{T}_{t_0} \to \mathbb{R}$, $j \in \{1, \dots, k\}$, by

$$c_j(\tau) = \frac{g_j(\tau)}{\mathrm{W}(\xi_1, \dots, \xi_k)(\tau)} - \frac{\mathrm{W}(\xi_1, \dots, \xi_{j-1}, \xi_{\mathrm{p}}, \xi_{j+1}, \dots, \xi_k)(\tau)}{\mathrm{W}(\xi_1, \dots, \xi_k)(\tau)}.$$

Finally, take

$$G_{F,t_0}(t, \tau) = 0, \qquad t \leq \tau,$$

and

$$G_{F,t_0}(t, \tau) = \xi_{\mathrm{p}}(t) + \sum_{j=1}^{k} c_j(\tau)\xi_j(t), \qquad t > \tau.$$

The definition immediately gives part (iv).

We also immediately deduce part (v), by virtue of Proposition 2.3.5 and Theorem 2.3.3.

For a function $f \colon \mathbb{T}_{t_0} \times \mathbb{T}_{t_0} \to \mathbb{R}$, let us denote

$$f(\tau+, \tau) = \lim_{t \downarrow \tau} f(t, \tau).$$

Note that

$$\frac{\partial^l G_{F,t_0}}{\partial t^l}(t, \tau) = \frac{\mathrm{d}^l \xi_{\mathrm{p}}}{\mathrm{d}t^l}(t) + \sum_{j=1}^{l} c_j(\tau) \frac{\mathrm{d}^l \xi_j}{\mathrm{d}t^l}(t), \qquad l \in \{0, 1, \dots, k-1\}, \ t > \tau,$$

which gives

$$\frac{\partial^l G_{F,t_0}}{\partial t^l}(\tau+,\tau) = \frac{\mathrm{d}^l \xi_{\mathrm{p}}}{\mathrm{d}t^l}(\tau) + \sum_{j=1}^{l} c_j(\tau)\frac{\mathrm{d}^l \xi_j}{\mathrm{d}t^l}(\tau), \qquad l \in \{0,1,\dots,k-1\}.$$

We then have

$$\begin{bmatrix} \xi_1(t) & \xi_2(t) & \cdots & \xi_k(t) \\ \frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) & \frac{\mathrm{d}\xi_2}{\mathrm{d}t}(t) & \cdots & \frac{\mathrm{d}\xi_k}{\mathrm{d}t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^{k-1}\xi_1}{\mathrm{d}t^{k-1}}(t) & \frac{\mathrm{d}^{k-1}\xi_2}{\mathrm{d}t^{k-1}}(t) & \cdots & \frac{\mathrm{d}^{k-1}\xi_k}{\mathrm{d}t^{k-1}}(t) \end{bmatrix} \begin{bmatrix} c_1(\tau) \\ c_2(\tau) \\ \vdots \\ c_k(\tau) \end{bmatrix} = \begin{bmatrix} G_{F,t_0}(t,\tau) - \xi_{\mathrm{p}}(t) \\ \frac{\partial G_{F,t_0}}{\partial t}(t,\tau) - \frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t) \\ \vdots \\ \frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(t,\tau) - \frac{\mathrm{d}^{k-1}\xi_{\mathrm{p}}}{\mathrm{d}t^{k-1}}(t) \end{bmatrix}. \tag{2.17}$$

Let us denote by $\mathrm{W}_{F,t_0,j}(\xi_1,\dots,\xi_k)(t,\tau)$ the determinant of the matrix on the left-hand side of (2.17), but with the jth column replaced by the vector on the right-hand side of (2.16). By Cramer's Rule, (2.17) is equivalent to

$$c_j(\tau) = \frac{\mathrm{W}_{F,t_0,j}(\xi_1,\dots,\xi_k)(\tau+,\tau)}{\mathrm{W}(\xi_1,\dots,\xi_k)(\tau)}, \qquad j \in \{1,\dots,k\}.$$

By our observation (2.16) and the definition of the functions $c_1,\dots,c_k$, this implies that (2.17) is equivalent to

$$\begin{aligned} \frac{\partial^l G_{F,t_0}}{\partial t^l}(\tau+,\tau) &= 0, \qquad l \in \{0,1,\dots,k-2\}, \\ \frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(\tau+,\tau) &= 1. \end{aligned}$$

This gives the "$t \downarrow \tau$" limits for part (iii). The "$t \uparrow \tau$" limits for part (iii) follow since $G_{F,t_0}(t,\tau) = 0$ for $t \leq \tau$.

Since $\tau \mapsto c_j(\tau)$, $j \in \{1,\dots,k\}$, and the first k derivatives of $t \mapsto \xi_{\mathrm{p}}(t)$ and $t \mapsto \xi_j(t)$, $j \in \{1,\dots,k\}$, are continuous for $t > \tau$, and since $G_{F,t_0}(t,\tau) = 0$ for $t \leq \tau$, we conclude that $\frac{\partial^l G_{F,t_0}}{\partial t^l}$, $l \in \{0,1,\dots,n\}$, is continuous away from points where $t = \tau$. This gives part (ii). Combining parts (iii) and (ii) gives part (i).

For part (vi), we must show that, given $b \in \mathrm{C}^0(\mathbb{T}_{t_0};\mathbb{R})$, the function $\xi_{\mathrm{p},b}$ in the statement of the theorem is an element of $\mathrm{Sol}(F_{t_0,b})$. We shall use the notation, for $f\colon \mathbb{T}_{t_0} \times \mathbb{T}_{t_0} \to \mathbb{R}$,

$$f(t,t-) = \lim_{\tau \uparrow t} f(t,\tau).$$

We then define

$$\xi_{\mathrm{p},b}(t) = \int_{t_0}^{t} G_{F,t_0}(t,\tau)b(\tau)\,\mathrm{d}\tau$$

and compute, for $\tau \in [t_0, t)$,

$$\frac{\mathrm{d}^l \xi_{\mathrm{p},b}}{\mathrm{d}t^l}(t) = \frac{\partial^{l-1} G_{F,t_0}}{\partial t^{l-1}}(t, t-)b(t) + \int_{t_0}^{t} \frac{\partial^l G_{F,t_0}}{\partial t^l}(t, \tau)\,\mathrm{d}\tau, \qquad l \in \{0, 1, \dots, k\}.$$

Since

$$\frac{\partial^{l-1} G_{F,t_0}}{\partial t^{l-1}}(t, t-) = \frac{\partial^{l-1} G_{F,t_0}}{\partial t^{l-1}}(t+, t) = 0, \qquad l \in \{0, 1, \dots, k-1\}$$

by parts (i) and (iii), we have

$$\frac{\mathrm{d}^l \xi_{\mathrm{p},b}}{\mathrm{d}t^l}(t) = \int_{t_0}^{t} \frac{\partial^l G_{F,t_0}}{\partial t^l}(t, \tau)\,\mathrm{d}\tau, \qquad l \in \{0, 1, \dots, k-1\}. \tag{2.18}$$

Also by parts (i) and (iii), we have

$$\frac{\mathrm{d}^k \xi_{\mathrm{p},b}}{\mathrm{d}t^k}(t) = b(t) + \int_{t_0}^{t} \frac{\partial^k G_{F,t_0}}{\partial t^k}(t, \tau)\,\mathrm{d}\tau. \tag{2.19}$$

Combining (2.18) and (2.19), and using part (v), we have, for $t \in \mathbb{T}_{t_0}$,

$$\frac{\partial^k \xi_{\mathrm{p},b}}{\partial t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1} \xi_{\mathrm{p},b}}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_{\mathrm{p},b}}{\mathrm{d}t}(t) + a_0(t)\xi_{\mathrm{p},b}(t) = b(t),$$

giving (vi).

The final uniqueness assertion of the theorem is obtained from the following observations:

1. for $t \leq \tau$, $t \mapsto G_{F,t_0}(t, \tau)$ is the unique element of $\mathrm{Sol}(F_{\mathrm{h}})$ with initial conditions

$$\frac{\partial^l G_{F,t_0}}{\partial t^l}(t_0, \tau) = 0, \qquad l \in \{0, 1, \dots, k-1\};$$

2. for $t > \tau$, $t \mapsto G_{F,t_0}(t, \tau)$ is the unique element of $\mathrm{Sol}(F_{\mathrm{h}})$ with initial conditions

$$\frac{\partial^l G_{F,t_0}}{\partial t^l}(\tau, \tau) = 0, \qquad l \in \{0, 1, \dots, k-2\},$$
$$\frac{\partial^{k-1} G_{F,t_0}}{\partial t^{k-1}}(\tau, \tau) = 1.$$

These, combined with Proposition 2.3.1, give the theorem. ■

Of course, we can give a name to the function G_{F,t_0} from the preceding theorem.

2.3.8 Definition (Green's function) Consider the linear homogeneous ordinary differential equation F with right-hand side equation (2.11) and suppose that the functions $a_0, a_1, \dots, a_{k-1}, b\colon \mathbb{T} \to \mathbb{R}$ are continuous. Let $t_0 \in \mathbb{T}$ and denote $\mathbb{T}_{t_0} = \mathbb{T} \cap [t_0, \infty)$. The function G_{F,t_0} of Theorem 2.3.7 is the ***Green's function*** for F at time t_0. •

There are a few observations one can make about the Green's function.

2.3.9 Remarks (Attributes of the Green's function)

1. As we observed in Remark 2.3.4, the mapping

$$\begin{aligned} L_F \colon \mathsf{C}^k(\mathbb{T}_{t_0}; \mathbb{R}) &\to \mathsf{C}^0(\mathbb{T}_{t_0}; \mathbb{R}) \\ \xi &\mapsto F_{\mathrm{h}}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t)\right) \end{aligned}$$

is surjective, and so, for any $b \in \mathsf{C}^0(\mathbb{T}_{t_0}; \mathbb{R})$, there exists one (indeed, many by Theorem 2.3.3), solution of the differential equation with solutions

$$F_{\mathrm{h}}\left(t, \xi(t), \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^k\xi}{\mathrm{d}t^k}(t)\right) = b(t).$$

One can think of the mapping

$$b \mapsto \left(t \mapsto \int_{t_0}^{t} G_{F,t_0}(t, \tau) b(\tau)\, \mathrm{d}\tau\right) \tag{2.20}$$

as prescribing a right-inverse of L_F. Of course, the prescription of a particular right-inverse amounts to a prescription for choosing initial conditions, since initial conditions are what distinguish elements of $\mathrm{Sol}(F)$. We refer the reader to Exercise 2.3.2 for just what initial condition choice is being made by the assignment (2.20).

2. There is also a physical interpretation of the mapping $t \mapsto G_{F,t_0}(t, \tau)$. The initial conditions are zero at t_0, so the solution is at rest, until something happens at $t = \tau$. At $t = \tau$, we imagine the system being given an "impulse" i.e., a short duration, large magnitude input. If the area under the graph of this impulse is 1, this will give a jolt to the kth derivative of G_{F,t_0} at $t = \tau$. This discontinuity when integrated, will give an input of 1 to the $(k-1)$st derivative, resulting in the initial conditions of part (iii) of Theorem 2.3.7.

   This nonsense can be made precise using the theory of distributions, and using the so-called "delta-function" as the "b." However, a detailed discussion of this, in this text, will take us too far afield.

   Nonetheless, it does motivate calling G_{F,t_0} the ***impulse response*** for F_{h}. In system theory, this impulse response plays an important rôle. •

Let us give the simplest possible example to illustrate the use of the Green's function.

2.3.10 Example (Green's function for first-order scalar linear ordinary differential equation) We consider the first order equation F with right-hand side

$$\widehat{F}(t, x) = -a(t)x.$$

Let us take $\mathbb{T}$ to be the time-domain for the equation, and let $t_0 \in \mathbb{T}$. One then takes $\mathbb{T}_{t_0}$, just as in the statement of Theorem 2.3.7 to be the set of points in $\mathbb{T}$ larger than t_0. The way one determines the Green's function is by first taking $\tau \in \mathbb{T}_{t_0}$ and then solving the initial value problem

$$\dot{\xi}(t) + a(t)\xi(t) = 0, \quad \xi(\tau) = 1,$$

just as prescribed in part (iii) of Theorem 2.3.7. However, in Example 2.2.5 we obtained the solution to this initial value problem as

$$\xi(t) = \mathrm{e}^{-\int_\tau^t a(s)\,\mathrm{d}s}.$$

Then the Green's function is given by

$$G_{F,t_0}(t,\tau) = \begin{cases} 0, & t \le \tau, \\ \mathrm{e}^{-\int_\tau^t a(s)\,\mathrm{d}s}, & t > \tau. \end{cases}$$

Therefore, given $b \in \mathsf{C}^0(\mathbb{T}_{t_0};\mathbb{R})$, a particular solution to the ordinary differential equation $F_{t_0,b}$ with right-hand side

$$\widehat{F}(t,x) = -a(t)x + b(t)$$

is given by

$$\xi_{\mathrm{p},b}(t) = \int_{t_0}^t \mathrm{e}^{-\int_\tau^t a(s)\,\mathrm{d}s} b(\tau)\,\mathrm{d}\tau.$$

Note that this, in general, is a *different* particular solution than that obtained in Example 2.3.6 using the Wronskian method of Proposition 2.3.5.

We plot the graph of G_{F,t_0} in the case of $\mathbb{T} = [0,\infty)$, $t_0 = 0$, and $a(t) = 1$ in Figure 2.3. •

2.3.11 Remark (Green's function for constant coefficient equations and convolution) Suppose that F is a kth-order scalar linear inhomogeneous ordinary differential equation with constant coefficients, and take $\mathbb{T} = [0,\infty)$ and $t_0 = 0$. As in the statement of Theorem 2.3.7, for each $\tau \in \mathbb{T}$, $t \mapsto G_{F,0}(t,\tau)$ is a solution for F satisfying the initial conditions

$$\begin{aligned} \frac{\partial^j G_{F,0}}{\partial t^j}(\tau) &= 0, \qquad j \in \{0,1,\dots,k-2\}, \\ \frac{\partial^{k-1} G_{F,0}}{\partial t^{k-1}}(\tau) &= 1. \end{aligned}$$

Since F has constant coefficients, it is autonomous, and so by Exercise 1.3.19 there exists $H_F\colon \mathbb{T} \to \mathbb{R}$ such that $G_{F,0}(t,\tau) = H_F(t-\tau)$. Then, if we add an inhomogeneous term b to F, the particular solution of Theorem 2.3.7(vi) is

$$\xi_{\mathrm{p},b}(t) = \int_0^t H_F(t-\tau)b(\tau)\,\mathrm{d}\tau.$$

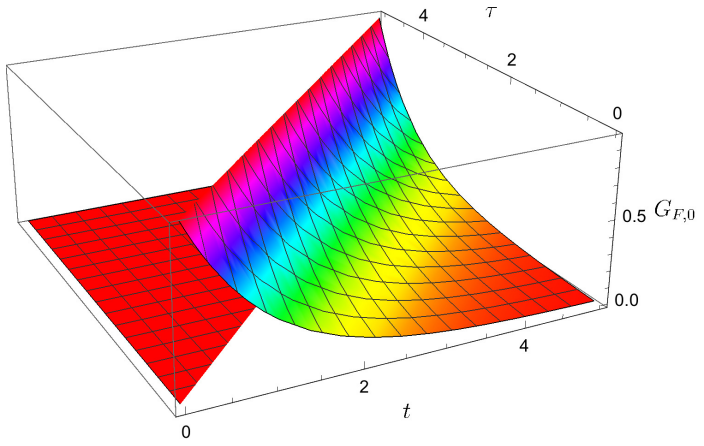


Figure 2.3 The Green's function for a scalar linear ordinary differential equation with constant coefficients

Integrals of the type

$$\int f(t-\tau)g(\tau)\,\mathrm{d}\tau$$

are known as ***convolution integrals***. These arise in system theory, Fourier theory, and approximation theory, for example. We shall consider convolution in the context of transform theory in Section 5.1. •

2.3.2 Equations with constant coefficients

We now specialise the general discussion from the preceding section to equations with constant coefficients. Thus we are looking at scalar linear inhomogeneous ordinary differential equations with right-hand sides given by

$$\widehat{F}(t,x,x^{(1)},\dots,x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1x^{(1)} - a_0x + b(t) \tag{2.21}$$

for $a_0, a_1, \dots, a_{k-1} \in \mathbb{R}$ and $b\colon \mathbb{T} \to \mathbb{R}$. Thus a solution $t \mapsto \xi(t)$ satisfies the equation

$$\frac{\mathrm{d}^k\xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0\xi(t) = b(t). \tag{2.22}$$

These equations are, of course, a special case of the equations considered in Section 2.3.1, and so all statements made about the general case of time-varying coefficients hold in the special case of constant coefficients. In particular, Propositions 2.3.1 and 2.3.2, and Theorem 2.3.3 hold for equations of the form (2.22). However, for these constant coefficient equations, it is possible to say some things a little more explicitly, and this is what we undertake to do.

2.3.2.1 The "method of undetermined coefficients" We present in this section a so-called method for solving scalar linear inhomogeneous ordinary differential equations with constant coefficients. With this method, one guesses a form of particular solution based on the form of the function b, and this does algebra to determine the precise solution. The advantages to this method are

1. it does not require first finding a fundamental set of solutions, as in Proposition 2.3.5,
2. it is in principle possible for a brainless monkey to apply the method, and
3. it is an excellent source of mindless computations that students can be forced to do for marks in homework and on exams.

The disadvantages of the method are

1. it only works for *very* specific functions b, and so does not work most of the time,
2. even when it does work, it is tedious and likely to produce errors when used in the hands of most humans,
3. it is 2016, for crying out loud, and there are computer packages that do this sort of thing in their sleep!

What we shall do is (1) describe when the method applies, (2) describe how one uses the method, and (3) reiterate the silliness of the method at the end of the discussion.

First let us indicate the sorts of "b's" we allow.

2.3.12 Definition (Pretty uninteresting function) Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval. A function $f : \mathbb{T} \to \mathbb{R}$ is ***pretty uninteresting*** if it has one of the following three forms:

(i) $f(t) = t^m e^{rt}$ for $m \in \mathbb{Z}_{\geq 0}$ and $r \in \mathbb{R}$;

(ii) $f(t) = t^m e^{\sigma t} \cos(\omega t)$ for $m \in \mathbb{Z}_{\geq 0}$, $\sigma \in \mathbb{R}$, and $\omega \in \mathbb{R}_{>0}$;

(iii) $f(t) = t^m e^{\sigma t} \sin(\omega t)$ for $m \in \mathbb{Z}_{\geq 0}$, $\sigma \in \mathbb{R}$, and $\omega \in \mathbb{R}_{>0}$.

The nonnegative integer m in the above forms is the ***order*** of f and is denoted by $o(f)$. If $f : \mathbb{T} \to \mathbb{R}$ has the form

$$f(t) = c_1 f_1(t) + \cdots + c_r f_r(t)$$

where $c_1, \ldots, c_r \in \mathbb{R}$ and each of $f_1, \ldots, f_r$ is pretty uninteresting, then f is ***also pretty uninteresting***. •

Here are some examples of useful pretty uninteresting functions.

2.3.13 Examples (Examples of interesting pretty uninteresting functions)

1. Consider the function $1 : [0, \infty) \to \mathbb{R}$ defined by $1(t) = 1$ for all $t \in [0, \infty)$. This is a "step function" and is pretty uninteresting. Often it is taken to be defined on all of $\mathbb{R}$, and to be zero for negative times. The idea is that it gives an input to a differential equation that "switches on" at $t = 0$. Among the many particular

solutions for a differential equation with $b = 1$, there is one that is known as the "step response," and it is determined by a specific choice of initial condition. Students going on to take a course in system theory will learn about this.

2. Next consider the function $H_\omega\colon [0,\infty) \to \mathbb{R}$ defined by $H_\omega(t) = \sin(\omega t)$ for $\omega \in \mathbb{R}_{>0}$. This is an example of an "harmonic" function, and specifically is a "sinusoid." In this case, one can think of prescribing a "b" of this form as "shaking" a differential equation. It can be interesting to know how the behaviour of the system will vary as we change ω. This gives rise in system theory to something called the "frequency response." •

We now state a few elementary properties of pretty uninteresting functions.

2.3.14 Lemma (Properties of pretty uninteresting functions) *Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, let $f, f_1, \dots, f_r\colon \mathbb{T} \to \mathbb{R}$ be pretty uninteresting functions, and consider a scalar linear homogeneous ordinary differential equation F with constant coefficients with right-hand side of the form (2.21). Define normalised scalar linear inhomogeneous ordinary differential equations F_j, $j \in \{1, \dots, r\}$, by*

$$F_j(t, x, x^{(1)}, \dots, x^{(k)}) = x^{k)} - \widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) - f_j(t).$$

Then the following statements hold:

(i) there exists unique normalised scalar linear homogeneous ordinary differential equation F_f of order $o(f)$ such that

$$F_f\left(t, f(t), \frac{df}{dt}(t), \dots, \frac{d^{o(f)}f}{dt^{o(f)}}\right) = 0, \qquad t \in \mathbb{T};$$

(ii) if $\xi_j \in \mathrm{Sol}(F_j)$, $j \in \{1, \dots, r\}$, and if

$$g = c_1 f_1 + \cdots + c_r f_r$$

is also pretty uninteresting, then, if $\xi = c_1\xi_1 + \cdots + c_r\xi_r$, then $\xi \in \mathrm{Sol}(F_g)$, where

$$F_j(t, x, x^{(1)}, \dots, x^{(k)}) = x^{k)} - \widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) - g(t).$$

Proof (i) An examination of Procedure 2.2.18 and the attendant Theorem 2.2.19 shows that F_f can be defined by defining their characteristic polynomials as follows

1. $f(t) = t^m e^{rt}$: take
$$P_{F_f} = (X - r)^{m+1};$$
2. $f(t) = t^m e^{\sigma t}\cos(\omega t)$ or $f(t) = t^m e^{\sigma t}\sin(\omega t)$: take
$$P_{F_f} = ((X - \sigma)^2 + \omega^2)^{m+1}.$$

(ii) This is a mere verification, once one understands the symbols involved. ■

The differential equation F_f in the first part of the lemma we call the ***annihilator*** of the pretty uninteresting function f. The following examples illustrate how one finds the annihilator in practice, based on the proof of the first part of the lemma.

2.3.15 Examples (Annihilator)

1. Consider the function $f(t) = 1$. This is the pretty uninteresting function $t \mapsto t^k e^{\sigma t} \cos(\omega t)$ with $k = 0$, $\sigma = 0$, and ω. This corresponds, from Procedure 2.2.18, to a root $r = 0$ of a polynomial with multiplicity 1. Thus $P_{F_f} = X$, and so
$$F(t, x, x^{(1)}) = X.$$
2. Now consider $f(t) = e^{-2t}$. This is the pretty uninteresting function $t \mapsto t^k e^{\sigma t} \cos(\omega t)$ with $k = 0$, $\sigma = -2$, and $\omega = 0$. This corresponds to a root $r = -2$ of a polynomial with multiplicity 1. Thus $P_{F_f} = X + 2$ and so
$$F_f(t, x, x^{(1)}) = x^{(1)} + 2x.$$
3. Next we take $f(t) = 2e^{3t} \sin(2t) + t^2$. This is an also pretty uninteresting function, being a linear combination of $f_1(t) = e^{3t} \sin(2t)$ and $f_2(t) = t^2$.
   Note that f_1 is the pretty uninteresting function $t \mapsto t^k e^{\sigma t} \sin(\omega t)$ with $k = 0$, $\sigma = 3$, and $\omega = 2$. This function is associated, via Procedure 2.2.18, with a root $\rho = 3 + 2\mathrm{i}$ of a polynomial with multiplicity 1. Of course, we must also have the root $\bar{\rho} = 3 - 2\mathrm{i}$.
   Note that f_2 is the pretty uninteresting function $t \mapsto t^k e^{\sigma t} \cos(\omega t)$ with $k = 2$, $\sigma = 0$, and $\omega = 0$. This is associated with a root $r = 0$ with multiplicity 3.
   Putting this all together,
$$P_{F_f} = (X - (3 + 2\mathrm{i}))(X - (3 - 2\mathrm{i}))X^3 = X^5 - 6X^4 + 13X^3.$$ •

The second part of the lemma points out, in short, the obvious fact that if "b" is also pretty uninteresting, then one can obtain a particular solution by obtaining a particular solution for each of its pretty uninteresting components, and then summing these with the same coefficients as in the also pretty uninteresting function. The point of this is that, to obtain a particular solution for an also pretty uninteresting "b," it suffices to know how to do this for a pretty uninteresting b. Thus we deliver the following construction.

2.3.16 Procedure (Method of undetermined coefficients) We let F be a normalised scalar linear inhomogeneous ordinary differential equation with constant coefficients with right-hand side
$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x + f(t),$$
where f is pretty uninteresting. Do the following.

1. Let F_f be the annihilator of f.
2. Let G_f be the normalised scalar linear homogeneous ordinary differential equation whose characteristic polynomial is $P_{G_f} = P_{F_f} P_{F_h}$.

3. Using Procedure 2.2.18, find
   (a) pretty uninteresting functions $\xi_1, \dots, \xi_k$ for which $\{\xi_1, \dots, \xi_k\}$ is a fundamental set of solutions for F_{h} and
   (b) pretty uninteresting functions $\eta_1, \dots, \eta_{o(f)+1}$ for which $\{\xi_1, \dots, \xi_k, \eta_1, \dots, \eta_{o(f)+1}\}$ is a fundamental set of solutions for G_f.
4. For (as yet) undetermined coefficients $c_1, \dots, c_{o(f)+1} \in \mathbb{R}$, denote
$$\xi_{\mathrm{p}} = c_1\eta_1 + \cdots + c_{o(f)+1}\eta_{o(f)+1}.$$
5. Determine $c_1, \dots, c_{o(f)+1}$ by demanding that ξ_{p} be a particular solution for F.

We shall show that this procedure makes sense and defines a particular solution for F. •

Let us verify that the preceding procedure gives what we want.

2.3.17 Proposition (Validity of the method of undetermined coefficients) *Let* F *be a normalised scalar linear inhomogeneous ordinary differential equation with constant coefficients with right-hand side*
$$\widehat{\mathrm{F}}(\mathrm{t}, \mathrm{x}, \mathrm{x}^{(1)}, \dots, \mathrm{x}^{(\mathrm{k}-1)}) = -\mathrm{a}_{\mathrm{k}-1}\mathrm{x}^{(\mathrm{k}-1)} - \cdots - \mathrm{a}_1\mathrm{x}^{(1)} - \mathrm{a}_0\mathrm{x} + \mathrm{f}(\mathrm{t}),$$
where f *is pretty uninteresting. Then all steps in Procedure 2.3.16 are unambiguously defined, and the result is a particular solution for* F.

Proof In the proof we shall assume that $f(t) = t^{o(f)}\mathrm{e}^{rt}$ for $r \in \mathbb{R}$. Entirely similar reasoning works for the other two sorts of pretty uninteresting functions.

From Procedure 2.2.18 we know that $P_{F_f} = (X - r)^{o(f)+1}$. Let us suppose that
$$P_{F_{\mathrm{h}}} = (X - r)^{m(r)}P,$$
where P does not have r as a root. Therefore,
$$P_{G_f} = (X - r)^{m(r)+o(f)+1}P.$$
Then, according to Procedure 2.2.18, among the pretty uninteresting solutions for F_{h} are
$$t \mapsto t^j\mathrm{e}^{rt}, \qquad j \in \{0, 1, \dots, m(r) - 1\}.$$
The rest of the pretty uninteresting solutions for F_{h} have nothing to do with the root "r" of the characteristic polynomial, and are not interesting to us here. Now the $o(f) + 1$ pretty uninteresting solutions for G_f that are added to those for F_{h} are
$$t^j\mathrm{e}^{rt}, \qquad j \in \{m(r), \dots, m(r) + o(f)\},$$
again according to Procedure 2.2.18. This demonstrates the viability of the first three steps of Procedure 2.2.18. We now need to show that one can solve for the coefficients $c_1, \dots, c_{o(f)+1}$ to obtain a particular solution for F. If
$$\xi_{\mathrm{p}}(t) = c_1t^{m(r)}\mathrm{e}^{rt} + \cdots + c_{o(f)+1}t^{m(r)+o(f)}$$

then Lemma 1 from the proof of Theorem 2.2.19 shows that

$$\left(\frac{\mathrm{d}^{m}(r)}{\mathrm{d}t^{m(r)}} - r\right)\xi_{\mathrm{p}}(t)$$

is an also pretty uninteresting function whose highest order (as a pretty uninteresting function) term is of order $o(f)$. By Corollary 2.2.17, and since the derivative of a pretty uninteresting function of order m is an also pretty uninteresting function of order m, we have that

$$F_{\mathrm{h}}\left(t, \xi_{\mathrm{p}}(t), \frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t), \ldots, \frac{\mathrm{d}^{k}\xi_{\mathrm{p}}}{\mathrm{d}t^{k}}(t)\right)$$

is an also pretty uninteresting function of order $o(f)$. Therefore, we can use the equality

$$F_{\mathrm{h}}\left(t, \xi_{\mathrm{p}}(t), \frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t), \ldots, \frac{\mathrm{d}^{k}\xi_{\mathrm{p}}}{\mathrm{d}t^{k}}(t)\right) = c_1 t^{m(r)}\mathrm{e}^{rt} + \cdots + c_{o(f)+1}t^{m(r)+o(f)}$$

to solve for the coefficients $c_1, \ldots, c_{o(f)+1}$, as asserted in Procedure 2.2.18. ■

While the preceding discussion does indeed provide a means of solving, in principle, scalar linear inhomogeneous ordinary differential equations with also pretty uninteresting "b's," it does tend to be a lot of work, cf. Example 2.3.18, and there are precisely zero equations that can be solved by this procedure that cannot far more easily be solved with a computer.

2.3.2.2 Some examples

We carry on with the three examples of Section 2.2.2.4. Thus we first give an "academic" example to illustrate Procedure 2.3.16. Then we consider a first- and second-order system with specific "b's," in order to discuss some features of the solutions in these cases.

2.3.18 Example ("Academic" example) We continue the example of Example 2.2.20, now adding an inhomogeneous term. Specifically, we consider the 4th-order scalar linear homogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}, x^{(2)}, x^{(3)}) = -5x + 8x^{(1)} - 2x^{(2)} + t\mathrm{e}^{t} + 2\cos(2t).$$

Thus solutions $t \mapsto \xi(t)$ to this equation satisfy

$$\frac{\mathrm{d}^{4}\xi}{\mathrm{d}t^{4}}(t) + 2\frac{\mathrm{d}^{2}\xi}{\mathrm{d}t^{2}}(t) - 8\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + 5\xi(t) = t\mathrm{e}^{t} + 2cos(2t).$$

The right-hand side of this equation has the form $b(t) = f_1(t) + 2f_2(t)$ for the two pretty uninteresting functions

$$f_1(t) = t\mathrm{e}^{t}, \quad f_2(t) = \cos(2t).$$

We find two particular solutions $\xi_{\mathrm{p},1}$ and $\xi_{\mathrm{p},2}$, satisfying

$$\frac{\mathrm{d}^4\xi_{\mathrm{p},1}}{\mathrm{d}t^4}(t) + 2\frac{\mathrm{d}^2\xi_{\mathrm{p},1}}{\mathrm{d}t^2}(t) - 8\frac{\mathrm{d}\xi_{\mathrm{p},1}}{\mathrm{d}t}(t) + 5\xi_{\mathrm{p},1}(t) = t\mathrm{e}^t$$

and

$$\frac{\mathrm{d}^4\xi_{\mathrm{p},2}}{\mathrm{d}t^4}(t) + 2\frac{\mathrm{d}^2\xi_{\mathrm{p},2}}{\mathrm{d}t^2}(t) - 8\frac{\mathrm{d}\xi_{\mathrm{p},2}}{\mathrm{d}t}(t) + 5\xi_{\mathrm{p},2}(t) = \cos(2t),$$

and then, by Lemma 2.3.14(ii),

$$\xi_{\mathrm{p}} = \xi_{\mathrm{p},1} + 2\xi_{\mathrm{p},2}$$

is a particular solution.

Let us find $\xi_{\mathrm{p},1}$ corresponding to $f_1(t) = t\mathrm{e}^t$. The annihilator F_{f_1} of f_1 has characteristic polynomial $P_{F_{f_1}} = (X-1)^2$. We have

$$P_{F_{f_1}}P_{F_{\mathrm{h}}} = (X-1)^2(X-1)^2(X^2+2X+5) = (X-1)^4(X^2+2X+5)$$

as the characteristic polynomial of $F_{f_1} \circ F_{\mathrm{h}}$. According to Procedure 2.2.18, a fundamental set of solutions, each of which is a pretty uninteresting function, is given by

$$\mathrm{e}^{-t}\cos(2t),\ \mathrm{e}^{-t}\sin(2t),\ \mathrm{e}^t,\ t\mathrm{e}^t,\ t^2\mathrm{e}^t,\ t^3\mathrm{e}^t.$$

The first four of these are solutions for F_{h}. So we form our candidate particular solution from the last two functions:

$$\xi_{\mathrm{p},1}(t) = c_1t^2\mathrm{e}^t + c_2t^3\mathrm{e}^t.$$

To determine c_1 and c_2, we compute

$$\left(\frac{\mathrm{d}^4}{\mathrm{d}t^4} + 2\frac{\mathrm{d}^2}{\mathrm{d}t^2} - 8\frac{\mathrm{d}}{\mathrm{d}t} + 5\right)\xi_{\mathrm{p},1}(t) = (16c_1 + 24c_2)\mathrm{e}^t + 48c_2t\mathrm{e}^t.$$

Thus we have

$$16c_1 + 24c_2 = 0,\ 48c_2 = 1 \quad\Longrightarrow\quad c_1 = -\frac{1}{32},\ c_2 = \frac{1}{48}.$$

Thus

$$\xi_{\mathrm{p},1}(t) = -\frac{t^2\mathrm{e}^t}{32} + \frac{t^3\mathrm{e}^t}{48}.$$

Now we find $\xi_{\mathrm{p},2}$ corresponding to $f_2 = \cos(2t)$. Here the annihilator F_{f_2} of f_2 has characteristic polynomial $P_{F_{f_2}} = X^2 + 4$. We have

$$P_{F_{f_2}}P_{F_{\mathrm{h}}} = (X^2+4)(X^4+2X^2-8X+5).$$

Thus the fundamental set of solutions for $F_{f_2} \circ F_{\mathrm{h}}$ is given by

$$\mathrm{e}^{-t}\cos(2t),\ \mathrm{e}^{-t}\sin(2t),\ \mathrm{e}^{t},\ t\mathrm{e}^{t},\ \cos(2t),\ \sin(2t).$$

Since the first four of these are solutions for F_{h}, we have

$$\xi_{\mathrm{p},2}(t) = c_1\cos(2t) + c_2\sin(2t).$$

To determine c_1 and c_2 we compute

$$\left(\frac{\mathrm{d}^4}{\mathrm{d}t^4} + 2\frac{\mathrm{d}^2}{\mathrm{d}t^2} - 8\frac{\mathrm{d}}{\mathrm{d}t} + 5\right)\xi_{\mathrm{p},2}(t) = (13c_1 - 16c_2)\cos(2t) + (16c_1 + 13c_2)sin(2t).$$

Therefore,

$$13c_1 - 16c_2 = 1,\ 16c_1 + 13c_2 = 0 \quad\Longrightarrow\quad c_1 = \frac{13}{425},\ c_2 = \frac{16}{425}.$$

Thus

$$\xi_{\mathrm{p},2} = \frac{13}{425}\cos(2t) + \frac{16}{425}\sin(2t).$$

Finally, we have the particular

$$\xi_{\mathrm{p}}(t) = -\frac{t^2\mathrm{e}^t}{32} + \frac{t^3\mathrm{e}^t}{48} + \frac{13}{425}\cos(2t) + \frac{16}{425}\sin(2t).$$

Thus, as per Theorem 2.3.3, *any* solution ξ of F can be written we

$$\xi(t) = c_1\mathrm{e}^t + c_2t\mathrm{e}^t + c_3\mathrm{e}^{-t}\cos(2t) + c_4\mathrm{e}^{-t}\sin(2t) - \frac{t^2\mathrm{e}^t}{32} + \frac{t^3\mathrm{e}^t}{48} + \frac{26}{425}\cos(2t) + \frac{32}{425}\sin(2t).$$

To determine the constants c_1, c_2, c_3, c_4, we use the initial conditions

$$\xi(0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(0) = x + 0^{(1)},\ \frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(0) = x_0^{(2)},\ \frac{\mathrm{d}^3\xi}{\mathrm{d}t}(0) = x_0^{(3)}.$$

These do *not* have the same solution as in Example 2.2.20 because of the presence of the particular solution. Some unpleasant computation gives the equations

$$\begin{aligned}
c_1 + c_3 &= -\frac{26}{425} + x_0,\\
c_1 + c_2 - c_3 + 2c_4 &= -\frac{64}{425} + x_0^{(1)},\\
c_1 + 2c_2 - 3c_3 - 4c_4 &= \frac{2089}{6800} + x_0^{(2)},\\
c_1 + 3c_2 + 11c_3 - 2c_4 &= \frac{4521}{6800} + x_0^{(3)}
\end{aligned}$$

that have to be solved. Here's what you get:

$$\begin{aligned}
c_1 &= \frac{15}{16}x_0 + \frac{1}{16}x_0^{(1)} + \frac{1}{16}x_0^{(2)} - \frac{1}{16}x_0^{(3)} - \frac{303}{3400},\\
c_2 &= -\frac{5}{8}x_0 + \frac{3}{8}x_0^{(1)} + \frac{1}{8}x_0^{(2)} + \frac{1}{8}x_0^{(3)} + \frac{2809}{27200},\\
c_3 &= \frac{1}{16}x_0 - \frac{1}{16}x_0^{(1)} - \frac{1}{16}x_0^{(2)} + \frac{1}{16}x_0^{(3)} + \frac{19}{680},\\
c_4 &= -\frac{1}{8}x_0 + \frac{1}{4}x_0^{(1)} - \frac{1}{8}x_0^{(2)} - \frac{3721}{54400}.
\end{aligned}$$

Alternatively, one can use MATHEMATICA® as illustrated in Section 2.4.2. You will then get back a reliable answer after about 15 seconds of typing. You can decide which method you think is best in practice. •

The next two examples give an illustration of where pretty uninteresting functions are interesting in application.

2.3.19 Example (First-order system with step input) The differential equation we consider here is an inhomogeneous version of the equation considered in Example 2.2.21. We take the first-order scalar linear inhomogeneous ordinary differential equation F with constant coefficients and with right-hand side

$$\widehat{F}(t,x) = -\frac{x}{\tau} + 1.$$

Thus solutions $t \mapsto \xi(t)$ to this differential equation satisfy

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + \frac{1}{\tau}\xi(t) = 1.$$

We have already determined that a solution to the homogeneous equation will have the form $\xi(t) = c\mathrm{e}^{-t/\tau}$, taking the convention that $\frac{1}{\tau} = 0$ when "$\tau = \infty$."

So next we find a particular solution. The annihilator F_f of the pretty uninteresting function $f(t) = 1$ has characteristic polynomial $P_{F_f} = X$. The characteristic polynomial for F_{h} is $P_{F_{\mathrm{h}}} = X + \frac{1}{\tau}$. Thus we must list the fundamental solutions for G_f, where

$$P_{G_f} = X(X - \tfrac{1}{\tau}).$$

There are two cases.

First, when $\tau \neq \infty$, the fundamental solutions are $t \mapsto \mathrm{e}^{-t/\tau}$ and $t \mapsto 1$, using Procedure 2.2.18. The first of these is a solution for the homogeneous solution, so we take a particular solution to be a multiple of the second: $\xi_{\mathrm{p}}(t) = c$. To find c we substitute into the differential equation:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \frac{1}{\tau}\right)\xi = \frac{c}{\tau}.$$

To be a particular solution, we must have $\frac{c}{\tau} = 1$ and so $c = \tau$. Thus $\xi_{\mathrm{p}}(t) = \tau$.

The other case arises when $\tau = \infty$, and in this case the fundamental solutions for G_f are $t \mapsto 1$ and $t \mapsto t$, again using Procedure 2.2.18. In this case, the first of these functions is a solution for the homogeneous system, and so a multiple of the second will be a particular solution, i.e., $\xi_{\mathrm{p}}(t) = ct$. To determine c we require that ξ_{p} be a particular solution:

$$\frac{\mathrm{d}}{\mathrm{d}t}\xi_{\mathrm{p}}(t) = c,$$

from which we deduce that $c = 1$. Thus $\xi_{\mathrm{p}}(t) = t$.

In summary, a particular solution is

$$\xi_{\mathrm{p}}(t) = \begin{cases} \tau, & \tau \neq \infty, \\ t, & t = \infty. \end{cases}$$

Therefore, *any* solution has the form

$$\xi(t) = c\mathrm{e}^{-ttau} + \xi_{\mathrm{p}}(t).$$

In case $\tau \neq \infty$, one normally takes the initial condition $\xi(0) = 0$ to get $c = -\tau$ and so

$$\xi(t) = \tau(1 - \mathrm{e}^{-t/\tau}).$$

To allow a fruitful comparison of the effects of changing τ, let us normalise this solution by multiplying by $\frac{1}{\tau}$ to get the ***step response***

$$1_F(t) = 1 - \mathrm{e}^{-t/\tau}.$$

In Figure 2.4 we graph this step response for varying values of $\tau \in \mathbb{R}_{>0}$. We note

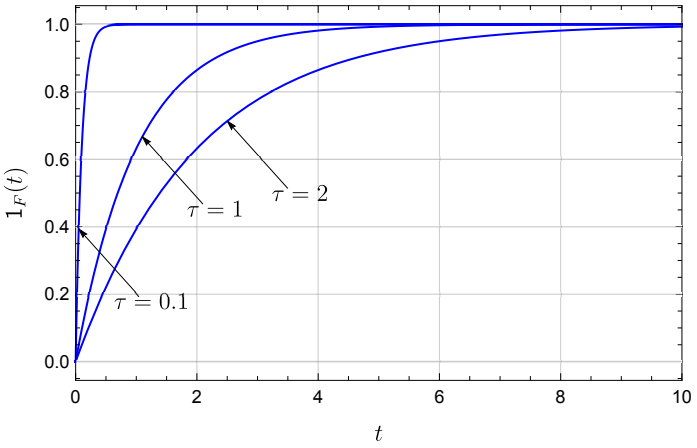


Figure 2.4 The step response of a first-order system

that as τ gets smaller, the step response rises more quickly, i.e., responds faster. •

2.3.20 Example (Second-order system with sinusoidal input) Next we consider the second-order differential equation of Example 2.2.22, but with a sinusoidal inhomogeneous term. Thus we take the second-order scalar linear inhomogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}) = -\omega_0^2 x - 2\zeta\omega_0 x^{(1)} + A\sin(\omega t)$$

for $A, \omega \in \mathbb{R}_{>0}$. Solutions $t \mapsto \xi(t)$ then satisfy

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2}(t) + 2\zeta\omega_0\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + \omega_0^2\xi(t) = A\sin(\omega t).$$

In Example 2.2.22 we carefully and thoroughly investigated the nature of the solutions for the homogeneous system. There we saw, for example, that as long as $\zeta > 0$, solutions to the homogeneous equation decay to zero as $t \to \infty$. For $\zeta = 0$, solutions were periodic. Here we will thus focus on $\zeta \in \mathbb{R}_{\geq 0}$ and on the nature of the particular solution. When $\zeta \in \mathbb{R}_{>0}$ this means that we are looking at the "steady-state" behaviour of the system, i.e., what we see after a long time. When $\zeta = 0$, we do not have this steady-state interpretation, but nonetheless we will interpret these solutions in light of our understanding of what happens when $\zeta \in \mathbb{R}_{>0}$.

The annihilator F_f for the pretty uninteresting function $f(t) = A\sin(\omega t)$ has characteristic polynomial $P_{F_f} = X^2+\omega^2$. We have two cases to consider for particular solutions.

The first case is when $\zeta \in \mathbb{R}_{>0}$ or when $\zeta = 0$ and $\omega \neq \omega_0$. Here the characteristic polynomial for G_f in Procedure 2.3.16 is

$$P_{G_f} = (X^2 + \omega^2)(X^2 + 2\zeta\omega_0 X + \omega_0^2)$$

The fundamental solutions for G_f associated to this polynomial, according to Procedure 2.2.18, are

$$\xi_1(t),\ \xi_2(t),\ \cos(\omega t),\ \sin(\omega t),$$

where ξ_1 and ξ_2 are homogeneous solutions as determined in Example 2.2.22. Thus a particular solution will be of the form

$$\xi_{\mathrm{p}}(t) = c_1\cos(\omega t) + c_2\sin(\omega t).$$

To determine c_1 and c_2 we require that ξ_{p} be a particular solution. Thus we compute

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\zeta\omega_0\frac{\mathrm{d}}{\mathrm{d}t} + \omega_0^2\right)\xi_{\mathrm{p}}(t)$$
$$= (c_1(\omega_0^2 - \omega^2) + c_2 2\zeta\omega_0\omega)\cos(\omega t) + (-c_2 2\zeta\omega_0\omega + c_2(\omega_0^2 - \omega^2))\sin(\omega t).$$

We must, therefore, have

$$\begin{aligned} c_1(\omega_0^2 - \omega^2) + c_2 2\zeta\omega_0\omega = 0, \\ -c_2 2\zeta\omega_0\omega + c_2(\omega_0^2 - \omega^2) = A, \end{aligned} \implies \begin{aligned} c_1 &= \frac{2\zeta\omega_0\omega A}{\omega^4 + \omega_0^4 - 2\omega_0^2\omega^2(1-2\zeta^2)}, \\ c_2 &= \frac{(\omega_0^2 - \omega^2)A}{\omega^4 + \omega_0^4 - 2\omega_0^2\omega^2(1-2\zeta^2)} \end{aligned}$$

Thus a particular solution is

$$\xi_{\mathrm{p}}(t) = \frac{2\zeta\omega_0\omega A}{\omega^4 + \omega_0^4 - 2\omega_0^2\omega^2(1-2\zeta^2)}\cos(\omega t) + \frac{(\omega_0^2 - \omega^2)A}{\omega^4 + \omega_0^4 - 2\omega_0^2\omega^2(1-2\zeta^2)}\sin(\omega t).$$

The other case is when $\zeta = 0$ and $\omega = \omega_0$. Here the characteristic polynomial for G_f in Procedure 2.3.16 is

$$P_{G_f} = (X^2 + \omega^2)^2$$

The fundamental solutions for G_f associated to this polynomial, according to Procedure 2.2.18, are

$$\xi_1(t),\ \xi_2(t),\ t\cos(\omega t),\ t\sin(\omega t),$$

where ξ_1 and ξ_2 are homogeneous solutions as determined in Example 2.2.22. Therefore, a particular solution will have the form

$$\xi_{\mathrm{p}}(t) = c_1 t\cos(\omega_0 t) + c_2 t\sin(\omega_0 t).$$

To determine c_1 and c_2 we ask that this be a particular solution. Thus we compute

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \omega_0^2\right)\xi_{\mathrm{p}}(t) = 2c_2\omega_0\cos(\omega_0 t) - 2c_1\omega_0\sin(\omega_0 t).$$

Therefore, we must have

$$2c_2\omega_0 = 0,\ 2c_1\omega_0 = A, \implies c_1 = \frac{A}{2\omega_0},\ c_2 = 0,$$

and so the particular solution we obtain is

$$\xi_{\mathrm{p}}(t) = \frac{At}{2\omega_0}\cos(\omega_0 t).$$

Therefore, in summary, a particular solution is

$$\xi_{\mathrm{p}}(t) = \begin{cases} \frac{At}{2\omega_0}\cos(\omega_0 t), & \zeta = 0,\ \omega = \omega_0, \\ \frac{2\zeta\omega_0\omega A}{\omega^4+\omega_0^4-2\omega_0^2\omega^2(1-2\zeta^2)}\cos(\omega t) + \frac{(\omega_0^2-\omega^2)A}{\omega^4+\omega_0^4-2\omega_0^2\omega^2(1-2\zeta^2)}\sin(\omega t), & \text{otherwise.} \end{cases}$$

Any solution will be a sum of this solution, plus some solution to the homogeneous equation as determined in Example 2.2.22.

In Figure 2.5 we graph particular solutions for various ζ's and ω_0's, keeping A and ω_0 fixed. We make the following observations.

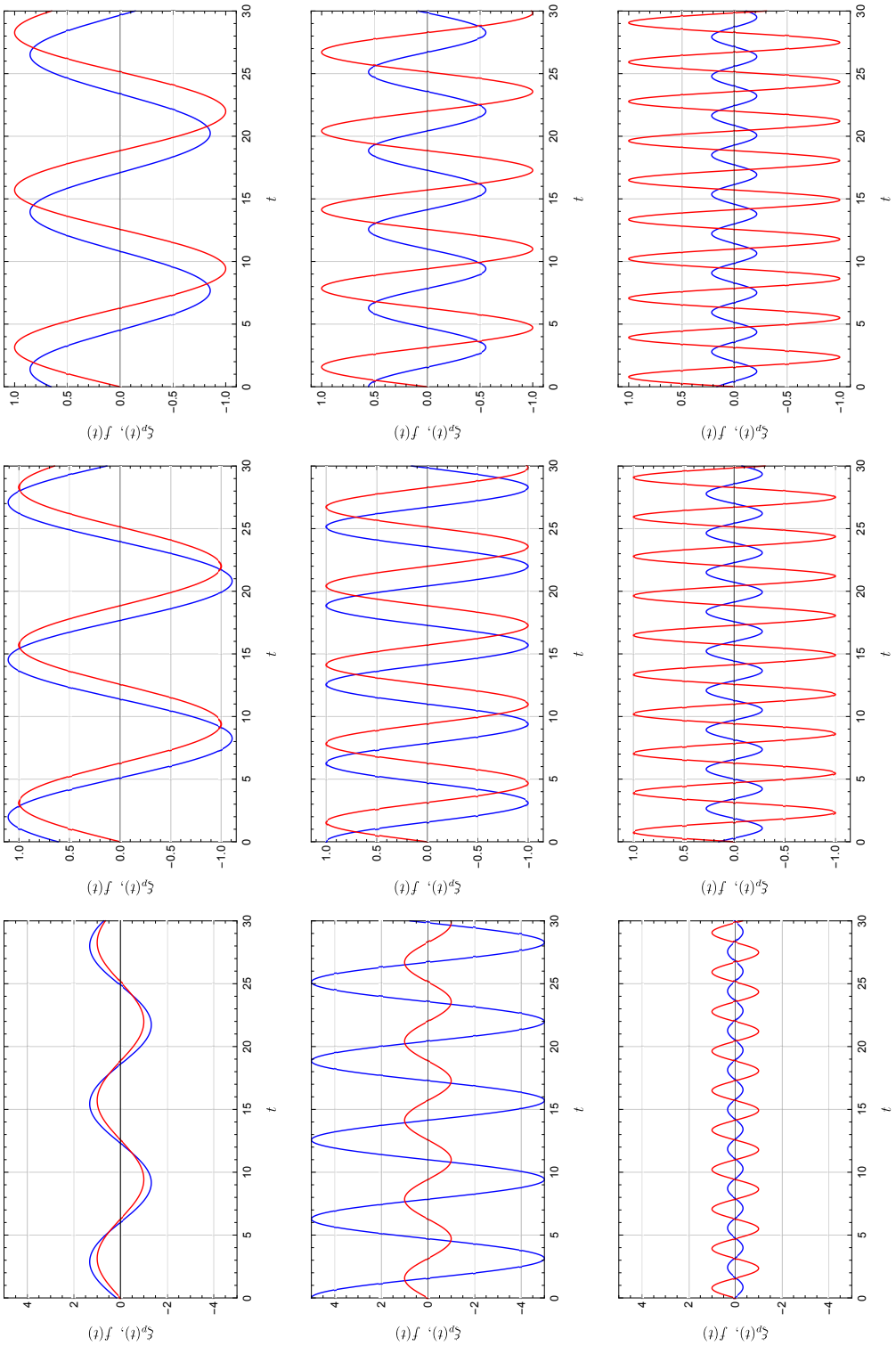


Figure 2.5 Response (in blue) of a second-order system with $\omega_0 = 1$ to a sinusoidal input with $A = 1$ (in red) for varying ζ and ω (left: $\zeta = 0.1$, $\omega \in \{0.5, 1, 2\}$; middle: $\zeta = 0.5$, $\omega \in \{0.5, 2, 1\}$; right: $\zeta = 0.9$, $\omega \in \{0.5, 1, 2\}$

1. For small values of ω (compared to ω_0), the response $\xi_p(t)$ is quite closely aligned in amplitude and phase with the input $f(t)$.
2. For small values of ζ, i.e., small damping, as $\omega \to \omega_0$ the response gets large in amplitude and the phase shift is about $\frac{1}{4}$ of a period.
3. For not so small values of ζ, the amplitude as $\omega \to \omega_0$ does not grow so much, but the phase still shifts by about $\frac{1}{4}$ of a period.
4. As the frequency ω gets large (compared to ω_0), the amplitude decays to zero, and the response and input are out of phase, i.e., the phase shift is about $\frac{1}{2}$ of a period.

One can see in the previous description the genesis of what happens when $\zeta = 0$, i.e., the response amplitude grows over time. This phenomenon is called "resonance," meaning that the excitation from the inhomogeneous term has the same frequency as the natural frequency of the system.

The matters touched above in the preceding discussion are captured in system theory by the notion of "frequency response." •

Exercises

2.3.1 Consider the ordinary differential equation F with right-hand side given by (2.11).

(a) Convert this to a first-order equation with k states, following Exercise 1.3.23.

(b) Show that, if the functions $a_0, a_1, \dots, a_k$ are continuous, then the resulting first-order equation satisfies the conditions of Theorem 1.4.8 for existence of a unique solution $t \mapsto \xi(t)$ satisfying the initial conditions

$$\xi(t_0) = x_0, \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) = x_0^{(1)}, \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t)(t_0) = x_0^{(k-1)}$$

at time $t_0 \in \mathbb{T}$.

2.3.2 Consider the ordinary differential equation F with right-hand side given by (2.11). Answer the following questions.

(a) Show that the particular particular solution

$$\xi_{p,b}(t) = \int_{t_0}^{t} G_{F,t_0}(t,\tau) b(\tau)\, \mathrm{d}\tau$$

satisfies the initial value problem

$$\frac{\mathrm{d}^k\xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t),$$
$$\xi(t_0) = 0, \ \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) = 0, \dots, \ \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t_0) = 0.$$

(b) Show that the solution to the initial value problem

$$\frac{\mathrm{d}^k\xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0(t)\xi(t) = b(t),$$
$$\xi(t_0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) = x_0^{(1)}, \ldots,\ \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t_0) = x_0^{(k-1)}$$

is given by $\xi(t) = \xi_{\mathrm{h}} + \xi_{\mathrm{p},b}$, where ξ_{h} is the solution to the homogeneous initial value problem

$$\frac{\mathrm{d}^k\xi_{\mathrm{h}}(t)}{\mathrm{d}t^k}(t) + a_{k-1}(t)\frac{\mathrm{d}^{k-1}\xi_{\mathrm{h}}}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1(t)\frac{\mathrm{d}\xi_{\mathrm{h}}}{\mathrm{d}t}(t) + a_0(t)\xi_{\mathrm{h}}(t) = 0,$$
$$\xi_{\mathrm{h}}(t_0) = x_0,\ \frac{\mathrm{d}\xi_{\mathrm{h}}}{\mathrm{d}t}(t_0) = x_0^{(1)}, \ldots,\ \frac{\mathrm{d}^{k-1}\xi_{\mathrm{h}}}{\mathrm{d}t^{k-1}}(t_0) = x_0^{(k-1)}.$$

2.3.3 Find the annihilator for each of the following also pretty uninteresting functions f:

(a) $f(t) = 2t^2 + 3t - 5$;
(b) $f(t) = (t^2 + 2t + 1)\mathrm{e}^t$;
(c) $f(t) = t\mathrm{e}^{2t}\cos(t) + \mathrm{e}^{2t}\sin(t)$;
(d) $f(t) = t^3\mathrm{e}^{-t}\sin(3t) + t^2\mathrm{e}^{-t}\cos(3t)$.

2.3.4 For the following scalar linear inhomogeneous ordinary differential equations F, determine the general form of their solutions:

(a) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 2x^{(1)} + x - 3\mathrm{e}^t$;
(b) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 5x^{(1)} + 6x - 2\mathrm{e}^{3t} - \cos(t)$;
(c) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 2x^{(1)} + 5x - t\mathrm{e}^t\sin(2t)$;
(d) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 4x - t\cos(2t) + \sin(2t)$;
(e) $F(t, x, x^{(1)}, x^{(2)}, x^{(3)}) = x^{(3)} - x - t\mathrm{e}^t$;
(f) $F(t, x, x^{(1)}, \ldots, x^{(4)}) = x^{(4)} + 4x^{(2)} + 4x - \cos(2t) - \sin(2t)$.

2.3.5 Solve the initial value problem for the following scalar linear inhomogeneous differential equations F with the stated initial conditions:

(a) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 2x^{(1)} + x - 3\mathrm{e}^t$, and $\xi(0) = 1$, $\dot{\xi}(0) = 1$;
(b) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 5x^{(1)} + 6x - 2\mathrm{e}^{3t} - \cos(t)$, and $\xi(0) = 0$, $\dot{\xi}(0) = 1$;
(c) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 2x^{(1)} + 5x - t\mathrm{e}^t\sin(2t)$, and $\xi(0) = 1$, $\dot{\xi}(0) = 0$;
(d) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 4x - t\cos(2t) + \sin(2t)$, and $\xi(0) = 2$, $\dot{\xi}(0) = 1$;
(e) $F(t, x^{(1)}, x^{(2)}, x^{(3)}) = x^{(3)} - x - t\mathrm{e}^t$, and $\xi(0) = 1$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 1$;
(f) $F(t, x, x^{(1)}, \ldots, x^{(4)}) = x^{(4)} + 4x^{(2)} + 4x - \cos(2t) - \sin(2t)$, and $\xi(0) = 0$, $\dot{\xi}(0) = 0$, $\ddot{\xi}(t) = 0$, $\dddot{\xi}(t) = 0$.

2.3.6 Suppose a mass m falls under the influence of gravity with gravitational acceleration a_g and suppose that the force due to air resistance is proportional to velocity, i.e., given by ρv, where v is the velocity.

(a) Use Newton's laws of force balance to write the equations governing the falling velocity of the mass.
(b) Obtain the solution to the differential equation from part (a), supposing the mass is at rest at $t = 0$.
(c) What is the terminal velocity of the mass?
(d) What are the units of m, a_g, and ρ, in terms of mass, length, and time units?
(e) Combine the physical constants m, a_g, and ρ in such a way that the units for the combined expression are "length/time," i.e., velocity. How does this constant compare to the terminal velocity you computed in part (c)?

2.3.7 Let $P \in \mathbb{R}[X]$ be given by

$$P = X^k + a_{k-1}X^{k-1} + \cdots + a_1 X + a_0,$$

and suppose that $r \in \mathbb{R}$ is not a root of P. Show that

$$\xi_{\mathrm{p}}(t) = \frac{\mathrm{e}^{rt}}{\widehat{P}(r)}$$

is a particular solution of the differential equation

$$F(t, x, x^{(1)}, \ldots, x^{(k)}) = x^{(k)} + a_{k-1}x^{(k-1)} + \cdots + a_1 x^{(1)} + a_0 x - \mathrm{e}^{rt}.$$

2.3.8 For the following scalar linear homogeneous ordinary differential equations with time-domain $\mathbb{T} = [0, \infty)$ and with $t_0 = 0$, find their Green's function:
(a) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + x^{(1)}$;
(b) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + \omega^2 x$, $\omega \in \mathbb{R}_{>0}$;
(c) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 2x^{(1)} + x$.

Section 2.4

Using a computer to work with scalar ordinary differential equations

We thank Jack Horn for putting together the MATHEMATICA® *and* MATLAB® *results in this section.*

In Sections 2.2 and 2.3 we have discussed the character of, and solved very specific examples of, scalar linear ordinary differential equations. This, however, represents a tiny subset (but, arguably, an important tiny subset) of the differential equations one might encounter in practice. Moreover, even in the simple examples where the analytical methods we have learnt *are* applicable, to apply them is often extremely tedious and error-prone. Therefore, in this section we illustrate how computers can make working with differential equations, specifically scalar ordinary differential equations, a bearable undertaking.

In the preface we listed a couple of computer packages—some symbolic, some numerical, some both—available for working with differential equations. We shall not attempt to illustrate how all of these work, but choose two as illustrative. We choose MATHEMATICA® to illustrate a computer algebra package[3] and MATLAB® to illustrate a numerical package. There is no reason for this choice, other than personal familiarity (in the case of MATHEMATICA®) and ease of access (in the case of MATLAB®).

2.4.1 The basic idea of numerically solving differential equations

While this is definitely *not* a text on numerical methods, it is worth understanding a little bit of what is under the hood when one is using a computer package to obtain numerical solutions to differential equations.

The basic step in converting a differential equation into something that can be worked with numerically is to replace derivatives with algebraic approximations. Suppose that one has a function $t \mapsto \xi(t)$. The obvious thing to do to approximate the derivative of ξ is to work with the standard difference quotient:

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) \approx \frac{\xi(t+h) - \xi(t)}{h}.$$

Here, $h \in \mathbb{R}_{>0}$ is to be thought of as small (in the limit as $h \to 0$ we get the actual derivative, if it exists), and is known as the ***time step***. Even here, there are multiple ways in which one might work with such a difference quotient; for example, here are two:

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) \approx \frac{\xi(t) - \xi(t-h)}{h}, \qquad \frac{\mathrm{d}\xi}{\mathrm{d}t}(t) \approx \frac{\xi(t+\frac{h}{2}) - \xi(t-\frac{h}{2})}{h}.$$

[3] MATHEMATICA® also does numerical computations, and indeed was used to produce the numerical results used in the book.

The first rule is call the "forward difference," the second the "backward difference," and the third the "midpoint rule." If one knows the value of ξ at time t_0, one can then get an approximation for the value of ξ at time $t_0 + h$ by

$$\xi(t_0 + h) = h\frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0) + \xi(t + 0),$$

then the value at time $t_0 + 2h$ by

$$\xi(t_0 + 2h) = h\frac{\mathrm{d}\xi}{\mathrm{d}t}(t_0 + h) + \xi(t_0 + h).$$

Then can, of course, be repeated, provided one has values for the derivatives. However, if ξ is the solution to a first-order scalar ordinary differential equation F with right-hand side $\widehat{F}$,

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)),$$

then one indeed does have the values for the derivatives. Indeed, one have

$$\begin{aligned}
\xi(t_0 + h) &= h\widehat{F}(t_0, \xi(t_0)) + \xi(t_0),\\
\xi(t_0 + 2h) &= h\widehat{F}(t_0 + h, \xi(t_0 + h)) + \xi(t_0 + h),\\
&\vdots
\end{aligned}$$

Thus we have determined a simple means of numerically generating an approximation for a solution for F given an initial condition!

We note, however, that any numerical computation package will use a much more sophisticated method for approximating derivatives than the forward difference method we have used above. Nonetheless, the basic principle is as we have outlined it in our simple illustration above.

2.4.2 Using MATHEMATICA® to obtain analytical and/or numerical solutions

For some ordinary differential equations, one can simply plug them into a computer algebra package, and out will pop the answer. So, this is always worth a shot.

Our first example illustrates this in MATHEMATICA®.

2.4.1 Example (Solving simple scalar ordinary differential equation) The first ordinary differential equation we will solve is the simple first order equation:

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) = \frac{-ty(t)}{2 - y(t)}.$$

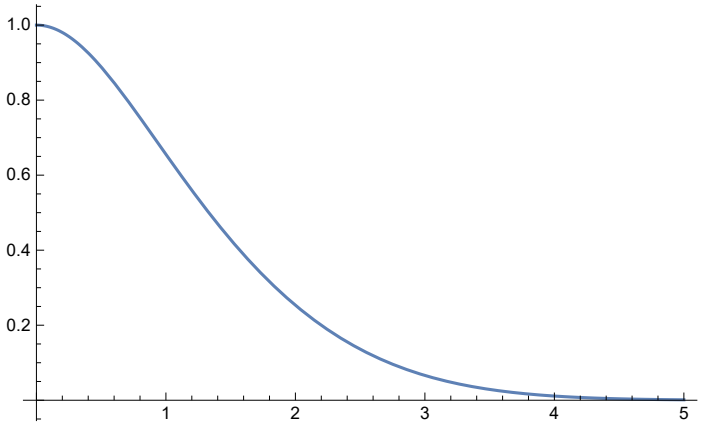
The following MATHEMATICA® script will use the `DSolve` command to solve this ordinary differential equation, then plot the solution.

soln = DSolve[{y'[t] == (−t * y[t])/(2 − y[t]), y[0] == 1}, y[t], t]

Plot[y[t]/.soln, {t, 0, 5}]

This gives the output

$$\left\{\left\{y[t] \to -2\text{ProductLog}\left[-\tfrac{1}{2}\sqrt{e^{-1-\frac{t^2}{2}}}\right]\right\}\right\}$$



Note that, as arguments to `DSolve` we give the conditions for a solution to the differential, as well as initial conditions. The syntax `y[t]/.soln` simply means that one should replace `y[t]` with its value as determined by the assignment `soln`. Also, the "`;`" at the end of a MATHEMATICA® command line means that the output will be suppressed. •

While `DSolve` is a useful command, it is also possibe to solve ordinary differential equations using MATHEMATICA® as an assistive tool, rather than just having it belt out solutions.

2.4.2 Example (Solving ordinary differential equations without using `DSolve`) We illustrate Procedure 2.2.18 for the fourth-order equation

$$\frac{\mathrm{d}^4 s}{\mathrm{d}x}(x) - \frac{\mathrm{d}^2 s}{\mathrm{d}x}(x) + 9s(x) = 0.$$

First we must find the roots of the characteristic polynomial.

CharPoly = a^4 − 10a^2 + 9 == 0;

roots = Solve[CharPoly, a];

Next, we will find the general solution.

S1 = C1 * Exp[a * x]/.roots[[1]];

S2 = C2 * Exp[a * x]/.roots[[2]];

S3 = C3 * Exp[a * x]/.roots[[3]];

```
S4 = C4 * Exp[a * x]/.roots[[4]];

GenSol = S1 + S2 + S3 + S4;
```

Once we have the general solution, we will create a system of equations using the given initial conditions to find the values for $C1$, $C2$, $C3$, and $C4$.

```
A1 = GenSol == 5/.x → 0;

A2 = D[GenSol, x] == −1/.x → 0;

A3 = D[GenSol, {x, 2}] == 21/.x → 0;

A4 = D[GenSol, {x, 3}] == −49/.x → 0;

Const = Solve[{A1, A2, A3, A4}, {C1, C2, C3, C4}];

Sol = GenSol/.Const
```

This gives the solution

$$\{2e^{-3x} - e^{-x} + 4e^{x}\}$$

We can verify this by using `DSolve`:

```
expr = s''''[x] − 10s''[x] + 9s[x] == 0;

DSolve[{expr, s[0] == 5, s'[0] == −1, s''[0] == 21, s'''[0] == −49}, s[x], x]
```

$$\left\{\left\{s[x] \to e^{-3x}\left(2 - e^{2x} + 4e^{4x}\right)\right\}\right\}$$

As you can see, both methods give the same result. •

Let us now work with a particular example with some physical motivation.

2.4.3 Example (Skydiver) Next we will look at another example, this time a second-order equation. Consider a skydiver jumping from a plane. Using Newton's laws of force balance, the governing equation is found to be:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t}(t) = -a_g + \frac{\rho}{m}\left(\frac{\mathrm{d}y}{\mathrm{d}t}(t)\right)^2.$$

The following script will solve the ordinary differential equation, and plot the jumpers position and velocity for the first twenty seconds.

```
m = 80;

g = 9.81;

p = 1.225;
```

```
sol = DSolve[{y''[t] == -g + (p * y'[t]^2) * (1/m), y[0] == 500, y'[0] == 0}, y[t], t];
a[t] = y[t]/.sol;
b[t] = D[a[t], t];
position = Plot[a[t], {t, 0, 20}]
velocity = Plot[Evaluate[b[t]], {t, 0, 20}]
```

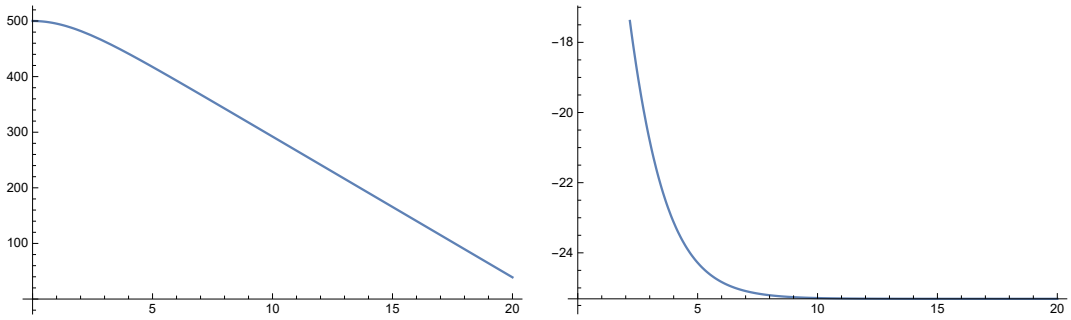


Figure 2.6 Parachuter's position (left) and velocity (right)

As can be seen from the plots, the parachuter's velocity asymptotically reaches a value determined as the inertial forces balance the aerodynamic drag forces. •

In the above examples, we obtained analytical solutions for the differential equations. Typically this is not possible, and one must obtain numerical solutions.

2.4.4 Example (Solving ordinary differential equations numerically) In this example we will show that mathematica also has the ability to solve differential equations numerically as well, again modelling a parachuter jumping from a plane. The `NDSolve` command works very similarly to the `DSolve` command, however it solves the ordinary differential equation, returning a numerical solution. We work again with the parachuter equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t}(t) = -a_g + \frac{\rho}{m}\left(\frac{\mathrm{d}y}{\mathrm{d}t}(t)\right)^2.$$

The MATHEMATICA® code is as follows.

```
NumericalSol = NDSolve[{y''[t] == -g + (p * y'[t]^2) * (1/m),
        y[0] == 500, y'[0] == 0}, y, {t, 0, 20}];
Plot[Evaluate[y[t]/.NumericalSol], {t, 0, 20}]
Plot[Evaluate[y'[t]/.NumericalSol], {t, 0, 20}]
```

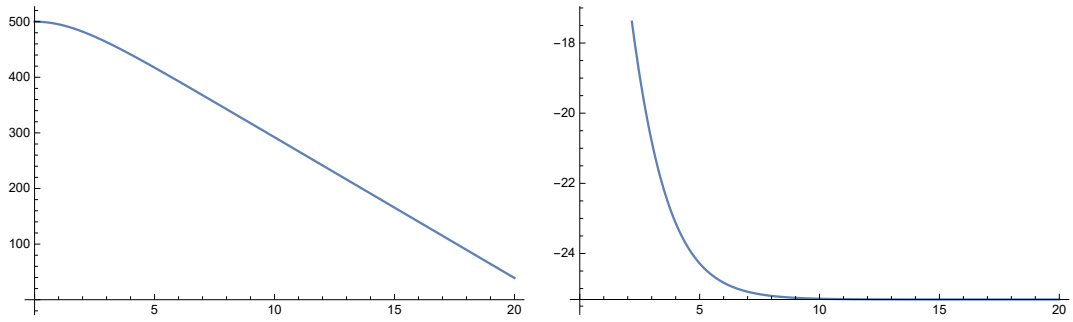


Figure 2.7 Parachuter's position (left) and velocity (right)

As you can see, the results are nearly identical when compared to the analytically obtained solutions. •

2.4.3 Using MATLAB® to obtain numerical solutions

MATLAB® is a very powerful tool for solving complicated differential equations. However, the process is not quite as simple as MATHEMATICA®. To use the `ode45` solver, you must first create a function that is your ordinary differential equation in the form $\frac{dy}{dt}(t) = F(t, y(t))$. This function must then be passed into another script that will solve it. If one types

```
odeexamples
```

at the MATLAB® prompt, you will be given you a list of examples and from these you can easily figure out how to do commonplace things using MATLAB®. To edit an example file named `foo.m`, type

```
edit foo.m
```

To run this file type

```
foo
```

in MATLAB®.

We will now consider the same two examples we covered in the section on MATHEMATICA®.

2.4.5 Example (Solving simple scalar ordinary differential equation) Below is the function that contains the same ordinary differential equation from Exercise 2.4.1. We will pass this into the following main script that will find the solution.

```
1 function [ dydt ] = Example1( t,y )
2
3 dydt = (-t*y)/(2-y);
4
5 end
```

Next we have the main script that will solve this ordinary differential equation. Note that `ode45` has three input arguments: the ordinary differential equation iteslf, time, and initial conditions. The plot that is produced by this script can be found in Figure 2.8.

```
clc
clear all
close all
%% Solving Numerically

t = linspace(0,5);
y0 = 1;

solution = ode45(@(t,y)Example1(t,y),t,y0);

%% Plotting

figure(1)
plot(solution.x,solution.y,'b')
xlabel('Time [s]');
ylabel('y(t)');

print -deps Example1Plot
```

Of course, the numerical result here agrees closely with the plot of the analytical result produced in Exercise 2.4.1. •

Next we consider the parachuter example initiated in Exercise 2.4.3.

2.4.6 Example (Skydiver) Next we will consider the same skydiver example as in Exercise 2.4.3. Again we must create a function containing the ordinary differential equation that will then be passed into the main script.

```
function [ dydt ] = Parachute(t,y)

    m = 80; %Mass, in kg, of the parachuter and their gear
    g = 9.81; %Gravitational constant
    p = 1.225; %Density  of  air in kg/m^3

    dydt = [y(2); -g+p*y(2).^2*(1/m)];
end
```

Here is the main script. The plots generated by this script can be found in Figure 2.9.

```
clc
```

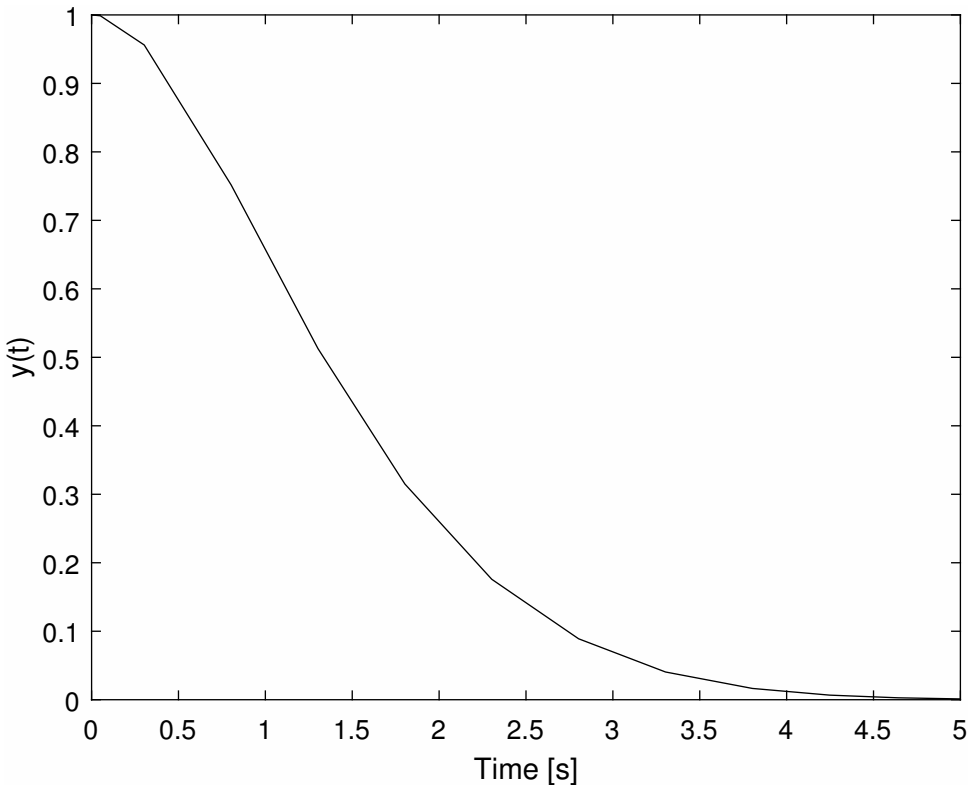


Figure 2.8 Plot generated by MATLAB® for Exercise 2.4.5

```
close all
clear all

t = linspace(0,20);

y0 = [500 0];

y = ode45(@(t,y)Parachute(t,y),t,y0);

figure(1)

subplot(2,1,1)
plot(y.x,y.y(1,:))
ylabel('Height [m]');
xlabel('Time [s]');

subplot(2,1,2)
plot(y.x,y.y(2,:))
ylabel('Velocity [m/s]');
```

```
21 xlabel('Time [s]');
```

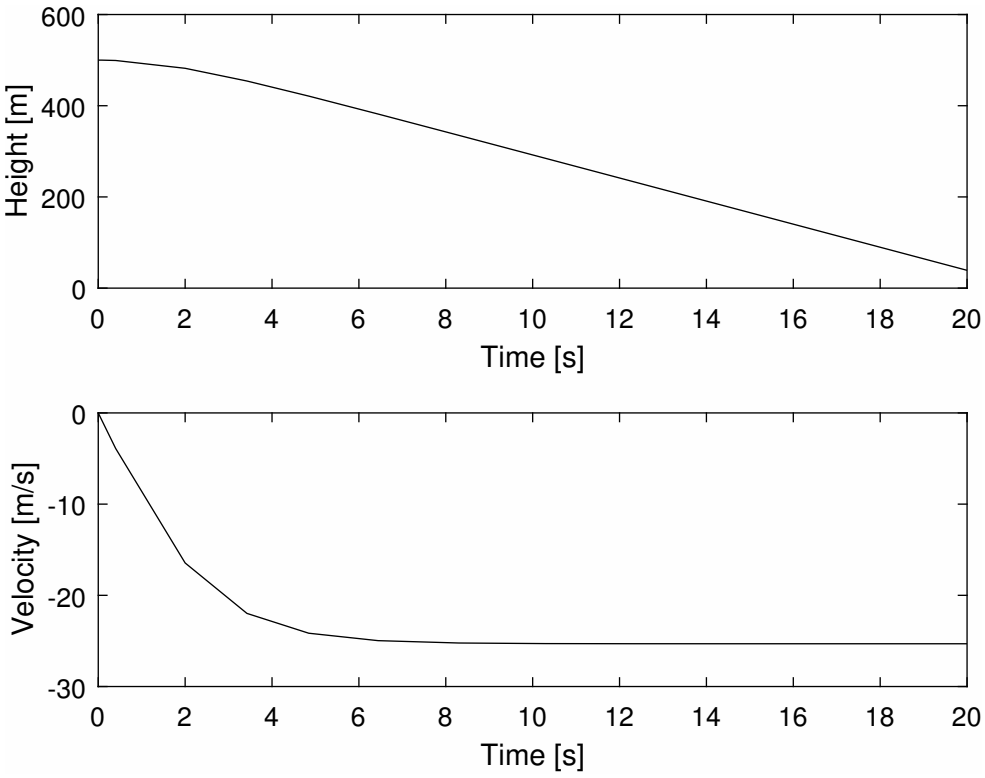


Figure 2.9 Position and velocity graphs of the parachuter Exercise 2.4.6

Again, of course, the numerical results agree with those produced by MATHEMATICA®, both analytically and numerically. •

Chapter 3

Systems of ordinary differential equations

In this chapter we extend our discussion of scalar differential equations in Chapter 2 to systems of equations. Thus, in the notation of Section 1.3.3, we consider an ordinary differential equation with time-domain $\mathbb{T} \subseteq \mathbb{R}$, state space $U \subseteq \mathbb{R}^m$, and with right-hand side

$$\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \times \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}^m) \to \mathbb{R}^m$$

giving the equation

$$\frac{\mathrm{d}^k \xi}{\mathrm{d}t^k}(t) = \widehat{\boldsymbol{F}}\left(t, \xi, \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \dots, \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t)\right)$$

for solutions $t \mapsto \xi(t)$. When we studied scalar equations in Chapter 2, we retained this higher-order form of the equations, because doing so allowed us to continue working with scalar equations. However, every scalar equation of order k can be represented as a first-order equation with k unknowns, cf. Exercise 1.3.23. In fact, in that exercise we see how to convert a kth-order differential equation in m unknowns into a first-order equation in km unknowns. The point is that, since in this chapter we are working already with vector equations, we will always suppose that our equations are first-order. Also, we will swap around our lettering from Section 1.3.3 and suppose that U is an open subset of $\mathbb{R}^n$. Thus we have a right-hand side

$$\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \to \mathbb{R}^n \tag{3.1}$$

and solutions satisfy

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)).$$

Note, however, that physically it may still be interesting to retain the higher-order form, even for vector equations, cf. the equation (1.2) modelling a coupled mass-spring system.

As with scalar ordinary differential equations, there is little that one can say in much generality about general systems of ordinary differential equations. Therefore, we focus almost entirely on linear equations in this chapter. One of the reasons that linear systems are so important is that, even for systems that are not linear,

a first step towards understanding them is often to linearise them. Thus we shall begin in Section 3.1 with a discussion of linearisation. The next two sections, 3.2 and 3.3, deal with linear systems of equations in some detail. In Section 3.4 we study, essentially, graphical representations for two-dimensional systems of ordinary differential equations, not necessarily linear. While the planar nature of the systems we consider limits the generality of the ideas we discuss, it is nonetheless the case that the ideas seen here form the basis for any serious further study of ordinary differential equations in more advanced treatments of the subject. In Section 3.5 we introduce numerical consideration of systems of ordinary differential equations.

Contents

Section 3.1

Linearisation

As we have said, if one is given a completely general system of ordinary differential equations, there is little that one can do. However, sometimes one might be able to find an isolated solution to the differential equation, and then it becomes interesting to know what one can say given this information. The first thing one typically tries is linearisation, i.e., look at the "first-order" variation of solutions from the given solution. In this section we present this method in some detail. We shall not at this point say much about what one can do after linearisation; our main objective is to understand why it might be interesting to focus our attention on linear systems, which is exactly what we do in the subsequent two sections.

3.1.1 Linearisation along solutions

Suppose that we have a system of ordinary differential equations $\boldsymbol{F}$ with right-hand side $\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n$ and that we have a solution $\xi_0\colon \mathbb{T}' \to U$ for $\boldsymbol{F}$. We wish to understand what happens to solutions "nearby" this fixed solution ξ_0.

To do this, we suppose that the map

$$\begin{aligned} \widehat{\boldsymbol{F}}_t\colon U &\to \mathbb{R}^n \\ \boldsymbol{x} &\mapsto \widehat{\boldsymbol{F}}(t, \boldsymbol{x}) \end{aligned}$$

is of class C^1. We denote

$$\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x}) = \boldsymbol{D}\widehat{\boldsymbol{F}}_t(\boldsymbol{x}), \qquad t \in \mathbb{T}.$$

We then suppose that we have a solution $\xi\colon \mathbb{T} \to U$ for $\boldsymbol{F}$ for which the deviation $\boldsymbol{\nu} \triangleq \xi - \xi_0$ is small. Let us try to understand the behaviour of $\boldsymbol{\nu}$. Naïvely, we can do this as follows:

$$\dot{\xi}(t) = \frac{\mathrm{d}(\xi_0 + \boldsymbol{\nu})}{\mathrm{d}t}(t) = \widehat{\boldsymbol{F}}(t, \xi_0(t) + \boldsymbol{\nu}) = \widehat{\boldsymbol{F}}(t, \xi_0(t)) + \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0) \cdot \boldsymbol{\nu}(t) + \cdots .$$

We will not here try to be precise about what "$\cdots$" might mean, but merely say that the idea of the preceding equation is that we approximate using the constant and first-order terms in the Taylor expansion, and then pray that this gives us something meaningful. Note that, since ξ_0 is a solution for $\boldsymbol{F}$, the approximation we arrive at is

$$\dot{\xi}(t) \approx \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0) \cdot (\xi(t) - \xi_0(t)).$$

Meaningful or not, we make the following definition.

3.1.1 Definition (Linearisation of an ordinary differential equation along a solution) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

supposing that $\widehat{\boldsymbol{F}}_t$ is of class C^1 for every $t \in \mathbb{T}$. For a solution $\xi_0\colon \mathbb{T}' \to U$ for $\boldsymbol{F}$, the ***linearisation of*** $\mathbf{F}$ ***along*** $\boldsymbol{\xi_0}$ is the linear ordinary differential equation $\boldsymbol{F}_{L,\xi_0}$ with right-hand side

$$\begin{aligned}\widehat{\boldsymbol{F}}_{L,\xi_0}\colon \mathbb{T}' \times \mathbb{R}^n &\to \mathbb{R}^n\\ (t, v) &\mapsto \widehat{\boldsymbol{DF}}(\xi_0(t)) \cdot v.\end{aligned}$$

•

Note that a solution $t \mapsto \nu(t)$ for the linearisation of $\boldsymbol{F}$ along ξ_0 satisfies

$$\dot{\nu}(t) = \boldsymbol{A}(t)(\nu(t)),$$

where

$$\boldsymbol{A}(t) = \widehat{\boldsymbol{DF}}(t, \xi_0(t)).$$

This is indeed a linear ordinary differential equation. We note that, even when $\boldsymbol{F}$ is autonomous, the linearisation will generally be nonautonomous, due to the dependence of the reference solution ξ_0 on time.

Note that there is an alternative view of linearisation that can be easily developed, one where linearisation is of the *equation*, not just along a solution. The construction we make is the following.

3.1.2 Definition (Linearisation of an ordinary differential equation) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

supposing that $\widehat{\boldsymbol{F}}_t$ is of class C^1 for every $t \in \mathbb{T}$. The ***linearisation*** of $\boldsymbol{F}$ is the ordinary differential equation $\boldsymbol{F}_L$ with right-hand side

$$\begin{aligned}\widehat{\boldsymbol{F}}_L\colon \mathbb{T} \times (U \times \mathbb{R}^n) &\to \mathbb{R}^n \oplus \mathbb{R}^n\\ (t, (x, v)) &\mapsto (\widehat{\boldsymbol{F}}(t, x), \widehat{\boldsymbol{DF}}(t, x)(v)).\end{aligned}$$

•

Solutions of the linearisation of $\boldsymbol{F}$ are then curves $t \mapsto (\xi(t), \nu(t))$ satisfying

$$\begin{aligned}\dot{\xi}(t) &= \widehat{\boldsymbol{F}}(t, \xi(t)),\\ \dot{\nu}(t) &= \widehat{\boldsymbol{DF}}(t, \xi(t)) \cdot \nu(t).\end{aligned}$$

We see, then, that in this version of linearisation we carry along the original differential equation $\boldsymbol{F}$ as part of the linearisation. This is, in no way, incompatible with the definition of linearisation along a solution ξ_0, since one needs $\boldsymbol{F}$ to provide the solution.

Let us illustrate how this works in an example. Finding nonlinear ordinary differential equations whose nontrivial solutions we can explicitly compute is not easy,[1] so we are sort of stuck with systems with one state. However, this will suffice for the illustrative purposes here.

3.1.3 Example (Linearisation of an ordinary differential equation along a solution) We work here with the logistical population model of (1.8). This is the scalar first-order ordinary differential equation with right-hand side

$$\widehat{F}(t, x) = kx(1 - x).$$

Solutions $t \mapsto \xi(t)$, therefore, satisfy

$$\dot{\xi}(t) = k\xi(t)(1 - \xi(t)).$$

This equation is separable and so can be solved using the method from Section 2.1. We skip the details, and instead just say that

$$\xi_0(t) = \frac{x_0 \mathrm{e}^{kt}}{1 + x_0(\mathrm{e}^{kt} - 1)}$$

is the solution for F satisfying $\xi_0(0) = x_0$, as long as $x_0 \notin \{0, 1\}$ (we shall consider the cases $x_0 \in \{0, 1\}$ in Example 3.1.7–1). We have

$$\boldsymbol{D}\widehat{F}(t, x) \cdot v = k(1 - 2x)v,$$

and so the linearisation F_{L,ξ_0} about the solution ξ_0 has the right-hand side

$$\widehat{F}_{L,\xi_0}(t, v) = \frac{k(1 - x_0(\mathrm{e}^{kt} + 1))}{1 + x_0(\mathrm{e}^{kt} - 1)} v.$$

Thus a solution $t \mapsto \nu(t)$ for the linearisation satisfies

$$\dot{\nu}(t) = \underbrace{\frac{k(1 - x_0(\mathrm{e}^{kt} + 1))}{1 + x_0(\mathrm{e}^{kt} - 1)}}_{a(t)} \nu(t).$$

This equation can actually be solved, as we saw in Example 2.2.5:

$$\nu(t) = v_0 \mathrm{e}^{-\int_0^t a(\tau)\,\mathrm{d}\tau} = v_0 \mathrm{e}^{k(t-t_0)} \frac{(1 + x_0(\mathrm{e}^{kt_0} - 1))^2}{(1 + x_0(\mathrm{e}^{-kt} - 1))^2},$$ [2]

where $\nu(t_0) = v_0$. Just what conclusions we can draw from this are not clear. . . nor should they be. . . The connection between a differential equation and its linearisation are not so clear at the moment. In Section 3.1.3 we shall describe the flow of the linearisation in some detail, and in doing so will arrive at a precise interpretation of linearisation. •

[1] We shall see in the next section that working with trivial solutions is easier.

[2] Integration courtesy of MATHEMATICA®.

3.1.2 Linearisation about equilibria

In this section we consider what amounts to a special case of linearisation about a solution. The solution we consider is a very particular sort of solution, as given by the following definition.

3.1.4 Definition (Equilibrium state for an ordinary differential equation) Let F be an ordinary differential equation with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n.$$

A state $x_0 \in U$ is an ***equilibrium state*** if $\widehat{F}(t, x_0) = \mathbf{0}$ for every $t \in \mathbb{T}$. •

The following result gives the relationship between equilibrium states and solutions.

3.1.5 Proposition (Equilibrium states and constant solutions) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}.$$

A state $\mathbf{x}_0 \in \mathrm{U}$ *is an equilibrium state if and only if the constant function* $\mathrm{t} \mapsto \mathbf{x}_0$ *is a solution for* $\mathbf{F}$.

Proof Let us denote by ξ_{x_0} the constant function $t \mapsto x_0$.

First suppose that x_0 is an equilibrium state. Then $\dot{\xi}_0(t) = 0$ for every $t \in \mathbb{T}$ and $\widehat{F}(t, \xi_0(t)) = \mathbf{0}$ and so

$$\dot{\xi}_0(t) = \widehat{F}(t, \xi_0(t)), \qquad t \in \mathbb{T},$$

and thus ξ_{x_0} is a solution.

Next suppose that ξ_{x_0} is a solution. Then

$$\mathbf{0} = \dot{\xi}_{x_0}(t) = \widehat{F}(t, \xi_0(t)) = \widehat{F}(t, x_0), \qquad t \in \mathbb{T},$$

so giving that x_0 is an equilibrium state. ■

Note that, as a consequence of the preceding simple result, we can linearise about the constant solution $t \mapsto x_0$ in the event that x_0 is an equilibrium state. Let us, however, use some particular language in this case.

3.1.6 Definition (Linearisation of an ordinary differential equation about an equilibrium state) Let F be an ordinary differential equation with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

supposing that $\widehat{F}_t$ is of class C^1 for every $t \in \mathbb{T}$, and let x_0 be an equilibrium state. The ***linearisation of*** $\mathbf{F}$ ***about*** $\mathbf{x_0}$ is the linear ordinary equation F_{L,x_0} with right-hand side

$$\begin{aligned} \widehat{F}_{L,x_0}\colon \mathbb{T} \times \mathbb{R}^n &\to \mathbb{R}^n \\ (t, v) &\mapsto D\widehat{F}(t, x_0) \cdot v. \end{aligned}$$

•

A solution $t \mapsto \boldsymbol{\nu}(t)$ for $\boldsymbol{F}_{L,x_0}$ satisfies

$$\dot{\boldsymbol{\nu}}(t) = \boldsymbol{A}(t)(\boldsymbol{v}),$$

where

$$\boldsymbol{A}(t) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x}_0).$$

Thus we see that the linearisation about an equilibrium point is indeed a linear ordinary differential equation, just as it should be since the same is true of the linearisation about an arbitrary solution. What is special here, however, is that the linearisation is autonomous if $\boldsymbol{F}$ is autonomous. Thus the linearisation when $\boldsymbol{F}$ is autonomous is a linear ordinary differential equation with constant coefficients.

3.1.7 Examples (Linearisation of an ordinary differential equation about an equilibrium state)

1. Let us first return to the linearisation of the logistical population model of Example 3.1.3. We have

   $$\widehat{F}(t, x) = kx(1 - x),$$

   and so there are two equilibrium states, $x_0 = 0$ and $x_0 = 1$. In Example 3.1.3 we computed the derivative of $\widehat{F}$ to be $\boldsymbol{D}\widehat{F}(t, x) \cdot v = k(1 - 2x)v$. We thus have the linearisations about $x_0 = 0$ and $x_0 = 1$ given by

   $$\widehat{F}_{L,0}(t, v) = kv, \quad \widehat{F}_{L,1}(t, v) = -kv.$$

   The solutions then satisfy the equations

   $$\dot{v}_0(t) = kv_0(t), \quad \dot{v}_1(t) = -kv_1(t),$$

   respectively. These are easily solved using Procedure 2.2.18 to give

   $$v_0(t) = v_0(0)\mathrm{e}^{kt}, \quad v_1(t) = v_1(0)\mathrm{e}^{-kt}.$$

   We see that we have exponential growth for the solutions of the linearisation about $x_0 = 0$ and exponential decay for the solutions about $x_0 = 1$.
   It turns out that this behaviour of the linearisations about the equilibrium state *is* an accurate approximation of the behaviour of the actual system near these states. We do not develop this here, but will address matters such as this in ***missing stuff***.
2. Let us consider the simple pendulum model of (1.3). This is a scalar second-order equation F whose right-hand side is

   $$\widehat{F}(t, x, x^{(1)}) = -\frac{a_g}{\ell} \sin(x).$$

In order to fit this differential equation into our linearisation framework, we must convert it into a first-order equation, as in Exercise 1.3.23. Doing this gives the first-order ordinary differential equation $\boldsymbol{F}$ with right-hand side

$$\boldsymbol{F}(t,(x_1,x_2)) = \Big(x_2, -\frac{a_g}{\ell}\sin(x)\Big).$$

This differential equation has equilibria $\boldsymbol{x}_n = (n\pi, 0)$, $n \in \mathbb{Z}$, corresponding to periodically repeated copies of the "down" and "up" rest positions of the pendulum. We shall work with two of these, $\boldsymbol{x}_0 = (0,0)$ and $\boldsymbol{x}_1 = (\pi, 0)$, as they are representative. We compute

$$\boldsymbol{D}\widehat{\boldsymbol{F}}(t,(x_1,x_2)) \cdot (v_1,v_2) = \begin{bmatrix} 0 & 1 \\ -\frac{a_g}{\ell}\cos(x) & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Now, if we compute this at the two equilibria, we have

$$\boldsymbol{DF}(t,\boldsymbol{x}_0)\cdot\boldsymbol{v} = \begin{bmatrix} 0 & 1 \\ -\frac{a_g}{\ell} & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad \boldsymbol{DF}(t,\boldsymbol{x}_1)\cdot\boldsymbol{v} = \begin{bmatrix} 0 & 1 \\ \frac{a_g}{\ell} & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

A solution $t \mapsto \boldsymbol{\nu}(t)$ of the linearisations satisfies

$$\begin{bmatrix} \dot\nu_1(t) \\ \dot\nu_2(t) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -\frac{a_g}{\ell} & 0 \end{bmatrix} \begin{bmatrix} \nu_1(t) \\ \nu_2(t) \end{bmatrix}, \quad \begin{bmatrix} \dot\nu_1(t) \\ \dot\nu_2(t) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \frac{a_g}{\ell} & 0 \end{bmatrix} \begin{bmatrix} \nu_1(t) \\ \nu_2(t) \end{bmatrix}.$$

It is possible to solve these equation using Procedure 3.2.45 below, and it turns out that the solutions are

$$\begin{aligned} \begin{bmatrix} \nu_1(t) \\ \nu_2(t) \end{bmatrix} &= \begin{bmatrix} \cos(\sqrt{a_g/\ell}t) & \sqrt{\ell/a_g}\sin(\sqrt{a_g/\ell}t) \\ -\sqrt{a_g/\ell}\sin(\sqrt{a_g/\ell}t) & \cos(\sqrt{a_g/\ell}t) \end{bmatrix} \begin{bmatrix} \nu_1(0) \\ \nu_2(0) \end{bmatrix}, \\ \begin{bmatrix} \nu_1(t) \\ \nu_2(t) \end{bmatrix} &= \begin{bmatrix} \cosh(\sqrt{a_g/\ell}t) & \sqrt{\ell/a_g}\sinh(\sqrt{a_g/\ell}t) \\ \sqrt{a_g/\ell}\sinh(\sqrt{a_g/\ell}t) & \cosh(\sqrt{a_g/\ell}t) \end{bmatrix} \begin{bmatrix} \nu_1(0) \\ \nu_2(0) \end{bmatrix}. \end{aligned}$$

In Section 1.1.2 we said a few quite informal things about how this process of linearisation is reflected in the behaviour of the pendulum near the "down" and "up" equilibria. This is reflected in the behaviour of the linearisations, in that, about the "down" equilibrium, the motion for the linearisation is periodic, and, about the "up" equilibrium, the motion diverges from $(0,0)$ most of the time. We shall be more rigorous about this in ***missing stuff***. •

Summary of linearisation constructions In this section we have illustrated the idea of linearisation in a few different contexts. The take away from these constructions is as follows.

1. The linearisation of an ordinary differential equation $\boldsymbol{F}$ about a solution ξ_0 gives rise to a linear ordinary differential equation that will generally be time-varying, even when $\boldsymbol{F}$ is autonomous.

2. It is possible to linearise an equation with n states in its entirety, to give an ordinary differential equation with $2n$ states.
3. The linearisation of an ordinary differential equation about an equilibrium state gives rise to a linear ordinary differential equation, and this linear equation is autonomous if $\boldsymbol{F}$ is autonomous.
4. At this point, we know nothing about what the linearisation of $\boldsymbol{F}$ says about $\boldsymbol{F}$. However, what is true is that linear ordinary differential equations, even with constant coefficients, arise naturally in the context of linearisation, and so are worthy of some study.

3.1.3 The flow of the linearisation

In this section, in contrast with the preceding sections, we give a very precise characterisation of linearisation. It has the benefit of being precise, but the drawback of being complicated. However, the constructions we give in this section are of some importance in subjects like optimal control theory. We shall do three things: (1) provide conditions under which the flow of an ordinary differential equation is differentiable in state and initial time, as well as final time with respect to which it is always differentiable; (2) give explicit formulae for the derivatives; (3) give an interpretation of these derivatives in terms of "wiggling" of initial conditions in state and time.

We shall first investigate thoroughly the properties of the flow of an ordinary differential equation that has more regularity properties than are required for the basic existence and uniqueness theorem, Theorem 1.4.8. In order to state the result we want, we will make use of some ideas that we will not develop fully until Section 3.2. Let us suppose that we have a system of ordinary differential equations $\boldsymbol{F}$ with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

and let $(t_0, \boldsymbol{x}_0) \in \mathbb{T} \times U$. We then have the solution

$$t \mapsto \xi_0(t) \triangleq \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0)$$

defined for $t \in J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. We then define

$$\begin{aligned} \boldsymbol{A}_{(t_0,\boldsymbol{x}_0)}\colon J_{(t_0,\boldsymbol{x}_0)} &\to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \\ t &\mapsto \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0)). \end{aligned}$$

Now consider the linear time-varying differential equation $\boldsymbol{F}^T_{(t_0,\boldsymbol{x}_0)}$ with right-hand side

$$\begin{aligned} \widehat{\boldsymbol{F}}^L_{(t_0,\boldsymbol{x}_0)}\colon J_{\boldsymbol{F}} \times \mathbb{R}^n &\to \mathbb{R}^n \\ (t, \boldsymbol{v}) &\mapsto \boldsymbol{A}_{(t_0,\boldsymbol{x}_0)}(t) \cdot \boldsymbol{v}. \end{aligned}$$

To describe solutions of this linear ordinary differential equation, we consider first the following ordinary differential equation. For $t \in J_F(t_0, x_0) \times \mathbb{R}^n$, we consider the following initial value problem:

$$\frac{d\Psi}{ds}(s) = A_{(t_0,x_0)}(s) \circ \Psi(s), \quad \Psi(t) = I_n.$$

As we shall see in the proof of the theorem immediately following, this initial value problem has solutions defined for all $s \in J_F(t_0, x_0)$. Moreover, we denote the solution at time s by $\Phi_{A_{(t_0,x_0)}}(s, t)$; the associated map

$$\Phi_{A_{(t_0,x_0)}} : J_F(t_0, x_0) \times J_{(t_0,x_0)} \to L(\mathbb{R}^n; \mathbb{R}^n)$$

is what we shall call the "state transition map" in Section 3.2.2.2, and we shall use some of the results from this section in the proof below. In particular, we shall use the fact that the solution to the initial value problem

$$\frac{dv}{ds}(s) = A_{(t_0,x_0)}(s) \cdot v(s), \quad v(t) = v_0$$

is

$$v(s) = \Phi_{A_{(t_0,x_0)}}(s, t) \cdot v_0, \qquad s \in J_F(t_0, x_0).$$

With the preceding background, we can now state the theorem.

3.1.8 Theorem (Differentiability of flows) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}} : \mathbb{T} \times U \to \mathbb{R}^n,$$

and make the following assumptions:

(i) the map $t \mapsto \widehat{\mathbf{F}}(t, \mathbf{x})$ *is continuous for each* $\mathbf{x} \in U$*;*

(ii) the map $\mathbf{x} \mapsto \widehat{\mathbf{F}}(t, \mathbf{x})$ *is continuously differentiable for each* $t \in \mathbb{T}$*;*

(iii) for each $\mathbf{x} \in U$*, there exist* $r \in \mathbb{R}_{>0}$ *and continuous functions* $g_0, g_1 : \mathbb{T} \to \mathbb{R}_{\geq 0}$ *such that*

(a) $\|\widehat{\mathbf{F}}(t, \mathbf{y})\| \leq g_0(t)$ *for* $(t, \mathbf{y}) \in \mathbb{T} \times B(r, \mathbf{x})$ *and*

(b) $\left|\frac{\partial \widehat{F}_j}{\partial x_k}(t, \mathbf{y})\right| \leq g_1(t)$ *for* $(t, \mathbf{y}) \in \mathbb{T} \times B(r, \mathbf{x})$ *and* $j, k \in \{1, \dots, n\}$.

Then the following statements hold:

(iv) for $t, t_0 \in \mathbb{T}$, $\Phi^{\mathbf{F}}_{t,t_0} : D_{\mathbf{F}}(t, t_0) \to U$ *is a* C^1*-diffeomorphism onto its image and its derivative is given by* $D\Phi^{\mathbf{F}}_{t,t_0}(\mathbf{x}_0) = \Phi_{\mathbf{A}_{(t_0,x_0)}}$.

(v) the map

$$D\Phi^{\mathbf{F}} : D_{\mathbf{F}} \to L(\mathbb{R}^n; \mathbb{R}^n)$$
$$(t, t_0, \mathbf{x}) \mapsto D\Phi^{\mathbf{F}}_{t,t_0}(\mathbf{x})$$

is continuous;

(vi) for $(t_0, \mathbf{x}_0) \in \mathbb{T} \times U$, *the set*

$$I_{\mathbf{F}}(t_0, \mathbf{x}_0) = \{t \in \mathbb{T} \mid \ t_0 \in J_{\mathbf{F}}(t, \mathbf{x}_0)\}$$

is an open interval, the map

$$\begin{aligned} \iota_{\mathbf{F},t_0,\mathbf{x}_0} \colon I_{\mathbf{F}}(t_0, \mathbf{x}_0) &\to U \\ t &\mapsto \Phi^{\mathbf{F}}(t_0, t, \mathbf{x}_0) \end{aligned}$$

is differentiable, and its derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^{\mathbf{F}}(t_0, t, \mathbf{x}_0) = -\Phi_{(t_0,\mathbf{x}_0)}(t_0, t) \cdot \widehat{\mathbf{F}}(t, \mathbf{x}_0).$$

Proof Let us first show that the hypotheses of the theorem imply those of Theorem 1.4.8(ii). Let $\boldsymbol{x} \in U$ and let $r \in \mathbb{R}_{>0}$ and $g_0, g_1 \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ be as in the statement of the theorem. For $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathsf{B}(r, \boldsymbol{x})$, the Mean Value Theorem***missing stuff*** gives

$$\begin{aligned} \left\|\widehat{\boldsymbol{F}}(t, \boldsymbol{y}_1) - \widehat{\boldsymbol{F}}(t, \boldsymbol{y}_2)\right\| &\leq \sup\{\|\boldsymbol{D}\widehat{\boldsymbol{F}}(\boldsymbol{y})\| \mid \ \boldsymbol{y} \in \mathsf{B}(r, \boldsymbol{x})\}\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| \\ &\leq g_1(t)\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|, \end{aligned}$$

giving the desired conclusion.***missing stuff***

(iv) By virtue of the proof of Theorem 1.4.13 there exists $r, r', \alpha \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$ and $t \in [t_0 - \alpha, t_0 + \alpha]$, then $\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x})$ is defined and takes values in $\mathsf{B}(r', \boldsymbol{x}_0)$. Moreover, we have

$$\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}) = \boldsymbol{x} + \int_{t_0}^{t} \widehat{\boldsymbol{F}}(s, \Phi^{\boldsymbol{F}}(s, t_0, \boldsymbol{x}))\,\mathrm{d}s$$

in this case. We note that r', r, and α depend on g_0 and L_0 according to the required inequalities

$$\left|\int_{t_0}^{t} g_0(s)\,\mathrm{d}s\right| < \frac{r'}{2}, \quad \left|\int_{t_0}^{t} L_0(s)\,\mathrm{d}s\right| < \lambda$$

for some $\lambda \in (0, 1)$.

By choosing r' small enough, there exists $g_1 \colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ locally integrable such that

$$\left|\frac{\partial \widehat{F}_j}{\partial x_k}(t, \boldsymbol{x})\right| \leq g_1(t), \qquad (t, \boldsymbol{x}) \in \mathbb{T} \times \mathsf{B}(r', \boldsymbol{x}_0).$$

We claim that, if $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$, then the ordinary differential equation $\boldsymbol{F}^T_{(t_0,\boldsymbol{x}_0)}$ with right-hand side

$$\begin{aligned} \widehat{\boldsymbol{F}}^T_{(t_0,\boldsymbol{x}_0)} \colon (t_0 - \alpha, t_0 + \alpha) \times \mathbb{R}^n &\to \mathbb{R}^n \\ (t, \boldsymbol{v}) &\mapsto \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x})) \cdot \boldsymbol{v} \end{aligned}$$

possesses unique solutions on $(t_0 - \alpha, t_0 + \alpha)$. To show this, we note by Lemma 1 from the proof of Theorem 1.4.8 that

$$t \mapsto \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^F(t, t_0, \boldsymbol{x}))$$

is locally integrable. Our assertion then follows from Proposition 3.2.5 below.

Now we show that, for each $t \in (t_0 - \alpha, t_0 + \alpha)$, Φ^F_{t,t_0} is differentiable at each $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$. Let $\rho \in (0, r)$ be small enough that $\mathsf{B}(\rho, \boldsymbol{x}) \subseteq \overline{\mathsf{B}}(r, \boldsymbol{x}_0)$ for every $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$. Let $\boldsymbol{h} \in \mathsf{B}(\rho, \boldsymbol{0})$. By the Mean Value Inequality, for $\boldsymbol{x} \in \mathsf{B}(r - \rho, \boldsymbol{x}_0)$, we have

$$\int_0^1 \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x} + s\boldsymbol{h}) \cdot \boldsymbol{h}\, \mathrm{d}s = \widehat{\boldsymbol{F}}(t, \boldsymbol{x} + \boldsymbol{h}) - \widehat{\boldsymbol{F}}(t, \boldsymbol{x}).$$

Therefore,

$$\widehat{\boldsymbol{F}}(t, \boldsymbol{x} + \boldsymbol{h}) - \widehat{\boldsymbol{F}}(t, \boldsymbol{x}) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x}) \cdot \boldsymbol{h} = \int_0^1 (\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x} + s\boldsymbol{h}) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x})) \cdot \boldsymbol{h}\, \mathrm{d}s \tag{3.2}$$

Define

$$M_t(\boldsymbol{h}) = \sup \left\{ \int_0^1 |||\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x} + s\boldsymbol{h}) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x})|||\, \mathrm{d}s \;\middle|\; \boldsymbol{x} \in \mathsf{B}(r - \rho, \boldsymbol{x}_0) \right\},$$

and note that M_t is continuous and that $M_t(\boldsymbol{0}) = 0$. For $\boldsymbol{x} \in \mathsf{B}(r-\rho, \boldsymbol{x}_0)$ and $\boldsymbol{h} \in \mathsf{B}(\rho, \boldsymbol{0})$, consider the initial value problems

$$\dot{\xi}_0(t) = \widehat{\boldsymbol{F}}(t, \xi_0(t)), \quad \xi_0(t_0) = \boldsymbol{x},$$

and

$$\dot{\xi}_1(t) = \widehat{\boldsymbol{F}}(t, \xi_1(t)), \quad \xi_1(t_0) = \boldsymbol{x} + \boldsymbol{h}.$$

Denote $\boldsymbol{\delta}(t) = \xi_1(t) - \xi_0(t)$. We then have

$$\begin{aligned} \dot{\boldsymbol{\delta}}(t) &= \widehat{\boldsymbol{F}}(t, \xi_0(t) + \boldsymbol{\delta}(t)) - \widehat{\boldsymbol{F}}(t, \xi_0(t)) \\ &= \underbrace{\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t))}_{A_{(t_0,\boldsymbol{x})}(t)} \cdot \boldsymbol{\delta}(t) + \underbrace{\int_0^1 (\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t) + s\boldsymbol{\delta}(t)) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t))) \cdot \boldsymbol{\delta}(t)\, \mathrm{d}s}_{\boldsymbol{e}(t)}, \end{aligned}$$

using (3.2). Note that

$$\begin{aligned} \|\boldsymbol{e}(t)\| &\le \int_0^1 \|\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t) + s\boldsymbol{\delta}(t)) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t)) \cdot \boldsymbol{\delta}(t)\|\, \mathrm{d}s \\ &\le \int_0^1 |||\boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t) + s\boldsymbol{\delta}(t)) - \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \xi_0(t))|||\, \|\boldsymbol{\delta}(t)\|\, \mathrm{d}s \\ &\le \|\boldsymbol{\delta}(t)\| M_t(\boldsymbol{\delta}(t)). \end{aligned}$$

Let $\boldsymbol{\nu}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\nu}}(t) = \boldsymbol{A}_{(t_0,\boldsymbol{x})}(t) \cdot \boldsymbol{\nu}(t), \quad \boldsymbol{\nu}(t_0) = \boldsymbol{h}.$$

Now, for fixing $t \in (t_0 - \alpha, t_0 + \alpha)$, we have

$$\boldsymbol{\delta}(t) = \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t, t_0) \cdot \boldsymbol{h} + \int_{t_0}^{t} \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau, t_0)\boldsymbol{e}(\tau)\,\mathrm{d}\tau,$$

by Corollary 3.3.3, noting that $\boldsymbol{\delta}(t_0) = \boldsymbol{h}$. Here $\Phi_{\boldsymbol{A}}$ is the state transition map from Section 3.2.2.2. Thus

$$\boldsymbol{\delta}(t) = \boldsymbol{\nu}(t) + \int_{t_0}^{t} \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau, t_0)\boldsymbol{e}(\tau)\,\mathrm{d}\tau.$$

Thus

$$\begin{aligned}\|\boldsymbol{\delta}(t) - \boldsymbol{\nu}(t)\| &\leq \int_{t_0}^{t} |||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau, t_0)||| \, \|\boldsymbol{e}(\tau)\|\,\mathrm{d}\tau \leq (t - t_0)|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\cdot, t_0)|||_\infty \|\boldsymbol{e}\|_\infty \\ &\leq (t - t_0)|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\cdot, t_0)|||_\infty \|\boldsymbol{\delta}(t)\| M_t(\boldsymbol{\delta}(t)),\end{aligned}$$

where the ∞-norm is over the interval $[t_0, \tau]$. As in the proof of Lemma 2(i), we have

$$\|\boldsymbol{\delta}(t)\| \leq C\|\boldsymbol{h}\|$$

for some $C \in \mathbb{R}_{>0}$. Therefore,

$$\|\boldsymbol{\delta}(t) - \boldsymbol{\nu}(t)\| \leq C'\|\boldsymbol{h}\| M_t(\boldsymbol{\delta}(t)),$$

where $C' = C(t - t_0)|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\cdot, t_0)|||_\infty$. Restoring the pre-abbreviation notation, this reads

$$\frac{\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x} + \boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}) - \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t, t_0) \cdot \boldsymbol{h}}{\|\boldsymbol{h}\|} \leq C' M_t(\boldsymbol{\delta}(t)).$$

Since $\lim_{\boldsymbol{h}\to\boldsymbol{0}} \boldsymbol{\delta}(t) = \boldsymbol{0}$ by continuity of solutions with respect to initial conditions and by definition of M_t, we have

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x} + \boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}) - \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t, t_0) \cdot \boldsymbol{h}}{\|\boldsymbol{h}\|} = 0,$$

which shows that $\Phi^{\boldsymbol{F}}_{t,t_0}$ is differentiable on $\mathsf{B}(r, \boldsymbol{x}_0)$ and for every $t \in (t_0 - \alpha, t_0 + \alpha)$, and that, moreover, the derivative satisfies the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{D}\Phi^{\boldsymbol{F}}_{t,t_0}(\boldsymbol{x}) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x})) \circ \boldsymbol{D}\Phi^{\boldsymbol{F}}_{t,t_0}(\boldsymbol{x}), \quad \boldsymbol{D}\Phi^{\boldsymbol{F}}_{t_0,t_0}(\boldsymbol{x}) = \boldsymbol{I}_n.$$

Next we show that Φ^F_{t,t_0} is *continuously* differentiable. To show this, let $\boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0)$ and let ρ be such that $\boldsymbol{x} + \boldsymbol{h} \in \mathsf{B}(r, \boldsymbol{x}_0)$. As we showed in the preceding part of the proof,

$$\boldsymbol{D}\Phi^F_{t,t_0}(\boldsymbol{x}+\boldsymbol{h}) = \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(t,t_0) = \boldsymbol{I}_n + \int_{t_0}^t \boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(\tau,t_0)\,\mathrm{d}\tau.$$

We have

$$\begin{aligned}
|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(t,t_0) - \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t,t_0)||| &\\
&\leq \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(\tau,t_0) - \boldsymbol{A}_{(t_0,\boldsymbol{x})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau,t_0)|||\,\mathrm{d}\tau \\
&\leq \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(\tau,t_0) - \boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau,t_0)|||\,\mathrm{d}\tau \\
&\quad + \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau,t_0) - \boldsymbol{A}_{(t_0,\boldsymbol{x})}(\tau) \circ \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(\tau,t_0)|||\,\mathrm{d}\tau \\
&\leq \int_{t_0}^t g_1(\tau)|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(t,t_0) - \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t,t_0)|||\,\mathrm{d}\tau \\
&\quad + |||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}|||_\infty \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) - \boldsymbol{A}_{(t_0,\boldsymbol{x})}(\tau)|||\,\mathrm{d}\tau.
\end{aligned}$$

By Lemma 1 from the proof of Theorem 1.4.13, we have

$$|||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}}(t,t_0) - \Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}(t,t_0)||| \leq |||\Phi_{\boldsymbol{A}_{(t_0,\boldsymbol{x})}}|||_\infty \mathrm{e}^{\int_{t_0}^t g_1(\tau)\,\mathrm{d}\tau} \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) - \boldsymbol{A}_{(t_0,\boldsymbol{x})}(\tau)|||\,\mathrm{d}\tau.$$

By the Dominated Convergence Theorem,

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \int_{t_0}^t |||\boldsymbol{A}_{(t_0,\boldsymbol{x}+\boldsymbol{h})}(\tau) - \boldsymbol{A}_{(t_0,\boldsymbol{x})}(\tau)|||\,\mathrm{d}\tau = 0,$$

which gives

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} |||\boldsymbol{D}\Phi^F_{t,t_0}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{D}\Phi^F_{t,t_0}(\boldsymbol{x})||| = 0,$$

which, for $t \in (t_0 - \alpha, t_0 + \alpha)$, gives the continuity of the derivative of Φ^F_{t,t_0} on $\mathsf{B}(r, \boldsymbol{x}_0)$.

The final part of the proof of the local part of the proof is to show that Φ^F_{t,t_0} is invertible with a continuously differentiable inverse. Let $r', \alpha' \in \mathbb{R}_{>0}$ and let $r \in (0, r']$ and $\alpha \in (0, \alpha']$ as above, and so such that

$$\Phi^F_{t,t_0}(\boldsymbol{x}) \in \mathsf{B}(r', \boldsymbol{x}_0), \qquad \boldsymbol{x} \in \mathsf{B}(r, \boldsymbol{x}_0),\ t \in [t_0 - \alpha, t_0 + \alpha].$$

Let $t \in (t_0 - \alpha, t_0 + \alpha) \cap \mathbb{T}$ and denote

$$V = \Phi^F_{t,t_0}(\mathsf{B}(r, \boldsymbol{x}_0)) \subseteq \mathsf{B}(r', \boldsymbol{x}_0).$$

Let $x \in \mathsf{B}(r, x_0)$. Since $y \triangleq \Phi^F_{t,t_0}(x) \in \mathsf{B}(r', x_0)$ and $t \in [t_0 - \alpha', t_0 + \alpha'] \cap \mathbb{T}$, there exists $\rho \in \mathbb{R}_{>0}$ such that, if $y' \in \mathsf{B}(\rho, y)$, then $(t_0, t, y') \in D_F$. Moreover, since $\Phi^F_{t_0,t}$ is continuous (indeed, continuously differentiable) and $\Phi^F_{t_0,t}(y) = x$, we may choose ρ sufficiently small that $\Phi^F_{t_0,t}(y') \in \mathsf{B}(r, x_0)$ if $y' \in \mathsf{B}(\rho, y)$. By the preceding part of the proof, $\Phi^F_{t_0,t}|\mathsf{B}(\rho, y)$ is continuously differentiable. Thus there is a neighbourhood of x on which the restriction of $\Phi^F_t t_0$ is invertible, continuously differentiable, and with a continuously differentiable inverse.

To complete this part of the proof, we need to prove the statement globally. To this end, let $(t_0, x_0) \in \mathbb{T} \times U$ and denote by $J_+(t_0, x_0) \subseteq \mathbb{T}$ the set of $b > t_0$ such that, for each $b' \in [t_0, b)$, there exists a relatively open interval $J \subseteq \mathbb{T}$ and a $r \in \mathbb{R}_{>0}$ such that

1. $b' \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r, x_0) \subseteq D_F$, and
3. for each $t \in J$, $\mathsf{B}(r, x_0) \ni x \mapsto \Phi^F(t, t_0, x)$ is a C^1-diffeomorphism onto its image.

By the local part of the proof above, $J_+(t_0, x_0) \neq \emptyset$. We then consider two cases.

The first case is $J_+(t_0, x_0) \cap [t_0, \infty) = \mathbb{T} \cap [t_0, \infty)$. In this case, for each $t \in \mathbb{T}$ with $t \geq t_0$, there exists a relatively open interval $J \subseteq \mathbb{T}$ and $r \in \mathbb{R}_{>0}$ such that

1. $t \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r, x_0) \subseteq D_F$, and
3. for each $\tau \in J$, $\mathsf{B}(r, x_0) \ni x \mapsto \Phi^F(\tau, t_0, x)$ is a C^1-diffeomorphism onto its image.

The second case is $J_+(t_0, x_0) \cap [t_0, \infty) \subset \mathbb{T} \cap [t_0, \infty)$. In this case we let $t_1 = \sup J_+(t_0, x_0)$ and note that $t_1 \neq \sup \mathbb{T}$. We claim that $t_1 \in J_F(t_0, x_0)$. Were this not the case, then we must have $b \triangleq \sup J_F(t_0, x_0) < t_1$. Since $b \in J_+(t_0, x_0)$, there must be a relatively open interval $J \subseteq \mathbb{T}$ containing b such that $t \in J_F(t_0, x_0)$ for all $t \in J$. But, since there are t's in J larger than b, this contradicts the definition of $J_F(t_0, x_0)$, and so we conclude that $t_1 \in J_F(t_0, x_0)$. Let us denote $x_1 = \Phi^F(t_1, t_0, x_0)$. By our local conclusions from the first part of the proof, there exists $\alpha_1, r_1 \in \mathbb{R}_{>0}$ such that $(t, t_1, x) \in D_F$ for every $t \in (t_1 - \alpha_1, t_1 + \alpha_1)$ and $x \in \mathsf{B}(r_1, x_1)$, and such that the map

$$\mathsf{B}(r_1, x_1) \ni x \mapsto \Phi^F(t, t_1, x)$$

is a C^1-diffeomorphism onto its image for every $t \in (t_1 - \alpha_1, t_1 + \alpha_1)$. Since $t \mapsto \Phi^F(t, t_0, x_0)$ is continuous and $\Phi^F(t_1, t_0, x_0) = x_1$, let $\delta \in \mathbb{R}_{>0}$ be such that $\delta < \frac{\alpha_1}{2}$ and $\Phi^F(t, t_0, x_0) \in \mathsf{B}(r_1/4, x_1)$ for $t \in (t_1 - \delta, t_1)$. Now let $\tau_1 \in (t_1 - \delta, t_1)$ and, by our hypotheses on t_1, there exists an open interval J and $r_1' \in \mathbb{R}_{>0}$ such that

1. $\tau_1 \in J$,
2. $J \times \{t_0\} \times \mathsf{B}(r_1', x_0) \subseteq D_F$, and
3. for each $\tau \in J$, $\mathsf{B}(r_1', x_0) \ni x \mapsto \Phi^F(\tau, t_0, x)$ is a C^1-diffeomorphism onto its image.

We also choose J and r_1' sufficiently small that

$$\{\Phi^F(t, t_0, x) \mid t \in J,\ x \in \mathsf{B}(r_1', x_0)\} \subseteq \mathsf{B}(r_1/2, x_1).$$

Now we claim that

$$(\tau_1 - \alpha_1, \tau_1 + \alpha_1) \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_{\boldsymbol{F}}.$$

We first show that

$$[\tau_1, \tau_1 + \alpha_1) \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_{\boldsymbol{F}}. \tag{3.3}$$

Indeed, we have $(\tau_1, t_0, \boldsymbol{x}) \in D_{\boldsymbol{F}}$ for every $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$ since $\tau_1 \in J$. By definition of J, $\Phi^{\boldsymbol{F}}(\tau_1, t_0, \boldsymbol{x}) \in \mathsf{B}(r_1/2, \boldsymbol{x}_1)$. By definition of τ_1, $t_1 - \tau_1 < \delta < \frac{\alpha_1}{2}$. Then, by definition of α_1 and r_1,

$$(t_1, \tau_1, \Phi^{\boldsymbol{F}}(\tau_1, t_0, \boldsymbol{x})) \in D_{\boldsymbol{F}}$$

for every $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. From this we conclude that $(t_1, t_0, \boldsymbol{x}) \in D_{\boldsymbol{F}}$ for every $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. Now, since

$$t \in [\tau_1, \tau_1 + \alpha_1) \quad \Longrightarrow \quad t \in (t_1 - \alpha_1, t_1 + \alpha_1),$$

we have $(t, t_1, \Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x})) \in D_{\boldsymbol{F}}$ for every $t \in (\tau_1 - \alpha_1, \tau_1 + \alpha_1)$ and $\boldsymbol{x} \in \mathsf{B}(r_1', \boldsymbol{x}_0)$. Since

$$\Phi^{\boldsymbol{F}}(t, t_1, \Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x})) = \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}),$$

we conclude (3.3). A similar but less complicated argument gives

$$(\tau_1 - \alpha_1, \tau_1) \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_{\boldsymbol{F}}.$$

Next we claim that the map

$$\mathsf{B}(r_1', \boldsymbol{x}_0) \ni \boldsymbol{x} \mapsto \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x})$$

is a C^1-diffeomorphism onto its image for every $t \in (\tau_1 - \alpha_1, \tau_1 + \alpha_1)$. By definition of τ_1, the map

$$\Phi^{\boldsymbol{F}}_{t,t_0} \colon \mathsf{B}(r_1', \boldsymbol{x}_0) \to \mathsf{B}(r_1/2, \boldsymbol{x}_1)$$

is a C^1-diffeomorphism onto its image for $t \in (\tau_1 - \alpha_1, \tau_1]$. We also have that

$$\Phi^{\boldsymbol{F}}_{t,\tau_1} \colon \mathsf{B}(r_1, \boldsymbol{x}_1) \to U$$

is a C^1-diffeomorphism onto its image for $t \in (\tau_1, \tau_1 + \alpha_1)$. Since the composition of C^1-diffeomorphisms onto their image is a C^1-diffeomorphism onto its image, our assertion follows.

By our above arguments, we have an open interval J' and $r_1' \in \mathbb{R}_{>0}$ such that

1. $t_1 \in J'$,
2. $J' \times \{t_0\} \times \mathsf{B}(r_1', \boldsymbol{x}_0) \subseteq D_{\boldsymbol{F}}$, and
3. for each $t \in J'$, $\mathsf{B}(r_1', \boldsymbol{x}_0) \ni \boldsymbol{x} \mapsto \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x})$ is a C^1-diffeomorphism onto its image.

This contradicts the fact that $t_1 = \sup J_+(t_0, \boldsymbol{x}_0)$ and so the condition

$$J_+(t_0, \boldsymbol{x}_0) \cap [t_0, \infty) \subset \mathbb{T} \cap [t_0, \infty)$$

cannot obtain.

One similarly shows that it must be the case that $J_-(t_0, \boldsymbol{x}_0) \cap (-\infty, t_0] = \mathbb{T} \cap (-\infty, t_0]$¡ where $J_-(t_0, \boldsymbol{x}_0)$ has the obvious definition.

Now we note that $\Phi^{\boldsymbol{F}}_{t,t_0}$ injective by uniqueness of solutions for $\boldsymbol{F}$. Now, assertion (iv) of the theorem now follows since the notion of "C^1-diffeomorphism" can be tested locally, i.e., in a neighbourhood of any point.

To conclude, we must show that the derivative satisfies the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0)) \circ \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0), \quad \boldsymbol{D}\Phi^{\boldsymbol{F}}(t_0, t_0, \boldsymbol{x}_0) = \boldsymbol{I}_n,$$

on $J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. Let $J_+(t_0, \boldsymbol{x}_0)$ (reusing the notation from the preceding part of the proof) be the set of $t \geq t_0$ such that $\tau \mapsto \boldsymbol{D}\Phi^{\boldsymbol{F}}(\tau, t_0, \boldsymbol{x}_0)$ satisfies the preceding initial value problem on $[t_0, t_1]$. Note that $J_+(t_0, \boldsymbol{x}_0) \neq \emptyset$ by our arguments in the first part of the proof. Let $t_1 = \sup J_+(t_0, \boldsymbol{x}_0)$. We claim that $t_1 = \sup J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. If $t_1 = t_0$ there is nothing to prove. So suppose that $t_1 > t_0$ and suppose that $t_1 \neq \sup J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. Therefore, $t_1 \in J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$ and so there exists $\alpha_1 \in \mathbb{R}_{>0}$ such that $(t_1 - \alpha_1, t_1 + \alpha_2) \subseteq J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. Let $\boldsymbol{x}_1 = \Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_0)$. Note that our arguments from the first part of the proof show that, on $(t_1 - \alpha_1, t_1 + \alpha_1)$, $t \mapsto \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1)$ satisfies the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{D}\boldsymbol{F}(t, t_1, \boldsymbol{x}_1) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1)) \circ \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1), \quad \boldsymbol{D}\Phi^{\boldsymbol{F}}(t_1, t_1, \boldsymbol{x}_1) = \boldsymbol{I}_n.$$

Now define $\boldsymbol{\Xi} \colon [t_0, t_1 + \alpha_1) \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ by

$$\boldsymbol{\Xi}(t) = \begin{cases} \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0), & t \in [t_0, t_1], \\ \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1), & t \in (t_1, t_1 + \alpha_1). \end{cases}$$

As we showed in the first part of the proof, if we denote $\boldsymbol{A}(t) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0))$, then, since

$$\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0) = \Phi^{\boldsymbol{F}}(t, t_1, \Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_0)) = \Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1)$$

for $t \in [t_1, t_1 + \alpha_1)$, we have $\boldsymbol{\Xi}(t) = \Phi_{\boldsymbol{A}_{(t_0, \boldsymbol{x}_0)}}(t, t_0)$ for $t \in [t_0, t_1 + \alpha_1)$. Thus we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{D}\boldsymbol{F}(t, t_1, \boldsymbol{x}_1) = \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1)) \circ \boldsymbol{D}\Phi^{\boldsymbol{F}}(t, t_1, \boldsymbol{x}_1), \quad \boldsymbol{D}\Phi^{\boldsymbol{F}}(t_1, t_1, \boldsymbol{x}_1) = \boldsymbol{I}_n,$$

on $[t_0, t_1 + \alpha_1)$, which contradicts the definition of $J_+(t_0, \boldsymbol{x}_0)$. Thus we must have $t_1 = \sup J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$. A similar argument can be made for $t < t_0$, and we have thus completed this part of the proof.

(v) Let us consider the ordinary differential equation $\boldsymbol{F}_1$ with right-hand side

$$\widehat{\boldsymbol{F}}_1 \colon \mathbb{T} \times U \times \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \to \mathbb{R}^n \times \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$$
$$(t, \boldsymbol{x}, \boldsymbol{X}) \mapsto (\widehat{\boldsymbol{F}}(t, \boldsymbol{x}), \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \boldsymbol{x}) \circ \boldsymbol{X}).$$

This ordinary differential equation satisfies the conditions of Theorem 1.4.8(ii). Moreover, as we saw from the previous part of the proof, $J_{F_1}(t_0, (\boldsymbol{x}_0, \boldsymbol{X}_0)) = J_{\boldsymbol{F}}(t_0, \boldsymbol{x}_0)$ for every $\boldsymbol{X}_0 \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$. Thus

$$D_{F_1} = \{(t, t_0, (\boldsymbol{x}_0, \boldsymbol{X}_0)) \mid (t, t_0, \boldsymbol{x}_0) \in D_{\boldsymbol{F}}\}.$$

From Theorem 1.4.13 we know that Φ^{F_1} is continuous. Moreover, from the first part of the proof,

$$\Phi^{F_1}(t, t_0, (\boldsymbol{x}_0, \boldsymbol{X}_0)) = (\Phi^{\boldsymbol{F}}(t, t_0, \boldsymbol{x}_0), \boldsymbol{D}\Phi^{\boldsymbol{F}}_{t,t_0}(\boldsymbol{x}_0) \circ \boldsymbol{X}).$$

From this, the desired conclusion follows.

(vi) We will show something more than is stated in this part of the theorem. The set up we will make is the following. We suppose that we have $a, b \in \mathbb{T}$ with $a < b$ and $\boldsymbol{x}_0 \in U$, and we suppose that, for some $\rho \in \mathbb{R}_{>0}$, we have a solution

$$[a - \rho, b + \rho] \ni t \mapsto \Phi^{\boldsymbol{F}}(t, a, \boldsymbol{x}_0).$$

Let us abbreviate $\boldsymbol{\xi}_0(t) = \Phi^{\boldsymbol{F}}(t, a, \boldsymbol{x}_0)$. Then, according to Theorem 1.4.13, there exists $r \in \mathbb{R}_{>0}$ such that, if $\tau \in [a, b]$ and if $(t, \boldsymbol{x}) \in (\tau - r, \tau + r) \times \mathsf{B}(r, \boldsymbol{\xi}_0(\tau))$, then the solution

$$s \mapsto \Phi^{\boldsymbol{F}}(s, t, \boldsymbol{x})$$

is defined for $s \in [a - \rho, b - \rho]$.[3] We denote

$$W_r = \cup_{\tau \in [a,b]} (\tau - r, \tau + r) \times \mathsf{B}(r, \boldsymbol{\xi}_0(\tau)).$$

We shall show that, for $t_0, t_1 \in [a, b]$, if $\boldsymbol{x}_0 = \boldsymbol{\xi}_0(t_0)$ and if $\boldsymbol{\xi}_0$ is differentiable at t_0, then the function

$$W_r \ni (t, \boldsymbol{x}) \mapsto \Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_0)$$

is differentiable at $(t_0, \boldsymbol{x}_0)$, and that its derivative is the linear map

$$(\sigma, \boldsymbol{v}) \mapsto \Phi_{\boldsymbol{A}_{(t_0, \boldsymbol{x}_0)}}(t_1, t_0) \cdot (\boldsymbol{v} - \sigma \dot{\boldsymbol{\xi}}_0(t_0)).$$

This implies the conclusions of the theorem, since the conclusions of the theorem are only about the function of t, not of t and $\boldsymbol{x}$.

We make some preliminary constructions. Let $B \in \mathbb{R}_{>0}$ be such that

$$\|\Phi_{\boldsymbol{A}_{(t_0, \boldsymbol{\xi}_0(t_0))}}(t_1, t_0) \cdot \boldsymbol{v}\| \leq B\|\boldsymbol{v}\|, \qquad t_1, t_0 \in [a - \rho, b + \rho],$$

[3] The existence of such $r \in \mathbb{R}_{>0}$ follows from a compactness argument, using compactness of $\{(\tau, \boldsymbol{\xi}_0(\tau)) \mid \tau \in [a, b]\}$.

this being possible by part (v). Now define

$$\sigma(\tau) = \sup\{|||\Phi_{A_{(t_0+\tau,x_0)}}(t_1, t_0+\tau) - \Phi_{A_{(t_0,x_0)}(t_1,t_0)}(t_1, t_0)||| \mid t_0, t_1 \in [a,b]\}.$$

By uniform continuity, σ is continuous and $\lim_{\tau\to 0}\sigma(\tau) = 0$. Now let $t_0, t_1 \in [a,b]$, let $\boldsymbol{x}_0 = \boldsymbol{\xi}_0(t_0)$, and suppose that $\boldsymbol{\xi}_0$ is differentiable at t_0. Denote

$$v_0(\tau) = \frac{\|\boldsymbol{\xi}_0(t_0+\tau) - \boldsymbol{\xi}_0(t_0) - \tau\dot{\boldsymbol{\xi}}_0(t_0)\|}{|\tau|},$$

and note that v_0 is continuous for small τ and that $\lim_{\tau\to 0} v_0(\tau) = 0$. Next denote

$$D(\tau, \boldsymbol{h}) = \sup\left\{ \frac{\|\Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0+\boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t_1, t_0, \boldsymbol{x}_0) - \Phi_{A_{(t_0+\tau,x_0)}}(t_1, t_0+\tau)\cdot\boldsymbol{h}\|}{\|\boldsymbol{h}\|} \;\middle|\; t_1 \in [a,b] \right\}.$$

Note that D is continuous and that $\lim_{(\tau,\boldsymbol{h})\to(0,\boldsymbol{0})} D(\tau,\boldsymbol{h}) = 0$.

Now we estimate

$$\begin{aligned}
&\|\Phi_{A_{(t_0+\tau,x_0)}}(t_1, t_0+\tau)\cdot(\boldsymbol{x}_0+\boldsymbol{h}-\boldsymbol{\xi}_0(t_0+\tau)) - \Phi_{A_{(t_0,x_0)}}(t_1,t_0)\cdot(\boldsymbol{h}-\tau\dot{\boldsymbol{\xi}}_0(t_0))\| \\
&\qquad \le \|\Phi_{A_{(t_0+\tau,x_0)}}(t_1, t_0+\tau)\cdot(\boldsymbol{\xi}_0(t_0+\tau) - \boldsymbol{x}_0 - \tau\dot{\boldsymbol{\xi}}_0(t_0))\| \\
&\qquad\quad + \|\Phi_{A_{(t_0+\tau,x_0)}}(t_1, t_0+\tau)\cdot(\boldsymbol{h}-\tau\dot{\boldsymbol{\xi}}_0(t_0)) - \Phi_{A_{(t_0,x_0)}}(t_1,t_0)\cdot(\boldsymbol{h}-\tau\dot{\boldsymbol{\xi}}_0(t_0))\| \\
&\qquad \le f_1(\tau)(|\tau| + \|\boldsymbol{h}\|),
\end{aligned}$$

where

$$f_1(\tau) = Bv_0(\tau) + (1 + \|\dot{\boldsymbol{\xi}}_0(t_0)\|)\sigma(\tau).$$

Note that f_1 is continuous for small τ and $\lim_{\tau\to 0} f_1(\tau) = 0$.

Now we estimate

$$\begin{aligned}
&\|\Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0+\boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t_1,t_0,\boldsymbol{x}_0) - \Phi_{A_{(t_0+\tau,x_0)}}(t_1,t_0+\tau)\cdot(\boldsymbol{x}_0+\boldsymbol{h}-\boldsymbol{\xi}_0(t_0+\tau))\| \\
&\qquad = \|\Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0+\boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{\xi}_0(t_0+\tau)) \\
&\qquad\quad - \Phi_{A_{(t_0+\tau,x_0)}}(t_1,t_0+\tau)\cdot(\boldsymbol{x}_0+\boldsymbol{h}-\boldsymbol{\xi}_0(t_0+\tau))\| \\
&\qquad \le \|\Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0+\boldsymbol{h}) - \Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0) - \Phi_{A_{(t_0+\tau,x_0)}}(t_1,t_0+\tau)\cdot\boldsymbol{h}\| \\
&\qquad\quad + \|\Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{\xi}_0(t_0+\tau)) - \Phi^{\boldsymbol{F}}(t_1, t_0+\tau, \boldsymbol{x}_0) \\
&\qquad\quad - \Phi_{A_{(t_0+\tau,x_0)}}(t_1,t_0+\tau)\cdot(\boldsymbol{\xi}_0(t_0+\tau)-\boldsymbol{x}_0)\| \\
&\qquad \le D(\tau,\boldsymbol{h})(|\tau| + \|\boldsymbol{h}\|) + D(\tau, \boldsymbol{\xi}_0(t_0+\tau)-\boldsymbol{x}_0)(|\tau| + \|\boldsymbol{\xi}_0(t_0+\tau)-\boldsymbol{x}_0\|).
\end{aligned}$$

By Taylor's Theorem, we have

$$\boldsymbol{\xi}_0(t_0+\tau) - \boldsymbol{x}_0 = \tau(R(\tau) + \dot{\boldsymbol{\xi}}(t_0))$$

for a continuous function R for which $\lim_{\tau\to 0} R(\tau) = 0$. Thus, for small τ,

$$\|\Phi^F(t_1, t_0+\tau, x_0+h) - \Phi^F(t_1, t_0, x_0) - \Phi_{A_{(t_0+\tau, x_0)}}(t_1, t_0+\tau)\cdot(x_0+h-\xi_0(t_0+\tau))\| \\ \leq f_2(\tau, h)(|\tau| + \|h\|),$$

where

$$f_2(\tau, h) = D(\tau, h) + (1 + \|\dot{\xi}(t_0)\|)D(\tau, \xi_0(t_0+\tau) - x_0).$$

We note that f_2 is continuous and that $\lim_{(\tau,\|h\|)} f_2(\tau, h) = 0$.

Combining the preceding two estimates we have

$$\|\Phi^F(t_1, t_0+\tau, x_0+h) - \Phi^F(t_1, t_0, x_0) - \Phi_{A_{(t_0,x_0)}}(t_1, t_0)\cdot(h-\tau\dot{\xi}_0(t_0))\| \\ \leq (f_1(\tau) + f_2(\tau, h))(|\tau| + \|h\|).$$

We thus conclude this part of the theorem. ■

The proof of the theorem immediately gives the following result.

3.1.9 Corollary (Flow of ordinary differential equations of class C^1) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times U \to \mathbb{R}^n.$$

If $\widehat{\mathbf{F}}$ *is of class* C^1*, then* $\Phi^{\mathbf{F}}\colon D_{\mathbf{F}} \to U$ *is of class* C^1*.*

Proof From the proof of part (vi) of the preceding theorem, we have

$$\|\Phi^F(t_1, t_0+\tau_0, x_0+h) - \Phi^F(t_1, t_0, x_0) - \Phi_{A_{(t_0,x_0)}}(t_1, t_0)\cdot(h-\tau_0\dot{\xi}_0(t_0))\| \\ \leq f(\tau_0, h)(|\tau_0| + \|h\|)$$

for a continuous function f satisfying $\lim_{(\tau_0,h)\to(0,0)} f(\tau_0, h) = 0$. Note that, under the hypotheses of the corollary, this conclusion holds for every $(t_1, t_0, x_0) \in D_F$ since solutions for F are of class C^1 in this case.

Now we have

$$\begin{aligned}\Phi^F(t_1+\tau_1, t_0+\tau_0, x_0+h) &- \Phi^F(t_1, t_0, x_0) \\ &= \Phi^F(t_1+\tau_1, t_0+\tau_0, x_0+h) - \Phi^F(t_1+\tau_1, t_0, x_0) \\ &\quad + \Phi^F(t_1+\tau_1, t_0, x_0) - \Phi^F(t_1, t_0, x_0).\end{aligned}$$

This then gives

$$\begin{aligned}&\|\Phi^F(t_1+\tau_1, t_0+\tau_0, x_0+h) - \Phi^F(t_1, t_0, x_0) - \Phi_{A_{(t_0,x_0)}}(t_1, t_0)\cdot(h-\tau_0\dot{\xi}_0(t_0)) - \tau_1\dot{\xi}_0(t_1)\| \\ &\leq \|\Phi^F(t_1+\tau_1, t_0+\tau_0, x_0+h) - \Phi^F(t_1+\tau_1, t_0, x_0) - \Phi_{A_{(t_0,x_0)}}(t_1+\tau_1, t_0)\cdot(h-\tau_0\dot{\xi}_0(t_0))\| \\ &\quad + \left|\left|\left|\Phi_{A_{(t_0,x_0)}}(t_1+\tau_1, t_0) - \Phi_{A_{(t_0,x_0)}}(t_1+\tau_1, t_0)\right|\right|\right| \|h-\tau_0\dot{\xi}_0(t_0)\| \\ &\quad + \|\Phi^F(t_1+\tau_1, t_0, x_0) - \Phi^F(t_1, t_0, x_0) - \tau_1\dot{\xi}_0(t_1)\|\end{aligned}$$

Arguments like those from the proof of part (vi) of the preceding theorem then give

$$\|\Phi^F(t_1+\tau_1,t_0+\tau_0,\boldsymbol{x}_0+\boldsymbol{h})-\Phi^F(t_1,t_0,\boldsymbol{x}_0)-\Phi_{A_{(t_0,\boldsymbol{x}_0)}}(t_1,t_0)\cdot(\boldsymbol{h}-\tau_0\dot{\xi}_0(t_0))-\tau_1\dot{\xi}_0(t_1)\| \\ \leq f(\tau_1,\tau_0,\boldsymbol{h})(|\tau_1|+|\tau_0|+\|\boldsymbol{h}\|),$$

where f is a continuous function satisfying

$$\lim_{(\tau_1,\tau_0,\boldsymbol{h})\to(0,0,\boldsymbol{0})} f(\tau_1,\tau_0,\boldsymbol{h})=0.$$

From this we conclude that the Φ^F is differentiable, and, moreover, that the derivative at $(t_1,t_0,\boldsymbol{x}_0)\in D_F$ is given by the linear map

$$(\sigma_1,\sigma_0,\boldsymbol{v})\mapsto\Phi_{A_{(t_0,\boldsymbol{x}_0)}}(t_1,t_0)\cdot(\boldsymbol{v}-\sigma_0\dot{\xi}_0(t_0))-\sigma_1\dot{\xi}_0(t_1).$$

In the proof of part (iv) of the preceding theorem we showed that $(t_1,t_0,\boldsymbol{x}_0)\mapsto\Phi_{A_{(t_0,\boldsymbol{x}_0)}}(t_1,t_0)$ is continuous. Since the map

$$(t_1,t_0,\boldsymbol{x}_0)\mapsto\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_1}\Phi^F(t,t_0,\boldsymbol{x}_0)=\widehat{\boldsymbol{F}}(t_1,\Phi^F(t_1,t_0,\boldsymbol{x}_0))$$

is also continuous, we conclude in this case that Φ^F is *continuously* differentiable. ■

The next construction is a natural one, intuitively; it involves "wiggling" the initial data for an ordinary differential equation.

3.1.10 Definition (Variation of initial data) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}}\colon\mathbb{T}\times U\to\mathbb{R}^n,$$

and let $\xi_0\colon\mathbb{T}'\to U$ be a solution for $\boldsymbol{F}$ satisfying $\xi_0(t_0)=\boldsymbol{x}_0$ for some $t_0\in\mathbb{T}'$ and $\boldsymbol{x}_0\in U$. A ***variation*** of the initial data $(t_0,\boldsymbol{x}_0)$ in the direction of $(\tau,\boldsymbol{v})\in\mathbb{R}\times\mathbb{R}^n$ is the curve

$$s\mapsto(t_0+s\tau,\boldsymbol{x}_0+s\boldsymbol{v}),$$

which we assume takes values in $\mathbb{T}\times U$ for small $s\in\mathbb{R}_{>0}$. •

For s small, one can then consider "perturbations" of the solution $t\mapsto\xi_0(t)=\Phi^F(t,t_0,\boldsymbol{x}_0)$, by which we mean the solutions $t\mapsto\Phi^F(t,t_0+s\tau,\boldsymbol{x}_0+s\boldsymbol{v})$. Note, by Theorem 1.4.13(ix), that if $(t,t_0,\boldsymbol{x}_0)\in D_F$, then $(t,t_0+s\tau,\boldsymbol{x}_0+s\boldsymbol{v})\in D_F$ for s sufficiently small. Thus we can ask for the "first-order effect" of the variation of the initial data on the solution at the final time t. Precisely, this is

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\Phi^F(t,t_0+s\tau,\boldsymbol{x}_0+s\boldsymbol{v})\in\mathbb{R}^n.$$

This is sufficiently interesting a quantity that we give it a name.

3.1.11 Definition (Infinitesimal variation corresponding to variation of initial data) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

and let $\xi_0\colon \mathbb{T}' \to U$ be a solution for $\boldsymbol{F}$ satisfying $\xi_0(t_0) = x_0$ for some $t_0 \in \mathbb{T}'$ and $x_0 \in U$. The ***infinitesimal variation*** associated with the variation of the initial data (t_0, x_0) in the direction of $(\tau, v) \in \mathbb{R} \times \mathbb{R}^n$ is

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \Phi^{\boldsymbol{F}}(t, t_0 + s\tau, x_0 + sv) \in \mathbb{R}^n. \qquad \bullet$$

The following result, which is an immediate consequence of Theorem 3.1.8, gives the formula for this first-order effect.

3.1.12 Corollary (The infinitesimal variation corresponding to a variation of initial data) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}},$$

and let $\xi_0\colon \mathbb{T}' \to \mathrm{U}$ *be a solution for* $\mathbf{F}$ *satisfying* $\xi_0(\mathrm{t}_0) = \mathbf{x}_0$ *for some* $\mathrm{t}_0 \in \mathbb{T}'$ *and* $\mathbf{x}_0 \in \mathrm{U}$. *The infinitesimal variation associated with the variation of the initial data* $(\mathrm{t}_0, \mathbf{x}_0)$ *in the direction of* $(\tau, \mathbf{v}) \in \mathbb{R} \times \mathbb{R}^{\mathrm{n}}$ *is given by*

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \Phi^{\mathbf{F}}(\mathrm{t}, \mathrm{t}_0 + \mathrm{s}\tau, \mathbf{x}_0 + \mathrm{s}\mathbf{v}) = \Phi_{\mathbf{A}_{(\mathrm{t}_0,\mathbf{x}_0)}}(\mathrm{t}, \mathrm{t}_0) \cdot \mathbf{v} - \tau\Phi_{\mathbf{A}_{(\mathrm{t}_0,\mathbf{x}_0)}}(\mathrm{t}, \mathrm{t}_0) \cdot \widehat{\mathbf{F}}(\mathrm{t}_0, \mathbf{x}_0).$$

Proof This follows from Theorem 3.1.8 and the Chain Rule. ■

3.1.4 While we're at it: ordinary differential equations of class C^m

In the previous section we considered ordinary differential equations depending continuously differentiably on state (Theorem 3.1.8) and on state and time (Corollary 3.1.9). In this section we extend these result to case where we assume more differentiability.

Let us start with just differentiability in state.

3.1.13 Theorem (Higher-order differentiability of flows) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}},$$

let $\mathrm{m} \in \mathbb{Z}_{>0}$, *and make the following assumptions:*

(i) *the map* $\mathrm{t} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})$ *is continuous for each* $\mathbf{x} \in \mathrm{U}$;

(ii) *the map* $\mathbf{x} \mapsto \widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})$ *is of class* C^{m} *for each* $\mathrm{t} \in \mathbb{T}$;

(iii) *for each* $\mathbf{x} \in \mathrm{U}$, *there exist* $\mathrm{r} \in \mathbb{R}_{>0}$ *and continuous functions* $\mathrm{g}_0, \mathrm{g}_1, \ldots, \mathrm{g}_{\mathrm{m}}\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ *such that*

(a) $\|\widehat{\mathbf{F}}(t, \mathbf{y})\| \leq g_0(t)$ *for* $(t, \mathbf{y}) \in \mathbb{T} \times \mathsf{B}(r, \mathbf{x})$ *and*

(b) $\left| \dfrac{\partial^l \widehat{F}_j}{\partial x_{k_1} \cdots \partial x_{k_l}}(t, \mathbf{y}) \right| \leq g_l(t)$ *for* $(t, \mathbf{y}) \in \mathbb{T} \times \mathsf{B}(r, \mathbf{x})$, $j, k_1, \ldots, k_l \in \{1, \ldots, n\}$, *and* $l \in \{1, \ldots, m\}$.

Then, for $t, t_0 \in \mathbb{T}$, $\Phi^F_{t,t_0} : \mathrm{D}_F(t, t_0) \to U$ *is a* C^m*-diffeomorphism onto its image.*

Proof It suffices to prove the theorem locally, since once this is done, one can use an argument like that in the proof of Theorem 3.1.8(iv) to get the global result.

We prove the result by induction on m, the result for $m = 1$ having been proved in Theorem 3.1.8. So suppose the result true for $m = r \in \mathbb{Z}_{>0}$, and that $\boldsymbol{F}$ satisfies the hypotheses of the theorem for $m = r + 1$. Then, for $(t_0, \boldsymbol{x}_0)$, the ordinary differential equation $\boldsymbol{F}_{1,(t_0,\boldsymbol{x}_0)}$ with right-hand side

$$\begin{aligned} \widehat{\boldsymbol{F}}_{1,(t_0,\boldsymbol{x}_0)} : \mathbb{T} \times \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) &\to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \\ (t, \boldsymbol{X}) &\mapsto \boldsymbol{D}\widehat{\boldsymbol{F}}(t, \Phi^F(t, t_0, \boldsymbol{x}_0)) \circ \boldsymbol{X} \end{aligned}$$

satisfies the hypotheses of the theorem for $m = r$.

missing stuff

According to the induction hypotheses and Theorem 3.1.8(iv), we conclude that $\boldsymbol{D}\Phi^F_{t,t_0}$ is of class C^r, i.e., Φ^F_{t,t_0} is of class C^{r+1}, as desired. ■

Exercises

3.1.1 Let F be a kth-order scalar ordinary differential equation with right-hand side

$$\widehat{F} : \mathbb{T} \times U \times \mathrm{L}^{\leq k-1}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}.$$

Let $\boldsymbol{F}_1$ be the first-order ordinary differential equation with k states, as in Exercise 1.3.23. Denote the state for F by $x \in U$ and the state for $\boldsymbol{F}_1$ by $\boldsymbol{y} \in U_1 = U \times \mathbb{R}^{k-1}$, as in Exercise 1.3.23.

(a) Argue that the correct definition of an equilibrium state for the kth-order ordinary differential equation F is a state $x_0 \in U$ such that

$$\widehat{F}(t, x_0, 0, \ldots, 0) = 0.$$

(b) Show that $x_0 \in U$ is an equilibrium for F as in part (a) if and only if $(x_0, 0, \ldots, 0)$ is an equilibrium state for $\boldsymbol{F}_1$.

Now let $x_0 \in U$ be an equilibrium state for F, as in part (a), with $\boldsymbol{y}_0 = (x_0, 0, \ldots, 0) \in U_1$ the associated equilibrium state for $\boldsymbol{F}_1$.

(c) Determine the linearisation of $\boldsymbol{F}_1$ about an equilibrium state $\boldsymbol{y}_0 = (x_0, 0, \ldots, 0)$.

(d) Show that the linearisation of $\boldsymbol{F}_1$ is a first-order linear ordinary differential equation with k states that comes from a kth-order scalar linear ordinary differential equation, and determine explicitly the coefficients in this scalar equation in terms of $\widehat{\boldsymbol{F}}$.

3.1.2 For the ordinary differential equations $\boldsymbol{F}$ with the given time-domains, state spaces, and right-hand sides, determine their equilibrium states and the linearisations about these equilibrium states:

(a) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}$, and $\widehat{F}(t, x) = x - x^3$;

(b) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}$, and $\widehat{F}(t, x) = a(t)x$, $a \in \mathsf{C}^0(\mathbb{T}; \mathbb{R})$ not identically zero;

(c) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}$, and $\widehat{F}(t, x) = \cos(x)$;

(d) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}^2$, and $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (x_2, x_1 - x_1^3)$;

(e) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}^2$, and $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (x_2, a(t)x_1)$, $a \in \mathsf{C}^0(\mathbb{T}; \mathbb{R})$ not identically zero;

(f) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}^2$, and $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (x_2, \cos(x_1))$;

(g) $\mathbb{T} = \mathbb{R}$, $U = \mathbb{R}^2_{>0}$, and $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (\alpha x_1 - \beta x_1 x_2, \delta x_1 x_2 - \gamma x_2)$, $\alpha, \beta, \delta, \gamma \in \mathbb{R}_{>0}$.

Section 3.2

Systems of linear homogeneous ordinary differential equations

In this section we shall begin our study of systems of linear ordinary differential equations by working with homogeneous systems. Having just specified that we will work with first-order ordinary differential equations with right-hand sides of form (3.1), for linear systems we immediately abandon this form by working with systems whose state space is a general finite-dimensional vector space V. To do this requires a tiny bit of effort to do the requisite calculus.

3.2.1 Working with general vector spaces

Let us make a few simple definitions.

3.2.1 Definition (Vector space-valued functions) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, let V be a finite-dimensional $\mathbb{F}$-vector space, and let $\xi \colon \mathbb{T} \to \mathsf{V}$ be a map. Let $\{e_1, \dots, e_n\}$ be a basis for V and write

$$\xi(t) = \xi_1(t)e_1 + \cdots + \xi_n(t)e_n$$

for maps $\xi_1, \dots, \xi_n \colon \mathbb{T} \to \mathbb{F}$.

(i) The map ξ is of ***class*** C^r, $r \in \mathbb{Z}_{\geq 0}$, if $\xi_1, \dots, \xi_n \in \mathsf{C}^r(\mathbb{T}; \mathbb{F})$.

(ii) We denote by $\mathsf{C}^r(\mathbb{T}; \mathsf{V})$ the set of mappings of class C^r.

(iii) If ξ is of class C^r, then the ***rth derivative*** of ξ is the map $\frac{\mathrm{d}^r\xi}{\mathrm{d}t^r} \colon \mathbb{T} \to \mathsf{V}$ defined by

$$\frac{\mathrm{d}^r\xi}{\mathrm{d}t^r}(t) = \sum_{j=1}^{n} \frac{\mathrm{d}^r\xi_j}{\mathrm{d}t^r} e_j, \qquad t \in \mathbb{T}. \qquad \bullet$$

One may verify that these definitions are independent of the basis chosen to make them (see Exercise 3.2.1).

We note that $\mathsf{C}^r(\mathbb{T}; \mathsf{V})$ is a vector space with vector addition

$$(\xi_1 + \xi_2)(t) = \xi_1(t) + \xi_2(t)$$

and scalar multiplication

$$(a\xi)(t) = a(\xi(t))$$

for $\xi, \xi_1, \xi_2 \in \mathsf{C}^r(\mathbb{T}; \mathsf{V})$ and $a \in \mathbb{F}$.

As we have always done, we will need notation for representing derivatives as variables for vector space-valued maps. This we do just as in the case of $\mathbb{R}^n$-valued maps. Here we only need first derivatives since we work only with first-order ordinary differential equations. We shall also keep in mind Remark 1.3.4 regarding the simpler nature of derivatives for functions of a single variable. All that being

said... we represent the variable for the first derivative for maps from $\mathbb{T}$ to V by $x^{(1)} \in \mathsf{V}$.

We will also require similar notions for linear maps.

3.2.2 Definition (Linear map-valued functions) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, let U and V be finite-dimensional $\mathbb{F}$-vector spaces, and let $L \colon \mathbb{T} \to \mathrm{L}(\mathsf{U}; \mathsf{V})$ be a map. Let $\{f_1, \dots, f_m\}$ and $\{e_1, \dots, e_n\}$ be bases for U and V, respectively, and write

$$L(t)(f_a) = \sum_{j=1}^{n} L_{ja}(t) e_j, \qquad a \in \{1, \dots, m\},$$

for maps $L_{ja} \colon \mathbb{T} \to \mathbb{F}$, $j \in \{1, \dots, n\}$, $a \in \{1, \dots, m\}$.

(i) The map L is of ***class*** $\mathbf{C^r}$, $r \in \mathbb{Z}_{\geq 0}$, if $L_{ja} \in \mathrm{C}^r(\mathbb{T}; \mathbb{F})$, $j \in \{1, \dots, n\}$, $a \in \{1, \dots, m\}$.

(ii) If L is of class C^r, then the ***rth derivative*** of L is the map $\frac{\mathrm{d}^r L}{\mathrm{d}t^r} \colon \mathbb{T} \to \mathrm{L}(\mathsf{U}; \mathsf{V})$ defined by

$$\frac{\mathrm{d}^r L}{\mathrm{d}t^r}(t)(f_a) = \sum_{j=1}^{n} \frac{\mathrm{d}^r L_{ja}}{\mathrm{d}t^r}(e_j), \qquad a \in \{1, \dots, m\},\ t \in \mathbb{T}. \qquad \bullet$$

Again, one may verify that these definitions are independent of the bases chosen to make them (see Exercise 3.2.2).

We shall make use of the "dot" notation for derivatives when it is convenient to do so. Thus we shall write

$$\dot{\xi}(t) = \frac{\mathrm{d}\xi}{\mathrm{d}t}(t), \quad \dot{L}(t) = \frac{\mathrm{d}L}{\mathrm{d}t}(t)$$

for V- and $\mathrm{L}(\mathsf{U}; \mathsf{V})$-valued functions ξ and L.

With these definitions, we can then make sense of a linear ordinary differential equation in a vector space.

3.2.3 Definition (System of linear ordinary differential equations) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let V be an n-dimensional $\mathbb{F}$-vector space.

(i) A ***system of linear ordinary differential equations*** in V is a map $F \colon \mathbb{T} \times \mathsf{V} \oplus \mathsf{V} \to \mathsf{V}$ of the form

$$F(t, x, x^{(1)}) = A_1(t)(x^{(1)}) + A_0(t)(x) - b_0(t)$$

for maps $A_0, A_1 \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ and $b_0 \colon \mathbb{T} \to \mathsf{V}$, where $A_1(t)$ is invertible for every $t \in \mathbb{T}$.

(ii) The ***right-hand side*** of a system of linear ordinary differential equations F is the map $\widehat{F} \colon \mathbb{T} \times \mathsf{V} \to \mathsf{V}$ is the map defined by

$$\widehat{F}(t, x) = -A_1(t)^{-1} \circ A_0(t)(x) + A_1(t)^{-1}(b_0(t)).$$

We shall typically denote $A(t) = -A_1(t)^{-1} \circ A_0(t)$ and $b(t) = A_1(t)^{-1}(b_0(t))$.

(iii) The system of linear ordinary differential equations F
  (a) is ***homogeneous*** if $b(t) = 0$ for every $t \in \mathbb{T}$,
  (b) is ***inhomogeneous*** if $b(t) \neq 0$ for some $t \in \mathbb{T}$, and
  (c) has ***constant coefficients*** if A is a constant map.

(iv) A ***solution*** for a system of linear ordinary differential equations F is a map $\xi \in \mathsf{C}^1(\mathbb{T}'; \mathsf{V})$ defined on a subinterval $\mathbb{T}' \subseteq \mathbb{T}$ and satisfying

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) = A(t)(\xi(t)) + b(t), \qquad t \in \mathbb{T}'.$$

Having gone to the effort of making the above revisions of our usual definition of a linear ordinary differential equation, we will say a few words about why we did this. When we study in detail linear ordinary differential equations with constant coefficients in Section 3.2.3, we shall make some rather detailed constructions with linear algebra. It is actually less confusing to do this in the setting of general vector spaces, since the coordinates of $\mathbb{R}^n$ are a mere distraction. That being said, a reader will come to no great harm if, in their mind, they replace "V" with "$\mathbb{R}^n$," as indeed all of our examples will come in this form. Indeed, in practice, even if a specific application does *not* come with $\mathsf{V} = \mathbb{R}^n$, one will normally choose a basis in which to represent the differential equation, after which one will be in the standard situation. In Exercise 3.2.3 we show that this is a valid thing to do.

3.2.2 Equations with time-varying coefficients

In this section we work with a system of linear homogeneous ordinary differential equations F in a finite-dimensional $\mathbb{R}$-vector space V, whose right-hand side, therefore, takes the form

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V} \\ (t, x) &\mapsto A(t)(x) \end{aligned} \tag{3.4}$$

for a map $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$. Thus we are looking at differential equations whose solutions $t \mapsto \xi(t)$ satisfy

$$\dot{\xi}(t) = A(t)(\xi(t)).$$

In this section we shall examine the basic properties of these solutions, and the set of all solutions, just as we did for scalar equations in Section 2.2.1.

3.2.2.1 Solutions and their properties First let us verify that the basic existence and uniqueness result holds for the differential equations we are considering.

3.2.4 Proposition (Local existence and uniqueness of solutions for systems of linear homogeneous ordinary differential equations) *Consider the system of linear homogeneous ordinary differential equations* F *with right-hand side (3.4) and suppose that* $A\colon \mathbb{T} \to L(V;V)$ *is continuous. Let* $(t_0, x_0) \in \mathbb{T} \times V$. *Then there exists an interval* $\mathbb{T}' \subseteq \mathbb{T}$ *and a map* $\xi\colon \mathbb{T}' \to V$ *of class* C^1 *that is a solution for* F *and which satisfies* $\xi(t_0) = x_0$. *Moreover, if* $\tilde{\mathbb{T}}' \subseteq \mathbb{T}$ *is another subinterval and if* $\tilde{\xi}\colon \tilde{\mathbb{T}}' \to V$ *is another* C^1*-solution for* F *satisfying* $\tilde{\xi}(t_0) = x_0$, *then* $\tilde{\xi}(t) = \xi(t)$ *for every* $t \in \tilde{\mathbb{T}}' \cap \mathbb{T}'$.

Proof By choosing a basis for V, we can take $V = \mathbb{R}^n$ so that A is an $n \times n$ matrix-valued function, which we denote as $\boldsymbol{A}$ in the usual way. (This is legitimate by Exercise 3.2.3.) We denote the components of $\boldsymbol{A}(t)$ by $A_{jk}(t)$, $j,k \in \{1,\dots,n\}$. The following technical lemma will be useful.

1 Lemma *For* $(v_1,\dots,v_n) \in \mathbb{R}^n$,

$$\sum_{j=1}^{n} |v_j| \le \sqrt{n}\left(\sum_{j=1}^{n} |v_j|^2\right)^{1/2}.$$

Proof Note that

$$\alpha = \frac{1}{n}\sum_{j=1}^{n} |v_j|$$

is the average of the positive numbers $|v_1|,\dots,|v_n|$. Thus we can write each of these numbers as this average divided by n plus the difference: $|v_j| = \alpha + \delta_j$. Note that $\sum_{j=1}^{n} \delta_j = 0$. Now compute

$$\left(\sum_{j=1}^{n} |v_j|^2\right)^{1/2} = \left(\sum_{j=1}^{n} (\alpha+\delta_j)^2\right)^{1/2} = \left(\sum_{j=1}^{n} (\alpha^2 + 2\alpha\delta_j + \delta_j^2)\right)^{1/2} \ge \left(\sum_{j=1}^{n} \alpha^2\right)^{1/2} = \sqrt{n}\alpha,$$

using the fact that $\sum_{j=1}^{n} \delta_j = 0$. This is the desired result upon employing the definition of α. ▼

We shall prove the proposition under the hypothesis that A is locally integrable, meaning that there exists a locally integrable function $g\colon \mathbb{T} \to \mathbb{R}_{\ge 0}$ such that $A_{jk}(t) \le g(t)$ for $t \in \mathbb{T}$. In this case, we must relax the conclusions of the theorem from assert only that solutions are locally absolutely continuous, not necessarily continuously differentiable. Let $a, b \in \mathbb{T}$, $a < b$, be such that $t_0 \in [a,b]$. The following estimate

will be useful for us: for any $x_1, x_2 \in \mathbb{R}^n$ and $t \in [a,b]$,

$$\begin{aligned}
\|\widehat{\boldsymbol{F}}(t,x_1) - \widehat{\boldsymbol{F}}(t,x_2)\| &= \|\boldsymbol{A}(t)(x_1) - \boldsymbol{A}(t)(x_2)\| = \|\boldsymbol{A}(t)(x_1 - x_2)\| \\
&= \left(\sum_{j=1}^n \left(\sum_{k=1}^n A_{jk}(t)(x_{1,k} - x_{2,k})\right)^2\right)^{1/2} \\
&\leq \left(\sum_{j=1}^n \left(\sum_{k=1}^n |A_{jk}(t)(x_{1,k} - x_{2,k})|\right)^2\right)^{1/2} \\
&\leq \left(\sum_{j=1}^n \left(g(t)\sum_{k=1}^n |x_{1,k} - x_{2,k}|\right)^2\right)^{1/2} = \left(g(t)^2 \sum_{j=1}^n \left(\sum_{k=1}^n |x_{1,k} - x_{2,k}|\right)^2\right)^{1/2} \\
&\leq \left(g(t)^2 \sum_{j=1}^n \sum_{k=1}^n |x_{1,k} - x_{2,k}|^2\right)^{1/2} \leq \left(n g(t)^2 \sum_{k=1}^n |x_{1,k} - x_{2,k}|^2\right)^{1/2} \\
&= \sqrt{n} g(t) \left(\sum_{k=1}^n |x_{1,k} - x_{2,k}|^2\right)^{1/2} = \sqrt{n} g(t) \|x_1 - x_2\|.
\end{aligned}$$

Let us take $h(t) = \sqrt{n} g(t)$, noting that h is locally integrable. We consider the Banach space $\mathsf{C}^0([a,b];\mathbb{R}^n)$ with the norm

$$\|\boldsymbol{f}\|_{\infty,h,t_0} = \sup\left\{ \left\| \boldsymbol{f}(t) \mathrm{e}^{-2\int_{t_0}^t h(s)\,\mathrm{d}s} \right\| \;\middle|\; t \in [a,b] \right\}.$$

Let us define

$$F_+ \colon \mathsf{C}^0([a,b];\mathbb{R}^n) \to \mathsf{C}^0([a,b];\mathbb{R}^n)$$

by

$$F_+(\xi)(t) = x_0 + \int_{t_0}^t \boldsymbol{A}(s)(\xi(s))\,\mathrm{d}s.$$

We now estimate, for $t \in [a,b]$,

$$\begin{aligned}
\|F_+(\xi_1)(t) - F_+(\xi_2)(t)\| &= \left\| \int_{t_0}^t \boldsymbol{A}(s)(\xi_1(s) - \xi_2(s))\,\mathrm{d}s \right\| \\
&\leq \int_{t_0}^t \|\boldsymbol{A}(s)(\xi_1(s) - \xi_2(s))\|\,\mathrm{d}s \\
&\leq \int_{t_0}^t \|\xi_1(s) - \xi_2(s)\| \mathrm{e}^{-2\int_{t_0}^s h(\tau)\,\mathrm{d}\tau} h(s) \mathrm{e}^{2\int_{t_0}^s h(\tau)\,\mathrm{d}\tau}\,ds \\
&\leq \frac{1}{2}\|\xi_1 - \xi_2\|_{\infty,h,t_0} \int_{t_0}^t \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{e}^{2\int_{t_0}^s h(\tau)\,\mathrm{d}\tau}\,\mathrm{d}s \\
&\leq \frac{1}{2}\|\xi_1 - \xi_2\|_{\infty,h,t_0} \mathrm{e}^{2\int_{t_0}^t h(s)\,\mathrm{d}s}.
\end{aligned}$$

From this we conclude that

$$\|F_+(\xi_1) - F_+(\xi_2)\|_{\infty,L} \leq \frac{1}{2}\|\xi_1 - \xi_2\|_{\infty,L}.$$

Now one argues just as in the proof of Theorem 1.4.8(ii), using the Contraction Mapping Theorem to conclude the existence of a unique solution ξ_+ for F on $[a, b]$. Moreover, since

$$\xi(t) = x_0 + \int_{t_0}^{t} A(s)(\xi(s))\,\mathrm{d}s,$$

we see that ξ is locally absolutely continuous and satisfies the initial conditions. ■

Next, as for scalar linear ordinary differential equations, we show that solutions exist for all time.

3.2.5 Proposition (Global existence of solutions for systems of linear homogeneous ordinary differential equations) *Consider the system of linear homogeneous ordinary differential equations* F *with right-hand side (3.4) and suppose that* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ *is continuous. If* $\xi\colon \mathbb{T}' \to \mathsf{V}$ *is a solution for* F*, then there exists a solution* $\overline{\xi}\colon \mathbb{T} \to \mathsf{V}$ *for which* $\overline{\xi}|\mathbb{T}' = \xi$.

Proof Note that in the proof of Proposition 3.2.4 we showed that solutions of the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x_0,$$

exist on any interval $[a, b] \subseteq \mathbb{T}$ containing t_0. So let $t \in \mathbb{T}$ and let $[a, b]$ be an interval containing both t_0 and t. We then have a solution for the initial value problem that is defined at t. Since $t \in \mathbb{T}$ is arbitrary, the result follows. ■

Now we can discuss the set of all solutions of a system of linear homogeneous ordinary differential equation F with right-hand side

$$\begin{aligned}\widehat{F}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V}\\ (t, x) &\mapsto A(t)(x).\end{aligned}$$

To this end, we denote by

$$\mathrm{Sol}(F) = \left\{\xi \in \mathsf{C}^1(\mathbb{T};\mathsf{V}) \;\middle|\; \dot{\xi}(t) = A(t)(\xi(t))\right\}$$

the set of solutions for F. The following result is then the main structural result about the set of solutions to a system of linear homogeneous ordinary differential equations.

3.2.6 Theorem (Vector space structure of sets of solutions) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (2.1) and suppose that the map* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ *is continuous. Then* Sol(F) *is an* n*-dimensional subspace of* $\mathsf{C}^1(\mathbb{T};\mathbb{R})$.

Proof Fix $t_0 \in \mathbb{T}$ and define

$$\begin{aligned}\sigma_{t_0}\colon \mathrm{Sol}(F) &\to \mathsf{V}\\ \xi &\mapsto \xi(t_0).\end{aligned}$$

We claim that σ_{t_0} is an isomorphism of vector spaces. First, the verification of the linearity of σ_{t_0} follows from the equalities

$$(\xi_1 + \xi_2)(t_0) = \xi_1(t_0) + \xi_2(t_0), \quad (a\xi)(t_0) = a(\xi(t_0)),$$

which themselves follow from the definition of the vector space structure in $\mathsf{C}^1(\mathbb{T};\mathsf{V})$. Next let us show that σ_{t_0} is injective by showing that $\ker(\sigma_{t_0}) = \{0\}$. Indeed, suppose that $\sigma_{t_0}(\xi) = 0$. Then, by the uniqueness assertion of Proposition 3.2.4, it follows that $\xi(t) = 0$ for every $t \in \mathbb{T}$, as desired. To show that σ_{t_0} is surjective, let $x_0 \in \mathsf{V}$. Then, by the existence assertion of Proposition 3.2.4, there exists $\xi \in \mathrm{Sol}(F)$ such that $\xi(t_0) = x_0$, i.e., such that $\sigma_{t_0}(\xi) = x_0$. ■

The following corollary, immediate from the proof of the theorem, gives an easy check on the linear independence of subsets of Sol(F).

3.2.7 Corollary (Linear independence in Sol(F)) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (2.1) and suppose that the map* $\mathrm{A}\colon \mathbb{T} \to \mathbb{R}$ *is continuous. Then a subset* $\{\xi_1, \dots, \xi_k\} \subseteq \mathrm{Sol(F)}$ *is linearly independent if and only if, for some* $t_0 \in \mathbb{T}$*, the subset* $\{\xi_1(t_0), \dots, \xi_k(t_0)\} \subseteq \mathsf{V}$ *is linearly independent.*

As with scalar linear homogeneous ordinary differential equations, the theorem allows us to give a special name to a basis for Sol(F).

3.2.8 Definition (Fundamental set of solutions) Consider the system of linear homogeneous ordinary differential equations F in the n-dimensional $\mathbb{R}$-vector space V with right-hand side (2.1) and suppose that the map $A\colon \mathbb{T} \to \mathbb{R}$ is continuous. A set $\{\xi_1, \dots, \xi_n\}$ of linearly independent elements of Sol(F) is a ***fundamental set of solutions*** for F. •

3.2.2.2 The state transition map We now present a particular way of organising a fundamental set of solutions into one object that, for all intents and purposes, completely characterises Sol(F). This we organise as the following theorem.

3.2.9 Theorem (Existence of, and properties of, the state transition map) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (2.1) and suppose that the map* $A\colon \mathbb{T} \to \mathbb{R}$ *is continuous. Then there exists a unique map* $\Phi_A\colon \mathbb{T}\times\mathbb{T} \to \mathsf{V}$ *with the following properties:*

(i) for each $t_0 \in \mathbb{T}$*, the function*

$$\begin{aligned}\Phi_{A,t_0}\colon \mathbb{T} &\to \mathrm{L}(\mathsf{V};\mathsf{V})\\ t &\mapsto \Phi_A(t,t_0)\end{aligned}$$

is differentiable and satisfies the initial value problem

$$\dot{\Phi}_{A,t_0}(t) = A(t)\circ\Phi_{A,t_0}(t), \quad \Phi_{A,t_0}(t_0) = \mathrm{id}_{\mathsf{V}};$$

(ii) the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x_0$$

is $t \mapsto \Phi_A(t,t_0)(x_0)$*;*

(iii) $\det(\Phi(t,t_0)) = e^{\int_{t_0}^{t}\operatorname{tr}(A(s))\,ds}$ *(the* ***Abel–Jacobi–Liouville formula****);*

(iv) for $t, t_0, t_1 \in \mathbb{T}$, $\Phi_A(t,t_0) = \Phi_A(t,t_1)\circ\Phi_A(t_1,t_0)$*;*

(v) for each $t, t_0 \in \mathbb{T}$, $\Phi_A(t,t_0)$ *is invertible and* $\Phi_A(t,t_0)^{-1} = \Phi(t_0,t)$.

Proof First of all, we define Φ by the condition in part (i). That is to say, we define Φ by

$$\frac{\partial\Phi_A}{\partial t}(t,t_0) = A(t)\circ\Phi_A(t,t_0), \quad \Phi_A(t_0,t_0) = \mathrm{id}_{\mathsf{V}}\,.$$

Note that this is an initial value problem associated with the system of linear homogeneous ordinary differential equations F_A in $\mathrm{L}(\mathsf{V};\mathsf{V})$ with right-hand side

$$\begin{aligned}\widehat{F_A}\colon \mathbb{T}\times\mathrm{L}(\mathsf{V};\mathsf{V}) &\to \mathrm{L}(\mathsf{V};\mathsf{V})\\ \Phi &\mapsto A(t)\circ\Phi;\end{aligned}$$

note the mapping $\Phi \mapsto A(t)\circ\Phi$ is linear.[4] Thus, by Proposition 3.2.4, it possesses

[4]The general setting here is this. Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and let U, V, and W be $\mathbb{F}$-vector spaces. Given $L \in \mathrm{L}(\mathsf{V};\mathsf{W})$, define a map ("composition with L") by

$$\begin{aligned}C_L\colon \mathrm{L}(\mathsf{U};\mathsf{V}) &\to \mathrm{L}(\mathsf{U};\mathsf{W})\\ M &\mapsto L\circ M.\end{aligned}$$

It is then straightforward to verify that this is a linear map:

$$\begin{aligned}C_L(M_1+M_2)(u) &= L\circ(M_1+M_2)(u) = L(M_1(u)+M_2(u))\\ &= L\circ M_1(u) + L\circ M_2(u) = C_L(M_1)(u) + C_L(M_2)(u)\end{aligned}$$

which implies that $C_L(M_1+M_2) = C_L(M_1) + C_L(M_2)$, and

$$C_L(aM)(u) = L\circ(aM)(u) = L(a(M(u))) = aL\circ M(u) = aC_L(M)(u),$$

which implies that $C_L(aM) = aC_L(M)$.

a unique solution which, by Proposition 3.2.5, exists in all of $\mathbb{T}$. This proves the existence and uniqueness and part (i).

(ii) We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t,t_0)(x_0) = \frac{\partial\Phi_A}{\partial t}(t,t_0)(x_0) = A(t)\circ\Phi_A(t,t_0)(x_0)$$

and $\Phi_A(t_0,t_0)(x_0) = x_0$, which shows that $t \mapsto \Phi_A(t,x_0)(x_0)$ solves the stated initial value problem. By uniqueness of such solutions, this part of the theorem follows.

(iii) We start with a lemma.

1 Lemma *Let $\mathbb{T} \subseteq \mathbb{R}$ be an interval and let $\mathbf{A}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ be a differentiable map. For $j,k \in \{1,\dots,n\}$, let $C_{jk}(t)$ be the (j,k)th cofactor of $\mathbf{A}(t)$, i.e., $(-1)^{j+k}$ times the determinant of the $(n-1)\times(n-1)$ matrix formed by deleting the jth row and kth column from $\mathbf{A}(t)$. Then*

$$\frac{\mathrm{d}(\det\mathbf{A})}{\mathrm{d}t}(t) = \sum_{j,k=1}^{n} C_{jk}(t)\dot{A}_{jk}(t).$$

Proof The row/column expansion rule for determinants gives

$$\det \boldsymbol{A}(t) = \sum_{k=1}^{n} A_{jk}(t)C_{jk}(t)$$

for any $j \in \{1,\dots,n\}$. Using the Chain Rule,

$$\frac{\mathrm{d}(\det \boldsymbol{A})}{\mathrm{d}t}(t) = \sum_{j,k=1}^{n} \frac{\partial(\det \boldsymbol{A})}{\partial A_j k}\dot{A}_{jk}(t) = \sum_{j,k=1}^{n} C_{jk}(t)\dot{A}_{jk}(t),$$

because C_{jk} does not depend on the (j,k)th component of $\boldsymbol{A}$. ▼

We choose a basis $\{e_1,\dots,e_n\}$ for V and denote by $\boldsymbol{A}(t)$ the matrix representative of $A(t)$ and by $\boldsymbol{\Phi}_A(t,t_0)$ the matrix representative of $\Phi_A(t,t_0)$. (That we can reduce to $\mathsf{V} = \mathbb{R}^n$ is justified by Exercises 3.2.3 and 3.2.4.) For $j,k \in \{1,\dots,n\}$, denote by $C_{jk}(t,t_0)$ the (j,k)th cofactor of $\boldsymbol{\Phi}_A(t,t_0)$, i.e., $(-1)^{j+k}$ times the determinant of the matrix $\boldsymbol{\Phi}_A(t,t_0)$ with the jth row and kth column removed. Also let $\boldsymbol{C}(t,t_0)$ be the matrix formed from these cofactors. Denote by $\Phi_{jk}(t,t_0)$ the (j,k)th component of $\boldsymbol{\Phi}_A$. Using the lemma,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\det\boldsymbol{\Phi}_A(t,t_0) &= \sum_{j,k=1}^{n} C_{jk}(t,t_0)\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{jk}(t,t_0)\\ &= \operatorname{tr}\left(\boldsymbol{C}(t,t_0)^T\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Phi}_A(t,t_0)\right)\\ &= \operatorname{tr}(\boldsymbol{\Phi}_A(t,t_0)\boldsymbol{C}(t,t_0)^T\boldsymbol{A}(t)),\end{aligned}$$

using part (i), the definition of trace and transpose, and the easily verified fact that $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$ for $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Now we note that

$$\boldsymbol{\Phi}_A \boldsymbol{C}(t, t_0)^T = \det \Phi_A \boldsymbol{I}_n$$

using Cramer's Rule for matrix inversion. Thus we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t} \det \boldsymbol{\Phi}_A(t, t_0) = \det \boldsymbol{\Phi}_A(t, t_0) \boldsymbol{A}(t).$$

This equation is a first-order scalar linear homogeneous ordinary differential equation, and we have seen how to solve these in Example 2.2.5. Applying the computations there to the present equation, and using the fact that $\det \boldsymbol{\Phi}_A(t, t_0) = \det \boldsymbol{I}_n = 1$, we get this part of the theorem.

(iv) We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_A(t, t_1) \circ \Phi_A(t_1, t_0)) = A(t) \circ \Phi_A(t, t_0) \circ \Phi_A(t_1, t_0)$$

and

$$\Phi_A(t_1, t_1) \circ \Phi_A(t_1, t_0) = \Phi(t_1, t_0).$$

We also have

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi_A(t, t_0) = A(t) \circ \Phi_A(t, t_0).$$

That is to say, both $t \mapsto \Phi_A(t, t_0)$ and $t \mapsto \Phi_A(t, t_1) \circ \Phi_A(t_1, t_0)$ satisfy the initial problem

$$\dot{\Phi}(t) = A(t) \circ \Phi(t), \quad \Phi(t_1) = \Phi_A(t_1, t_0).$$

By uniqueness of solutions for systems of linear homogeneous ordinary differential equations, we conclude that $\Phi_A(t, t_0) = \Phi_A(t, t_1) \circ \Phi_A(t_1, t_0)$, as desired.

(v) The invertibility of $\Phi(t, t_0)$ follows from part (iii). The specific formula for the inverse follows from the formula

$$\mathrm{id}_{\mathsf{V}} = \Phi_A(t_0, t_0) = \Phi(t_0, t) \circ \Phi(t, t_0),$$

which itself follows from part (iv). ■

Let us formally name the mapping Φ_A defined in the theorem.

3.2.10 Definition Consider the system of linear homogeneous ordinary differential equations F in the n-dimensional $\mathbb{R}$-vector space V with right-hand side (2.1) and suppose that the map $A \colon \mathbb{T} \to \mathbb{R}$ is continuous. The map $\Phi_A \colon \mathbb{T} \times \mathbb{T} \to \mathsf{V}$ from Theorem 3.2.9 is the ***state transition map***. •

One imagines that it is possible to compute the state transition map if one is given a fundamental set of solutions. The following procedure gives an explicit means of doing this.

3.2.11 Procedure (Determining the state transition map from a fundamental set of solutions) Given a system of linear homogeneous ordinary differential equations F in the n-dimensional $\mathbb{R}$-vector space V with right-hand side equation

$$\widehat{F}(t, x) = A(t)(x),$$

with map $A\colon \mathbb{T} \to \mathbb{R}$ is continuous, and given a fundamental set of solutions $\{\xi_1, \dots, \xi_n\}$, do the following.

1. Choose a basis $\{e_1, \dots, e_n\}$.
2. Let $\boldsymbol{\xi}_j\colon \mathbb{T} \to \mathbb{R}^n$ be the components of ξ_j, $j \in \{1, \dots, n\}$, i.e.,

   $$\xi_j(t) = \xi_{1,j}(t)e_1 + \cdots + \xi_{j,n}(t)e_n.$$

   If $\mathsf{V} = \mathbb{R}^n$, one can just take the components of ξ_j, $j \in \{1, \dots, n\}$, in the standard basis, as usual.
3. Assemble the matrix function $\boldsymbol{\Xi}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ by making the components of $\boldsymbol{\xi}_1(t), \dots, \boldsymbol{\xi}_j(t)$ the columns of $\boldsymbol{\Xi}(t)$:

   $$\boldsymbol{\Xi}(t) = \begin{bmatrix} \xi_{1,1}(t) & \xi_{2,1}(t) & \cdots & \xi_{n,1}(t) \\ \xi_{1,2}(t) & \xi_{2,2}(t) & \cdots & \xi_{n,2}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{1,n}(t) & \xi_{2,n}(t) & \cdots & \xi_{n,n}(t) \end{bmatrix}.$$

   (Be sure you understand that $\xi_{j,k}(t)$ is the kth component of $\boldsymbol{\xi}_j(t)$.) We call the matrix-valued function $\boldsymbol{\Xi}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ a ***fundamental matrix*** for F.
4. Define $\boldsymbol{\Phi}(t, t_0) = \Xi(t)\Xi(t_0)^{-1}$.
5. Then $\boldsymbol{\Phi}(t, t_0)$ is the matrix representative of $\Phi_A(t, t_0)$ in the basis $\{e_1, \dots, e_n\}$. •

Let us verify that the preceding procedure does indeed yield the state transition map.

3.2.12 Proposition (Determining the state transition map from a fundamental set of solutions) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (2.1) and suppose that the map* $\mathrm{A}\colon \mathbb{T} \to \mathbb{R}$ *is continuous. Then Procedure 3.2.11 will produce the state transition map.*

Proof By choosing a basis $\{e_1, \dots, e_n\}$ as in Procedure 3.2.11, we can assume that $\mathsf{V} = \mathbb{R}^n$. (This is legitimate by virtue of Exercises Exercise 3.2.3 and 3.2.4.) Let us denote by $\boldsymbol{A}(t)$ the matrix representative of $A(t)$. Defining $\boldsymbol{\Phi}(t, t_0)$ as in the given procedure, we have

$$\frac{\partial \boldsymbol{\Phi}}{\partial t}(t, t_0) = \dot{\boldsymbol{\Xi}}(t)\boldsymbol{\Xi}(t_0)^{-1}.$$

Noting that each of $\boldsymbol{\xi}_j$, $j \in \{1, \dots, n\}$, is a solution for F, we have

$$\dot{\xi}_{j,k}(t) = \sum_{l=1}^{n} A_{kl}(t)\xi_{j,l}(t), \qquad j \in \{1, \dots, n\},\ t \in \mathbb{T}.$$

Therefore, in matrix notation,

$$\begin{bmatrix} \dot{\xi}_1(t) \,\big|\, \cdots \,\big|\, \dot{\xi}(t) \end{bmatrix} = A(t) \begin{bmatrix} \xi_1(t) \,\big|\, \cdots \,\big|\, \xi(t) \end{bmatrix} \implies \dot{\Xi}(t) = A(t)\Xi(t), \qquad t \in \mathbb{T}.$$

Therefore,

$$\frac{\partial \Phi}{\partial t}(t, t_0) = A(t)\Xi(t)\Xi(t_0)^{-1} = A(t)\Phi(t, t_0).$$

Moreover, $\Phi(t_0, t_0) = I_n$. This $t \mapsto \Phi(t, t_0)$ satisfies the matrix representative of the initial value problem satisfied by $t \mapsto \Phi_A(t, t_0)$, i.e., $\Phi(t, t_0)$ is the matrix representative of $\Phi_A(t, t_0)$. ■

In general, it cannot be expected to find the state transition map for a system of linear homogeneous ordinary differential equations. However, to illustrate Procedure 3.2.11, let us give a "cooked" example.

3.2.13 Example (Computing the state transition map) We take the system of linear homogeneous ordinary differential equations F in $\mathbb{R}^2$ with right-hand side

$$\begin{aligned} \widehat{F}\colon (0, \infty) \times \mathbb{R}^2 &\to \mathbb{R}^2 \\ (t, (x_1, x_2)) &\mapsto \left(\frac{1}{t} x_1 - x_2, \frac{1}{t^2} x_1 + \frac{2}{t} x_2 \right). \end{aligned}$$

Solutions $t \mapsto (x_1(t), x_2(t))$ satisfy

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{t} & -1 \\ \frac{1}{t^2} & \frac{2}{t} \end{bmatrix}}_{A(t)} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}.$$

A direct verification shows that the functions $\xi_1, \xi_2\colon (0, \infty) \to \mathbb{R}^2$ defined by

$$\xi_1 = (t^2, -t), \quad \xi_2(t) = (-t^2 \ln(t), t + t \ln(t))$$

are solutions of F. To verify that these are linearly independent we compute

$$\det \begin{bmatrix} t^2 & -t^2 \ln(t) \\ -t & t + t \ln(t) \end{bmatrix} = t^3.$$

As this determinant is nowhere zero, we conclude the desired linear independence.

Now we determine the state transition map in this case. In the notation of Procedure 3.2.11, we have

$$\Xi(t) = \begin{bmatrix} t^2 & -t^2 \ln(t) \\ -t & t + t \ln(t) \end{bmatrix},$$

and then a tedious computation gives

$$\Phi_A(t, t_0) = \Xi(t)\Xi(t_0)^{-1} = \begin{bmatrix} -\frac{t^2(\ln(t/t_0) - 1)}{t_0^2} & -\frac{t^2 \ln(t/t_0)}{t_0} \\ \frac{t \ln(t/t_0)}{t_0^2} & \frac{t(\ln(t/t_0) + 1)}{t_0} \end{bmatrix}.$$

•

3.2.2.3 The Peano–Baker series In this section we will provide a series representation for the state transition map for a system of linear ordinary differential equations. This is presented for two reasons: (1) as an illustration of series methods in ordinary differential equations, as these arise in many important contexts; (2) as an illustration, in an elementary setting, of iterative procedure used in the proof of Theorem 1.4.8. It is by no means being suggested that the series representation we give for the state transition map is useful for computation.

We let $\mathbb{T} \subseteq \mathbb{R}$ be an interval and let $A \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ be continuous. By its definition, the state transition map $(t, t_0) \mapsto \Phi_A(t, t_0)$ is determined from the initial value problem

$$\dot{\Phi}(t) = A(t) \circ \Phi(t), \quad \Phi(t_0) = \mathrm{id}_{\mathsf{V}}.$$

Let us fix $t, t_0 \in \mathbb{T}$ and take $t > t_0$, for concreteness. By the Fundamental Theorem of Calculus (assuming, as we are, that $t \mapsto A(t)$ is continuous), this is equivalent to

$$\Phi(t) = \mathrm{id}_{\mathsf{V}} + \int_{t_0}^{t} A(\tau) \circ \Phi(s) \, \mathrm{d}s. \tag{3.5}$$

Let us informally iterate to find a solution. We define $\Phi_0 \colon [t_0, t] \to \mathrm{L}(\mathsf{V};)$ by $\Phi_0(\tau) = \mathrm{id}_{\mathsf{V}}$. This will, generally, not satisfy the integral equation (3.5). So, let us substitute this zeroth-order approximation into the same integral equation to get (hopefully) a better approximation $\Phi_1 \colon [t_0, t] \to \mathrm{L}(\mathsf{V}; \mathsf{V})$:

$$\Phi_1(\tau) = \Phi_0(\tau) + \int_{t_0}^{t} A(\tau) \circ \Phi_0(\tau) \, \mathrm{d}\tau = \mathrm{id}_{\mathsf{V}} + \int_{t_0}^{t} A(\tau) \, \mathrm{d}\tau.$$

We now continue this process iteratively, assuming that, if we have defined $\Phi_k \colon [t_0, t] \to \mathrm{L}(\mathsf{V}; \mathsf{V})$, we define $\Phi_{k+1} \colon [t_0, t] \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ by

$$\Phi_{k+1}(\tau) = \Phi_k(\tau) + \int_{t_0}^{t} A(\tau) \circ \Phi_k(\tau) \, \mathrm{d}\tau.$$

It is pretty clear that

$$\Phi_k(t) - \Phi_{k-1}(t) = \underbrace{\int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{k-1}} A(t_1) \circ A(t_2) \circ \cdots \circ A(t_k) \, \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1}_{I_k(t, t_0)}.$$

Thus we can make the following definition.

3.2.14 Definition (Peano–Baker series) For an interval $\mathbb{T} \subseteq \mathbb{R}$, for $t_0 \in \mathbb{T}$, and for a continuous map $A \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$, the series

$$I_\infty(t, t_0) = \mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} I_k(t, t_0)$$

is the $\mathbf{t_0}$***-Peano–Baker series*** for A. •

Of course, the definition is quite meaningless without addressing whether the series converges. The main result of this section is now the following.

3.2.15 Theorem (Convergence of the Peano–Baker series) *Let* V *be a finite-dimensional* $\mathbb{R}$*-vector space. For an interval* $\mathbb{T} \subseteq \mathbb{R}$*, for* $\mathrm{t}_0 \in \mathbb{T}$*, and for a continuous map* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$*, the* t_0*-Peano–Baker series converges uniformly on every compact subinterval of* $\mathbb{T}$*, and, moreover,* $\mathrm{I}_\infty(\mathrm{t},\mathrm{t}_0) = \Phi_{\mathrm{A}}(\mathrm{t},\mathrm{t}_0)$.

Proof Let $T_+ > t_0$. We will show that the t_0-Peano–Baker series converges uniformly to $t \mapsto \Phi_A(t, t_0)$ on $[t_0, T_+]$. A similar proof can be concocted for $T_- < t_0$. Then, given a compact subinterval $\mathbb{T}' \subseteq \mathbb{T}$. the theorem follows by taking T_- and T_+ such that $\mathbb{T}' \subseteq [T_-, T_+]$.

We let $\{e_1, \dots, e_n\}$ be a basis for V. We let $\boldsymbol{A}(t)$ be the matrix representative of $A(t)$ and let $\boldsymbol{I}_k(t, t_0)$ be the matrix representative for $I_k(t, t_0)$. Note that, because the matrix representation for a composition of linear maps is the product of the matrix representations, we have

$$\int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{k-1}} \boldsymbol{A}(t_1)\boldsymbol{A}(t_2) \cdots \boldsymbol{A}(t_k) \, \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1.$$

For $\boldsymbol{B} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ let us define

$$\|\boldsymbol{B}\| = \left(\sum_{j,k=1}^{n} |B_{jk}|^2 \right)^{1/2}.$$

We claim that

$$\|\boldsymbol{BC}\| \leq \|\boldsymbol{B}\|\|C\|. \tag{3.6}$$

Let us denote by $\boldsymbol{c}_j(\boldsymbol{B})$ the jth column of $\boldsymbol{B}$. In this case

$$\|\boldsymbol{B}\| = \left(\sum_{j=1}^{n} \|\boldsymbol{c}_j(\boldsymbol{B})\|^2 \right)^{1/2}.$$

Now we can verify that

$$\begin{aligned}
\|\boldsymbol{BC}\| &= \left(\sum_{j=1}^{n} \|\boldsymbol{c}_j(\boldsymbol{BC})\|^2 \right)^{1/2} = \left(\sum_{j=1}^{n} \|\boldsymbol{B}\boldsymbol{c}_j(\boldsymbol{C})\|^2 \right)^{1/2} \\
&\leq \left(\sum_{j=1}^{n} \|\boldsymbol{B}\|^2 \|\boldsymbol{c}_j(\boldsymbol{C})\|^2 \right)^{1/2} \leq \|\boldsymbol{B}\| \left(\sum_{j=1}^{n} \|\boldsymbol{c}_j(\boldsymbol{C})\|^2 \right)^{1/2} \\
&= \|\boldsymbol{B}\|\|\boldsymbol{C}\|.
\end{aligned}$$

We also use the equality

$$\int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{k-1}} \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1 = \frac{(t - t_0)^k}{k!}, \qquad k \in \mathbb{Z}_{>0}. \tag{3.7}$$

This we prove by induction on k. For $k = 1$ it is certainly true. So suppose it true for $k = m$ and then compute

$$\int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_m} \mathrm{d}t_{m+1} \cdots \mathrm{d}t_2 \mathrm{d}t_1 = \int_{t_0}^{t} \frac{(t_1 - t_0)^m}{m!} \mathrm{d}t_1 = \frac{(t - t_0)^{m+1}}{(m+1)!},$$

as desired.

Now let

$$M = \sup\{\|\boldsymbol{A}(\tau)\| \mid \tau \in [t_0, T_+]\}.$$

Let $\epsilon \in \mathbb{R}_{>0}$. Since the series of numbers

$$\sum_{k=0}^{\infty} \frac{M^k (T_+ - t_0)^k}{k!}$$

converges (it is equal to $\mathrm{e}^{M(T_+ - t_0)}$), there exists $N \in \mathbb{Z}_{>0}$ such that, if $r, s \geq N$ with $r > s$,

$$\sum_{k=s+1}^{r} \frac{M^k (T_+ - t_0)^k}{k!} < \epsilon.$$

Therefore, for $r, s \geq N$ with $r > s$, we have

$$\begin{aligned}
\left\| \sum_{k=1}^{r} \boldsymbol{I}_k(t, t_0) - \sum_{k=1}^{s} \boldsymbol{I}_k(t, t_0) \right\| &\leq \sum_{k=s+1}^{r} \|\boldsymbol{I}_k(t, t_0)\| \\
&\leq \sum_{k=s+1}^{r} \int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{k-1}} \|\boldsymbol{A}(t_1)\boldsymbol{A}(t_2) \cdots \boldsymbol{A}(t_k)\| \, \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1 \\
&\leq \sum_{k=s+1}^{r} \int_{t_0}^{t} \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{k-1}} \|\boldsymbol{A}(t_1)\| \, \|\boldsymbol{A}(t_2)\| \cdots \|\boldsymbol{A}(t_k)\| \, \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1 \\
&\leq \sum_{k=s+1}^{r} M^k \left(\int_{t_0}^{t} \mathrm{d}\tau \right)^k \leq \sum_{k=s+1}^{r} \frac{M^k (T_+ - t_0)^k}{k!} < \epsilon
\end{aligned}$$

using (3.6) and (3.7). This shows that the sequence of functions

$$t \mapsto \mathrm{id}_V + \sum_{k=1}^{m} \boldsymbol{I}_k(t, t_0), \qquad m \in \mathbb{Z}_{>0},$$

is uniformly Cauchy, and so uniformly convergent.***missing stuff***

Finally, we show that $I_\infty(t, t_0) = \Phi_A(t, t_0)$. By the Fundamental Theorem of Calculus, the function

$$t \mapsto \boldsymbol{I}_k(t, t_0)$$

is of class C^1. Moreover, a direct calculation using the definitions gives

$$\dot{\boldsymbol{I}}_{k+1}(t, t_0) = \boldsymbol{A}(t)\boldsymbol{I}_k(t, t_0), \qquad k \in \mathbb{Z}_{>0}.$$

Therefore, the series

$$\sum_{k=1}^{\infty} \dot{\boldsymbol{I}}_k(t, t_0) = \boldsymbol{A}(t) \sum_{k=1}^{\infty} \boldsymbol{I}_{k-1}(t, t_0),$$

with the convention that $\boldsymbol{I}_0(t, t_0) = \boldsymbol{I}_n$, converges uniformly. Thus the series of term-by-term derivatives converges uniformly, and so term-by-term differentiation of $\boldsymbol{I}_\infty(t, t_0)$ is permissible. Moreover,

$$\dot{\boldsymbol{I}}_\infty(t, t_0) = \boldsymbol{A}(t)\boldsymbol{I}_\infty(t, t_0)$$

and $\boldsymbol{I}_\infty(t_0, t_0) = \boldsymbol{I}_n$. Thus the matrix representative of $t \mapsto I_\infty(t, t_0)$ satisfies the same initial value problem as $t \mapsto \Phi_A(t, t_0)$, and the uniqueness assertion of Proposition 3.2.4 gives the result, at least for matrix representatives. That the conclusion also holds in V is a consequence of Exercise 3.2.4. ■

3.2.2.4 The adjoint equation In this section we consider a system of linear ordinary differential equations related to a given one. It is, in a very precise sense, dual to the original equation. So we start with what exactly "dual" means.

3.2.16 Definition (Dual of a vector space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. The ***dual***[5] of V is the $\mathbb{F}$-vector space $\mathsf{V}^* = \mathrm{L}(\mathsf{V}; \mathbb{F})$. •

Let us consider the notion of duality in a simple setting, indeed one where the notion of duality is so simple it is confusing.

3.2.17 Example ($(\mathbb{F}^n)^*$) We consider the n-dimensional $\mathbb{F}$-vector space $\mathbb{F}^n$. Thus, by definition and by the usual conflation of linear maps with matrices, $(\mathbb{F}^n)^*$ is *exactly* the set of $1 \times n$ matrices. Thus we can *represent* an element $\boldsymbol{\alpha} \in (\mathbb{R}^n)^*$ by

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix}$$

for $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Now, we often write elements of $\mathbb{F}^n$ as columns, i.e., as

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Note that $\boldsymbol{v}$ is not "equal" to this $n \times 1$ matrix, we merely use this $n \times 1$ matrix to *represent* $\boldsymbol{v}$ for the purposes of doing matrix-vector multiplication. (Of course, in

[5]There are many places where one wishes to refine the notion of "dual" we give here to include some form of continuity. Here we will only work with finite-dimensional vector spaces where such nuances do not materialise.

the very literal sense, $v = (v_1, v_2, \dots, v_n)$.) Therefore, for example, the product of the $1 \times n$ matrix representing $\boldsymbol{\alpha}$ and the $n \times 1$ matrix representing $\boldsymbol{v}$ is

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

This is merely the matrix/vector multiplication representation of what one would write as $\boldsymbol{\alpha}(\boldsymbol{v})$, thinking of elements of $(\mathbb{R}^n)^*$ as what they are: *linear functions on* V. •

The upshot of the preceding example is: the dual of an n-dimensional $\mathbb{F}$-vector space is also an n-dimensional $\mathbb{F}$-vector space. However, it is definitely not the case that $\mathsf{V}^* = \mathsf{V}$. We shall use the notation "$\alpha(v)$," "$\alpha \cdot v$," or "$\langle \alpha; v \rangle$" to denote the same thing.

Next we see how linear maps behave relative to duality.

3.2.18 Definition (Dual of a linear map) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let U and V be $\mathbb{F}$-vector spaces. The ***dual*** of $L \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ is the linear map $L^* \colon \mathsf{V}^* \to \mathsf{U}^*$ defined by

$$\langle L^*(\beta); u \rangle = \langle \beta; L(u) \rangle,$$

for $u \in \mathsf{U}$ and $\beta \in \mathsf{V}^*$. •

Note that this *does* define L^* since, for each $\beta \in \mathsf{V}^*$, it tells us what $L(\beta)$ does to $u \in \mathsf{U}$.

Let us work with our simple example above to understand the dual of a linear map.

3.2.19 Example (Dual of a linear map between Euclidean space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $\boldsymbol{A} \in \mathrm{L}(\mathbb{F}^m; \mathbb{F}^n)$. Thus, in the usual way, $\boldsymbol{A}$ is represented by a matrix

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix}.$$

Let $\boldsymbol{\beta} \in (\mathbb{F}^n)^*$ and $\boldsymbol{u} \in \mathbb{F}^m$. The relation

$$\langle \boldsymbol{A}^*(\boldsymbol{\beta}); \boldsymbol{u} \rangle = \langle \boldsymbol{\beta}; \boldsymbol{A}(\boldsymbol{u}) \rangle,$$

when expressed in matrix/vector notation, reads

$$(\boldsymbol{A}^*(\boldsymbol{\beta}))\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{A}\boldsymbol{u},$$

from which we conclude that $\boldsymbol{A}^*(\boldsymbol{\beta}) = \boldsymbol{\beta}\boldsymbol{A}$. That is, $\boldsymbol{A}^*$ is still a matrix, but multiplication is on the left. If we do something unnatural, we can write elements of $(\mathbb{F}^n)^*$ as columns by transposing them. In this case

$$(\boldsymbol{\beta}\boldsymbol{A})^T = \boldsymbol{A}^T\boldsymbol{\beta}^T.$$

In this way, we can think of $\boldsymbol{A}^* \in (\mathbb{F}^n)^*$ as being the transpose of $\boldsymbol{A}$. Indeed, sometimes $\boldsymbol{A}^*$ is called the "transpose" of $\boldsymbol{A}$. •

With all of the above as backdrop, we can now define the adjoint equation.

3.2.20 Definition (Adjoint of a system of linear homogeneous ordinary differential equations) Consider the system of linear homogeneous ordinary differential equations F in the n-dimensional $\mathbb{R}$-vector space V with right-hand side (3.4). The ***adjoint equation*** for F is the system F^* of linear homogeneous ordinary differential equations in V^* with right-hand side

$$\begin{aligned}\widehat{F^*}\colon \mathbb{T} \times \mathsf{V}^* &\to \mathsf{V}^*\\ (t, p) &\mapsto -A^*(t)(p).\end{aligned}$$

•

Thus solutions $t \mapsto p(t)$ for the adjoint equation satisfy

$$\dot{p}(t) = -A^*(t)(p(t)).$$

Let us give the state transition map for the adjoint equation.

3.2.21 Proposition (State transition map for the adjoint equation) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (3.4) and suppose that* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *is continuous. Then* $\mathrm{A}^*\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}^*; \mathsf{V}^*)$ *is continuous and the state transition map for the adjoint equation is defined by* $\Phi_{-\mathrm{A}^*}(\mathrm{t}, \mathrm{t}_0) = \Phi_{\mathrm{A}}(\mathrm{t}_0, \mathrm{t})^*$ *for* $\mathrm{t}, \mathrm{t}_\mathrm{o} \in \mathbb{T}$.

Proof The continuity of A^* follows from choosing a basis for V so that A becomes the matrix-valued function $\boldsymbol{A}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$. In this case, $A^*(t)$ has the matrix representative $\boldsymbol{A}(t)^T$, which is shows that the matrix representative of A is continuous of and only if the matrix representative of A^* is continuous.

By Theorem 3.2.9(v) we have

$$\Phi_A(t, t_0) \circ \Phi_A(t_0, t) = \mathrm{id}_{\mathsf{V}}.$$

Differentiating this with respect to time we get

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t, t_0) \circ \Phi_A(t_0,) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t, t_0)\right) \circ \Phi_A(t_0, t) + \Phi_A(t, t_0) \circ \left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t_0, t)\right),$$

from which we derive

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t_0, t) &= -\Phi_A(t_0, t) \circ \left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t, t_0)\right) \circ \Phi_A(t_0, t)\\ &= -\Phi_A(t_0, t) \circ A(t) \circ \Phi_A(t, t_0) \circ \Phi_A(t_0, t)\\ &= -\Phi_A(t_0, t) \circ A(t). \qquad (3.8)\end{aligned}$$

Taking the dual of this equation, and using Exercise 3.2.7, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t_0, t)^* = -A^*(t) \circ \Phi_A(t_0, t)^*.$$

Since $\Phi_A^*(t_0, t_0) = \mathrm{id}_{V^*}$, we thus see that $t \mapsto \Phi_A(t, t_0)^*$ satisfies the initial value problem that defines the state transition map for the adjoint equation, and so the uniqueness assertion of Proposition 3.2.4 gives the result. ■

We have not yet addressed the important question, "Why should one care about the adjoint equation?" We convert this question into another question with the following result.

3.2.22 Proposition (A property of the adjoint equation) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (3.4) and suppose that* $A\colon \mathbb{T} \to \mathrm{L}(V; V)$ *is continuous. Let* $t_0 \in \mathbb{T}$, $x_0 \in V$, *and* $p_0 \in V^*$, *and denote* $x(t) = \Phi_A(t, t_0)(x_0)$ *and* $p(t) = \Phi_A(t_0, t)^*(p_0)$. *Then*

$$\langle p(t); x(t)\rangle = \langle p_0; x_0\rangle.$$

Proof We compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle p(t); x(t)\rangle &= \langle \dot{p}(t); x(t)\rangle + \langle p(t); \dot{x}(t)\rangle\\ &= -\langle A^*(t)(p(t)); \Phi_A(t, t_0)(x_0)\rangle + \langle \Phi_A(t_0, t)^*(p_0); A(t)(x(t))\rangle\\ &= -\langle A^*(t) \circ \Phi_A(t_0, t)^*(p_0); \Phi(t, t_0)(x_0)\rangle + \langle \Phi_A(t_0, t)^*(p_0); A(t) \circ \Phi_A(t, t_0)(x_0)\rangle\\ &= 0.\end{aligned}$$

Since the function $t \mapsto \langle p(t); x(t)\rangle$ is of class C^1, it follows that this function is constant. ■

When $\alpha \in V^*$ and $v \in V$ satisfy $\alpha(v) = 0$, we say that α ***annihilates*** v. This is a sort of "orthogonality condition," although it most definitely is not an actual orthogonality condition, there being no inner product in sight. One of the upshots of the preceding result is the following corollary, saying that the adjoint equation preserves the annihilation condition.

3.2.23 Corollary (The geometric meaning of the adjoint equation) *Consider the system of linear homogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (3.4) and suppose that* $A\colon \mathbb{T} \to \mathrm{L}(V; V)$ *is continuous. Let* $t_0 \in \mathbb{T}$, $x_0 \in V$, *and* $p_0 \in V^*$, *and denote* $x(t) = \Phi_A(t, t_0)(x_0)$ *and* $p(t) = \Phi_A(t_0, t)^*(p_0)$. *If* $\langle p_0; x_0\rangle = 0$, *then* $\langle p(t); x(t)\rangle = 0$ *for all* $t \in \mathbb{T}$.

It is this property of the adjoint equation that makes it an important tool in optimal control theory, but this is not a subject into which we shall dwell deeply here.

3.2.3 Equations with constant coefficients

We now consider the special case of systems of linear homogeneous equations with constant coefficients, i.e., those systems of linear ordinary differential equations F in a vector space V with right-hand sides

$$\widehat{F}(t, x) = A(x), \tag{3.9}$$

for $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. As with the scalar version of such equations that we studied in Section 2.2.2, there is a great deal more that we can say about such equations, beyond the general assertions in the preceding section. Indeed, one can say that, in principle, one can "solve" such equations, and we shall present a procedure for doing so.

Before we do so, however, we reiterate that the ordinary differential equations we are considering in this section are special cases of the time-varying equations of the preceding section, so all of the general statements made there apply here as well. In particular, Propositions 3.2.4 and 3.2.5, and Theorem 3.2.6 hold for equations of the form (3.9).

We have already seen in Theorem 3.2.6 that linear algebra plays a rôle in the theory of systems of linear homogeneous ordinary differential equations. We shall see in this section that this rôle is amplified for equations with constant coefficients. Therefore, the next two sections have to do with linear algebra.

3.2.3.1 Invariant subspaces associated with eigenvalues We assume that the reader is familiar with the basic theory of eigenvalues and eigenvectors for linear transformations of finite-dimensional $\mathbb{F}$-vector spaces. In this section and the next, we shall expand this elementary theory into a comprehensive understanding of the invariant subspaces of a linear transformation of a finite-dimensional $\mathbb{R}$-vector space.

One of the complications that arise in the study of eigenvalues is that of multiplicity. We shall see that there are two, generally distinct, notions of multiplicity, and the distinction between these is the source of some beautiful, and somewhat complicated, mathematics.

We begin by associating a specific invariant subspace to a given linear transformation of a vector space. The construction is a little involved, and seems a little pointless at present. However, it will form the essential part of the definition of algebraic multiplicity in Definition 3.2.26. We consider some constructions involving the kernels of powers of an endomorphism. Thus we let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, V be an $\mathbb{F}$-vector space, and $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. We denote

$$L^j = \underbrace{L \circ \cdots \circ L}_{j \text{ times}}, \qquad j \in \mathbb{Z}_{>0},$$

and consider the subspaces $\ker(L^j) \subseteq \mathsf{V}$, $j \in \mathbb{Z}_{>0}$. We denote by U_L the subspace spanned by $\cup_{j\in\mathbb{Z}_{>0}} \ker(L^j)$. Since the sequence

$$\ker(L) \subseteq \ker(L^2) \subseteq \cdots \subseteq \ker(L^j) \subseteq \cdots \tag{3.10}$$

is increasing, in fact we simply have $\mathsf{U}_L = \cup_{j\in\mathbb{Z}_{>0}} \ker(L^j)$. Since the subspace U_L will be essential in our definition of algebraic multiplicity, let us make a few comments on its properties and its computation in practice.

3.2.24 Proposition (Characterisation of U_L) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $\mathrm{L} \in \mathrm{L}(\mathsf{V};\mathsf{V})$. The subspace U_L is the smallest subspace of V with the properties that*

(i) *U_L is L-invariant and*

(ii) *$\ker(\mathrm{L}) \subseteq \mathsf{U}_\mathrm{L}$.*

Proof Let us first prove that U_L has the two properties stated in the proposition. Let $v \in \mathsf{U}_L$ so that $v \in \ker(L^k)$ for some $k \in \mathbb{Z}_{>0}$. Then $L \circ L^k(v) = L^k(L(v)) = 0$, and so $L(v) \in \ker(L^k) \subseteq \mathsf{U}_L$, showing that U_L is L-invariant. It is also clear that $\ker(L) \subseteq \mathsf{U}_L$.

Now we show that U_L is the smallest subspace with the two stated properties. Thus we let U'_L be a subspace with the two properties. We claim that $\ker(L^j) \subseteq \mathsf{U}'_L$ for $j \in \mathbb{Z}_{>0}$. This is clearly true for $j = 1$, so suppose it true for $j \in \{1, \ldots, k\}$ and let $v \in \ker(L^{k+1})$. Thus $L^{k+1}(v) = L^k(L(v)) = 0$. Thus $L(v) \in \ker(L^k) \subseteq \mathsf{U}'_L$ by the induction hypothesis. Therefore, by definition of U_L and since U'_L is a subspace, we have $\mathsf{U}_L \subseteq \mathsf{U}'_L$, which completes the proof. ■

This result has the following corollary which is useful in limiting the computations one must do in practice when computing the algebraic multiplicity. The result says, roughly, that if the sequence

$$\ker(L) \subseteq \ker(L^2) \subseteq \cdots \subseteq \ker(L^j) \subseteq \cdots$$

has two neighbouring terms which are equal, then all remaining terms in the sequence are also equal. This makes the computation of U_L simpler in these cases.

3.2.25 Corollary (Computation of U_L) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $\mathrm{L} \in \mathrm{L}(\mathsf{V};\mathsf{V})$. If, for some $\mathrm{k} \in \mathbb{Z}_{>0}$, $\ker(\mathrm{L}^\mathrm{k}) = \ker(\mathrm{L}^{\mathrm{k}+1})$, then $\ker(\mathrm{L}^\mathrm{j}) = \ker(\mathrm{L}^\mathrm{j})$ for all $\mathrm{j} \geq \mathrm{k}$, and, moreover, $\mathsf{U}_\mathrm{L} = \ker(\mathrm{L}^\mathrm{k})$. Moreover, if V is finite-dimensional, then it will always be the case that $\mathsf{U}_\mathrm{L} = \ker(\mathrm{L}^\mathrm{k})$ for some $\mathrm{k} \in \mathbb{Z}_{>0}$.*

Proof The result will follow from the definition of U_L if we can show that $\mathsf{U}_L = \ker(L^k)$. Since $\ker(L^k) \subseteq \mathsf{U}_L$, this will follow if we can show that $\ker(L^k)$ is L-invariant, since clearly $\ker(L) \subseteq \ker(L^k)$. First let $v \in \ker(L^k)$. Then, since $\ker(L^{k+1}) = \ker(L^k)$, $L^{k+1}(v) = L^k(L(v)) = 0$, showing that $L(v) \in \ker(L^k)$. Thus $\ker(L^k)$ is L-invariant, and the corollary follows. The final assertion of the corollary is merely the statement that subspaces of a finite-dimensional vector space are finite-dimensional. ■

We now use the definition of the subspace U_L above to talk about invariant subspaces associated with eigenvalues. To do so, for $\lambda \in \mathbb{F}$ we denote

$$L_\lambda = \lambda \operatorname{id}_{\mathsf{V}} - L.$$

We have the following obvious facts:

1. λ is an eigenvalue if and only if $\ker(L_\lambda) \neq \{0\}$;
2. eigenvectors are nonzero vectors in $\ker(L_\lambda)$.

We may now characterise the various multiplicities associated with an eigenvalue.

3.2.26 Definition (Eigenspaces, algebraic and geometric multiplicity) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, let $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, and let $\lambda \in \mathbb{F}$ be an eigenvalue for L.

(i) The ***eigenspace*** for λ is the subspace $\mathsf{W}(\lambda, L) = \ker(L_\lambda)$.

(ii) The ***generalised eigenspace*** for λ is the subspace $\overline{\mathsf{W}}(\lambda, L) = \cup_{j \in \mathbb{Z}_{>0}} \ker(L_\lambda^j)$.

(iii) The ***geometric multiplicity*** of λ is $m_{\mathrm{g}}(\lambda, L) = \dim_{\mathbb{F}}(\mathsf{W}(\lambda, L))$.

(iv) The ***algebraic multiplicity*** of λ is $m_{\mathrm{a}}(\lambda, L) = \dim_{\mathbb{F}}(\overline{\mathsf{W}}(\lambda, L))$. •

The definitions immediately lead to the following facts.

3.2.27 Remarks (Properties of geometric and algebraic multiplicity)

1. Note that both the geometric and algebraic multiplicity are nonzero.
2. The algebraic and geometric multiplicities are always finite if V is finite-dimensional.
3. It always holds that $m_{\mathrm{a}}(\lambda, L) \geq m_{\mathrm{g}}(\lambda, L)$. •

It will be useful to know that eigenspaces and generalised eigenspaces are invariant.

3.2.28 Proposition (Invariance of eigenspaces and generalised eigenspaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, let $\mathrm{L} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, and let λ be an eigenvalue for L. Then, for any $\mathrm{j} \in \mathbb{Z}_{>0}$, $\ker(\mathrm{L}_\lambda^{\mathrm{j}})$ is an L-invariant subspace. As a consequence, $\mathsf{W}(\lambda, \mathrm{L})$ and $\overline{\mathsf{W}}(\lambda, \mathrm{L})$ are L-invariant subspaces.*

Proof We first claim that $L \circ L_\lambda^j = L_\lambda^j \circ L$. We prove this by induction. For $j = 1$ we simply have

$$L \circ (\lambda \operatorname{id}_{\mathsf{V}} - L) = \lambda L \circ \operatorname{id}_{\mathsf{V}} - L \circ L = \lambda \operatorname{id}_{\mathsf{V}} \circ L - L \circ L = (\lambda \operatorname{id}_{\mathsf{V}} - L) \circ L.$$

Now suppose the claim true for $j \in \{1, \dots, k\}$ and compute

$$\begin{aligned} L \circ (\lambda \operatorname{id}_{\mathsf{V}} - L)^{k+1} &= L \circ (\lambda \operatorname{id}_{\mathsf{V}} - L) \circ (\lambda \operatorname{id}_{\mathsf{V}} - L)^k = (\lambda \operatorname{id}_{\mathsf{V}} - L) \circ L \circ (\lambda \operatorname{id}_{\mathsf{V}} - L)^k \\ &= (\lambda \operatorname{id}_{\mathsf{V}} - L) \circ (\lambda \operatorname{id}_{\mathsf{V}} - L)^k \circ L = (\lambda \operatorname{id}_{\mathsf{V}} - L)^{k+1} \circ L, \end{aligned}$$

giving our claim.

The first assertion of the proposition now follows easily. If $v \in \ker(L_\lambda^j)$ then we have

$$L_\lambda^j(v) = 0 \quad \Longrightarrow \quad L \circ L_\lambda^j(v) = L_\lambda^j(L(v)) = 0$$

so that $L \in \ker(L_\lambda^j)$. For the second assertion, it immediately follows that $\mathsf{W}(\lambda, L)$ is L-invariant. The L-invariance of $\overline{\mathsf{W}}(\lambda, L)$ follows since, if $v \in \overline{\mathsf{W}}(\lambda, L)$, then $v \in \ker(L_\lambda^j)$ for some $j \in \mathbb{Z}_{>0}$. ■

It is fairly clear that, if λ_1 and λ_2 are distinct eigenvalues for $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, then $\mathsf{W}(\lambda_1, L) \cap \mathsf{W}(\lambda_2, L) = \{0\}$. It is less clear, although still true, that the corresponding statement for the generalised eigenspaces also holds.

3.2.29 Proposition (Intersections of generalised eigenspaces are zero) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $\mathrm{L} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. If λ_1 and λ_2 are distinct eigenvalues for L then $\overline{\mathsf{W}}(\lambda_1, \mathrm{L}) \cap \overline{\mathsf{W}}(\lambda_2, \mathrm{L}) = \{0\}$.*

Proof We first prove a lemma characterising the intersections of generalised eigenspaces.

1 Lemma $\overline{\mathsf{W}}(\lambda_1, \mathrm{L}) \cap \overline{\mathsf{W}}(\lambda_2, \mathrm{L}) = \cup_{\mathrm{j} \in \mathbb{Z}_{>0}}(\ker(\mathrm{L}_{\lambda_1}^{\mathrm{j}}) \cap \ker(\mathrm{L}_{\lambda_2}^{\mathrm{j}}))$.

Proof By definition we have

$$\overline{\mathsf{W}}(\lambda_1, L) \cap \overline{\mathsf{W}}(\lambda_2, L) = \left(\cup_{j \in \mathbb{Z}_{>0}} \ker(L_{\lambda_1}^j)\right) \cap \left(\cup_{k \in \mathbb{Z}_{>0}} \ker(L_{\lambda_2}^k)\right).$$

By standard properties of union and intersection we have

$$\begin{aligned} \overline{\mathsf{W}}(\lambda_1, L) \cap \overline{\mathsf{W}}(\lambda_2, L) &= \cup_{k \in \mathbb{Z}_{>0}} \left(\left(\cup_{j \in \mathbb{Z}_{>0}} \ker(L_{\lambda_1}^j)\right) \cap \ker(L_{\lambda_2}^k)\right) \\ &= \cup_{k \in \mathbb{Z}_{>0}} \left(\cup_{j \in \mathbb{Z}_{>0}} \Big(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^k)\Big)\right) \end{aligned}$$

It is clear that the inclusion

$$\cup_{j \in \mathbb{Z}_{>0}}(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j)) \subseteq \cup_{k \in \mathbb{Z}_{>0}} \Big(\cup_{j \in \mathbb{Z}_{>0}} \Big(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^k)\Big)\Big)$$

holds. If

$$v \in \cup_{k \in \mathbb{Z}_{>0}} \Big(\cup_{j \in \mathbb{Z}_{>0}} \Big(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^k)\Big)\Big),$$

then there exists $j, k \in \mathbb{Z}_{>0}$ such that $v \in \ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^k)$. If $j = k$ then we immediately have

$$v \in \cup_{j \in \mathbb{Z}_{>0}}(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j)).$$

So suppose, without loss of generality, that $j > k$. Then

$$\ker(L_{\lambda_2}^k) \subseteq \ker(L_{\lambda_2}^j),$$

and so we again arrive at

$$v \in \cup_{j \in \mathbb{Z}_{>0}}(\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j)),$$

so giving our claim. ▼

We next claim that $\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j) = \{0\}$ for each $j \in \mathbb{Z}_{>0}$. We prove this by induction on j. For $j = 1$, let $v \in \ker(L_{\lambda_1}) \cap \ker(L_{\lambda_2})$. Then

$$L(v) = \lambda_1 v = \lambda_2 v \quad \Longrightarrow \quad (\lambda_1 - \lambda_2)v = 0 \quad \Longrightarrow \quad v = 0.$$

Now suppose that $\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j) = \{0\}$ for $j \in \{1, \dots, k\}$ and let $v \in \ker(L_{\lambda_1}^{k+1}) \cap \ker(L_{\lambda_2}^{k+1})$. Then

$$L_{\lambda_1}^{k+1}(v) = L_{\lambda_2}^{k+1}(v) = 0.$$

This means that $L_{\lambda_1}(v) \in \ker(L_{\lambda_1}^k)$ and $L_{\lambda_2}(v) \in \ker(L_{\lambda_2}^k)$.

We now use a lemma.

2 Lemma *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $\mathrm{L}, \mathrm{M} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. Show that, if L and M commute, i.e., $\mathrm{L} \circ \mathrm{M} = \mathrm{M} \circ \mathrm{L}$, then $\ker(\mathrm{L})$ is M-invariant and that $\ker(\mathrm{M})$ is L-invariant.*

Proof Let $v \in \ker(L)$. Then $0 = M \circ L(v) = L \circ M(v)$, i.e., $M(v) \in \ker(L)$. Thus $\ker(L)$ is M-invariant. The other assertion, of course, follows in the same way. ▼

A direct computation shows that L_{λ_1} and L_{λ_2} commute. Thus, by the lemma, $\ker(L_{\lambda_2}^k)$ and $\ker(L_{\lambda_1}^k)$ are invariant under L_{λ_1} and L_{λ_2}, respectively. Thus we have

$$L_{\lambda_1}(L_{\lambda_2}(v)) \in \ker(L_{\lambda_2}^k), \quad L_{\lambda_2}(L_{\lambda_1}(v)) \in \ker(L_{\lambda_1}^k).$$

Therefore, by the induction hypothesis,

$$L_{\lambda_1}(L_{\lambda_2}(v)) = L_{\lambda_2}(L_{\lambda_1}(v)) = 0,$$

since L_{λ_1} and L_{λ_2} commute. Therefore,

$$L_{\lambda_2}(v) \in \ker(L_{\lambda_1}) \subseteq \ker(L_{\lambda_1}^k), \quad L_{\lambda_1}(v) \in \ker(L_{\lambda_2}) \subseteq \ker(L_{\lambda_2}^k).$$

That is, $L_{\lambda_1}(v), L_{\lambda_2}(v) \in \ker(L_{\lambda_1}^k) \cap \ker(L_{\lambda_2}^k)$. Again by the induction hypothesis, this gives $L_{\lambda_1}(v) = 0$ and $L_{\lambda_2}(v) = 0$. Thus $v \in \ker(L_{\lambda_1}) \cap \ker(L_{\lambda_2}) = \{0\}$, so giving our claim that $\ker(L_{\lambda_1}^j) \cap \ker(L_{\lambda_2}^j) = \{0\}$ for each $j \in \mathbb{Z}_{>0}$.

The result now easily follows from this and Lemma 1. ■

Let us give an example that exhibit the character of and relationship between algebraic and geometric multiplicity.

3.2.30 Example (Algebraic and geometric multiplicity) For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ take $\mathsf{V} = \mathbb{F}^3$ and define $L_1, L_2 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ by the two 3×3 matrices

$$L_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

These linear maps both have eigenvalues 0 and -1. We can readily see that

$$\begin{aligned}\ker(0\,\mathrm{id}_V - L_1) &= \operatorname{span}_{\mathbb{F}}((1,0,0),(0,1,0)),\\ \ker(0\,\mathrm{id}_V - L_1) &= \operatorname{span}_{\mathbb{F}}((1,0,0)),\\ \ker(-1\,\mathrm{id}_V - L_1) &= \operatorname{span}_{\mathbb{F}}((0,0,1)),\\ \ker(-1\,\mathrm{id}_V - L_2) &= \operatorname{span}_{\mathbb{F}}((0,0,1)).\end{aligned}$$

From this we deduce that for L_1, $m_g(0, L_1) = 2$ and $m_g(-1, L_1) = 1$, and that for L_2, $m_g(0, L_2) = 1$ and $m_g(-1, L_2) = 1$. To compute the algebraic multiplicities, we must compute the powers of the matrices $\lambda\,\mathrm{id}_V - L$ where λ runs over the eigenvalues, and L is either L_1 or L_2. For this purpose it is sufficient to compute

$$\begin{aligned}\dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_1)) &= 2, & \dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_2)) &= 1,\\ \dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_1)^2) &= 2, & \dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_2)^2) &= 2,\\ \dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_1)^3) &= 2, & \dim_{\mathbb{F}}(\ker(0\,\mathrm{id}_V - L_2)^3) &= 2,\\ \dim_{\mathbb{F}}(\ker(-1\,\mathrm{id}_V - L_1)) &= 1, & \dim_{\mathbb{F}}(\ker(-1\,\mathrm{id}_V - L_2)) &= 1,\\ \dim_{\mathbb{F}}(\ker(-1\,\mathrm{id}_V - L_1)^2) &= 1, & \dim_{\mathbb{F}}(\ker(-1\,\mathrm{id}_V - L_2)^2) &= 1.\end{aligned}$$

We then conclude that $m_a(0, L_1) = m_a(0, L_2) = 2$ and $m_a(-1, L_1) = m_a(-1, L) = 1$. •

The definition of the algebraic multiplicity that we give is interesting, because it is geometric. However, it is not very useful in that we do not know *a priori* how far along we need to go in the sequence (3.10) before it terminates. If V is finite-dimensional, it is certainly the case that we will be able to stop after $\dim_{\mathbb{F}}(V)$ terms. But in this case, we can give a *precise* upper bound for the $k \in \mathbb{Z}_{>0}$ for which $\ker(L_\lambda^j) = \ker(L_\lambda^k)$ for all $j \geq k$. A proof of this upper bound requires a deeper understanding of linear transformations than we are willing to undertake just now, so we content ourselves with a mere statement of the required estimate.

3.2.31 Fact (Determining algebraic multiplicity) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be a finite-dimensional $\mathbb{F}$ vector space, let $L \in \mathrm{L}(V; V)$, and let $\lambda \in \mathbb{F}$ be an eigenvalue for L. Let $P_L \in \mathbb{F}[X]$ be the characteristic polynomial for L and let $m(\lambda, P_L)$ be the multiplicity of λ as a root of P_L. Then*

(i) $\ker(L_\lambda^j) = \ker(L_\lambda^{m(\lambda,P_L)})$ *for* $j \geq m(\lambda, P_L)$ *and*

(ii) $m_a(\lambda, L) = m(\lambda, P_L)$.

A corollary of this fact, Proposition 3.2.29, and the Fundamental Theorem of Algebra (that every polynomial over $\mathbb{C}$ has a root) is the following.

3.2.32 Theorem (Decomposition into generalised eigenspaces for $\mathbb{C}$-linear transformations) *Let V be a finite-dimensional $\mathbb{C}$ vector space, let $L \in \mathrm{L}(V; V)$, and let $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ be the distinct eigenvalues for L. Then*

$$V = \overline{W}(\lambda_1, L) \oplus \cdots \oplus \overline{W}(\lambda_k, L),$$

and each of the subspaces $\overline{W}(\lambda_j, L)$, $j \in \{1, \ldots, k\}$, are L-invariant.

Note that the theorem does not hold, in general, for $\mathbb{R}$-vector spaces, since a $\mathbb{R}$-linear transformation may not possess *any* eigenvalues; see Exercise 1.2.3. Thus for $\mathbb{R}$-linear transformations we have to work a little harder.

3.2.3.2 Invariant subspaces of $\mathbb{R}$-linear maps associated with complex eigenvalues The primary reason for complexifying a $\mathbb{R}$-vector space and then a $\mathbb{R}$-linear map is for the purpose of studying eigenvalues of $\mathbb{R}$-linear transformations. Thus we let V be a $\mathbb{R}$-vector space, let $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$, with $\mathsf{V}^{\mathbb{C}}$ and $L^{\mathbb{C}}$ the associated complexifications. We are interested in studying how the eigenvalues of L and $L^{\mathbb{C}}$ are related. The following result gives the relationships we seek.

3.2.33 Proposition (Eigenvalues and eigenspaces of a linear transformation and its complexification) *If V is a $\mathbb{R}$-vector space and if $\mathrm{L} \in \mathrm{L}(\mathsf{V};\mathsf{V})$ with $\mathrm{L}^{\mathbb{C}} \in \mathrm{L}(\mathsf{V}^{\mathbb{C}};\mathsf{V}^{\mathbb{C}})$ its complexification, then the following statements hold:*

(i) $\lambda \in \mathbb{R}$ *is an eigenvalue for* L *if and only if* λ *is an eigenvalue for* $\mathrm{L}^{\mathbb{C}}$*;*

(ii) if $\lambda \in \mathbb{C}$ *is an eigenvalue for* $\mathrm{L}^{\mathbb{C}}$*, then* $\bar{\lambda}$ *is an eigenvalue for* $\mathrm{L}^{\mathbb{C}}$*;*

(iii) if $\lambda \in \mathbb{R}$ *is an eigenvalue for* L *then*

$$\mathsf{W}(\lambda, \mathrm{L}^{\mathbb{C}}) = \{(\mathrm{u}, \mathrm{v}) \mid \mathrm{u}, \mathrm{v} \in \mathsf{W}(\lambda, \mathrm{L})\};$$

(iv) if $\lambda \in \mathbb{R}$ *is an eigenvalue for* L *then*

$$\overline{\mathsf{W}}(\lambda, \mathrm{L}^{\mathbb{C}}) = \{(\mathrm{u}, \mathrm{v}) \mid \mathrm{u}, \mathrm{v} \in \overline{\mathsf{W}}(\lambda, \mathrm{L})\};$$

(v) if $\lambda \in \mathbb{C}$ *is an eigenvalue for* $\mathrm{L}^{\mathbb{C}}$ *then*

$$\mathsf{W}(\bar{\lambda}, \mathrm{L}^{\mathbb{C}}) = \{(\mathrm{u}, \mathrm{v}) \in \mathsf{V}^{\mathbb{C}} \mid (\mathrm{u}, -\mathrm{v}) \in \mathsf{W}(\lambda, \mathrm{L}^{\mathbb{C}})\};$$

(vi) if $\lambda \in \mathbb{C}$ *is an eigenvalue for* $\mathrm{L}^{\mathbb{C}}$ *then*

$$\overline{\mathsf{W}}(\bar{\lambda}, \mathrm{L}^{\mathbb{C}}) = \{(\mathrm{u}, \mathrm{v}) \in \mathsf{V}^{\mathbb{C}} \mid (\mathrm{u}, -\mathrm{v}) \in \overline{\mathsf{W}}(\lambda, \mathrm{L}^{\mathbb{C}})\}.$$

Proof (i) First suppose that λ is an eigenvalue for L and denote by $\mathsf{W}(\lambda, L)$ the eigenspace. We claim that

$$\ker(L_{\lambda}^{\mathbb{C}}) = \{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid u, v \in \mathsf{W}(\lambda, L)\}.$$

Indeed, by definition of the complexification of a linear map, $L_{\lambda}^{\mathbb{C}}(u, v) = (0, 0)$ if and only if $L(u) = \lambda u$ and $L(v) = \lambda v$. This shows that λ is an eigenvalue of $L^{\mathbb{C}}$ and that the eigenspace is $\{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid u, v \in \mathsf{W}(\lambda, L)\}$.

Now suppose that $\lambda \in \mathbb{R}$ is an eigenvalue for $L^{\mathbb{C}}$ and let $\mathsf{W}(\lambda, L^{\mathbb{C}})$ be the eigenspace. Thus, by definition of the complexification of a linear map we have

$$\ker(L_{\lambda}^{\mathbb{C}}) = \{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid u, v \in \ker(L_{\lambda})\},$$

so giving λ as an eigenvalue for L and also giving

$$\mathsf{W}(\lambda, L^{\mathbb{C}}) = \{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid u, v \in \mathsf{W}(\lambda, L)\}.$$

(ii) Before we get to the proof of this part of the result, let us make some constructions with complexifications. For $M \in \mathrm{L}(\mathsf{V}^{\mathbb{C}}; \mathsf{V}^{\mathbb{C}})$, let us write

$$M(u, v) = (M_1(u) + M_2(v), M_3(u) + M_4(v))$$

for $(u, v) \in \mathsf{V}^{\mathbb{C}}$, and for some $M_1, M_2, M_3, M_4 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. For this map to be $\mathbb{C}$-linear, we must have $M(\mathrm{i}(u, v)) = \mathrm{i}M(u, v)$ for every $(u, v) \in \mathsf{V}^{\mathbb{C}}$. Using the definition of scalar multiplication in $\mathsf{V}^{\mathbb{C}}$, this reads

$$(-M_1(v) + M_2(u), -M_3(v) + M_4(u)) = (-M_3(u) - M_4(v), M_1(u) + M_2(v)).$$

This holds for every $(u, v) \in \mathsf{V}^{\mathbb{C}}$ if and only if $M_4 = M_1$ and $M_3 = -M_2$. Thus we can write

$$M(u, v) = (M_1(u) + M_2(v), -M_2(u) + M_1(v)).$$

Now define the ***conjugate*** of M as $\bar{M} \in \mathrm{L}(\mathsf{V}^{\mathbb{C}}; \mathsf{V}^{\mathbb{C}})$ defined by

$$\bar{M}(u, v) = (M_1(u) - M_2(v), M_2(u) + M_1(v)).$$

Note that

$$\begin{aligned}
& \bar{M} = M \\
\iff\ & M_1(u) - M_2(u) = M_1(u) + M_2(v),\ M_2(u) + M_1(v) = -M_2(u) + M_1(v), \\
& \qquad (u, v) \in \mathsf{V}^{\mathbb{C}} \\
\iff\ & M_2 = 0 \\
\iff\ & M = M_1^{\mathbb{C}}.
\end{aligned}$$

That it to say, M is the complexification of a $\mathbb{R}$-linear map if and only if $\bar{M} = M$.

With these constructions at hand, we proceed with the proof. We may as well suppose that λ is not real. Thus we write $\lambda = \sigma + \mathrm{i}\omega$ for $\sigma, \omega \in \mathbb{R}$ and with $\omega \neq 0$.

We first claim that $L^{\mathbb{C}}_{\bar{\lambda}} = \bar{L}^{\mathbb{C}}_{\lambda}$. Indeed

$$\bar{L}^{\mathbb{C}}_{\lambda} = \overline{L^{\mathbb{C}} - \lambda \,\mathrm{id}_{\mathsf{V}}} = \bar{L}^{\mathbb{C}} - \bar{\lambda}\,\bar{\mathrm{id}}_{\mathsf{V}} = L - \bar{\lambda}\,\mathrm{id}_{\mathsf{V}} = L^{\mathbb{C}}_{\bar{\lambda}}.$$

The following lemma gives us a useful characterisation of the kernel and image of the conjugate of a linear map, and this characterisation will be used several times in the remainder of the proof.

1 Lemma *If* U *and* V *are* $\mathbb{R}$*-vector spaces and if* $\mathrm{L} \in \mathrm{L}(\mathsf{U}^{\mathbb{C}}; \mathsf{V}^{\mathbb{C}})$*, then*

(i) $\ker(\bar{\mathrm{L}}) = \{(\mathrm{u}, \mathrm{v}) \in \mathsf{U}^{\mathbb{C}} \mid (\mathrm{u}, -\mathrm{v}) \in \ker(\mathrm{L})\}$ *and*

(ii) $\mathrm{image}(\bar{\mathrm{L}}) = \{(\mathrm{u}, \mathrm{v}) \in \mathsf{V}^{\mathbb{C}} \mid (\mathrm{u}, -\mathrm{v}) \in \mathrm{image}(\mathrm{L})\}$.

Proof As above, we may write

$$L(u, v) = (L_1(u) + L_2(v), -L_2(u) + L_1(v))$$

for $L_1, L_2 \in \mathrm{L}(\mathsf{U}; \mathsf{V})$. We then compute

$$\begin{aligned}
\ker(\bar{L}) &= \{(u, v) \in \mathsf{U}^{\mathbb{C}} \mid (L_1(u) - L_2(v), L_2(u) + L_1(v)) = (0, 0)\} \\
&= \{(u, -v) \in \mathsf{U}^{\mathbb{C}} \mid (L_1(u) + L_2(v), L_2(u) - L_1(v)) = (0, 0)\} \\
&= \{(u, -v) \in \mathsf{U}^{\mathbb{C}} \mid (L_1(u) + L_2(v), -L_2(u) + L_1(v)) = (0, 0)\} \\
&= \{(u, v) \in \mathsf{U}^{\mathbb{C}} \mid (u, -v) \in \ker(L)\},
\end{aligned}$$

giving the first part of the lemma.

For the second part we write

$$\begin{aligned}
\mathrm{image}(\bar{L}) &= \{(L_1(u) - L_2(v), L_2(u) + L_1(v)) \mid (u, v) \in \mathsf{U}^{\mathbb{C}}\} \\
&= \{(L_1(u) + L_2(v), L_2(u) - L_1(v)) \mid (u, -v) \in \mathsf{U}^{\mathbb{C}}\} \\
&= \{(L_1(u) + L_2(v), L_2(u) - L_1(v)) \mid (u, v) \in \mathsf{U}^{\mathbb{C}}\} \\
&= \{(u', v') \mid (u', -v') \in \mathrm{image}(L)\},
\end{aligned}$$

so giving the second part of the lemma. ▼

Now we proceed with the proof. Let us first consider the case when λ is an eigenvalue for $L^{\mathbb{C}}$. By the lemma we have

$$\ker(L^{\mathbb{C}}_{\bar{\lambda}}) = \{(u, -v) \mid (u, v) \in \ker(L^{\mathbb{C}}_{\lambda})\}.$$

Thus $\bar{\lambda}$ is an eigenvalue for $L^{\mathbb{C}}$ and

$$\mathsf{W}(\bar{\lambda}, \mathsf{V}^{\mathbb{C}}) = \{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid (u, -v) \in \mathsf{W}(\lambda, L^{\mathbb{C}})\}.$$

(iii) This was proved during the course of proving (i).

(iv) We have $(L^{\mathbb{C}}_{\lambda})^j(u, v) = (L^j_{\lambda}(u), L^j_{\lambda}(v))$ for each $j \in \mathbb{Z}_{>0}$ and $(u, v) \in \mathsf{V}^{\mathbb{C}}$. Therefore,

$$\ker((L^{\mathbb{C}}_{\lambda})^j) = \{(u, v) \in \mathsf{V}^{\mathbb{C}} \mid u, v \in \ker(L^j_{\lambda})\}.$$

From this we infer that

$$\cup_{j \in \mathbb{Z}_{>0}} \ker((L^{\mathbb{C}}_{\lambda})^j) = \left\{(u, v) \in \mathsf{V}^{\mathbb{C}} \;\middle|\; u, v \in \cup_{j \in \mathbb{Z}_{>0}} \ker(L^j_{\lambda})\right\},$$

which is the desired result.

(v) This was proved during the course of the proof of part (ii).

(vi) Since $L^{\mathbb{C}}_{\bar\lambda} = \bar L^{\mathbb{C}}_{\lambda}$, it follows that $(L^{\mathbb{C}}_{\bar\lambda})^j = (\bar L^{\mathbb{C}}_{\lambda})^j$ for each $j \in \mathbb{Z}_{>0}$. From the lemma above we then conclude that, for each $j \in \mathbb{Z}_{>0}$,

$$\ker((L^{\mathbb{C}}_{\bar\lambda})^j) = \left\{(u, v) \in \mathsf{V}^{\mathbb{C}} \;\middle|\; (u, -v) \in \ker((L^{\mathbb{C}}_{\lambda})^j)\right\}.$$

It follows that

$$\cup_{j\in\mathbb{Z}_{>0}} \ker((L^{\mathbb{C}}_{\bar\lambda})^j) = \left\{(u, v) \in \mathsf{V}^{\mathbb{C}} \;\middle|\; (u, -v) \in \cup_{j\in\mathbb{Z}_{>0}} \ker((L^{\mathbb{C}}_{\lambda})^j)\right\},$$

which is exactly the claim. ■

The proposition tells us that every eigenvalue of L is also an eigenvalue of $L^{\mathbb{C}}$. Of course, it is not generally the case that eigenvalues of $L^{\mathbb{C}}$ are also eigenvalues of L, since the former are allowed to be complex, whereas the latter are always real. Nonetheless, one can wonder what implications the existence of non-real eigenvalues for $L^{\mathbb{C}}$ has on the structure of L. The following result addresses precisely this point. The essential idea is that eigenspaces for $L^{\mathbb{C}}$ give rise to invariant subspaces for L of twice the dimension.

3.2.34 Proposition (Real invariant subspaces for complex eigenvalues) *Let* V *be a finite-dimensional* $\mathbb{R}$*-vector space, let* $\mathrm{L} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$*, and let* $\mathsf{V}^{\mathbb{C}}$ *and* $\mathrm{L}^{\mathbb{C}}$ *be the corresponding complexifications. Suppose that* $\lambda = \sigma + \mathrm{i}\omega$*,* $\sigma, \omega \in \mathbb{R}$*,* $\omega \neq 0$*, is a complex eigenvalue for* $\mathrm{L}^{\mathbb{C}}$ *and let* $\mathscr{B}_\lambda$ *and* $\overline{\mathscr{B}}_\lambda$ *be bases for the eigenspace* $\mathsf{W}(\lambda, \mathrm{L}^{\mathbb{C}})$ *and the generalised eigenspace* $\overline{\mathsf{W}}(\lambda, \mathrm{L}^{\mathbb{C}})$*, respectively. Then the following statements hold:*

(i) the sets

$$\begin{aligned} \mathscr{B}'_\lambda &= \{\mathrm{u} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \mathscr{B}_\lambda\} \cup \{\mathrm{v} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \mathscr{B}_\lambda\}, \\ \overline{\mathscr{B}}'_\lambda &= \{\mathrm{u} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \overline{\mathscr{B}}_\lambda\} \cup \{\mathrm{v} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \overline{\mathscr{B}}_\lambda\} \end{aligned} \tag{3.11}$$

are linearly independent;

(ii) if $(\mathrm{u}, \mathrm{v}) \in \mathscr{B}_\lambda$ *then*

$$\mathrm{L}(\mathrm{u}) = \sigma\mathrm{u} - \omega\mathrm{v}, \quad \mathrm{L}(\mathrm{v}) = \omega\mathrm{u} + \sigma\mathrm{v},$$

and so, in particular, the two-dimensional subspace $\mathrm{span}_{\mathbb{R}}(\mathrm{u}, \mathrm{v})$ *is* L*-invariant;*

(iii) the subspaces $\mathrm{span}_{\mathbb{R}}(\mathscr{B}'_\lambda)$ *and* $\mathrm{span}_{\mathbb{R}}(\overline{\mathscr{B}}'_\lambda)$ *are* L*-invariant;*

(iv) relative to the partition given in (3.11) for the basis $\mathscr{B}'_\lambda$*, the restriction of* L *to* $\mathrm{span}_{\mathbb{R}}(\mathscr{B}'_\lambda)$ *has the matrix representative*

$$\begin{bmatrix} \sigma \mathbf{I}_{\mathrm{k}} & \omega \mathbf{I}_{\mathrm{k}} \\ -\omega \mathbf{I}_{\mathrm{k}} & \sigma \mathbf{I}_{\mathrm{k}} \end{bmatrix},$$

where k *is the number of basis vectors in* $\mathscr{B}_\lambda$*.*

Proof (i) We shall prove the result for $\overline{\mathscr{B}}'_\lambda$, the proof for $\mathscr{B}'_\lambda$ being entirely similar. Let us define

$$\overline{\mathscr{B}}_{\bar\lambda} = \{(u,-v) \in \mathsf{V}^{\mathbb{C}} \mid (u,v) \in \overline{\mathscr{B}}_\lambda\},$$

noting by Proposition 3.2.29 that $\operatorname{span}_{\mathbb{C}}(\overline{\mathscr{B}}_\lambda) \cap \operatorname{span}_{\mathbb{C}}(\overline{\mathscr{B}}_{\bar\lambda}) = \{(0,0)\}$. Moreover, by Proposition 3.2.33(vi) we also know that $\overline{\mathscr{B}}_{\bar\lambda}$ is a basis for $\overline{\mathsf{W}}(\bar\lambda, L^{\mathbb{C}})$. These facts together ensure that $\overline{\mathscr{B}}_\lambda \cup \overline{\mathscr{B}}_{\bar\lambda}$ is a basis for $\overline{\mathsf{W}}(\lambda, L^{\mathbb{C}}) \oplus \overline{\mathsf{W}}(\bar\lambda, L^{\mathbb{C}})$. Now define $(2k) \times (2k)$-matrix $\boldsymbol{P}$ by

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{I}_k \\ \boldsymbol{I}_k & -\boldsymbol{I}_k \end{bmatrix}.$$

This matrix is invertible as one can see by checking that it has an inverse given by

$$\boldsymbol{P}^{-1} = \frac{1}{2}\begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{I}_k \\ \boldsymbol{I}_k & -\boldsymbol{I}_k \end{bmatrix}.$$

Thus $\boldsymbol{P}$ is a change of basis matrix from the basis $\overline{\mathscr{B}}_\lambda \cup \overline{\mathscr{B}}_{\bar\lambda}$ for $\overline{\mathsf{W}}(\lambda, L^{\mathbb{C}}) \oplus \overline{\mathsf{W}}(\bar\lambda, L^{\mathbb{C}})$ to another basis for $\overline{\mathsf{W}}(\lambda, L^{\mathbb{C}}) \oplus \overline{\mathsf{W}}(\bar\lambda, L^{\mathbb{C}})$. Using the definition of the change of basis matrix, one can further check that this new basis is exactly

$$\{(u,0) \mid (u,v) \in \overline{\mathscr{B}}_\lambda\} \cup \{(0,v) \mid (u,v) \in \overline{\mathscr{B}}_\lambda\}. \tag{3.12}$$

Using the fact that this is a basis, and so linearly independent, we now prove that $\overline{\mathscr{B}}'_\lambda$ is linearly independent. Let

$$\{u \in \mathsf{V} \mid (u,v) \in \overline{\mathscr{B}}_\lambda\} = \{u_1, \dots, u_k\}, \quad \{u \in \mathsf{V} \mid (u,v) \in \overline{\mathscr{B}}_\lambda\} = \{v_1, \dots, v_k\},$$

and suppose that

$$a_1u_1 + \cdots + a_ku_k + b_1v_1 + \cdots + b_kv_k = 0$$

for $a_1, \dots, a_k, b_1, \dots, b_k \in \mathbb{R}$. Using the definition of scalar multiplication in $\mathsf{V}^{\mathbb{C}}$ this implies that

$$(a_1 + \mathrm{i}0)(u_1,0) + \cdots + (a_k + \mathrm{i}0)(u_k,0) + (b_1 + \mathrm{i}0)(0,v_1) + \cdots + (b_k + \mathrm{i}0)(0,v_k) = (0,0).$$

Since the set (3.12) is linearly independent, we must have $a_j + \mathrm{i}0 = 0 + \mathrm{i}0$ and $b_j + \mathrm{i}0 = 0 + \mathrm{i}0$ for $j \in \{1, \dots, k\}$. This gives linear independence of $\overline{\mathscr{B}}'_\lambda$.

(ii) If $(u,v) \in \mathscr{B}_\lambda$ then we have

$$(L(u), L(v)) = L^{\mathbb{C}}(u,v) = (\sigma + \mathrm{i}\omega)(u,v) = (\sigma u - \omega v, \omega u + \sigma v),$$

as claimed. Since $L(u), L(v) \in \operatorname{span}_{\mathbb{R}}(u,v)$, it follows that $\operatorname{span}_{\mathbb{R}}(u,v)$ is L-invariant.

(iii) To prove this part of the result it is useful to employ a lemma that captures the essence of what is going on.

1 Lemma *Let* V *be a* $\mathbb{R}$*-vector space with complexification* $\mathsf{V}^{\mathbb{C}}$ *and let* $\mathrm{L} \in \mathrm{L}(\mathsf{V};\mathsf{V})$ *have complexification* $\mathrm{L}^{\mathbb{C}}$*. If* U *is a subspace of* $\mathsf{V}^{\mathbb{C}}$ *which is invariant under* $\mathrm{L}^{\mathbb{C}}$ *then*

(i) the subspace

$$\overline{\mathsf{U}} = \{(\mathrm{u}, -\mathrm{v}) \mid (\mathrm{u}, \mathrm{v}) \in \mathsf{U}\}$$

is $\mathrm{L}^{\mathbb{C}}$*-invariant and*

(ii) the subspaces

$$\{\mathrm{u} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \mathsf{U} + \overline{\mathsf{U}}\}, \quad \{\mathrm{v} \in \mathsf{V} \mid (\mathrm{u}, \mathrm{v}) \in \mathsf{U} + \overline{\mathsf{U}}\}$$

of V *are* L*-invariant.*

Proof (i) Let $(u, -v) \in \overline{\mathsf{U}}$ for $(u, v) \in \mathsf{U}$. Then $(L(u), L(v)) \in \mathsf{U}$ since U is $L^{\mathbb{C}}$-invariant. Therefore,

$$L^{\mathbb{C}}(u, -v) = (L(u), -L(v)) \in \overline{\mathsf{U}},$$

giving invariance of $\overline{\mathsf{U}}$ under $L^{\mathbb{C}}$ as desired.

(ii) Let $u \in \{u' \in \mathsf{V} \mid (u', v') \in \mathsf{U} + \overline{\mathsf{U}}\}$ and let $v \in \mathsf{V}$ have the property that $(u, v) \in \mathsf{U} + \overline{\mathsf{U}}$. Then $(u, -v) \in \mathsf{U} + \overline{\mathsf{U}}$. Since $\mathsf{U} + \overline{\mathsf{U}}$ is $L^{\mathbb{C}}$-invariant,

$$(L(u), L(v)), (L(u), -L(v)) \in \mathsf{U} + \overline{\mathsf{U}}.$$

Therefore, $(2L(u), 0) \in \mathsf{U} + \overline{\mathsf{U}}$ and so $L(u) \in \{u' \in \mathsf{V} \mid (u', v') \in \mathsf{U} + \overline{\mathsf{U}}\}$, giving invariance of $\{u' \in \mathsf{V} \mid (u', v') \in \mathsf{U} + \overline{\mathsf{U}}\}$ under L. A similar computation gives invariance of $\{v' \in \mathsf{V} \mid (u', v') \in \mathsf{U} + \overline{\mathsf{U}}\}$ under L. ▼

By applying the lemma with $\mathsf{U} = \mathsf{W}(\lambda, L^{\mathbb{C}})$ and then with $\mathsf{U} = \overline{\mathsf{W}}(\lambda, L^{\mathbb{C}})$, this part of the proposition follows.

(iv) Let us write

$$\mathscr{B}_\lambda = \{(u_1, v_1), \dots, (u_k, v_k)\}.$$

The basis $\mathscr{B}'_\lambda$ can then be written as

$$\mathscr{B}'_\lambda = \{u_1, \dots, u_k, v_1, \dots, v_k\}.$$

We then have

$$L(u_j) = \sigma u_j - \omega v_j, \quad L(v_j) = \omega u_j + \sigma v_j \qquad j \in \{1, \dots, k\}.$$

Using the definition of matrix representative, this then gives the matrix representative of $L|\operatorname{span}_{\mathbb{R}}(\mathscr{B}'_\lambda)$ to be

$$\begin{bmatrix} \sigma \boldsymbol{I}_I & \omega \boldsymbol{I}_I \\ -\omega \boldsymbol{I}_I & \sigma \boldsymbol{I}_I \end{bmatrix},$$

as desired. ■

The idea is that, for every $\mathbb{C}$-subspace of $\mathsf{V}^{\mathbb{C}}$ that is invariant under $L^{\mathbb{C}}$, there corresponds a $\mathbb{R}$-subspace of V of twice the dimension that is invariant under L. Moreover, one can choose as a basis for this $\mathbb{R}$-subspace the real and imaginary parts of the basis for the $\mathbb{C}$-subspace. Finally, if the invariant $\mathbb{C}$-subspace is an eigenspace for $\mathsf{V}^{\mathbb{C}}$, then the representation of L is related to the complex eigenvalue in a simple way (i.e., as in part (ii)).

It is useful to develop some notation for capturing all of this. To this end, for a complex eigenvalue $\lambda \in \mathbb{C} \setminus \mathbb{R}$, we denote

$$\mathsf{W}(\lambda, L) = \{u \in \mathsf{V} \mid (u, v) \in \mathsf{W}(\lambda, \mathsf{V}^{\mathbb{C}})\} + \{v \in \mathsf{V} \mid (u, v) \in \mathsf{W}(\lambda, L)\}$$

and

$$\overline{\mathsf{W}}(\lambda, L) = \{u \in \mathsf{V} \mid (u, v) \in \overline{\mathsf{W}}(\lambda, \mathsf{V}^{\mathbb{C}})\} + \{v \in \mathsf{V} \mid (u, v) \in \overline{\mathsf{W}}(\lambda, L)\}. \tag{3.13}$$

Note that, despite the notation, $\mathsf{W}(\lambda, L)$ is *not* an eigenspace and $\overline{\mathsf{W}}(\lambda, L)$ is *not* a generalised eigenspace, simply because λ is *not* an eigenvalue.

Let us consider a simple example of how this works.

3.2.35 Example (Complex eigenvalues for $\mathbb{R}$-linear transformations) We take the linear transformation $L \in \mathrm{L}(\mathbb{R}^3; \mathbb{R}^3)$ defined by the 3×3 matrix

$$\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}.$$

The characteristic polynomial of L is

$$P_L = \det \begin{bmatrix} X-1 & 0 & 2 \\ 0 & X-1 & 0 \\ -2 & 0 & X-1 \end{bmatrix} = X^3 - 3X^2 + 7X - 5.$$

This may be determined to have roots

$$\ell_1 = 1, \ \lambda_1 = 1 + 2\mathrm{i}, \ \lambda_3 = 1 - 2\mathrm{i}.$$

Let us first consider the invariant subspaces of $L^{\mathbb{C}}$. The generalised eigenspace (which is the same as the eigenspace, since the geometric multiplicity is 1) for L associated to the real eigenvalue $\ell = 1$ is

$$\overline{\mathsf{W}}(1, L) = \ker(1 \cdot \mathrm{id}_{\mathsf{V}} - L) = \ker \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix} = \mathrm{span}_{\mathbb{R}}((0, 1, 0)).$$

Therefore, by Proposition 3.2.33(iv), the generalised eigenspace for $L^{\mathbb{C}}$ associated to the real eigenvalue $\ell = 1$ is

$$\overline{\mathsf{W}}(1, L^{\mathbb{C}}) = \mathrm{span}_{\mathbb{R}}(((0, 1, 0), (0, 0, 0)), ((0, 0, 0), (0, 1, 0))).$$

Written more clearly, maybe,

$$\overline{\mathsf{W}}(1, L^{\mathbb{C}}) = \operatorname{span}_{\mathbb{R}}((0,1,0), \mathrm{i}(0,1,0)).$$

The generalised eigenspace (which is the same as the eigenspace, since the geometric multiplicity is 1) for $L^{\mathbb{C}}$ associated to the complex eigenvalue $\lambda_1 = 1 + 2\mathrm{i}$ is

$$\overline{\mathsf{W}}(1 + 2\mathrm{i}, L^{\mathbb{C}}) = \ker((1 + 2\mathrm{i}) \cdot \mathrm{id}_{\mathsf{V}} - L) = \ker \begin{bmatrix} 2\mathrm{i} & 0 & 2 \\ 0 & 2\mathrm{i} & 0 \\ -2 & 0 & 2\mathrm{i} \end{bmatrix} = \operatorname{span}_{\mathbb{C}}(((0,0,1), (1,0,0))).$$

Written in a perhaps clearer way,

$$\overline{\mathsf{W}}(1 + 2\mathrm{i}, L^{\mathbb{C}}) = \operatorname{span}_{\mathbb{C}}((0,0,1) + \mathrm{i}(1,0,0)).$$

From Proposition 3.2.33(vi) we immediately have

$$\overline{\mathsf{W}}(1 - 2\mathrm{i}, L^{\mathbb{C}}) = \operatorname{span}_{\mathbb{C}}((0,0,1) - \mathrm{i}(1,0,0)).$$

We note that

$$(\mathbb{R}^3)^{\mathbb{C}} = \mathbb{C}^3 = \overline{\mathsf{W}}(1, L^{\mathbb{C}}) \oplus \overline{\mathsf{W}}(1 + 2\mathrm{i}, L^{\mathbb{C}}) \oplus \overline{\mathsf{W}}(1 - 2\mathrm{i}, L^{\mathbb{C}}).$$

Now we can think about the invariant subspaces of L. Here we work with the real eigenvalue and *one* of the complex eigenvalues (the other being conjugate and so essentially redundant). As above, we have the generalised eigenspace corresponding to the real eigenvalue $\ell = 1$ given by

$$\overline{\mathsf{W}}(1, L) = \operatorname{span}_{\mathbb{R}}((0,1,0)).$$

For the 2-dimensional invariant subspace corresponding to $\lambda_1 = 1 + 2\mathrm{i}$, by Proposition 3.2.34(i) we take the real and imaginary parts of the basis for $\overline{\mathsf{W}}(1 + 2\mathrm{i}, L^{\mathbb{C}})$, i.e., the 2-dimensional subspace

$$\overline{\mathsf{W}}(1 + 2\mathrm{i}) = \operatorname{span}_{\mathbb{R}}((0,0,1), (1,0,0)).$$

Thus we have the invariant subspace decomposition

$$\mathbb{R}^3 = \overline{\mathsf{W}}(1, L) \oplus \overline{\mathsf{W}}(1 + 2\mathrm{i}, L). \qquad \bullet$$

3.2.36 Remark (Summary of linear algebraic constructions) The preceding two sections developed a fairly complicated picture of the structure of linear transformations associated with eigenvalues. In this remark, we summarise the take-away message from all of this. We let V be a finite-dimensional $\mathbb{R}$-vector space and let $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. We suppose we have distinct real eigenvalues

$$\ell_1, \dots, \ell_r$$

for L and distinct complex eigenvalues

$$\lambda_1 = \sigma_1 + \mathrm{i}\omega_1, \dots, \lambda_s = \sigma_2 + \mathrm{i}\omega_s,$$

along with their complex conjugates. We let $m_{\mathrm{a}}(\ell_j, L)$, $j \in \{1, \dots, r\}$, and $m_{\mathrm{a}}(\lambda_j, L)$, $j \in \{1, \dots, s\}$, be the algebraic multiplicities.

1. We have
$$\sum_{j=1}^{r} m_{\mathrm{a}}(\ell_j, L) + 2 \sum_{j=1}^{s} m_{\mathrm{a}}(\lambda_j, L) = \dim_{\mathbb{R}}(\mathsf{V}).$$
2. For each $j \in \{1, \dots, r\}$, there is a subspace
$$\overline{\mathsf{W}}(\ell_j, L) = \ker((\ell_j \mathrm{id}_{\mathsf{V}} - L)^{m_{\mathrm{a}}(\ell_j, L)})$$
of V of $\mathbb{R}$-dimension $m_{\mathrm{a}}(\ell_j, L)$ that is L-invariant.
3. For each $j \in \{1, \dots, s\}$, there is a subspace
$$\overline{\mathsf{W}}(\lambda_j, L^{\mathbb{C}}) = \ker((\lambda_j \mathrm{id}_{\mathsf{V}} - L^{\mathbb{C}})^{m_{\mathrm{a}}(\lambda_j, L)})$$
of $\mathsf{V}^{\mathbb{C}}$ of $\mathbb{C}$-dimension $m_{\mathrm{a}}(\lambda_j, L)$ that is $L^{\mathbb{C}}$-invariant.
4. For each $j \in \{1, \dots, s\}$, there is a subspace $\overline{\mathsf{W}}(\lambda_j, L)$ of V as in (3.13) of $\mathbb{R}$-dimension $2m_{\mathrm{a}}(\lambda_j, L)$ that is L-invariant.
5. We have
$$\mathsf{V} = \overline{\mathsf{W}}(\ell_1, L) \oplus \cdots \oplus \overline{\mathsf{W}}(\ell_r, L) \oplus \overline{\mathsf{W}}(\lambda_1, L) \oplus \cdots \oplus \overline{\mathsf{W}}(\lambda_s, L). \tag{3.14}$$
This decomposition of V into L-invariant subspaces will form the basis for Procedure 3.2.45 where we determine the state transition map for a system of linear homogeneous ordinary differential equations with constant coefficients.

•

3.2.3.3 The Jordan canonical form In the preceding two sections we described a collection of invariant subspaces of a linear transformation L of a finite-dimensional $\mathbb{R}$-vector space associated with the eigenvalues of L. It turns out that the resulting invariant subspace decomposition (3.14) can be further refined. We shall present this refinement without proof since it is interesting but not ultimately useful for us.

The key idea to organise the discussion is the following.

3.2.37 Definition (Jordan blocks) Let $k \in \mathbb{Z}_{>0}$, let $\ell \in \mathbb{R}$, and let $\lambda = \sigma + \mathrm{i}\omega$ with $\omega \neq 0$. Denote

$$\boldsymbol{B}(\sigma, \omega) = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}.$$

We have the following constructions.

(i) The $k \times k$ matrix

$$\boldsymbol{J}(\ell, k) \triangleq \begin{bmatrix} \ell & 1 & 0 & \cdots & 0 & 0 \\ 0 & \ell & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \ell & 1 \\ 0 & 0 & 0 & \cdots & 0 & \ell \end{bmatrix}$$

is the $\mathbb{R}$-***Jordan block*** associated with k and ℓ.

(ii) The $2k \times 2k$-matrix

$$\boldsymbol{J}(\sigma, \omega, k) \triangleq \begin{bmatrix} \boldsymbol{B}(\sigma, \omega) & \boldsymbol{I}_2 & \boldsymbol{0}_{2\times 2} & \cdots & \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} \\ \boldsymbol{0}_{2\times 2} & \boldsymbol{B}(\sigma, \omega) & \boldsymbol{I}_2 & \cdots & \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} & \cdots & \boldsymbol{B}(\sigma, \omega) & \boldsymbol{I}_2 \\ \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times 2} & \cdots & \boldsymbol{0}_{2\times 2} & \boldsymbol{B}(\sigma, \omega) \end{bmatrix}$$

is the $\mathbb{R}$-***Jordan block*** associated with k and $\lambda = \sigma + \mathrm{i}\omega \in \mathbb{C}$.

(iii) A ***Jordan arrangement*** for ℓ is a matrix of the form

$$\boldsymbol{J}(\ell, \boldsymbol{k}) = \begin{bmatrix} \boldsymbol{J}(\ell, k_1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}(\ell, k_2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{J}(\ell, k_r) \end{bmatrix}$$

for some $\boldsymbol{k} = (k_1, \ldots, k_r) \in \mathbb{Z}_{>0}$.

(iv) A ***Jordan arrangement*** for $\lambda = \sigma + \mathrm{i}\omega$ is a matrix of the form

$$\boldsymbol{J}(\sigma, \omega, \boldsymbol{k}) = \begin{bmatrix} \boldsymbol{J}(\sigma, \omega, k_1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}(\sigma, \omega, k_2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{J}(\sigma, \omega, k_r) \end{bmatrix}$$

for some $\boldsymbol{k} = (k_1, \ldots, k_r) \in \mathbb{Z}_{>0}$. •

With this notation we can state the following canonical form for $\mathbb{R}$-endomorphisms.

3.2.38 Theorem ($\mathbb{R}$-Jordan canonical form) *Let* V *be a finite-dimensional* $\mathbb{R}$*-vector space. For* $\mathrm{L} \in \mathrm{L}(\mathsf{V};\mathsf{V})$ *suppose that* $\ell_j \in \mathbb{R}$, $j \in \{1,\dots,r\}$, *and* $\sigma_j, \omega_j \in \mathbb{R}$, $\omega_j > 0$, $j \in \{1,\dots,s\}$, *are such that*

$$\ell_1,\dots,\ell_r,\sigma_1+\mathrm{i}\omega_1,\dots,\sigma_s+\mathrm{i}\omega_s,\sigma_1-\mathrm{i}\omega_1,\dots,\sigma_s-\mathrm{i}\omega_s$$

are the distinct eigenvalues of L. *Then there exists*

(i) $p_j \in \mathbb{Z}_{>0}$, $j \in \{1,\dots,r\}$,

(ii) $\mathbf{k}_j \in \mathbb{Z}_{>0}^{p_j}$, $j \in \{1,\dots,r\}$,

(iii) $q_j \in \mathbb{Z}_{>0}$, $j \in \{1,\dots,s\}$,

(iv) $\mathbf{l}_j \in \mathbb{Z}_{>0}^{q_j}$, $j \in \{1,\dots,s\}$, *and*

(v) a basis $\mathscr{B}$ *for* V

such that

$$[\mathrm{L}]_{\mathscr{B}}^{\mathscr{B}} = \begin{bmatrix} \mathbf{J}(\ell_1,\mathbf{k}_1) & \cdots & \mathbf{0} & \mathbf{0}\cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{J}(\ell_r,\mathbf{k}_r) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}(\sigma_1,\omega_1,\mathbf{l}_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}(\sigma_s,\omega_s,\mathbf{l}_s) \end{bmatrix}.$$

Moreover, this form of the matrix representative is unique up to reordering of the diagonal blocks.

3.2.3.4 Complexification of systems of linear ordinary differential equations

In Section 2.2.2.1 we complexified a scalar linear homogeneous ordinary differential equation with constant coefficients. The reason we had to do so was that the characteristic polynomial for such an equation will generally have complex roots, and these complex roots lead naturally to complex solutions of the differential equation. It is only after taking real and imaginary parts of a complex solution that we recover the real solutions. The same sort of thing happens with systems of linear homogeneous ordinary differential equations with constant coefficients. In this case, the issue that arises, as ought to be clear from the discussion in the preceding two sections, is that one will generally have complex eigenvalues.

The process of complexification is an easy one, and requires no words like "everything we have done in the real case also works in the complex case," since we are working with systems defined on abstract $\mathbb{R}$-vector spaces, and $\mathsf{V}^{\mathbb{C}}$ is certainly a $\mathbb{R}$-vector space.

3.2.39 Definition (Complexification of a system of linear ordinary differential equation) Consider the system of linear homogeneous ordinary differential equations F with constant coefficients and with right-hand side (3.9). The ***complexification***

of F is the system of linear homogeneous ordinary differential equations $F^{\mathbb{C}}$ with constant coefficients given by

$$\begin{aligned} F^{\mathbb{C}} \colon \mathbb{T} \times \mathsf{V}^{\mathbb{C}} \times \mathsf{V}^{\mathbb{C}} &\to \mathsf{V}^{\mathbb{C}} \\ (t, z, w) &\mapsto w - A^{\mathbb{C}}(z). \end{aligned}$$

•

A ***solution*** for $F^{\mathbb{C}}$ is a C^1-map $\zeta \colon \mathbb{T} \to \mathsf{V}^{\mathbb{C}}$ that satisfies

$$\dot{\zeta}(t) = A^{\mathbb{C}}(\zeta(t)).$$

Note that, as $\mathsf{V}^{\mathbb{C}} = \mathsf{V} \times \mathsf{V}$, we can write $\zeta(t) = (\xi(t), \eta(y))$ for C^1-maps $\xi, \eta \colon \mathbb{T} \to \mathsf{V}$ that are the ***real part*** and ***imaginary part*** of ζ, respectively.

As in the scalar case, the real and imaginary parts of a solution separately satisfy the uncomplexified differential equation.

3.2.40 Lemma (Real and imaginary parts of complex solutions are solutions) *Consider the system of linear homogeneous ordinary differential equations* F *with constant coefficients, with right-hand side (3.9) and with complexification* $\mathrm{F}^{\mathbb{C}}$*. If* $\zeta \colon \mathbb{T} \to \mathsf{V}^{\mathbb{C}}$ *is a solution for* $\mathrm{F}^{\mathbb{C}}$*, then* $\mathrm{Re}(\zeta)$ *and* $\mathrm{Im}(\zeta)$ *are solutions for* F*.*

Proof Given $\zeta \colon \mathbb{T} \to \mathsf{V}^{\mathbb{C}}$ we write $\zeta(t) = (\xi(t), \eta(t))$ so that $\xi = \mathrm{Re}(\zeta)$ and $\eta = \mathrm{Im}(\zeta)$. Since ζ is a solution for $F^{\mathbb{C}}$, we have

$$\dot{\zeta}(t) = (\dot{\xi}(t), \dot{\eta}(t)) = A^{\mathbb{C}}(\zeta(t)) = (A(\xi(t)), A(\eta(t))$$

by definition of $A^{\mathbb{C}}$. Equating the second and fourth terms in this string of equalities gives the lemma. ■

3.2.3.5 The operator exponential

In this section we consider the constant coefficient version of the state transition map.

3.2.41 Definition (Operator exponential) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, V be a finite-dimensional $\mathbb{F}$-vector space, and let $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. The ***operator exponential*** of L is the linear map $\mathrm{e}^L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ defined by $\mathrm{e}^L = \Phi_A(1, 0)$, where $A \colon [0, 1] \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ is defined by $A(t) = L$ for all $t \in [0, 1]$. •

What we call the "operator exponential" will almost universally be called the "matrix exponential" because it is defined as we have defined it, but in the case where $\mathsf{V} = \mathbb{R}^n$ and so L is an $n \times n$ matrix. Since we work with abstract vector spaces, our terminology in perhaps better suited to our setting.

Let us give some alternative characterisations and properties of the operator exponential.

3.2.42 Theorem (Properties of the operator exponential) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be a finite-dimensional $\mathbb{R}$-vector space, and let $\mathrm{L}, \mathrm{M} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. Then the following statements hold:*

(i) $\mathrm{e}^{\mathrm{L}} = \mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} \frac{\mathrm{L}^k}{k!}$;

(ii) if $\mathbb{F} = \mathbb{C}$, then e^{L} is a $\mathbb{C}$-linear map;

(iii) $\frac{\mathrm{d}}{\mathrm{dt}} \mathrm{e}^{\mathrm{Lt}} = \mathrm{L} \circ \mathrm{e}^{\mathrm{Lt}} = \mathrm{e}^{\mathrm{Lt}} \circ \mathrm{L}$;

(iv) $\mathrm{e}^0 = \mathrm{id}_{\mathsf{V}}$;

(v) for $\alpha \in \mathbb{F}$, $\mathrm{e}^{\alpha} \mathrm{id}_{\mathsf{V}} = \mathrm{e}^{\alpha \mathrm{id}_{\mathsf{V}}}$;

(vi) $\mathrm{e}^{\mathrm{Lt}} \circ \mathrm{e}^{\mathrm{Mt}} = \mathrm{e}^{(\mathrm{L}+\mathrm{M})\mathrm{t}}$ *for all $\mathrm{t} \in \mathbb{R}$ if and only if* $\mathrm{L} \circ \mathrm{M} = \mathrm{M} \circ \mathrm{L}$;

(vii) e^{L} *is invertible and* $(\mathrm{e}^{\mathrm{L}})^{-1} = \mathrm{e}^{-\mathrm{L}}$;

(viii) if $\mathsf{U} \subseteq \mathsf{V}$ is L-invariant, then it is also e^{L}-invariant;

(ix) the solution to the initial value problem

$$\dot{\xi}(\mathrm{t}) = \mathrm{L}(\xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathrm{x}_0,$$

is $\xi(\mathrm{t}) = \mathrm{e}^{\mathrm{L}(\mathrm{t}-\mathrm{t}_0)}(\mathrm{x}_0)$.

Proof (i) Let $A \colon [0,1] \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ be defined by $A(t) = L$ for $t \in [0,1]$. Adapting the notation of Section 3.2.2.3, if we define

$$I_k = \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{k-1}} A(t_1) \circ A(t_2) \circ \cdots \circ A(t_k) \, \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1,$$

then

$$\mathrm{e}^A = \mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} I_k,$$

and we know the series converges by virtue of Theorem 3.2.15. Note that

$$I_k = L^k \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{k-1}} \mathrm{d}t_k \cdots \mathrm{d}t_2 \mathrm{d}t_1 = \frac{L^k}{k!}$$

by (3.7), and this part of the result then follows.

(ii) Since e^L is $\mathbb{R}$-linear, we have

$$\mathrm{e}^L(v_1 + v_2) = \mathrm{e}^L(v_1) + \mathrm{e}^L(v_2).$$

Now let $v \in \mathsf{V}$ and $a \in \mathbb{C}$. We have

$$\mathrm{e}^L(av) = av + \sum_{k=1}^{\infty} \frac{L^k}{k!}(av) = a\left(v + \sum_{k=1}^{\infty} \frac{L^k}{k!}(v)\right) = a \exp^L(v),$$

using part (i) and $\mathbb{C}$-linearity of L, and hence also of L^k for every $k \in \mathbb{Z}_{>0}$.

(iii) As we say in the proof of Theorem 3.2.15, both series

$$\sum_{k=0}^{\infty} \frac{L^k t^k}{k!},$$

and the series

$$\sum_{k=1}^{\infty} \frac{L^k t^{k-1}}{(k-1)!} = L \circ \left(\sum_{k=0}^{\infty} \frac{L^k t^k}{k!} \right) = \left(\sum_{k=0}^{\infty} \frac{L^k t^k}{k!} \right) \circ L.$$

of term-by-term derivatives with respect to t, converge uniformly on any bounded time-domain. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{Lt} = L \circ \mathrm{e}^{Lt} = \mathrm{e}^{Lt} \circ L.$$

(iv) This follows from part (i).

(v) By part (i) we have

$$\mathrm{e}^{\alpha \, \mathrm{id}_{\mathsf{V}}} = \mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} \frac{\mathrm{id}_{\mathsf{V}}^k \, \alpha^k}{k!} = \left(1 + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} \right) \mathrm{id}_{\mathsf{V}} = \mathrm{e}^{\alpha} \, \mathrm{id}_{\mathsf{V}},$$

as desired.

(vi) Suppose that $L \circ M = M \circ L$. This gives

$$(L + M)^k = \sum_{j=0}^{k} \binom{k}{j} L^j M^{k-j},$$

using the Binomial Formula, where $\binom{k}{j} = \frac{k!}{j!(k-j)!}$. (Note that this *does* require that $L \circ M = M \circ L$.) Then

$$\begin{aligned} \mathrm{e}^{(L+M)t} &= \mathrm{id}_{\mathsf{V}} + \sum_{k=1} \frac{(L+M)^k t^k}{k!} \\ &= \mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} \sum_{j=0}^{k} \frac{L^j t^j M^{k-j} t^{k-j}}{j!(k-j)!} \\ &= \left(\mathrm{id}_{\mathsf{V}} + \sum_{j=1}^{\infty} \frac{L^j t^j}{j!} \right) \left(\mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{\infty} \frac{M^k t^k}{k!} \right) \end{aligned}$$

for all $t \in \mathbb{R}$.

Now suppose that $\mathrm{e}^{(L+M)t} = \mathrm{e}^{Lt} \circ \mathrm{e}^{Mt}$ for all $t \in \mathbb{R}$. We then compute

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{(L+M)t} = (L + M) \circ \mathrm{e}^{(L+M)t}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{Lt}\circ\mathrm{e}^{Mt} = L\circ\mathrm{e}^{Lt}\circ\mathrm{e}^{Mt} + \mathrm{e}^{Lt}\circ M\circ\mathrm{e}^{Mt}.$$

Next

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathrm{e}^{(L+M)t} = (L+M)^2\mathrm{e}^{(L+M)t}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathrm{e}^{Lt}\circ\mathrm{e}^{Mt} = L^2\circ\mathrm{e}^{Lt}\circ\mathrm{e}^{Mt} + L\circ\mathrm{e}^{Lt}\circ M\circ\mathrm{e}^{Mt} + L\circ\mathrm{e}^{Lt}\circ M\circ\mathrm{e}^{Mt} + \mathrm{e}^{Lt}\circ M^2\circ\mathrm{e}^{Mt}.$$

Evaluating the two second-derivatives at $t = 0$ and equating them gives

$$\begin{aligned}
&(L+M)^2 = L^2 + 2L\circ M + M^2\\
\implies\ &L^2 + L\circ M + M\circ L + M^2 = L^2 + 2L\circ M + M^2\\
\implies\ &M\circ L = L\circ M,
\end{aligned}$$

as desired.

(vii) That e^L is invertible is a consequence of its definition and Theorem 3.2.9(v). By parts (iv) and (vi), we have

$$\mathrm{id}_{\mathsf{V}} = \mathrm{e}^{L-L} = \mathrm{e}^L\mathrm{e}^{-L},$$

from which we conclude that $(\mathrm{e}^L)^{-1} = \mathrm{e}^{-L}$.

(viii) Let U be an L-invariant subspace of V and let $u \in \mathsf{U}$. We claim that U is also L^k-invariant for every $k \in \mathbb{Z}_{>0}$. This we prove by induction, it obviously being true when $k = 1$. Suppose it true for $k = m$ and let $u \in \mathsf{U}$. Then $L^{m+1}(u) = L\circ L^m(u)$. Since $L^m(u) \in \mathsf{U}$ and since U is L-invariant, we immediately have $L^{m+1}(u) \in \mathsf{U}$, showing that, indeed, U is L^k-invariant for every $k \in \mathbb{Z}_{>0}$. Using part (i) we then have

$$\left(\mathrm{id}_{\mathsf{V}} + \sum_{k=1}^{m}\frac{L^k}{k!}\right)(u) = u + \sum_{k=1}^{m}\frac{L^k(u)}{k!} \in \mathsf{U}.$$

Thus we have the sequence $(u_m)_{m\in\mathbb{Z}_{>0}}$ in V given by

$$u + \sum_{k=1}^{m}\frac{L^k(u)}{k!}.$$

Since U is closed,[6] we have

$$\mathrm{e}^L(u) = u + \sum_{k=1}^{\infty}\frac{L^k(u)}{k!} = \lim_{m\to\infty} u_m \in \mathsf{U},$$

[6]Let us sketch why a subspace U of a finite-dimensional vector space V is closed. Suppose that we have a sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ in U that converges in V to some u. Let $\{f_1,\dots,f_r\}$ be a basis for U that extends to a basis $\{f_1,\dots,f_r,e_1,\dots,e_{n-r}\}$ for V. Write

$$u = u_1f_1 + \cdots + u_rf_r + v_1e_1 + \cdots + v_{n-r}e_{n-r}$$

as desired.

(ix) Using (iii) we compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{L(t-t_0)}(x_0) = L \circ \mathrm{e}^{L(t-t_0)}(x_0).$$

We also have $\mathrm{e}^{L(t-t_0)}(x_0)$, when evaluated at $t = t_0$, is x_0 by part (iv). Thus $t \mapsto \mathrm{e}^{L(t-t_0)}(x_0)$ does indeed satisfy the stated initial value problem. ■

3.2.43 Remark ($\mathrm{e}^{\mathsf{L}} \circ \mathrm{e}^{\mathsf{M}} = \mathrm{e}^{\mathsf{L}+\mathsf{M}}$ does not imply $\mathsf{L} \circ \mathsf{M} = \mathsf{M} \circ \mathsf{L}$) Let $\mathsf{V} = \mathbb{R}^3$ and define $L, M \in \mathrm{L}(\mathbb{R}^3; \mathbb{R}^3)$ by the matrices

$$\begin{bmatrix} 0 & 6\pi & 0 \\ -6\pi & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 8\pi \\ 0 & -8\pi & 0 \end{bmatrix},$$

respectively. Using Procedure 3.2.48 below, we can compute $\mathrm{e}^L = \mathrm{e}^M = \mathrm{e}^{L+M} = \mathrm{id}_{\mathbb{R}^3}$, and so $\mathrm{e}^L\mathrm{e}^M = \mathrm{e}^{L+M}$. However, we do not have $L \circ M = M \circ L$, as may be verified by a direct computation. •

Let us consider the representation of the operator exponential in a basis.

3.2.44 Proposition (The matrix representation of the operator exponential is the operator exponential of the matrix representation) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an n-dimensional $\mathbb{F}$-vector space, let $\mathrm{L} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, and let $\mathscr{B} = \{\mathrm{e}_1, \dots, \mathrm{e}_n\}$ be a basis for V. Then*

$$[\mathrm{e}^{\mathrm{L}}]^{\mathscr{B}}_{\mathscr{B}} = \mathrm{e}^{[\mathrm{L}]^{\mathscr{B}}_{\mathscr{B}}}.$$

Proof This follows from the definition of the operator exponential and Exercise 3.2.4. ■

3.2.3.6 Bases of solutions Now, for equations with constant coefficients, we construct "explicitly" a basis for $\mathrm{Sol}(F)$.

---

and

$$u_j = u_{j,1}f_1 + \cdots + u_{j,r}f_r, \qquad j \in \mathbb{Z}_{>0}.$$

Then

$$\begin{aligned}
& \lim_{j\to\infty} u_j = u \\
\implies \quad & \lim_{j\to\infty}(u_{j,1}f_1 + \cdots + u_{j,r}f_r) = u_1f_1 + \cdots + u_rf_r + v_1e_1 + \cdots + v_{n-r}e_{n-r} \\
\implies \quad & \left(\lim_{j\to\infty} u_{j,1}\right)f_1 + \cdots + \left(\lim_{j\to\infty} u_{j,r}\right)f_r = u_1f_1 + \cdots + u_rf_r + v_1e_1 + \cdots + v_{n-r}e_{n-r} \\
\implies \quad & \lim_{j\to\infty} u_{j,a} = u_a,\ v_b = 0, \qquad a \in \{1, \dots, r\},\ b \in \{1, \dots, n-r\}.
\end{aligned}$$

Thus $u \in \mathsf{U}$, as claimed. Perhaps a reader will need a little more analysis that they have to fully understand this proof, but the main ideas are suggestive as concerns the truth of the assertion.

3.2.45 Procedure (Basis of solutions for a system of linear homogeneous ordinary differential equations with constant coefficients) Given a system of linear homogeneous ordinary differential equations

$$F\colon \mathbb{T} \times \mathsf{V} \oplus \mathsf{V} \to \mathsf{V}$$

in an n-dimensional $\mathbb{R}$-vector space V and with right-hand side

$$\widehat{F}(t, x) = A(x),$$

do the following.

1. Choose a basis $\{e_1, \dots, e_n\}$ for V. Let $\boldsymbol{A}$ be the matrix representative of A with respect to this basis. If $\mathsf{V} = \mathbb{R}^n$, one can just take $\boldsymbol{A}$ to be the usual matrix associated with $A \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$.
2. Compute the characteristic polynomial $P_A = \det(X\boldsymbol{I}_n - \boldsymbol{A})$.
3. Compute the roots of P_A, i.e., the eigenvalues of $A^{\mathbb{C}}$, and organise them as follows. We have distinct real roots

   $$\ell_1, \dots, \ell_r$$

   and distinct complex roots

   $$\lambda_1 = \sigma_1 + \mathrm{i}\omega_1, \dots, \lambda_s = \sigma_2 + \mathrm{i}\omega_s,$$

   $\omega_1, \dots, \omega_s \in \mathbb{R}_{>0}$, along with their complex conjugates.
4. Let $m_j = m_{\mathrm{a}}(\ell_j, A)$, $j \in \{1, \dots, r\}$, and $\mu_j = m_{\mathrm{a}}(\lambda_j, A)$, $j \in \{1, \dots, s\}$, be the algebraic multiplicities.
5. For $j \in \{1, \dots, r\}$, let $\{\boldsymbol{x}_{j,1}, \dots, \boldsymbol{x}_{j,m_j}\}$ be a basis for

   $$\overline{\mathsf{W}}(\ell_j, \boldsymbol{A}) = \ker((\ell_j \boldsymbol{I}_n - \boldsymbol{A})^{m_j}).$$

6. For $j \in \{1, \dots, s\}$, let $\{\boldsymbol{z}_{j,1}, \dots, \boldsymbol{z}_{j,\mu_j}\}$ be a basis for

   $$\overline{\mathsf{W}}(\lambda_j, \boldsymbol{A}^{\mathbb{C}}) = \ker((\lambda_j \boldsymbol{I}_n - \boldsymbol{A}^{\mathbb{C}})^{\mu_j}).$$

   Write $\boldsymbol{z}_{j,k} = \boldsymbol{a}_{j,k} + \mathrm{i}\boldsymbol{b}_{j,k}$ for each $k \in \{1, \dots, \mu_j\}$. Then

   $$\{\boldsymbol{a}_{j,1}, \boldsymbol{b}_{j,1}, \dots, \boldsymbol{a}_{j,\mu_j}, \boldsymbol{b}_{j,\mu_j}\}$$

   is a basis for $\overline{\mathsf{W}}(\lambda_j, \boldsymbol{A})$.
7. For $j \in \{1, \dots, r\}$ and $k \in \{1, \dots, m_j\}$, define

   $$\boldsymbol{\xi}_{j,k}(t) = \mathrm{e}^{\ell_j t}\left(\boldsymbol{I}_n + \frac{(\boldsymbol{A} - \ell_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A} - \ell_j \boldsymbol{I}_n)^{m_j - 1} t^{m_j - 1}}{(m_j - 1)!}\right)\boldsymbol{x}_{j,k}.$$

8. For $j \in \{1, \dots, s\}$ and $k \in \{1, \dots, \mu_j\}$, define

$$\boldsymbol{\alpha}_{j,k}(t) = \mathrm{e}^{\sigma_j t}\left(\left(\sum_{m=0}^{\mu_j-1}\sum_{l=0}^{\lceil\frac{m-1}{2}\rceil}\frac{(-1)^l\omega^{2l}t^m}{(2l)!(m-2l)!}(\boldsymbol{A}-\sigma_j\boldsymbol{I}_n)^{m-2l}\right)(\cos(\omega_j t)\boldsymbol{a}_{j,k}-\sin(\omega_j t)\boldsymbol{b}_{j,k})\right.$$
$$\left.-\left(\sum_{m=0}^{\mu_j-1}\sum_{l=0}^{\lfloor\frac{m-1}{2}\rfloor}\frac{(-1)^{j+1}\omega^{2l+1}t^m}{(2l+1)!(m-2l-1)!}(\boldsymbol{A}-\sigma_j\boldsymbol{I}_n)^{m-2l-1}\right)(cos(\omega_j t)\boldsymbol{b}_{j,k}+\sin(\omega_j t)\boldsymbol{a}_{j,k})\right) \tag{3.15}$$

and

$$\boldsymbol{\beta}_{j,k}(t) = \mathrm{e}^{\sigma_j t}\left(\left(\sum_{m=0}^{\mu_j-1}\sum_{l=0}^{\lceil\frac{m-1}{2}\rceil}\frac{(-1)^l\omega^{2l}t^m}{(2l)!(m-2l)!}(\boldsymbol{A}-\sigma_j\boldsymbol{I}_n)^{m-2l}\right)(\cos(\omega_j t)\boldsymbol{b}_{j,k}+\sin(\omega_j t)\boldsymbol{a}_{j,k})\right.$$
$$\left.+\left(\sum_{m=0}^{\mu_j-1}\sum_{l=0}^{\lfloor\frac{m-1}{2}\rfloor}\frac{(-1)^{j+1}\omega^{2l+1}t^m}{(2l+1)!(m-2l-1)!}(\boldsymbol{A}-\sigma_j\boldsymbol{I}_n)^{m-2l-1}\right)(\cos(\omega_j t)\boldsymbol{a}_{j,k}-\sin(\omega_j t)\boldsymbol{b}_{j,k})\right), \tag{3.16}$$

where, for $x \in \mathbb{R}$, $\lfloor x \rfloor$ is greatest integer less than or equal to x and $\lceil x \rceil$ is smallest integer greater than or equal to x.

9. For $j \in \{1, \dots, r\}$ and $k \in \{1, \dots, m_j\}$, let $\xi_{j,k} \colon \mathbb{T} \to \mathsf{V}$ be the function whose components with respect to the basis $\{e_1, \dots, e_n\}$ are the components of $\boldsymbol{\xi}_{j,k}$.
10. For $j \in \{1, \dots, s\}$ and $k \in \{1, \dots, \mu_j\}$, let $\alpha_{j,k}, \beta_{j,k} \colon \mathbb{T} \to \mathsf{V}$ be the functions whose components with respect to the basis $\{e_1, \dots, e_n\}$ are the components of $\boldsymbol{\alpha}_{j,k}$ and $\boldsymbol{\beta}_{j,k}$, respectively.
11. Then the n functions

$$\begin{aligned} &\xi_{j,k}, && j \in \{1, \dots, r\},\ k \in \{1, \dots, m_j\},\\ &\alpha_{j,k}, \beta_{j,k}, && j \in \{1, \dots, s\},\ k \in \{1, \dots, \mu_j\}, \end{aligned}$$

are a basis for $\mathrm{Sol}(F)$. •

Of course, we should verify that the procedure does, indeed, produce a basis for $\mathrm{Sol}(F)$.

3.2.46 Theorem (Basis of solutions for a system of linear homogeneous ordinary differential equations with constant coefficients) *Given a system of linear homogeneous ordinary differential equations*

$$\mathrm{F} \colon \mathbb{T} \times \mathsf{V} \oplus \mathsf{V} \to \mathsf{V}$$

in an n*-dimensional* $\mathbb{R}$*-vector space* V *and with right-hand side*

$$\widehat{\mathrm{F}}(\mathrm{t}, \mathrm{x}) = \mathrm{A}(\mathrm{x}),$$

define n *functions as in Procedure 3.2.45. Then these functions for a basis for* Sol(F).

Proof By virtue of Exercise 3.2.3 we can choose a basis $\{e_1, \dots, e_n\}$ for V and so assume that $\mathsf{V} = \mathbb{R}^n$.

Let us first fix $j \in \{1, \dots, r\}$ and show that $\xi_{j,k}$, $k \in \{1, \dots, m_j\}$, are solutions for $\boldsymbol{F}$. Let $t_0 \in \mathbb{T}$. Let us also fix $k \in \{1, \dots, m_j\}$. By Theorem 3.2.42(ix), the unique solution to the initial value problem

$$\dot{\xi}(t) = \boldsymbol{A}\xi(t), \quad \xi(t_0) = \mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}_{j,k},$$

is

$$t \mapsto \mathrm{e}^{\boldsymbol{A}(t-t_0)}\mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}_{j,k} = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_{j,k},$$

using Theorem 3.2.42(vi) and the obvious fact that the matrices $t\boldsymbol{A}$ and $t_0\boldsymbol{A}$ commute. Now we have

$$\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_{j,k} = \mathrm{e}^{\ell_j t\boldsymbol{I}_n}\mathrm{e}^{(\boldsymbol{A}-\ell_j\boldsymbol{I}_n)t} = \mathrm{e}^{\ell_j t}\mathrm{e}^{(\boldsymbol{A}-\ell_j\boldsymbol{I}_n)t}\boldsymbol{x}_{j,k}$$

using parts (v) and (vi) of Theorem 3.2.42. Now, since $\boldsymbol{x}_{j,k} \in \overline{\mathsf{W}}(\ell_j, \boldsymbol{A})$,

$$\mathrm{e}^{\ell_j t}\mathrm{e}^{(\boldsymbol{A}-\ell_j\boldsymbol{I}_n)t}\boldsymbol{x}_{j,k} = \mathrm{e}^{\ell_j t}\sum_{m=0}^{m_j-1}\frac{(A-\ell_j\boldsymbol{I}_n)^m t^m}{m!}\boldsymbol{x}_{j,k},$$

using Theorem 3.2.42(i). However, this last expression is exactly $\xi_{j,k}(t)$, showing that this is indeed a solution for $\boldsymbol{F}$.

Next we show that, still keeping $j \in \{1, \dots, r\}$ fixed, the m_j solutions $\xi_{j,k}$, $k \in \{1, \dots, m_j\}$, are linearly independent. As we have seen,

$$\xi_{j,k}(t_0) = \mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}_{j,k}, \qquad k \in \{1, \dots, m_j\}.$$

Thus, for $c_1, \dots, c_{m_j} \in \mathbb{R}$, we have

$$\begin{aligned}
& c_1\xi_{j,k}(t_0) + \cdots + c_{m_j}\xi_{j,m_j}(t_0) = \boldsymbol{0} \\
\implies \quad & c_1\mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}_{j,1} + \cdots + c_{m_j}\mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}_{j,m_j} = \boldsymbol{0} \\
\implies \quad & \mathrm{e}^{\boldsymbol{A}t_0}(c_1\boldsymbol{x}_{j,1} + \cdots + c_{m_j}\boldsymbol{x}_{j,m_j}) = \boldsymbol{0} \\
\implies \quad & c_1\boldsymbol{x}_{j,1} + \cdots + c_{m_j}\boldsymbol{x}_{j,m_j} = \boldsymbol{0} \\
\implies \quad & c_1 = \cdots = c_{m_j} = 0,
\end{aligned}$$

since $\boldsymbol{x}_{j,1}, \dots, x_{j,m_j}$ are constructed as being linearly independent. By Corollary 3.2.7 we conclude that $\xi_{j,1}, \dots, \xi_{j,m_j}$ are indeed linearly independent.

Now we fix $j \in \{1, \dots, s\}$ and work with the complex eigenvalue $\lambda_j = \sigma_j + \mathrm{i}\omega_j$. First of all, let us define $\zeta_{j,k} \colon \mathbb{T} \to \mathbb{C}^n$, $k \in \{1, \dots, \mu_j\}$, by

$$\zeta_{j,k} = \mathrm{e}^{\boldsymbol{A}^{\mathbb{C}}t}\boldsymbol{z}_{j,k}.$$

Then, exactly as above for the real eigenvalues, we have

$$\zeta_{j,k}(t) = \mathrm{e}^{\lambda_j t} \sum_{m=0}^{\mu_j - 1} \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^m t^m}{m!} \boldsymbol{z}_{j,k}.$$

Moreover, $\zeta_{j,k}$, $k \in \{1, \dots, \mu_j\}$, are solutions for $\boldsymbol{F}^{\mathbb{C}}$. Therefore, by Lemma 3.2.40, the real and imaginary parts of $\zeta_{j,k}$ are solutions for $\boldsymbol{F}$. To determine the real and imaginary parts, we first make use of the following lemma.

1 Lemma *For a $\mathbb{C}$-vector space V, for $\mathrm{L} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, for $\mathrm{b} \in \mathbb{C}$, and for $\mathrm{m} \in \mathbb{Z}_{\geq 0}$,*

$$(\mathrm{L} + \mathrm{i}\,\mathrm{b}\,\mathrm{id}_{\mathsf{V}})^{\mathrm{m}} = \sum_{\mathrm{j}=0}^{\lceil \frac{\mathrm{m}-1}{2} \rceil} \binom{\mathrm{m}}{2\mathrm{j}} (-1)^{\mathrm{j}} \mathrm{b}^{2\mathrm{j}} \mathrm{L}^{\mathrm{m}-2\mathrm{j}} + \mathrm{i} \sum_{\mathrm{j}=0}^{\lfloor \frac{\mathrm{m}-1}{2} \rfloor} \binom{\mathrm{m}}{2\mathrm{j}+1} (-1)^{\mathrm{j}} (\mathrm{b}^{2\mathrm{j}+1} \mathrm{L}^{\mathrm{m}-2\mathrm{j}-1}),$$

where, for $\mathrm{r}, \mathrm{s} \in \mathbb{Z}$ with $\mathrm{r} \geq \mathrm{s}$, $\binom{\mathrm{r}}{\mathrm{s}} = \frac{\mathrm{r}!}{\mathrm{s}!(\mathrm{r}-\mathrm{s})!}$.

Proof By the Binomial Formula, and since L and id_{V} commute, we have

$$(L + \mathrm{i}\, b\, \mathrm{id}_{\mathsf{V}}) = \sum_{j=0}^{m} \binom{m}{j} (\mathrm{i}\, b)^j L^{m-1}.$$

The stated formula is obtained by separating this expression into its real and imaginary parts. ▼

With the lemma, and some tedious manipulations, one can then verify that

$$\begin{aligned}
\boldsymbol{\alpha}_{j,k}(t) &= \mathrm{Re}\left(\mathrm{e}^{\lambda_j t} \left(\boldsymbol{I}_n + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^{\mu_j - 1}}{(\mu_j - 1)!} \right) \boldsymbol{z}_{j,k} \right), \\
\boldsymbol{\beta}_{j,k}(t) &= \mathrm{Im}\left(\mathrm{e}^{\lambda_j t} \left(\boldsymbol{I}_n + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^{\mu_j - 1}}{(\mu_j - 1)!} \right) \boldsymbol{z}_{j,k} \right)
\end{aligned}$$

for $k \in \{1, \dots, \mu_j\}$. This shows that $\boldsymbol{\alpha}_{j,k}$ and $\boldsymbol{\beta}_{j,k}$ are solutions for $\boldsymbol{F}$ for $k \in \{1, \dots, \mu_j\}$.

Now we verify that

$$\boldsymbol{\alpha}_{j,1}, \dots, \boldsymbol{\alpha}_{j,\mu_j}, \boldsymbol{\beta}_{j,1}, \dots, \boldsymbol{\beta}_{j,\mu_j}$$

are linearly independent. As above in the real case, the complex solutions $\zeta_{j,1}, \dots, \zeta_{j,\mu_j}$ for $\boldsymbol{F}^{\mathbb{C}}$ are linearly independent. Now let $t_0 \in \mathbb{T}$ and

$c_1, \dots, c_{\mu_j}, d_1, \dots, d_{\mu_j} \in \mathbb{R}$, and note that

$$\begin{aligned}
&\sum_{k=1}^{\mu_j} (c_k \boldsymbol{a}_{j,k}(t_0) + d_k \boldsymbol{b}_{j,k}(t_0)) = \boldsymbol{0} \\
\implies &\sum_{k=1}^{\mu_j} (c_k \operatorname{Re}(\boldsymbol{\zeta}_{j,k})(t_0) + d_k \operatorname{Im}(\boldsymbol{\zeta}_{j,k})(t_0)) = \boldsymbol{0} \\
\implies &\sum_{k=1}^{\mu_j} (c_k \operatorname{Re}(\mathrm{e}^{A^{\mathbb{C}} t_0} \boldsymbol{z}_{j,k}) + d_k \operatorname{Im}(\mathrm{e}^{A^{\mathbb{C}} t_0} \boldsymbol{z}_{j,k})) \\
\implies &\sum_{k=1}^{\mu_j} (c_k \mathrm{e}^{A^{\mathbb{C}} t_0} \boldsymbol{a}_{j,k} + d_k \mathrm{e}^{A^{\mathbb{C}} t_0} \boldsymbol{b}_{j,k}) = \boldsymbol{0} \\
\implies &\sum_{k=1}^{\mu_j} (c_k \boldsymbol{a}_{j,k} + d_k \boldsymbol{b}_{j,k}) = \boldsymbol{0} \\
\implies &c_1 = \cdots = c_{\mu_j} = d_1 = \cdots = d_{\mu_j} = 0,
\end{aligned}$$

using the fact that, since $\boldsymbol{A}$ is real, $\mathrm{e}^{A^{\mathbb{C}} t_0}$ is also real and using Proposition 3.2.34(i). This gives the linear independence of

$$\boldsymbol{\alpha}_{j,1}, \dots, \boldsymbol{\alpha}_{j,\mu_j}, \boldsymbol{\beta}_{j,1}, \dots, \boldsymbol{\beta}_{j,\mu_j},$$

as claimed.

Now we have $m_1 + \cdots + m_2 + 2(\mu_1 + \cdots + \mu_2) = n$ solutions for $\boldsymbol{F}$. It remains to show that the collection of all of these solutions are linearly independent. Let us suppose that

$$\begin{aligned}
&\underbrace{c_{1,1}\boldsymbol{\xi}_{1,1}(t) + \cdots + c_{1,m_1}\boldsymbol{\xi}_{1,m_1}(t)}_{\in \overline{\mathsf{W}}(\ell_1, \boldsymbol{A})} + \cdots + \underbrace{c_{r,1}\boldsymbol{\xi}_{r,1}(t) + \cdots + c_{r,m_r}\boldsymbol{\xi}_{r,m_r}(t)}_{\in \overline{\mathsf{W}}(\ell_r, \boldsymbol{A})} \\
&\quad + \underbrace{d_{1,1}\boldsymbol{a}_{1,1}(t) + \cdots + d_{1,\mu_1}\boldsymbol{a}_{1,\mu_1}(t)}_{\in \overline{\mathsf{W}}(\lambda_1, \boldsymbol{A})} + \cdots + \underbrace{d_{s,1}\boldsymbol{a}_{s,1}(t) + \cdots + d_{s,\mu_s}\boldsymbol{a}_{s,\mu_s}(t)}_{\in \overline{\mathsf{W}}(\lambda_s, \boldsymbol{A})} \\
&\quad + \underbrace{e_{1,1}\boldsymbol{b}_{1,1}(t) + \cdots + e_{1,\mu_1}\boldsymbol{b}_{1,\mu_1}(t)}_{\in \overline{\mathsf{W}}(\lambda_1, \boldsymbol{A})} + \cdots + \underbrace{e_{s,1}\boldsymbol{b}_{s,1}(t) + \cdots + e_{s,\mu_s}\boldsymbol{b}_{s,\mu_s}(t)}_{\in \overline{\mathsf{W}}(\lambda_s, \boldsymbol{A})} = \boldsymbol{0},
\end{aligned}$$

for suitable scalar coefficients. Since the generalised eigenspaces intersect in $\{0\}$ by Proposition 3.2.29, and since the generalised eigenspaces are invariant under $\mathrm{e}^{\boldsymbol{A}t}$ for all $t \in \mathbb{T}$ by Theorem 3.2.42(viii), for the preceding equation to hold, each of its components in each of the eigenspaces must be zero, i.e.,

$$c_{j,1}\boldsymbol{\xi}_{j,1}(t) + \cdots + c_{j,m_j}\boldsymbol{\xi}_{j,m_j}(t) = \boldsymbol{0}, \qquad j \in \{1, \dots, r\},$$

and

$$d_{j,1}\boldsymbol{a}_{j,1}(t) + \cdots + d_{j,\mu_j}\boldsymbol{a}_{j,\mu_j}(t) + e_{j,1}\boldsymbol{b}_{j,1}(t) + \cdots + e_{j,\mu_j}\boldsymbol{b}_{j,\mu_j}(t) = \boldsymbol{0}, \qquad j \in \{1, \dots, s\}.$$

This implies that all coefficients must be zero, since we have already shown the linear independence of the solutions with initial conditions in each of the subspaces $\overline{\mathsf{W}}(\ell_l, \boldsymbol{A})$, $j \in \{1, \dots, r\}$, and $\overline{\mathsf{W}}(\lambda_j, \boldsymbol{A})$, $j \in \{1, \dots, s\}$. Thus we have the desired linear independence, and thus the theorem follows. ■

From the proof of the theorem, we provide the following comment on how one might deal with complex eigenvalues in practice.

3.2.47 Remark (Computing solutions associated with complex eigenvalues) The formulae (3.15) and (3.16) of Procedure 3.2.45, while fun to look at, are typically not the best ways to work out solutions associated with complex eigenvalues. However, the proof of the preceding theorem tells us an alternative that is easier in easy examples (although using a computer algebra package is even easier). Indeed, in the proof we saw that

$$\begin{aligned}
\boldsymbol{\alpha}_{j,k}(t) &= \operatorname{Re}\left(\mathrm{e}^{\lambda_j t}\left(\boldsymbol{I}_n + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^{\mu_j - 1}}{(\mu_j - 1)!}\right)\boldsymbol{z}_{j,k}\right),\\
\boldsymbol{\beta}_{j,k}(t) &= \operatorname{Im}\left(\mathrm{e}^{\lambda_j t}\left(\boldsymbol{I}_n + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^{\mu_j - 1}}{(\mu_j - 1)!}\right)\boldsymbol{z}_{j,k}\right)
\end{aligned}$$

for $k \in \{1, \dots, \mu_j\}$. Thus, in practice, one might simply compute

$$\boldsymbol{\zeta}_{j,k}(t) = \mathrm{e}^{\lambda_j t}\left(\boldsymbol{I}_n + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)t}{1!} + \cdots + \frac{(\boldsymbol{A}^{\mathbb{C}} - \lambda_j \boldsymbol{I}_n)^{\mu_j - 1}}{(\mu_j - 1)!}\right)\boldsymbol{z}_{j,k},$$

$k \in \{1, \dots, s\}$, and simply takes its real and imaginary parts as linearly independent solutions. •

We can now give an algorithm for computing, in principle, the operator exponential. The following procedure, while given for computing e^A, obviously may be used as well to compute the state transition matrix $\Phi_A(t, t_0) = \mathrm{e}^{A(t-t_0)}$ for a system of linear homogeneous ordinary differential equations with constant coefficients.

3.2.48 Procedure (Operator exponential) Given an n-dimensional $\mathbb{R}$-vector space V and $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, do the following.

1. Choose a basis $\{e_1, \dots, e_n\}$ and let $\boldsymbol{A}$ be the matrix representative of A. If $\mathsf{V} = \mathbb{R}^n$, one can just take $\boldsymbol{A}$ to be the usual matrix associated with $A \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$.
2. Using Procedure 3.2.45, determine a fundamental set of solutions $\xi_1, \dots, \xi_n$, defined on all of $\mathbb{R}$, for the system of linear homogeneous ordinary differential equations $\boldsymbol{F}$ in $\mathbb{R}^n$ with right-hand side

$$\widehat{\boldsymbol{F}}(t, \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}.$$

3. Define
$$\Xi(t) = \begin{bmatrix} \xi_{1,1}(t) & \xi_{2,1}(t) & \cdots & \xi_{n,1}(t) \\ \xi_{1,2}(t) & \xi_{2,2}(t) & \cdots & \xi_{n,2}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{1,n}(t) & \xi_{2,n}(t) & \cdots & \xi_{n,n}(t) \end{bmatrix},$$
where $\xi_{j,k}$ is the kth component of ξ_j.
4. Using Procedure 3.2.11 calculate
$$\mathrm{e}^{At} = \Phi_A(t, 0) = \Xi(t)\Xi(0)^{-1}.$$
5. We then have e^A as the linear map whose matrix representative is $\mathrm{e}^{\boldsymbol{A}}$. •

3.2.3.7 Some examples Obviously, carrying out Procedure 3.2.45 for a moderately complicated linear transformation A is not something one would want to do more than once a day, and that once a day for at most a week or so. Because I am very manly, I did this four times in one day.

3.2.49 Example (Simple 2 × 2 example) We take $\mathsf{V} = \mathbb{R}^2$ and let $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^2; \mathbb{R}^2)$ be defined by the matrix
$$\boldsymbol{A} = \begin{bmatrix} -7 & 4 \\ -6 & 3 \end{bmatrix}.$$
The characteristic polynomial for $\boldsymbol{A}$ is
$$P_A = X^2 + 4X + 3 = (X + 1)(X + 3).$$
Thus the eigenvalues for $\boldsymbol{A}$ are $\ell_1 = -1$ and $\ell_2 = -3$. Since the eigenvalues are distinct, the algebraic and geometric multiplicities will be equal, and the generalised eigenvectors will simply be eigenvectors. An eigenvector for $\ell_1 = -1$ is $\boldsymbol{x}_{1,1} = (2, 3)$ and an eigenvector for $\ell_2 = -3$ is $\boldsymbol{x}_{2,1} = (1, 1)$. Procedure 3.2.45 then gives two linearly independent solutions as
$$\xi_{1,1}(t) = \mathrm{e}^{-t}\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \xi_{2,1}(t) = \mathrm{e}^{-3t}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$
Thus we have determined a fundamental matrix to be
$$\Xi(t) = \begin{bmatrix} 2\mathrm{e}^{-t} & \mathrm{e}^{-3t} \\ 3\mathrm{e}^{-2t} & \mathrm{e}^{-3t} \end{bmatrix},$$
by assembling the linearly independent solutions into the columns of this matrix. It is then an easy calculation to arrive at
$$\mathrm{e}^{At} = \Xi(t)\Xi(0)^{-1} = \begin{bmatrix} 3\mathrm{e}^{-3t} - 2\mathrm{e}^{-t} & -2\mathrm{e}^{-3t} + 2\mathrm{e}^{-t} \\ 3\mathrm{e}^{-3t} - 3\mathrm{e}^{-t} & -2\mathrm{e}^{-3t} + 3\mathrm{e}^{-t} \end{bmatrix}$$
•

3.2.50 Example (A 3 × 3 example with multiplicity) We take $V = \mathbb{R}^3$ with the linear map $A \in L(\mathbb{R}^3; \mathbb{R}^3)$ determined by the matrix A more interesting case is the following:

$$A = \begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Since the matrix is upper triangular, the eigenvalues are the diagonal elements: $\ell_1 = -2$ and $\ell_2 = -1$. The algebraic multiplicity of ℓ_1 is 2. However, we readily see that $\dim_{\mathbb{R}}(\ker(\ell_1 I_3 - A)) = 1$ and so the geometric multiplicity is 1. So we need to compute generalised eigenvectors in this case. We have

$$(A - \lambda_1 I_3)^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and the generalised eigenvectors span the kernel of this matrix, and so we may take $x_{1,1} = (1, 0, 0)$ and $x_{1,2} = (0, 1, 0)$ as generalised eigenvectors. Applying Procedure 3.2.45 gives

$$\begin{aligned} \xi_{1,1}(t) &= \mathrm{e}^{-2t} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\mathrm{e}^{-2t} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{e}^{-2t} \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \xi_{1,2}(t) &= \mathrm{e}^{-2t} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t\mathrm{e}^{-2t} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} t\mathrm{e}^{-2t} \\ \mathrm{e}^{-2t} \\ 0 \end{bmatrix}. \end{aligned}$$

Finally we determine that $x_{2,1} = (0, 0, 1)$ is an eigenvector corresponding to $\ell_2 = -1$, and so this gives the solution

$$\xi_{2,1}(t) = \begin{bmatrix} 0 \\ 0 \\ \mathrm{e}^{-t} \end{bmatrix}.$$

Thus we arrive at our three linearly independent solutions. We assemble these into the columns of a matrix to determine a fundamental matrix

$$\Xi(t) = \begin{bmatrix} \mathrm{e}^{-2t} & t\mathrm{e}^{-2t} & 0 \\ 0 & \mathrm{e}^{-2t} & 0 \\ 0 & 0 & \mathrm{e}^{-t} \end{bmatrix}.$$

It so happens that in this example we lucked out and $\mathrm{e}^{At} = \Xi(t)$ since $\Xi(0) = I_3$. •

3.2.51 Example (A simple example with complex roots) Here we take $\mathsf{V} = \mathbb{R}^3$ with $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^3; \mathbb{R}^3)$ determined by the matrix

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

The characteristic polynomial is

$$P_{\boldsymbol{A}} = X^3 + 4X^2 + 6X + 4.$$

One ascertains that the eigenvalues are then $\lambda_1 = -1 + \mathrm{i}$, $\bar{\lambda}_1 = -1 - \mathrm{i}$, $\ell_1 = -2$. Let's deal with the complex root first, using Remark 3.2.47 rather than the complicated formulae (3.15) and (3.16) of Procedure 3.2.45 for complex eigenvalues. We have

$$\boldsymbol{A} - \lambda_1 \boldsymbol{I}_3 = \begin{bmatrix} -\mathrm{i} & 1 & 0 \\ -1 & -\mathrm{i} & 0 \\ 0 & 0 & -1 - \mathrm{i} \end{bmatrix},$$

from which we glean that an eigenvector is $\boldsymbol{z}_{1,1} = (-\mathrm{i}, 1, 0)$. Using Remark 3.2.47, the complex solution is then

$$\boldsymbol{\zeta}_{1,1}(t) = \mathrm{e}^{(-1+\mathrm{i})t} \begin{bmatrix} -\mathrm{i} \\ 1 \\ 0 \end{bmatrix}.$$

Using Euler's formula, $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$, we have

$$\boldsymbol{\zeta}_{1,1}(t) = \mathrm{e}^{-t} \begin{bmatrix} -\mathrm{i}\cos t + \sin t \\ \cos t + \mathrm{i}\sin t \\ 0 \end{bmatrix} = \mathrm{e}^{-t} \begin{bmatrix} \sin t \\ \cos t \\ 0 \end{bmatrix} + \mathrm{i}\mathrm{e}^{-t} \begin{bmatrix} -\cos t \\ \sin t \\ 0 \end{bmatrix},$$

thus giving

$$\boldsymbol{\alpha}_{1,1}(t) = \mathrm{e}^{-t} \begin{bmatrix} \sin t \\ \cos t \\ 0 \end{bmatrix}, \quad \boldsymbol{\beta}_{1,1}(t) = \mathrm{e}^{-t} \begin{bmatrix} -\cos t \\ \sin t \\ 0 \end{bmatrix}.$$

Corresponding to the real eigenvalue ℓ_1 we readily determine that

$$\boldsymbol{\xi}_{1,1} = \mathrm{e}^{-2t} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

is a corresponding solution. This gives three linearly independent real solutions $\boldsymbol{\alpha}_{1,1}(t)$, $\boldsymbol{\beta}_{1,1}(t)$, and $\boldsymbol{\xi}_{1,1}(t)$. Putting these into the columns of a matrix gives a fundamental matrix

$$\boldsymbol{\Xi}(t) = \begin{bmatrix} \mathrm{e}^{-t}\sin t & -\mathrm{e}^{-t}\cos t & 0 \\ \mathrm{e}^{-t}\cos t & \mathrm{e}^{-t}\sin t & 0 \\ 0 & 0 & \mathrm{e}^{-2t} \end{bmatrix}.$$

A straightforward computation yields

$$\mathrm{e}^{At} = \Xi(t)\Xi(0)^{-1} = \begin{bmatrix} \mathrm{e}^{-t}\cos t & \mathrm{e}^{-t}\sin t & 0 \\ -\mathrm{e}^{-t}\sin t & \mathrm{e}^{-t}\cos t & 0 \\ 0 & 0 & \mathrm{e}^{-2t} \end{bmatrix}. \qquad \bullet$$

3.2.52 Example (An example of complex roots with multiplicity) Our final example has $\mathsf{V} = \mathbb{R}^4$ and $A \in \mathrm{L}(\mathbb{R}^4; \mathbb{R}^4)$ determined by the matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

The eigenvalues are determined to be $\lambda_1 = \mathrm{i}$ and $\bar{\lambda}_1 = -\mathrm{i}$, both with algebraic multiplicity 2. One readily determines that the kernel of $i\mathbf{I}_4 - A$ is one-dimensional, and so the geometric multiplicity of these eigenvalues is just 1. Thus we need to compute complex generalised eigenvectors. We compute

$$(A - \mathrm{i}\mathbf{I}_4)^2 = 2\begin{bmatrix} -1 & -\mathrm{i} & -\mathrm{i} & 1 \\ \mathrm{i} & -1 & -1 & -\mathrm{i} \\ 0 & 0 & -1 & -\mathrm{i} \\ 0 & 0 & \mathrm{i} & -1 \end{bmatrix}$$

and one checks that $z_{1,1} = (0, 0, -\mathrm{i}, 1)$ and $z_{1,2} = (-\mathrm{i}, 1, 0, 0)$ are two linearly independent generalised eigenvectors. We compute

$$(A - \mathrm{i}\mathbf{I}_4)z_{1,1} = \begin{bmatrix} -\mathrm{i} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad (A - \mathrm{i}\mathbf{I}_4)z_{1,2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We now determine the two linearly independent real solutions corresponding to $z_{1,1}$. We have

$$\begin{aligned} \zeta_{1,1}(t) &= \mathrm{e}^{\mathrm{i}t}(u_1 + t(A - \mathrm{i}\mathbf{I}_4)z_{1,1}) = \mathrm{e}^{\mathrm{i}t}\begin{bmatrix} 0 \\ 0 \\ -\mathrm{i} \\ 1 \end{bmatrix} + t\mathrm{e}^{\mathrm{i}t}\begin{bmatrix} -\mathrm{i} \\ 1 \\ 0 \\ 0 \end{bmatrix} \\ &= (\cos t + \mathrm{i}\sin t)\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \mathrm{i}\begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathrm{i}t\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} t\sin t \\ t\cos t \\ \sin t \\ \cos t \end{bmatrix} + \mathrm{i}\begin{bmatrix} -t\cos t \\ t\sin t \\ -\cos t \\ \sin t \end{bmatrix}. \end{aligned}$$

Therefore,

$$\boldsymbol{\alpha}_{1,1}(t) = \begin{bmatrix} t\sin t \\ t\cos t \\ \sin t \\ \cos t \end{bmatrix}, \quad \boldsymbol{\beta}_{1,1}(t) = \begin{bmatrix} -t\cos t \\ t\sin t \\ -\cos t \\ \sin t \end{bmatrix}.$$

For $\boldsymbol{z}_{2,1}$ we have

$$\begin{aligned} \boldsymbol{\zeta}_{1,2}(t) &= \mathrm{e}^{\mathrm{i}t}(\boldsymbol{u}_2 + t(\boldsymbol{A} - \mathrm{i}\boldsymbol{I}_4)\boldsymbol{u}_2) = \mathrm{e}^{\mathrm{i}t}\begin{bmatrix} -\mathrm{i} \\ 1 \\ 0 \\ 0 \end{bmatrix} \\ &= (\cos t + \mathrm{i}\sin t)\left(\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathrm{i}\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \sin t \\ \cos t \\ 0 \\ 0 \end{bmatrix} + \mathrm{i}\begin{bmatrix} -\cos t \\ \sin t \\ 0 \\ 0 \end{bmatrix}, \end{aligned}$$

and so we have

$$\boldsymbol{\alpha}_{1,2}(t) = \begin{bmatrix} \sin t \\ \cos t \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\beta}_{1,2}(t) = \begin{bmatrix} -\cos t \\ \sin t \\ 0 \\ 0 \end{bmatrix}.$$

Thus we have the four real linearly independent solutions $\boldsymbol{\alpha}_{1,1}$, $\boldsymbol{\alpha}_{1,2}$, $\boldsymbol{\beta}_{1,1}$, and $\boldsymbol{\beta}_{1,2}$. The corresponding fundamental matrix is

$$\boldsymbol{\Xi}(t) = \begin{bmatrix} t\sin t & -t\cos t & \sin t & -\cos t \\ t\cos t & t\sin t & \cos t & \sin t \\ \sin t & -\cos t & 0 & 0 \\ \cos t & \sin t & 0 & 0 \end{bmatrix}.$$

A little manipulation gives

$$\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{\Xi}(t)\boldsymbol{\Xi}(0)^{-1} = \begin{bmatrix} \cos t & \sin t & t\cos t & t\sin t \\ -\sin t & \cos t & -t\sin t & t\cos t \\ 0 & 0 & \cos t & \sin t \\ 0 & 0 & -\sin t & \cos t \end{bmatrix}. \qquad \bullet$$

Exercises

3.2.1 Show that the definition of "class C^r" and "rth-derivative" in Definition 3.2.1 do not depend on the basis chosen.
Hint: *Use the change of basis formula (1.24).*

3.2.2 Show that the definition of "class C^r" and "rth-derivative" in Definition 3.2.2 do not depend on the basis chosen.
Hint: *Use the change of basis formula (1.25).*

3.2.3 Let V be an n-dimensional $\mathbb{R}$-vector space and let F be a system of linear ordinary differential equations in V with right-hand side

$$\widehat{F}(t,x) = A(t)(x) + b(t)$$

for $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ and $b\colon \mathbb{T} \to \mathsf{V}$. Let $\{e_1,\dots,e_n\}$ be a basis for V and write

$$b(t) = \sum_{j=1}^{n} b_j(t)e_j, \quad A(t)(e_j) = \sum_{k=1}^{n} A_{kj}(t)e_k, \qquad j \in \{1,\dots,n\},$$

for functions $b_j\colon \mathbb{T} \to \mathbb{R}$, $j \in \{1,\dots,n\}$, and $A_{kj}\colon \mathbb{T} \to \mathbb{R}$, $j,k \in \{1,\dots,n\}$. This defines $\boldsymbol{b}\colon \mathbb{T} \to \mathbb{R}^n$ and $\boldsymbol{A}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$. Denote by $\boldsymbol{F}$ the system of linear ordinary differential equations in $\mathbb{R}^n$ given by

$$\boldsymbol{F}(t,\boldsymbol{x},\boldsymbol{x}^{(1)}) = \boldsymbol{x}^{(1)} - \boldsymbol{A}(t)\boldsymbol{x} - \boldsymbol{b}(t).$$

Answer the following questions.

(a) Show that $\xi\colon \mathbb{T}' \to \mathsf{V}$ is a solution for F if and only if the function $\boldsymbol{\xi}\colon \mathbb{T}' \to \mathbb{R}^n$, defined by

$$\xi(t) = \sum_{j=1}^{n} \xi_j(t)e_j,$$

is a solution for $\boldsymbol{F}$.

Now let $\{\tilde{e}_1,\dots,\tilde{e}_n\}$ be another basis for V and let $\boldsymbol{P}$ be the change of basis matrix defined by

$$\tilde{e}_j = \sum_{k=1}^{n} P_{kj}e_k, \qquad j \in \{1,\dots,n\}.$$

Define $\tilde{\boldsymbol{b}}\colon \mathbb{T} \to \mathbb{R}^n$, $\tilde{\boldsymbol{A}}\colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$, and $\tilde{\boldsymbol{F}}$ as above, for this new basis.

(b) Show that $\tilde{\boldsymbol{b}}(t) = \boldsymbol{P}\boldsymbol{b}(t)$ and $\tilde{\boldsymbol{A}}(t) = \boldsymbol{P}^{-1}\boldsymbol{A}(t)\boldsymbol{P}$ for every $t \in \mathbb{T}$.
Hint: *Use the change of basis formulae (1.24) and (1.26).*

(c) Show that, if $\boldsymbol{\xi}\colon \mathbb{T}' \to \mathbb{R}^n$ is a solution for $\boldsymbol{F}$, then $\tilde{\boldsymbol{\xi}}\colon \mathbb{T}' \to \mathbb{R}^n$ is a solution for $\tilde{\boldsymbol{F}}$ if and only if $\tilde{\boldsymbol{\xi}}(t) = \boldsymbol{P}^{-1}\boldsymbol{\xi}(t)$ for every $t \in \mathbb{T}$.

3.2.4 Let V be an n-dimensional $\mathbb{R}$-vector space and let F be a system of linear homogeneous ordinary differential equations with right-hand side

$$\widehat{F}(t,x) = A(t)(x)$$

for a continuous map $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$. Let $\{e_1,\dots,e_n\}$ be a basis for V and let $\boldsymbol{A}(t)$ be the matrix representative for $A(t)$, $t \in \mathbb{T}$, and let $\boldsymbol{F}$ be the corresponding system of linear homogeneous ordinary differential equations in $\mathbb{R}^n$ with right-hand side

$$\widehat{\boldsymbol{F}}(t,\boldsymbol{x}) = \boldsymbol{A}(t)\boldsymbol{x}.$$

cf. Exercise 3.2.3.

(a) Show that, for every $t, t_0 \in \mathbb{T}$, the matrix representative of $\Phi_A(t, t_0)$ is $\Phi_A(t, t_0)$.

Now let $\{\tilde{e}_1, \dots, \tilde{e}_n\}$ be another basis for V and let $\boldsymbol{P}$ be the change of basis matrix defined by

$$\tilde{e}_j = \sum_{k=1}^{n} P_{kj} e_k, \qquad j \in \{1, \dots, n\}.$$

Define $\tilde{\boldsymbol{A}} \colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ and $\tilde{\boldsymbol{F}}$ as above, for this new basis.

(b) Show that, for every $t, t_0 \in \mathbb{T}$,

$$\Phi_{\tilde{A}}(t, t_0) = \boldsymbol{P}^{-1} \Phi_A(t, t_0) \boldsymbol{P}.$$

3.2.5 Consider the system of linear homogeneous ordinary differential equations F with right-hand side equation (3.4) and suppose that $A \colon \mathbb{T} \to \mathbb{R}$ is continuous. Recall from the proof of Theorem 3.2.6 the maps

$$\begin{aligned} \sigma_t \colon \mathrm{Sol}(F) &\to \mathsf{V} \\ \xi &\mapsto \xi(t) \end{aligned}, \qquad t \in \mathbb{T},$$

that were shown to be isomorphisms.

(a) Show that

$$\Phi_A(t, t_0) = \sigma_t \circ \sigma_{t_0}^{-1}$$

for each $t, t_0 \in \mathbb{T}$.

(b) Use this to give alternative proofs of parts (iv) and (v) of Theorem 3.2.9.

3.2.6 Consider a scalar linear homogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_0(t)x - a_1(t)x^{(1)} - \cdots - a_{k-1}(t)x^{(k-1)},$$

for continuous functions $a_0, a_1, \dots, a_{k-1} \colon \mathbb{T} \to \mathbb{R}$.

(a) Following Exercise 1.3.23, convert this kth order scalar system into a first order system $\boldsymbol{F}_1$ of linear homogeneous ordinary differential equations in $\mathbb{R}^k$, i.e., find the matrix function $\boldsymbol{A} \colon \mathbb{T} \to \mathrm{L}(\mathbb{R}^k; \mathbb{R}^k)$ in this case.

(b) For a solution $t \mapsto \xi(t)$ for F, what is the corresponding solution $t \mapsto \boldsymbol{\xi}(t)$ for $\boldsymbol{F}_1$?

(c) Show that, given a fundamental set of solutions $\{\xi_1, \dots, \xi_k\}$ for F, the solutions $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_k\}$ for $\boldsymbol{F}_1$ from part (b) are a fundamental set of solutions for $\boldsymbol{F}_1$.

(d) Show that

$$\det \boldsymbol{\Phi}_A(t, t_0) = \frac{\mathrm{W}(\xi_1, \dots, \xi_n)(t)}{\mathrm{W}(\xi_1, \dots, \xi_n)(t_0)}.$$

(e) Show that

$$\mathrm{W}(\xi_1,\dots,\xi_k)(t) = \mathrm{W}(\xi_1,\dots,\xi_k)(t_0)\mathrm{e}^{-\int_{t_0}^{t} a_{k-1}(\tau)\,\mathrm{d}\tau}.$$

3.2.7 Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and let U, V, and W be $\mathbb{F}$-vector spaces. For $L \in \mathrm{L}(\mathsf{U};\mathsf{V})$ and $M \in \mathrm{L}(\mathsf{V};\mathsf{W})$, show that $(M \circ L)^* = L^* \circ M^*$.

3.2.8 Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, let V be an n-dimensional $\mathbb{F}$-vector space, and let $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$. Let $\mathscr{B} = \{e_1,\dots,e_n\}$ and $\mathscr{B}' = \{\tilde{e}_1,\dots,\tilde{e}_n\}$ be bases for V and let $\boldsymbol{P}$ be the change of basis matrix defined by

$$\tilde{e}_j = \sum_{k=1}^{n} P_{kj} e_k, \qquad j \in \{1,\dots,n\}.$$

Let $\boldsymbol{L}$ and $\tilde{\boldsymbol{L}}$ be the matrix representatives for L in the $\mathscr{B}$ and $\tilde{\mathscr{B}}$, respectively.

(a) Use part (b) of Exercise 3.2.4 to show that

$$[\mathrm{e}^L]^{\tilde{\mathscr{B}}}_{\tilde{\mathscr{B}}} = \boldsymbol{P}^{-1}[\mathrm{e}^L]^{\mathscr{B}}_{\mathscr{B}}\boldsymbol{P}.$$

(b) Use Theorem 3.2.42(i) and Proposition 3.2.44 to arrive at the same conclusion.

3.2.9 Consider the first-order scalar linear homogeneous ordinary differential equation with right-hand side $\widehat{F}(t,x) = a(t)x$ for a continuous function $a\colon \mathbb{T} \to \mathbb{R}$. Determine the state-transition map in this case, thinking of this as a system of linear homogeneous ordinary differential equations in the one-dimensional vector space $\mathbb{R}$.

3.2.10 Let $\lambda \in \mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and consider the linear map $\boldsymbol{A} \in \mathrm{L}(\mathbb{F}^n;\mathbb{F}^n)$ determined by the $n \times n$-matrix

$$\boldsymbol{A} = \left[\begin{array}{ccccc|ccccc}
\lambda & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
0 & \lambda & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & \cdots & \lambda & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
0 & 0 & \cdots & 0 & \lambda & 0 & 0 & 0 & \cdots & 0 & 0\\
\hline
0 & 0 & \cdots & 0 & 0 & \lambda & 1 & 0 & \cdots & 0 & 0\\
0 & 0 & \cdots & 0 & 0 & 0 & \lambda & 1 & \cdots & 0 & 0\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \lambda & \cdots & 0 & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & \lambda & 1\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \lambda
\end{array}\right].$$

We suppose the lower right block is a $k \times k$-matrix and the upper left block, therefore, is a $(n-k) \times (n-k)$-matrix.

Answer the following questions.

(a) What are the eigenvalues of $\boldsymbol{A}$?
(b) For each of the eigenvalues of $\boldsymbol{A}$, determine its algebraic multiplicity.
(c) For each of the eigenvalues of $\boldsymbol{A}$, determine its eigenspace.
(d) For each of the eigenvalues of $\boldsymbol{A}$, determine its geometric multiplicity.
(e) For each of the eigenvalues of $\boldsymbol{A}$, determine its generalised eigenspace.
(f) For each of the eigenvalues ℓ of $\boldsymbol{A}$, determine the smallest $m \in \mathbb{Z}_{>0}$ for which $\overline{\mathsf{W}}(\ell, \boldsymbol{A}) = \ker((\boldsymbol{A} - \ell \boldsymbol{I}_n)^m)$.

3.2.11 For each of the following linear maps $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$, given by an $n \times n$-matrix, determine the

1. eigenvalues,
2. eigenvectors,
3. generalised eigenvectors,
4. algebraic multiplicities of each eigenvalue, and
5. geometric multiplicities of each eigenvalue.

Here are the linear maps:

(a) $\boldsymbol{A} = \begin{bmatrix} 2 & -5 \\ 0 & 3 \end{bmatrix}$;

(b) $\boldsymbol{A} = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix}$;

(c) $\boldsymbol{A} = \begin{bmatrix} 4 & -1 \\ 4 & 0 \end{bmatrix}$;

(d) $\boldsymbol{A} = \begin{bmatrix} 5 & 0 & -6 \\ 0 & 2 & 0 \\ 3 & 0 & -4 \end{bmatrix}$;

(e) $\boldsymbol{A} = \begin{bmatrix} 5 & 0 & -6 \\ 1 & 2 & -1 \\ 3 & 0 & -4 \end{bmatrix}$;

(f) $\boldsymbol{A} = \begin{bmatrix} 4 & 2 & -4 \\ 2 & 0 & -4 \\ 2 & 2 & -2 \end{bmatrix}$;

(g) $\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 3 & -1 & 3 \\ 0 & 1 & 2 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$;

(h) $\boldsymbol{A} = \begin{bmatrix} -7 & 0 & 0 & -4 \\ -13 & -2 & -1 & -8 \\ 6 & 1 & 0 & 4 \\ 15 & 1 & 0 & 9 \end{bmatrix}$;

(i) $\boldsymbol{A} = \begin{bmatrix} 1 & 4 & -2 & 0 & 9 \\ 0 & -2 & 1 & 2 & -6 \\ -2 & 4 & -1 & 3 & 0 \\ -9 & 4 & 1 & 0 & 2 \\ 4 & 0 & 3 & -1 & 3 \end{bmatrix}$.

3.2.12 For each of the following $\mathbb{R}^n$-valued functions ξ of time, indicate whether they can be the solution of a system of linear homogeneous ordinary differential equations with constant coefficients. If they can be, find a matrix $\boldsymbol{A}$ for which the function satisfies $\dot{\xi}(t) = \boldsymbol{A}\xi(t)$. If they cannot be, explain why not.

(a) $\xi(t) = (\mathrm{e}^t, \mathrm{e}^{-t})$;
(b) $\xi(t) = (\cos(t) - \sin(t), \cos(t) + \sin(t))$;
(c) $\xi(t) = (\mathrm{e}^t + \mathrm{e}^{2t}, 0, 0)$;
(d) $\xi(t) = (t, 0, 1)$;

(e) $\xi(t) = (e^t, e^t + e^{2t}, 0)$;
(f) $\xi(t) = (\cos(t) + \sin(t), \cos(t) + \sin(t))$.

3.2.13 Let F be a scalar linear homogeneous ordinary differential equation with right-hand side

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x,$$

for $a_0, a_1, \dots, a_{k-1} \in \mathbb{R}$.

(a) Following Exercise 1.3.23, convert F into a first-order system of linear homogeneous ordinary differential equations $\boldsymbol{F}_1$ in $\mathbb{R}^k$ and with right-hand side

$$\widehat{\boldsymbol{F}}_1(t, \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x},$$

explicitly identifying $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^k; \mathbb{R}^k)$.
(b) Show that the characteristic polynomial P_F of F is the same as the characteristic polynomial P_A of $\boldsymbol{A}$.

3.2.14 Determine $e^{\boldsymbol{A}t}$ for the following linear transformations $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ from Exercise 3.2.11.

3.2.15 For the linear transformations $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ of Exercise 3.2.14, determine the solution to the initial value problem

$$\dot{\xi}(t) = \boldsymbol{A}\xi(t), \quad \xi(0) = \boldsymbol{x}_0,$$

with $\boldsymbol{x}_0$ as follows:

(a) $\boldsymbol{x}_0 = (0, 1)$;
(b) $\boldsymbol{x}_0 = (2, -3)$;
(c) $\boldsymbol{x}_0 = (1, 1)$;
(d) $\boldsymbol{x}_0 = (-3, -1, 0)$;
(e) $\boldsymbol{x}_0 = (1, 0, 1)$;
(f) $\boldsymbol{x}_0 = (4, 1, 2)$;
(g) $\boldsymbol{x}_0 = (1, -1, 0, 1)$;
(h) $\boldsymbol{x}_0 = (-1, -1, 3, -2)$;
(i) $\boldsymbol{x}_0 = (0, 0, 0, 0, 0)$.

3.2.16 For the scalar linear homogeneous ordinary differential equations of Exercise 2.2.10, do the following:

(a) convert these to a first-order system of linear homogeneous ordinary differential equations, explicitly identifying A;
(b) using the fundamental solutions obtained during the solution of the problems from Exercise 2.2.10, compute $e^{\boldsymbol{A}t}$;
(c) solve the initial value problems from Exercise 2.2.10 using the operator exponential.

3.2.17 Let $\ell \in \mathbb{R}$ and $k \in \mathbb{Z}_{>0}$. Consider the Jordan block

$$\boldsymbol{J}(\ell, k) = \begin{bmatrix} \ell & 1 & 0 & \cdots & 0 & 0 \\ 0 & \ell & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \ell & 1 \\ 0 & 0 & 0 & \cdots & 0 & \ell \end{bmatrix}.$$

Do the following.

(a) Solve the initial value problems

$$\dot{\xi}_j(t) = \boldsymbol{J}(\ell, k)\xi_j(t), \quad \xi_j(0) = \boldsymbol{e}_j, \qquad \text{IVP}_j$$

$j \in \{1, \dots, k\}$, recursively, first by solving IVP_k, then by solving IVP_{k-1}, and so on. At each stage you should be solving a scalar linear, possibly inhomogeneous, ordinary differential equation, and so the methods of Sections 2.2.2 and 2.3.2 can be used.

(b) Use your calculations to determine $\mathrm{e}^{\boldsymbol{J}(\ell,k)t}$.

Alternatively, compute $\mathrm{e}^{\boldsymbol{J}(\ell,k)t}$ as follows.

(c) What are the eigenvalues of $\boldsymbol{J}(\ell, k)$?

(d) What are the geometric and algebraic multiplicities of the eigenvalues?

(e) Compute

$$(\boldsymbol{J}(\ell, k) - \ell \boldsymbol{I}_n)^j, \qquad j \in \{0, 1, \dots, k-1\},$$

probably using mathematical induction on j.

(f) Use your answers from the preceding three questions to explicitly compute $\mathrm{e}^{\boldsymbol{J}(\ell,k)t}$ using Procedure 3.2.45.

Section 3.3

Systems of linear inhomogeneous ordinary differential equations

In this section we extend our discussion of homogeneous equations in Section 3.2 to inhomogeneous equations. Thus we are talking about systems of linear ordinary differential equations F in a finite-dimensional $\mathbb{R}$-vector space V with right-hand side

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V} \\ (t, x) &\mapsto A(t)(x) + b(t) \end{aligned} \tag{3.17}$$

for maps $b\colon \mathbb{T} \to \mathsf{V}$ and $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$. In our treatment of scalar equations in Section 3.2, we were gave no fewer than three methods for working with inhomogeneous equations, two general methods (using Wronskians in Section 2.3.1.2 and the theory of Green's function in Section 2.3.1.3) and one method that only works for inhomogeneous terms that are pretty uninteresting (the "method of undetermined coefficients in Section 2.3.2.1). We shall not be so expansive for systems of linear inhomogeneous equations, and shall really only consider "the" method for working with such equations, since this method is as tractable as any other method in practice (which is to say, not very tractable at all, barring the use of a computer algebra package), and is exceptionally powerful in developing the theory of systems of linear ordinary differential equations.

As we have done in all preceding developments of linear ordinary differential equations, we work first in the general time-varying case, and then in the case of constant coefficients.

3.3.1 Equations with time-varying coefficients

We state the, by now, more or less obvious results concerning existence and uniqueness, now for systems of linear inhomogeneous ordinary differential equations.

3.3.1 Proposition (Local existence and uniqueness of solutions for systems of linear inhomogeneous ordinary differential equations) *Consider the system of linear inhomogeneous ordinary differential equations* F *with right-hand side (3.17) and suppose that* $\mathrm{b}\colon \mathbb{T} \to \mathsf{V}$ *and* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *are continuous. Let* $(\mathrm{t}_0, \mathrm{x}_0) \in \mathbb{T} \times \mathsf{V}$. *Then there exists an interval* $\mathbb{T}' \subseteq \mathbb{T}$ *and a map* $\xi\colon \mathbb{T}' \to \mathsf{V}$ *of class* C^1 *that is a solution for* F *and which satisfies* $\xi(\mathrm{t}_0) = \mathrm{x}_0$. *Moreover, if* $\tilde{\mathbb{T}}' \subseteq \mathbb{T}$ *is another subinterval and if* $\tilde{\xi}\colon \tilde{\mathbb{T}}' \to \mathsf{V}$ *is another* C^1*-solution for* F *satisfying* $\tilde{\xi}(\mathrm{t}_0) = \mathrm{x}_0$, *then* $\tilde{\xi}(\mathrm{t}) = \xi(\mathrm{t})$ *for every* $\mathrm{t} \in \tilde{\mathbb{T}}' \cap \mathbb{T}'$.

Proof By Proposition 3.2.4, there exists a compact interval $\mathbb{T}' \subseteq \mathbb{T}$ and a solution

$\xi_{\mathrm{h}} \colon \mathbb{T} \to \mathsf{V}$ for F_{h} satisfying $\xi_{\mathrm{h}}(t_0) = x_0$. Now define

$$\begin{aligned} \xi \colon \mathbb{T}' &\to \mathsf{V} \\ t &\mapsto \Phi_A(t, t_0)(x_0) + \int_{t_0}^{t} \Phi(t, \tau)(b(\tau))\, \mathrm{d}\tau. \end{aligned}$$

Note that the integral defining ξ exists since both $\tau \mapsto \Phi_A(t, \tau)$ and $\tau \mapsto b(\tau)$ are continuous, the first holding for every $t \in \mathbb{T}'$. In order to verify that ξ so defined is a solution for F, we will use the following lemma.

1 Lemma *Let $\mathbb{T} \subseteq \mathbb{R}$ be a compact interval and let $\mathbf{g} \colon \mathbb{T} \times \mathbb{T} \to \mathbb{R}^{\mathrm{n}}$ have the following properties:*

(i) for $\mathrm{t} \in \mathbb{T}$, the map $\tau \mapsto \mathbf{g}(\mathrm{t}, \tau)$ is continuous;

(ii) for $\tau \in \mathbb{T}$, the map $\mathrm{t} \mapsto \mathbf{g}(\mathrm{t}, \tau)$ is differentiable;

(iii) there exists $\mathrm{M}_1 \in \mathbb{R}_{>0}$ such that $\|\mathbf{g}(\mathrm{t}, \tau)\| \leq \mathrm{M}_1$ for every $\mathrm{t}, \tau \in \mathbb{T}$;

(iv) there exists $\mathrm{M}_2 \in \mathbb{R}_{>0}$ such that $\|\frac{\partial \mathrm{g}_{\mathrm{j}}}{\partial \mathrm{t}}(\mathrm{t}, \tau)\| \leq \mathrm{M}_2$ for every $\mathrm{j} \in \{1, \ldots, \mathrm{n}\}$ and $\mathrm{t}, \tau \in \mathbb{T}$.

Then, for any $\mathrm{t}_0 \in \mathbb{T}$, the function

$$\begin{aligned} \mathbf{G} \colon \mathbb{T} &\to \mathbb{R}^{\mathrm{n}} \\ \mathrm{t} &\mapsto \int_{\mathrm{t}_0}^{\mathrm{t}} \mathbf{g}(\mathrm{t}, \tau)\, \mathrm{d}\tau \end{aligned}$$

is differentiable and

$$\frac{\mathrm{d}\mathbf{G}}{\mathrm{dt}}(\mathrm{t}) = \int_{\mathrm{t}_0}^{\mathrm{t}} \frac{\partial \mathbf{g}}{\partial \mathrm{t}}(\mathrm{t}, \tau)\, \mathrm{d}\tau + \mathbf{g}(\mathrm{t}, \mathrm{t}).$$

Proof Continuity of $\tau \mapsto g(t, \tau)$ ensures that the integral in the definition of $\boldsymbol{G}$ exists. Consider the function

$$\begin{aligned} \tilde{\boldsymbol{G}} \colon \mathbb{T} \times \mathbb{T} &\to \mathbb{R}^{n} \\ (t_1, t_2) &\mapsto \int_{t_0}^{t_1} \boldsymbol{g}(t_2, \tau)\, \mathrm{d}\tau. \end{aligned}$$

By the Fundamental Theorem of Calculus, $\tilde{\boldsymbol{G}}$ is differentiable with respect to t_1 and

$$\frac{\partial \tilde{\boldsymbol{G}}}{\partial t_1}(t_1, t_2) = \boldsymbol{g}(t_2, t_1).$$

The assumptions on $\boldsymbol{g}$ ensure that we can differentiate $\tilde{\boldsymbol{G}}$ with respect to t_2 inside the integral:

$$\frac{\partial \tilde{\boldsymbol{G}}}{\partial t_2}(t_1, t_2) = \int_{t_0}^{t_1} \frac{\partial \boldsymbol{g}}{\partial t_2}(t_2, \tau)\, \mathrm{d}\tau.$$

Now define

$$\begin{aligned}\boldsymbol{\delta}\colon \mathbb{T} &\to \mathbb{T}\times\mathbb{T}\\ t &\mapsto (t,t).\end{aligned}$$

Clearly $\boldsymbol{\delta}$ is differentiable and

$$\boldsymbol{G}(t) = \tilde{\boldsymbol{G}}\circ\boldsymbol{\delta}(t).$$

Using the Chain Rule,

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{G}}{\mathrm{d}t}(t) &= \frac{\partial\tilde{\boldsymbol{G}}}{\partial t_1}(\boldsymbol{\delta}(t))\circ\frac{\mathrm{d}\delta_1}{\mathrm{d}t}(t) + \frac{\partial\tilde{\boldsymbol{G}}}{\partial t_t}(\boldsymbol{\delta}(t))\circ\frac{\mathrm{d}\delta_2}{\mathrm{d}t}(t)\\ &= \boldsymbol{g}(t,t) + \int_{t_0}^{t}\frac{\partial\boldsymbol{g}}{\partial t}(t,\tau)\,\mathrm{d}\tau,\end{aligned}$$

as claimed. ▼

Let us verify that the hypotheses of the lemma hold for $(t,\tau)\mapsto\Phi_A(t,\tau)(b(\tau))$. First of all, we certainly have the first two hypotheses of the lemma. Moreover, writing $\Phi_A(t,\tau)(b(\tau)) = \Phi_A(t,t_0)\circ\Phi_A(\tau,t_0)^{-1}(b(\tau))$ and noting that (1) $\tau\mapsto b(\tau)$ is continuous (and so bounded on the compact interval $\mathbb{T}'$), (2) $t\mapsto\Phi_A(t,t_0)$ is continuous (and so also bounded on the compact interval $\mathbb{T}'$), and (3) $\tau\mapsto\Phi_A(\tau,t_0)^{-1}$ is also continuous (and so also bounded), we conclude that the third hypothesis of the lemma holds. Finally, using Theorem 3.2.9(i), $\frac{\partial\Phi_A}{\partial t}(t,\tau) = A(t)\circ\Phi_A(t,\tau)$, and this is bounded by continuity of A and our observation about that $\Phi_A(t,\tau)$ is bounded. Thus we can use the lemma to calculate

$$\begin{aligned}\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) &= A(t)\circ\Phi_A(t,t_0)(x_0) + A(t)\circ\int_{t_0}^{t}\Phi_A(t,\tau)(b(\tau))\,\mathrm{d}\tau + b(t)\\ &= A(t)(\xi(t)) + b(t),\end{aligned}$$

i.e., ξ is a solution of F. Moreover, we also clearly have $\xi(t_0) = x_0$.

To conclude uniqueness, suppose that we have two solutions ξ_1 and ξ_2 defined on the same interval $\mathbb{T}'$. Then

$$\frac{\mathrm{d}\xi_1}{\mathrm{d}t}(t) = A(t)(\xi_1(t)) + b(t),\qquad \frac{\mathrm{d}\xi_s}{\mathrm{d}t}(t) = A(t)(\xi_s(t)) + b(t),$$

and $\xi_1(t_0) = \xi_2(t_0) = x_0$. Therefore,

$$\frac{\mathrm{d}(\xi_1-\xi_2)}{\mathrm{d}t}(t) = A(t)(\xi_1(t)-\xi_2(t)),\quad (\xi_1-\xi_2)(t_0) = 0.$$

By the uniqueness assertion of Proposition 3.2.4, we conclude that $\xi_1-\xi_2 = 0$, i.e., $\xi_1 = \xi_2$. ■

We also have a global existence result in this case, just as for homogeneous systems.

3.3.2 Proposition (Global existence of solutions for systems of linear inhomogeneous ordinary differential equations) *Consider the system of linear inhomogeneous ordinary differential equations* F *with right-hand side (3.17) and suppose that* $\mathrm{b}\colon \mathbb{T} \to \mathsf{V}$ *and* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ *are continuous. If* $\xi\colon \mathbb{T}' \to \mathsf{V}$ *is a solution for* F*, then there exists a solution* $\overline{\xi}\colon \mathbb{T} \to \mathsf{V}$ *for which* $\overline{\xi}|\mathbb{T}' = \xi$.

Proof In the proof of Proposition 3.3.1, we showed that a unique solution exists on any compact interval containing t_0. Just as in the proof of Proposition 3.2.5, this implies that a solution exists at any $t \in \mathbb{T}$. ■

Since, in the proof of Proposition 3.3.1, we gave an explicit formula for solutions to initial value problems, it is worth extracting this explicit formula.

3.3.3 Corollary (An explicit solution for systems of linear inhomogeneous ordinary differential equations) *Consider the system of linear inhomogeneous ordinary differential equations* F *with right-hand side (3.17) and suppose that* $\mathrm{b}\colon \mathbb{T} \to \mathsf{V}$ *and* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ *are continuous. Given* $\mathrm{t}_0 \in \mathbb{T}$ *and* $\mathrm{x}_0 \in \mathsf{V}$*, the unique solution* $\xi\colon \mathbb{T} \to \mathsf{V}$ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \mathrm{A}(\mathrm{t})(\xi(\mathrm{t})) + \mathrm{b}(\mathrm{t})$$

is

$$\xi(\mathrm{t}) = \Phi_{\mathrm{A}}(\mathrm{t},\mathrm{t}_0)(\mathrm{x}_0) + \int_{\mathrm{t}_0}^{\mathrm{t}} \Phi_{\mathrm{A}}(\mathrm{t},\tau)(\mathrm{b}(\tau))\,\mathrm{d}\tau, \qquad \mathrm{t} \in \mathbb{T}. \tag{3.18}$$

The formula (3.18) for solutions to systems of linear inhomogeneous ordinary differential equations is often called the ***variation of constants formula***.

We note that this solution bears a strong resemblance in form to the Green's function solution for scalar systems given in Theorem 2.3.7; indeed, one can think of the state transition map as playing the rôle of a Green's function in this case. In particular, given $b \in \mathsf{V}$ (a constant vector, note) the physical interpretation of Remark 2.3.9–2 applies to the map $t \mapsto \Phi_A(t,\tau)(b)$, and leads us to think of this as being the result of applying an impulse at time τ with (vector) magnitude b. This leads to the important notion in system theory of the impulse response.

Now we can discuss the set of all solutions of a system of linear inhomogeneous ordinary differential equation F with right-hand side

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V} \\ (t,x) &\mapsto A(t)(x). \end{aligned}$$

To this end, we denote by

$$\mathrm{Sol}(F) = \left\{ \xi \in \mathsf{C}^1(\mathbb{T};\mathsf{V}) \;\middle|\; \dot{\xi}(t) = A(t)(\xi(t)) \right\}$$

the set of solutions for F. While $\mathrm{Sol}(F)$ was a vector space in the homogeneous case, in the inhomogeneous case this is no longer the case. However, the set of all solutions for the homogeneous case plays an important rôle, even in the

homogeneous case. To organise this discussion, we let F_{h} be the "homogeneous part" of F. Thus the right-hand side of F_{h} is

$$\widehat{F_{\mathrm{h}}}(t, x) = A(t)(x).$$

As in Theorem 2.3.3, $\mathrm{Sol}(F_{\mathrm{h}})$ is a $\mathbb{R}$-vector space of dimension $\dim_{\mathbb{R}}(\mathsf{V})$. The following result is then the main structural result about the set of solutions to a system of linear inhomogeneous ordinary differential equations, mirroring Theorem 2.3.3 for scalar systems.

3.3.4 Theorem (Affine space structure of sets of solutions) *Consider the system of linear inhomogeneous ordinary differential equations* F *in the* n*-dimensional* $\mathbb{R}$*-vector space* V *with right-hand side (2.11) and suppose that the maps* $\mathrm{b}\colon \mathbb{T} \to \mathsf{V}$ *and* $\mathrm{A}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *are continuous. Let* $\xi_{\mathrm{p}} \in \mathrm{Sol}(\mathrm{F})$. *Then*

$$\mathrm{Sol}(\mathrm{F}) = \{\xi + \xi_{\mathrm{p}} \mid \xi \in \mathrm{Sol}(\mathrm{F}_{\mathrm{h}})\}.$$

Proof First note that, by Theorem 3.2.6, $\mathrm{Sol}(F) \neq \emptyset$ and so there exists some $\xi_{\mathrm{p}} \in \mathrm{Sol}(F)$. We have, of course,

$$\frac{\mathrm{d}\xi_{\mathrm{p}}}{\mathrm{d}t}(t) = A(t)(\xi_{\mathrm{p}}(t)) + b(t). \tag{3.19}$$

Next let $\xi \in \mathrm{Sol}(F)$ so that

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) = A(t)(\xi(t)) + b(t). \tag{3.20}$$

Subtracting (3.19) from (3.20) we get

$$\frac{\mathrm{d}(\xi - \xi_{\mathrm{p}})}{\mathrm{d}t}(t) = A(t)(\xi(t) - \xi_{\mathrm{p}}(t)),$$

i.e., $\xi - \xi_{\mathrm{p}} \in \mathrm{Sol}(F_{\mathrm{h}})$. In other words, $\xi = \tilde{\xi} + \xi_{\mathrm{p}}$ for $\tilde{\xi} \in \mathrm{Sol}(F_{\mathrm{h}})$.

Conversely, suppose that $\xi = \tilde{\xi} + \xi_{\mathrm{p}}$ for $\tilde{\xi} \in \mathrm{Sol}(F_{\mathrm{h}})$. Then

$$\frac{\mathrm{d}\tilde{\xi}}{\mathrm{d}t}(t) = A(t)(\tilde{\xi}(t)). \tag{3.21}$$

Adding (3.19) and (3.21) we get

$$\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) = A(t)(\xi(t)) + b(t),$$

i.e., $\xi \in \mathrm{Sol}(F)$. ■

As with scalar linear inhomogeneous ordinary differential equations, there is an insightful correspondence to be made between the situation described in Theorem 3.3.4 and that of systems of linear algebraic equations described in Proposition 1.2.4.

3.3.5 Remark (Comparison of Theorem 3.3.4 with systems of linear algebraic equations) Let us compare here the result of Theorem 3.3.4 with the situation in Proposition 1.2.4 concerning linear algebraic equations of the form $L(u) = v_0$, for vector spaces U and W, a linear map $L \in \mathrm{L}(\mathsf{U}; \mathsf{W})$, and a fixed $w_0 \in \mathsf{W}$. In the setting of systems of linear inhomogeneous ordinary differential equations in a $\mathbb{R}$-vector space V, we have

$$\begin{aligned} \mathsf{U} &= \mathrm{C}^1(\mathbb{T}; \mathsf{V}), \\ \mathsf{W} &= \mathrm{C}^0(\mathbb{T}; \mathsf{V}), \\ L(f)(t) &= \dot{f}(t) - A(t)(f(t)), \\ w_0 &= b. \end{aligned}$$

Then Propositions 3.3.1 and 3.3.2 tell us that L is surjective, and so $w_0 \in \mathrm{image}(L)$. Thus we are in case (ii) of Proposition 1.2.4, which exactly the statement of Theorem 3.3.4. Note that L is not injective, since Theorem 3.2.6 tells us that $\dim_{\mathbb{R}}(\ker(L)) = \dim_{\mathbb{R}}(\mathsf{V})$. •

3.3.6 Remark (What happened to the Wronskian?) In Section 2.3.1.2 we described how the Wronskian can be used for scalar linear inhomogeneous ordinary differential equations to generate a particular solution. A similar development is possible for systems of equations, but we shall not pursue it here. It is worth recording the reasons for not doing so.

1. In Corollary 3.3.3 we produce a specific and natural "particular solution" for a system of linear inhomogeneous ordinary differential equations, namely the function that assigns to the inhomogeneous term "b," the solution

$$\xi_{\mathrm{p}}(t) = \int_{t_0}^{t} \Phi_A(t, \tau)(b(\tau)) \, \mathrm{d}\tau.$$

   Then the form of the solution of Corollary 3.3.3 is $\xi = \xi_{\mathrm{h}} + \xi_{\mathrm{p}}$, where $\xi_{\mathrm{h}} \in \mathrm{Sol}(F_{\mathrm{h}})$ satisfies the initial conditions. This is just so cool. . . why would you want to do more?

2. In Section 2.2.1 we discussed the notion of a fundamental set of solutions for scalar linear homogeneous ordinary differential equations. There is no really distinguished fundamental set of solutions, and the Wronskian-related constructions were developed for an *arbitrary* fundamental set of solutions. This has its benefits in this setting, as the results are general in this respect.

   However, in Section 3.2.2.2 we saw that there was *one* object that naturally describes the solutions for a system of linear homogeneous ordinary differential equations, the state transition map. Note that in Procedure 3.2.11 we indicate how to build the state transition map from a fundamental set of solutions for a system of equations, through the fundamental matrix-function Ξ that we build

after choosing a basis. It is the fundamental matrix, and its determinant, that would be involved in Wronskian-type constructions for systems of equations. However, these are only arrived at after choosing a basis, and so seem quite unnatural in our setting of general vector spaces. •

Given that we will not be pursuing any Wronskian-type constructions, it only remains to illustrate how one might use the about constructions in practice.

3.3.7 Example (System of linear inhomogeneous ordinary differential equations) We take $\mathsf{V} = \mathbb{R}^2$ and the linear inhomogeneous ordinary differential equation $\boldsymbol{F}$ with right-hand side

$$\begin{aligned}\widehat{\boldsymbol{F}}\colon (0,\infty)\times\mathbb{R}^2 &\to \mathbb{R}^2\\ (t,(x_1,x_2)) &\mapsto \left(\frac{1}{t}x_1 - x_2 + t, \frac{1}{t^2}x_1 + \frac{2}{t}x_2 - t^2\right).\end{aligned}$$

A solution $t \mapsto (\xi_1(t), \xi_2(t))$ satisfies

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = \underbrace{\begin{bmatrix}\frac{1}{t} & -1\\ \frac{1}{t^2} & \frac{2}{t}\end{bmatrix}}_{\boldsymbol{A}(t)}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \underbrace{\begin{bmatrix}t\\ -t^2\end{bmatrix}}_{\boldsymbol{b}(t)}.$$

Note that the homogeneous system $\boldsymbol{F}_{\mathrm{h}}$ was examined in Example 3.2.13, where we computed the state transition matrix to be

$$\boldsymbol{\Phi}_{\boldsymbol{A}}(t,t_0) = \begin{bmatrix}-\frac{t^2(\ln(t/t_0)-1)}{t_0^2} & -\frac{t^2\ln(t/t_0)}{t_0}\\ \frac{t\ln(t/t_0)}{t_0^2} & \frac{t(\ln(t/t_0)+1)}{t_0}\end{bmatrix}.$$

We then compute[7]

$$\begin{aligned}\int_{t_0}^{t}\boldsymbol{\Phi}_{\boldsymbol{A}}(t,\tau)\boldsymbol{b}(\tau)\,\mathrm{d}\tau &= \int_{t_0}^{t}\begin{bmatrix}-\frac{t^2(\ln(t/\tau)-1)}{\tau^2} & -\frac{t^2\ln(t/\tau)}{\tau}\\ \frac{t\ln(t/\tau)}{\tau^2} & \frac{t(\ln(t/\tau)+1)}{\tau}\end{bmatrix}\begin{bmatrix}\tau\\ -\tau^2\end{bmatrix}\mathrm{d}\tau\\ &= \begin{bmatrix}\frac{1}{4}t^2(t^2 - 2t_0^2\ln(t/t_0) - 2\ln(t/t_0)^2 + 4\ln(t/t_0) - t_0^2)\\ \frac{1}{4}t(2\ln(t/t_0)(\ln(t/t_0) + t_0^2) - 3(t-t_0)(t+t_0))\end{bmatrix}.\end{aligned}$$

If we now wish to find the solution for $\boldsymbol{F}$ with initial condition $\boldsymbol{x}_0 = (x_{10}, x_{20})$ at time t_0, we use the explicit form of Corollary 3.3.3:

$$\begin{aligned}\xi(t) &= \boldsymbol{\Phi}_A(t,t_0)\boldsymbol{x}_0 + \int_{t_0}^{t}\boldsymbol{\Phi}_A(t,\tau)\boldsymbol{b}(\tau)\,\mathrm{d}\tau\\ &= \begin{bmatrix}-\frac{t^2(\ln(t/t_0)-1)}{t_0^2}x_{10} - \frac{t^2\ln(t/t_0)}{t_0}x_{20} + \frac{1}{4}t^2(t^2 - 2t_0^2\ln(t/t_0) - 2\ln(t/t_0)^2 + 4\ln(t/t_0) - t_0^2)\\ \frac{t\ln(t/t_0)}{t_0^2}x_{10} + \frac{t(\ln(t/t_0)+1)}{t_0}x_{20} + \frac{1}{4}t(2\ln(t/t_0)(\ln(t/t_0) + t_0^2) - 3(t-t_0)(t+t_0))\end{bmatrix}.\end{aligned}$$

[7] Integration courtesy of MATHEMATICA®.

As with pretty much any method for solving systems of linear inhomogeneous (or, indeed, homogeneous) ordinary differential equations, tedious computations and generally impossible integrals render the explicit formula of Corollary 3.3.3 of questionable value as a computational tool. •

3.3.2 Equations with constant coefficients

We now specialise the discussion in the preceding section to systems of linear inhomogeneous ordinary differential equations with constant coefficients. Thus we are looking at a system of linear inhomogeneous ordinary differential equations F in a finite-dimensional $\mathbb{R}$-vector space V and with right-hand side given by

$$\widehat{F}(t, x) = A(x) + b(t) \tag{3.22}$$

for $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ and $b \colon \mathbb{T} \to \mathsf{V}$. Of course, all general results concerning the existence and uniqueness of solutions (i.e., Propositions 3.3.1 and 3.3.2), and of the structure of the set of solutions (i.e., Theorem 3.3.4) apply in the constant coefficient case. Here, however, we can refine a little the explicit solution of Corollary 3.3.3 because, as per Theorem 3.2.42(ix), $\Phi_A(t, t_0) = \mathrm{e}^{A(t-t_0)}$ in this case. We can thus summarise the situation in the following theorem.

3.3.8 Theorem (An explicit solution for systems of linear inhomogeneous ordinary differential equations with constant coefficients) *Consider the system of linear inhomogeneous ordinary differential equations* F *with constant coefficients and right-hand side (3.22), and suppose that* $\mathsf{b} \colon \mathbb{T} \to \mathsf{V}$ *is continuous. Given* $\mathsf{t}_0 \in \mathbb{T}$ *and* $\mathsf{x}_0 \in \mathsf{V}$*, the unique solution* $\xi \colon \mathbb{T} \to \mathsf{V}$ *to the initial value problem*

$$\dot{\xi}(\mathsf{t}) = \mathrm{A}(\xi(\mathsf{t})) + \mathrm{b}(\mathsf{t})$$

is

$$\xi(\mathsf{t}) = \mathrm{e}^{\mathrm{A}(\mathsf{t}-\mathsf{t}_0)}(\mathsf{x}_0) + \int_{\mathsf{t}_0}^{\mathsf{t}} \mathrm{e}^{\mathrm{A}(\mathsf{t}-\tau)}(\mathrm{b}(\tau))\,\mathrm{d}\tau, \qquad \mathsf{t} \in \mathbb{T}.$$

We comment that our observations Remark 2.3.11 about the particular solution

$$\xi_{\mathrm{p},b} = \int_{t_0}^{t} \mathrm{e}^{A(t-\tau)}(b(\tau))\,\mathrm{d}\tau$$

for constant coefficient systems and its relation to convolution integrals is also valid here.

3.3.9 Remark (What happened to the "method of undetermined coefficients"?) In Section 2.3.2.1 we spent some time describing a rather *ad hoc* method, the "method of undetermined coefficients," for finding particular solutions for scalar linear inhomogeneous ordinary differential equations with constant coefficients. A similar strategy is possible for systems of linear inhomogeneous ordinary differential equations with constant coefficients, but we shall not pursue it here. Here is why.

1. The rationale of Remark 3.3.6–1 is equally valid here: we have such a nice characterisation in Corollary 3.3.3 of a particular solution that to mess this up with an *ad hoc* procedure that only works for pretty uninteresting functions is simply not a worthwhile undertaking.
2. While for scalar equations it might be argued that there is some reason for being able to quickly bang out particular solutions for specific pretty uninteresting functions—see, particular, the notion of "step response" in Example 2.3.19 and the notion of "frequency response" in Example 2.3.20—for systems of equations the benefit of this is not so clear, given the complexity of doing computation in any example. •

All that remains, since we have discharged ourselves of the responsibility of providing any analogies to the various methods we used for scalar equations in Section 2.2, is to give an example of how to apply the explicit formula of Theorem 3.3.8.

3.3.10 Example (A second-order scalar equation as a system of equations) We consider here the second-order scalar linear inhomogeneous ordinary differential equation F with right-hand side

$$\widehat{F}(t, x, x^{(1)}) = -\omega_0^2 x - 2\zeta\omega_0 x^{(1)} + A\sin(\omega t)$$

that was considered in detail in Example 2.3.20. First we convert this into a system of linear inhomogeneous ordinary differential equations, following Exercise 1.3.23. Thus we introduce the variables $x_1 = x$ and $x_2 = x^{(1)}$ so that

$$\begin{aligned} x_1^{(1)} &= x^{(1)} = x_2, \\ x_2^{(1)} &= x^{(2)} = -\omega_0^2 x - 2\zeta\omega_0 x^{(1)} + A\sin(\omega t) = -\omega_0^2 x_1 - 2\zeta\omega_0 x_2 + A\sin(\omega t). \end{aligned}$$

That is to say

$$\widehat{\boldsymbol{F}}_1(t, (x_1, x_2)) = (x_2, -\omega_0^2 x_1 - 2\zeta\omega_0 x_2 + A\sin(\omega t)).$$

Solutions $t \mapsto (\xi_1(t), \xi_2(t))$ then satisfy

$$\begin{bmatrix} \dot{\xi}_1(t) \\ \dot{\xi}_2(t) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ A\sin(\omega t) \end{bmatrix}}_{\boldsymbol{b}(t)}.$$

To illustrate, we suppose that $\zeta^2 \leq 1$ and $\omega_0 > 0$.

We will first compute $\mathrm{e}^{\boldsymbol{A}t}$ in this case, following Procedure 3.2.48, making use of the notation in Procedure 3.2.45. The characteristic polynomial of $\boldsymbol{A}$ is

$$P_{\boldsymbol{A}} = X^2 + 2\zeta\omega_0 X + \omega_0^2,$$

and so the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = \omega_0(-\zeta + \mathrm{i}\sqrt{1-\zeta^2})$, along with its complex conjugate $\bar{\lambda}_1$. This eigenvalue necessarily has algebraic and geometric multiplicity 1. We compute that

$$\ker(\boldsymbol{A}^{\mathbb{C}} - \lambda_1 \boldsymbol{I}_2) = \operatorname{span}_{\mathbb{R}}((-\zeta, \omega_0) + \mathrm{i}(\sqrt{1-\zeta^2}, \omega_0)).$$

Thus we take

$$\boldsymbol{\zeta}_{1,1} = (-\zeta, \omega_0) + \mathrm{i}(\sqrt{1-\zeta^2}, \omega_0)$$

and, therefore,

$$\boldsymbol{a}_{1,1} = (-\zeta, \omega_0), \quad \boldsymbol{b}_{1,1} = (-\sqrt{1-\zeta^2}, 0).$$

Thus

$$\boldsymbol{\alpha}_{1,1}(t) = \mathrm{e}^{-\omega_0\zeta t}\cos(\omega_0\sqrt{1-\zeta^2}t)\boldsymbol{a}_{1,1} - \mathrm{e}^{-\omega_0\zeta t}\sin(\omega_0\sqrt{1-\zeta^2}t)\boldsymbol{b}_{1,1}$$

and

$$\boldsymbol{\beta}_{1,1}(t) = \mathrm{e}^{-\omega_0\zeta t}\cos(\omega_0\sqrt{1-\zeta^2}t)\boldsymbol{b}_{1,1} + \mathrm{e}^{-\omega_0\zeta t}\sin(\omega_0\sqrt{1-\zeta^2}t)\boldsymbol{a}_{1,1}.$$

Thus a fundamental matrix is then determined to be

$$\boldsymbol{\Xi}(t) = \mathrm{e}^{-\omega_0\zeta t}\begin{bmatrix} -\zeta\cos(\omega_0\sqrt{1-\zeta^2}t) + \sqrt{1-\zeta^2}\sin(\omega_0\sqrt{1-\zeta^2}t) & -\sqrt{1-\zeta^2}\cos(\omega_0\sqrt{1-\zeta^2}t) - \zeta\sin(\omega_0\sqrt{1-\zeta^2}t) \\ \omega_0\cos(\omega_0\sqrt{1-\zeta^2}t) & \omega_0\sin(\omega_0\sqrt{1-\zeta^2}t) \end{bmatrix}.$$

Then we calculate

$$\begin{aligned} \mathrm{e}^{At} &= \boldsymbol{\Xi}(t)\boldsymbol{\Xi}(0)^{-1} \\ &= \mathrm{e}^{-\omega_0\zeta t}\begin{bmatrix} \cos(\omega_0\sqrt{1-\zeta^2}t) + \frac{\zeta\sin(\omega_0\sqrt{1-\zeta^2}t)}{\sqrt{1-\zeta^2}} & \frac{\sin(\omega_0\sqrt{1-\zeta^2}t)}{\omega_0\sqrt{1-\zeta^2}} \\ -\frac{\omega_0\sin(\omega_0\sqrt{1-\zeta^2}t)}{\sqrt{1-\zeta^2}} & \cos(\omega_0\sqrt{1-\zeta^2}t) - \frac{\zeta\sin(\omega_0\sqrt{1-\zeta^2}t)}{\sqrt{1-\zeta^2}} \end{bmatrix}. \end{aligned}$$

Now we can calculate[8]

$$\begin{aligned}\int_0^t \mathrm{e}^{A(t-\tau)}\boldsymbol{b}(\tau)\,\mathrm{d}\tau = \Bigg(&\mathrm{e}^{-\omega_0\zeta t}\frac{2A\omega\omega_0\zeta}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\cos(\omega_0\sqrt{1-\zeta^2}t)\\ &+\mathrm{e}^{-\omega_0\zeta t}\frac{A\omega(\omega^2+\omega_0^2(2\zeta^2-1))}{\sqrt{1-\zeta^2}\omega_0(\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4)}\sin(\omega_0\sqrt{1-\zeta^2}t)\\ &-\frac{2A\omega\omega_0\zeta}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\cos(\omega t)\\ &+\frac{A(\omega_0^2-\omega^2)}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\sin(\omega t),\\ &\mathrm{e}^{-\omega_0\zeta t}\frac{A\omega(\omega^2-\omega_0^2)}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\cos(\omega_0\sqrt{1-\zeta^2}t)\\ &-\mathrm{e}^{-\omega\zeta t}\frac{A\omega\zeta(\omega^2+\omega_0^2)}{\sqrt{1-\zeta^2}(\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4)}\sin(\omega_0\sqrt{1-\zeta^2}t)\\ &+\frac{A\omega(\omega_0^2-\omega^2)}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\cos(\omega t)\\ &+\frac{2A\zeta\omega^2\omega_0}{\omega^4+2\omega^2\omega_0^2(2\zeta^2-1)+\omega_0^4}\sin(\omega t)\Bigg),\end{aligned}\tag{3.23}$$

assuming that $\zeta \neq 0$. If $\zeta = 0$ and $\omega \neq \omega_0$, the preceding expression is still valid. When $\zeta = 0$ and $\omega = \omega_0$, a different computation must be done, and in this case we compute

$$\int_0^t \mathrm{e}^{A(t-\tau)}\boldsymbol{b}(\tau)\,\mathrm{d}\tau = \left(\tfrac{A}{2\omega_0^2}(\sin(\omega_0 t)-\omega_0 t\cos(\omega_0 t)),\tfrac{A}{2}t\sin(\omega_0 t)\right).\tag{3.24}$$

Note that, in all cases, the preceding expressions give the solution to the ordinary differential equation when the initial conditions are $(0,0)$. Let us make some comments on this solution.

1. $\zeta \neq 0$: Note that (3.23) is *not* the steady-state response of the system, as was the particular solution obtained for this problem in Example 2.3.20 using the method of undetermined coefficients. The reason for the disparity is that the expression above has the property that its initial conditions at $t = 0$ are $(0,0)$.

[8] Integration courtesy of MATHEMATICA®.

Note that, as $t \to \infty$, we have

$$\int_0^t \mathrm{e}^{A(t-\tau)}\boldsymbol{b}(\tau)\,\mathrm{d}\tau \approx \begin{pmatrix} -\dfrac{2A\omega\omega_0\zeta}{\omega^4 + 2\omega^2\omega_0^2(2\zeta^2-1) + \omega_0^4}\cos(\omega t) \\ \quad + \dfrac{A(\omega_0^2 - \omega^2)}{\omega^4 + 2\omega^2\omega_0^2(2\zeta^2-1) + \omega_0^4}\sin(\omega t), \\ \dfrac{A\omega(\omega_0^2 - \omega^2)}{\omega^4 + 2\omega^2\omega_0^2(2\zeta^2-1) + \omega_0^4}\cos(\omega t) \\ \quad + \dfrac{2A\zeta\omega^2\omega_0}{\omega^4 + 2\omega^2\omega_0^2(2\zeta^2-1) + \omega_0^4}\sin(\omega t) \end{pmatrix}.$$

Notice that the first component of this is exactly the particular solution of Example 2.3.20, while the second component is its time-derivative. This is as it should be, given our conversion of the scalar second-order equation into a vector first-order equation.

2. $\zeta = 0$ and $\omega \neq \omega_0$: In this case, there is no steady-state solution since the homogeneous solution does not decay to zero as $t \to \infty$, and is indeed periodic itself. Nonetheless, the solution (3.23) does have two components, one with frequency ω and one with frequency ω_0. While this does not quite disambiguate the particular from the homogeneous solution[9], we can nonetheless see from the expression (3.23) that the particular solution of Example 2.3.20 is comprised on the last two terms in the first component.
3. $\zeta = 0$ and $\omega = \omega_0$: In this case, there is again no steady-state solution; indeed the solution "blows up" as $t \to \infty$. This is as we saw in Example 2.3.20, and is due to the physical phenomenon of "resonance." Moreover, the first component of (3.24) is *not* the particular solution from Example 2.3.20; the particular particular solution (3.24) is prescribed to have initial condition $(0, 0)$, whereas, in the method of undetermined coefficients, it is the *form* of the solution that is determined. •

Exercises

3.3.1 Consider the first-order scalar linear homogeneous ordinary differential equation with right-hand side $\widehat{F}(t, x) = a(t)x + b(t)$ for continuous functions $a, b\colon \mathbb{T} \to \mathbb{R}$. Using your result from Exercise 3.2.9, use Corollary 3.3.3 to determine the solution to the initial value problem

$$\dot{\xi}(t) = a(t)\xi(t) + b(t), \quad \xi(t_0) = x_0,$$

thinking of this as a system of linear inhomogeneous ordinary differential equations in the one-dimensional vector space $\mathbb{R}$.

[9] A periodic function can have more than one frequency.

3.3.2 Consider the scalar linear inhomogeneous ordinary differential equation F given by

$$F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + \omega^2 x - \sin(\omega t)$$

for $\omega \in \mathbb{R}_{>0}$. Answer the following questions.

(a) Use the method of undetermined coefficients to obtain a particular solution for $\boldsymbol{F}$.

(b) Convert $\boldsymbol{F}$ into a system of linear inhomogeneous ordinary differential equations $\boldsymbol{F}_1$ in $\mathbb{R}^2$ with right-hand side

$$\begin{aligned} \widehat{\boldsymbol{F}} \colon \mathbb{T} \times \mathbb{R}^2 &\to \mathbb{R}^2 \\ (t, (x_1, x_2)) &\mapsto \boldsymbol{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \boldsymbol{b}(t), \end{aligned}$$

giving explicit formulae for $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^2; \mathbb{R}^2)$ and $\boldsymbol{b} \colon \mathbb{T} \to \mathbb{R}^2$.

(c) Show that

$$\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} \cos(\omega t) & \frac{1}{\omega} \sin(\omega t) \\ -\omega \sin(\omega t) & \cos(\omega t) \end{bmatrix}.$$

(d) Compute

$$\xi_{\mathrm{p},b}(t) = \int_0^t \mathrm{e}^{\boldsymbol{A}(t-\tau)} \boldsymbol{b}(\tau) \, \mathrm{d}\tau.$$

Use your answer to give a particular solution for the scalar equation $\boldsymbol{F}$.

(e) Explain how the particular solutions from parts (a) and (d) are the same, and explain how to describe the difference between them.

3.3.3 For the linear transformations $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ of Exercise 3.2.14, use Theorem 3.3.8 to determine the solution to the initial value problem

$$\dot{\xi}(t) = \boldsymbol{A}\xi(t) + \boldsymbol{b}(t), \quad \xi(0) = \boldsymbol{0},$$

with $\boldsymbol{b}$ as follows:

(a) $\boldsymbol{b}(t) = (0, 1)$;

(b) $\boldsymbol{b}(t) = (\cos(t), 0)$;

(c) $\boldsymbol{b}(t) = (\mathrm{e}^{2t}, 0)$;

(d) $\boldsymbol{b}(t) = (\sin(t), 0, 1)$;

(e) $\boldsymbol{b}(t) = (0, \mathrm{e}^{-t}, 0)$;

(f) $\boldsymbol{b}(t) = (\sin(2t), 0, 1)$;

(g) $\boldsymbol{b}(t) = (1, 0, 0, 1)$;

(h) $\boldsymbol{b}(t) = (\sin(t), 0, 0, \cos(t))$;

(i) $\boldsymbol{b}(t) = (0, 0, 0, 0, 0)$.

Section 3.4

Phase-plane analysis

In this section we consider a way of representing the behaviour of ordinary differential equations whose state space is a subset of $\mathbb{R}^2$ via their "phase portraits." We have already used this method informally on a number of occasions, and in this section we shall be a little more systematic. We begin in Section 3.4.1 by exhaustively examining phase portraits for linear systems in two variables. In Section 3.4.2 we consider phenomenon that can happen for nonlinear systems. In this case, the presentation is essentially example driven, and we give little by way of rigorous methodology. This analysis appears a little *ad hoc*, however, the methods can give more insight into what is "really happening" with a differential equation. Also, the ideas that we encounter in the simple two-dimensional setting suggest techniques that may be profitably applied in higher-dimensions. These ideas are discussed in Section 3.4.3.

3.4.1 Phase portraits for linear systems

We begin our discussion with a consideration of phase portraits for systems of linear ordinary differential equations in $\mathbb{R}^2$ with constant coefficients. Thus we are considering differential equations $\boldsymbol{F}$ with

$$\begin{aligned}\widehat{\boldsymbol{F}}\colon &\mathbb{T}\times\mathbb{R}^2\to\mathbb{R}^2\\ &(t,(x_1,x_2))\mapsto \underbrace{\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix}}_{A}\begin{bmatrix}x_1\\x_2\end{bmatrix}.\end{aligned}$$

In Section 3.2.3 we learned that the solution to the initial value problem

$$\begin{bmatrix}\dot\xi_1(t)\\\dot\xi_2(t)\end{bmatrix}=\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix}\begin{bmatrix}\xi_1(t)\\\xi_2(t)\end{bmatrix},\quad\begin{bmatrix}\xi_1(0)\\\xi_2(0)\end{bmatrix}=\boldsymbol{x}_0=\begin{bmatrix}x_{0,1}\\x_{2,0}\end{bmatrix},$$

is $\xi(t)=\mathrm{e}^{At}\boldsymbol{x}_0$. What we shall do in this section is represent these solutions in a particular way, such as we initially discussed in Example 1.3.23. To be specific, we shall plot the solutions as parameterised curves in the (x_1,x_2)-plane. In doing this, we shall represent, not just one solution, but the entirety of solutions with various initial conditions. By doing this, one gets a qualitative understanding of the behaviour of solutions that is simply not achievable by looking at a closed-form solution or by looking at plots of $t\mapsto\xi_1(t)$ and $t\mapsto\xi_2(t)$ of *fixed* solutions with a single initial condition. The resulting collection of solutions, represented as parameterised curves, is called the ***phase portrait***.

We shall break down the analysis into various cases, based on the character of eigenvalues and eigenvectors.

3.4.1.1 Stable nodes We first consider the case where there are two negative real eigenvalues. In this case, there are a few cases to consider, but all fall into the general category of what we call a ***stable node***, since, as we shall see, all solutions tend to $(0,0)$ as $t \to \infty$.

Distinct eigenvalues

Here we suppose that we have eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 < \lambda_2 < 0$. The behaviour in the case is then determined by the eigenvectors. Let us first look at the simple case where the eigenvectors are the standard basis vectors $e_1 = (1,0)$ and $e_2 = (0,1)$. In this case, A is given by

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Then

$$\begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} = \mathrm{e}^{At} \begin{bmatrix} \xi_1(0) \\ \xi_2(t) \end{bmatrix} = \begin{bmatrix} \xi_1(0)\mathrm{e}^{\lambda_1 t} \\ \xi_2(0)\mathrm{e}^{\lambda_2 t} \end{bmatrix}.$$

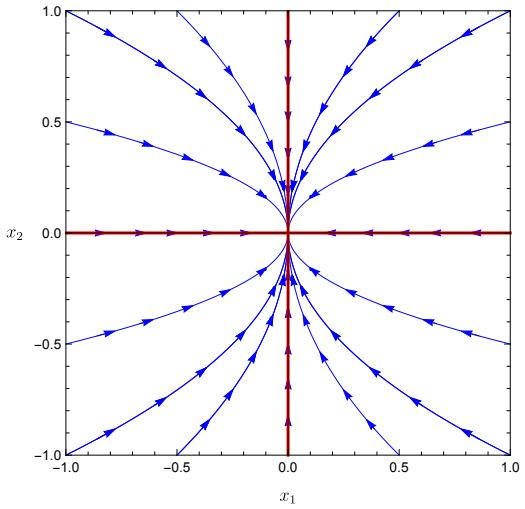
In Figure 3.1a we show the phase portrait, i.e., the family of solutions plotted as parameterised curves in the (x_1, x_2)-plane. Let us make a few comments about the nature of the phase portrait so as to explain the nature of its essential features.

1. The eigenvectors, which are e_1 and e_2 in this case, show up as lines through the origin with the property that solutions that start on these lines remain on these lines. These are, then, ***invariant subspaces*** for the dynamics. In Figure 3.1a these are indicated in red. In this case, because the eigenvalues are negative, the solutions along these lines approach $(0,0)$ as $t \to \infty$, as can be seen from the direction of the arrows.
2. Solutions corresponding to other initial conditions also approach $(0,0)$ as $t \to \infty$. From Figure 3.1a we can see that all of these other solutions approach $(0,0)$ tangent to the eigenvector e_2. The reason for this is that the eigenvalue λ_1 is the "more negative" eigenvalue, and so solutions decay to zero more quickly in the direction of the corresponding eigenvector e_1.

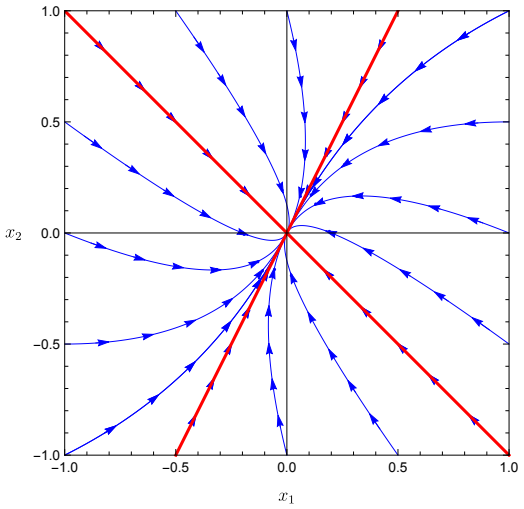
In the phase portrait of Figure 3.1a the eigenvectors are the standard basis vectors, and this was selected to make the process easier to visualise and explain. However, typically the eigenvectors are *not* the standard basis vectors, of course. However, the same ideas apply: (1) the eigenvectors represent invariant subspaces for the dynamics and (2) solutions approach $(0,0)$ more quickly in the direction of the "more negative" eigenvector. Let us illustrate this with an example, taking

$$A = \begin{bmatrix} -\frac{5}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{4}{3} \end{bmatrix}.$$

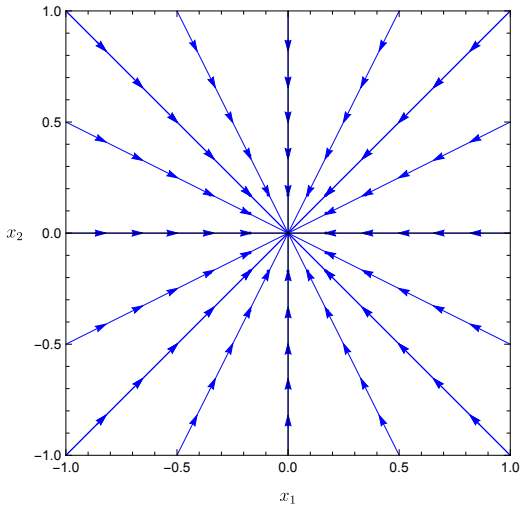
In this case we compute the eigenvalues of A to be $\lambda = -1$ and $\lambda_2 = -2$, i.e., the same eigenvalues as in the example illustrated in Figure 3.1a. Corresponding eigenvectors are $v_1 = (1,-1)$ and $v_2 = (1,2)$. In Figure 3.1b we show the phase portrait. In



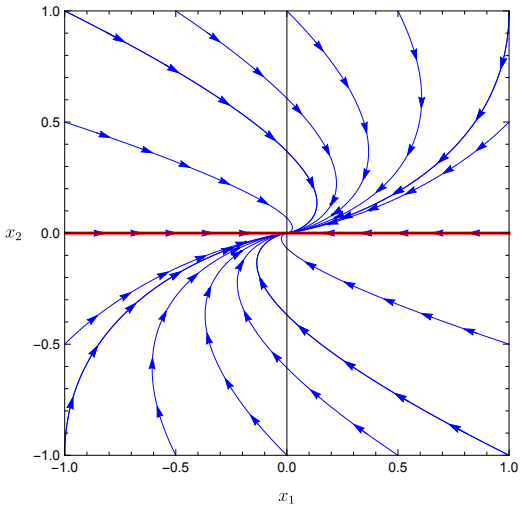
(a) Stable node with the distinct eigenvalues $\lambda_1 = -2$ and $\lambda_2 = -1$, and standard basis vectors as eigenvectors



(b) Stable node with distinct eigenvalues $\lambda_1 = -2$ and $\lambda_2 = -1$ and eigenvectors $v_1 = (1, -1)$, and $v_2 = (1, 2)$



(c) Stable node with repeated eigenvalue $\lambda = -1$ and geometric multiplicity 2



(d) Stable node with repeated eigenvalue $\lambda = -1$ and geometric multiplicity 1

Figure 3.1 Stable nodes

red we denote the invariant subspaces corresponding to the eigenvectors. Note that, essentially, once one understand the phase portrait in Figure 3.1a with the standard basis vectors as eigenvectors, it is a matter of "distortion" to produce the phase portrait of Figure 3.1b with its different eigenvectors.

Repeated eigenvalue with geometric multiplicity 2

Next we consider the case where $\boldsymbol{A}$ has a single eigenvalue $\lambda \in \mathbb{R}_{<0}$ with $m_{\mathrm{a}}(\lambda, \boldsymbol{A}) = m_{\mathrm{g}}(\lambda, \boldsymbol{A}) = 2$. In this case note that we simply have

$$\boldsymbol{A} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \implies \mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} \mathrm{e}^{\lambda t} & 0 \\ 0 & \mathrm{e}^{\lambda t} \end{bmatrix}.$$

That is to say, all vectors are eigenvectors. Thus the phase portrait of Figure 3.1c is perhaps not surprising.

Repeated eigenvalue with geometric multiplicity 1

Here we again consider the case where $\boldsymbol{A}$ has a single eigenvalue $\lambda \in \mathbb{R}_{<0}$ with $m_{\mathrm{a}}(\lambda, \boldsymbol{A}) = 2$. But in this case we assume that $m_{\mathrm{g}}(\lambda, \boldsymbol{A}) = 1$. A representative example is given by

$$\boldsymbol{A} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \implies \mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} \mathrm{e}^{\lambda t} & t \\ 0 & \mathrm{e}^{\lambda t} \end{bmatrix}.$$

The phase portrait is shown in Figure 3.1d, with the single invariant subspace indicated in red.

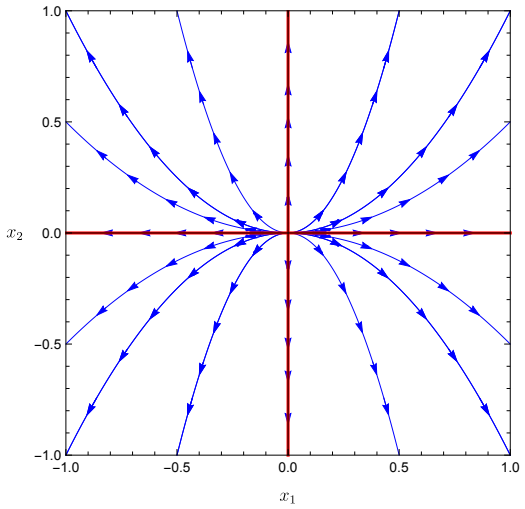
3.4.1.2 Unstable nodes The cases we consider in this section are rather like those in the previous section, except that here we will work with positive eigenvalues. In this case we have an ***unstable node*** since all solutions, except the one with initial condition $(0, 0)$, diverge to infinity as $t \to \infty$. The analysis is quite like that for stable nodes, so we will be briefer.

Distinct eigenvalues

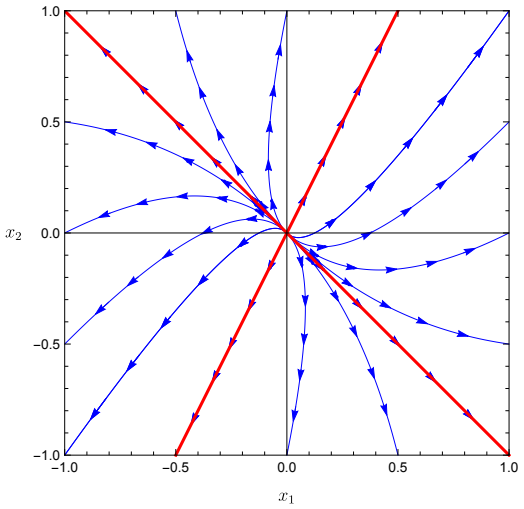
We first consider the case where $\boldsymbol{A}$ has distinct negative real eigenvalues. In this case, there will be two linearly independent eigenvectors that will each span a one-dimensional invariant subspace for the differential equation. Consider first the case where

$$\boldsymbol{A} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \implies \mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} \mathrm{e}^{\lambda_1 t} & 0 \\ 0 & \mathrm{e}^{\lambda_2 t} \end{bmatrix}$$

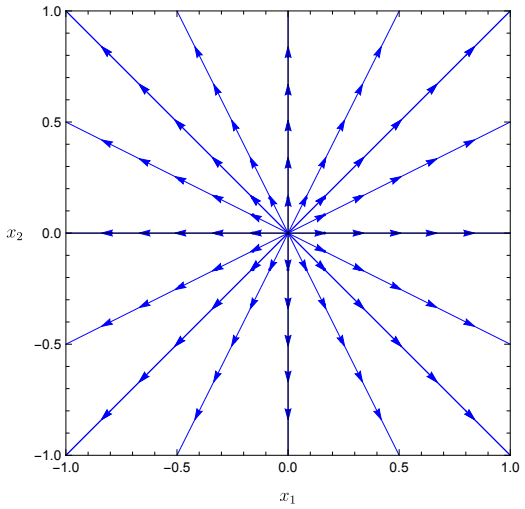
for $0 < \lambda_1 < \lambda_2$. In this case the eigenvectors are the standard basis vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. The phase portrait is shown in Figure 3.2a for this case. We see that, the phase portrait is, in some sense, the "opposite" of that in Figure 3.1a for a stable node. One still has the invariant subspaces, but now the parameterised curves for solutions are diverging from the equilibrium at $(0, 0)$. Note that, since the divergence from $(0, 0)$ is faster in the direction of $\boldsymbol{e}_2$, solution curves approach $(0, 0)$ faster going backwards in time. This is why solutions approach $(0, 0)$ tangent to the $\boldsymbol{e}_1$-direction.



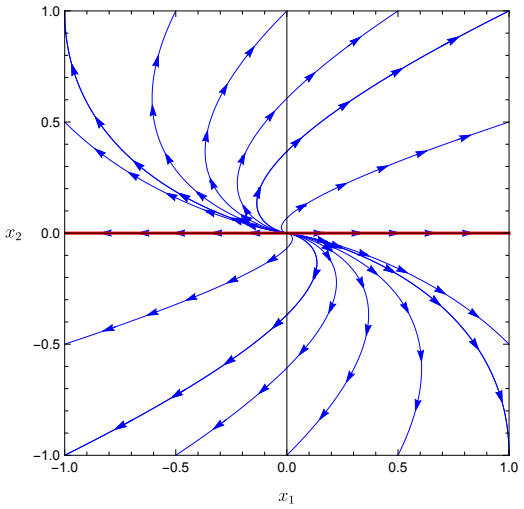
(a) Unstable node with the distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, and standard basis vectors as eigenvectors



(b) Unstable node with distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$ and eigenvectors $v_1 = (1, -1)$, and $v_2 = (1, 2)$



(c) Unstable node with repeated eigenvalue $\lambda = 1$ and geometric multiplicity 2



(d) Unstable node with repeated eigenvalue $\lambda = 1$ and geometric multiplicity 1

Figure 3.2 Unstable nodes

Let us also consider a case where the eigenvectors are not the standard basis vectors. Here we take

$$A = \begin{bmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{5}{3} \end{bmatrix},$$

which has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. Associated eigenvectors are $v_1 = (1, -1)$ and $v_2 = (2, 1)$. As we see in Figure 3.2b, the phase portrait is the expected "distortion" of the phase portrait from Figure 3.2a.

Repeated eigenvalue with geometric multiplicity 2

Next we consider the case of a positive real eigenvalue λ with $m_a(\lambda, A) = m_g(\lambda, A) = 2$. In this case, A is necessarily given by

$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \quad \Longrightarrow \quad e^{At} = \begin{bmatrix} e^{\lambda t} & 0 \\ 0 & e^{\lambda t} \end{bmatrix}.$$

In this case, every one-dimensional subspace is an invariant subspace along which solutions diverge to ∞. The phase portrait is shown in Figure 3.2c, and shows the expected features.

Repeated eigenvalue with geometric multiplicity 1

The final unstable node is associated to a positive eigenvalue λ with $m_a(\lambda, A) = 2$ and $m_g(\lambda, A) = 1$. In this case, we have only one one-dimensional invariant subspace associated to an eigenvector. In Figure 3.2d we show the phase portrait for this case associated with the typical example

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \quad \Longrightarrow \quad e^{At} = \begin{bmatrix} e^{\lambda t} & t \\ 0 & e^{\lambda t} \end{bmatrix}.$$

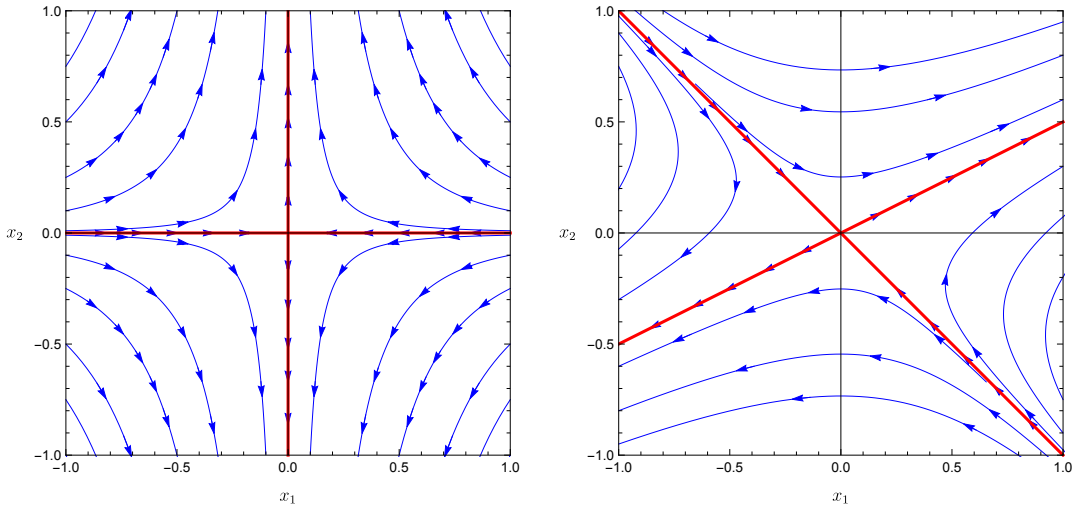
Again, we note that all solution curves, except for the one at the equilibrium $(0, 0)$, diverge to ∞ as $t \to \infty$.

3.4.1.3 Saddle points The next case we consider is where the real eigenvalues λ_1 and λ_2 satisfy $\lambda_1 < 0 < \lambda_2$. In this case we have what is called a ***saddle point***, in reference to the setting of a function of two variables at a point where the derivative of the function vanishes and its Hessian has one positive and one negative eigenvalue.

In this case, eigenvectors for the distinct eigenvalues are necessarily linearly independent, so we do not have to carefully consider cases of differing algebraic and geometric multiplicities. Let us begin with the special case

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \Longrightarrow \quad e^{At} = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix},$$

where the eigenvectors are the standard basis vectors e_1 and e_2. In Figure 3.3a we show the phase portrait in this case. Let us make a few comments on what we see.



(a) Saddle point with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$, and standard basis vectors as eigenvectors

(b) Saddle point with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$, and eigenvectors $v_1 = (1, -1)$ and $v_2 = (2, 1)$

Figure 3.3 Saddle points

1. There are two invariant subspaces corresponding to the linearly independent eigenvectors. On the invariant subspace associated with the negative eigenvalue, the solutions converge to $(0, 0)$ as $t \to \infty$. On the invariant subspace associated with the positive eigenvalue, solutions diverge to ∞ as $t \to \infty$.
2. All other solutions, except for that at the equilibrium point $(0, 0)$, diverge to ∞ as $t \to \infty$, but do so after possibly falling temporarily under the influence of the negative eigenvalue.

We can, of course, adapt this to situations where the eigenvectors are not the standard basis vectors. To illustrate, let us take

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}.$$

Then the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 2$, and the associated eigenvectors $v_1 = (1, -1)$ and $v_2 = (2, 1)$. The phase portrait here we depict in Figure 3.3b. It is, as expected, a "distortion" of the phase portrait in Figure 3.3a with the standard basis vectors as eigenvectors.

3.4.1.4 Centres We next consider cases where A has complex eigenvalues, first looking at the case where the eigenvalues of A are purely imaginary, say $\lambda_1 = i\omega$ and $\lambda_2 = -i\omega$, with $\omega \in \mathbb{R}_{>0}$. In this case we say we have a ***centre***. The prototypical

case here is

$$\boldsymbol{A} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}.$$

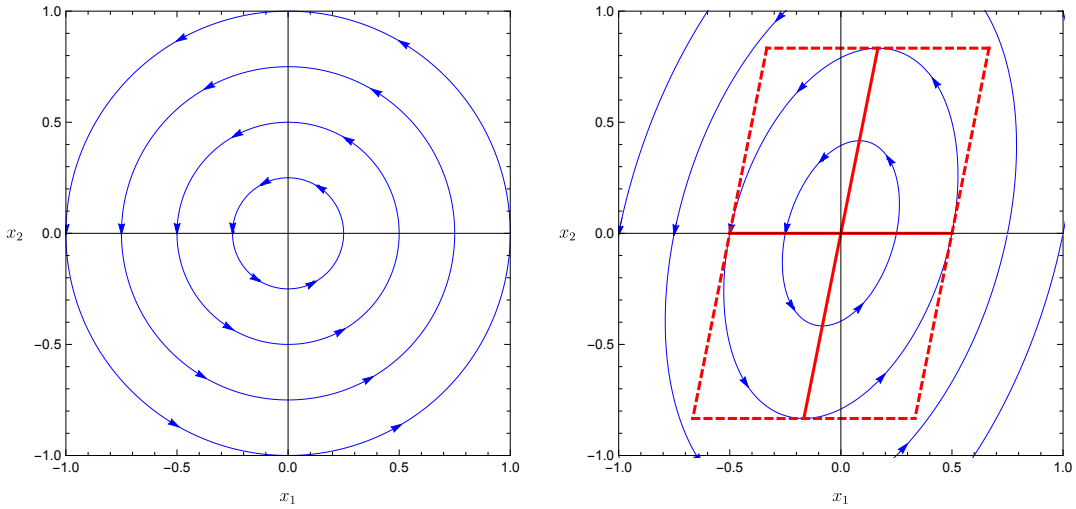
In this case we have, using Procedure 3.2.45,

$$\mathrm{e}^{At} = \begin{bmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{bmatrix}.$$

Note that, if

$$\begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} = \mathrm{e}^{At} \begin{bmatrix} \xi_1(0) \\ \xi_2(0) \end{bmatrix},$$

then $\|\boldsymbol{\xi}(t)\| = \|\boldsymbol{\xi}(0)\|$. Thus the parameterised solution curves reside in circles centred at $(0,0)$, and this is illustrated in Figure 3.4a.



(a) "Canonical" centre (b) "Noncanonical" centre

Figure 3.4 Centres

For more generic cases, the solutions will still be periodic, and the solution curves will then live on ellipses. To describe the ellipses, we suppose that we have eigenvalues $\lambda_1 = \mathrm{i}\omega$ and $\lambda_2 = -\mathrm{i}\omega$. We suppose that the associated eigenvectors are $\boldsymbol{w}_1 = \boldsymbol{u} + \mathrm{i}\boldsymbol{v}$ and $\boldsymbol{w}_2 = \boldsymbol{u} - \mathrm{i}\boldsymbol{v}$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$. To illustrate how $\boldsymbol{u}$ and $\boldsymbol{v}$ prescribe the ellipses traced out by solutions, we shall consider an example:

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} \\ \frac{5}{3} & -\frac{1}{3} \end{bmatrix}.$$

The eigenvalues in this case are $\lambda_1 = \mathrm{i}$ and $\lambda_2 = -\mathrm{i}$. The eigenvectors are $\boldsymbol{w}_1 = \boldsymbol{u} + \mathrm{i}\boldsymbol{v}$ and $\boldsymbol{w}_2 = \boldsymbol{u} - \mathrm{i}\boldsymbol{v}$, where

$$\boldsymbol{u} = (1, 5), \quad \boldsymbol{v} = (3, 0).$$

In Figure 3.4b we illustrate the phase portrait in this case, and also show scaled eigenvectors in red, and a box centred at $(0, 0)$ whose sides are parallel to the eigenvectors. As one can see, the ellipse along which solution curves evolve is the unique ellipse tangent to an appropriately scaled box.

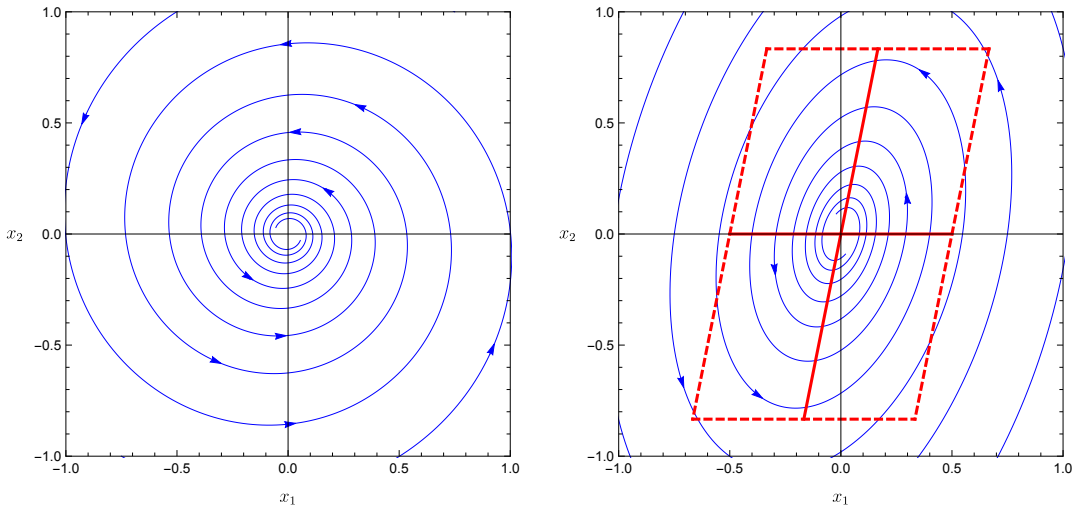
3.4.1.5 Stable spirals We continue thinking about cases with complex eigenvalues, but now we consider eigenvalues with nonzero real part. First we consider the situation where the real part is negative, this being called a ***stable spiral***. First let us consider the prototypical case where

$$A = \begin{bmatrix} \sigma & -\omega \\ \omega & \sigma \end{bmatrix},$$

with eigenvalues $\lambda_1 = \sigma + \mathrm{i}\omega$ and $\lambda_2 = \sigma - \mathrm{i}\omega$, where we take $\sigma \in \mathbb{R}_{<0}$. We have, using Procedure 3.2.45,

$$\mathrm{e}^{At} = \mathrm{e}^{\sigma t} \begin{bmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{bmatrix}.$$

The phase portrait in this case we depict in Figure 3.5a, and one can see why the



(a) "Canonical" stable spiral (b) "Noncanonical" stable spiral

Figure 3.5 Stable spirals

name "stable spiral" is applied in this case.

We can also consider a more generic case to illustrate, as in the case of centres, the rôle of the eigenvectors. We take

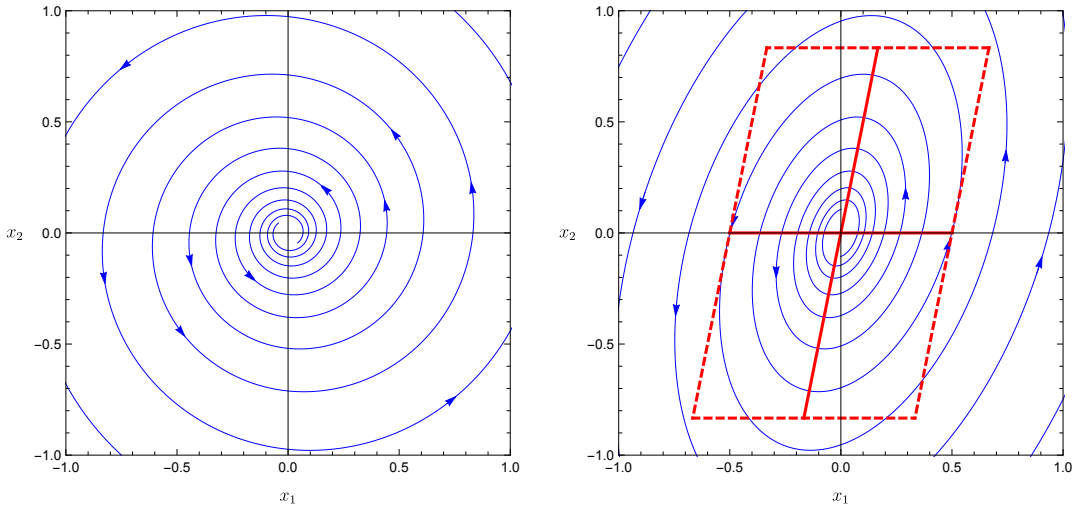
$$A = \begin{bmatrix} \frac{7}{30} & -\frac{2}{3} \\ \frac{5}{3} & -\frac{13}{30} \end{bmatrix},$$

and determine the eigenvalues to be $\lambda_1 = -\frac{1}{10} + \mathrm{i}$ and $\lambda_2 = -\frac{1}{10} - \mathrm{i}$. The eigenvectors are $w_1 = u + \mathrm{i}v$ and $w_2 = u - \mathrm{i}v$, where

$$u = (1, 5), \quad v = (3, 0),$$

i.e., the eigenvectors are the same as for the centre in the previous section. In Figure 3.5b we depict the phase plane in this case, and also overlay the box used to illustrate the rôle of the eigenvectors in the case of a centre.

3.4.1.6 Unstable spirals Next we consider the case where A has complex eigenvalues with positive real part, this being the case of an ***unstable spiral***. The "canonical" case is exactly like that for a stable spiral, except now $\sigma \in \mathbb{R}_{>0}$. The phase portrait in this case is depicted in Figure 3.6a. The situation is the "opposite" of



(a) "Canonical" unstable spiral (b) "Noncanonical" unstable spiral

Figure 3.6 Unstable spirals

that for the stable spiral in Figure 3.5a.

We can also give a more generic case by considering

$$A = \begin{bmatrix} \frac{13}{30} & -\frac{2}{3} \\ \frac{5}{3} & -\frac{7}{30} \end{bmatrix}.$$

In this case, the eigenvalues are $\lambda_1 = \frac{1}{10} + \mathrm{i}$ and $\lambda_2 = \frac{1}{10} - \mathrm{i}$ and the eigenvectors are $w_1 = u + \mathrm{i}v$ and $w_2 = u - \mathrm{i}v$, where

$$u = (1, 5), \quad v = (3, 0).$$

Note that these are the same eigenvalues as for the centre and the stable spiral considered above. In Figure 3.6b we show the phase portrait in this case, along with a box determined by the eigenvectors as in our discussion of the spiral above.

3.4.1.7 Nonisolated equilibria The remaining situations we consider are "degenerate" and do not arise as frequently as the preceding cases (although they *do* arise). All of these correspond to cases of a zero eigenvalue. Note that, if one has a zero eigenvalue and if v is any corresponding eigenvector, then any multiple of v is an equilibrium point for the differential equation. Thus, when one is considering cases with zero eigenvalues, the equilibrium point at $(0, 0)$ is not isolated.

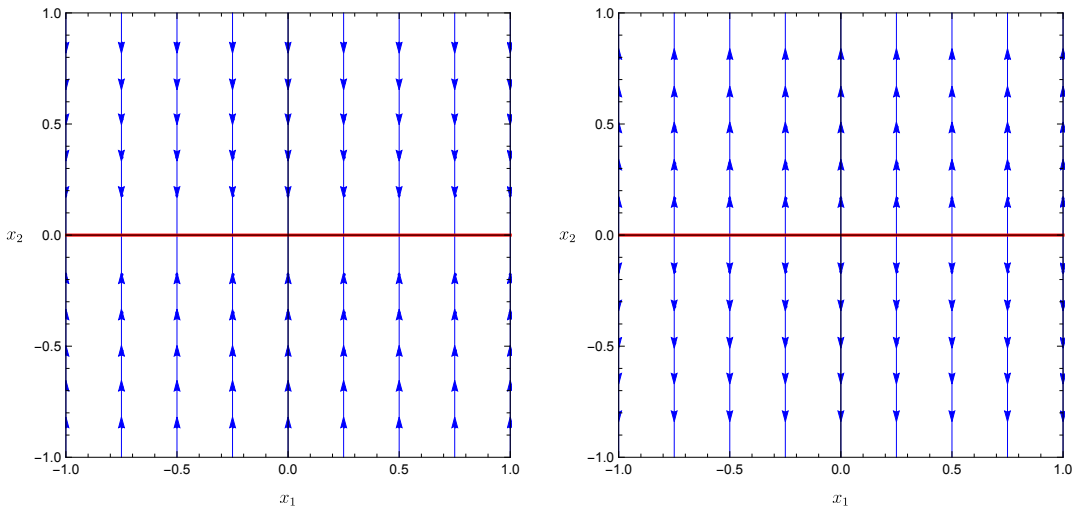
Let us consider the various cases.

Zero eigenvalue with algebraic multiplicity 1

We begin by supposing that A has eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \lambda \neq 0$. In this case, we suppose that

$$A = \begin{bmatrix} 0 & 0 \\ 0 & \lambda \end{bmatrix}.$$

The behaviour of the solution curves in the phase portrait depends on whether λ is positive or negative. In Figure 3.7a we depict the case when $\lambda \in \mathbb{R}_{<0}$. We see, in this



(a) One zero eigenvalue and one negative eigenvalue

(b) One zero eigenvalue and one positive eigenvalue

Figure 3.7 Zero eigenvalue with algebraic multiplicity 1

case, that the subspace (in red) generated by the eigenvector e_1 for the eigenvalue 0 is populated with equilibria, and that, because λ is negative, all solution curves approach one of these equilibria as $t \to \infty$.

The situation for $\lambda \in \mathbb{R}_{>0}$ is rather similar, and is depicted in Figure 3.7b.

Zero eigenvalue with algebraic multiplicity 2

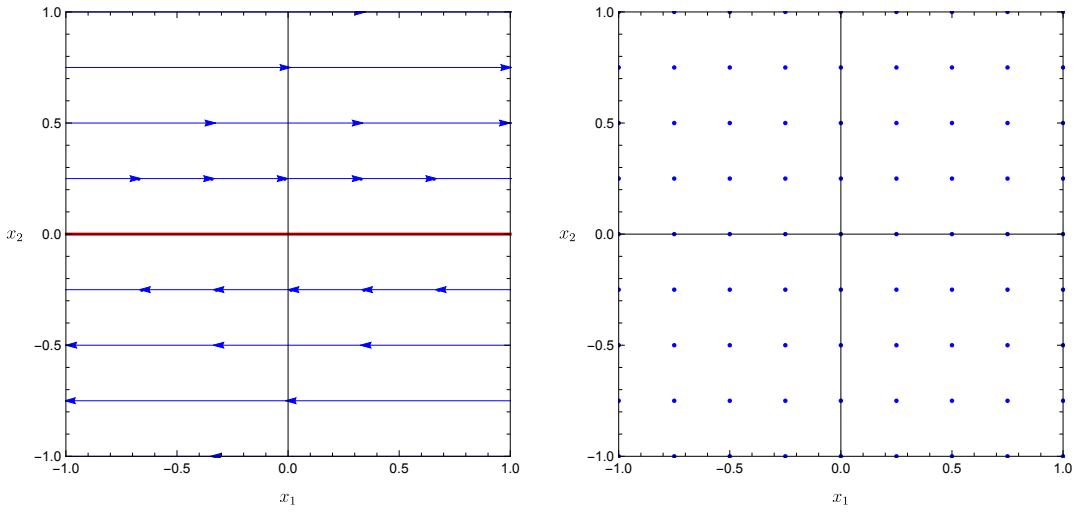
Finally we consider the case of a repeated zero eigenvalue. There are two situations to consider here, one when $m_{\mathrm{g}}(0, \boldsymbol{A}) = 1$ and another when $m_{\mathrm{g}}(0, \boldsymbol{A}) = 2$. In the former case, we consider

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and in the latter case we must have $\boldsymbol{A} = \boldsymbol{0}$. In the former case we have

$$\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} 0 & t \\ 0 & 0 \end{bmatrix}$$

and in the latter case we have $\mathrm{e}^{\boldsymbol{0}t} = \boldsymbol{I}_n$. In Figure 3.8a and Figure 3.8b we show



(a) Zero eigenvalue with algebraic multiplicity 2 and geometric multiplicity 1

(b) Zero eigenvalue with algebraic and geometric multiplicity 2

Figure 3.8 Zero eigenvalue with algebraic multiplicity 2

the phase portraits. Of course, the phase portrait in Figure 3.8b is spectacularly uninteresting, since it consists entirely of equilibria!

3.4.2 An introduction to phase portraits for nonlinear systems

The analysis of the preceding section for planar linear ordinary differential equations with constant coefficients was quite comprehensive, exactly because the setting was so simple. Extensions to either higher-dimensions than planar and/or to nonlinear ordinary differential equations are difficult, the former for reasons of difficulty of representation, the latter for reasons of plain ol' difficulty. In this section we consider some *ad hoc* techniques for understanding phase portraits for planar nonlinear ordinary differential equations.

3.4.2.1 Phase portraits near equilibrium points

3.4.2.2 Periodic orbits

3.4.2.3 Attractors

3.4.3 Extension to higher dimensions

3.4.3.1 Behaviour near equilibria

3.4.3.2 Attractors

Exercises

3.4.1 For the scalar linear homogeneous ordinary differential equations in $\mathbb{R}^2$ defined by the following 2×2 matrices, do the following:

1. determine what type of planar linear system this is, i.e., "stable node," "unstable node," "saddle point," etc.;
2. sketch the phase portrait, clearly indicating the essential features (knowing what these are is part of the question).

(a) $A = \begin{bmatrix} 2 & -5 \\ 0 & 3 \end{bmatrix}$;

(b) $A = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$;

(c) $A = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix}$;

(d) $A = \begin{bmatrix} 4 & -1 \\ 4 & 0 \end{bmatrix}$;

(e) $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$;

(f) $A = \begin{bmatrix} -4 & 6 \\ -1 & 1 \end{bmatrix}$;

(g) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$;

(h) $A = \begin{bmatrix} 2 & 4 \\ -2 & 6 \end{bmatrix}$;

(i) $A = \begin{bmatrix} -4 & 9 \\ -1 & 2 \end{bmatrix}$.

Section 3.5

Using a computer to work with systems of ordinary differential equations

We thank Jack Horn for putting together the MATHEMATICA® *and* MATLAB® *results in this section.*

In this section we illustrate how to use computer packages to obtain analytical and numerical solutions for systems of ordinary differential equations. We restrict our attention to attention to linear equations with constant coefficients, since these are really the only significant class of equations that one can work with analytically. For numerical solutions, the techniques here are extended in the obvious way to nonlinear or time-varying systems. As in Section 2.4, we restrict our attention to illustrating the use of MATHEMATICA® and MATLAB®.

3.5.1 Using MATHEMATICA® to obtain analytical and/or numerical solutions

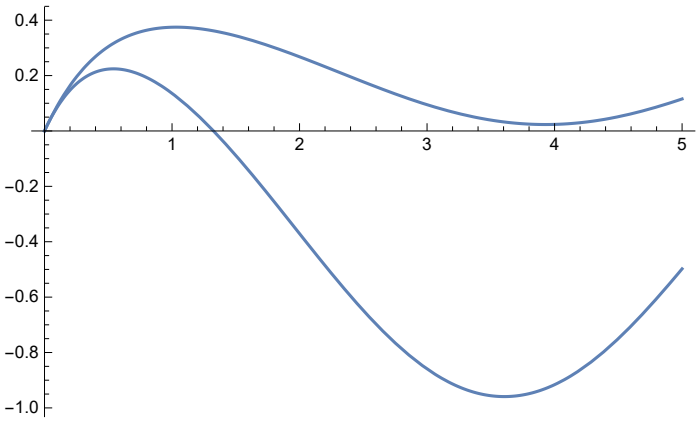
Solving systems of differential equations in MATHEMATICA® requires a similar procedure as solving a single ordinary differential equation. You must use the `DSolve` command, while keeping your system in the form $\frac{dx}{dt}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$, for a given matrix A and vector function $\boldsymbol{f}$.

3.5.1 Example (Using `DSolve` to solve systems of ordinary differential equations) The first system we will consider is:

$$\frac{d\boldsymbol{y}}{dt}(t) = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix} \boldsymbol{y}(t) + \begin{bmatrix} \cos(t) \\ 1 \end{bmatrix}$$

The following script will find and plot the solutions to this system.

```
A = {{-1, -2}, {1, -3}};

Y[t_] = {y1[t], y2[t]};

solution = DSolve[{Y'[t] == A.Y[t] + {Cos[t], 1}, Y[0] == {0, 0}}, {y1, y2}, t];

Plot[{y1[t], y2[t]}/.solution, {t, 0, 5}]
```



Note that the "." in MATHEMATICA® means matrix-vector multiplication in the above code. •

3.5.2 Example (Matrix exponential in MATHEMATICA®) MATHEMATICA® is also an incredibly handy software for various aspects of linear algebra. In this example we will work with the matrix

$$A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

and will compute matrix exponentials, first using the `MatrixExp` command, then by following the process in Procedure 3.2.48.

A = {{−1, 1, 0}, {−1, −1, 0}, {0, 0, 2}};

MatrixExp[t * A]//MatrixForm

$$\begin{pmatrix} e^{-t}\text{Cos}[t] & e^{-t}\text{Sin}[t] & 0 \\ -e^{-t}\text{Sin}[t] & e^{-t}\text{Cos}[t] & 0 \\ 0 & 0 & e^{2t} \end{pmatrix}$$

Now we will follow the steps from class, and compare the results.

Eigenvals = Eigenvalues[A];

Eigenvect = Eigenvectors[A];

F1 = Exp[t * Eigenvals[[1]]] * Eigenvect[[1]];

F2 = Exp[t * Eigenvals[[2]]] * Eigenvect[[2]];

F3 = Exp[t * Eigenvals[[3]]] * Eigenvect[[3]];

Fund = Transpose[{F1, F2, F3}];

FundInv = Inverse[Fund];

B = FundInv/.$t \to 0$;

Indirect = Fund.B//MatrixForm

This "indirect" method gives us the ugly looking matrix shown below:

$$\begin{pmatrix} \frac{1}{2}e^{(-1-i)t} + \frac{1}{2}e^{(-1+i)t} & \frac{1}{2}ie^{(-1-i)t} - \frac{1}{2}ie^{(-1+i)t} & 0 \\ -\frac{1}{2}ie^{(-1-i)t} + \frac{1}{2}ie^{(-1+i)t} & \frac{1}{2}e^{(-1-i)t} + \frac{1}{2}e^{(-1+i)t} & 0 \\ 0 & 0 & e^{2t} \end{pmatrix}$$

However, this is equivalent to the matrix found by using the `MatrixExp` command, which can be seen by applying the `ComplexExpand` command.

ComplexExpand[Indirect]//MatrixForm

$$\begin{pmatrix} e^{-t}\text{Cos}[t] & e^{-t}\text{Sin}[t] & 0 \\ -e^{-t}\text{Sin}[t] & e^{-t}\text{Cos}[t] & 0 \\ 0 & 0 & e^{2t} \end{pmatrix}$$

Sometimes it is not so easy to see that identical symbolic expressions in MATHEMATICA® are, in fact, identical. For things that are not excessively disgusting to look at, sometimes the `Simplify` command is useful. For complex things, `ComplexExpand` is sometimes useful. •

Next we consider inhomogeneous equations, using Corollary 3.3.3.

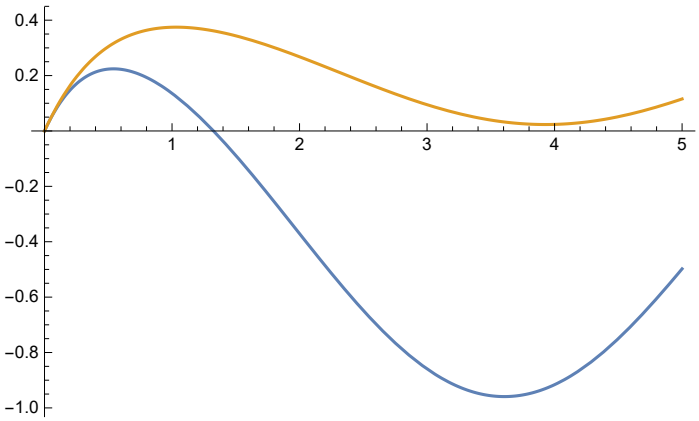
3.5.3 Example (Inhomogeneous linear systems of equations using MATHEMATICA®) Now that we are comfortable with commands such as `MatrixExp`, we will see how it is also possible to solve systems of ordinary differential equations using the formula

$$x(t) = \mathrm{e}^{At}x_0 + \int_0^{t_0} \mathrm{e}^{A(t_0-\tau)}f(\tau)\,\mathrm{d}\tau.$$

We will show this by solving the same system given Exercise 3.5.1.

x[t] = MatrixExp[t * A].{0, 0} + Integrate[MatrixExp[A * (t − T)].{Cos[T], 1}, {T, 0, t}];

Plot[x[t], {t, 0, 5}]



As you can see, the plots are identical to the direct results in Exercise 3.5.1. •

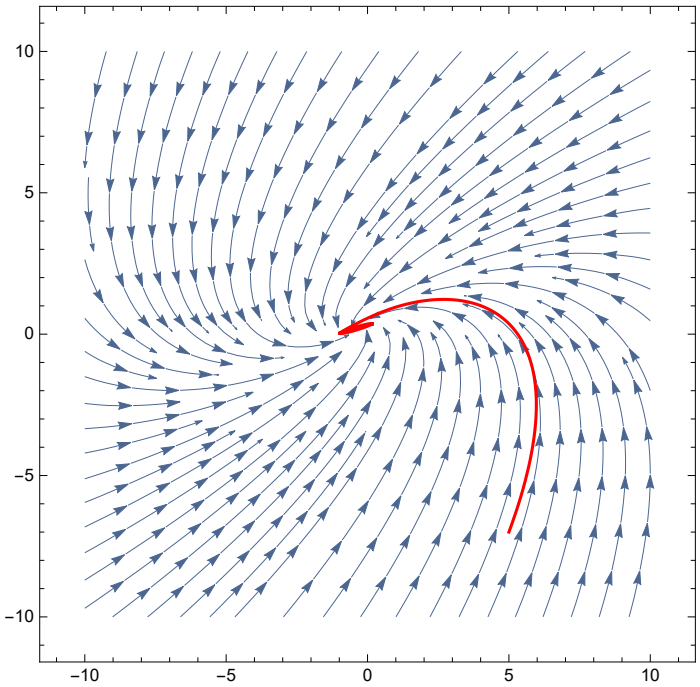
One can also use MATHEMATICA® to produce phase portraits. There are sophisticated MATHEMATICA® packages for doing this (we used DynPac for the plots from Section 3.4), and here we shall indicate how to do this with standard MATHEMATICA® commands.

3.5.4 Example (Phase plane using MATHEMATICA®) We consider the planar system of linear equations

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \end{bmatrix} = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} \cos(t) \\ 1 \end{bmatrix}.$$

We use the commands `StreamPlot` and `ParametricPlot`.

```
splot = StreamPlot[{−x − 2y, x − 3y}, {x, −10, 10}, {y, −10, 10}];

Show[splot,

ParametricPlot[

Evaluate[

First[

{x[t], y[t]}/.DSolve[{x'[t] == −x[t] − 2 * y[t] + Cos[t], y'[t] == x[t] − 3 * y[t] + 1,

{x[0], y[0]} == {5, −7}}, {x[t], y[t]}, t]]], {t, 0, 10}, PlotStyle → Red]]
```



We have plotted, using `StreamPlot`, the phase plane for the homogeneous system, and superimposed in red one solution for the inhomogeneous system. •

3.5.2 Using MATLAB® to obtain numerical solutions

In MATLAB®, solving systems of differential equations is not much different than solving a single ordinary differential equation. You must create a function for your system, which must then be passed into a script that will use the `ode45` solver.

3.5.5 Example (Using `ode45` to solve systems of ordinary differential equations) We will once again be considering the same examples as we did in MATHEMATICA®, this time using MATLAB®. First we will solve the following system:

$$\frac{d\boldsymbol{y}}{dt}(t) = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix} \boldsymbol{y}(t) + \begin{bmatrix} \cos(t) \\ 1 \end{bmatrix}.$$

```
function [ dydt ] = Example2( t,y )

A = [-1 -2;1 -3];

dydt = A*y + [cos(t); 1];

end
```

Below is the main script that will plot the solution to this system. See Figure 3.9 for the MATLAB® generated plots.

```
clc
clear all
close all
%% Solving Numerically

t  = linspace(0,5);
y0 = [0 0];

y = ode45(@(t,y)Example2(t,y),t,y0);

plot(y.x,y.y)
xlabel('Time [s]');
ylabel('y(t)');
legend('y1(t)','y2(t)');
```

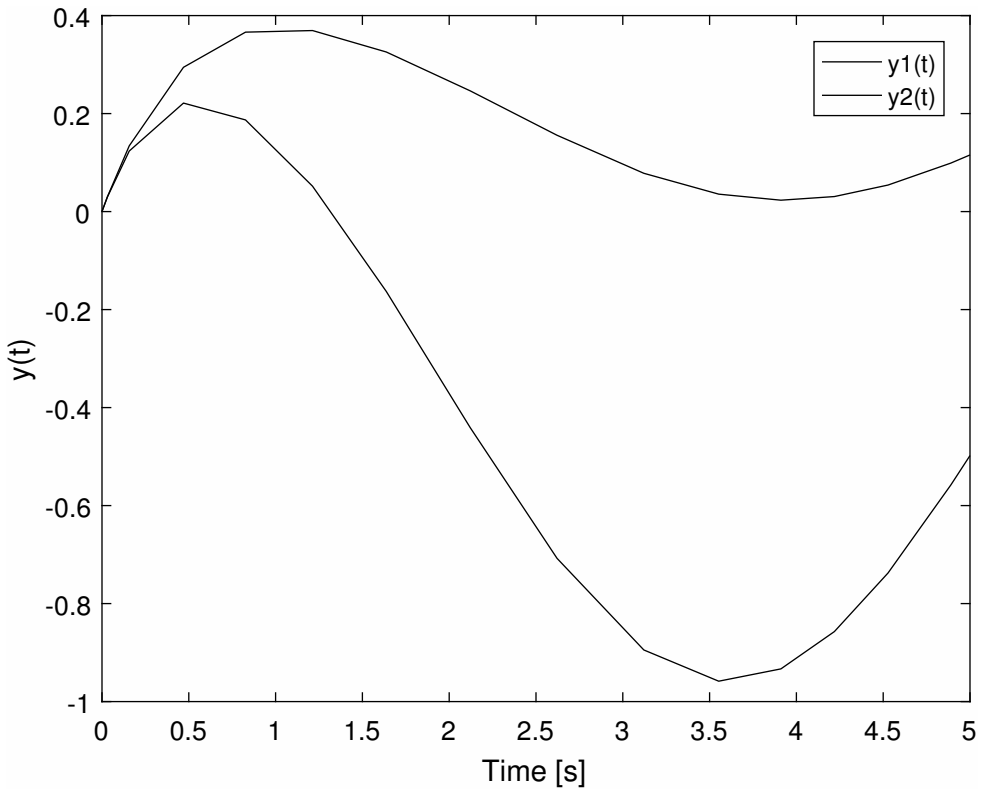


Figure 3.9 Plots generated by MATLAB® for Exercise 3.5.5

One can see that the solutions are quite similar to those from Exercise 3.5.1 using MATHEMATICA®. The jagged character of the plots is indicative of the fact that the time step for `ode45` can be decreased. This can be done by specifying

```
t_int = tinit:tstep:tfinal
```

where the meaning of `tinit`, `tfinal`, and `tstep` is just what you think they are. •

MATLAB® is also very useful for linear algebra.

3.5.6 Example (Matrix exponential in MATLAB®) We will consider the same matrix exponential example

$$A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

as in Example 3.5.2. Again, it is possible to calculate the matrix exponential both directly (using the `expm` command), or you can follow the steps from Procedure 3.2.48.

```
clc
clear all
close all

%% Calculating Matrix Exponential Directly
A = [-1 1 0; -1 -1 0; 0 0 -2];
syms t
MatrixExpDirect = expm(t*A)
%% Calculating Matrix Exponential Using Procedure from Class

[EigenVectors,EigenValues] = eig(t*A);

F1 = exp(EigenValues(1,1)).*EigenVectors(:,1);
F2 = exp(EigenValues(2,2)).*EigenVectors(:,2);
F3 = exp(EigenValues(3,3)).*EigenVectors(:,3);

Fund = [F1 F2 F3];
FundInv = inv(Fund);
B = subs(FundInv,0); %Here we are evaluating the fundamental
    matrix at t = 0

MatrixExponential = Fund*B
```

Here is the output from the MATLAB® code

```
MatrixExponential =
[exp(t*(-1-1i))/2+exp(t*(-1+1i))/2,
(exp(t*(-1-1i))*1i)/2-(exp(t*(-1+1i))*1i)/2, 0]
[-(exp(t*(-1-1i))*1i)/2+(exp(t*(-1+1i))*1i)/2,
exp(t*(-1-1i))/2+exp(t*(-1+1i))/2, 0]
[0, 0, exp(-2*t)]
```

Of course, the result here is the same as we saw using MATHEMATICA®. •

Finally, let us see how MATLAB® can be used to create phase portraits.

3.5.7 Example To create phase portraits in MATLAB®, you must use the `meshgrid` command, and evaluate the first derivatives of y_1 and y_2 at each point for $t = 0$. Once you have done this, use the `quiver` command to plot the vector field. To plot a specific solution, simply use the `ode45` command, and plot the first and second columns of the outputted matrix. See Figure 3.10 for the result.

```
clc
clear all
close all

[x,y] = meshgrid(-10:1:10,-10:1:10);

u = zeros(size(x));
v = zeros(size(x));

t = 0;
for i = 1:numel(x)
    dydt = Example2(t,[x(i);y(i)]);
    u(i) = dydt(1);
    v(i) = dydt(2);
end

quiver(x,y,u,v,'b');
xlabel('y1(t)');
ylabel('y2(t)');

t = linspace(0,5);
y0 = [5,-7];

hold on
y = ode45(@(t,y)Example2(t,y),t,y0);
plot(y.y(1,:),y.y(2,:))
hold off

print -deps PhasePortrait
```

It is possible to customise MATLAB® output to look prettier, but this is something we leave to the reader as they progress through their professional lives. •

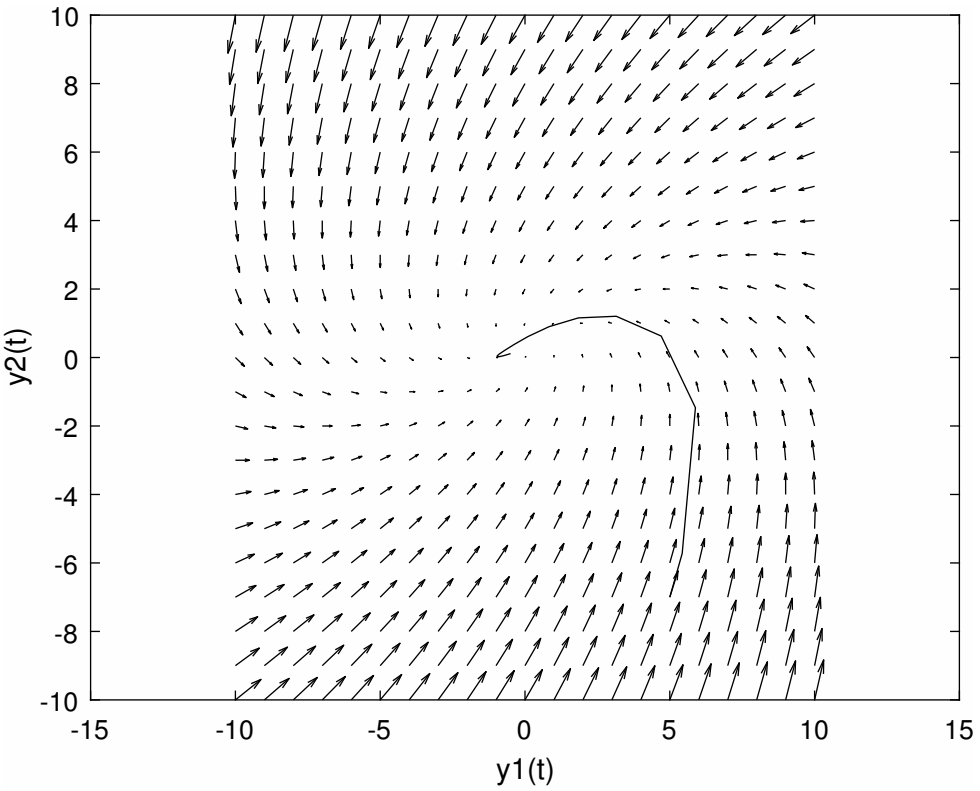


Figure 3.10 Phase portrait generated by MATLAB®

Chapter 4

Stability theory for ordinary differential equations

In the preceding two chapters we considered some methods for solving ordinary differential equations, dealing almost exclusively with linear equations. In Section 3.1 we motivated our rationale for this by illustrating that systems of ordinary differential equations can be linearised, although we did not at that time indicate how this process of linearisation might be useful. In this chapter we shall see, among other things, a concrete illustration of why one is interested in linear ordinary differential equations, namely that understanding them can help one understand the stability of systems that are not necessarily linear. Indeed, in this chapter we shall engage in a general discussion of stability, and this connection to linear ordinary differential equations will be just one of the topics considered.

We shall begin our general presentation in Section 4.1 with definitions of various types of stability and examples that illustrate these. We shall give many definitions here, and shall only consider a few of them in any detail subsequently. However, the full slate of definitions is useful for establishing context. In Section 4.2 we consider the stability of systems of linear ordinary differential equations, where the extra structure, especially in the case of systems with constant coefficients, allows a complete description of stability. Two methods, called "Lyapunov's First and Second Method," for stability analysis for systems of (not necessarily linear) ordinary differential equations are considered in Sections 4.3 and 4.4. Lyapunov's First Method allows the determination of the stability of a system of differential equations from its linearisation in some cases.

Contents

Section 4.1

Stability definitions

In this section we state the standard stability definitions for a system of ordinary differential equations. Thus we are working with an ordinary differential equation $\boldsymbol{F}$ with right-hand side

$$\widehat{\boldsymbol{F}}\colon \mathbb{T} \times U \to \mathbb{R}^n,$$

where $U \subseteq \mathbb{R}^n$ is an open subset of $\mathbb{R}^n$. In order to ensure local existence and uniqueness of solutions, we shall make the following assumptions on $\boldsymbol{F}$.

4.1.1 Assumption (Right-hand side assumptions for stability definitions) We suppose that

(i) the map $t \mapsto \widehat{\boldsymbol{F}}(t, \boldsymbol{x})$ is continuous for each $\boldsymbol{x} \in U$,

(ii) the map $\boldsymbol{x} \mapsto \widehat{\boldsymbol{F}}(t, \boldsymbol{x})$ is Lipschitz for each $t \in \mathbb{T}$, and

(iii) for each $\boldsymbol{x} \in U$ and for each $r \in \mathbb{R}_{>0}$, there exist continuous functions $g, L\colon \mathbb{T} \to \mathbb{R}_{\geq 0}$ such that

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{y})\| \leq g(t), \qquad (t, \boldsymbol{y}) \in \mathbb{T} \times \mathsf{B}(r, \boldsymbol{x}),$$

and

$$\|\widehat{\boldsymbol{F}}(t, \boldsymbol{y}_1) - \widehat{\boldsymbol{F}}(t, \boldsymbol{y}_2)\| \leq L(t)\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|, \qquad t \in \mathbb{T},\ \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathsf{B}(r, \boldsymbol{x}).$$ •

4.1.1 Definitions

The first thing one should address when talking about stability is "stability of what?" Almost always—and always for us—we will be thinking about stability of a solution $t \mapsto \xi_0(t)$ of a system of ordinary differential equations $\boldsymbol{F}$. In all cases, stability of a solution intuitively means that other solutions starting nearby remain nearby at $t \to \infty$. However, this intuitive idea needs to be made precise. As part of this, we make the following definitions.

4.1.2 Definition (ϵ-neighbourhood of a curve) Let $U \subseteq \mathbb{R}^n$ be open, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let $\gamma\colon \mathbb{T} \to U$ be a curve. The set

$$\mathcal{N}(\gamma, \epsilon) = \{\boldsymbol{x} \in U \mid \|\boldsymbol{x} - \gamma(t)\| < \epsilon \text{ for some } t \in \mathbb{T}\}$$

is the ***ϵ-neighbourhood*** of γ. •

4.1.3 Definition (Distance to a set) Let $U \subseteq \mathbb{R}^n$ be open and let $S \subseteq U$. The function

$$\begin{aligned} \mathrm{d}_S\colon U &\to \mathbb{R}_{\geq 0} \\ \boldsymbol{x} &\mapsto \inf\{\|\boldsymbol{x} - \boldsymbol{y}\| \mid \boldsymbol{y} \in S\} \end{aligned}$$

is the ***distance function to*** $\mathbf{S}$. •

We can now state our stability definitions.

4.1.4 Definition (Stability of solutions) Let $\boldsymbol{F}$ be a system of ordinary differential equations satisfying Assumption 4.1.1 and suppose that $\sup \mathbb{T} = \infty$.[1] Let $\xi_0 \colon \mathbb{T}' \to U$ be a solution for $\boldsymbol{F}$, supposing that $\sup \mathbb{T}' = \infty$. The solution ξ_0 is:

(i) ***Lyapunov stable***, or merely ***stable***, if, for any $\epsilon \in \mathbb{R}_{>0}$ and $t_0 \in \mathbb{T}'$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| < \epsilon$ for $t \geq t_0$;

(ii) ***asymptotically stable*** if it is stable and if, for every $t_0 \in \mathbb{T}'$, there exists $\delta \in \mathbb{R}_{>0}$ such that, for $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| < \epsilon$ for $t \geq t_0 + T$;

(iii) ***exponentially stable*** if it is stable and if, for every $t_0 \in \mathbb{T}'$, there exists $M, \delta, \sigma \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| \leq M\mathrm{e}^{-\sigma(t-t_0)}$;

(iv) ***orbitally stable*** if, for any $\epsilon \in \mathbb{R}_{>0}$ and $t_0 \in \mathbb{T}'$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\xi(t) \in \mathcal{N}(\xi_0, \epsilon)$ for $t \geq t_0$;

(v) ***asymptotically orbitally stable*** if it is orbitally stable and if, for every $t_0 \in \mathbb{T}'$, there exists $\delta \in \mathbb{R}_{>0}$ such that, for $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\mathrm{d}_{\mathrm{image}(\xi_0)}(\xi(t)) < \epsilon$ for $t \geq t_0 + T$;

(vi) ***exponentially orbitally stable*** if it is orbitally stable and if, for every $t_0 \in \mathbb{T}'$, there exists $M, \sigma, \delta \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{x} \in U$ satisfies $\|\xi_0(t_0) - \boldsymbol{x}\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

is defined on $[t_0, \infty)$ and satisfies $\mathrm{d}_{\mathrm{image}(\xi_0)}(\xi(t)) \leq M\mathrm{e}^{-\sigma(t-t_0)}$;

[1] Thus $\mathbb{T}$ is a time-interval that is unbounded on the right, i.e., either $\mathbb{T} = [a, \infty)$ or $\mathbb{T} = (a, \infty)$ for some $a \in \mathbb{R}$.

(vii) ***uniformly Lyapunov stable***, or merely ***uniformly stable***, if, for any $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| < \epsilon$ for $t \geq t_0$;

(viii) ***uniformly asymptotically stable*** if it is uniformly stable and if there exists $\delta \in \mathbb{R}_{>0}$ such that, for $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| < \epsilon$ for $t \geq t_0 + T$;

(ix) ***uniformly exponentially stable*** if it is uniformly stable and if there exists $M, \sigma, \delta \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\|\xi(t) - \xi_0(t)\| \leq M e^{-\sigma(t-t_0)}$;

(x) ***uniformly orbitally stable*** if, for any $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\xi(t) \in \mathcal{N}(\xi_0, \epsilon)$ for $t \geq t_0$;

(xi) ***uniformly asymptotically orbitally stable*** if it is uniformly orbitally stable and if there exists $\delta \in \mathbb{R}_{>0}$ such that, for $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\mathrm{d}_{\mathrm{image}(\xi_0)}(\xi(t)) < \epsilon$ for $t \geq t_0 + T$;

(xii) ***uniformly exponentially orbitally stable*** if it is uniformly orbitally stable and if there exists $M, \sigma, \delta \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T}' \times U$ satisfies $\|\xi_0(t_0) - x\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

is defined on $[t_0, \infty)$ and satisfies $\mathrm{d}_{\mathrm{image}(\xi_0)}(\xi(t)) \leq M e^{-\sigma(t-t_0)}$;

(xiii) ***unstable*** if it is not stable. •

While this seems like an absurdly large number of definitions, it is made to appear larger by there being a few concepts, represented in all possible combinations. Let us describe the essential dichotomies and trichotomies.

1. *Stable/(asymptotically stable)/(exponentially stable).* The idea of the dichotomy of stable/(asymptotically stable) is that stability has to do with solutions remaining close if their initial conditions are close, while asymptotic stability has to do with solutions with close initial conditions getting closer and closer as time goes by. The notion of exponential stability is similar to that of asymptotic stability, but places some constraints on the rate at which solutions with nearby initial conditions approach one another.
2. *Stable/(orbitally stable).* The stable/(orbitally stable) dichotomy has to do with how one measures the "closeness" of solutions with nearby initial conditions. When dealing with stability, as opposed to orbital stability, one asks that, at all times, solutions remain close. Orbital stability is weaker in that we do not ask that solutions at the same time are close, but rather that one solution at one time is close to another solution, but possibly at a different time.
3. *Stable/(uniformly stable).* The dichotomy here here has to do with the rôle of the initial time t_0 in the definition. In uniform stability, the parameters δ, M, and σ are independent of the initial time t_0, whereas with (nonuniform) stability, these parameters depend on t_0. This is a more or less standard occurrence of the notion of "uniform," and if a reader is encountering this notion for the first time, it is best to acquire a feeling for what it represents.

Now that we have presented our definitions and tried to understand what they mean, let us explore them a little. First let us consider the relationships between the various notions of stability. To do this it is most convenient to arrange the various definitions in a diagram. To control the clutter in the diagram and other places, we use some obvious abbreviations:

(U)S	(uniformly) stable
(U)AS	(uniformly) asymptotically stable
(U)ES	(uniformly) exponentially stable
(U)OS	(uniformly) orbitally stable
(U)AOS	(uniformly) asymptotically orbitally stable
(U)EOS	(uniformly) exponentially orbitally stable

With these abbreviations, we have the diagram in Figure 4.1 illustrating the relationships between the various forms of stability. All of the implications in the diagram follow more or less immediately from the definitions.

Next let us see that, in the case of most interest to us where the solution ξ_0 is an equilibrium solution, the preceding definitions simplify by a factor of $\frac{1}{2}$. Thus, in this discussion, we have an equilibrium state x_0 for F, i.e., $\widehat{F}(t, x_0) = 0$ for all $t \in \mathbb{T}$. In this case, as per Proposition 3.1.5, we have the equilibrium solution ξ_0 defined by $\xi_0(t) = 0$, $t \in \mathbb{T}$. The usual linguistic simplification is to speak, not of the stability of this equilibrium solution, but of the stability of the equilibrium state x_0 since the latter prescribes the former.

The next result records the simplifications that occur in the stability definitions in this case.

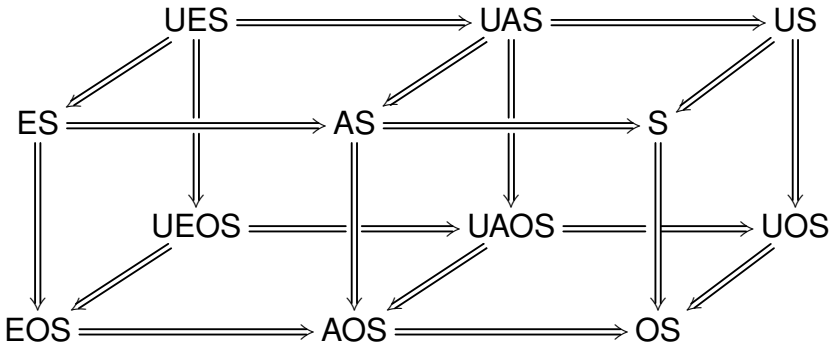


Figure 4.1 Relationships between the various forms of stability

4.1.5 Proposition (Collapsing of stability definitions for equilibria) *Let* $\mathbf{F}$ *be a system of ordinary differential equations satisfying Assumption 4.1.1 and suppose that* $\sup \mathbb{T} = \infty$. *For an equilibrium state* $\mathbf{x}_0$ *for* $\mathbf{F}$, *we have the following implications:*

(i) OS $\implies$ S; *(iv)* UOS $\implies$ US;
(ii) AOS $\implies$ AS; *(v)* UAOS $\implies$ AOS;
(iii) EOS $\implies$ ES; *(vi)* UEOS $\implies$ UES.

In short, all forms of orbital stability are implied by their nonorbital counterparts in the case of equilibrium solutions.

Moreover, if $\mathbf{F}$ *is autonomous, then we additionally have the following implications:*

(vii) S $\implies$ US;
(viii) AS $\implies$ UAS;
(ix) ES $\implies$ UES.

Proof In all cases, this amounts to the observation that, if ξ_0 is the equilibrium solution $\xi_0(t) = x_0$, then $\mathcal{N}(\xi_0, \epsilon) = \mathsf{B}(\epsilon, x_0)$, and so

1. $x \in \mathcal{N}(\xi_0, \epsilon)$ if and only if $\|x - x_0\| < \epsilon$ and
2. $\mathrm{d}_{\mathrm{image}(\xi_0)}(x) = \|x - x_0\|$. ■

For the final assertion of the proposition, we shall explicitly give the proof that S $\implies$ US, the other implications following using the same idea. Let $\epsilon \in \mathbb{R}_{>0}$. Since x_0 is stable, for $t_0 \in \mathbb{T}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $x \in U$ satisfies $\|x - x_0\| < \delta$, the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

exists for $t \geq t_0$ and satisfies $\|\xi(t) - x_0\| < \epsilon$ for $t \geq t_0$. Now let $\hat{t}_0 \in \mathbb{T}$. Then, let $x \in U$ be such that $\|x - x_0\| < \delta$ and let $\xi \colon \mathbb{T} \to U$ and $\hat{\xi} \colon \mathbb{T} \to U$ be the solutions to the initial value problems

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

and

$$\dot{\hat{\xi}}(t) = \widehat{F}(t, \hat{\xi}(t)), \quad \hat{\xi}(t_0') = x,$$

respectively. By Exercise 1.3.19 we have $\hat{\xi}(t) = \xi(t-(\hat{t}_0-t_0))$. Therefore, $\hat{\xi}$ is defined for $t \geq \hat{t}_0$ and

$$\|\hat{x}(t) - x_0\| = \|x(t - (\hat{t}_0 - t_0)) - x_0\| < \epsilon$$

for $t \geq \hat{t}_0$. This shows that the choice of δ can be made independently of the initial time t_0, and so x_0 is uniformly stable.

We conclude our discussion of stability definitions with a warning of some lurking dangers in these definitions.

4.1.6 Remarks (Caveats concerning stability definitions)

1. First let us provide some good news. For stability of equilibria—by far the most widely used and interesting case—the definitions we give are completely standard and coherent and offer no difficulties in their use.
2. It is often possible to reduce the study of stability of nonequilibrium solutions to the study of equilibria. Let us illustrate how this is done. We suppose that we have an ordinary differential equation $\boldsymbol{F}$ with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n$$

with $\sup \mathbb{T} = \infty$. Let us suppose that we have a solution $\xi_0\colon \mathbb{T} \to U$ for $\boldsymbol{F}$, whose stability we wish to examine. In order to do this, we suppose that there exists $r \in \mathbb{R}_{>0}$ such that the "tube"

$$T(r, \xi_0) = \{\xi_0(t) + x' \mid t \in \mathbb{T},\ x' \in \mathsf{B}(r, \mathbf{0})\}$$

of radius r about ξ_0 is a subset of U. We then define a "time-varying change of coordinates"

$$\begin{aligned} \boldsymbol{\Phi}\colon \mathbb{T} \times T(r, \xi_0) &\to \mathbb{T} \times \mathsf{B}(r, \mathbf{0}) \\ (t, x) &\mapsto (t, x - \xi_0(t)). \end{aligned}$$

We then define a differential equation $\boldsymbol{G}$ with right-hand side

$$\begin{aligned} \widehat{G}\colon \mathbb{T} \times \mathsf{B}(r, \mathbf{0}) &\to \mathbb{R}^n \\ (t, y) &\mapsto \widehat{F} \circ \boldsymbol{\Phi}^{-1}(t, y), \end{aligned}$$

whose state space is $\mathsf{B}(r, \mathbf{0})$. Note that, if $\xi\colon \mathbb{T}' \to U$ is a solution for $\boldsymbol{F}$ for which $\xi(t) - \xi_0(t) \in \mathsf{B}(r, \mathbf{0})$, then the function $\eta(t) = \xi(t) - \xi_0(t)$ is a solution for $\boldsymbol{G}$. Indeed,

$$\dot{\eta}(t) = \dot{\xi}(t) - \dot{\xi}_0(t) = \widehat{F}(t, \xi(t)) = \widehat{F}(t, \eta(t) + \xi_0(t)) = \widehat{G}(t, \eta(t)).$$

Moreover, since $\boldsymbol{\Phi} \circ \xi_0(t) = (t, \mathbf{0})$ for every $t \in \mathbb{T}$, the solution ξ_0 is mapped to the equilibrium solution $\eta_0\colon t \mapsto \mathbf{0}$. Therefore, the study of the stability of solution

ξ_0 is reduced to the study of the equilibrium solution at $\mathbf{0}$. In this way, the study of nonequilibrium solutions can sometimes be reduced to the study of equilibrium solutions. Note, also, that, even if $\boldsymbol{F}$ is autonomous, the resulting differential equation $\boldsymbol{G}$ will be nonautonomous.

3. Now for the bad news. For stability of nonequilibrium solutions, there are some possible problems with the definitions that need to be understood. The problems manifest themselves in at least two different ways, and these two ways are not unrelated.
   (a) The ϵ-neighbourhood of a solution is measured using a specific notion of distance coming from the Euclidean norm. It is possible that this is not the most meaningful way of measuring distance, and that, upon choosing another way of measuring distance, one can get inconsistent conclusions when applying stability tests. For example, one might use one method of measuring distance and conclude stability, while another method of measuring distance yields instability. To see examples of where this can happen requires understanding "other ways of measuring distance," and this is not something we shall do here.
   (b) The definitions we give can vary with coordinate systems. That is, one can render a stable (or unstable) system unstable (or stable) by using different coordinates. The reader is asked to explore this in Exercise 4.1.1.

   These caveats need to be kept in mind when working with the stability of nonequilibrium solutions. •

4.1.2 Examples

In this section, we give some examples to illustrate some of the ways in which the different notions of stability are separated in practice.

4.1.7 Example (Stable versus unstable versus asymptotically stable I) We consider the ordinary differential equation F with state space $U = \mathbb{R}$ and with right-hand side $\widehat{F}(t, x) = ax$ with $a \in \mathbb{R}$. This is a simple linear ordinary differential equation and has solution $\xi(t) = \xi(t_0)\mathrm{e}^{a(t-t_0)}$. We shall consider the stability of the equilibrium point $x_0 = 0$. We have three cases.

1. $a < 0$: In this case we note two things. First of all, $|\xi(t)| \leq |\xi(t_0)|$ for $t \geq t_0$, from which we conclude that the equilibrium at $x_0 = 0$ is stable. (Formally, let $\epsilon \in \mathbb{R}_{>0}$. Then, if we take $\delta = \epsilon$, we have

$$|\xi(t_0) - 0| \leq \delta \implies |\xi(t) - 0| < \epsilon, \qquad t \geq t_0.$$

   which is what is required to prove stability of the equilibrium $x_0 = 0$.) Also, $\lim_{t\to\infty}|\xi(t) - 0| = 0$, which gives asymptotic stability of $x_0 = 0$. Moreover, in this case we also have $|\xi(t)| = |\xi(t_0)|\mathrm{e}^{a(t-t_0)}$, and so we further have exponential stability.

2. $a = 0$: Here we have $\xi(t) = \xi(t_0)$ for all t. Therefore, we have stability, but not asymptotic stability of the equilibrium $x_0 = 0$. (Formally, let $\epsilon \in \mathbb{R}_{>0}$. Then, taking $\delta = \epsilon$, we have

$$|\xi(t_0) - 0| < \delta \implies |\xi(t) - 0| < \epsilon, \qquad t \geq t_0.)$$

3. $a > 0$: Here, as long as $\xi(t_0) \neq 0$, we have $\lim_{t\to\infty}|\xi(t)| = \infty$, and this suffices to show that the equilibrium $x_0 = 0$ is unstable. (Formally, we must show that there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for any $\delta \in \mathbb{R}_{>0}$ there exists $\xi(t_0) \in \mathbb{R}$ and $T \in \mathbb{R}_{>0}$ such that, $|\xi(t_0)| < \delta$ and $|\xi(t_0 + T)| \geq \epsilon$. We can take $\epsilon = 1$ and, given $\delta \in \mathbb{R}_{>0}$, we can take $\xi(t_0) = \frac{\delta}{2}$ and $T \in \mathbb{R}_{>0}$ such that $\mathrm{e}^{aT} \geq \frac{2}{\delta}$.) •

4.1.8 Example (Stable versus unstable versus asymptotically stable II) We consider another example illustrating the same trichotomy as the preceding example, but one that generates some pictures that one can keep in mind when thinking about concepts of stability. We consider the ordinary differential equation $\boldsymbol{F}$ with state space $U = \mathbb{R}^2$ and with right-hand side $\widehat{F}(t, (x_1, x_2)) = (x_2, -x_1 - 2\delta x_2)$ for $|\delta| < 1$. We shall be concerned with the stability of the equilibrium point $\boldsymbol{x}_0 = (0, 0)$. Solutions $\boldsymbol{\xi}\colon \mathbb{T} \to \mathbb{R}^2$ satisfy

$$\begin{aligned}\dot{\xi}_1(t) &= \xi_2(t),\\ \dot{\xi}_2(t) &= -\xi_1(t) - 2\delta\xi_2(t).\end{aligned}$$

This is a linear homogeneous ordinary differential equation with constant coefficients determined by the matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & -\delta \end{bmatrix}.$$

We compute the eigenvalues of $\boldsymbol{A}$ to be

$$\lambda_1 = -\delta + \mathrm{i}\sqrt{1 - \delta^2}, \quad \lambda_2 = -\delta - \mathrm{i}\sqrt{1 - \delta^2}.$$

Thus we have two distinct complex eigenvalues. We can then apply Procedures 3.2.45 and 3.2.48 to compute

$$\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{-\delta t}\begin{bmatrix} \cos(\sqrt{1-\delta^2}t) + \frac{\delta}{\sqrt{1-\delta^2}}\sin(\sqrt{1-\delta^2}t) & \frac{1}{\sqrt{1-\delta^2}}\sin(\sqrt{1-\delta^2}t) \\ -\frac{1}{\sqrt{1-\delta^2}}\sin(\sqrt{1-\delta^2}t) & \cos(\sqrt{1-\delta^2}t) + \frac{\delta}{\sqrt{1-\delta^2}}\sin(\sqrt{1-\delta^2}t) \end{bmatrix}.$$

In Figure 4.2 we plot the parameterised curves in (x_1, x_2)-space in what we shall in Section 3.4 call "phase portraits. Without going through the details of the analysis, we shall simply make the following observations.

1. $\delta > 0$: Here we see that $\boldsymbol{x}_0 = (0, 0)$ is asymptotically stable.
2. $\delta = 0$: Here we see that $\boldsymbol{x}_0 = (0, 0)$ is stable, but not asymptotically stable.

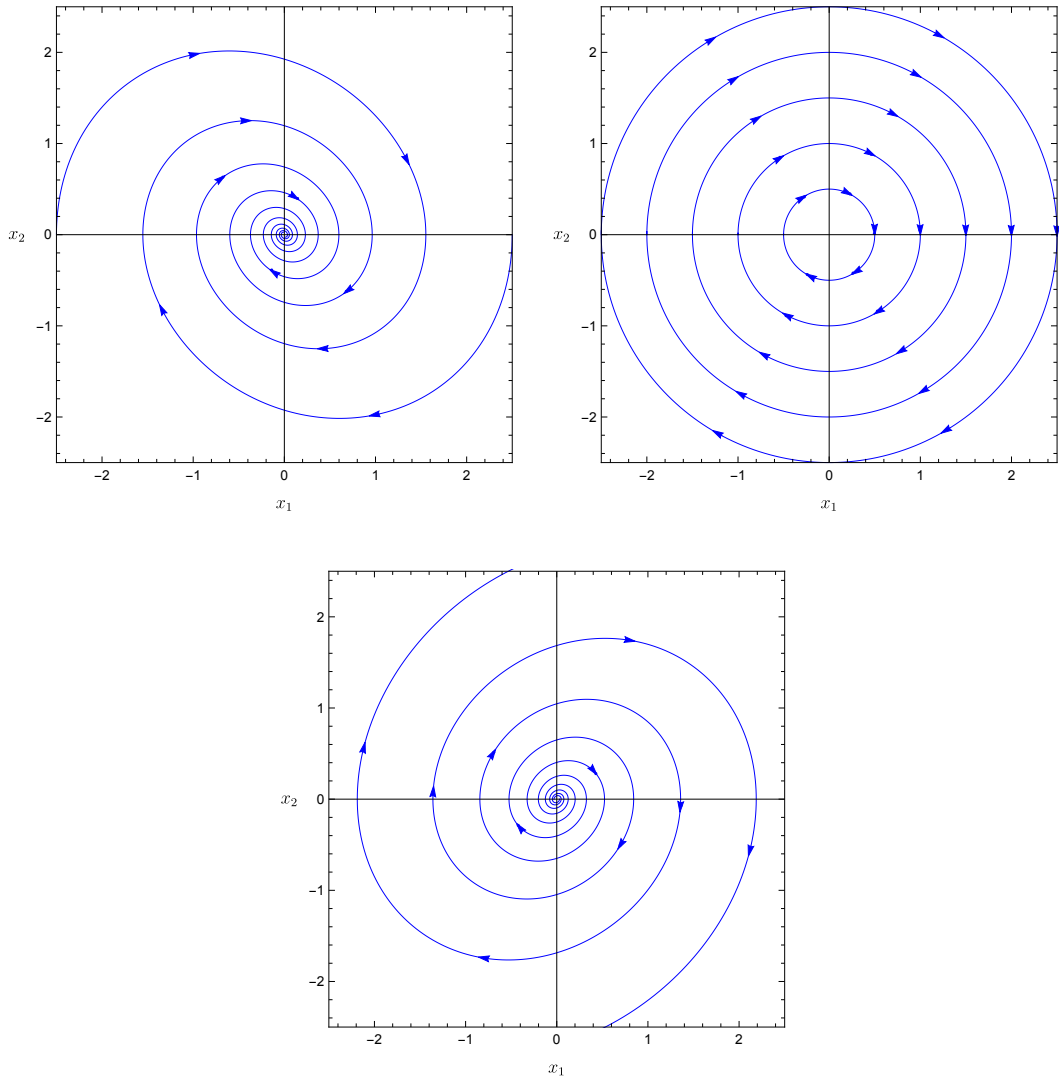


Figure 4.2 Phase portraits for $\widehat{F}(t,(x_1,x_2)) = (x_2, -x_1 - \delta x_2)$ for $\delta < 0$ (top left), $\delta = 0$ (top right), and $\delta > 0$ (bottom)

3. $\delta < 0$: Here we see that $\boldsymbol{x}_0 = (0,0)$ is unstable.

One can look at the behaviour of solutions in Figure 4.2 to convince oneself of the validity of these conclusions. •

The definitions we give in Definition 4.1.4 are "local.[2]" This means that they only give conclusions about the behaviour of solutions nearby the reference solution. Our preceding two examples might give one the impression that they hold globally, but this is not the case, as we illustrate in the next two examples.

[2]Indeed, the definitions we give are often prefixed by "local."

4.1.9 Example (Stable does not mean "globally stable" I) Here we consider the ordinary differential equation F with state space $U = \mathbb{R}$ and right-hand side $\widehat{F}(t, x) = x - x^3$. We will look at the stability of the equilibria for this differential equation. According to Proposition 3.1.5, a state $x_0 \in \mathbb{R}$ is an equilibrium state if and only if $x_0 - x_0^3 = 0$, which gives the three equilibria $x_- = -1$, $x_0 = 0$, and $x_+ = 0$. We shall subsequently see how to rigorously prove the stability of these three equilibria, but here we shall argue heuristically. In Figure 4.3 we graph

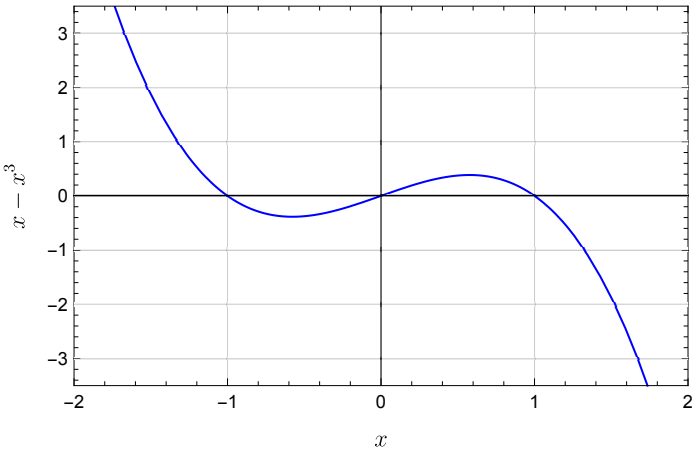


Figure 4.3 The right-hand side $x - x^3$

the right-hand side as a function of x. From this graph, we make the following conclusions.

1. x_0 is unstable: We see that, when $x > x_0 = 0$ and x is nearby $x_0 = 0$, that $\widehat{F}(t, x) > 0$, Therefore, if $\xi(t_0) > 0$ and is nearby 0, then $\xi(t)$ will become "more positive." In similar manner, if $\xi(t_0) < 0$ and is nearby 0, then $\xi(t)$ will become "more negative." Thus all solutions nearby 0 "move away" from 0.
2. $x_\pm$ are asymptotically stable: Here the opposite phenomenon occurs as compared to x_0. When $x > x_\pm$ and x is nearby $x_\pm$, then $\widehat{F}(t, x) < 0$. Therefore, if $\xi(t_0) > x_\pm$ and is nearby $x_\pm$, then $\xi(t)$ will "move towards" $x_\pm$. In similar manner, if $\xi(t_0) < x_\pm$ and is nearby $x_\pm$, then $\xi(t)$ will again "move towards" $x_\pm$.

The point is that our conclusions about stability for all three equilibria hold only for initial conditions nearby the equilibria. Moreover, the stability is different for different equilibria. •

4.1.10 Example (Stable does not mean "globally stable" II) The example here illustrates a similar phenomenon as the preceding example, but does so while producing some useful pictures. The ordinary differential equation we consider has state space $U = \mathbb{R}^2$ with right-hand side $\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = (x_2, x_1 - x_1^3 - \frac{1}{2}x_2)$. Thus solutions

$\xi\colon \mathbb{T} \to \mathbb{R}^2$ satisfy

$$\begin{aligned}\dot{\xi}_1(t) &= \xi_2(t)\\ \dot{\xi}_2(t) &= \xi_1(t) - \xi_1(t)^3 - \tfrac{1}{2}\xi_2(t).\end{aligned}$$

We will consider the stability of the equilibria for $\boldsymbol{F}$. By Proposition 3.1.5, an equilibrium $\boldsymbol{x}_0 = (x_{01}, x_{02})$ will satisfy

$$\begin{aligned}0 &= x_{02},\\ 0 &= x_{01} - x_{01}^3 - \frac{1}{2}x_{01},\end{aligned}$$

which gives the three equilibrium points $\boldsymbol{x}_0 = (0,0)$, $\boldsymbol{x}_- = (-1,0)$, and $\boldsymbol{x}_+ = (0,1)$. In Figure 4.4 we show a few parameterised solutions for $\boldsymbol{F}$ in the (x_1, x_2)-plane. From

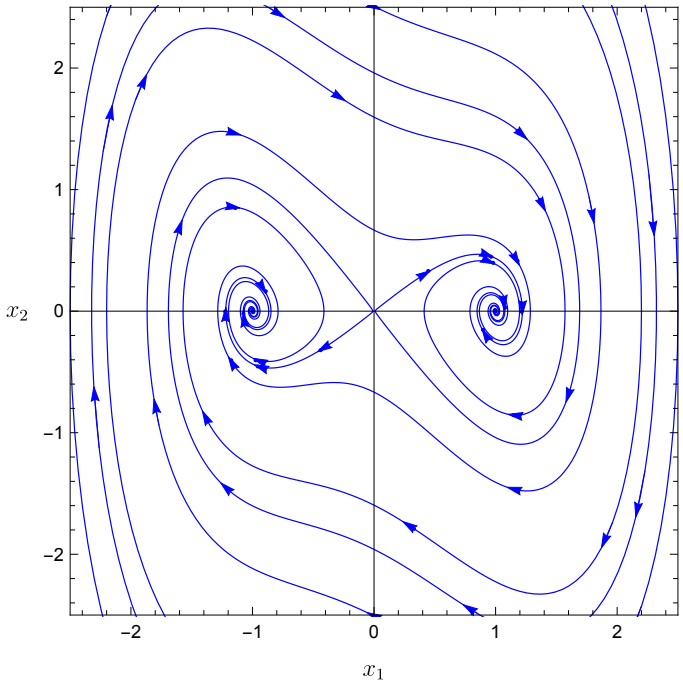


Figure 4.4 Phase portrait for $\widehat{F}(t,(x_1,x_2)) = (x_2, x_1 - x_1^3 - \frac{1}{2}x_2)$

this figure we deduce that $\boldsymbol{x}_0$ is unstable and $\boldsymbol{x}_\pm$ is asymptotically stable. •

The reader will have noticed that "stable" is included in the definition of "asymptotically stable." It seems like this might be redundant, but it is not as the next example indicates.

4.1.11 Example (Why "stable" is part of the definition of "asymptotically stable") We work with the ordinary differential equation $\boldsymbol{F}$ with state space $\mathbb{R}^2$ and with

$$\widehat{\boldsymbol{F}}(t,(x_1,x_2)) = \left(\frac{x_1^2(x_2-x_1)+x_2^5}{(x_1^2+x_2^2)(1+(x_1^2+x_2^2)^2)}, \frac{x_2^2(x_2-2x_1)}{(x_1^2+x_2^2)(1+(x_1^2+x_2^2)^2)}\right)$$

This solutions $\boldsymbol{\xi}\colon \mathbb{T} \to \mathbb{R}^2$ for $\boldsymbol{F}$ satisfy

$$\begin{aligned}\dot{\xi}_1(t) &= \frac{\xi_1(t)^2(\xi_2(t)-\xi_1(t))+\xi_2(t)^5}{(\xi_1(t)^2+\xi_2(t)^2)(1+(\xi_1(t)^2+\xi_2(t)^2)^2)}\\ \dot{\xi}_2(t) &= \frac{\xi_2(t)^2(\xi_2(t)-2\xi_1(t))}{(\xi_1(t)^2+\xi_2(t)^2)(1+(\xi_1(t)^2+\xi_2(t)^2)^2)}.\end{aligned}$$

We are interested in the stability of the equilibrium point $\boldsymbol{x}_0 = (0,0)$. In Figure 4.5 we depict the phase portrait for the equation. From the phase portrait, we can

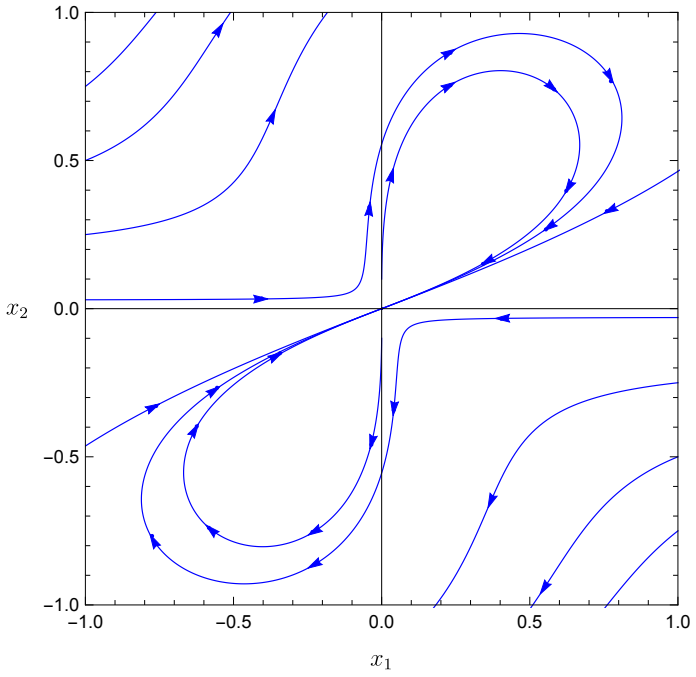


Figure 4.5 Phase portrait for $\widehat{F}(t,x) = \left(\frac{x_1^2(x_2-x_1)+x_2^5}{(x_1^2+x_2^2)(1+(x_1^2+x_2^2)^2)}, \frac{x_2^2(x_2-2x_1)}{(x_1^2+x_2^2)(1+(x_1^2+x_2^2)^2)}\right)$

reasonable say that (1) for any initial condition $\boldsymbol{\xi}(t_0) \in \mathbb{R}^2$, we have $\lim_{t\to\infty} \boldsymbol{\xi}(t) = (0,0)$ and (2) $\boldsymbol{x}_0 = (0,0)$ is not stable. The former can be seen straightaway from Figure 4.5. For the latter, we note that, for any $\epsilon \in \mathbb{R}_{>0}$, no matter how small we choose δ, there is an initial condition satisfying $\|\boldsymbol{\xi}(t_0) - \boldsymbol{x}_0\| < \delta$ for which the

corresponding solution leaves the ball of radius ϵ centred at x_0. Thus stability is required as part of the definition of asymptotic stability in order to rule out this "large deviation" behaviour.[3] •

4.1.12 Example (Asymptotically stable versus exponentially stable) In some of our examples above where an equilibrium is asymptotically stable, it is also exponentially stable. However, this need not be the case always. To illustrate this, we consider the ordinary differential equation with state space $U = \mathbb{R}$ and right-hand side $\widehat{F}(t, x) = -x^3$. In this case, we can argue as in Example 4.1.9 that the equilibrium state at $x_0 = 0$ is asymptotically stable. Let us show that it is not exponentially stable. For $\xi(t_0) \in \mathbb{R}$, we can use the technique of Section 2.1 to obtain the solution with this initial condition as

$$\xi(t) = \operatorname{sign}(\xi(t_0)) \left(\frac{1 + 2(t - t_0)\xi(t_0)^2}{\xi(t_0)^2} \right)^{-1/2},$$

where $\operatorname{sign}\colon \mathbb{R} \to \{-1, 0, 1\}$ returns the sign of a real number. The observation we make is that, as $t \to \infty$, $\xi(t)$ decays to zero like $(t - t_0)^{-1/2}$, which prohibits exponential stability. •

4.1.13 Example (Stable versus orbitally stable) As we saw in Proposition 4.1.5, one cannot distinguish between "stable" and "orbitally stable" for equilibria. Therefore, necessarily, if we wish to consider a distinction between these sorts of stability, we need to work with a nonequilibrium solution. The example we give is one that is easily imagined, and we do not rigorously prove our assertions.

We consider the motion of a simple pendulum. This can be thought of as a first-order system of ordinary differential equations with state space $U = \mathbb{R}^2$ and with right-hand side

$$\widehat{\boldsymbol{F}}(t, (x_1, x_2)) = \left(x_2, -\frac{a_g}{\ell} \sin(x_1) \right).$$

Here a_g is acceleration due to gravity and ℓ is the length of the pendulum. Solutions $\boldsymbol{\xi}\colon \mathbb{T} \to \mathbb{R}^2$ satisfy

$$\begin{aligned} \dot{\xi}_1(t) &= \xi_2(t) \\ \dot{\xi}_2(t) &= -\frac{a_g}{\ell} \sin(x_1). \end{aligned}$$

Let us make some (mathematically unproved, but physically "obvious") observations about this equation.

[3] One very often sees the following definition.

Definition A solution $\boldsymbol{\xi}_0$ of an ordinary differential equation is ***attractive*** if there exists $\delta \in \mathbb{R}_{>0}$ such that, for any ϵ, there exists $T \in \mathbb{R}_{>0}$ for which, if $\|\boldsymbol{\xi}(t_0) - \boldsymbol{\xi}_0(t_0)\| < \delta$, then $\|\boldsymbol{\xi}(t) - \boldsymbol{\xi}_0(t)\| < \epsilon$ for $t \geq t_0 + T$. •

One can then say that "asymptotic stability" means "stable" and "attractive."

1. For small oscillations of the pendulum, the period of the oscillation is $2\pi\sqrt{\frac{\ell}{a_g}}$.
2. As the amplitude of the oscillation becomes large (approaching π), the period becomes large. Indeed, for oscillations with amplitude exactly π, the period is "∞." Let us be clear what this means. There is a motion of the pendulum where, at "$t = -\infty$," the pendulum is upright at rest, and then begins to fall. It will fall and then swing up to the upright configuration at rest, getting there at "$t = \infty$."
3. For amplitudes between 0 and π, the period will grow monotonically from $2\pi\sqrt{\frac{\ell}{a_g}}$ to ∞. There is, in fact, a precise formula for this, and it is

$$T(\theta_0) = 4\sqrt{\frac{\ell}{a_g}}\int_0^{\pi/2}\frac{1}{(1-\sin^2(\frac{\theta_0}{2})\sin^2(\phi))^{1/2}}\,\mathrm{d}\phi,$$

where θ_0 is the amplitude of the oscillation. This integral, while not expressible in terms of anything you know about, *is* expressible in terms of what is known as an "elliptic function." The formula itself can be derived using conservation of energy. In Figure 4.6 we plot the period of oscillation versus the amplitude.

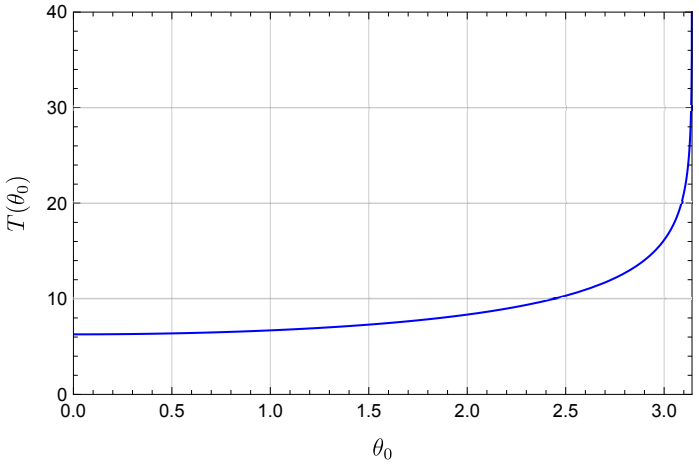


Figure 4.6 Normalised (by $\frac{\ell}{a_g}$) period of a pendulum oscillation as a function of the amplitude

Now let us make use of the preceding observations. We will consider the stability of some nontrivial periodic motion of the pendulum with amplitude between 0 and π. We claim that such a solution is orbitally stable, but not stable. In Figure 4.7 we show a periodic motion of the pendulum as a parameterised curve in the (x_1, x_2)-plane. In the figure we plot three solutions. The middle of the three solutions is the solution ξ_0 whose stability we are referencing. It has initial condition θ_0 and is defined on the time interval $[0, T(\theta_0)]$. The other two solutions have

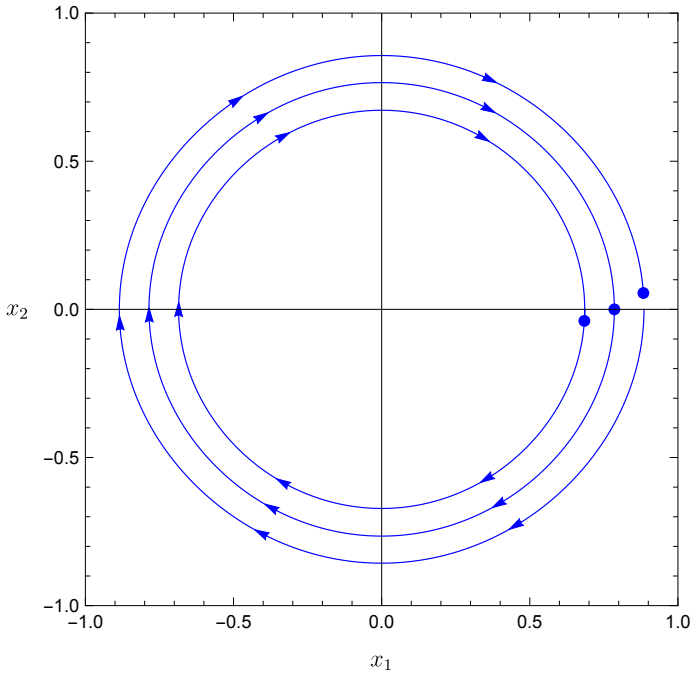


Figure 4.7 Orbital stability, but not stability, of the nontrivial periodic motions of a pendulum; the middle curve is the nominal solution whose stability is being determined

nearby initial conditions, and are defined on the same time interval, and a dot is placed at the final point of the solution. We make the following observations.

1. ξ_0 *is not stable:* The reasoning here is this. In Figure 4.7 we see that the periodic solutions nearby ξ_0 do not undergo exactly one period in the time it takes ξ_0 to undergo exactly one period; the inner solution travels more than one period and the outer solution travels less than one period. Now imagine letting the trajectory ξ_0 undergo an increasing number of periods. The inner and outer solutions will drift further and further from ξ_0 when compared at the same times. This prohibits stability of ξ_0 since nearby initial conditions will produce solutions that are eventually not close.
2. ξ_0 *is orbitally stable:* The reasoning here is this. While solutions with nearby initial conditions will drift apart in time, the solutions themselves remain close in the sense that any point on one solution is nearby some point (not at the same time) on the other solution. More viscerally, the images of solutions for nearby initial solutions are close. •

4.1.14 Example (Stable versus uniformly stable I) Here we take the linear homoge-

neous ordinary differential equation F in $\mathsf{V} = \mathbb{R}$ defined by the right-hand side

$$\widehat{F}(t, x) = -\frac{x}{1+t}$$

for $t \in \mathbb{T} = [0, \infty)$. We will consider the stability of the equilibrium point $x_0 = 0$. We can explicitly solve this ordinary differential equation (for example, using the method of Section 2.1) to give

$$\xi(t) = \frac{\xi(t_0)(1+t_0)}{1+t}.$$

From this we can make the following observations.

1. $x_0 = 0$ *is asymptotically stable:* This follows since, for any initial condition $\xi(t_0)$, we have $\lim_{t\to\infty} \xi(t) = 0$.
2. $x_0 = 0$ *is uniformly stable:* Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \epsilon$. If $|\xi(t_0) - 0| < \delta$, then

   $$|\xi(t)| \leq |\xi(t_0)| < \epsilon$$

   for $t \geq t_0$. This gives the desired uniform stability.
3. $x_0 = 0$ *is not uniformly asymptotically stable:* We must show that, for every $\delta \in \mathbb{R}_{>0}$ and $T \in \mathbb{R}_{>0}$, there exists $\epsilon \in \mathbb{R}_{>0}$, $t_0 \in \mathbb{T}$, and $x \in \mathbb{R}$ satisfying $|x - 0| < \delta$, such that the solution $\xi \colon \mathbb{T} \to \mathbb{R}$ to the initial value problem

   $$\dot{\xi}(t) = -\frac{\xi(t)}{1+t}, \quad \xi(t_0) = x,$$

   satisfies $|\xi(t_0 + T)| \geq \epsilon$. We take $x = \frac{\delta}{2}$, $\epsilon = 1$, $T \in \mathbb{R}_{>0}$, and $t_0 \in \mathbb{T}$ such that

   $$\frac{1+t_0}{1+t_0+T} \geq \frac{2}{\delta};$$

   this is possible since $\lim_{t_0\to\infty} \frac{1+t_0}{1+t_0+T} = 1$ for any $T \in \mathbb{R}_{>0}$. Now let $x \in \mathbb{R}$ satisfy $|x - 0| < \delta$, and let $\xi \colon \mathbb{T} \to \mathbb{R}$ be the solution to the initial value problem

   $$\dot{\xi}(t) = -\frac{\xi(t)}{1+t}, \quad \xi(t_0) = x.$$

   Then

   $$|\xi(t_0 + T)| = \left| \frac{x(1+t_0)}{1+t_0+T} \right| = \frac{\delta}{2} \left| \frac{1+t_0}{1+t_0+T} \right| \geq 1,$$

   which gives the desired lack uniform asymptotic convergence. •

4.1.15 Example (Stable versus uniformly stable II) We again consider a linear homogeneous ordinary differential equation in $\mathbb{R}$, this one with right-hand side

$$\widehat{F}(t, x) = \sin(\ln(t)) + \cos(\ln(t)) - \alpha$$

for some $\alpha \in (1, \sqrt{2})$. Here we consider $\mathbb{T} = (0, \infty)$. Again we consider stability of the equilibrium point at $x_0 = 0$. In this case, an application of the method of Section 2.1 gives the solution

$$\xi(t) = \mathrm{e}^{-\alpha(t-t_0)+t\sin(\ln(t))-t_0\sin(\ln(t_0))}\xi(t_0).$$

We make the following observations.

1. $x_0 = 0$ *is asymptotically stable:* Here we note that, since

$$\lim_{t\to\infty}(-\alpha(t - t_0) + t\sin(\ln(t)) - t_0\sin(\ln(t_0))) = -\infty$$

since $\alpha > 1$, we must have $\lim_{t\to\infty}\xi(t) = 0$ for any initial condition $\xi(t_0)$. This gives asymptotic stability. In fact, we can refine this conclusion a little.

2. $x_0 = 0$ *is not uniformly stable:* This is more difficult to prove. We choose $\beta \in (\alpha, \sqrt{2})$ and $\theta_1 \in (0, \frac{\pi}{4})$ and $\theta_2 \in (\frac{\pi}{4}, \frac{\pi}{2})$ such that

$$\sin\theta + \cos\theta \geq \beta, \qquad \theta \in [\theta_1, \theta_2].^{4}$$

Then, for $j \in \mathbb{Z}_{>0}$, define

$$t_j = \mathrm{e}^{2j\pi+\theta_2}, \quad t_{0,j} = \mathrm{e}^{2j\pi+\theta_1},$$

and compute, for $j \in \mathbb{Z}_{>0}$,

$$\begin{aligned}\int_{t_{0,j}}^{t_j}(\sin(\ln(t)) + \cos(\ln(t)) - \alpha)\,\mathrm{d}t &= \int_{2j\pi+\theta_1}^{2j\pi+\theta_2}(\sin\theta + \cos\theta - \alpha)\mathrm{e}^{2j\pi+\theta}\,\mathrm{d}\theta\\ &= \int_{\theta_1}^{\theta_2}(\sin\theta + \cos\theta - \alpha)\mathrm{e}^{2j\pi+\theta}\,\mathrm{d}\theta\\ &\geq (\beta - \alpha)\mathrm{e}^{2j\pi}\int_{\theta_1}^{\theta_2}\mathrm{e}^{\theta}\,\mathrm{d}\theta\\ &= (\beta - \alpha)\mathrm{e}^{2j\pi}(\mathrm{e}^{\theta_2} - \mathrm{e}^{\theta_1}),\end{aligned}$$

[4] To see why this is possible, first note that

$$\sqrt{2}\cos(\theta - \tfrac{\pi}{4}) = \sin\theta\sin\tfrac{\pi}{4} + \cos\theta\cos\tfrac{\pi}{4} = \sin\theta + \cos\theta,$$

using standard trigonometric identities. Then note that the function

$$\theta \mapsto \sqrt{2}\cos(\theta - \tfrac{\pi}{4})$$

has a local maximum at $\theta = \frac{\pi}{4}$ with value $\sqrt{2}$. Thus, since $\alpha < \beta < \sqrt{2}$, we can choose $\theta_1 < \frac{\pi}{4}$ and $\theta_2 > \frac{\pi}{4}$ sufficiently close to $\frac{\pi}{4}$ to ensure that $\sin\theta + \cos\theta \geq \beta$ for $\theta \in [\theta_1, \theta_2]$.

where we have used the change of variable $t = \mathrm{e}^{2j\pi+\theta}$ in the second line. Note, then, that

$$\lim_{j\to\infty} \int_{t_{0,j}}^{t_j} (\sin(\ln(t)) + \cos(\ln(t)) - \alpha)\,\mathrm{d}t = \infty.$$

Now, using this fact, we claim that $x_0 = 0$ is not uniformly stable. We must show that there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for every $\delta \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$, $t_0 \in \mathbb{T}$, and $x \in \mathbb{R}$ satisfying $|x - 0| < \delta$ and for which the solution to the initial value problem

$$\dot{\xi}(t) = (\sin(\ln(t)) + \cos(\ln(t)) - \alpha)\xi(t), \quad \xi(t_0) = x,$$

satisfies $|\xi(t_0 + T) - 0| \geq \epsilon$. We take $\epsilon = 1$. Let $\delta \in \mathbb{R}_{>0}$ and $x = \frac{\delta}{2}$. Let $j \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\int_{t_{0,j}}^{t_j} (\sin(\ln(t)) + \cos(\ln(t)) - \alpha)\,\mathrm{d}t \geq \frac{2}{\delta}.$$

Then take $t_0 = t_{0,j}$ and $T = t_j$. We then have

$$|\xi(t_0 + T)| = \left| \int_{t_{0,j}}^{t_j} (\sin(\ln(t)) + \cos(\ln(t)) - \alpha)\,\mathrm{d}t \right| |x| \geq 1,$$

giving the desired absence of uniform stability. •

While the preceding examples do not cover all of the possible gaps in the stability definitions of Definition 4.1.4, they do hopefully sufficiently illustrate the essence of the difference in the various definitions that a reader can have a picture in their mind of these differences as we proceed to study stability in more detail in the sequel.

Exercises

4.1.1 Let us consider the system of ordinary differential equations $\boldsymbol{F}$ with state-space $\mathbb{R}^2$ defined by the right-hand side

$$\begin{aligned} \widehat{F}\colon \mathbb{R} \times \mathbb{R}^2 &\to \mathbb{R}^2 \\ (t, (x_1, x_2)) &\mapsto (1, 0). \end{aligned}$$

Answer the following questions.

(a) Show that

$$\begin{aligned} \boldsymbol{\xi}_0\colon \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto (t, 0) \end{aligned}$$

is the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(0) = \boldsymbol{0}.$$

(b) Show that the solution ξ_0 is stable but not asymptotically stable.

Now consider a change of coordinates from $(x_1, x_2) \in \mathbb{R}^2$ to $(y_1, y_2) \in \mathbb{R}^2$ defined by

$$y_1 = x_1, \quad y_2 = \mathrm{e}^{x_1} x_2,$$

and let $\boldsymbol{G}$ be the ordinary differential equation $\boldsymbol{F}$, represented in these coordinates.

(c) Use the Chain Rule to compute $\dot{y}_1$ and $\dot{y}_2$,

$$\dot{y}_1(t) = \frac{\partial y_1}{\partial x_1}\dot{x}_1(t) + \frac{\partial y_1}{\partial x_2}\dot{x}_2(t),$$
$$\dot{y}_2(t) = \frac{\partial y_2}{\partial x_1}\dot{x}_1(t) + \frac{\partial y_2}{\partial x_2}\dot{x}_2(t),$$

and so give the right-hand side $\widehat{G}$ for $\boldsymbol{G}$.

Hint: *Write everything in terms of the coordinates* $(\mathrm{y}_1, \mathrm{y}_2)$.

(d) Show that the solution ξ_0 is mapped, under the change of coordinates, to the solution $\eta_0 \colon \mathbb{R} \to \mathbb{R}^2$ given by $\eta_0(t) = (t, 0)$.

(e) Show that η_0 is not stable.

Now consider a change of coordinates from $(x_1, x_2) \in \mathbb{R}^2$ to $(z_1, z_2) \in \mathbb{R}^2$ defined by

$$z_1 = x_1, \quad z_2 = \mathrm{e}^{-x_1} x_2,$$

and let $\boldsymbol{H}$ be the ordinary differential equation $\boldsymbol{F}$, represented in these coordinates.

(f) Use the Chain Rule to compute $\dot{z}_1$ and $\dot{z}_2$,

$$\dot{z}_1(t) = \frac{\partial z_1}{\partial x_1}\dot{x}_1(t) + \frac{\partial z_1}{\partial x_2}\dot{x}_2(t),$$
$$\dot{z}_2(t) = \frac{\partial z_2}{\partial x_1}\dot{x}_1(t) + \frac{\partial z_2}{\partial x_2}\dot{x}_2(t),$$

and so give the right-hand side $\widehat{H}$ for $\boldsymbol{H}$.

Hint: *Write everything in terms of the coordinates* $(\mathrm{z}_1, \mathrm{z}_2)$.

(g) Show that the solution ξ_0 is mapped, under the change of coordinates, to the solution $\zeta_0 \colon \mathbb{R} \to \mathbb{R}^2$ given by $\zeta_0(t) = (t, 0)$.

(h) Show that ζ_0 is asymptotically stable.

Section 4.2

Stability of linear ordinary differential equations

In this section we devote ourselves specially to the theory of stability for linear systems. We shall see that, for linear systems, there are a few natural places where one can refine the general definitions of stability from Definition 4.1.4, taking advantage of the linearity of the dynamics. Moreover, there are also equivalent characterisations of stability that hold for linear equations that do not hold in general.

As we did in Chapter 3 when dealing with linear systems, we shall work with linear systems whose state space is a finite-dimensional vector space V. Our stability definitions from Definition 4.1.4 all involve the measure of distance provided by the Euclidean norm on $\mathbb{R}^n$. An abstract vector space does not have a natural norm, but one can always be provided by, for example, choosing a basis $\mathscr{B} = \{e_1, \dots, e_n\}$ and then defining $\|v\|_{\mathscr{B}} = \|(v_1, \dots, v_n)\|$, where $v = v_1 e_1 + \cdots + v_n e_n$. The fact of the matter is that nothing we do depends in any way on the choice of this norm,[5] and so we shall simply use the symbol "$\|\cdot\|$" to represent some choice of norm, possibly arising from the Euclidean norm by a choice of basis as described above. For readers following the "all vector spaces are $\mathbb{R}^n$" path, this is not anything of concern so you can resume sleeping.

4.2.1 Special stability definitions for linear equations

We begin with some definitions for stability that are suitable for linear equations.

4.2.1 Definition (Stability for linear systems) Let F be a system of linear homogeneous ordinary differential equations in an n-dimensional $\mathbb{R}$-vector space V and with right-hand side $\widehat{F}(t, x) = A(t)(x)$ for $A \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$. Suppose that $\sup \mathbb{T} = \infty$. Let $\zeta \colon \mathbb{T} \to \mathsf{V}$ be the zero solution $\zeta(t) = 0$, $t \in \mathbb{T}$.

(i) The equation F is S (resp. AS, ES, US, UAS, UES) if the zero solution ζ is S (resp. AS, ES, US, UAS, UES).

The equation F is:

(ii) ***globally stable*** if, for each $t_0 \in \mathbb{T}$, there exists $C \in \mathbb{R}_{>0}$ such that, for $x \in \mathsf{V}$, the

[5] The "big fact" here is that if we have two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ for a vector space V, then there exists $C \in \mathbb{R}_{>0}$ such that

$$C\|v\|_2 \leq \|v\|_1 \leq C^{-1}\|v\|_2, \qquad v \in \mathsf{V}.$$

Thus, if a reader goes through our definitions where a norm is used, she will see that using a different norm will only have the effect of change constants in the definition, while not materially altering the meaning of the definition.

solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq C\|x\|$ for $t \geq t_0$;

(iii) ***globally asymptotically stable*** if, for each $t_0 \in \mathbb{T}$ and each $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, for $x \in \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq \epsilon\|x\|$ for $t \geq t_0 + T$;

(iv) ***globally exponentially stable*** if, for each $t_0 \in \mathbb{T}$, there exists $M, c \in \mathbb{R}_{>0}$ such that, for $x \in \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq M\|x\|\mathrm{e}^{-c(t-t_0)}$ for $t \geq t_0$;

(v) ***globally uniformly stable*** if there exists $C \in \mathbb{R}_{>0}$ such that, for $(t_0, x) \in \mathbb{T} \times \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq C\|x\|$ for $t \geq t_0$;

(vi) ***globally uniformly asymptotically stable*** if it is globally uniformly stable and if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, for $(t_0, x) \in \mathbb{T} \times \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq \epsilon\|x\|$ for $t \geq t_0 + T$;

(vii) ***globally uniformly exponentially stable*** if there exists $M, c \in \mathbb{R}_{>0}$ such that, for $(t_0, x) \in \mathbb{T} \times \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq M\|x\|\mathrm{e}^{-c(t-t_0)}$ for $t \geq t_0$. •

Part (i) of the definition is merely the statement of the convention that, when talking about stability for linear ordinary differential equations, one is interested in the stability of the equilibrium state at 0. For this reason, given Proposition 4.1.5, we do not discuss orbital stability for linear equations. The remaining six definitions above are quite particular to linear equations.

We can add obviously to our list of abbreviations.

(U)GS	(uniformly) globally stable
(U)GAS	(uniformly) globally asymptotically stable
(U)GES	(uniformly) globally exponentially stable

There is a little subtlety to the preceding definitions that merits exploration, and this is that (1) the definition of GAS does not include GS as part of the definition, (2) the definition of UGES does not include UGS, whereas (3) for UGAS and UGES, we *do* include the requirement that the equation also be UGS. As we shall see in the proof of Theorem 4.2.3 below, it is the case that GAS $\implies$ GS. It is obvious from the definition that UGES $\implies$ UGS. However, it is *not* true that UGS can be omitted in the definitions of UGAS and UGES, as the following example shows.

4.2.2 Example (UGS must be a part of the definition of UGAS and UGES) We shall construct a system of linear homogeneous ordinary differential equations F in $\mathsf{V} = \mathbb{R}$ with right hand-side $\widehat{F}(t, x) = a(t)x$ and with the following properties:

1. F is not UGS;
2. for $\epsilon \in \mathbb{R}_{>0}$ there exists $T \in \mathbb{R}_{>0}$ with the property that, for $(t_0, x) \in \mathbb{T} \times \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = a(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $|\xi(t)| < \epsilon|x|$ for $t \geq t_0 + T$.

The example is a little convoluted.

We take $\mathbb{T} = \mathbb{R}_{\geq 0}$ and define $a \colon \mathbb{T} \to \mathbb{R}$ in the following way.

1. Define sequences $(a_k)_{k\in\mathbb{Z}_{\geq 0}}$, $(b_k)_{k\in\mathbb{Z}_{\geq 0}}$, and $(\Delta_k)_{k\in\mathbb{Z}_{\geq 0}}$ as follows:
   (a) $\Delta_k = 2^{-k-1}$, $k \in \mathbb{Z}_{\geq 0}$;
   (b) $b_k = k2^{k+1}$, $k \in \mathbb{Z}_{\geq 0}$;
   (c) define $a_1 = 1$ and then define a_k, $k \geq 2$, by

$$b_{k-1}\Delta_{k-1} - a_k(1 - \Delta_k) + b_k\Delta_k + b_{k+1}\Delta_{k+1} = -1.$$

2. If $t \in \mathbb{T}$, let $k \in \mathbb{Z}_{\geq 0}$ be such that $t \in [k, k + 1)$, and then define

$$a(t) = \begin{cases} -a_k, & t \in [k, k + \Delta_{k+1}), \\ b_k, & t \in [k + \Delta_{k+1}, k + 1). \end{cases}$$

Note that a is not continuous, however, it can be modified to be continuous and still have the desired properties.

To show that F, defined by a, has the desired properties, we first show that F has the property 1 above. For $k \in \mathbb{Z}_{\geq 0}$ define $t_k = k + 1$ and $t_{0,k} = k + \Delta_k$. Let $x = 1 \in \mathsf{V}$ and let $\xi_k \colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}_k(t) = a(t)\xi_k(t), \quad \xi_k(t_{0,k}) = x,$$

for $k \in \mathbb{Z}_{\geq 0}$. Note that

$$|\xi_k(t_k)| = \left| x\mathrm{e}^{-\int_{t_{0,k}}^{t_k} a(\tau)\,\mathrm{d}\tau} \right| = |x|\mathrm{e}^k.$$

This prohibits uniform global stability for F.

Next we show that F has the property 2 above. Thus let $\epsilon \in \mathbb{R}_{>0}$ and define $T \in \mathbb{Z}_{>0}$ such that $\mathrm{e}^{-(T-3)} < \epsilon$. Let $t_0 \in \mathbb{T}$ and let $t \geq t_0 + T$. Let $k_1 \in \mathbb{Z}_{\geq 0}$ be such that $t_0 \in [k_1, k_1 + 1)$, let $k_2 \in \mathbb{Z}_{>0}$ be such that $t \in [k_2, k_2 + 1)$. Note that

$$t - t_0 \geq T \implies k_2 - k_1 + 1 > T \implies k_2 - k_1 - 2 > T - 3.$$

Now we estimate

$$\begin{aligned}
\int_{t_0}^{t_0+t} a(\tau)\,\mathrm{d}\tau &= \int_{t_0}^{k_1+1} a(\tau)\,\mathrm{d}\tau + \sum_{k=k_1+1}^{k_2-1} \int_{k}^{k+1} a(\tau)\,\mathrm{d}\tau + \int_{k_2}^{t} a(\tau)\,\mathrm{d}\tau \\
&\leq b_{k_1}\Delta_{k_1} + \sum_{k=k_1+1}^{k_2-1} (-a_k(1-\Delta_k) + b_k\Delta_k) + b_{k_2}\Delta_{k_2} \\
&\leq \sum_{k_1+1}^{k_2-1} (b_{k-1}\Delta_{k-1} - a_k(1-\Delta_k) + b_k\Delta_k + b_{k+1}\Delta_{k+1}) \\
&= -\sum_{k=k_1+1}^{k_2-1} 1 = -(k_2 - k_1 - 2) < -(T-3).
\end{aligned}$$

Now let $x \in \mathsf{V}$ and let $\xi \colon \mathbb{T} \to \mathsf{V}$ satisfy the initial value problem

$$\dot{\xi}(t) = a(t)\xi(t), \quad \xi(t_0) = x.$$

Then

$$|\xi(t)| = \left| x\mathrm{e}^{-\int_{t_0}^{t} a(\tau)\,\mathrm{d}\tau} \right| \leq |x|\mathrm{e}^{(T-3)} < \epsilon|x|,$$

for $t \geq t_0 + T$, giving the desired conclusion. •

Let us further explore these definitions by (1) exploring their relationships with the notions of stability from Definition 4.1.4 and (2) exploring the relationships between these new notions.

First the first...

4.2.3 Theorem (Equivalence of stability and global stability for linear ordinary differential equations) *Consider the system of linear homogeneous ordinary differential equations* F *with right-hand side (4.5) and suppose that* $\mathrm{A} \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *is continuous. Suppose that* $\sup \mathbb{T} = \infty$. *Then* F *is* S *(resp.* AS, ES, US, UAS, UES*) if and only if it is* GS *(resp.* GAS, GES, GUS, GUAS, GUES*).*

Proof (GS $\implies$ S) Let $t_0 \in \mathbb{T}$ and let $C \in \mathbb{R}_{>0}$ be such that the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq C\|x\|$ for $t \geq t_0$. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{C}$. Now let $x \in \mathsf{V}$ satisfy $\|x\| < \delta$ and let $\xi \colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

We then have

$$\|\xi(t)\| \leq C\|x\| = \frac{\epsilon}{\delta}\|x\| \leq \epsilon,$$

for $t \geq t_0$, giving stability of F.

($\mathsf{S} \implies \mathsf{GS}$) Let $t_0 \in \mathbb{T}$ and let $\delta \in \mathbb{R}_{>0}$ have the property that, if $\|x\| \leq \delta$, then the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq 1$ for $t \geq t_0$. Define $C = \delta^{-1}$. Now let $x \in \mathsf{V}$ and let $\xi \colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x. \tag{4.1}$$

First suppose that $x \neq 0$ and define $\hat{x} = \delta \frac{x}{\|x\|}$ so that $\|\hat{x}\| = \delta$. Thus the solution $\hat{\xi} \colon \mathbb{T} \to \mathsf{V}$ to the initial value problem

$$\dot{\hat{\xi}}(t) = A(t)(\hat{\xi}(t)), \quad \hat{\xi}(t_0) = \hat{x},$$

satisfies $\|\hat{\xi}(t)\| \leq 1$ for $t \geq t_0$. However,

$$\xi(t) = \Phi_A(t, t_0)(x) = \Phi_A(t, t_0)\left(\frac{\|x\|}{\delta}\hat{x}\right) = \frac{\|x\|}{\delta}\Phi_A(t, t_0)(\hat{x}) = C\|x\|\hat{\xi}(t).$$

Therefore,

$$\|\xi(t)\| = C\|x\|\,\|\hat{\xi}(t)\| \leq C\|x\|.$$

If $x = 0$ this relation clearly holds since the solution to the initial value problem (4.1) is simply $\xi(t) = 0$, $t \in \mathbb{T}$. Thus F is globally stable.

($\mathsf{GAS} \implies \mathsf{AS}$) First we show that $\mathsf{GAS} \implies \mathsf{GS}$ (which implies S as we have already proved). Let $t_0 \in \mathbb{T}$, let $x \in \mathsf{V}$, and let $\xi \colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

First suppose that $x \neq 0$. Since

$$\lim_{t \to \infty} \frac{\|\xi(t)\|}{\|x\|} = 0$$

and since ξ is continuous (indeed, of class C^1), it follows that $t \mapsto \frac{\|\xi(t)\|}{\|x\|}$ is bounded, i.e., there exists $C \in \mathbb{R}_{>0}$ such that

$$\frac{\|\xi(t)\|}{\|x\|} \leq C \quad \implies \quad \|\xi(t)\| \leq C\|x\|.$$

This relationship also holds when $x = 0$, we conclude global stability of F.

Now let $t_0 \in \mathbb{T}$ and take $\delta = \frac{1}{2}$. Let $\epsilon \in \mathbb{R}_{>0}$, and take $T \in \mathbb{R}_{>0}$ such that the solution $\xi\colon \mathbb{T} \to \mathsf{V}$ to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq \epsilon\|x\|$ for $t \geq t_0 + T$. Now suppose that $\|x\| < \delta = \frac{1}{2}$, and let $\xi\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

Then, for $t \geq t_0 + T$,

$$\|\xi(t)\| \leq \epsilon\|x\| < \epsilon.$$

This shows that F is asymptotically stable.

(AS $\implies$ GAS) Let $t_0 \in \mathbb{T}$ and let $\delta \in \mathbb{R}_{>0}$ have the property that, given $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, if $\|x\| < \delta$, then the solution $\xi\colon \mathbb{T} \to \mathsf{V}$ to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| < \epsilon$ for $t \geq t_0 + T$.

Let $\epsilon \in \mathbb{R}_{>0}$ and let $T \in \mathbb{R}_{>0}$ be such that, if $\|x\| < \delta$, then the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| < \frac{\epsilon\delta}{2}$ for $t \geq t_0 + T$. Let $x \in \mathsf{V}$ and let $\xi\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

Let $\hat{x} = \delta\frac{x}{2\|x\|}$ and let $\hat{\xi}\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\hat{\xi}}(t) = A(t)(\hat{\xi}(t)), \quad \hat{\xi}(t_0) = \hat{x}.$$

Since $\|\hat{x}\| = \frac{\delta}{2} < \delta$, $\|\hat{\xi}(t)\| < \frac{\epsilon\delta}{2}$ for $t \geq t_0 + T$. We also have

$$\xi(t) = \Phi_A(t, t_0)(x) = \Phi_A(t, t_0)\left(\frac{2\|x\|}{\delta}\hat{x}\right) = \frac{2\|x\|}{\delta}\Phi_A(t, t_0)(\hat{x}) = \frac{2\|x\|}{\delta}\|x\|\hat{\xi}(t).$$

Thus

$$\|\xi(t)\| \leq \frac{2}{\delta}\|x\|\,\|\hat{\xi}(t)\| < \epsilon\|x\|,$$

for $t \geq t_0 + T$, and so F is globally asymptotically stable.

(GES $\implies$ ES) First we note that GES $\implies$ GS (which implies S, as we have already seen). Indeed, the proof that GAS $\implies$ GS we gave above also applies if we replace "GAS" with "GES."

Now let $t_0 \in \mathbb{T}$ and let $\tilde{M}, \tilde{\sigma} \in \mathbb{R}_{>0}$ be such that, for $v \in \mathsf{V}$, the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq \tilde{M}\|x\|\mathrm{e}^{-\tilde{\sigma}(t-t_0)}$ for $t \geq t_0$. Now let $\delta = \frac{1}{2}$ and take $M = \tilde{M}$ and $\sigma = \tilde{\sigma}$. Then, for $\|x\| < \delta = \frac{1}{2}$, let $\xi\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

We then have

$$\|\xi(t)\| \leq \tilde{M}\|x\|\mathrm{e}^{\tilde{\sigma}(t-t_0)} \leq M\mathrm{e}^{-\sigma(t-t_0)},$$

showing that F is exponentially stable.

(ES $\implies$ GES) Let $t_0 \in \mathbb{T}$ and let $\tilde{M}, \delta, \tilde{\sigma} \in \mathbb{R}_{>0}$ be such that, if $\|x\| < \delta$, then the solution $\xi\colon \mathbb{T} \to \mathsf{V}$ to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq \tilde{M}\mathrm{e}^{-\tilde{\sigma}(t-t_0)}$ for $t \geq t_0$.

Take $M = \frac{2\tilde{M}}{\delta}$ and $\sigma = \tilde{\sigma}$. Now let $x \in \mathsf{V}$ and let $\xi\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

Let $\hat{x} = \delta\frac{x}{2\|x\|}$ and let $\hat{\xi}\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\hat{\xi}}(t) = A(t)(\hat{\xi}(t)), \quad \hat{\xi}(t_0) = \hat{x}.$$

Since $\|\hat{x}\| = \frac{\delta}{2} < \delta$, $\|\hat{\xi}(t)\| \leq \tilde{M}\mathrm{e}^{-\tilde{\sigma}(t-t_0)}$ for $t \geq t_0$. Then, as in the proof that AS $\implies$ GAS,

$$\xi(t) = \frac{2\|x\|}{\delta}\hat{\xi}(t),$$

and so

$$\|\xi(t)\| = \frac{2}{\delta}\|x\|\,\|\hat{\xi}(t)\| \leq \frac{2\tilde{M}}{\delta}\|x\|\mathrm{e}^{-\tilde{\sigma}(t-t_0)} = M\|x\|\mathrm{e}^{-\sigma(t-t_0)},$$

for $t \geq t_0$, showing that F is globally exponentially stable.

The remainder of the proof concerns the results we have already proved, but with the property "uniform" being applied to all hypotheses and conclusions. The proofs are entirely similar to those above. We shall, therefore, only work this out in one of the three cases, the other two following in an entirely similar manner.

(GUS $\implies$ US) Let $C \in \mathbb{R}_{>0}$ be such that the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq C\|x\|$ for $t \geq t_0$. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{C}$. Now let $x \in \mathsf{V}$ satisfy $\|x\| < \delta$ and let $\xi\colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x.$$

We then have

$$\|\xi(t)\| \leq C\|x\| = \frac{\epsilon}{\delta}\|x\| \leq \epsilon,$$

for $t \geq t_0$, giving stability of F.

(US $\implies$ GUS) Let $\delta \in \mathbb{R}_{>0}$ have the property that, if $\|x\| \leq \delta$, then the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t)\| \leq 1$ for $t \geq t_0$. Define $C = \delta^{-1}$. Now let $x \in \mathsf{V}$ and let $\xi \colon \mathbb{T} \to \mathsf{V}$ be the solution to the initial value problem

$$\dot{\xi}(t) = A(t)(\xi(t)), \quad \xi(t_0) = x. \tag{4.2}$$

First suppose that $x \neq 0$ and define $\hat{x} = \delta \frac{x}{\|x\|}$ so that $\|\hat{x}\| = \delta$. Thus the solution $\hat{\xi} \colon \mathbb{T} \to \mathsf{V}$ to the initial value problem

$$\dot{\hat{\xi}}(t) = A(t)(\hat{\xi}(t)), \quad \hat{\xi}(t_0) = \hat{x},$$

satisfies $\|\hat{\xi}(t)\| \leq 1$ for $t \geq t_0$. However,

$$\xi(t) = \Phi_A(t, t_0)(x) = \Phi_A(t, t_0)\left(\frac{\|x\|}{\delta}\hat{x}\right) = \frac{\|x\|}{\delta}\Phi_A(t, t_0)(\hat{x}) = C\|x\|\hat{\xi}(t).$$

Therefore,

$$\|\xi(t)\| = C\|x\|\,\|\hat{\xi}(t)\| \leq C\|x\|.$$

If $x = 0$ this relation clearly holds since the solution to the initial value problem (4.2) is simply $\xi(t) = 0$, $t \in \mathbb{T}$. Thus F is globally stable. ■

Now let us examine some relationships between these special notions of stability for linear equations.

4.2.4 Theorem (Equivalence of uniform asymptotic and uniform exponential stability for linear ordinary differential equations) *Consider the system of linear homogeneous ordinary differential equations* F *with right-hand side (4.5) and suppose that* $\mathrm{A} \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *is continuous. Suppose that* $\sup \mathbb{T} = \infty$. *Then* F *is UGAS if and only if it is UGES.*

Proof It is clear that UGES implies UGAS, so we will only prove the converses.

(UGAS $\implies$ UGES) By definition of uniform asymptotic stability, there exists $C, T \in \mathbb{R}_{>0}$ such that

$$\|\Phi_A(t, t_0)(x)\| \leq C\|x\|$$

and

$$\|\Phi_A(t, t_0)(x)\| \leq \frac{1}{2}\|x\|, \qquad t \geq t_0 + T,$$

for all $(t_0, x) \in \mathbb{T} \times \mathsf{V}$. Then, for $k \in \mathbb{Z}_{>0}$, $(t_0, x) \in \mathbb{T} \times \mathsf{V}$, and $t \geq t_0 + kT$,

$$\begin{aligned}&\|\Phi_A(t, t_0)(x)\| \\ &\quad = \|\Phi_A(t, t_0 + kT) \circ \Phi_A(t_0 + kT, t_0 + (k-1)T) \circ \cdots \circ \Phi_A(t_0 + T, t_0)(x)\| \leq \frac{C}{2^k}\|x\|.\end{aligned}$$

Now define $M = C$ and $\sigma = \frac{\ln 2}{T}$ and let $(t_0, x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. Then $t \in [t_0, t_0 + kT)$ for some uniquely defined $k \in \mathbb{Z}_{>0}$, and then

$$\|\Phi_A(t, t_0)(x)\| \leq \frac{C}{2^k}\|x\| = Me^{-\sigma kT} \leq Me^{\sigma(t-t_0)},$$

as desired. ■

Note that the conclusions of the theorem are not true if we eliminate "uniform" in the hypotheses.

4.2.5 Example (Global asymptotic stability does not imply global exponential stability) We consider the system of linear homogeneous ordinary differential equations F in $\mathsf{V} = \mathbb{R}$ and with

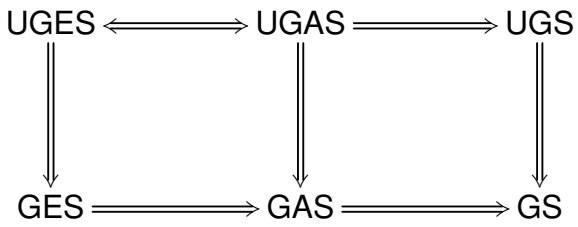
$$\widehat{F}(t, x) = -\frac{x}{t},$$

and we take $\mathbb{T} = [1, \infty)$. This equation can be solved using the methods of Section 2.1 to give

$$\xi(t) = \frac{t_0\xi(t_0)}{t},$$

and from this we conclude that, for any initial condition $\xi(t_0)$, $\lim_{t\to\infty} \xi(t) = 0$ (i.e., we have GAS) but that we do not have exponential stability. •

Let us summarise the relationships between the various notions of stability for systems of linear homogeneous ordinary differential equations in a diagram:



The arrows not present in the diagram represent implications that do not, in fact, hold.

4.2.2 Stability theorems for linear equations

Now we turn to some results concerning the stability of systems of linear homogeneous ordinary differential equations. We proceed in a manner contrary to our approach in Sections 2.2, 2.3, 3.2, and 3.3, and first consider in Section 4.2.2.1 equations with constant coefficients. The rationale is that, for equations with constant coefficients, there are easily understandable characterisations for all of the various sorts of stability. When we turn in Section 4.2.2.2 to general equations, the constant coefficient characterisations give us something with which to compare. Much of what can be said for the stability of linear equations with constant

coefficients has to do with the roots of the characteristic polynomial of the linear transformation associated to the equation. In Section 4.2.2.3 we give some methods for understanding the roots of polynomials without having to compute them.

4.2.2.1 Equations with constant coefficients We shall study the stability of systems of linear homogeneous ordinary differential equations F with constant coefficients in a finite-dimensional $\mathbb{R}$-vector space V. Such an equation will have right-hand side

$$\begin{aligned}\widehat{F}\colon \mathbb{T}\times\mathsf{V} &\to \mathsf{V}\\ (t,x) &\mapsto A(x)\end{aligned}$$

for $A \in \mathrm{L}(\mathsf{V};\mathsf{V})$.

First we observe that the general stability definitions of Definition 4.2.1 for linear homogeneous ordinary differential equations collapse.

4.2.6 Proposition (Collapsing of stability definitions for linear homogeneous equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *and with right-hand side* $\widehat{\mathrm{F}}(\mathrm{t},\mathrm{x}) = \mathrm{A}(\mathrm{x})$ *for* $\mathrm{A} \in \mathrm{L}(\mathsf{V};\mathsf{V})$. *Suppose that* $\sup\mathbb{T} = \infty$. *Then* F *is* GS *(resp.* GAS*,* GES*) if and only if it is* UGS *(resp.* UGAS*,* UGES*). Moreover,* F *is* GAS *if and only if it is* GES*.*

Proof The first assertion follows from Proposition 4.1.5 and the second follows from Theorem 4.2.4. ■

Now we turn to providing a useful characterisation of stability for linear homogeneous ordinary differential equations with constant coefficients. To do this we first make a definition.

4.2.7 Definition (Spectrum of a linear transformation) Let V be a finite-dimensional $\mathbb{R}$-vector space and let $A \in \mathrm{L}(\mathsf{V};\mathsf{V})$. The ***spectrum*** of A is the set

$$\mathrm{spec}(A) = \{\lambda \in \mathbb{C} \mid \lambda \text{ is an eigenvalue for } A^{\mathbb{C}}\}$$

of eigenvalues of the complexification of A. •

Our characterisations of stability will be given in terms of the location of $\mathrm{spec}(A)$. It will be convenient to introduce the following notation:

$$\begin{aligned}\mathbb{C}_- &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) < 0\}, \quad \mathbb{C}_+ = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0\},\\ \overline{\mathbb{C}}_- &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) \le 0\}, \quad \overline{\mathbb{C}}_+ = \{z \in \mathbb{C} \mid \mathrm{Re}(z) \ge 0\},\\ \mathrm{i}\mathbb{R} &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) = 0\}.\end{aligned}$$

With this notation, we state the following theorem, which is the main result of this section.

4.2.8 Theorem (Stability of systems of linear homogeneous ordinary differential equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional vector space* V *with constant coefficients and with* $\widehat{\mathrm{F}}(\mathrm{t},\mathrm{x}) = \mathrm{A}(\mathrm{x})$ *for* $\mathrm{A} \in \mathrm{L}(\mathsf{V};\mathsf{V})$*. The following statements hold.*

(i) F *unstable if* $\operatorname{spec}(\mathrm{A}) \cap \mathbb{C}_+ \neq \emptyset$*.*

(ii) F *is GAS if* $\operatorname{spec}(\mathrm{A}) \subseteq \mathbb{C}_-$*.*

(iii) F *is GS if* $\operatorname{spec}(\mathrm{A}) \cap \mathbb{C}_+ = \emptyset$ *and if* $\mathrm{m_g}(\lambda, \mathrm{A}) = \mathrm{m_a}(\lambda, \mathrm{A})$ *for* $\lambda \in \operatorname{spec}(\mathrm{A}) \cap (\mathrm{i}\mathbb{R})$*.*

(iv) F *is unstable if* $\mathrm{m_g}(\lambda, \mathrm{A}) < \mathrm{m_a}(\lambda, \mathrm{A})$ *for* $\lambda \in \operatorname{spec}(\mathrm{A}) \cap (\mathrm{i}\mathbb{R})$*.*

Proof (i) In this case there is an eigenvalue $\sigma + \mathrm{i}\omega \in \mathbb{C}_+$ and a corresponding eigenvector $u + \mathrm{i}v \in \mathsf{V}^{\mathbb{C}}$ which gives rise to real solutions

$$\xi_1(t) = \mathrm{e}^{\sigma t}(\cos(\omega t)u - \sin(\omega t)v), \quad \xi_2(t) = \mathrm{e}^{\sigma t}(\sin(\omega t)u + \cos(\omega t)v).$$

Clearly these solutions are unbounded as $t \to \infty$ since $\sigma > 0$.

(ii) If all eigenvalues lie in $\mathbb{C}_-$, then any solution of F will be a linear combination of n linearly independent vector functions of the form

$$t^k \mathrm{e}^{-\alpha t}u \quad \text{or} \quad t^k \mathrm{e}^{-\sigma t}(\cos(\omega t)u - \sin(\omega t)v) \quad \text{or} \quad t^k \mathrm{e}^{-\sigma t}(\sin(\omega t)u + \cos(\omega t)v) \tag{4.3}$$

for $\alpha, \sigma > 0$. Note that all such functions tend in length to zero as $t \to \infty$. Suppose that we have a collection $\xi_1, \dots, \xi_n$ of such vector functions. Then, for any solution ξ we have, for some constants $c_1, \dots, c_n$,

$$\begin{aligned} \lim_{t\to\infty} \|\xi(t)\| &= \lim_{t\to\infty} \|c_1\xi_1(t) + \cdots + c_n\xi_n(t)\| \\ &\leq |c_1| \lim_{t\to\infty} \|\xi_1(t)\| + \cdots + |c_n| \lim_{t\to\infty} \|\xi_n(t)\| \\ &= 0, \end{aligned}$$

where we have used the triangle inequality, and the fact that the solutions $\xi_1, \dots, \xi_n$ all tend to zero as $t \to \infty$.

(iii) If $\operatorname{spec}(A) \cap \mathbb{C}_+ = \emptyset$ and if, further, $\operatorname{spec}(A) \subseteq \mathbb{C}_-$, then we are in case (ii), so F is GAS, and so GS. Thus we need only concern ourselves with the case when we have eigenvalues on the imaginary axis. In this case, provided that all such eigenvalues have equal geometric and algebraic multiplicities, all solutions will be linear combinations of functions like those in (4.3) or functions like

$$\sin(\omega t)u \quad \text{or} \quad \cos(\omega t)u. \tag{4.4}$$

Let $\xi_1, \dots, \xi_\ell$ be ℓ linearly independent functions of the form (4.3), and let $\xi_{\ell+1}, \dots, \xi_n$ be linearly independent functions of the form (4.4), so that $\xi_1, \dots, \xi_n$ forms a set of linearly independent solutions for F. Thus we will have, for some constants

$c_1, \dots, c_n$,

$$\begin{aligned}\limsup_{t\to\infty}\|\xi(t)\| &= \limsup_{t\to\infty}\|c_1\xi_1(t) + \cdots + c_n\xi_n(t)\| \\ &\le |c_1|\limsup_{t\to\infty}\|\xi_1(t)\| + \cdots + |c_\ell|\limsup_{t\to\infty}\|\xi_\ell(t)\| + \\ &\quad |c_{\ell+1}|\limsup_{t\to\infty}\|\xi_{\ell+1}(t)\| + \cdots + |c_n|\limsup_{t\to\infty}\|\xi_n(t)\| \\ &= |c_{\ell+1}|\limsup_{t\to\infty}\|\xi_{\ell+1}(t)\| + \cdots + |c_n|\limsup_{t\to\infty}\|\xi_n(t)\|.\end{aligned}$$

Since each of the terms $\|\xi_{\ell+1}(t)\|, \dots, \|\xi_n(t)\|$ are bounded as functions of t, their lim sup's will exist, which is what we wish to show.

(iv) If A has an eigenvalue $\lambda = \mathrm{i}\omega$ on the imaginary axis for which $m_{\mathrm{g}}(\lambda, A) < algmult(\lambda, A)$, then there will be solutions for F that are linear combinations of vector functions of the form $t^k \sin(\omega t)u$ or $t^k \cos(\omega t)v$. Such functions are unbounded as $t \to \infty$, and so F is unstable. ■

4.2.9 Remarks (Stability and eigenvalues)

1. A matrix $\boldsymbol{A}$ is ***Hurwitz*** if $\mathrm{spec}(A) \subseteq \mathbb{C}_-$. Thus A is Hurwitz if and only if F is GAS.
2. We see that stability is almost completely determined by the eigenvalues of A. Indeed, one says that F is ***spectrally stable*** if A has no eigenvalues in $\mathbb{C}_+$. It is only in the case where there are repeated eigenvalues on the imaginary axis that one gets to distinguish spectral stability from stability. •

The notion of stability for systems of linear homogeneous ordinary differential equations with constant coefficients is, in principle, an easy one to check, as we see from an example.

4.2.10 Example (Stability of system of linear homogeneous ordinary differential equations with constant coefficients)

We look at a system of linear homogeneous ordinary differential equations $\boldsymbol{F}$ in $\mathbb{R}^2$ with constant coefficients, and determined by the 2×2-matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}.$$

The eigenvalues of $\boldsymbol{A}$ are the roots of the characteristic polynomial $P_A = X^2 + aX + b$, and these are

$$-\tfrac{a}{2} \pm \tfrac{1}{2}\sqrt{a^2 - 4b}.$$

The situation with the eigenvalue placement can be broken into cases.

1. $a = 0$ and $b = 0$: In this case there is a repeated zero eigenvalue. Thus we have spectral stability, but we need to look at eigenvectors to determine stability. One readily verifies that there is only one linearly independent eigenvector for the zero eigenvalue, so the system is unstable.

2. $a = 0$ and $b > 0$: In this case the eigenvalues are purely imaginary. Since the roots are also distinct, they will have equal algebraic and geometric multiplicity. Thus the system is GS, but not GAS.
3. $a = 0$ and $b < 0$: In this case both roots are real, and one will be positive. Thus the system is unstable.
4. $a > 0$ and $b = 0$: There will be one zero eigenvalue if $b = 0$. If $a > 0$ the other root will be real and negative. In this case then, we have a root on the imaginary axis. Since it is distinct, the system will be GS, but not GAS.
5. $a > 0$ and $b > 0$: One may readily ascertain (in Section 4.2.2.3 we'll see an easy way to do this) that all eigenvalues are in $\mathbb{C}_-$ if $a > 0$ and $b > 0$. Thus when a and b are strictly positive, the system is GAS.
6. $a > 0$ and $b < 0$: In this case both eigenvalues are real, one being positive and the other negative. Thus the system is unstable.
7. $a < 0$ and $b = 0$: We have one zero eigenvalue. The other, however, will be real and positive, and so the system is unstable.
8. $a < 0$ and $b > 0$: We play a little trick here. If s_0 is a root of $s^2 + as + b$ with $a, b < 0$, then $-s_0$ is clearly also a root of $s^2 - as + b$. From the previous case, we know that $-s_0 \in \mathbb{C}_-$, which means that $s_0 \in \mathbb{C}_+$. So in this case all eigenvalues are in $\mathbb{C}_+$, and so we have instability.
9. $a < 0$ and $b < 0$: In this case we are guaranteed that all eigenvalues are real, and furthermore it is easy to see that one eigenvalue will be positive, and the other negative. Thus the system will be unstable. •

4.2.2.2 Equations with time-varying coefficients We work in this section with a system F of linear homogeneous ordinary differential equations with right-hand side

$$\begin{aligned} \widehat{F} \colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V} \\ (t, x) &\mapsto A(t)(x) \end{aligned} \tag{4.5}$$

for some function $A \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$.

4.2.2.3 Hurwitz polynomials From Theorem 4.2.8 we see that it is important to be able to determine when the roots of a polynomial lie in the negative half-plane. However, checking that such a condition holds may not be so easy; one should regard the problem of computing the roots of a polynomial as being impossible for polynomials of degree 5 or more, and annoyingly complicated for polynomials of degree 3 or 4. However, one may establish conditions on the coefficients of a polynomial. In this section, we present three methods for doing exactly this. We also look at a test for the roots to lie in $\mathbb{C}_-$ when we only approximately know the coefficients of the polynomial. We shall generally say that a polynomial all of whose roots lie in $\mathbb{C}_-$ is ***Hurwitz***.

The Routh criterion

For the method of Routh, we construct an array involving the coefficients of the polynomial in question. The array is constructed inductively, starting with the first two rows. Thus suppose one has two collections $a_{11}, a_{12}, \dots$ and $a_{21}, a_{22}, \dots$ of numbers. In practice, this is a finite collection, but let us suppose the length of each collection to be indeterminate for convenience. Now construct a third row of numbers $a_{31}, a_{32}, \dots$ by defining $a_{3k} = a_{21}a_{1,k+1} - a_{11}a_{2,k+1}$. Thus a_{3k} is minus the determinant of the matrix $\left[\begin{smallmatrix} a_{11} & a_{1,k+1} \\ a_{21} & a_{2,k+1} \end{smallmatrix}\right]$. In practice, one writes this down as follows:

$$\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1k} & \cdots \\ a_{21} & a_{22} & \cdots & a_{2k} & \cdots \\ a_{21}a_{12} - a_{11}a_{22} & a_{21}a_{13} - a_{11}a_{23} & \cdots & a_{21}a_{1,k+1} - a_{11}a_{2,k+1} & \cdots \end{array}$$

One may now proceed in this way, using the second and third row to construct a fourth row, the third and fourth row to construct a fifth row, and so on. To see how to apply this to a given polynomial $P \in \mathbb{R}[X]$. Define two polynomials $P_+, P_- \in \mathbb{R}[X]$ as the even and odd part of P. To be clear about this, if

$$P = p_0 + p_1X + p_2X^2 + p_3X^3 + \cdots + p_{n-1}X^{n-1} + p_nX^n,$$

then

$$P_+ = p_0 + p_2X + p_4X^2 + \dots, \quad P_- = p_1 + p_3X + p_5X^2 + \dots.$$

Note then that $P(X) = P_+(X^2) + XP_-(X^2)$. Let $R(P)$ be the array constructed as above, with the first two rows being comprised of the coefficients of P_+ and P_-, respectively, starting with the coefficients of lowest powers of X, and increasing to higher powers of X. Thus the first three rows of $R(P)$ are

$$\begin{array}{cccc} p_0 & p_2 & \cdots & p_{2k} & \cdots \\ p_1 & p_3 & \cdots & p_{2k+1} & \cdots \\ p_1p_2 - p_0p_3 & p_1p_4 - p_0p_5 & \cdots & p_1p_{2k+2} - p_0p_{2k+3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

In making this construction, a zero is inserted whenever an operation is undefined. It is readily determined that the first column of $R(P)$ has at most $n + 1$ nonzero components. The ***Routh array*** is then the first column of the first $n + 1$ rows.

With this as setup, we may now state a criterion for determining whether a polynomial is Hurwitz.

4.2.11 Theorem (Routh's criterion) *A polynomial*

$$\mathrm{P} = \mathrm{X}^{\mathrm{n}} + \mathrm{p}_{\mathrm{n}-1}\mathrm{X}^{\mathrm{n}-1} + \cdots + \mathrm{p}_1\mathrm{X} + \mathrm{p}_0 \in \mathbb{R}[\mathrm{X}]$$

is Hurwitz if and only if all elements of the Routh array corresponding to R(P) *are positive.*

Proof Let us construct a sequence of polynomials as follows. We let $P_0 = P_+$ and $P_1 = P_-$ and let

$$P_2(X) = X^{-1}\Big(P_1(0)P_0(X) - P_0(0)P_1(X)\Big).$$

Note that the constant coefficient of $P_1(0)P_0(X) - P_0(0)P_1(X)$ is zero, so this does indeed define P_2 as a polynomial. Now inductively define

$$P_k(X) = X^{-1}\Big(P_{k-1}(0)P_{k-2}(X) - P_{k-2}(0)P_{k-1}(X)\Big)$$

for $k \geq 3$. With this notation, we have the following lemma that describes the statement of the theorem.

1 Lemma *The* $(k+1)$*st row of* $R(P)$ *consists of the coefficients of* P_k *with the constant coefficient in the first column. Thus the hypothesis of the theorem is equivalent to the condition that* $P_0(0), P_1(0), \dots, P_n(0)$ *all be positive.*

Proof We have $P_0(0) = p_0$, $P_1(0) = p_1$, and $P_2(0) = p_1p_2 - p_0p_3$, directly from the definitions. Thus the lemma holds for $k \in \{0, 1, 2\}$. Now suppose that the lemma holds for $k \geq 3$. Thus the kth and the $(k+1)$st rows of $R(P)$ are the coefficients of the polynomials

$$P_{k-1}(X) = p_{k-1,0} + p_{k-1,1}X + \cdots$$

and

$$P_k(X) = p_{k,0} + p_{k,1}X + \cdots,$$

respectively. Using the definition of P_{k+1} we see that $P_{k+1}(0) = p_{k,0}p_{k-1,1} - p_{k-1,0}p_{k,1}$. However, this is exactly the term as it would appear in first column of the $(k+2)$nd row of $R(P)$. ▼

Now note that $P(X) = P_0(X^2) + XP_1(X^2)$ and define $Q \in \mathbb{R}[X]$ by $Q(X) = P_1(X^2) + XP_2(X^2)$. One may readily verify that $\deg(Q) \leq n - 1$. Indeed, in the proof of Theorem 4.2.13, a formula for Q will be given. The following lemma is key to the proof. Let us suppose for the moment that p_n is not equal to 1.

2 Lemma *The following statements are equivalent:*

(i) P *is Hurwitz and* $p_n > 0$*;*

(ii) Q *is Hurwitz,* $q_{n-1} > 0$*, and* $P(0) > 0$*.*

Proof We have already noted that $P(X) = P_0(X^2) + XP_1(X^2)$. We may also compute

$$Q(X) = P_1(X^2) + X^{-1}\Big(P_1(0)P_0(X^2) - P_0(0)P_1(X^2)\Big). \tag{4.6}$$

For $\lambda \in [0, 1]$ define $Q_\lambda(X) = (1-\lambda)P(X) + \lambda Q(X)$, and compute

$$Q_\lambda(X) = \Big((1-\lambda) + X^{-1}\lambda P_1(0)\Big)P_0(X^2) + \Big((1-\lambda)X + \lambda - X^{-1}\lambda P_0(0)\Big)P_1(X^2).$$

The polynomials $P_0(X^2)$ and $P_1(X^2)$ are even, so that when evaluated on the imaginary axis they are real. Now we claim that the roots of Q_λ that lie on the imaginary

axis are independent of λ, provided that $P(0) > 0$ and $Q(0) > 0$. First note that, if $P(0) > 0$ and $Q(0) > 0$, then 0 is not a root of Q_λ. Now, if $\mathrm{i}\omega_0$ is a nonzero imaginary root, then we must have

$$\Big((1-\lambda) - \mathrm{i}\omega_0^{-1}\lambda P_1(0)\Big)P_0(-\omega_0^2) + \Big((1-\lambda)i\omega_0 + \lambda + \mathrm{i}\omega_0^{-1}\lambda P_0(0)\Big)P_1(-\omega_0^2) = 0.$$

Balancing real and imaginary parts of this equation gives

$$\begin{aligned}&(1-\lambda)P_0(-\omega_0^2) + \lambda P_1(-\omega_0^2) = 0\\ &\lambda\omega_0^{-1}\Big(P_0(0)P_1(-\omega_0^2) - P_1(0)P_0(-\omega_0^2)\Big) + \omega_0(1-\lambda)P_1(-\omega_0^2).\end{aligned} \tag{4.7}$$

If we think of this as a homogeneous linear equation in $P_0(-\omega_0^2)$ and $P_1(\omega_0^2)$, one determines that the determinant of the coefficient matrix is

$$\omega_0^{-1}\Big((1-\lambda)^2\omega_0^2 + \lambda((1-\lambda)P_0(0) + \lambda P_1(0))\Big).$$

This expression is positive for $\lambda \in [0,1]$ since $P(0), Q(0) > 0$ implies that $P_0(0), P_1(0) > 0$. To summarise, we have shown that, provided that $P(0) > 0$ and $Q(0) > 0$, all imaginary axis roots $\mathrm{i}\omega_0$ of Q_λ satisfy $P_0(-\omega_0^2) = 0$ and $P_1(-\omega_0^2) = 0$. In particular, the imaginary axis roots of Q_λ are independent of $\lambda \in [0,1]$ in this case.

(i) $\Longrightarrow$ (ii) For $\lambda \in [0,1]$ let

$$N(\lambda) = \begin{cases} n, & \lambda \in [0,1) \\ n-1, & \lambda = 1. \end{cases}$$

Thus $N(\lambda)$ is the number of roots of Q_λ. Now let

$$Z_\lambda = \{z_{\lambda,i} \mid \ i \in \{1,\dots,N(\lambda)\}\}$$

be the set of roots of Q_λ. Since P is Hurwitz, $Z_0 \subseteq \mathbb{C}_-$. Our previous computations then show that $Z_\lambda \cap \mathrm{i}\mathbb{R} = \emptyset$ for $\lambda \in [0,1]$. Now, if $Q = Q_1$ were to have a root in $\overline{\mathbb{C}}_+$, this would mean that, for some value of λ, one of the roots of Q_λ would have to lie on the imaginary axis, using the (nontrivial) fact that the roots of a polynomial are continuous functions of its coefficients. This then shows that all roots of Q must lie in $\mathbb{C}_-$. That $P(0) > 0$ is a consequence of Exercise 4.2.3 and P being Hurwitz. One may check that $q_{n-1} = p_1 \cdots p_n$, so that $q_{n-1} > 0$ follows from Exercise 4.2.3 and $p_n > 0$.

(ii) $\Longrightarrow$ (i) Let us adopt the notation $N(\lambda)$ and Z_λ from the previous part of the proof. Since Q is Hurwitz, $Z_1 \subseteq \mathbb{C}_-$. Furthermore, since $Z_\lambda \cap \mathrm{i}\mathbb{R} = \emptyset$, it follows that, for $\lambda \in [0,1]$, the number of roots of Q_λ within $\mathbb{C}_-$ must equal $n-1$ as $\deg(Q) = n-1$. In particular, P can have at most one root in $\mathbb{C}_+$. This root, then, must be real, and let us denote it by $z_0 > 0$. Thus $P(X) = \tilde{P}(X)(X - z_0)$ where $\tilde{P}$ is Hurwitz. By Exercise 4.2.3 it follows that all coefficients of $\tilde{P}$ are positive. If we write

$$\tilde{P} = \tilde{p}_{n-1}X^{n-1} + \tilde{p}_{n-2}X^{n-2} + \cdots + \tilde{p}_1 X + \tilde{p}_0,$$

then

$$P(X) = \tilde{p}_{n-1}X^n + (\tilde{p}_{n-2} - z_0\tilde{p}_{n-1})X^{n-1} + \cdots + (\tilde{p}_0 - z_0\tilde{p}_1)X - \tilde{p}_0 z_0.$$

Thus the existence of a root $z_0 \in \mathbb{C}_+$ contradicts the fact that $P(0) > 0$. Note that we have also shown that $p_n > 0$. ▼

Now we proceed with the proof proper. First suppose that P is Hurwitz. By successive applications of Lemma 2, it follows that the polynomials

$$Q_k(X) = P_k(X^2) + XP_{k+1}(X^2), \quad k \in \{1, \dots, n\},$$

are Hurwitz and that $\deg(Q_k) = n - k$, $k \in \{1, \dots, n\}$. What's more, the coefficient of X^{n-k} is positive in Q_k. Now, by Exercise 4.2.3, we have $P_0(0) > 0$ and $P_1(0) > 0$. Now suppose that $P_0(0), P_1(0), \dots, P_k(0)$ are all positive. Since Q_k is Hurwitz with the coefficient of the highest power of X being positive, from Exercise 4.2.3 it follows that the coefficient of X in Q_k should be positive. However, this coefficient is exactly $P_{k+1}(0)$. Thus we have shown that $P_k(0) > 0$ for $k = 0, 1, \dots, n$. From Lemma 1 it follows that the elements of the Routh array are positive.

Now suppose that one element of the Routh array is nonpositive and that P is Hurwitz. By Lemma 2, we may suppose that $P_{k_0}(0) \leq 0$ for some $k_0 \in \{2, 3, \dots, n\}$. Furthermore, since P is Hurwitz, as above the polynomials Q_k, $k \in \{1, \dots, n\}$, must also be Hurwitz, with $\deg(Q_k) = n - k$ where the coefficient of X^{n-k} in Q_k is positive. In particular, by Exercise 4.2.3, all coefficients of Q_{k_0-1} are positive. However, since $Q_{k_0-1}(X) = P_{k_0-1}(X^2) + XP_{k_0}(X^2)$ it follows that the coefficient of X in Q_{k_0-1} is negative, and hence we arrive at a contradiction, and the theorem follows. ■

The Routh criterion is simple to apply, and we illustrate it in the simple case of a degree two polynomial.

4.2.12 Example (The Routh criterion) Let us apply the criteria to the simplest nontrivial example possible: $P = X^2 + aX + b$. We compute the Routh table to be

$$R(P) = \begin{matrix} b & 1 \\ a & 0. \\ a & 0 \end{matrix}$$

Thus the Routh array is $\begin{bmatrix} b & a & a \end{bmatrix}$, and its entries are all positive if and only if $a, b > 0$. Let us see how this compares to what we know doing the calculations "by hand." The roots of P are $r_1 = -\frac{a}{2} + \frac{1}{2}\sqrt{a^2 - 4b}$ and $r_2 = -\frac{a}{2} - \frac{1}{2}\sqrt{a^2 - 4b}$. Let us consider the various cases.

1. If $a^2 - 4b < 0$, then the roots are complex with nonzero imaginary part, and with real part $-a$. Thus the roots in this case lie in the negative half-plane if and only if $a > 0$. We also have $b > \frac{a^2}{4}$ and so $b > 0$, and hence $ab > 0$ as in the Routh criterion.
2. If $a^2 - 4b = 0$, then the roots are both $-a$, and so lie in the negative half-plane if and only if $a > 0$. In this case $b = \frac{a^2}{4}$ and so $b > 0$. Thus $ab > 0$ as predicted.

3. Finally we have the case when $a^2 - 4b > 0$. We have two subcases.
   (a) When $a > 0$, then we have negative half-plane roots if and only if $a^2 - 4b < a^2$ which means that $b > 0$. Therefore, we have negative half-plane roots if and only $a > 0$ and $ab > 0$.
   (b) When $a < 0$, then we will never have all negative half-plane roots since $-a + \sqrt{a^2 - 4b}$ is always positive.

So we see that the Routh criterion provides a very simple encapsulation of the necessary and sufficient conditions for all roots to lie in the negative half-plane, even for this simple example. •

The Hurwitz criterion

We consider in this section another test for a polynomial to be Hurwitz. The key ingredient in the Hurwitz construction we consider is a matrix formed from the coefficients of a polynomial

$$P = X^n + p_{n-1}X^{n-1} + \cdots + p_1 X + p_0 \in \mathbb{R}[X].$$

We denote the ***Hurwitz matrix*** by $\boldsymbol{H}(P) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ and define it by

$$\boldsymbol{H}(P) = \begin{bmatrix} p_{n-1} & 1 & 0 & 0 & \cdots & 0 \\ p_{n-3} & p_{n-2} & p_{n-1} & 1 & \cdots & 0 \\ p_{n-5} & p_{n-4} & p_{n-3} & p_{n-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_0 \end{bmatrix}.$$

Any terms in this matrix that are not defined are taken to be zero. Of course, we also take $p_n = 1$. Now define $\boldsymbol{H}(P)_k \in \mathrm{L}(\mathbb{R}^k; \mathbb{R}^k)$, $k \in \{1, \ldots, n\}$, to be the matrix of elements $\boldsymbol{H}(P)_{ij}$, $i, j \in \{1, \ldots, k\}$. Thus $\boldsymbol{H}(P)_k$ is the matrix formed by taking the "upper left $k \times k$ block from $\boldsymbol{H}(P)$." Also define $\Delta_k = \det \boldsymbol{H}(P)_k$.

With this notation, the Hurwitz criterion is as follows.

4.2.13 Theorem (Hurwitz's criterion) *A polynomial*

$$\mathrm{P} = \mathrm{X}^{\mathrm{n}} + \mathrm{p}_{\mathrm{n}-1}\mathrm{X}^{\mathrm{n}-1} + \cdots + \mathrm{p}_1\mathrm{X} + \mathrm{p}_0 \in \mathbb{R}[\mathrm{X}]$$

is Hurwitz if and only if the n ***Hurwitz determinants*** $\Delta_1, \ldots, \Delta_{\mathrm{n}}$ *are positive.*

Proof Let us begin by resuming with the notation from the proof of Theorem 4.2.11. In particular, we recall the definition of $Q(X) = P_1(X^2) + XP_2(X^2)$. We wish to compute $\boldsymbol{H}(Q)$, so we need to compute Q in terms of the coefficients of P. A computation using the definition of Q and P_2 gives

$$Q(X) = p_1 + (p_1p_2 - p_0p_3)X + p_3X^2 + (p_1p_4 - p_0p_5)X^3 + \cdots.$$

One can then see that, when n is even, we have

$$\boldsymbol{H}(Q) = \begin{bmatrix} p_{n-1} & p_1 p_n & 0 & 0 & \cdots & 0 & 0 \\ p_{n-3} & p_1 p_{n-2} - p_0 p_{n-1} & p_{n-1} & p_1 p_n & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_1 p_2 - p_0 p_3 & p_3 \\ 0 & 0 & 0 & 0 & \cdots & 0 & p_1 \end{bmatrix}$$

and, when n is odd, we have

$$\boldsymbol{H}(Q) = \begin{bmatrix} p_1 p_{n-1} - p_0 p_n & p_n & 0 & 0 & \cdots & 0 & 0 \\ p_1 p_{n-3} - p_0 p_{n-2} & p_{n-2} & p_1 p_{n-1} - p_0 p_n & p_n & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_1 p_2 - p_0 p_3 & p_3 \\ 0 & 0 & 0 & 0 & \cdots & 0 & p_1 \end{bmatrix}.$$

Now define $\boldsymbol{T} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ by

$$\boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & p_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -p_0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -p_0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

when n is even and by

$$\boldsymbol{T} = \begin{bmatrix} p_1 & 0 & \cdots & 0 & 0 & 0 \\ -p_0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & p_1 & 0 & 0 \\ 0 & 0 & \cdots & -p_0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

when n is odd. One then verifies by direct calculation that

$$\boldsymbol{H}(P)\boldsymbol{T} = \begin{bmatrix} & & & \vdots \\ & \boldsymbol{H}(Q) & & p_4 \\ & & & p_2 \\ 0 & \cdots & 0 & p_0 \end{bmatrix}. \tag{4.8}$$

We now let $\Delta_1, \ldots, \Delta_n$ be the determinants defined above and let $\tilde{\Delta}_1, \ldots, \tilde{\Delta}_{n-1}$ be the similar determinants corresponding to $\boldsymbol{H}(Q)$. A straightforward computation

using (4.8) gives the following relationships between the Δ's and the $\tilde{\Delta}$'s:

$$\begin{aligned}\Delta_1 &= p_1\\ \Delta_{k+1} &= \begin{cases} p_1^{-\lfloor \frac{k}{2} \rfloor}\tilde{\Delta}_k, & k \text{ even}\\ p_1^{-\lceil \frac{k}{2} \rceil}\tilde{\Delta}_k, & k \text{ odd}\end{cases}, \qquad k = 1, \dots, n-1,\end{aligned} \tag{4.9}$$

where $\lfloor x \rfloor$ gives the greatest integer less than or equal to x and $\lceil x \rceil$ gives the smallest integer greater than or equal to x.

With this background notation, let us proceed with the proof, first supposing that P is Hurwitz. In this case, by Exercise 4.2.3, it follows that $p_1 > 0$ so that $\Delta_1 > 0$. By Lemma 2 of Theorem 4.2.11, it also follows that Q is Hurwitz. Thus $\tilde{\Delta}_1 > 0$. A trivial induction argument on $n = \deg(P)$ then shows that $\Delta_2, \dots, \Delta_n > 0$.

Now suppose that one of $\Delta_1, \dots, \Delta_n$ is nonpositive and that P is Hurwitz. Since Q is then Hurwitz by Lemma 2 of Theorem 4.2.11, we readily arrive at a contradiction, and this completes the proof. ■

The Hurwitz criterion is simple to apply, and we illustrate it in the simple case of a degree two polynomial.

4.2.14 Example (The Hurwitz criterion) Let us apply the criteria to our simple example of $P = X^2 + aX + b$. We then have

$$\boldsymbol{H}(P) = \begin{bmatrix} a & 1\\ 0 & b\end{bmatrix}$$

We then compute $\Delta_1 = a$ and $\Delta_2 = ab$. Thus $\Delta_1, \Delta_2 > 0$ if and only if $a, b > 0$. This agrees with our application of the Routh method to the same polynomial in Example 4.2.12. •

The Hermite criterion

We next look at a manner of determining whether a polynomial is Hurwitz which makes contact with the Lyapunov methods of Section 4.3.6. Let us consider, as usual, a monic polynomial of degree n:

$$P(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0.$$

Corresponding to such a polynomial, we construct its ***Hermite matrix*** as the $n \times n$ matrix $\boldsymbol{P}(P)$ given by

$$\boldsymbol{P}(P)_{ij} = \begin{cases} \sum_{k=1}^{i} (-1)^{k+i} p_{n-k+1} p_{n-i-j+k}, & j \geq i,\ i+j \text{ even}\\ \boldsymbol{P}(P)_{ji}, & j < i,\ i+j \text{ even}\\ 0, & i+j \text{ odd}.\end{cases}$$

As usual, in this formula we take $p_i = 0$ for $i < 0$. One can get an idea of how this matrix is formed by looking at its appearance for small values of n. For $n = 2$ we have

$$\boldsymbol{P}(P) = \begin{bmatrix} p_1p_2 & 0 \\ 0 & p_0p_1 \end{bmatrix},$$

for $n = 3$ we have

$$\boldsymbol{P}(P) = \begin{bmatrix} p_2p_3 & 0 & p_0p_3 \\ 0 & p_1p_2 - p_0p_3 & 0 \\ p_0p_3 & 0 & p_0p_1 \end{bmatrix},$$

and for $n = 4$ we have

$$\boldsymbol{P}(P) = \begin{bmatrix} p_3p_4 & 0 & p_1p_4 & 0 \\ 0 & p_2p_3 - p_1p_4 & 0 & p_0p_3 \\ p_1p_4 & 0 & p_1p_2 - p_0p_3 & 0 \\ 0 & p_0p_3 & 0 & p_0p_1 \end{bmatrix}.$$

The following theorem gives necessary and sufficient conditions for P to be Hurwitz based on its Hermite matrix.

4.2.15 Theorem (Hermite's criterion) *A polynomial*

$$\mathrm{P}(\mathrm{s}) = \mathrm{s}^{\mathrm{n}} + \mathrm{p}_{\mathrm{n}-1}\mathrm{s}^{\mathrm{n}-1} + \cdots + \mathrm{p}_1\mathrm{s} + \mathrm{p}_0 \in \mathbb{R}[\mathrm{s}]$$

is Hurwitz if and only if $\mathbf{P}(\mathrm{P})$ *is positive-definite.*

Proof Let

$$\boldsymbol{A}(P) = \begin{bmatrix} -p_{n-1} & -p_{n-2} & \cdots & -p_1 & -p_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \boldsymbol{b}(P) = \begin{bmatrix} p_{n-1} \\ 0 \\ p_{n-3} \\ 0 \\ \vdots \end{bmatrix}.$$

An unenjoyable computation gives

$$\boldsymbol{P}(P)\boldsymbol{A}(P) + \boldsymbol{A}(P)^T\boldsymbol{P}(P) = -\boldsymbol{b}(P)\boldsymbol{b}(P)^T.$$

First suppose that $\boldsymbol{P}(P)$ is positive-definite. By Theorem 4.3.27(i), since $\boldsymbol{b}(P)\boldsymbol{b}(P)^T$ is positive-semidefinite, $\boldsymbol{A}(P)$ is Hurwitz. Conversely, if $\boldsymbol{A}(P)$ is Hurwitz, then there is only one symmetric $\boldsymbol{P}$ so that

$$\boldsymbol{P}\boldsymbol{A}(P) + \boldsymbol{A}(P)^T\boldsymbol{P} = -\boldsymbol{b}(P)\boldsymbol{b}(P)^T,$$

this by Theorem 4.3.59(i). Since $\boldsymbol{P}(P)$ satisfies this relation even when $\boldsymbol{A}(P)$ is not Hurwitz, it follows that $\boldsymbol{P}(P)$ is positive-definite. The theorem now follows since the characteristic polynomial of $\boldsymbol{A}(P)$ is P. ■

Let us apply this theorem to our favourite example.

4.2.16 Example (Hermite's criterion) We consider the polynomial $P(s) = s^2 + as + b$ which has the Hermite matrix

$$\boldsymbol{P}(P) = \begin{bmatrix} a & 0 \\ 0 & ab \end{bmatrix}.$$

Since this matrix is diagonal, it is positive-definite if and only if the diagonal entries are zero. Thus we recover the by now well established condition that $a, b > 0$. •

The Hermite criterion, Theorem 4.2.15, does indeed record necessary and sufficient conditions for a polynomial to be Hurwitz. However, it is more computationally demanding than it needs to be, especially for large polynomials. Part of the problem is that the Hermite matrix contains so many zero entries. To get conditions involving smaller matrices leads to the so-called ***reduced Hermite criterion*** which we now discuss. Given a degree n polynomial P with its Hermite matrix $\boldsymbol{P}(P)$, we define ***reduced Hermite matrices*** $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$ as follows:

1. $\boldsymbol{C}(P)$ is obtained by removing the even numbered rows and columns of $\boldsymbol{P}(P)$ and
2. $\boldsymbol{D}(P)$ is obtained by removing the odd numbered rows and columns of $\boldsymbol{P}(P)$.

Thus, if n is even, $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$ are $\frac{n}{2} \times \frac{n}{2}$, and if n is odd, $\boldsymbol{C}(P)$ is $\frac{n+1}{2} \times \frac{n+1}{2}$ and $\boldsymbol{D}(P)$ is $\frac{n-1}{2} \times \frac{n-1}{2}$. Let us record a few of these matrices for small values of n. For $n = 2$ we have

$$\boldsymbol{C}(P) = \begin{bmatrix} p_1 p_2 \end{bmatrix}, \quad \boldsymbol{D}(P) = \begin{bmatrix} p_0 p_1 \end{bmatrix},$$

for $n = 3$ we have

$$\boldsymbol{C}(P) = \begin{bmatrix} p_2 p_3 & p_0 p_3 \\ p_0 p_3 & p_0 p_1 \end{bmatrix}, \quad \boldsymbol{D}(P) = \begin{bmatrix} p_1 p_2 - p_0 p_3 \end{bmatrix},$$

and for $n = 4$ we have

$$\boldsymbol{C}(P) = \begin{bmatrix} p_3 p_4 & p_1 p_4 \\ p_1 p_4 & p_1 p_2 - p_0 p_3 \end{bmatrix}, \quad \boldsymbol{D}(P) = \begin{bmatrix} p_2 p_3 - p_1 p_4 & p_0 p_3 \\ p_0 p_3 & p_0 p_1 \end{bmatrix}.$$

Let us record a useful property of the matrices $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$, noting that they are symmetric.

4.2.17 Lemma (A property of reduced Hermite matrices) P(P) *is positive-definite if and only if both* C(P) *and* D(P) *are positive-definite.*

Proof For $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, denote $\boldsymbol{x}_{\text{odd}} = (x_1, x_3, \ldots)$ and $\boldsymbol{x}_{\text{even}} = (x_2, x_4, \ldots)$. A simple computation then gives

$$\boldsymbol{x}^T \boldsymbol{P}(P) \boldsymbol{x} = \boldsymbol{x}_{\text{odd}}^T \boldsymbol{C}(P) \boldsymbol{x}_{\text{odd}} + \boldsymbol{x}_{\text{even}}^T \boldsymbol{D}(P) \boldsymbol{x}_{\text{even}}. \tag{4.10}$$

Clearly, if $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$ are both positive-definite, then so too is $\boldsymbol{P}(P)$. Conversely, suppose that one of $\boldsymbol{C}(P)$ or $\boldsymbol{D}(P)$, say $\boldsymbol{C}(P)$, is not positive-definite. Thus there exists $\boldsymbol{x} \in \mathbb{R}^n$ so that $\boldsymbol{x}_{\text{odd}} \neq \boldsymbol{0}$ and $\boldsymbol{x}_{\text{even}} = \boldsymbol{0}$, and for which

$$\boldsymbol{x}_{\text{odd}}^T \boldsymbol{C}(P) \boldsymbol{x}_{\text{odd}} \leq 0.$$

From (4.10), it now follows that $\boldsymbol{P}(P)$ is not positive-definite. ■

The Hermite criterion then tells us that P is Hurwitz if and only if both $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$ are positive-definite. The remarkable fact is that we need only check one of these matrices for definiteness, and this is recorded in the following theorem.

4.2.18 Theorem (Reduced Hermite criterion) *A polynomial*

$$\mathrm{P} = \mathrm{X}^{\mathrm{n}} + \mathrm{p}_{\mathrm{n}-1}\mathrm{X}^{\mathrm{n}-1} + \cdots + \mathrm{p}_1\mathrm{X} + \mathrm{p}_0 \in \mathbb{R}[\mathrm{X}]$$

is Hurwitz if and only if any one of the following conditions holds:

(i) $\mathrm{p}_{2\mathrm{k}} > 0$, $\mathrm{k} \in \{0, 1, \ldots, \lfloor \frac{\mathrm{n}-1}{2} \rfloor\}$ *and* $\mathbf{C}(\mathrm{P})$ *is positive-definite;*

(ii) $\mathrm{p}_{2\mathrm{k}} > 0$, $\mathrm{k} \in \{0, 1, \ldots, \lfloor \frac{\mathrm{n}-1}{2} \rfloor\}$ *and* $\mathbf{D}(\mathrm{P})$ *is positive-definite;*

(iii) $\mathrm{p}_0 > 0$, $\mathrm{p}_{2\mathrm{k}+1} > 0$, $\mathrm{k} \in \{0, 1, \ldots, \lfloor \frac{\mathrm{n}-2}{2} \rfloor\}$ *and* $\mathbf{C}(\mathrm{P})$ *is positive-definite;*

(iv) $\mathrm{p}_0 > 0$, $\mathrm{p}_{2\mathrm{k}+1} > 0$, $\mathrm{k} \in \{0, 1, \ldots, \lfloor \frac{\mathrm{n}-2}{2} \rfloor\}$ *and* $\mathbf{D}(\mathrm{P})$ *is positive-definite.*

Proof First suppose that P is Hurwitz. Then all coefficients are positive (see Exercise 4.2.3) and $\boldsymbol{P}(P)$ is positive-definite by Theorem 4.2.15. This implies that $\boldsymbol{C}(P)$ and $\boldsymbol{D}(P)$ are positive-definite by Lemma 4.2.17, and thus conditions (i)–(iv) hold. For the converse assertion, the cases when n is even or odd are best treated separately. This gives eight cases to look at. As certain of them are quite similar in flavour, we only give details the first time an argument is encountered.

Case 1: We assume (i) and that n is even. Denote

$$\boldsymbol{A}_1(P) = \begin{bmatrix} -\frac{p_{n-2}}{p_n} & -\frac{p_{n-4}}{p_n} & \cdots & -\frac{p_2}{p_n} & -\frac{p_0}{p_n} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

A calculation then gives $\boldsymbol{C}(P)\boldsymbol{A}_1(P) = -\boldsymbol{D}(P)$. Since $\boldsymbol{C}(P)$ is positive-definite, there exists an orthogonal matrix $\boldsymbol{R}$ so that $\boldsymbol{R}\boldsymbol{C}(P)\boldsymbol{R}^T = \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is diagonal with strictly positive diagonal entries. Let $\boldsymbol{\Delta}^{1/2}$ denote the diagonal matrix whose diagonal entries are the square roots of those of $\boldsymbol{\Delta}$. Now denote $\boldsymbol{C}(P)^{1/2} = \boldsymbol{R}^T\boldsymbol{\Delta}^{1/2}\boldsymbol{R}$, noting that $\boldsymbol{C}(P)^{1/2}\boldsymbol{C}(P)^{1/2} = \boldsymbol{C}(P)$. Also note that $\boldsymbol{C}(P)^{1/2}$ is invertible, and we shall denote its inverse by $\boldsymbol{C}(P)^{-1/2}$. Note that this inverse is also positive-definite. This then gives

$$\boldsymbol{C}(P)^{1/2}\boldsymbol{A}_1(P)\boldsymbol{C}(P)^{-1/2} = -\boldsymbol{C}(P)^{-1/2}\boldsymbol{D}(P)\boldsymbol{C}(P)^{-1/2}. \tag{4.11}$$

The matrix on the right is symmetric, so this shows that $\boldsymbol{A}_1(P)$ is similar to a symmetric matrix, allowing us to deduce that $\boldsymbol{A}_1(P)$ has real eigenvalues. These eigenvalues are also roots of the characteristic polynomial

$$s^{n/2} + \frac{p_{n-2}}{p_n}s^{n/2-1} + \cdots + \frac{p_2}{p_n}s + \frac{p_0}{p_n}.$$

Our assumption (i) ensures that is s is real and nonnegative, the value of the characteristic polynomial is positive. From this we deduce that all eigenvalues of

$\boldsymbol{A}_1(P)$ are negative. From (4.11) it now follows that $\boldsymbol{D}(P)$ is positive-definite, and so P is Hurwitz by Lemma 4.2.17 and Theorem 4.2.15.

Case 2: We assume (ii) and that n is even. Consider the polynomial $P^{-1}(s) = s^n P(\frac{1}{s})$. Clearly the roots of P^{-1} are the reciprocals of those for P. Thus P^{-1} is Hurwitz if and only if P is Hurwitz (see Exercise 4.2.4). Also, the coefficients for P^{-1} are obtained by inverting those for P. Using this facts, one can see that $\boldsymbol{C}(P^{-1})$ is obtained from $\boldsymbol{D}(P)$ by reversing the rows and columns, and that $\boldsymbol{D}(P^{-1})$ is obtained from $\boldsymbol{C}(P)$ by reversing the rows and columns. One can then show that P^{-1} is Hurwitz just as in Case 1, and from this it follows that P is Hurwitz.

Case 3: We assume (iii) and that n is odd. In this case we let

$$\boldsymbol{A}_2(P) = \begin{bmatrix} -\frac{p_{n-2}}{p_n} & -\frac{p_{n-4}}{p_n} & \cdots & -\frac{p_1}{p_n} & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

and note that one can check to see that

$$\boldsymbol{C}(P)\boldsymbol{A}_2(P) = -\begin{bmatrix} \boldsymbol{D}(P) & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix}. \tag{4.12}$$

As in Case 1, we may define the square root, $\boldsymbol{C}(P)^{1/2}$, of $\boldsymbol{C}(P)$, and ascertain that

$$\boldsymbol{C}(P)^{1/2}\boldsymbol{A}_2(P)\boldsymbol{C}(P)^{-1/2} = -\boldsymbol{C}(P)^{-1/2}\begin{bmatrix} \boldsymbol{D}(P) & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix}\boldsymbol{C}(P)^{-1/2}.$$

Again, the conclusion is that $\boldsymbol{A}_2(P)$ is similar to a symmetric matrix, and so must have real eigenvalues. These eigenvalues are the roots of the characteristic polynomial

$$X^{(n+1)/2} + \frac{p_{n-2}}{p_n}X^{(n+1)/2-1} + \cdots + \frac{p_1}{p_n}X.$$

This polynomial clearly has a zero root. However, since (iii) holds, for positive real values of X, the characteristic polynomial takes on positive values, so the nonzero eigenvalues of $\boldsymbol{A}_2(P)$ must be negative, and there are $\frac{n+1}{2} - 1$ of these. From this and (4.12) it follows that the matrix

$$\begin{bmatrix} \boldsymbol{D}(P) & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix}$$

has one zero eigenvalue and $\frac{n+1}{2} - 1$ positive real eigenvalues. Thus $\boldsymbol{D}(P)$ must be positive-definite, and P is then Hurwitz by Lemma 4.2.17 and Theorem 4.2.15.

Case 4: We assume (i) and that n is odd. As in Case 2, define $P^{-1}(X) = X^n P(\frac{1}{X})$. In this case one can ascertain that $\boldsymbol{C}(P^{-1})$ is obtained from $\boldsymbol{C}(P)$ by reversing rows and

columns, and that $\boldsymbol{D}(P^{-1})$ is obtained from $\boldsymbol{D}(P)$ by reversing rows and columns. The difference from the situation in Case 2 arises because here we are taking n odd, while in Case 2 it was even. In any event, one may now apply Case 3 to P^{-1} to show that P^{-1} is Hurwitz. Then P is itself Hurwitz by Exercise 4.2.4.

Case 5: We assume (ii) and that n is odd. For $\epsilon > 0$ define $P_\epsilon \in \mathbb{R}[X]$ by $P_\epsilon(X) = (X + \epsilon)P(X)$. Thus the degree of P_ϵ is now even. Indeed,

$$P_\epsilon(X) = p_n X^{n+1} + (p_{n-1} + \epsilon p_n)X^n + \cdots + (p_0 + \epsilon p_1)X + \epsilon p_0.$$

One may readily determine that

$$\boldsymbol{C}(P_\epsilon) = \boldsymbol{C}(P) + \epsilon \boldsymbol{C}$$

for some matrix $\boldsymbol{C}$ which is independent of ϵ. In like manner, one may show that

$$\boldsymbol{D}(P_\epsilon) = \begin{bmatrix} \boldsymbol{D}(P) + \epsilon \boldsymbol{D}_{11} & \epsilon \boldsymbol{D}_{12} \\ \epsilon \boldsymbol{D}_{12} & \epsilon p_0^2 \end{bmatrix},$$

where $\boldsymbol{D}_{11}$ and $\boldsymbol{D}_{12}$ are independent of ϵ. Since $\boldsymbol{D}(P)$ is positive-definite and $a_0 > 0$, for ϵ sufficiently small we must have that $\boldsymbol{D}(P_\epsilon)$ is positive-definite. From the argument of Case 2, we may infer that P_ϵ is Hurwitz, from which it is obvious that P is also Hurwitz.

Case 6: We assume (iv) and that n is odd. We define $P^{-1}(X) = X^n P(\frac{1}{X})$ so that $\boldsymbol{C}(P^{-1})$ is obtained from $\boldsymbol{C}(P)$ by reversing rows and columns, and that $\boldsymbol{D}(P^{-1})$ is obtained from $\boldsymbol{D}(P)$ by reversing rows and columns. One can now use Case 5 to show that P^{-1} is Hurwitz, and so P is also Hurwitz by Exercise 4.2.4.

Case 7: We assume (iii) and that n is even. As with Case 5, we define $P_\epsilon(X) = (X + \epsilon)P(X)$ and in this case we compute

$$\boldsymbol{C}(P_\epsilon) = \begin{bmatrix} \boldsymbol{C}(P) + \epsilon \boldsymbol{C}_{11} & \epsilon \boldsymbol{C}_{12} \\ \epsilon \boldsymbol{C}_{12} & \epsilon p_0^2 \end{bmatrix}$$

and

$$\boldsymbol{D}(P_\epsilon) = \boldsymbol{D}(P) + \epsilon \boldsymbol{D},$$

where $\boldsymbol{C}_{11}$, $\boldsymbol{C}_{12}$, and $\boldsymbol{D}$ are independent of ϵ. By our assumption (iii), for $\epsilon > 0$ sufficiently small we have $\boldsymbol{C}(P_\epsilon)$ positive-definite. Thus, invoking the argument of Case 1, we may deduce that $\boldsymbol{D}(P_\epsilon)$ is also positive-definite. Therefore P_ϵ is Hurwitz by Lemma 4.2.17 and Theorem 4.2.13. Thus P is itself also Hurwitz.

Case 8: We assume (iv) and that n is even. Taking $P^{-1}(X) = X^n P(\frac{1}{X})$ we see that $\boldsymbol{C}(P^{-1})$ is obtained from $\boldsymbol{D}(P)$ by reversing the rows and columns, and that $\boldsymbol{D}(P^{-1})$ is obtained from $\boldsymbol{C}(P)$ by reversing the rows and columns. Now one may apply Case 7 to deduce that P^{-1}, and therefore P, is Hurwitz. ■

The Liénard–Chipart criterion

Although less well-known than the criterion of Routh and Hurwitz, the test we give next has the advantage of delivering fewer determinantal inequalities to test. This results from their being a dependence on some of the Hurwitz determinants.

4.2.19 Theorem (Liénard–Chipart criterion) *A polynomial*

$$P = X^n + p_{n-1}X^{n-1} + \cdots + p_1X + p_0 \in \mathbb{R}[X]$$

is Hurwitz if and only if any one of the following conditions holds:

(i) $p_{2k} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$ *and* $\Delta_{2k+1} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$;

(ii) $p_{2k} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$ *and* $\Delta_{2k} > 0$, $k \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$;

(iii) $p_0 > 0$, $p_{2k+1} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-2}{2} \rfloor\}$ *and* $\Delta_{2k+1} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$;

(iv) $p_0 > 0$, $p_{2k+1} > 0$, $k \in \{0, 1, \ldots, \lfloor \frac{n-2}{2} \rfloor\}$ *and* $\Delta_{2k} > 0$, $k \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$.

Here $\Delta_1, \ldots, \Delta_n$ *are the Hurwitz determinants.*

Proof The theorem follows immediately from ***missing stuff*** and Theorem 4.2.18, after one checks that the principal minors of $\boldsymbol{C}(P)$ are exactly the odd Hurwitz determinants $\Delta_1, \Delta_3, \ldots$, and that the principal minors of $\boldsymbol{D}(P)$ are exactly the even Hurwitz determinants $\Delta_2, \Delta_4, \ldots$. ■

The advantage of the Liénard–Chipart test over the Hurwitz test is that one will generally have fewer determinants to compute. Let us illustrate the criterion in the simplest case, when $n = 2$.

4.2.20 Example (Liénard–Chipart criterion) We consider the polynomial $P = X^2 + aX + b$. Recall that the Hurwitz determinants were computed in Example 4.2.14:

$$\Delta_1 = a, \quad \Delta_2 = ab.$$

Let us write down the four conditions of Theorem 4.2.19:

1. $p_0 = b > 0$, $\Delta_1 = a > 0$;
2. $p_0 = b > 0$, $\Delta_2 = ab > 0$;
3. $p_0 = b > 0$, $p_1 = a > 0$, $\Delta_1 = a > 0$;
4. $p_0 = b > 0$, $p_1 = a > 0$, $\Delta_2 = ab > 0$.

We see that all of these conditions are equivalent in this case, and imply that P is Hurwitz if and only if $a, b > 0$, as expected. This example is really too simple to illustrate the potential advantages of the Liénard-Chipart criterion, but we refer the reader to Exercise 4.2.5 to see how the test can be put to good use. •

Kharitonov's test

It is sometimes the case that one does not know exactly the coefficients for a given polynomial. In such instances, one may know bounds on the coefficients. That is, for a polynomial

$$P(s) = p_n s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0, \tag{4.13}$$

one may know that the coefficients satisfy inequalities of the form

$$p_i^{\min} \leq p_i \leq p_i^{\max}, \quad i = 0, 1, \ldots, n. \tag{4.14}$$

In this case, the following remarkable theorem gives a simple test for the stability of the polynomial for all possible values for the coefficients.

4.2.21 Theorem (Kharitonov's criterion) *Given a polynomial of the form (4.13) with the coefficients satisfying the inequalities (4.14), define four polynomials*

$$\begin{aligned}
Q_1(s) &= p_0^{\min} + p_1^{\min} s + p_2^{\max} s^2 + p_3^{\max} s^3 + \cdots \\
Q_2(s) &= p_0^{\min} + p_1^{\max} s + p_2^{\max} s^2 + p_3^{\min} s^3 + \cdots \\
Q_3(s) &= p_0^{\max} + p_1^{\max} s + p_2^{\min} s^2 + p_3^{\min} s^3 + \cdots \\
Q_4(s) &= p_0^{\max} + p_1^{\min} s + p_2^{\min} s^2 + p_3^{\max} s^3 + \cdots
\end{aligned}$$

Then P *is Hurwitz for all*

$$(p_0, p_1, \ldots, p_n) \in [p_0^{\min}, p_0^{\max}] \times [p_1^{\min}, p_1^{\max}] \times \cdots \times [p_n^{\min}, p_n^{\max}]$$

if and only if the polynomials Q_1, Q_2, Q_3, *and* Q_4 *are Hurwitz.*

Proof Let us first assume without loss of generality that $p_j^{\min} > 0$, $j = 0, \ldots, n$. Indeed, by Exercise 4.2.3, for a polynomial to be Hurwitz, its coefficients must have the same sign, and we may as well suppose this sign to be positive. If

$$\boldsymbol{p} = (p_0, p_1, \ldots, p_n) \in [p_0^{\min}, p_0^{\min}] \times [p_1^{\min}, p_1^{\min}] \times \cdots \times [p_n^{\min}, p_n^{\min}],$$

then let us say, for convenience, that $\boldsymbol{p}$ is ***allowable***. For $\boldsymbol{p}$ allowable denote

$$P_{\boldsymbol{p}}(s) = p_n s^n + p_{n-1} s^{n-1} + \cdots + p_1 s + p_0.$$

It is clear that if all polynomials $P_{\boldsymbol{p}}$ are allowable then the polynomials Q_1, Q_2, Q_3, and Q_4 are Hurwitz. Thus suppose for the remainder of the proof that Q_1, Q_2, Q_3, and Q_4 are Hurwitz, and we shall deduce that $P_{\boldsymbol{p}}$ is also Hurwitz for every allowable $\boldsymbol{p}$.

For $\omega \in \mathbb{R}$ define

$$R(\omega) = \{P_{\boldsymbol{p}}(i\omega) \mid \boldsymbol{p} \text{ allowable}\}.$$

The following property of $R(\omega)$ lies at the heart of our proof. It is first noticed by Dasgupta [1988].

1 Lemma *For each* $\omega \in \mathbb{R}$, $R(\omega)$ *is a rectangle in* $\mathbb{C}$ *whose sides are parallel to the real and imaginary axes, and whose corners are* $Q_1(i\omega)$, $Q_2(i\omega)$, $Q_3(i\omega)$, *and* $Q_4(i\omega)$.

Proof We note that for $\omega \in \mathbb{R}$ we have

$$\begin{aligned}
\operatorname{Re}(Q_1(i\omega)) &= \operatorname{Re}(Q_2(i\omega)) = p_0^{\min} - p^{\max}\omega^2 + p_4^{\min}\omega^4 + \cdots \\
\operatorname{Re}(Q_3(i\omega)) &= \operatorname{Re}(Q_4(i\omega)) = p_0^{\max} - p^{\min}\omega^2 + p_4^{\max}\omega^4 + \cdots \\
\operatorname{Im}(Q_1(i\omega)) &= \operatorname{Im}(Q_4(i\omega)) = \omega\left(p^{\min} - p^{\max}\omega^2 + p_4^{\min}\omega^4 + \cdots\right) \\
\operatorname{Im}(Q_2(i\omega)) &= \operatorname{Im}(Q_3(i\omega)) = \omega\left(p^{\max} - p^{\min}\omega^2 + p_4^{\max}\omega^4 + \cdots\right).
\end{aligned}$$

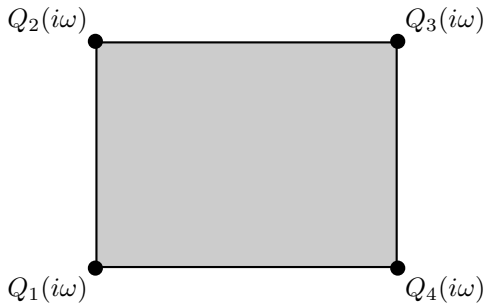


Figure 4.8 $R(\omega)$

From this we deduce that for any allowable $\boldsymbol{p}$ we have

$$\begin{aligned}\mathrm{Re}(Q_1(i\omega)) = \mathrm{Re}(Q_2(i\omega)) \le \mathrm{Re}(P_{\boldsymbol{p}}(i\omega)) \le \mathrm{Re}(Q_3(i\omega)) = \mathrm{Re}(Q_4(i\omega))\\ \mathrm{Im}(Q_1(i\omega)) = \mathrm{Im}(Q_4(i\omega)) \le \mathrm{Im}(P_{\boldsymbol{p}}(i\omega)) \le \mathrm{Im}(Q_2(i\omega)) = \mathrm{Im}(Q_3(i\omega)).\end{aligned}$$

This leads to the picture shown in Figure 4.8 for $R(\omega)$. The lemma follows immediately from this. ▼

Using the lemma, we now claim that if $\boldsymbol{p}$ is allowable, then $P_{\boldsymbol{p}}$ has no imaginary axis roots. To do this, we record the following useful property of Hurwitz polynomials.

2 Lemma *If* $\mathrm{P} \in \mathbb{R}[\mathrm{s}]$ *is monic and Hurwitz with* $\deg(\mathrm{P}) \ge 1$, *then* $\arg \mathrm{P}(\mathrm{i}\omega)$ *is a continuous and strictly increasing function of* ω.

Proof Write

$$P(s) = \prod_{j=1}^{n}(s - z_j)$$

where $z_j = \sigma_j + i\omega_j$ with $\sigma_j < 0$. Thus

$$\arg P(i\omega) = \sum_{j=1}^{n} \arg\left(i\omega + |\sigma_j| - i\omega_j\right) = \sum_{j=1}^{n} \arctan\Big(\frac{\omega - \omega_j}{|\sigma_j|}\Big).$$

Since $|\sigma_j| > 0$, each term in the sum is continuous and strictly increasing, and thus so too is $\arg P(i\omega)$. ▼

To show that $0 \notin R(\omega)$ for $\omega \in \mathbb{R}$, first note that $0 \notin R(0)$. Now, since the corners of $R(\omega)$ are continuous functions of ω, if $0 \in R(\omega)$ for some $\omega > 0$, then it must be the case that for some $\omega_0 \in [0, \omega]$ the point $0 \in \mathbb{C}$ lies on the boundary of $R(\omega_0)$. Suppose that 0 lies on the lower boundary of the rectangle $R(\omega_0)$. This means that $Q_1(i\omega_0) < 0$ and $Q_4(i\omega_0) > 0$ since the corners of $R(\omega)$ cannot pass through 0. Since Q_1 is Hurwitz, by Lemma 2 we must have $Q_1(i(\omega_0 + \delta))$ in the $(-,-)$ quadrant in $\mathbb{C}$ and $Q_4(i(\omega_0 + \delta))$ in the $(+,+)$ quadrant in $\mathbb{C}$ for $\delta > 0$ sufficiently small. However,

since $\text{Im}(Q_1(i\omega)) = \text{Im}(Q_4(i\omega))$ for all $\omega \in \mathbb{R}$, this cannot be. Therefore 0 cannot lie on the lower boundary of $R(\omega_0)$ for any $\omega_0 > 0$. Similar arguments establish that 0 cannot lie on either of the other three boundaries either. This then prohibits 0 from lying in $R(\omega)$ for any $\omega > 0$.

Now suppose that $P_{\boldsymbol{p}_0}$ is not Hurwitz for some allowable $\boldsymbol{p}_0$. For $\lambda \in [0, 1]$ each of the polynomials

$$\lambda Q_1 + (1 - \lambda)P_{\boldsymbol{p}_0} \tag{4.15}$$

is of the form $P_{\boldsymbol{p}_\lambda}$ for some allowable $\boldsymbol{p}_\lambda$. Indeed, the equation (4.15) defines a straight line from Q_1 to $P_{\boldsymbol{p}_0}$, and since the set of allowable $\boldsymbol{p}$'s is convex (it is a cube), this line remains in the set of allowable polynomial coefficients. Now, since Q_1 is Hurwitz and $P_{\boldsymbol{p}_0}$ is not, by continuity of the roots of a polynomial with respect to the coefficients, we deduce that for some $\lambda \in [0, 1)$, the polynomial $P_{\boldsymbol{p}_\lambda}$ must have an imaginary axis root. However, we showed above that $0 \notin R(\omega)$ for all $\omega \in \mathbb{R}$, denying the possibility of such imaginary axis roots. Thus all polynomials $P_{\boldsymbol{p}}$ are Hurwitz for allowable $\boldsymbol{p}$. ■

4.2.22 Remarks

1. Note the pattern of the coefficients in the polynomials Q_1, Q_2, Q_3, and Q_4 has the form $(\dots, \max, \max, \min, \min, \dots)$ This is charmingly referred to as the ***Kharitonov melody***.
2. One would anticipate that to check the stability for P one should look at all possible extremes for the coefficients, giving 2^n polynomials to check. That this can be reduced to four polynomial checks is an unobvious simplification. •

Let us apply the Kharitonov test in the simplest case when $n = 2$.

4.2.23 Example We consider

$$P(s) = s^2 + as + b$$

with the coefficients satisfying

$$(a, b) \in [a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}].$$

The polynomials required by Theorem 4.2.21 are

$$\begin{aligned} Q_1(s) &= s^2 + a_{\min}s + b_{\min} \\ Q_2(s) &= s^2 + a_{\max}s + b_{\min} \\ Q_3(s) &= s^2 + a_{\max}s + b_{\max} \\ Q_4(s) &= s^2 + a_{\min}s + b_{\max}. \end{aligned}$$

We now apply the Routh/Hurwitz criterion to each of these polynomials. This indicates that all coefficients of the four polynomials Q_1, Q_2, Q_3, and Q_4 should be positive. This reduces to requiring that

$$a_{\min}, a_{\max}, b_{\min}, b_{\max} > 0.$$

That is, $a_{\min}, b_{\min} > 0$. In this simple case, we could have guessed the result ourselves since the Routh/Hurwitz criterion are so simple to apply for degree two polynomials. Nonetheless, the simple example illustrates how to apply Theorem 4.2.21. •

Notes

It is interesting to note that the method of Edward John Routh (1831–1907) was developed in response to a famous paper of James Clerk Maxwell[6] (1831–1879) on the use of governors to control a steam engine. This paper of Maxwell [1868] can be regarded as the first paper in mathematical control theory.

Theorem 4.2.11 is due to Routh [1877].

Theorem 4.2.13 is due to Hurwitz [1895].

Theorem 4.2.15 is due to Charles Hermite (1822–1901) [see Hermite 1854]. The slick proof using Lyapunov methods comes from the paper of Parks [1962].

Our proof of Theorem 4.2.18 follows that of Anderson [1972].

Theorem 4.2.19 is from Liénard and Chipart [1914][7] This is given thorough discussion by Gantmacher [1959]. Here we state the result, and give a proof due to Anderson [1972] that is more elementary than that of Gantmacher. The observation in the proof of Theorem 4.2.19 is made by a computation which we omit, and appears to be first been noticed by Fujiwara [1915].

Theorem 4.2.21 is due to Kharitonov [1978]. Since the publication of Kharitonov's result, or more properly its discovery by the non-Russian speaking world, there have been many simplifications of the proof [e.g., Chapellat and Bhattacharyya 1989, Dasgupta 1988, Mansour and Anderson 1993]. The proof we give essentially follows Minnichelli, Anagnost, and Desoer [1989]. Anderson, Jury, and Mansour [1987] observe that for polynomials of degree 3, 4, or 5, it suffices to check not four, but one, two, or three polynomials, respectively, as being Hurwitz. A proof of Kharitonov's theorem, using Lyapunov methods (see Section 4.3.6), is given by Mansour and Anderson [1993].

Reference on operator norm.

Exercises

4.2.1

In the next exercise we shall make use of a norm $|||\cdot|||$ on the set $\mathrm{L}(\mathsf{V};\mathsf{V})$ of linear transformations induced by a norm $\|\cdot\|$ on V. The norm is defined by

$$|||L||| = \sup\left\{ \frac{\|L(v)\|}{\|v\|} \;\middle|\; v \in \mathsf{V} \setminus \{0\} \right\},$$

[6] Maxwell, of course, is better known for his famous equations of electromagnetism.

[7] Perhaps the relative obscurity of the test reflects that of its authors; I was unable to find a biographical reference for either Liénard or Chipart. I do know that Liénard did work in differential equations, with the ***Liénard equation*** being a well-studied second-order linear differential equation.

for $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$. It is easy to show that this is, in fact, a norm and we refer the reader to the references for this.

4.2.2 Let F be a system of linear homogeneous ordinary differential equations in an n-dimensional $\mathbb{R}$-vector space V and with right-hand side $\widehat{F}(t,x) = A(t)(x)$ for a continuous map $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$. Suppose that $\sup \mathbb{T} = \infty$.

(a) Show that F is stable if and only if, for every $t_0 \in \mathbb{T}$, there exists $C \in \mathbb{R}_{>0}$ such that $|||\Phi_A(t,t_0)||| \leq C$ for $t \geq t_0$.

(b) Show that F is asymptotically stable if and only if, for every $t_0 \in \mathbb{T}$ and $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that $|||\Phi_A(t,t_0)||| < \epsilon$ for $t \geq t_0 + T$.

(c) Show that F is exponentially stable if and only if, for every $t_0 \in \mathbb{T}$, there exist $M, \sigma \in \mathbb{R}_{>0}$ such that $|||\Phi_A(t,t_0)||| \leq Me^{-\sigma(t-t_0)}$ for $t \geq t_0$.

(d) Show that F is uniformly stable if and only if there exists $C \in \mathbb{R}_{>0}$ such that, for every $t_0 \in \mathbb{T}$, $|||\Phi_A(t,t_0)||| \leq C$ for $t \geq t_0$.

(e) Show that F is uniformly asymptotically stable if and only if,

1. there exists $C \in \mathbb{R}_{>0}$ such that, for every $t_0 \in \mathbb{T}$, $|||\Phi_A(t,t_0)||| \leq C$ for $t \geq t_0$ and
2. for every $\epsilon \in \mathbb{R}_{>0}$, there exists $T \in \mathbb{R}_{>0}$ such that, for every $t_0 \in \mathbb{T}$, $|||\Phi_A(t,t_0)||| < \epsilon$ for $t \geq t_0 + T$.

(f) Show that F is exponentially stable if and only if there exist $M, \sigma \in \mathbb{R}_{>0}$ such that, for every $t_0 \in \mathbb{T}$, $|||\Phi_A(t,t_0)||| \leq Me^{-\sigma(t-t_0)}$ for $t \geq t_0$.

4.2.3 A useful necessary condition for a polynomial to have all roots in $\mathbb{C}_-$ is given by the following theorem.

Theorem *If the polynomial*

$$P = X^n + p_{n-1}X^{n-1} + \cdots + p_1X + p_0 \in \mathbb{R}[X]$$

is Hurwitz, then the coefficients $p_0, p_1, \ldots, p_{n-1}$ *are all positive.*

(a) Prove this theorem.

(b) Is the converse of the theorem true? If so, prove it, if not, give a counterexample.

4.2.4 Consider a polynomial

$$P = p_nX^n + p_{n-1}X^{n-1} + \cdots + p_1X + p_0 \in \mathbb{R}[X]$$

with $p_0, p_n \neq 0$, and define $P^{-1} \in \mathbb{R}[X]$ by $P^{-1}(X) = X^nP(\frac{1}{X})$.

(a) Show that the roots for P^{-1} are the reciprocals of the roots for P.

(b) Show that P is Hurwitz if and only if P^{-1} is Hurwitz.

4.2.5 For the following two polynomials,

(a) $P = X^3 + aX^2 + bX + c$,

(b) $P = X^4 + aX^3 + bX^2 + cX + d$,

write down the four conditions of the Liénard–Chipart criterion, Theorem 4.2.19, and determine which is the least restrictive.

Section 4.3

Lyapunov's Second Method

Much of the basic stability theory used in practice originates with the work of Aleksandr Mikhailovich Lyapunov (1857–1918). In this section and the next we shall cover what are commonly called "Lyapunov's First Method" (also "Lyapunov's Indirect Method") and "Lyapunov's Second Method" (also "Lyapunov's Direct Method"). The First Method is a useful one in that it allows one to deduce stability from the linearisation, and often the stability of the linearisation can be determined by computing a polynomial (Section 4.2.2.1) and performing computations with its coefficients (Section 4.2.2.3). The Second Method, on the other hand, involves hypothesising a function—called a "Lyapunov function"—with certain properties. In practice and in general, it is to be regarded as impossible to find a Lyapunov function. However, the true utility of the Second Method is that, once one has a Lyapunov function, there is a great deal one can say about the differential equation. However, such matters lie beyond the scope of the present text, and we refer to the references for further discussion.

It goes without saying that we shall discuss the Second Method first. Lyapunov's Second Method, or Direct Method, is a little. . . er. . . indirect, since it has to do with considering functions with certain properties. We shall consider in the text four settings for Lyapunov's Second Method. We shall treat each of the four cases in a self-contained manner, so a reader does not have to understand the (somewhat complicated) most general setting in order to understand the (less complicated) less general settings. Therefore, let us provide a roadmap for these cases.

4.3.1 Road map for Lyapunov's Second Method We list the four settings for Lyapunov's Second Method, and what should be read to comprehend them, together or separately.

1. *General nonautonomous equations.* The most general setting is that of equations that are nonautonomous, i.e., time-varying, and not necessarily linear. Here one needs to carefully discriminate between uniform and nonuniform stability notions. The material required to access the result on these equations is:
   (a) class $\mathcal{K}$- and class $\mathcal{KL}$-functions in Section 4.3.1.1;
   (b) time-invariant definite and semidefinite functions in Section 4.3.1.2;
   (c) time-varying definite and semidefinite functions in Section 4.3.1.3;
   (d) characterisations of stability using class $\mathcal{K}$- and class $\mathcal{KL}$-functions in Section 4.3.2;
   (e) the results on Lyapunov's Second Method in Section 4.3.3;
   (f) the theorems of Sections 4.3.4, 4.3.5, and 4.3.6 are corollaries of the more general theorems, although we also give independent proofs.

2. *General autonomous equations.* Here we consider autonomous ordinary differential that are not necessarily linear. The simplifications assumed by not having to discriminate between uniform and nonuniform stability make the results here significantly simpler than those for nonautonomous equations. The material needed to understand the results in this case is:
   (a) understand Definition 4.3.6;
   (b) the results on Lyapunov's Second Method in Section 4.3.4;
   (c) the theorems of Section 4.3.6 are corollaries of the more general theorems, although we also give independent proofs. •
3. *Time-varying linear equations.* The next class of equations one can consider are linear homogeneous time-varying ordinary differential equations. Note that it is necessary to understand the results on Lyapunov's Second Method here in order to prove the results on Lyapunov's First Method for nonautonomous equations. In order to understand this material, the following material needs to be read:
   (a) time-invariant quadratic functions in Section 4.3.1.4;
   (b) time-varying quadratic functions in Section 4.3.1.5;
   (c) the results on Lyapunov's Second Method in Section 4.3.5.
4. *Time-invariant linear equations.* Our final setting concerns linear homogeneous time-invariant ordinary differential equations. Note that these results are required to understand the results on Lyapunov's First Method for autonomous equations. In this setting, one needs to read the following material:
   (a) time-invariant quadratic functions in Section 4.3.1.4;
   (b) the result on Lyapunov's Second Method in Section 4.3.6;
   (c) the theorems of Section 4.3.6 are corollaries of the more general theorems, although we also give independent proofs. •

The first thing we do is discuss the various classes of functions that appear in Lyapunov's Second Method.

4.3.1 Positive-definite and decrescent functions

We will be considering functions that, intuitively, have the equilibrium point x_0 as a maximum and whose derivative along solutions is nonincreasing. It is these notions of "maximum" and "nonincreasing" that we are concerned with here. It turns out that there is a great deal to say about these seemingly simple subjects.

4.3.1.1 Class $\mathcal{K}$-, class $\mathcal{L}$-, and class $\mathcal{KL}$-functions It is convenient for many of our characterisations and for many of our proofs concerning Lyapunov's Second Method to have at hand two classes of scalar functions of a real variable, which leads to another class of scalar functions of two real variables.

4.3.2 Definition (Class $\mathcal{K}$, class $\mathcal{L}$, and class $\mathcal{KL}$) Let $a \in \mathbb{R}$ and $b, b' \in \mathbb{R}_{>0} \cup \{\infty\}$.

(i) A function $\phi\colon [0,b) \to \mathbb{R}_{\geq 0}$ is of ***class*** $\mathcal{K}$ if
  (a) ϕ is continuous,
  (b) ϕ is strictly increasing, i.e., $\phi(x) < \phi(y)$ if $x < y$, and
  (c) $\phi(0) = 0$.

  By $\mathcal{K}([0,b);[0,b'))$ we denote the set of functions of class $\mathcal{K}$ with domain $[0,b)$ and codomain $[0,b')$.

(ii) A function $\psi\colon [a,\infty) \to \mathbb{R}_{\geq 0}$ is of ***class*** $\mathcal{L}$ if
  (a) ψ is continuous,
  (b) ψ is strictly decreasing, i.e., $\psi(x) > \psi(y)$ is $x < y$, and
  (c) $\lim_{x\to\infty} \psi(x) = \infty$.

  By $\mathcal{L}([a,\infty);[0,b'))$ we denote the set of functions of class $\mathcal{L}$ with domain $[a,\infty)$ and codomain $[0,b')$.

(iii) A function $\psi\colon [0,b) \times [a,\infty) \to \mathbb{R}_{\geq 0}$ is of ***class*** $\mathcal{KL}$ if
  (a) $x \mapsto \psi(x,y)$ is of class $\mathcal{K}$ for each $y \in [a,\infty)$ and
  (b) $y \mapsto \psi(x,y)$ is of class $\mathcal{L}$ for each $x \in [0,b)$.

  By $\mathcal{KL}([0,b) \times [a,\infty);[0,b'))$ we denote the set of functions of class $\mathcal{KL}$ with domain $[0,b) \times [a,\infty)$ and codomain $[0,b')$. •

These sorts of functions are often collectively referred to as "comparison functions."

Let $\phi \in \mathcal{K}([0,b);\mathbb{R}_{\geq 0})$. Since ϕ is strictly increasing, the limit $\lim_{x\to b} \phi(b)$ exists, allowing that the limit mat be ∞. For this reason, we can unambiguously write $\phi(b)$, although b is not in the domain of ϕ.

In Exercises 4.3.1, 4.3.3, and 4.3.4 the reader can sort through some examples of functions that are or are not in these classes. Here we shall enumerate a few useful properties of such functions.

4.3.3 Lemma (Properties of class $\mathcal{K}$-, class $\mathcal{L}$-, and class $\mathcal{KL}$-functions) *Let* $b, b' \in \mathbb{R}_{>0} \cup \{\infty\}$ *and* $a \in \mathbb{R}$. *Then the following statements hold:*

(i) if $\phi \in \mathcal{K}([0,b);\mathbb{R}_{\geq 0}$, *then* $\phi^{-1} \in \mathcal{K}([0,\phi(b));\mathbb{R}_{\geq 0})$ *is well-defined and is of class* $\mathcal{K}$;

(ii) if $\phi_1 \in \mathcal{K}([0,b);[0,b'))$ *and* $\phi_2 \in \mathcal{K}([0,b');\mathbb{R}_{\geq 0})$, *then* $\phi_2 \circ \phi_1$ *is of class* $\mathcal{K}$;

(iii) if $\phi_1\colon [0,b) \to [0,b')$ *and* $\phi_2\colon [0,b') \to \mathbb{R}_{\geq 0}$ *are of class* $\mathcal{K}$, *and if* $\psi\colon [0,b) \times [a,\infty) \to [0,b')$ *is of class* $\mathcal{KL}$, *then the function*

$$[0,b) \times [a,\infty) \ni (x,y) \mapsto \phi_2(\psi(\phi_1(x),y)) \in \mathbb{R}_{\geq 0}$$

is of class $\mathcal{KL}$.

Proof These are all just a matter of working through definitions, and we leave this to the reader as Exercise 4.3.2. ■

One often encounters functions that are "almost" of class $\mathcal{K}$, and in this case it is sometimes possible to bound them from below by a class $\mathcal{K}$-function.

4.3.4 Lemma (Bounding nondecreasing functions by strictly increasing functions) *Let* $\mathrm{b} \in \mathbb{R}_{>0} \cup \{\infty\}$ *and let* $\mathrm{f}\colon [0, \mathrm{b}) \to \mathbb{R}_{\geq 0}$ *have the following properties:*

(i) f *is continuous;*

(ii) f *is nondecreasing, i.e.,* $\mathrm{f}(\mathrm{x}_1) \leq \mathrm{f}(\mathrm{x}_2)$ *for* $\mathrm{x}_1 < \mathrm{x}_2$*;*

(iii) $\mathrm{f}(\mathrm{x}) \in \mathbb{R}_{>0}$ *for* $\mathrm{x} \in (0, \mathrm{b})$*;*

(iv) $\mathrm{f}(0) = 0$.

Then there exist $\phi_1, \phi_2 \in \mathcal{K}([0, \mathrm{b}); \mathbb{R}_{\geq 0})$ *such that* $\phi_1(\mathrm{x}) \leq \mathrm{f}(\mathrm{x}) \leq \phi_2(\mathrm{x})$ *for* $\mathrm{x} \in [0, \mathrm{b})$. *Moreover,* ϕ_1 *can be chosen to be locally Lipschitz.*

Proof Let $(x_j)_{j\in\mathbb{Z}}$ be the strictly increasing doubly infinite sequence in $(0, b)$ given by

$$x_j = \begin{cases} \frac{b}{2} 2^j, & j \leq 0, \\ b(1 - 2^{-j-1}), & j > 0, \end{cases}$$

noting that $\lim_{j\to-\infty} x_j = 0$ and $\lim_{j\to\infty} x_j = b$. Define a doubly infinite sequence $(\alpha_j)_{j\in\mathbb{Z}}$ by

$$\alpha_j = \begin{cases} 2^{j-1}, & j \leq 0, \\ 1 - 2^{-j-1}, & j > 0. \end{cases}$$

Note that both sequences are strictly increasing and that

$$\lim_{j\to-\infty} \alpha_j = 0, \quad \lim_{j\to\infty} \alpha_j = 1.$$

Let $N_1 \in \mathbb{Z}_{>0}$ be sufficiently large that $x_{j+1} - x_j < 1$ for $j \leq -N_1$. This is possible since $(x_{-j})_{j\in\mathbb{Z}_{>0}}$ converges to 0, and so is Cauchy. Let $N_2 \in \mathbb{Z}_{\geq 0}$ be the smallest positive integer such that

$$f(x_j) - f(x_{j-1}) < 1, \qquad j \leq -N_2.$$

This is possible since $(f(x_{-j}))_{j\in\mathbb{Z}_{>0}}$ converges to 0 (by continuity of f) and so is Cauchy. Let $N = \max\{N_1, N_2\}$. Now define

$$\phi_{1,j} = \begin{cases} (x_{j+1} - x_j)\alpha_j f(x_j), & |j| \geq N, \\ \alpha_j f(x_j), & |j| < N, \end{cases}$$

and

$$\phi_{2,j} = (1 + \alpha_j) f(x_j), \qquad j \in \mathbb{Z}.$$

Here are the key observations about the doubly infinite sequences $(\phi_{1,j})_{j\in\mathbb{Z}}$ and $(\phi_{2,j})_{j\in\mathbb{Z}}$.

1. We have $\phi_{1,j-1} < \phi_{1,j}$ for $j \in \mathbb{Z}$. This follows because
   (a) $x_j - x_{j-1} < x_{j+1} - x_j < 1$ for $j \leq -N$,
   (b) $\alpha_j < \alpha_{j-1}$ for $j \in \mathbb{Z}$, and
   (c) $f(x_{j-1}) \leq f(x_j)$ for $j \leq -N$.

2. $f(x) \geq \phi_{j,1}$ for $x \in [x_j, x_{j+1})$ and $j \in \mathbb{Z}$. This follows because
   (a) $x_{j+1} - x_j < 1$ for $j \leq -N$,
   (b) $\alpha_j < 1$ for $j \in \mathbb{Z}$, and
   (c) $f(x) \geq f(x_j)$ for $x \in [x_j, x_{j+1})$.
3. $\phi_{2,j} < \phi_{2,j}$ for $j \in \mathbb{Z}$. This follows because
   (a) $1 + \alpha_j < 1 + \alpha_{j+1}$ for $j \in \mathbb{Z}$ and
   (b) $f(x_j) \leq f(x_{j+1})$ for $j \in \mathbb{Z}$.
4. $f(x) \leq \phi_{2,j}$ for $x \in [x_{j-1}, x_j)$ and $j \in \mathbb{Z}$. This follows because
   (a) $1 + \alpha_j > 1$ for $j \in \mathbb{Z}$ and
   (b) $f(x) \leq f_j(x)$ for $x \in [x_{j-1}, x_j)$ and $j \in \mathbb{Z}$.

Now define

$$\phi_1(x) = \begin{cases} 0, & x = 0, \\ \phi_{1,j-1} + \frac{x - x_j}{x_{j+1} - x_j}(\phi_{1,j} - \phi_{1,j-1}), & x \in [x_j, x_{j+1}), \end{cases}$$

and

$$\phi_2(x) = \begin{cases} 0, & x = 0, \\ \phi_{2,j} + \frac{\phi_{2,j+1} - \phi_{2,j}}{x_j - x_{j-1}}(x - x_{j-1}), & x \in [x_{j-1}, x_j). \end{cases}$$

One can then directly verify that $\phi_1, \phi_2 \in \mathcal{K}([0, b); \mathbb{R}_{\geq 0})$ and that

$$\phi_2(x) \leq f(x) \leq \phi_2(x)$$

for all $x \in [0, b)$.

Let us now show that ϕ_1 is locally Lipschitz. Note that both ϕ_1 and ϕ_2 are piecewise linear on $(0, b)$, which means they are locally Lipschitz on $(0, b)$. In order to show that ϕ_1 can be chosen to be locally Lipschitz on $[0, b)$, we show that the slopes of the linear segments comprising ϕ_1 are bounded as we approach 0. The set of such slopes is

$$\left\{ \frac{\phi_{1,j} - \phi_{1,j-1}}{x_{j+1} - x_j} \;\middle|\; j \in \mathbb{Z} \right\},$$

and we will verify that

$$\limsup_{j \to -\infty} \frac{\phi_{1,j} - \phi_{1,j-1}}{x_{j+1} - x_j} < \infty.$$

We first note that all of these slopes are positive, as can be seen from the properties of $\phi_{1,j}$, $j \in \mathbb{Z}$. By definition of N, if $j \leq -N$,

$$\begin{aligned} \frac{\phi_{1,j} - \phi_{1,j-1}}{x_{j+1} - x_j} &= (1 - \alpha_j) f(x_j), 1 - (1 - \alpha_{j-1}) f(x_{j-1}) \\ &\leq (1 - \alpha_j) f(x_j) \leq (1 - \alpha_N) f(x_N). \end{aligned}$$

Therefore,

$$\limsup_{j\to-\infty} \frac{\phi_{1,j} - \phi_{1,j-1}}{x_{j+1} - x_j} < \infty,$$

as claimed. Now let $x', x'' \in [0, b)$ satisfy $x' < x''$ and let $N \in \mathbb{Z}$ be such that $[x', x''] \subseteq [0, x_N)$. Letting

$$M = \sup\left\{ \frac{\phi_{1,j} - \phi_{1,j-1}}{x_{j+1} - x_j} \;\middle|\; j \leq N \right\},$$

we have

$$|\phi_1(x_1) - \phi_1(x_2)| \leq M|x_1 - x_2|,$$

which gives the desired conclusion. ■

A useful relationship between functions of class $\mathcal{K}$ and class $\mathcal{KL}$ is given by the following lemma.

4.3.5 Lemma (Solutions of differential equations with class $\mathcal{K}$ right-hand side) *Let $\phi \in \mathcal{K}([0, b); \mathbb{R}_{\geq 0})$ be locally Lipschitz. Then there exists $\psi \in \mathcal{KL}([0, b) \times \mathbb{R}_{\geq 0}; \mathbb{R}_{\geq 0})$ such that, if $x \in [0, b)$ and $t_0 \in \mathbb{R}$, then the solution to the initial value problem*

$$\dot{\xi}(t) = -\phi(\xi(t)), \quad \xi(t_0) = x,$$

is $\psi(x, t - t_0)$ for $t \geq t_0$.

Proof Using the method of Section 2.1, for $x \in (0, b)$ and for $t_0 \in \mathbb{R}$, the solution to the initial value problem

$$\dot{\xi}(t) = \phi(\xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\int_{t_0}^{t} d\tau = -\int_{x}^{\xi(t)} \frac{dy}{\phi(y)}.$$

To encode the dependence of this solution on the initial data, we shall denote it by $\xi_{t_0,x}$. Let us fix $x_0 \in (0, b)$ and define

$$\begin{aligned} \alpha\colon [0, b) &\to \mathbb{R} \\ x &\mapsto -\int_{x_0}^{x} \frac{dy}{\phi(y),} \end{aligned}$$

and note that α has the following properties.

1. *α is continuously differentiable:* This is due to the Fundamental Theorem of Calculus.
2. *α is strictly decreasing:* This is because ϕ is positive on $(0, b)$.

3. $\lim_{x\to 0}\alpha(x) = \infty$*:* Here we note that $\alpha(\xi_{0,x_0}(t)) = t$. Because ϕ is positive on $(0, b)$, it follows that $\lim_{t\to\infty}\xi_{0,x_0}(t) = 0$. Moreover, again since ϕ is positive on $(0, b)$, we cannot have $\xi_{0,x_0}(t) = 0$ for any finite t. Thus we have

$$\lim_{x\to 0}\alpha(x) = \lim_{t\to\infty}\alpha(\xi_{0,x_0}(t)) = \lim_{t\to\infty} t = \infty,$$

as asserted.

Now let $c = -\lim_{x\to b}\alpha(x)$, allowing that $c = \infty$. Thus $\text{image}(\alpha) = (-c, \infty)$ and, since α is strictly decreasing, we have a well-defined map $\alpha^{-1}\colon (-c, \infty) \to (0, b)$. Since

$$\alpha(\xi_{t_0,x}(t)) - \alpha(x) = t - t_0,$$

we have

$$\xi_{t_0,x}(t) = \alpha^{-1}(\alpha(x) + t - t_0).$$

Then define

$$\psi(x, s) = \begin{cases} \alpha^{-1}(\alpha(x) + s), & x \in (0, b), \\ 0, & x = 0. \end{cases}$$

It is clear that ψ is continuous on $(0, b) \times \mathbb{R}_{>0}$. Moreover, since

$$\lim_{(x,s)\to(0,0)}\psi(x, s) = \lim_{(x,s)\to(0,0)}\alpha^{-1}(\alpha(x) + s) = \lim_{(x,s)\to(0,0)}\xi_{0,x}(s) = 0,$$

we conclude continuity of ψ on its domain, and so we have continuity is each argument. Because $\xi_{0,x}(s) = \psi(x, s)$, we have

$$\frac{\partial\psi}{\partial s}(x, s) = \dot{\xi}_{0,x}(s) = -\phi(\xi_{0,x}(s)) = -\phi(\psi(x, s)) < 0$$

for $s \in \mathbb{R}_{>0}$, and so ψ is strictly decreasing in its second argument. It is also strictly increasing in its first argument since

$$\begin{aligned} \frac{\partial\psi}{\partial x}(x, s) &= \frac{\partial\alpha^{-1}}{\partial y}(\alpha(x) + s)\frac{\partial\alpha}{\partial y}(x) \\ &= \left(\frac{\partial\alpha}{\partial y}(\alpha^{-1}(\alpha(x) + s))\right)^{-1}\frac{\partial\alpha}{\partial y}(x) \\ &= \frac{\phi(\psi(x, s))}{\phi(x)} > 0, \end{aligned}$$

using the Inverse Function Theorem (***missing stuff***). Finally,

$$\lim_{s\to\infty}\psi(x, s) = \lim_{t\to\infty}\xi(t) = 0,$$

and we have verified that $\psi \in \mathcal{KL}([0, b) \times \mathbb{R}_{\geq 0}; \mathbb{R}_{\geq 0})$ ■

4.3.1.2 General time-invariant functions Now we give some definitions that, while simple, are not as simple as they seem.

4.3.6 Definition (Locally definite, locally semidefinite, decrescent I) Let $U \subseteq \mathbb{R}^n$ be an open set and let $x_0 \in U$. A function $f\colon U \to \mathbb{R}$ is:

(i) ***locally positive-definite*** about x_0 if
  (a) it is continuous,
  (b) $f(x_0) = 0$,
  (c) there exists $r \in \mathbb{R}_{>0}$ such that $f(x) \in \mathbb{R}_{>0}$ for $x \in \mathsf{B}(r, x_0) \setminus \{x_0\}$;

(ii) ***locally positive-semi definite*** about x_0 if
  (a) it is continuous,
  (b) $f(x_0) = 0$,
  (c) there exists $r \in \mathbb{R}_{>0}$ such that $f(x) \in \mathbb{R}_{\geq 0}$ for $x \in \mathsf{B}(r, x_0) \setminus \{x_0\}$;

(iii) ***locally negative-definite about*** $\mathbf{x_0}$ if $-f$ is positive-definite about x_0;

(iv) ***locally negative-semidefinite about*** $\mathbf{x_0}$ if $-f$ is positive-semidefinite about x_0;

(v) ***locally decrescent about*** $\mathbf{x_0}$ if there exists a locally positive-definite function $g\colon U \to \mathbb{R}$ around x_0 and $r \in \mathbb{R}_{>0}$ such that $f(x) \leq g(x)$ for every $x \in \mathsf{B}(r, x_0)$. •

If $f\colon U \to \mathbb{R}$ is locally positive-definite (resp. locally positive-semidefinite) about x_0 and if $r \in \mathbb{R}_{>0}$ is such that $f(x) \in \mathbb{R}_{>0}$ for $x \in \mathsf{B}(r, x_0)$, we shall say that f is ***locally positive-semidefinite about*** $\mathbf{x_0}$ ***in*** $\mathbf{B(r, x_0)}$ (resp. ***locally positive-semidefinite about*** $\mathbf{x_0}$ ***in*** $\mathbf{B(r, x_0)}$). Similar terminology applies, of course, for functions that are locally negative-definite or locally negative-semidefinite. In like manner, if f is locally decrescent about x_0, and if $r \in \mathbb{R}_{>0}$ and g, locally positive-definite about x_0 in $\mathsf{B}(r, x_0)$, are such that $f(x) \leq g(x)$ for $x \in \mathsf{B}(r, x_0)$, then we say that f is ***locally decrescent about*** $\mathbf{x_0}$ ***in*** $\mathbf{B(r, x_0)}$.

We introduce the following notation:

$\mathrm{LPD}_r(x_0)$ set of locally positive-definite functions about x_0 in $\mathsf{B}(r, x)$;
$\mathrm{LPSD}_r(x_0)$ set of locally positive-semidefinite functions about x_0 in $\mathsf{B}(r, x_0)$;
$\mathrm{LD}_r(x_0)$ set of locally decrescent functions about x_0 in $\mathsf{B}(r, x_0)$

and we also denote

$$\mathrm{LPD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{LPD}_r(x_0), \quad \mathrm{LPSD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{LPSD}_r(x_0), \quad \mathrm{LD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{LD}_r(x_0).$$

The following lemma characterises some of the preceding types of functions by class $\mathcal{K}$-functions.

4.3.7 Lemma (Positive-definite and decrescent in terms of class $\mathcal{K}$ II) *For* $\mathrm{U} \subseteq \mathbb{R}^n$ *open, a continuous function* $\mathrm{f}\colon \mathrm{U} \to \mathbb{R}$, *and* $\mathrm{r} \in \mathbb{R}_{>0}$, *the following statements hold:*

(i) $\mathrm{f} \in \mathrm{LPD}_r(\mathbf{x}_0)$ *if and only if there exist* $\phi_1, \phi_2 \in \mathcal{K}([0, \mathrm{r}); \mathbb{R}_{\geq 0})$ *such that*

$$\phi_1(\|\mathbf{x} - \mathbf{x}_0\|) \leq \mathrm{f}(\mathbf{x}) \leq \phi_2(\|\mathbf{x} - \mathbf{x}_0\|)$$

for all $\mathbf{x} \in \mathsf{B}(\mathrm{r}, \mathbf{x}_0)$;

(ii) $\mathrm{f} \in \mathrm{LD}_{\mathrm{r}}(\mathbf{x}_0)$ *if and only if there exists* $\phi \in \mathcal{K}([0,\mathrm{r});\mathbb{R}_{\geq 0})$ *such that*

$$\mathrm{f}(\mathbf{x}) \leq \phi(\|\mathbf{x}-\mathbf{x}_0\|)$$

for all $\mathbf{x} \in \mathsf{B}(\mathrm{r},\mathbf{x}_0)$.

Proof (i) Suppose that $f \in \mathrm{LPD}_r(x_0)$. We first define $\psi_1 : [0,r) \to \mathbb{R}_{\geq 0}$ by

$$\psi_1(s) = \inf\{f(x) \mid \|x-x_0\| \in [s,r)\}.$$

We claim that (1) ψ_1 is continuous, (2) $\psi_1(0) = 0$, (3) $\psi_1(s) \in \mathbb{R}_{>0}$ for $s \in (0,r)$, (4) ψ_1 is nonincreasing, and (5) $f(x) \geq \psi_1(\|x-x_0\|)$. The only one of these that is not rather obvious is the continuity of ψ_1.

This we prove as follows. Let $s_0 \in [0,r)$ and let $\epsilon \in \mathbb{R}_{>0}$. For $x \in \mathsf{B}(r,x_0)$, let $\delta_x \in \mathbb{R}_{>0}$ be such that, if $x' \in \mathsf{B}(r,x_0)$ satisfies $\|x'-x\| < \delta_x$, then $|f(x')-f(x)| < \epsilon$. Now, by compactness of

$$S(s_0,x_0) = \{x \in \mathsf{B}(r,x_0) \mid \|x-x_0\| = s_0\},$$

let $x_1,\dots,x_k \in S(x_0,x_0)$ be such that $S(s_0,x_0) \subseteq \cup_{j=1}^k \mathsf{B}(\delta_{x_j},x_j)$. Define

$$\begin{aligned} d_{s_0} &: S(s_0,x_0) \to \mathbb{R}_{>0} \\ &\quad x \mapsto \min\{\|x-x_1\|,\dots,\|x-x_k\|\}. \end{aligned}$$

Being a min of continuous functions, d_{s_0} is continuous (by ***missing stuff***). Being a continuous function on a compact set, there exists $\delta \in \mathbb{R}_{>0}$ such that $d_{s_0}(x) \geq \delta$ for every $x \in S(s_0,x_0)$. Now, let $s \in [0,r)$ be such that $|s-s_0| < \delta$. First suppose that $s > s_0$. Since ψ_1 is nondecreasing, $\psi_1(s) - \psi_1(s_0) \geq 0$. Now, if $x \in S(s_0,x_0)$, there exists $x' \in S(s,x_0)$ such that $|f(x')-f(x)| < \epsilon$. Thus

$$-\epsilon < f(x') - f(x) < \epsilon.$$

Since

$$\psi_1(s) \leq f(x'), \quad -\psi(s_0) \geq -f(x),$$

we have

$$\psi_1(s) - \psi(s_0) \leq f(x') - f(x) < \epsilon.$$

In like manner, if $s < s_0$, we have

$$\psi(s_0) - \psi(s) < \epsilon,$$

which gives $|\psi(s)-\psi(s_0)| < \epsilon$. This gives the asserted continuity of ψ_1.

Now, by Lemma 4.3.4, there exists $\phi_1 \in \mathcal{K}([0,r);\mathbb{R}_{\geq 0})$ such that

$$\phi_1(\|x-x_0\|) \leq \psi_1(\|x-x_0\|) \leq f(x)$$

for $x \in \mathsf{B}(r,x_0)$.

Next define $\psi_2 : [0, r) \to \mathbb{R}_{\geq 0}$ by

$$\psi_2(s) = \sup\{f(x) \mid \|x - x_0\| \leq s\}.$$

We can see that (1) ψ_2 is continuous, (2) $\psi_2(0) = 0$, (3) $\psi_2(s) \in \mathbb{R}_{>0}$ for $s \in (0, r)$, (4) ψ_2 is nondecreasing, and (5) $f(x) \leq \psi_2(\|x - x_0\|)$. Again, continuity is the only not completely trivial assertion, and an argument like that above for ψ_1 can be easily made to prove this continuity assertion. Now, by Lemma 4.3.4, there exists $\phi_2 \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ such that

$$\phi_2(\|x - x_0\|) \geq \psi_1(\|x - x_0\|) \geq f(x)$$

for $x \in \mathsf{B}(r, x_0)$.

Next suppose that there exist $\psi_1, \phi_2 \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ such that

$$\phi_1(\|x - x_0\|) \leq f(x) \leq \phi_2(\|x - x_0\|)$$

for all $x \in \mathsf{B}(r, x_0)$. The left inequality immediately gives $f \in \mathrm{LPD}_r(x_0)$.

(ii) Suppose that $f \in \mathrm{LD}_r(x_0)$. Let $g \in \mathrm{LPD}_r(x_0)$ be such that $f(x) \leq g(x)$ for $x \in \mathsf{B}(r, x_0)$. By part (i) let $\phi \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ be such that

$$\phi(\|x - x\|) \geq g(x) \geq f(x),$$

as desired.

Finally, suppose that there exists $\phi \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ such that $f(x) \leq \phi(\|x - x_0\|)$ for $x \in \mathsf{B}(r, x_0)$. Since the function g defined on $\mathsf{B}(r, x_0)$ by $g(x) = \phi(\|x - x_0\|)$ is locally positive-definite about x_0 in $\mathsf{B}(r, x_0)$, the proof of the lemma is concluded. ■

4.3.1.3 General time-varying functions

Next we generalise the constructions of the preceding section to allow functions that depend on time.

4.3.8 Definition (Locally definite, locally semidefinite, decrescent II) Let $U \subseteq \mathbb{R}^n$ be an open set, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let $x_0 \in U$. A function $f : \mathbb{T} \times U \to \mathbb{R}$ is:

(i) ***locally positive-definite*** about x_0 if
   (a) it is continuous,
   (b) $f(t, x_0) = 0$ for all $t \in \mathbb{T}$, and
   (c) there exist $r \in \mathbb{R}_{>0}$ and $f_0 \in \mathrm{LPD}_r(x_0)$ such that $f(t, x) \geq f_0(x)$ for every $(t, x) \in \mathbb{T} \times \mathsf{B}(r, x)$.

(ii) ***locally positive-semi definite*** about x_0 if
   (a) it is continuous,
   (b) $f(t, x_0) = 0$ for all $t \in \mathbb{T}$, and
   (c) there exist $r \in \mathbb{R}_{>0}$ and $f_0 \in \mathrm{LPSD}_r(x_0)$ such that $f(t, x) \geq f_0(x)$ for every $(t, x) \in \mathbb{T} \times \mathsf{B}(r, x)$.

(iii) ***locally negative-definite about*** $\mathbf{x_0}$ if $-f$ is positive-definite about x_0;
(iv) ***locally negative-semidefinite about*** $\mathbf{x_0}$ if $-f$ is positive-semidefinite about x_0;
(v) ***locally decrescent about*** $\mathbf{x_0}$ if there exist $r \in \mathbb{R}_{>0}$ and $g \in \mathrm{LPD}_r(x_0)$ such that $f(t,x) \leq g(x)$ for every $(t,x) \in \mathbb{T} \times \mathsf{B}(r,x_0)$. •

Let us introduce some notation for these classes of functions. As for time–invariant functions, we have all of the preceding notions of definiteness about x_0 "in $\mathsf{B}(r,x_0)$," with the obvious meaning. Let us not use all of the words required to make this obvious terminology precise. We also have the following symbols, keeping in mind that functions now are defined on $\mathbb{T} \times U$:

$\mathrm{TVLPD}_r(x_0)$ set of locally positive-definite functions about x_0 in $\mathsf{B}(r,x)$;
$\mathrm{TVLPSD}_r(x_0)$ set of locally positive-semidefinite functions about x_0 in $\mathsf{B}(r,x_0)$;
$\mathrm{TVLD}_r(x_0)$ set of locally decrescent functions about x_0 in $\mathsf{B}(r,x_0)$

and we also denote

$$\mathrm{TVLPD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{TVLPD}_r(x_0), \quad \mathrm{TVLPSD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{TVLPSD}_r(x_0),$$
$$\mathrm{TVLD}(x_0) = \cup_{r \in \mathbb{R}_{>0}} \mathrm{TVLD}_r(x_0).$$

An application of the definitions and of Lemma 4.3.7 gives the following lemma.

4.3.9 Lemma (Positive-definite and decrescent in terms of class $\mathcal{K}$ I) *For $\mathrm{U} \subseteq \mathbb{R}^n$ open, an interval $\mathbb{T} \subseteq \mathbb{R}$, a continuous function $\mathrm{f}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}$, and $\mathrm{r} \in \mathbb{R}_{>0}$, the following statements hold:*

(i) $\mathrm{f} \in \mathrm{TVLPD}_\mathrm{r}(\mathbf{x}_0)$ *if and only if there exist $\phi_1, \phi_2 \in \mathcal{K}([0,\mathrm{r});\mathbb{R}_{\geq 0})$ such that*

$$\phi_1(\|\mathbf{x} - \mathbf{x}_0\|) \leq \mathrm{f}(\mathrm{t},\mathbf{x}) \leq \phi_2(\|\mathbf{x} - \mathbf{x}_0\|)$$

for all $\mathrm{t} \in \mathbb{T}$ and $\mathbf{x} \in \mathsf{B}(\mathrm{r},\mathbf{x}_0)$;

(ii) $\mathrm{f} \in \mathrm{TVLD}_\mathrm{r}(\mathbf{x}_0)$ *if and only if there exists $\phi \in \mathcal{K}([0,\mathrm{r});\mathbb{R}_{\geq 0})$ such that*

$$\mathrm{f}(\mathrm{t},\mathbf{x}) \leq \phi(\|\mathbf{x} - \mathbf{x}_0\|)$$

for all $\mathrm{t} \in \mathbb{T}$ and $\mathbf{x} \in \mathsf{B}(\mathrm{r},\mathbf{x}_0)$.

4.3.10 Remark (The uniformity in time of time-varying definitions) The reader will note that, in the definition of $\mathrm{TVLPD}(x_0)$, etc., the characterisations are in terms of *time-invariant* functions from $\mathrm{LPD}(x_0)$, etc., and are required to hold for every $t \in \mathbb{T}$. One says, in this case, that the bounds required for elements of $\mathrm{TVLPD}(x_0)$, etc., hold ***uniformly*** in t. One might imagine conditions that are *not* uniform in t, but just what is required of such a definition is rather complicated. Our lack of consideration of these cases reflected in Sections 4.3.3 and 4.3.5, where we only consider Lyapunov's Second Method for characterising *uniform* stability, since nonuniform counterparts are more complicated. •

4.3.1.4 Time-invariant quadratic functions When we apply Lyapunov's Second Method to linear differential equations, we will use locally positive-definite functions as in the general case. However, because of the extra structure of linear equations, it is natural to consider locally positive-definite functions of a very particular form. In this section we shall consider the time-invariant case.

As we do when talking about linear ordinary differential equations, we shall work with equations whose state space is a finite-dimensional $\mathbb{R}$-vector space V. In such a case, the definitions of locally positive-definite, etc., are modified to account for the fact that we are principally interested in what is happening with the zero vector when talking about linear systems. The appropriate definitions require having at hand an inner product that generalises the Euclidean inner product.[8] That is, we suppose that we assign to each pair of vectors $v_1, v_2 \in \mathsf{V}$ a number $\langle v_1, v_2 \rangle \in \mathbb{R}$, and this assignment has the following properties:

1. for fixed $v_2 \in \mathsf{V}$, the function $v_1 \mapsto \langle v_1, v_2 \rangle$ is linear;
2. for fixed $v_1 \in \mathsf{V}$, the function $v_2 \mapsto \langle v_1, v_2 \rangle$ is linear;
3. $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$ for all $v_1, v_2 \in \mathsf{V}$;
4. $\langle v, v \rangle \in \mathbb{R}_{\geq 0}$ for all $v \in \mathsf{V}$;
5. $\langle v, v \rangle = 0$ only if $v = 0$.

We think of $\langle v_1, v_2 \rangle$ as being the "angle" between v_1 and v_2. The following are terminology and facts we shall require about inner products.

1. The assignment $v \mapsto \sqrt{\langle v, v \rangle}$ defines a norm on V that we shall simply denote by $\|\cdot\|$.
2. Given $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$, the ***transpose*** of L is the linear map $L^T \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ defined by

$$\langle L^T(v_1), v_2 \rangle = \langle v_1, L(v_2) \rangle, \qquad v_1, v_2 \in \mathsf{V}.$$

   A linear map L is ***symmetric*** if $L^T = L$.
3. If V is n-dimensional and if $L \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ is symmetric, then
   (a) its eigenvalues are real and
   (b) there is an orthonormal basis $\{e_1, \ldots, e_n\}$ of eigenvectors, i.e., (i) each of the vectors e_j, $j \in \{1, \ldots, n\}$, is an eigenvector for some eigenvalue, (ii) $\langle e_j, e_k \rangle = 0$ for $j \neq k$, and (iii) $\|e_j\| = 1$, $j \in \{1, \ldots, n\}$.

The functions of interest to us are then those prescribed by the following definition.

[8] Children call the Euclidean inner product the "dot" product, and it is defined by

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto \sum_{j=1}^{n} x_{1,j} x_{2,j}.$$

The expression on the right is often denoted $\boldsymbol{x}_1 \cdot \boldsymbol{x}_2$. However, we eschew the "$\cdot$"-notation, which is for babies, and instead write it as $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n}$.

4.3.11 Definition (Quadratic function) Let V be an n-dimensional $\mathbb{R}$-vector space, let $\langle \cdot, \cdot \rangle$ be an inner product on V, and let $Q \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ be a symmetric linear map. The ***quadratic function*** associated to Q is

$$\begin{aligned} f_Q \colon \mathsf{V} &\to \mathbb{R} \\ v &\mapsto \langle Q(v), v \rangle. \end{aligned}$$

•

Now we classify various sorts of quadratic functions.

4.3.12 Definition (Locally definite, locally semidefinite, decrescent III) Let V be an n-dimensional $\mathbb{R}$-vector space, let $\langle \cdot, \cdot \rangle$ be an inner product on V, and let $Q \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ be a symmetric linear map. The linear map Q is:

(i) ***positive-definite*** if $f_Q(v) \in \mathbb{R}_{>0}$ for $v \in \mathsf{V} \setminus \{0\}$;

(ii) ***positive-semi definite*** if $f_Q(v) \in \mathbb{R}_{\geq 0}$ for $v \in \mathsf{V}$;

(iii) ***negative-definite*** if $-Q$ is positive-definite;

(iv) ***negative-semidefinite*** if $-Q$ is positive-semidefinite;

(v) ***decrescent*** if there exists a positive-definite symmetric linear map $Q_0 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ such that $f_Q(v) \leq f_{Q_0}(v)$ for $v \in \mathsf{V}$. •

Let us relate these notions to local definiteness notions for general functions, and also to the eigenvalues of Q.

4.3.13 Lemma (Characterisations of definite, semidefinite, and decrescent symmetric linear maps) *Let* V *be an* n*-dimensional* $\mathbb{R}$*-vector space, let* $\langle \cdot, \cdot \rangle$ *be an inner product on* V*, and let* $\mathrm{Q} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ *be a symmetric linear map. Then the following statements hold.*

(i) The following statements are equivalent:

(a) Q *is positive-definite;*

(b) $\mathrm{f_Q} \in \mathrm{LPD}(0)$*;*

(c) $\mathrm{spec}(\mathrm{Q}) \subseteq \mathbb{R}_{>0}$.

(ii) The following statements are equivalent:

(a) Q *is positive-semidefinite;*

(b) $\mathrm{f_Q} \in \mathrm{LPSD}(0)$*;*

(c) $\mathrm{spec}(\mathrm{Q}) \subseteq \mathbb{R}_{\geq 0}$.

(iii) Q *is decrescent.*

Proof First of all, because Q is symmetric, all eigenvalues of Q are real and there is an orthonormal basis $\{e_1, \dots, e_n\}$ of eigenvectors. Thus there exist $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that

$$Q(e_j) = \lambda_j e_j, \qquad j \in \{1, \dots, n\}.$$

Therefore, if $v = \sum_{j=1}^{n} v_j e_v$, then

$$f_Q(v) = \left\langle Q\left(\sum_{j=1}^{n} v_j e_j\right), \sum_{k=1}^{n} v_k e_k \right\rangle = \sum_{j,k=1}^{n} v_j v_k \langle Q(e_j), e_k \rangle$$
$$= \sum_{j,k=1}^{k} \lambda_j v_j v_k \langle e_j, e_k \rangle = \sum_{j=1}^{n} \lambda_j v_j^2.$$

With this formula in hand, we prove the lemma.

(i) If Q is positive-definite, then it is clear that f_Q is locally positive-definite, from the definition.

Now, we claim that, if $\operatorname{spec}(Q) \not\subset \mathbb{R}_{>0}$, then f_Q is not locally positive definite about 0. Indeed, suppose that $\lambda_j \leq 0$ for some $j \in \{1, \dots, n\}$. Then, for any $\epsilon \in \mathbb{R}_{>0}$,

$$f_Q(\epsilon e_j) = \lambda_j \epsilon^2 \leq 0.$$

Since, for any $r \in \mathbb{R}_{>0}$, we can choose $\epsilon = \frac{r}{2} \in \mathbb{R}_{>0}$ so that $\epsilon e_j \in \mathsf{B}(r, 0) \setminus \{0\}$, it cannot be the case that f_Q is locally positive-definite.

Finally, if $\operatorname{spec}(Q) \subseteq \mathbb{R}_{>0}$, then the formula

$$f_Q(v) = \sum_{j=1}^{n} \lambda_j v_j^2$$

ensures that Q is positive-definite.

(ii) The proof follows along the lines of the first part of the proof, *mutatis mutandis*.

(iii) As in the opening paragraph of the proof, we write

$$f_Q(v) = \lambda_j v_j^2,$$

where $\lambda_1, \dots, \lambda_j$ are the eigenvalues of Q. We then let

$$C = \max\{1, \lambda_1, \dots, \lambda_n\}$$

and define $Q_0 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ so that

$$f_{Q_0}(v) = C \sum_{j=1}^{n} v_j^2,$$

and observe that Q_0 is positive-definite (by part (i)) and that $f_Q(v) \leq f_{Q_0}(v)$ for all $v \in \mathsf{V}$. ■

The vacuous nature of the nature of decrescent symmetric linear maps (every symmetric linear map is decrescent) arises simply because this notion is not really a valuable one for time-invariant quadratic functions. We state the definition simply for the sake of preserving symmetry of the definitions.

Along these lines, the following result will be helpful to us in the next section.

4.3.14 Lemma (Upper and lower bounds for positive-definite quadratic functions) *Let* V *be an* n*-dimensional* $\mathbb{R}$*-vector space, let* $\langle \cdot, \cdot \rangle$ *be an inner product on* V*, and let* $\mathrm{Q} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ *be a positive-definite symmetric linear map. Then there exists* $\mathrm{C} \in \mathbb{R}_{>0}$ *such that, for every* $\mathrm{v} \in \mathsf{V}$*, we have*

$$\mathrm{C}\langle \mathrm{v}, \mathrm{v} \rangle \leq \mathrm{f}_{\mathrm{Q}}(\mathrm{v}) \leq \mathrm{C}^{-1}\langle \mathrm{v}, \mathrm{v} \rangle.$$

Proof As in the proof of Lemma 4.3.13, for an orthonormal basis of eigenvectors $\{e_1, \dots, e_n\}$, we have

$$f_Q(v) = \sum_{j=1}^{n} \lambda_j v_j^2$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues. The result follows by taking requiring that

$$C \leq \min\{\lambda_1, \dots, \lambda_n\}$$

and

$$C^{-1} \geq \max\{\lambda_1, \dots, \lambda_n\}.$$ ■

4.3.1.5 Time-varying quadratic functions

The final collection of functions we consider are those that are quadratic, as in the preceding section, and vary with time. A reader who has been paying attention while reading the preceding sections will likely be able to write down the definitions and characterisations we give next, as these follow quite naturally from what we have done already.

4.3.15 Definition (Time-varying quadratic function) Let V be an n-dimensional $\mathbb{R}$-vector space, let $\langle \cdot, \cdot \rangle$ be an inner product on V, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let $Q \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ be such that $Q(t)$ is a symmetric linear map for every $t \in \mathbb{T}$. The ***time-varying quadratic function*** associated to Q is

$$\begin{aligned} f_Q \colon \mathbb{T} \times \mathsf{V} &\to \mathbb{R} \\ (t, v) &\mapsto \langle Q(t)(v), v \rangle. \end{aligned}$$ •

4.3.16 Definition (Locally definite, locally semidefinite, decrescent IV) Let V be an n-dimensional $\mathbb{R}$-vector space, let $\langle \cdot, \cdot \rangle$ be an inner product on V, let $\mathbb{T} \subseteq \mathbb{R}$ be an interval, and let $Q \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ be such that $Q(t)$ is a symmetric linear map for every $t \in \mathbb{T}$. The function Q is:

(i) ***positive-definite*** if there exists a positive-definite symmetric linear map $Q_0 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ such that $f_Q(t, v) \geq f_{Q_0}(v)$ for $(t, v) \in \mathbb{T} \times \mathsf{V}$;

(ii) ***positive-semi definite*** if there exists a positive-definite symmetric linear map $Q_0 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ such that $f_Q(t, v) \geq f_{Q_0}(v)$ for $(t, v) \in \mathbb{T} \times \mathsf{V}$;

(iii) ***negative-definite*** if $-Q$ is positive-definite;

(iv) ***negative-semidefinite*** if $-Q$ is positive-semidefinite.

(v) ***decrescent*** if there exists a positive-definite symmetric linear map $Q_0 \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ such that $f_Q(t, v) \leq f_{Q_0}(v)$ for $(t, v) \in \mathbb{T} \times \mathsf{V}$. •

4.3.17 Lemma (Characterisations of definite, semidefinite, and decrescent time-varying symmetric linear maps) *Let* V *be an* n*-dimensional* $\mathbb{R}$*-vector space, let* $\langle \cdot, \cdot \rangle$ *be an inner product on* V*, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, and let* $Q \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *be such that* $Q(t)$ *is a symmetric linear map for every* $t \in \mathbb{T}$*. Then the following statements hold.*

(i) The following statements are equivalent:

(a) Q *is positive-definite;*

(b) $\mathrm{f}_Q \in \mathrm{TVLPD}(0)$*;*

(c) there exists $\ell \in \mathbb{R}_{>0}$ *such that*

$$\ell \leq \inf\{\lambda \in \mathbb{R} \mid \lambda \in \operatorname{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\}.$$

(ii) The following statements are equivalent:

(a) Q *is positive-semidefinite;*

(b) $\mathrm{f}_Q \in \mathrm{TVLPSD}(0)$*;*

(c) there exists $\ell \in \mathbb{R}_{\geq 0}$ *such that*

$$\ell \leq \inf\{\lambda \in \mathbb{R} \mid \lambda \in \operatorname{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\}.$$

(iii) The following statements are equivalent:

(a) Q *is decrescent;*

(b) $\mathrm{f}_Q \in \mathrm{TVLD}(0)$*;*

(c) there exists $\mu \in \mathbb{R}_{>0}$ *such that*

$$\mu \geq \sup\{\lambda \in \mathbb{R} \mid \lambda \in \operatorname{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\}.$$

Proof (i) First suppose that Q is positive-definite. By definition, by Lemma 4.3.13(i), and by Definition 4.3.8(i), $f_Q \in \mathrm{TVLPD}(0)$.

Next, suppose that

$$\inf\{\lambda \in \mathbb{R} \mid \lambda \in \operatorname{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\} \leq 0.$$

For $t \in \mathbb{T}$, let $\lambda_1(t), \dots, \lambda_n(t) \subseteq \mathbb{R}$ be the eigenvalues of $Q(t)$. Without loss of generality, suppose that

$$\lambda_1(t) = \min\{\lambda_1(t), \dots, \lambda_n(t)\}, \qquad t \in \mathbb{T}.$$

For $t \in \mathbb{T}$, let $v_1(t) \in \mathsf{V}$ be an eigenvector for the eigenvalue $\lambda_1(t)$, and suppose that $\|v_1(t)\| = 1$, and note that

$$f_Q(t, v_1(t)) = \langle Q(t)(v_1(t)), v_1(t) \rangle = \lambda_1(t) \langle v_1(t), v_1(t) \rangle = \lambda_1(t).$$

By assumption $\inf\{f_Q(t, v_1(t)) \mid t \in \mathbb{T}\} \leq 0$. This means that there exists a sequence $(t_j)_{j \in \mathbb{Z}_{>0}}$ such that

$$\lim_{j \to \infty} f_Q(t_j, v_1(t_j)) \leq 0.$$

Now let $r \in \mathbb{R}_{>0}$ and $g \in \mathrm{TVLPD}_r(0)$. By Lemma 4.3.9(i), let $\phi \in \mathcal{K}([0,r);\mathbb{R}_{\geq 0})$ be such that $\phi(|x|) \leq g(x)$ for all $x \in \mathsf{B}(r,0)$. For $\epsilon \in \mathbb{R}_{>0}$ such that $\epsilon^2 < r$, we have

$$\lim_{j\to\infty} f(t_j, \epsilon^2 v_1(t_j)) = \epsilon \lim_{j\to\infty} f(t_j, v_1(t_j)) \leq 0 < \phi(\epsilon^2) \leq g(\epsilon^2 v)$$

for every $v \in \mathsf{V}$ for which $\|v\| = 1$. This means that there exists $N \in \mathbb{Z}_{>0}$ such that

$$f(t_j, \epsilon^2 v_1(t_j)) < g(\epsilon^2 v_1(t_j)), \qquad j \geq N.$$

Since g and r were arbitrary, this prohibits f from being in TVLPD(0).

Finally, suppose that

$$\inf\{\lambda \in \mathbb{R} \mid \lambda \in \mathrm{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\} > 0.$$

Let $\ell \in \mathbb{R}_{>0}$ be such that

$$\inf\{\lambda \in \mathbb{R} \mid \lambda \in \mathrm{spec}(Q(t)) \text{ for some } t \in \mathbb{T}\} \geq \ell$$

and define the symmetric positive-definite linear map Q_0 so that $f_{Q_0}(v) = \ell\langle v, v\rangle$ for all $v \in \mathsf{V}$. Then, for $t \in \mathbb{T}$, let $\lambda_1(t), \ldots, \lambda_n(t)$ be the eigenvalues for $Q(t)$ and let $\{e_1(t), \ldots, e_n(t)\}$ be an orthonormal basis of eigenvectors. If $v \in \mathsf{V}$, write

$$v = \sum_{j=1}^{n} v_j(t) e_j(t)$$

for uniquely defined $v_1(t), \ldots, v_n(t) \in \mathbb{R}$. Then, recalling the calculations from the proof of Lemma 4.3.13,

$$f_Q(t, v) = \sum_{j=1}^{n} \lambda_j(t) v_j(t)^2 \geq \ell \sum_{j=1}^{n} v_j(t)^2 = \ell\langle v, v\rangle = f_{Q_0}(v),$$

and so Q is positive-definite.

(ii) This follows, *mutatis mutandis*, as does the preceding part of the lemma.

(iii) This also follows, *mutatis mutandis*, from the proof of part (i). ■

4.3.18 Lemma (Upper and lower bounds for time-varying positive-definite and decrescent quadratic functions) *Let* V *be an* n*-dimensional* $\mathbb{R}$*-vector space, let* $\langle\cdot,\cdot\rangle$ *be an inner product on* V*, let* $\mathbb{T} \subseteq \mathbb{R}$ *be an interval, and let* $\mathrm{Q}\colon \mathbb{T} \to \mathrm{L}(\mathsf{V};\mathsf{V})$ *be such that* $\mathrm{Q}(\mathrm{t})$ *is symmetric for every* $\mathrm{t} \in \mathbb{T}$*. Then the following statements hold:*

(i) Q *is positive-definite if and only if there exists* $\mathrm{C} \in \mathbb{R}_{>0}$ *such that* $\mathrm{C}\langle \mathrm{v}, \mathrm{v}\rangle \leq \mathrm{f}_{\mathrm{Q}}(\mathrm{t}, \mathrm{v})$ *for every* $(\mathrm{t}, \mathrm{v}) \in \mathbb{T} \times \mathsf{V}$*;*

(ii) Q *is decrescent if and only if there exists* $\mathrm{C} \in \mathbb{R}_{>0}$ *such that* $\mathrm{f}_{\mathrm{Q}}(\mathrm{t}, \mathrm{v}) \leq \mathrm{C}\langle \mathrm{v}, \mathrm{v}\rangle$ *for every* $(\mathrm{t}, \mathrm{v}) \in \mathbb{T} \times \mathsf{V}$*.*

Proof As in the proof of Lemma 4.3.17, for $t \in \mathbb{T}$ we let $\lambda_1(t), \dots, \lambda_j(t)$ be the eigenvalues for $Q(t)$ and let $\{e_1(t), \dots, e_n(t)\}$ be an orthonormal basis of eigenvectors for $Q(t)$. If we write

$$v = \sum_{j=1}^{n} v_j(t) e_j(t),$$

we then have

$$f_Q(t, v) = \sum_{j=1}^{n} \lambda_j(t) v_j(t)^2.$$

Then Q is positive-definite if and only if there exists $C \in \mathbb{R}_{>0}$ such that

$$C \leq \lambda_j(t), \qquad j \in \{1, \dots, n\},\ t \in \mathbb{T},$$

and Q is decrescent if and only if there exists $C \in \mathbb{R}_{>0}$ such that

$$\lambda_j(t) \leq C, \qquad j \in \{1, \dots, n\},\ t \in \mathbb{T}.$$

The result then follows by a simple computation, mirroring many we have already done. ■

4.3.2 Stability in terms of class $\mathcal{K}$- and class $\mathcal{KL}$-functions

In this section, whose content consists of a single lemma with its lengthy proof, we characterise various notions of stability in terms of class $\mathcal{K}$- and class $\mathcal{KL}$-functions. While it is possible to prove some of our results relating to Lyapunov's Second Method, the characterisations we give in the lemma are useful in capturing the essence of some of the proofs, and of uniting their style.

Here is the lemma of which we speak.

4.3.19 Lemma (Stability of equilibria for nonautonomous equations in terms of class $\mathcal{K}$- and class $\mathcal{KL}$-functions) *Let* $\mathbf{F}$ *be a system of ordinary differential equations with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^n$$

with $\sup \mathbb{T} = \infty$ *and satisfying Assumption 4.1.1. For an equilibrium point* $\mathbf{x}_0 \in \mathrm{U}$ *for* $\mathbf{F}$*, the following statements hold:*

(i) $\mathbf{x}_0$ *is stable if and only if, for each* $t_0 \in \mathbb{T}$*, there exist* $\delta \in \mathbb{R}_{>0}$ *and* $\alpha \in \mathcal{K}([0, \delta); \mathbb{R}_{\geq 0})$ *such that, for every* $\mathbf{x} \in \mathrm{U}$ *satisfying* $\|\mathbf{x} - \mathbf{x}_0\| < \delta$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(t) = \widehat{\mathbf{F}}(t, \xi(t)), \quad \xi(t_0) = \mathbf{x},$$

satisfies $\|\xi(t)\| \leq \alpha(\|\mathbf{x} - \mathbf{x}_0\|)$ *for* $t \geq t_0$*;*

(ii) $\mathbf{x}_0$ *is uniformly stable if and only if there exist* $\delta \in \mathbb{R}_{>0}$ *and* $\alpha \in \mathcal{K}([0,\delta);\mathbb{R}_{\geq 0})$ *such that, for every* $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T} \times \mathrm{U}$ *satisfying* $\|\mathbf{x} - \mathbf{x}_0\| < \delta$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x},$$

satisfies $\|\xi(\mathrm{t}) - \mathbf{x}_0\| \leq \alpha(\|\mathbf{x} - \mathbf{x}_0\|)$ *for* $\mathrm{t} \geq \mathrm{t}_0$*;*

(iii) $\mathbf{x}_0$ *is asymptotically stable if and only if, for every* $\mathrm{t}_0 \in \mathbb{T}'$*, there exist* $\delta \in \mathbb{R}_{>0}$ *and* $\beta \in \mathcal{KL}([0,\delta) \times [\mathrm{t}_0, \infty); \mathbb{R}_{\geq 0})$ *such that, if* $\mathbf{x} \in \mathrm{U}$ *satisfies* $\|\mathbf{x} - \mathbf{x}_0\| < \delta$*, then the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x},$$

satisfies $\|\xi(\mathrm{t}) - \xi_0(\mathrm{t})\| \leq \beta(\|\mathbf{x} - \mathbf{x}_0\|, \mathrm{t})$ *for* $\mathrm{t} \geq \mathrm{t}_0$*;*

(iv) $\mathbf{x}_0$ *is uniformly asymptotically stable if and only if there exist* $\delta \in \mathbb{R}_{>0}$ *and* $\beta \in \mathcal{KL}([0,\delta) \times [0,\infty); \mathbb{R}_{\geq 0})$ *such that, if* $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T} \times \mathrm{U}$ *satisfies* $\|\mathbf{x} - \mathbf{x}_0\| < \delta$*, then the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x},$$

satisfies $\|\xi(\mathrm{t}) - \xi_0(\mathrm{t})\| \leq \beta(\|\mathbf{x} - \mathbf{x}_0\|, \mathrm{t} - \mathrm{t}_0)$ *for* $\mathrm{t} \geq \mathrm{t}_0$.

Proof (i) First suppose that, for each $t_0 \in \mathbb{T}$, there exist $\delta \in \mathbb{R}_{>0}$ and $\alpha \in \mathcal{K}([0,\delta);\mathbb{R}_{\geq 0})$ such that, for every $\boldsymbol{x} \in U$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

satisfies $\|\xi(t)\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|)$ for $t \geq t_0$. Let $t_0 \in \mathbb{T}$ and let δ and α be as above. Let $\epsilon \in \mathbb{R}_{>0}$ and let $\epsilon' = \min\{\epsilon, \alpha(\delta)\}$. Let $\delta' = \min\{\delta, \alpha^{-1}(\frac{\epsilon'}{2})\}$. Let $\boldsymbol{x} \in U$ satisfy $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta' \leq \delta$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x}.$$

Since

$$\|\xi(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|) \leq \alpha(\delta') \leq \alpha(\alpha^{-1}(\tfrac{\epsilon'}{2})) = \tfrac{\epsilon'}{2} < \epsilon,$$

we conclude stability of $\boldsymbol{x}_0$.

Next suppose that $\boldsymbol{x}_0$ is stable and let $t_0 \in \mathbb{T}$. For $\epsilon \in \mathbb{R}_{>0}$, let $A(\epsilon) \subseteq \mathbb{R}_{>0}$ be the set of positive numbers δ such that, for every $\boldsymbol{x} \in U$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

satisfies $\|\xi(t)\| < \epsilon$ for $t \geq t_0$. Then denote $\overline{\delta}(\epsilon) = \sup A(\epsilon)$. This then defines, for some $\epsilon_0 \in \mathbb{R}_{>0}$, a function $\overline{\delta}\colon [0,\epsilon_0) \to \mathbb{R}_{\geq 0}$ that is nondecreasing. By ***missing stuff*** there exists $\overline{\alpha} \in \mathcal{K}([0,\epsilon_0);\mathbb{R}_{\geq 0})$ such that $\overline{\alpha}(\epsilon) \leq \overline{\delta}(\epsilon)$ for every $\epsilon \in [0,\epsilon_0)$. We can suppose that ϵ_0 is sufficiently small that $\mathrm{image}(\overline{\alpha}) = [0,\delta_0)$ for $\delta_0 \in \mathbb{R}_{>0}$. Define

$\alpha = \overline{\alpha}^{-1}$, which is of class $\mathcal{K}$ by Lemma 4.3.3(i). Now, let $\boldsymbol{x} \in U$ satisfies $\|\boldsymbol{x}-\boldsymbol{x}_0\| < \delta_0$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}.$$

Then let $\epsilon = \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|)$ note that

$$\|\boldsymbol{x} - \boldsymbol{x}_0\| = \overline{\alpha}(\epsilon) \leq \overline{\delta}(\epsilon).$$

Therefore,

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| < \epsilon = \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|),$$

completing this part of the proof.

(ii) First suppose that there exist $\delta \in \mathbb{R}_{>0}$ and $\alpha \in \mathcal{K}([0,\delta);\mathbb{R}_{\geq 0})$ such that, for every $(t_0, \boldsymbol{x}) \in \mathbb{T}\times U$ satisfying $\|\boldsymbol{x}-\boldsymbol{x}_0\| < \delta$, the solution $\boldsymbol{\xi}$ to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x},$$

satisfies $\|\boldsymbol{\xi}(t)-\boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x}-\boldsymbol{x}_0\|)$ for $t \geq t_0$. Let δ and α be as above. Let $\epsilon \in \mathbb{R}_{>0}$ and let $\epsilon' = \min\{\epsilon, \alpha(\delta)\}$. Let $\delta' = \min\{\delta, \alpha^{-1}(\frac{\epsilon'}{2})\}$. Let $(t_0, \boldsymbol{x}) \in \mathbb{T} \times U$ satisfy $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta' \leq \delta$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}.$$

Since

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|) \leq \alpha(\delta') \leq \alpha(\alpha^{-1}(\tfrac{\epsilon'}{2})) = \tfrac{\epsilon'}{2} < \epsilon,$$

we conclude uniform stability of $\boldsymbol{x}_0$.

Next suppose that $\boldsymbol{x}_0$ is uniformly stable. For $\epsilon \in \mathbb{R}_{>0}$, let $A(\epsilon) \subseteq \mathbb{R}_{>0}$ be the set of positive numbers δ such that, for every $(t_0, \boldsymbol{x}) \in \mathbb{T} \times U$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, the solution $\boldsymbol{\xi}$ to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x},$$

satisfies $\|\boldsymbol{\xi}(t)\| < \epsilon$ for $t \geq t_0$. Then denote $\overline{\delta}(\epsilon) = \sup A(\epsilon)$. This then defines, for some $\epsilon_0 \in \mathbb{R}_{>0}$, a function $\overline{\delta}\colon [0,\epsilon_0) \to \mathbb{R}_{\geq 0}$ that is nondecreasing. By ***missing stuff*** there exists $\overline{\alpha} \in \mathcal{K}([0,\epsilon_0);\mathbb{R}_{\geq 0})$ such that $\overline{\alpha}(\epsilon) \leq \overline{\delta}(\epsilon)$ for every $\epsilon \in [0,\epsilon_0)$. We can suppose that ϵ_0 is sufficiently small that image$(\overline{\alpha}) = [0,\delta_0)$ for $\delta_0 \in \mathbb{R}_{>0}$. Define $\alpha = \overline{\alpha}^{-1}$, which is of class $\mathcal{K}$ by Lemma 4.3.3(i). Now, let $\boldsymbol{x} \in U$ satisfy $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta_0$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}.$$

Then let $\epsilon = \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|)$ note that

$$\|\boldsymbol{x} - \boldsymbol{x}_0\| = \overline{\alpha}(\epsilon) \leq \overline{\delta}(\epsilon).$$

Therefore,

$$\|\xi(t) - x_0\| < \epsilon = \alpha(x - x_0),$$

completing this part of the proof.

(iii) First suppose that, for every $t_0 \in \mathbb{T}'$, there exist $\delta \in \mathbb{R}_{>0}$ and $\beta \mathcal{KL}([0, \delta) \times [t_0, \infty); \mathbb{R}_{\geq 0})$ such that, if $x \in U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t) - \xi_0(t)\| \leq \beta(\|x - x_0\|, t)$ for $t \geq t_0$. Let $t_0 \in \mathbb{T}$ and let δ and β be as above. If $x \in U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - \xi_0(t)\| \leq \beta(\|x - x_0\|, t) \leq \beta(\|x - x_0\|, t_0)$$

for $t \geq t_0$. By (ii) we conclude that x_0 is stable. Also, let $\epsilon \in \mathbb{R}_{>0}$ and let $T \in \mathbb{R}_{>0}$ be sufficiently large that $\beta(\frac{\delta}{2}, t_0 + T) < \epsilon$. Then, if $(t_0, x) \in \mathbb{T} \times U$ satisfy $\|x - x_0\| < \frac{\delta}{2}$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - x_0\| \leq \beta(\tfrac{\delta}{2}, t) \leq \beta(\tfrac{\delta}{2}, t_0 + T) < \epsilon$$

for $t \geq t_0 + T$. This gives asymptotic stability of x_0.

Next suppose that x_0 is asymptotically stable. Let $t_0 \in \mathbb{T}$. Since x_0 is stable (by definition), by part (i) there exists $\delta_0 \in \mathbb{R}_{>0}$ and $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$ such that, for $\delta \in [0, \delta_0]$, if $x \in U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ of the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - x_0\| \leq \alpha(\|x - x_0\|) < \alpha(r).$$

Now, if $(\delta, \epsilon) \in [0, \delta_0] \times \mathbb{R}_{>0}$, then let $A(\delta, \epsilon) \subseteq \mathbb{R}_{>0}$ be the set of $T \in \mathbb{R}_{>0}$ such that, if $x \in U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t) = x_0\| < \epsilon$ for $t \geq t_0 + T$, this being possible by asymptotic stability. Then define $\overline{T}(\delta, \epsilon) = \inf A(\delta, \epsilon)$.

Let us record some useful properties of $\overline{T}$.

1 Lemma

(i) $\overline{\mathrm{T}}(\delta,\epsilon) \in \mathbb{R}_{\geq 0}$ *for all* $(\delta,\epsilon) \in [0,\delta_0] \times \mathbb{R}_{>0}$;

(ii) $\delta \mapsto \overline{\mathrm{T}}(\delta,\epsilon)$ *is nondecreasing for every* $\epsilon \in \mathbb{R}_{>0}$, *i.e.,* $\overline{\mathrm{T}}(\delta_1,\epsilon) \leq \overline{\mathrm{T}}(\delta_2,\epsilon)$ *for* $\delta_1 < \delta_2$;

(iii) $\epsilon \mapsto \overline{\mathrm{T}}(\delta,\epsilon)$ *is nonincreasing for every* $\delta \in [0,\delta_0]$, *i.e.,* $\overline{\mathrm{T}}(\delta,\epsilon_1) \geq \overline{\mathrm{T}}(\delta,\epsilon_2)$ *for* $\epsilon_1 < \epsilon_2$;

(iv) $\overline{\mathrm{T}}(\delta,\epsilon) = 0$ *if* $\epsilon > \alpha(\delta)$.

Proof (i) This follows since, if $T \in A(\delta,\epsilon)$, then $T \in \mathbb{R}_{\geq 0}$.

(ii) Let $\delta_1 < \delta_2$. By definition, if $T \in A(\delta_2,\epsilon)$ then it is also the case that $T \in A(\delta_1,\epsilon)$. That is, $A(\delta_2,\epsilon) \subseteq A(\delta_1,\epsilon)$ and so $\inf A(\delta_1,\epsilon) \leq \inf A(\delta_2,\epsilon)$.

(iii) Let $\epsilon_1 < \epsilon_2$. Here, if $T \in A(\delta,\epsilon_1)$ then $T \in A(\delta,\epsilon_2)$, and this gives the result.

(iv) If $\epsilon > \alpha(\delta)$, then, if $(t_0,\boldsymbol{x}) \in \mathbb{T} \times U$ satisfies $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, the solution $\boldsymbol{\xi}$ of the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t,\boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x},$$

satisfies

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|) < \alpha(\delta) < \epsilon$$

for all $t \geq t_0$. Thus $0 \in A(\delta,\epsilon)$. ▼

For $(\delta,\epsilon) \in [0,\delta_0] \times \mathbb{R}_{>0}$, define

$$\tau(\delta,\epsilon) = \frac{2}{\epsilon} \int_{\epsilon/2}^{\epsilon} \overline{T}(\delta,x), \mathrm{d}x + \frac{\delta}{\epsilon}.^{9}$$

Let us record some properties of τ.

2 Lemma

(i) $\tau(\delta,\epsilon) \in \mathbb{R}_{>0}$ *for every* $(\delta,\epsilon) \in [0,\delta_0] \times \mathbb{R}_{>0}$;

(ii) $\epsilon \mapsto \tau(\delta,\epsilon)$ *is continuous for every* $\delta \in [0,\delta_0]$;

(iii) $\lim_{\epsilon\to\infty} \tau(\delta,\epsilon) = 0$ *for every* $\delta \in [0,\delta_0]$;

(iv) $\delta \mapsto \tau(\delta,\epsilon)$ *is strictly increasing for every* $\epsilon \in \mathbb{R}_{>0}$;

(v) $\epsilon \mapsto \tau(\delta,\epsilon)$ *is strictly decreasing for every* $\delta \in [0,\delta_0]$;

(vi) $\tau(\delta,\epsilon) \geq \overline{\mathrm{T}}(\delta,\epsilon) + \frac{\delta}{\epsilon}$.

Proof (i) This follows since $\overline{T}$ is $\mathbb{R}_{\geq 0}$-valued by Lemma 1(i).

(ii) By the Fundamental Theorem of Calculus, the function

$$\epsilon \mapsto \int_{\epsilon/2}^{\epsilon} \overline{T}(\delta,x), \mathrm{d}x$$

is continuous, and from this the continuity of τ follows.

[9] There is a fussy little point here about whether $\overline{T}$ is locally integrable in ϵ. This follows since $\overline{T}$ is nonincreasing, and so of "bounded variation."

(iii) For fixed δ, we have $\overline{T}(\delta,\epsilon) = 0$ for $\epsilon > \alpha(\delta)$ by Lemma 1(iv), and so

$$\lim_{\epsilon\to\infty} \tau(\delta,\epsilon) = \lim_{\epsilon\to\infty} \frac{\delta}{\epsilon} = 0.$$

(iv) This follows since

$$\delta \mapsto \frac{2}{\epsilon}\int_{\epsilon/2}^{\epsilon} \overline{T}(\delta,x)\,\mathrm{d}x$$

is nondecreasing by Lemma 1(ii) and since $\delta \mapsto \frac{\delta}{\epsilon}$ is strictly increasing.

(v) This follows since $\epsilon \mapsto \frac{2}{\epsilon}$ is strictly decreasing, since

$$\int_{\epsilon/2}^{\epsilon} \overline{T}(\delta,x)\,\mathrm{d}x$$

is nonincreasing by Lemma 1(iii) and since $\epsilon \mapsto \frac{\epsilon}{\delta}$ is strictly decreasing.

(vi) We have

$$\tau(\delta,\epsilon) \geq \frac{2}{\epsilon}\int_{\epsilon/2}^{\epsilon} \overline{T}(\delta,\epsilon)\,\mathrm{d}x + \frac{\delta}{\epsilon} \geq \overline{T}(\delta,\epsilon) + \frac{\delta}{\epsilon},$$

as claimed. ▼

Now, for $(\delta,s) \in [0,\delta_0]\times\mathbb{R}_{>0}$, define $\sigma(\delta,s) \in \mathbb{R}_{\geq 0}$ by asking that $\sigma(\delta,\tau(\delta,\epsilon)) = \epsilon$, i.e., $s \mapsto \sigma(\delta,s)$ is the inverse of $\epsilon \mapsto \tau(\delta,\epsilon)$. We have the following properties of σ.

3 Lemma

(i) $\delta \mapsto \sigma(\delta,\mathrm{s})$ *is strictly increasing for every* $\mathrm{s} \in \mathbb{R}_{>0}$*;*

(ii) $\mathrm{s} \mapsto \sigma(\delta,\mathrm{s})$ *is strictly decreasing for every* $\delta \in [0,\delta_0]$*;*

(iii) $\mathrm{s} \mapsto \sigma(\delta,\mathrm{s})$ *is continuous for every* $\delta \in [0,\delta_0]$*;*

(iv) $\lim_{\mathrm{s}\to\infty} \sigma(\delta,\mathrm{s}) = 0$ *for* $\delta \in [0,\delta_0]$*;*

(v) $\mathrm{s} = \tau(\delta,\sigma(\delta,\mathrm{s})) > \overline{\mathrm{T}}(\delta,\sigma(\delta,\mathrm{s}))$ *for every* $\delta \in [0,\delta_0]$.

Proof (i) and (ii) follows from parts (iv) and (v) of Lemma 2.

(iii) This follows from Lemma 2(ii).

(iv) This follows from Lemma 2(iii).

(v) This follows from Lemma 2(vi). ▼

To complete the proof, we let $\delta_0 \in \mathbb{R}_{>0}$ be as above and define

$$\begin{aligned}\beta\colon [0,\delta)\times[t_0,\infty) &\to \mathbb{R}_{\geq 0}\\ (\delta,t) &\mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2},t-t_0)}.\end{aligned}$$

The following lemma gives the essential feature of β.

4 Lemma $\beta \in \mathcal{KL}([0, \frac{\delta_0}{2}) \times [t_0, \infty); \mathbb{R}_{\geq 0})$.

Proof For fixed $t \in [t_0, \infty)$, the function

$$\delta \mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2}, t - t_0)}$$

is in $\mathcal{K}([0, \frac{\delta_0}{2}); \mathbb{R}_{\geq 0})$ because:

1. $\delta \mapsto \alpha(\delta)$ is continuous and strictly increasing since $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$;
2. the product of strictly increasing functions is and strictly increasing;
3. $x \mapsto \sqrt{x}$ is continuous and strictly increasing on $\mathbb{R}_{\geq 0}$;
4. the composition of continuous strictly increasing functions is continuous and strictly increasing;
5. $\alpha(0) = 0$ since $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$.

For fixed $\delta \in [0, \frac{\delta_0}{2})$, the function

$$t \mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2}, t - t_0)}$$

is in $\mathcal{L}([t_0, \infty); \mathbb{R}_{\geq 0})$ because:

1. $t \mapsto \sigma(\delta, t - t_0)$ is continuous and strictly decreasing by parts (ii) and (iii) of Lemma 3;
2. $\lim_{t\to\infty} \sigma(\delta, t - t_0) = 0$ by Lemma 3(iv). ▼

Now let $\boldsymbol{x} \in U$ satisfy $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \frac{\delta_0}{2}$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}.$$

Then

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|), \qquad t \geq t_0.$$

Also, for $t > t_0$ and $\delta \in [0, \frac{\delta_0}{2}]$, if $\boldsymbol{x} \in U$ satisfies $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, and if $\boldsymbol{\xi}$ is the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x},$$

then we have

$$t - t_0 = \tau(\delta, \sigma(\delta, t - t_0)) \geq \overline{T}(\delta, \sigma(\delta, t - t_0)) + \frac{\delta}{t - t_0} > \overline{T}(\delta, \sigma(\delta, t - t_0)).$$

By definition of $\overline{T}$, this means that

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \sigma(\delta, t - t_0).$$

Continuity of σ in the second argument means that this relation holds, not just for $t > t_0$, but for $t \geq t_0$. Combining the inequalities

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|), \quad \|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \sigma(\delta, t - t_0) < \sigma(\tfrac{\delta_0}{2}, t - t_0)$$

which we have shown to hold for $(t_0, x) \in \mathbb{T} \times U$ satisfying $\|x - x_0\| < \frac{\delta_0}{2}$ and for $t \geq t_0$, we have

$$\|\xi(t) - x_0\| \leq \sqrt{\alpha(\|x - x_0\|)\sigma(\tfrac{\delta_0}{2}, t - t_0)} = \beta(\|x - x_0\|, t - t_0),$$

which gives this part of the lemma.

(iv) First suppose that there exist $\delta \in \mathbb{R}_{>0}$ and $\beta \in \mathcal{KL}([0, \delta) \times [0, \infty); \mathbb{R}_{\geq 0})$ such that, if $(t_0, x) \in \mathbb{T} \times U$ satisfy $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t) - \xi_0(t)\| \leq \beta(\|x - x_0\|, t - t_0)$ for $t \geq t_0$. Let δ and β be as above. If $(t_0, x) \in \mathbb{T} \times U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - \xi_0(t)\| \leq \beta(\|x - x_0\|, t - t_0) \leq \beta(\|x - x_0\|, 0)$$

for $t \geq t_0$. By (ii) we conclude that x_0 is uniformly stable. Also, let $\epsilon \in \mathbb{R}_{>0}$ and let $T \in \mathbb{R}_{>0}$ be sufficiently large that $\beta(\frac{\delta}{2}, T) < \epsilon$. Then, if $(t_0, x) \in \mathbb{T} \times U$ satisfy $\|x - x_0\| < \frac{\delta}{2}$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - x_0\| \leq \beta(\tfrac{\delta}{2}, t - t_0) \leq \beta(\tfrac{\delta}{2}, T) < \epsilon$$

for $t \geq t_0 + T$. This gives uniform asymptotic stability of x_0.

Next suppose that x_0 is uniformly asymptotically stable. Since x_0 is uniformly stable (by definition), by part (ii) there exists $\delta_0 \in \mathbb{R}_{>0}$ and $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$ such that, for $\delta \in [0, \delta_0]$, if $(t_0, x) \in \mathbb{T} \times U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ of the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies

$$\|\xi(t) - x_0\| \leq \alpha(\|x - x_0\|) < \alpha(r).$$

Now, if $(\delta, \epsilon) \in [0, \delta_0] \times \mathbb{R}_{>0}$, then let $A(\delta, \epsilon) \subseteq \mathbb{R}_{>0}$ be the set of $T \in \mathbb{R}_{>0}$ such that, if $(t_0, x) \in \mathbb{T} \times U$ satisfies $\|x - x_0\| < \delta$, then the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies $\|\xi(t) = x_0\| < \epsilon$ for $t \geq t_0 + T$, this being possible by uniform asymptotic stability. Then define $\overline{T}(\delta, \epsilon) = \inf A(\delta, \epsilon)$.

The properties of Lemma 1 also hold for $\overline{T}$ in this case. For $(\delta, \epsilon) \in [0, \delta_0] \times \mathbb{R}_{>0}$, define

$$\tau(\delta, \epsilon) = \frac{2}{\epsilon} \int_{\epsilon/2}^{\epsilon} \overline{T}(\delta, x), \mathrm{d}x + \frac{\delta}{\epsilon}.$$

The properties of Lemma 2 also hold for τ in this case. Now, for $(\delta, s) \in [0, \delta_0] \times \mathbb{R}_{>0}$, define $\sigma(\delta, s) \in \mathbb{R}_{\geq 0}$ by asking that $\sigma(\delta, \tau(\delta, \epsilon)) = \epsilon$, i.e., $s \mapsto \sigma(\delta, s)$ is the inverse of $\epsilon \mapsto \tau(\delta, \epsilon)$. The properties of Lemma 3 also hold for σ in this case.

To complete the proof, we let $\delta_0 \in \mathbb{R}_{>0}$ be as above and define

$$\begin{aligned} \beta \colon [0, \delta) \times \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0} \\ (\delta, s) &\mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2}, s)}. \end{aligned}$$

The following lemma gives the essential feature of β.

5 Lemma $\beta \in \mathcal{KL}([0, \frac{\delta_0}{2}) \times \mathbb{R}_{\geq 0}; \mathbb{R}_{\geq 0})$.

Proof For fixed $s \in \mathbb{R}_{\geq 0}$, the function

$$\delta \mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2}, s)}$$

is in $\mathcal{K}([0, \frac{\delta_0}{2}); \mathbb{R}_{\geq 0})$ because:

1. $\delta \mapsto \alpha(\delta)$ is continuous and strictly increasing since $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$;
2. the product of strictly increasing functions is and strictly increasing;
3. $x \mapsto \sqrt{x}$ is continuous and strictly increasing on $\mathbb{R}_{\geq 0}$;
4. the composition of continuous strictly increasing functions is continuous and strictly increasing;
5. $\alpha(0) = 0$ since $\alpha \in \mathcal{K}([0, \delta_0); \mathbb{R}_{\geq 0})$.

For fixed $\delta \in [0, \frac{\delta_0}{2})$, the function

$$s \mapsto \sqrt{\alpha(\delta)\sigma(\tfrac{\delta_0}{2}, s)}$$

is in $\mathcal{L}(\mathbb{R}_{\geq 0}; \mathbb{R}_{\geq 0})$ because:

1. $s \mapsto \sigma(\delta, s)$ is continuous and strictly decreasing by parts (ii) and (iii) of Lemma 3;
2. $\lim_{s \to \infty} \sigma(\delta, s) = 0$ by Lemma 3(iv). ▼

Now let $(t_0, \boldsymbol{x}) \in \mathbb{T} \times U$ satisfy $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \frac{\delta_0}{2}$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}.$$

Then

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|), \qquad t \geq t_0.$$

Also, for $t > t_0$ and $\delta \in [0, \frac{\delta_0}{2}]$, if $(t_0, \boldsymbol{x}) \in \mathbb{T} \times U$ satisfies $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$, and if $\boldsymbol{\xi}$ is the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x},$$

then we have

$$t - t_0 = \tau(\delta, \sigma(\delta, t - t_0)) \geq \overline{T}(\delta, \sigma(\delta, t - t_0)) + \frac{\delta}{t - t_0} > \overline{T}(\delta, \sigma(\delta, t - t_0)).$$

By definition of $\overline{T}$, this means that

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \sigma(\delta, t - t_0).$$

Continuity of σ in the second argument means that this relation holds, not just for $t > t_0$, but for $t \geq t_0$. Combining the inequalities

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|), \quad \|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \sigma(\delta, t - t_0) < \sigma(\tfrac{\delta_0}{2}, t - t_0)$$

which we have shown to hold for $(t_0, \boldsymbol{x}) \in \mathbb{T} \times U$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \frac{\delta_0}{2}$ and for $t \geq t_0$, we have

$$\|\boldsymbol{\xi}(t) - \boldsymbol{x}_0\| \leq \sqrt{\alpha(\|\boldsymbol{x} - \boldsymbol{x}_0\|)\sigma(\tfrac{\delta_0}{2}, t - t_0)} = \beta(\|\boldsymbol{x} - \boldsymbol{x}_0\|, t - t_0),$$

which gives this part of the lemma. ■

4.3.3 The Second Method for nonautonomous equations

Now, after that lengthy diversion concerning sort of elementary properties of functions, we come to Lyapunov's Section Method. We shall consider this method in four settings, nonautonomous/autonomous and nonlinear/linear. We begin with the most general setting, that for nonautonomous nonlinear equations.

In Lyapunov's Second Method, we will need to evaluate the derivative of a function along the solutions of an ordinary differential equation. To facilitate this, we make the following definition.

4.3.20 Definition (Lie derivative of a function along an ordinary differential equation) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \to \mathbb{R}^n$$

and let $f \colon \mathbb{T} \times U \to \mathbb{R}$ be of class C^1. The ***Lie derivative*** of f along $\boldsymbol{F}$ is

$$\begin{aligned} \mathscr{L}_{\boldsymbol{F}} f \colon \mathbb{T} \times U &\to \mathbb{R} \\ (t, \boldsymbol{x}) &\mapsto \frac{\partial f}{\partial t}(t, \boldsymbol{x}) + \sum_{j=1}^{n} \widehat{F}_j(t, \boldsymbol{x}) \frac{\partial f}{\partial x_j}(t, \boldsymbol{x}). \end{aligned}$$

•

4.3.21 Lemma (Essential property of the Lie derivative I) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U}\to\mathbb{R}^{\mathrm{n}}$$

and let $\mathrm{f}\colon \mathbb{T}\times \mathrm{U}\to\mathbb{R}$ *be of class* C^1. *If* $\xi\colon \mathbb{T}'\to \mathrm{U}$ *is a solution for* $\mathbf{F}$, *then*

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{f}(\mathrm{t},\xi(\mathrm{t}))=\mathscr{L}_{\mathrm{F}}\mathrm{f}(\mathrm{t},\xi(\mathrm{t})).$$

Proof Using the Chain Rule and the fact that

$$\dot{\xi}(t)=\widehat{F}(t,\xi(t)),$$

we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f(t,\xi(t))&=\frac{\partial f}{\partial t}(t,\xi(t))+\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(t,\xi(t))\frac{\mathrm{d}\xi_j}{\mathrm{d}t}(t)\\&=\frac{\partial f}{\partial t}(t,\xi(t))+\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(t,\xi(t))\widehat{F}_j(t,\xi(t))\\&=\mathscr{L}_{\mathbf{F}}f(t,\xi(t)),\end{aligned}$$

as desired. ■

We collect our basic results on Lyapunov's Second Method in this case in the following result.

4.3.22 Theorem (Lyapunov's Second Method for nonautonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U}\to\mathbb{R}^{\mathrm{n}}$$

and let $\mathbf{x}_0\in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$. *Assume that* $\sup\mathbb{T}=\infty$ *and that* $\mathbf{F}$ *satisfies Assumption 4.1.1. Then the following statements hold.*

(i) The equilibrium point $\mathbf{x}_0$ *is stable if there exists* $\mathrm{V}\colon \mathbb{T}\times \mathrm{U}\to\mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1;

(b) $\mathrm{V}\in \mathrm{TVLPD}(\mathbf{x}_0)$;

(c) $-\mathscr{L}_{\mathbf{F}}\mathrm{V}\in \mathrm{TVLPSD}(\mathbf{x}_0)$.

(ii) The equilibrium point $\mathbf{x}_0$ *is uniformly stable if there exists* $\mathrm{V}\colon \mathbb{T}\times \mathrm{U}\to\mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1;

(b) $\mathrm{V}\in \mathrm{TVLPD}(\mathbf{x}_0)$;

(c) $\mathrm{V}\in \mathrm{TVLD}(\mathbf{x}_0)$;

(d) $-\mathscr{L}_{\mathbf{F}}\mathrm{V} \in \mathrm{TVLPSD}(\mathbf{x}_0)$.

(iii) *The equilibrium point* $\mathbf{x}_0$ *is asymptotically stable if there exists* $\mathrm{V}\colon \mathbb{T}\times\mathrm{U}\to\mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1*;*

(b) $\mathrm{V} \in \mathrm{TVLPD}(\mathbf{x}_0)$*;*

(c) $-\mathscr{L}_{\mathbf{F}}\mathrm{V} \in \mathrm{TVLPD}(\mathbf{x}_0)$.

(iv) *The equilibrium point* $\mathbf{x}_0$ *is uniformly asymptotically stable if there exists* $\mathrm{V}\colon \mathbb{T}\times\mathrm{U}\to\mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1*;*

(b) $\mathrm{V} \in \mathrm{TVLPD}(\mathbf{x}_0)$*;*

(c) $\mathrm{V} \in \mathrm{TVLD}(\mathbf{x}_0)$*;*

(d) $-\mathscr{L}_{\mathbf{F}}\mathrm{V} \in \mathrm{TVLPD}(\mathbf{x}_0)$.

Proof (i) Let $t_0 \in \mathbb{T}$. Let $r \in \mathbb{R}_{>0}$ be such that

1. $\overline{\mathsf{B}}(2r, x_0) \subseteq U$,
2. $V \in \mathrm{TVLPD}_{2r}(x_0)$, and
3. $-\mathscr{L}_{\mathbf{F}}V \in \mathrm{TVLPSD}_{2r}(x_0)$.

By definition of time-varying locally positive, let $f \in \mathrm{LPD}_{2r}(x_0)$ be such that

$$f(x) \le V(t,x) \tag{4.16}$$

for all $(t,x) \in \mathbb{T}\times\overline{\mathsf{B}}(r,x_0)$. Also let $g \in \mathrm{LPSD}_r(x_0)$ be such that

$$\mathscr{L}_F V(t,x) \le -g(x) \le 0$$

for $(t,x) \in \mathbb{T}\times\overline{\mathsf{B}}(r,x_0)$. Let $c \in \mathbb{R}_{>0}$ be such that

$$c < \inf\{f(x) \mid \|x - x_0\| = r\}$$

and then define

$$f^{-1}(\le c) = \{x \in \overline{\mathsf{B}}(r,x_0) \mid f(x) \le c\}.$$

Also, for $t \in \mathbb{T}$, denote

$$V_t^{-1}(\le c) = \{x \in \overline{\mathsf{B}}(r,x_0) \mid V(t,x) \le c\}.$$

By (4.16), we have

$$V_t^{-1}(\le c) \subseteq f^{-1}(\le c) \subseteq \mathsf{B}(r,x_0), \qquad t \in \mathbb{T}.$$

Define $\alpha_2\colon [0,2r] \to \mathbb{R}$ by

$$\beta(s) = \sup\{V(t_0,x) \mid \|x - x_0\| \le s\}.$$

A reference to the proof of Lemma 4.3.7(i) gives $\alpha_2 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ such that and

$$V(t_0, \boldsymbol{x}) \leq \beta(\|\boldsymbol{x} - \boldsymbol{x}_0\|) \leq \alpha_2(\|\boldsymbol{x} - \boldsymbol{x}_0\|), \qquad \boldsymbol{x} \in \overline{\mathsf{B}}(r, \boldsymbol{x}_0).$$

Note that $\lim_{s \to 0} \alpha_2(s) = 0$, and so there exists $\delta \in \mathbb{R}_{>0}$ such that $\alpha_2(s) < c$ for $s \in [0, \delta]$. Note that

$$\boldsymbol{x} \in \mathsf{B}(\delta, \boldsymbol{x}_0) \implies V(t_0, \boldsymbol{x}) \leq c.$$

Let $\boldsymbol{x} \in \mathsf{B}(\delta, \boldsymbol{x}_0)$ and let $\boldsymbol{\xi}$ be the solution to the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{F}}(t, \boldsymbol{\xi}(t)), \quad \boldsymbol{\xi}(t_0) = \boldsymbol{x}. \tag{4.17}$$

The following technical lemmata are required to proceed with the proof, and will recur a number of times for proofs relating to Lyapunov's Second Method.

1 Lemma *The solution $\boldsymbol{\xi}$ satisfies $\boldsymbol{\xi}(\mathrm{t}) \in \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0)$ for $\mathrm{t} \geq \mathrm{t}_0$.*

Proof Suppose this is not true. Then, by continuity of $\boldsymbol{\xi}$, there exists a largest $T \in \mathbb{R}_{>0}$ such that $\boldsymbol{\xi}(t) \in \overline{\mathsf{B}}(r, \boldsymbol{x}_0)$ for all $t \in [t_0, t_0 + T]$. This implies, by continuity of $t \mapsto V(t, \boldsymbol{\xi}(t))$, that

$$\|\boldsymbol{\xi}(T) - \boldsymbol{x}_0\| = r. \tag{4.18}$$

Using the facts that

$$\boldsymbol{x} \in \mathsf{B}(\delta, \boldsymbol{x}_0) \subseteq V_{t_0}^{-1}(\leq c) \subseteq \mathsf{B}(r, \boldsymbol{x}_0),$$

and that

$$\frac{\mathrm{d}}{\mathrm{d}t} V(t, \boldsymbol{\xi}(t)) = \mathscr{L}_{\boldsymbol{F}} V(t, \boldsymbol{\xi}(t)) \leq 0, \qquad t \in [t_0, t_0 + T]$$

(the leftmost equality by Lemma 4.3.21), we have

$$\begin{aligned} V(T, \boldsymbol{\xi}(T)) &= V(t_0, \boldsymbol{\xi}(t_0)) + \int_{t_0}^{T} V(t, \boldsymbol{\xi}(t)) \, \mathrm{d}t \\ &= V(t_0, \boldsymbol{\xi}(t_0)) + \int_{t_0}^{T} \mathscr{L}_{\boldsymbol{F}} V(t, \boldsymbol{\xi}(t)) \, \mathrm{d}t < c. \end{aligned} \tag{4.19}$$

However, this contradicts (4.18) and the definition of c, and so we conclude the lemma. ▼

The next lemma we state in some generality, since it asserts a generally useful fact.

2 Lemma *Let* **F** *be an ordinary differential equation whose right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$

satisfies $\sup \mathbb{T} = \infty$ *and Assumption 4.1.1. Let* $\mathrm{K} \subseteq \mathrm{U}$ *be compact and assume that, for every* $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T} \times \mathrm{K}$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x}_0,$$

satisfies $\xi(\mathrm{t}) \in \mathrm{K}$ *for* $\mathrm{t} \geq \mathrm{t}_0$.

Then, for every $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T} \times \mathrm{K}$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x}_0,$$

is defined on $[\mathrm{t}_0, \infty)$.

Proof Suppose the hypotheses of the lemma hold, but the conclusions do not. Thus there exists $(t_0, \boldsymbol{x}) \in \mathbb{T} \times K$ for which the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x}_0, \tag{4.20}$$

in not defined for all $t \in [t_0, \infty)$. Then there exists a largest $T \in \mathbb{R}_{>0}$ such that the solution of the initial value problem is defined on $[t_0, t_0 + T)$. Let $(t_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $[t_0, t_0 + T)$ converging to $t_0 + T$. By the Bolzano–Weierstrass Theorem, the sequence $(\xi(t_j))_{j \in \mathbb{Z}_{>0}}$ has a convergent subsequence $(\xi(t_{j_k}))_{k \in \mathbb{Z}_{>0}}$:

$$\lim_{k \to \infty} \xi(t_{j_k}) = \boldsymbol{y} \in K.$$

Now, by Theorem 1.4.8(ii), there exists $\epsilon \in \mathbb{R}_{>0}$ such that the solution η to the initial value problem

$$\dot{\eta}(t) = \widehat{\boldsymbol{F}}(t, \eta(t)), \quad \eta(t_0 + T) = \boldsymbol{y},$$

is defined on $t \in [t_0 + T - \epsilon, t_0 + T + \epsilon]$. Moreover, by assumption, $\eta(t) \in K$ for every $t \in [t_0 + T - \epsilon, t_0 + T + \epsilon]$. Define $\overline{\xi}\colon [t_0, t_0 + T + \epsilon] \to K$ by

$$\overline{\xi}(t) = \begin{cases} \xi(t), & t \in [t_0, t_0 + T), \\ \eta(t), & t \in [t_0 + T, t_0 + T + \epsilon]. \end{cases}$$

Note, then, that $\overline{\xi}$ is a solution to the differential equation and satisfies the initial condition $\overline{\xi}(t_0) = \boldsymbol{x}$. Thus we have arrived at a contradiction to the solution to the initial value problem (4.20) being defined only on $[t_0, t_0 + T)$. ▼

By combining the preceding two lemmata, we conclude that the solution ξ to the initial value problem (4.17) with $\boldsymbol{x} \in \mathsf{B}(\delta, x_0)$ satisfies (1) $\xi(t) \in \overline{\mathsf{B}}(r, \boldsymbol{x}_0)$ for all $t \geq t_0$ and (2) it is defined on $[t_0, \infty)$. Moreover, by the computation (4.19),

$$\xi(t) \in V_t^{-1}(\leq c) \subseteq f^{-1}(\leq c).$$

By Lemma 4.3.7, there exists $\alpha_1 \in \mathcal{K}([0,2r);\mathbb{R}_{\geq 0})$ such that

$$\alpha_1(\|x - x_0\|) \leq f(x), \qquad x \in \overline{\mathsf{B}}(r, x_0).$$

Let $x \in \mathsf{B}(\delta, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x.$$

Since $x \in \mathsf{B}(\delta, x_0)$, our arguments above imply that ξ is defined on $[t_0, \infty)$ and that $\xi(t) \in \overline{\mathsf{B}}(r, x_0)$ for $t \geq t_0$. Moreover,

$$\alpha_1(\|\xi(t) - x_0\|) \leq f_1(\xi(t)) \leq V(t, \xi(t)) \leq V(t_0, \xi(t_0)) \leq \alpha_2(\|\xi(t_0) - x_0\|)$$

for $t \geq t_0$. Thus

$$\|\xi(t) - x_0\| \leq \alpha_1^{-1} \circ \alpha_2(\|\xi(t_0) - x_0\|)$$

for $t \geq t_0$. Since $\alpha_1^{-1} \circ \alpha_2 \in \mathcal{K}([0,2r);\mathbb{R}_{\geq 0})$ by Lemma 4.3.3, we can now conclude uniform stability from Lemma 4.3.19(ii).

(ii) Let $r \in \mathbb{R}_{>0}$ be such that

1. $\overline{\mathsf{B}}(2r, x_0) \subseteq U$,
2. $V \in \mathrm{TVLPD}_{2r}(x_0)$,
3. $V \in \mathrm{TVLD}_{2r}(x_0)$, and
4. $-\mathscr{L}_F V \in \mathrm{TVLPSD}_{2r}(x_0)$.

By definition of time-varying locally positive and locally decrescent, let $f_1, f_2 \in \mathrm{LPD}_{2r}(x_0)$ be such that

$$f_1(x) \leq V(t,x) \leq f_2(x) \tag{4.21}$$

for all $(t,x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0)$. Also let $g \in \mathrm{LPSD}_r(x_0)$ be such that

$$\mathscr{L}_F V(t,x) \leq -g(x) \leq 0$$

for $(t,x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0)$. Let $c \in \mathbb{R}_{>0}$ be such that

$$c < \inf\{f_1(x) \mid \|x - x_0\| = r\}$$

and then define

$$f_1^{-1}(\leq c) = \{x \in \overline{\mathsf{B}}(r, x_0) \mid f_1(x) \leq c\}$$

and

$$f_2^{-1}(\leq c) = \{x \in \overline{\mathsf{B}}(r, x_0) \mid f_2(x) \leq c\}.$$

Also, for $t \in \mathbb{T}$, denote

$$V_t^{-1}(\leq c) = \{x \in \overline{\mathsf{B}}(r, x_0) \mid V(t,x) \leq c\}.$$

By (4.21), we have

$$f_2^{-1}(\leq c) \subseteq V_t^{-1}(\leq c) \subseteq f_1^{-1}(\leq c) \subseteq \mathsf{B}(r, x_0), \qquad t \in \mathbb{T}.$$

Let $x \in f_2^{-1}(\leq c)$, let $t_0 \in \mathbb{T}$, and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x. \tag{4.22}$$

The following lemma is an adaptation of Lemma 1 to our current setting.

3 Lemma *The solution ξ satisfies $\xi(t) \in \overline{B}(r, x_0)$ for $t \geq t_0$.*

Proof Suppose this is not true. Then, by continuity of ξ, there exists a largest $T \in \mathbb{R}_{>0}$ such that $\xi(t) \in \overline{B}(r, x_0)$ for all $t \in [t_0, t_0 + T]$. This implies, by continuity of $t \mapsto V(t, \xi(t))$, that

$$\|V(T, \xi(T)) - x_0\| = r. \tag{4.23}$$

Using the facts that

$$x \in f_2^{-1}(t_0, \leq c) \subseteq V_{t_0}^{-1}(\leq c) \subseteq B(r, x_0),$$

and that

$$\frac{d}{dt} V(t, \xi(t)) = \mathscr{L}_F V(t, \xi(t)) \leq 0, \qquad t \in [t_0, t_0 + T]$$

(the leftmost equality by Lemma 4.3.21), we have

$$\begin{aligned} V(T, \xi(T)) &= V(t_0, \xi(t_0)) + \int_{t_0}^{T} V(t, \xi(t)) \, dt \\ &= V(t_0, \xi(t_0)) + \int_{t_0}^{T} \mathscr{L}_F V(t, \xi(t)) \, dt < c. \end{aligned} \tag{4.24}$$

However, this contradicts (4.23) and the definition of c, and so we conclude the lemma. ▼

By combining the preceding lemma with Lemma 2, we conclude that the solution ξ to the initial value problem (4.22) with $x \in f_2^{-1}(\leq c)$ satisfies (1) $\xi(t) \in \overline{B}(r, x_0)$ for all $t \geq t_0$ and (2) it is defined on $[t_0, \infty)$. Moreover, by the computation (4.24),

$$\xi(t) \in V_t^{-1}(\leq c) \subseteq f_1^{-1}(\leq c).$$

By Lemma 4.3.7, there exist $\alpha_1, \alpha_2 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ such that

$$\alpha_1(\|x - x_0\|) \leq f_1(x), \quad f_2(x) \leq \alpha_2(\|x - x_0\|), \qquad x \in \overline{B}(r, x_0).$$

Now let $\delta \in (0, r]$ be sufficiently small that $\alpha_2(s) \leq c$ for $s \in [0, \delta]$. Note that

$$x \in B(\delta, x_0) \implies \alpha_2(\|x - x_0\|) \leq c \implies x \in f_2^{-1}(\leq c).$$

Let $(t_0, x) \in \mathbb{T} \times B(\delta, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x.$$

Since $x \in f_2^{-1}(\leq c)$, our arguments above imply that ξ is defined on $[t_0, \infty)$ and that $\xi(t) \in \overline{B}(r, x_0)$ for $t \geq t_0$. Moreover,

$$\alpha_1(\|\xi(t) - x_0\|) \leq f_1(\xi(t)) \leq V(t, \xi(t)) \leq V(t_0, \xi(t_0)) \leq f_2(\xi(t_0)) \leq \alpha_2(\|\xi(t_0) - x_0\|)$$

for $t \geq t_0$. Thus

$$\|\xi(t) - x_0\| \leq \alpha_1^{-1} \circ \alpha_2(\|\xi(t_0) - x_0\|)$$

for $t \geq t_0$. Since $\alpha_1^{-1} \circ \alpha_2 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ by Lemma 4.3.3, we can now conclude uniform stability from Lemma 4.3.19(ii).

(iii) Let $t_0 \in \mathbb{T}$. Let $r \in \mathbb{R}_{>0}$ be such that

1. $\overline{\mathsf{B}}(2r, x_0) \subseteq U$,
2. $V \in \mathrm{TVLPD}_{2r}(x_0)$,
3. $-\mathscr{L}_F V \in \mathrm{TVLPD}_{2r}(x_0)$.

As in the proof of part (i), we let $f_1 \in \mathrm{LPD}_{2r}(x_0)$ and $\alpha_1 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ be such that

$$\alpha_1(\|x - x_0\|) \leq f_1(x) \leq V(t, x)$$

for $(t, x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0)$. Also as in the proof of part (i), let $\alpha_2 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ be such that

$$V(t_0, x) \leq \alpha_2(\|x - x_0\|), \qquad x \in \overline{\mathsf{B}}(r, x_0).$$

Also let $f_3 \in \mathrm{LPD}_{2r}(x_0)$ and $\alpha_3 \in \mathcal{K}([0, 2r); \mathbb{R}_{\geq 0})$ be such that

$$\alpha_3(\|x - x_0\|) \leq f_3(x) \leq -\mathscr{L}_F V(t, x)$$

for $(t, x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0)$.

Of course, we then conclude stability of x_0 from part (i). We then have

$$V(t, x) \leq \alpha_2(\|x - x_0\|) \implies \alpha_3 \circ \alpha_2^{-1} \circ V(t, x) \leq \alpha_3(\|x - x_0\|) \tag{4.25}$$

for $(t, x) \in \mathbb{T} \times \overline{\mathsf{B}}(r, x_0)$. By Lemma 4.3.3 we have $\alpha_3 \circ \alpha_2^{-1} \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ and, therefore, by Lemma 4.3.4, there exists a locally Lipschitz $\alpha \in \mathcal{K}([0, r); \mathbb{R}_{\geq 0})$ such that $\alpha(s) \leq \alpha_3 \circ \alpha_2^{-1}(x)$ for all $x \in [0, r)$. Now let δ be as in the proof of part (i). Let $x \in \overline{\mathsf{B}}(\delta, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x.$$

Recall that

1. $\xi(t) \in \overline{\mathsf{B}}_r(x_0,)$ for all $t \in [t_0, \infty)$ by Lemma 1,
2. $V(t, \xi(t)) \leq c$ for all $t \in [t_0, \infty)$ by definition of δ.

Using Lemma 4.3.21 and (4.25), we then have

$$\frac{\mathrm{d}}{\mathrm{d}t} V(t, \xi(t)) = \mathscr{L}_F V(t, \xi(t)) \leq -\alpha_3(\|\xi(t) - x_0\|) \leq -\alpha \circ V(t, \xi(t)).$$

The following technical lemma is now required.

4 Lemma *Let* F *be a scalar ordinary differential equation with right-hand side*

$$\widehat{\mathrm{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}$$

where $\mathrm{U} \subseteq \mathbb{R}$ *is open. For* $(\mathrm{t}_0, \mathrm{y}_0) \in \mathbb{T} \times \mathrm{U}$*, let* $\xi, \eta\colon \mathbb{T}' \to \mathrm{U}$ *be of class* C^1 *and satisfy*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathrm{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathrm{y}_0$$

and

$$\dot{\eta}(\mathrm{t}) < \widehat{\mathrm{F}}(\mathrm{t}, \boldsymbol{\eta}(\mathrm{t})), \quad \eta(\mathrm{t}_0) = \mathrm{y}_0.$$

Then $\eta(\mathrm{t}) < \xi(\mathrm{t})$ *for* $\mathrm{t} > \mathrm{t}_0$.

Proof We have

$$\dot{\eta}(t_0) < \widehat{F}(t_0, y_0) = \dot{\xi}(t_0).$$

Therefore, by continuity of the derivatives, there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$\dot{\eta}(t) < \dot{\xi}(t), \qquad t \in [t_0, t_0 + \epsilon].$$

Therefore, for $t \in (t_0, t_0 + \epsilon]$,

$$\eta(t) = \int_{t_0}^{t} \dot{\eta}(\tau)\,\mathrm{d}\tau < \int_{t_0}^{t} \dot{\xi}(\tau)\,\mathrm{d}\tau = \xi(t).$$

Now suppose that it does not hold that $\eta(t) < \xi(t)$ for all $t \geq t_0$. Then let

$$T = \inf\{t \geq t_0 \mid \eta(t) \geq \xi(t)\} > t_0 + \epsilon.$$

By continuity, $\eta(T) = \xi(T)$. Thus

$$\begin{aligned} \dot{\eta}(T) &= \underbrace{\dot{\eta}(T) - \widehat{F}(T, \eta(T))}_{<0} + \widehat{F}(T, \eta(T)) \\ &< \underbrace{\dot{\xi}(t) - \widehat{F}(T, \xi(T))}_{=0} + \widehat{F}(T, \xi(T)) = \dot{\xi}(T). \end{aligned}$$

On the other hand, for $h \in \mathbb{R}_{>0}$ (sufficiently small for the expression to be defined) we have

$$\frac{\eta(T) - \eta(T-h)}{h} > \frac{\xi(T) - \xi(T-h)}{h},$$

and taking the limit as $h \to 0$ gives $\dot{\eta}(T) \geq \dot{\xi}(T)$, contradicting our computation just proceeding. ▼

By Lemma 4.3.5, there exists $\psi \in \mathcal{KL}([0, r) \times [t_0, \infty); \mathbb{R}_{\geq 0})$ such that, if $y \in [0, r)$, then the solution to the initial value problem

$$\dot{\eta}(t) = -\alpha(\eta(t)), \quad \eta(t_0) = y,$$

is $\psi(y,t)$ for $t \geq t_0$. By Lemma 4 we have

$$V(t,\xi(t)) \leq \psi(V(t_0,x),t), \qquad t \geq t_0.$$

Therefore,

$$\begin{aligned}\|\xi(t)-x_0\| &\leq \alpha_1^{-1} \circ \psi(V(t_0,x),t)\\ &\leq \alpha_1^{-1} \circ \psi(\alpha_2(\|x-x_0\|),t).\end{aligned}$$

By Lemma 4.3.3(iii), the mapping

$$\begin{aligned}\beta\colon [0,r) \times [t_0,\infty) &\to \mathbb{R}\\ (s,\tau) &\mapsto \alpha_1^{-1} \circ \psi(\alpha_2(s),\tau)\end{aligned}$$

is of class $\mathcal{KL}$. The asymptotic stability of x_0 now follows from Lemma 4.3.19(iii).

(iv) Let $r \in \mathbb{R}_{>0}$ be such that

1. $\overline{\mathsf{B}}(2r,x_0) \subseteq U$,
2. $V \in \mathrm{TVLPD}_{2r}(x_0)$,
3. $V \in \mathrm{TVLD}_{2r}(x_0)$, and
4. $-\mathscr{L}_F V \in \mathrm{TVLPD}_{2r}(x_0)$.

As in the proof of part (ii), we let $f_1, f_2 \in \mathrm{LPD}_{2r}(x_0)$ and $\alpha_1, \alpha_2 \in \mathcal{K}([0,2r);\mathbb{R}_{\geq 0})$ be such that

$$\alpha_1(\|x-x_0\|) \leq f_1(x) \leq V(t,x) \leq f_2(x) \leq \alpha_2(\|x-x_0\|)$$

for $(t,x) \in \mathbb{T} \times \overline{\mathsf{B}}(r,x_0)$. Also let $f_3 \in \mathrm{LPD}_{2r}(x_0)$ and $\alpha_3 \in \mathcal{K}([0,2r);\mathbb{R}_{\geq 0})$ be such that

$$\alpha_3(\|x-x_0\|) \leq f_3(x) \leq -\mathscr{L}_F V(t,x)$$

for $(t,x) \in \mathbb{T} \times \overline{\mathsf{B}}(r,x_0)$.

Of course, we then conclude uniform stability of x_0 from part (ii). We then have

$$V(t,x) \leq \alpha_2(\|x-x_0\|) \implies \alpha_3 \circ \alpha_2^{-1} \circ V(t,x) \leq \alpha_3(\|x-x_0\|) \tag{4.26}$$

for $(t,x) \in \mathbb{T} \times \overline{\mathsf{B}}(r,x_0)$. By Lemma 4.3.3 we have $\alpha_3 \circ \alpha_2^{-1} \in \mathcal{K}([0,r);\mathbb{R}_{\geq 0})$ and, therefore, by Lemma 4.3.4, there exists a locally Lipschitz $\alpha \in \mathcal{K}([0,r);\mathbb{R}_{\geq 0})$ such that $\alpha(s) \leq \alpha_3 \circ \alpha_2^{-1}(x)$ for all $x \in [0,r)$. Now let δ be as in the proof of part (ii). Let $(t_0,x) \in \mathbb{T} \times \overline{\mathsf{B}}(\delta,x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t,\xi(t)), \quad \xi(t_0) = x.$$

Recall that

1. $\xi(t) \in \overline{\mathsf{B}}_r(x_0,)$ for all $t \in [t_0,\infty)$ by Lemma 3,
2. $V(t,\xi(t)) \leq c$ for all $t \in [t_0,\infty)$ by definition of δ.

Using Lemma 4.3.21 and (4.26), we then have

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t,\xi(t)) = \mathscr{L}_F V(t,\xi(t)) \leq -\alpha_3(\|\xi(t)-x_0\|) \leq -\alpha \circ V(t,\xi(t)).$$

By Lemma 4.3.5, there exists $\psi \in \mathcal{KL}([0,r)\times\mathbb{R}_{\geq 0};\mathbb{R}_{\geq 0})$ such that, if $y \in [0,r)$ and $t_0 \in \mathbb{R}$, then the solution to the initial value problem

$$\dot{\eta}(t) = -\alpha(\eta(t)), \quad \eta(t_0) = y,$$

is $\psi(y, t-t_0)$ for $t \geq t_0$. By Lemma 4 we have

$$V(t,\xi(t)) \leq \psi(V(t_0,x), t-t_0), \qquad t \geq t_0.$$

Therefore,

$$\begin{aligned}\|\xi(t)-x_0\| &\leq \alpha_1^{-1} \circ \psi(V(t_0,x), t-t_0)\\ &\leq \alpha_1^{-1} \circ \psi(\alpha_2(\|x-x_0\|), t-t_0).\end{aligned}$$

By Lemma 4.3.3(iii), the mapping

$$\begin{aligned}\beta\colon [0,r)\times\mathbb{R}_{\geq 0} &\to \mathbb{R}\\ (s,\tau) &\mapsto \alpha_1^{-1} \circ \psi(\alpha_2(s),\tau)\end{aligned}$$

is of class $\mathcal{KL}$. The uniform asymptotic stability of x_0 now follows from Lemma 4.3.19(iv). ■

4.3.23 Terminology The function V in the statement of the preceding theorem is typically called a ***Lyapunov function***. It is not uncommon for this terminology to be used imprecisely, in the sense that when one sees the expression "Lyapunov function," it is clear only from context whether one is in case (i), (ii), (iii), or (iv) of the preceding theorem. Typically this is not to be thought of as confusing, as the context indeed makes this clear. •

We also have the following sufficient condition for exponential stability (as opposed to mere asymptotic stability) which comes with the flavour of Lyapunov's Second Method.

4.3.24 Theorem (Lyapunov's Second Method for exponential stability of nonautonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times\mathsf{U} \to \mathbb{R}^n$$

and let $\mathbf{x}_0 \in \mathsf{U}$ *be an equilibrium point for* $\mathbf{F}$. *Assume that* $\sup\mathbb{T} = \infty$ *and* $\mathbf{F}$ *satisfies Assumption 4.1.1. Then* $\mathbf{x}_0$ *is uniformly exponentially stable if there exists* $\mathsf{V}\colon \mathbb{T}\times\mathsf{U} \to \mathbb{R}$ *with the following properties:*

(i) V *is of class* C^1*;*

(ii) *there exists* $C_1, \alpha_1, r_1 \in \mathbb{R}_{>0}$ *such that*

$$C_1\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_1} \leq V(t,\mathbf{x}) \leq C_1^{-1}\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_1}$$

for all $(t,\mathbf{x}) \in \mathbb{T}\times B(r_1,\mathbf{x}_0)$;

(iii) *there exists* $C_2, \alpha_2, r_2 \in \mathbb{R}_{>0}$ *such that*

$$\mathscr{L}_{\mathbf{F}}V(t,\mathbf{x}) \leq -C_2\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_2}$$

for all $(t,\mathbf{x}) \in \mathbb{T}\times B(r_2,\mathbf{x}_0)$.

Proof Let $r, \alpha \in \mathbb{R}_{>0}$ be such that

1. $C_1\|x-x_0\|^{\alpha} \leq V(t,x) \leq C_1^{-1}\|x-x_0\|^{\alpha}$ for all $(t,x) \in \mathbb{T}\times B(2r,x_0)$ and
2. $-\mathscr{L}_F V(t,x) \geq C_2\|x-x_0\|^{\alpha}$ for all $(t,x) \in \mathbb{T}\times B(r,x_0)$.

Let $c \in \mathbb{R}_{>0}$ be such that

$$c < \inf\{C_1\|x-x_0\|^{\alpha} \mid \|x-x_0\| = r\}.$$

We then let $\delta \in \mathbb{R}_{>0}$ be such that, if $x \in B(\delta,x_0)$, then $C_2\|x-x_0\| \leq c$. Let $(t_0,x) \in \mathbb{T}\times B(\delta,x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t,\xi(t)), \quad \xi(t_0 = x.$$

We then argue as in Lemmata 3 and 2 from the proof of Theorem 4.3.22 that $\xi(t) \in \overline{B}(r,x_0)$ for $t \geq t_0$ and that ξ is defined for all $t \in [t_0,\infty)$. Now compute, using Lemma 4.3.21 and the definitions of C_1, C_2, and α,

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t,\xi(t)) = \mathscr{L}_F V(t,\xi(t)) \leq -C_2\|\xi(t)-x_0\|^{\alpha} \leq -C_1C_2V(t,\xi(t)).$$

By Lemma 4 of Theorem 4.3.22,

$$V(t,\xi(t)) \leq V(t_0,\xi(t_0))\mathrm{e}^{-C_1C_2(t-t_0)}$$

for $t \geq t_0$. Now, again using the definition of C_1 and α,

$$\begin{aligned}\|\xi(t)-x_0\| &\leq \left(\frac{V(t,\xi(t))}{C_1}\right)^{1/\alpha} \leq \left(\frac{V(t_0,\xi(t_0))\mathrm{e}^{-C_1C_2(t-t_0)}}{C_1}\right)^{1/\alpha}\\ &\leq \frac{\|x-x_0\|}{C_1^{2\alpha}}\mathrm{e}^{-C_1C_2(t-t_0)/\alpha}\end{aligned}$$

for all $t \geq t_0$. Recalling that the preceding estimates are valid for any $(t_0,x) \in \mathbb{T}\times B(\delta,x_0)$, we conclude uniform exponential stability of x_0. ■

4.3.4 The Second Method for autonomous equations

In the preceding section we gave a quite general version of Lyapunov's Second Method applied to nonautonomous ordinary differential equations. As can be seen, the proofs are lengthy and a little detailed. Here we consider the simpler autonomous case, for which we give a self-contained proof for a reader wishing for a "light" alternative. In stating the result in this case, we recall from Proposition 4.1.5 that "stability" and "uniform stability" are equivalent, and that "asymptotic stability" and "uniform asymptotic stability" are equivalent for nonautonomous ordinary differential equations.

Before we get to the statement of the main result, we first give the non-time-varying version of the definition of Lie derivative.

4.3.25 Definition (Lie derivative of a function along an autonomous ordinary differential equation) Let $\boldsymbol{F}$ be an autonomous ordinary differential equation with right-hand side

$$\begin{aligned}\widehat{\boldsymbol{F}}\colon \mathbb{T}\times U &\to \mathbb{R}^n\\ (t,\boldsymbol{x}) &\mapsto \widehat{\boldsymbol{F}}_0(\boldsymbol{x}),\end{aligned}$$

and let $f\colon U\to\mathbb{R}$ be of class C^1. The ***Lie derivative*** of f along $\boldsymbol{F}$ is

$$\begin{aligned}\mathscr{L}_{\boldsymbol{F}_0} f\colon U &\to \mathbb{R}\\ \boldsymbol{x} &\mapsto \sum_{j=1}^{n} \widehat{F}_{0,j}(\boldsymbol{x})\frac{\partial f}{\partial x_j}(\boldsymbol{x}).\end{aligned}$$

•

4.3.26 Lemma (Essential property of the Lie derivative II) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} &\to \mathbb{R}^{\mathrm{n}}\\ (\mathrm{t},\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathrm{t},\mathbf{x}),\end{aligned}$$

and let $\mathrm{f}\colon \mathrm{U}\to\mathbb{R}$ *be of class* C^1. *If* $\xi\colon \mathbb{T}'\to \mathrm{U}$ *is a solution for* $\mathbf{F}$, *then*

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{f}(\xi(\mathrm{t})) = \mathscr{L}_{\mathbf{F}_0}\mathrm{f}(\xi(\mathrm{t})).$$

Proof Using the Chain Rule and the fact that

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}_0(\xi(t)),$$

we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f(\xi(t)) &= \sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(\xi(t))\frac{\mathrm{d}\xi_j}{\mathrm{d}t}(t)\\ &= \sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(\xi(t))\widehat{F}_{0,j}(\xi(t))\\ &= \mathscr{L}_{F_0}f(\xi(t)),\end{aligned}$$

as desired. ■

We can now state the main concerning Lyapunov's Second Method in the nonautonomous case.

4.3.27 Theorem (Lyapunov's Second Method for autonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times U &\to \mathbb{R}^n\\ (t,x) &\mapsto \widehat{\mathbf{F}}_0(x),\end{aligned}$$

and let $x_0 \in U$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup\mathbb{T} = \infty$ *and that* $\mathbf{F}$ *satisfies Assumption 4.1.1. Then the following statements hold.*

(i) The equilibrium point x_0 *is stable if there exists* $V\colon U \to \mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1*;*

(b) $V \in \mathrm{LPD}(x_0)$*;*

(c) $-\mathscr{L}_{F_0}V \in \mathrm{LPSD}(x_0)$.

(ii) The equilibrium point x_0 *is asymptotically stable if there exists* $V\colon U \to \mathbb{R}$ *with the following properties:*

(a) V *is of class* C^1*;*

(b) $V \in \mathrm{LPD}(x_0)$*;*

(c) $-\mathscr{L}_{F_0}V \in \mathrm{LPD}(x_0)$.

We shall give two proofs of Theorem 4.3.27, one assuming Theorem 4.3.22 and one independent of that more general theorem.

Proof of Theorem 4.3.27, assuming Theorem 4.3.22 In this case, the theorem is an easy corollary of the more general Theorem 4.3.22. Indeed, the hypotheses of parts (i) and (ii) of Theorem 4.3.27 immediately imply those of parts (ii) and (iv), respectively, of Theorem 4.3.22. ■

Independent proof of Theorem 4.3.27 (i) Let $\epsilon \in \mathbb{R}_{>0}$. Let $r \in (0, \frac{\epsilon}{2}]$ be chosen so that

1. $\overline{\mathsf{B}}(2r, \boldsymbol{x}_0) \subseteq U$,
2. $V \in \mathrm{LPD}_{2r}(\boldsymbol{x}_0)$, and
3. $-\mathscr{L}_{F_0} V \in \mathrm{LPSD}_{2r}(\boldsymbol{x}_0)$.

Let $c \in \mathbb{R}_{>0}$ be such that

$$c < \inf\{V(\boldsymbol{x}) \mid \|\boldsymbol{x} - \boldsymbol{x}_0\| = r\}$$

and define

$$V^{-1}(\leq c) = \{\boldsymbol{x} \in \overline{\mathsf{B}}(r, \boldsymbol{x}_0) \mid V(\boldsymbol{x}) \leq c\}.$$

Then $V^{-1}(\leq c) \subseteq \mathsf{B}(r, \boldsymbol{x}_0)$ by continuity of V and the definition of c. By continuity of V, let $\delta \in \mathbb{R}_{>0}$ be such that, if $\boldsymbol{x} \in \mathsf{B}(\delta, \boldsymbol{x}_0)$, then $V(\boldsymbol{x}) < c$. Therefore, we have

$$\mathsf{B}(\delta, \boldsymbol{x}_0) \subseteq V^{-1}(\leq c) \subseteq \mathsf{B}(r, \boldsymbol{x}_0).$$

Let $(t_0, \boldsymbol{x}) \in \mathbb{T} \times \mathsf{B}(\delta, \boldsymbol{x}_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}_0(\xi(t)), \quad \xi(t_0) = \boldsymbol{x}.$$

The following lemmata, which essentially appear in the proof of Theorem 4.3.22, are repeated here for the purposes of making the proof self-contained.

1 Lemma *The solution ξ satisfies $\xi(\mathrm{t}) \in \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0)$ for $\mathrm{t} \geq \mathrm{t}_0$.*

Proof Suppose this is not true. Then, by continuity of ξ, there exists a largest $T \in \mathbb{R}_{>0}$ such that $\xi(t) \in \overline{\mathsf{B}}(r, \boldsymbol{x}_0)$ for all $t \in [t_0, t_0 + T]$. This implies, by continuity of $t \mapsto V(\xi(t))$, that

$$\|V(\xi(T)) - \boldsymbol{x}_0\| = r. \tag{4.27}$$

Using the facts that

$$\boldsymbol{x} \in \mathsf{B}(\delta, \boldsymbol{x}_0) \subseteq V^{-1}(\leq c) \subseteq \mathsf{B}(r, \boldsymbol{x}_0),$$

and that

$$\frac{\mathrm{d}}{\mathrm{d}t} V(\xi(t)) = \mathscr{L}_{F_0} V(\xi(t)) \leq 0, \qquad t \in [t_0, t_0 + T]$$

(the leftmost equality by Lemma 4.3.26), we have

$$\begin{aligned} V(\xi(T)) &= V(\xi(t_0)) + \int_{t_0}^{T} V(\xi(t))\,\mathrm{d}t \\ &= V(\xi(t_0)) + \int_{t_0}^{T} \mathscr{L}_{F_0} V(\xi(t))\,\mathrm{d}t < c. \end{aligned} \tag{4.28}$$

However, this contradicts (4.27) and the definition of c, and so we conclude the lemma. ▼

2 Lemma *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation whose right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$
$$(\mathrm{t}, \mathbf{x}) \mapsto \widehat{\mathbf{F}}_0(\mathbf{x})$$

satisfies $\sup \mathbb{T} = \infty$ *and Assumption 4.1.1. Let* $\mathrm{K} \subseteq \mathrm{U}$ *be compact and assume that, for every* $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T}\times \mathrm{K}$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}_0(\xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x}_0,$$

satisfies $\xi(\mathrm{t}) \in \mathrm{K}$ *for* $\mathrm{t} \geq \mathrm{t}_0$.

Then, for every $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T}\times \mathrm{K}$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}_0(\xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x}_0,$$

is defined on $[\mathrm{t}_0, \infty)$.

Proof Suppose the hypotheses of the lemma hold, but the conclusions do not. Thus there exists $(t_0, \boldsymbol{x}) \in \mathbb{T}\times K$ for which the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}_0(\xi(t)), \quad \xi(t_0) = \boldsymbol{x}_0, \tag{4.29}$$

in not defined for all $t \in [t_0, \infty)$. Then there exists a largest $T \in \mathbb{R}_{>0}$ such that the solution of the initial value problem is defined on $[t_0, t_0 + T)$. Let $(t_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $[t_0, t_0 + T)$ converging to $t_0 + T$. By the Bolzano–Weierstrass Theorem, the sequence $(\xi(t_j))_{j\in\mathbb{Z}_{>0}}$ has a convergent subsequence $(\xi(t_{j_k}))_{k\in\mathbb{Z}_{>0}}$:

$$\lim_{k\to\infty} \xi(t_{j_k}) = \boldsymbol{y} \in K.$$

Now, by Theorem 1.4.8(ii), there exists $\epsilon \in \mathbb{R}_{>0}$ such that the solution η to the initial value problem

$$\dot{\eta}(t) = \widehat{F}_0(\eta(t)), \quad \eta(t_0 + T) = \boldsymbol{y},$$

is defined on $t \in [t_0 + T - \epsilon, t_0 + T + \epsilon]$. Moreover, by assumption, $\eta(t) \in K$ for every $t \in [t_0 + T - \epsilon, t_0 + T + \epsilon]$. Define $\overline{\xi}\colon [t_0, t_0 + T + \epsilon] \to K$ by

$$\overline{\xi}(t) = \begin{cases} \xi(t), & t \in [t_0, t_0 + T), \\ \eta(t), & t \in [t_0 + T, t_0 + T + \epsilon]. \end{cases}$$

Note, then, that $\overline{\xi}$ is a solution to the differential equation and satisfies the initial condition $\overline{\xi}(t_0) = \boldsymbol{x}$. Thus we have arrived at a contradiction to the solution to the initial value problem (4.29) being defined only on $[t_0, t_0 + T)$. ▼

Since $r \leq \frac{\epsilon}{2} < \epsilon$, the preceding lemma immediately proves stability of $\boldsymbol{x}_0$.

(ii) Let $r, \delta \in \mathbb{R}_{>0}$ be chosen so that

1. $\overline{\mathsf{B}}(2r, x_0) \subseteq U$,
2. $V \in \mathrm{LPD}_{2r}(x_0)$,
3. $-\mathscr{L}_{F_0} V \in \mathrm{LPD}(x_0)$, and
4. if $(t_0, x) \in \mathbb{T} \times \mathsf{B}(\delta, x_0)$ and if ξ is the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}_0(\xi(t)), \quad \xi(t_0) = x,$$

then $\xi(t) \in \overline{\mathsf{B}}(r, x_0)$ for $t \geq t_0$ and ξ is defined on $[t_0, \infty)$.

The last condition is possible by virtue of our arguments in part (i).

Let $(t_0, x) \in \mathbb{T} \times \mathsf{B}(\delta, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}_0(\xi(t)), \quad \xi(t_0) = x.$$

Since $\frac{\mathrm{d}}{\mathrm{d}t} V(\xi(t)) < 0$ for all $t \geq t_0$, it follows that $t \mapsto V(\xi(t))$ is strictly decreasing. Thus, since V is nonnegative, there exists $\gamma \in \mathbb{R}_{\geq 0}$ such that

$$\lim_{t \to \infty} V(\xi(t)) = \gamma.$$

We claim that $\gamma = 0$. Suppose otherwise, and that $\gamma \in \mathbb{R}_{>0}$. Let $\alpha \in \mathbb{R}_{>0}$ be such that, if $x \in \mathsf{B}(\alpha, x_0)$, then $V(x) < \gamma$. Therefore, $\xi(t) \in \overline{\mathsf{B}}(r, x_0) \setminus \mathsf{B}(\alpha, x_0)$. Denote

$$\beta = \inf\{-\mathscr{L}_{F_0} V(x) \mid \|x - x_0\| \in [\alpha, r]\},$$

the infimum existing because it is over a compact set by ***missing stuff***. Moreover, since $\mathscr{L}_{F_0} V$ is negative definite, $\beta \in \mathbb{R}_{>0}$. Now we calculate

$$\begin{aligned} V(\xi(t)) &= V(\xi(t_0)) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\xi(\tau)) \, \mathrm{d}\tau \\ &= V(\xi(t_0)) + \int_{t_0}^{t} \mathscr{L}_{F_0} V(\xi(\tau)) \, \mathrm{d}\tau \\ &\leq V(\xi(t_0)) - \beta(t - t_0). \end{aligned}$$

This implies that $\lim_{t \to \infty} V(\xi(t)) = -\infty$. This contradiction leads us to conclude that $\gamma = 0$.

Finally, we must show that this implies that

$$\lim_{t \to \infty} \|\xi(t) - x_0\| = 0$$

(still supposing ξ to be the solution for initial condition $(t_0, x) \in \mathbb{T} \times \mathsf{B}(\delta, x_0)$). To this end, let $\epsilon \in \mathbb{R}_{>0}$ and let $b \in \mathbb{R}_{>0}$ be such that

$$b < \inf\{V(x) \mid \|x - x_0\| = \epsilon\}.$$

Then, as we argued above that $V^{-1}(\leq c) \subseteq \mathsf{B}(r, x_0)$, here we conclude that $V^{-1}(\leq b) \subseteq \mathsf{B}(\epsilon, x_0)$. Therefore, if we let $T \in \mathbb{R}_{>0}$ be sufficiently large that $V(\xi(t)) \leq b$ for $t \geq T$, then $\xi(t) \in \mathsf{B}(\epsilon, x_0)$ for all $t \geq T$. ■

4.3.28 Terminology The function V in the statement of the preceding theorem is typically called a ***Lyapunov function***. It is not uncommon for this terminology to be used imprecisely, in the sense that when one sees the expression "Lyapunov function," it is clear only from context whether one is in case (i) or (ii) of the preceding theorem. Typically this is not to be thought of as confusing, as the context indeed makes this clear. •

4.3.29 Remark (Automatic implications of Theorem 4.3.27) We recall from Proposition 4.1.5 that uniform stability and stability are equivalent for autonomous ordinary differential equations, and similarly that uniform asymptotic stability and asymptotic stability are equivalent. •

4.3.30 Theorem (Lyapunov's Second Method for exponential stability of autonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} &\to \mathbb{R}^{\mathrm{n}}\\ (\mathrm{t},\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathbf{x}),\end{aligned}$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$ *and* $\mathbf{F}$ *satisfies Assumption 4.1.1. Then* $\mathbf{x}_0$ *is exponentially stable if there exists* $\mathrm{V}\colon \mathrm{U} \to \mathbb{R}$ *with the following properties:*

(i) V *is of class* C^1*;*

(ii) there exist $\mathrm{C}_1, \alpha_1, \mathrm{r}_1 \in \mathbb{R}_{>0}$ *such that*

$$\mathrm{C}_1\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_1} \leq \mathrm{V}(\mathbf{x}) \leq \mathrm{C}_1^{-1}\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_1}$$

for all $\mathbf{x} \in \mathsf{B}(\mathrm{r}_1, \mathbf{x}_0)$*;*

(iii) there exist $\mathrm{C}_2, \alpha_2, \mathrm{r}_2 \in \mathbb{R}_{>0}$ *such that*

$$\mathscr{L}_{\mathbf{F}_0}\mathrm{V}(\mathbf{x}) \leq -\mathrm{C}_2\|\mathbf{x}-\mathbf{x}_0\|^{\alpha_2}$$

for all $\mathbf{x} \in \mathsf{B}(\mathrm{r}_2, \mathbf{x}_0)$.

Proof Let $r, \alpha \in \mathbb{R}_{>0}$ be such that

1. $C_1\|x-x_0\|^\alpha \leq V(x) \leq C_1^{-1}\|x-x_0\|^\alpha$ for all $x \in \mathsf{B}(2r, x_0)$ and
2. $-\mathscr{L}_{F_0}V(x) \geq C_2\|x-x_0\|^\alpha$ for all $x \in \mathsf{B}(r, x_0)$.

Let $c \in \mathbb{R}_{>0}$ be such that

$$c < \inf\{C_1\|x-x_0\|^\alpha \mid \|x-x_0\| = r\}.$$

We then let $\delta \in \mathbb{R}_{>0}$ be such that, if $x \in \mathsf{B}(\delta, x_0)$, then $C_2\|x-x_0\| \leq c$. Let $(t_0, x) \in \mathbb{T}\times\mathsf{B}(\delta, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}_0(\xi(t)), \quad \xi(t_0) = x.$$

We then argue as in Lemmata 1 and 2 from the proof of Theorem 4.3.27 that $\xi(t) \in \overline{B}(r, x_0)$ for $t \geq t_0$ and that ξ is defined for all $t \in [t_0, \infty)$. Now compute, using Lemma 4.3.26 and the definitions of C_1, C_2, and α,

$$\frac{\mathrm{d}}{\mathrm{d}t} V(\xi(t)) = \mathscr{L}_F V(\xi(t)) \leq -C_2 \|\xi(t) - x_0\|^\alpha \leq -C_1 C_2 V(\xi(t)).$$

The following technical lemma is now required.

1 Lemma *Let* F *be an autonomous scalar ordinary differential equation with right-hand side*

$$\widehat{\mathrm{F}} \colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}$$
$$(\mathrm{t}, \mathrm{x}) \mapsto \widehat{\mathrm{F}}_0(\mathrm{x}),$$

where $\mathrm{U} \subseteq \mathbb{R}$ *is open. For* $(\mathrm{t}_0, \mathrm{y}_0) \in \mathbb{T} \times \mathrm{U}$*, let* $\xi, \eta \colon \mathbb{T}' \to \mathrm{U}$ *be of class* C^1 *and satisfy*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathrm{F}}_0(\xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathrm{y}_0$$

and

$$\dot{\eta}'(\mathrm{t}) < \widehat{\mathrm{F}}_0(\boldsymbol{\eta}(\mathrm{t})), \quad \eta(\mathrm{t}_0) = \mathrm{y}_0.$$

Then $\eta(\mathrm{t}) < \xi(\mathrm{t})$ *for* $\mathrm{t} > \mathrm{t}_0$.

Proof We have

$$\dot{\eta}(t_0) < \widehat{F}(y_0) = \dot{\xi}(t_0).$$

Therefore, by continuity of the derivatives, there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$\dot{\eta}(t) < \dot{\xi}(t), \qquad t \in [t_0, t_0 + \epsilon].$$

Therefore, for $t \in (t_0, t_0 + \epsilon]$,

$$\eta(t) = \int_{t_0}^{t} \dot{\eta}(\tau)\,\mathrm{d}\tau < \int_{t_0}^{t} \dot{\xi}(\tau)\,\mathrm{d}\tau = \xi(t).$$

Now suppose that it does not hold that $\eta(t) < \xi(t)$ for all $t \geq t_0$. Then let

$$T = \inf\{t \geq t_0 \mid \eta(t) \geq \xi(t)\} > t_0 + \epsilon.$$

By continuity, $\eta(T) = \xi(T)$. Thus

$$\begin{aligned} \eta'(T) &= \underbrace{\eta'(T) - \widehat{F}(\eta(T))}_{<0} + \widehat{F}(T, \eta(T)) \\ &< \underbrace{\xi'(t) - \widehat{F}(\xi(T))}_{=0} + \widehat{F}(\xi(T)) = \xi'(T). \end{aligned}$$

On the other hand, for $h \in \mathbb{R}_{>0}$ (sufficiently small for the expression to be defined) we have

$$\frac{\eta(T) - \eta(T-h)}{h} > \frac{\xi(T) - \xi(T-h)}{h},$$

and taking the limit as $h \to 0$ gives $\eta'(T) \geq \xi'(T)$, contradicting our computation just proceeding. ▼

By the lemma,

$$V(\xi(t)) \leq V(\xi(t_0))\mathrm{e}^{-C_1C_2(t-t_0)}$$

for $t \geq t_0$. Now, again using the definition of C_1 and α,

$$\begin{aligned}\|\xi(t) - x_0\| &\leq \left(\frac{V(\xi(t))}{C_1}\right)^{1/\alpha} \leq \left(\frac{V(\xi(t_0))\mathrm{e}^{-C_1C_2(t-t_0)}}{C_1}\right)^{1/\alpha} \\ &\leq \frac{\|x - x_0\|}{C_1^{2\alpha}}\mathrm{e}^{-C_1C_2(t-t_0)/\alpha}\end{aligned}$$

for all $t \geq t_0$. Recalling that the preceding estimates are valid for any $(t_0, x) \in \mathbb{T} \times \mathsf{B}(\delta, x_0)$, we conclude exponential stability of x_0. ■

4.3.5 The Second Method for time-varying linear equations

The next two sections will be concerned with Lyapunov's Second Method for systems of linear homogeneous ordinary differential equations. In this section we treat the time-varying case, and in the next we treat the time-invariant case. While it is relatively easy to prove the theorems in this case using the general, not for linear equations, results of Sections 4.3.3 and 4.3.4, we instead give self-contained proofs that illustrate the special character of stability for linear differential equations that we studied in Section 4.2.

In the study of Lyapunov's Second Method for linear equations, one works with Lyapunov functions that are especially adapted to the linear structure of the equations, namely the quadratic functions of Sections 4.3.1.4 and 4.3.1.5. In working with such functions, the derivatives along solutions, called the "Lie derivative" in Definitions 4.3.20 and 4.3.25, take a particular form that leads to the following definition and associated following result.

4.3.31 Definition (Lyapunov pair for time-varying linear ordinary differential equations) Let F be a system of linear homogeneous ordinary differential equations in an n-dimensional $\mathbb{R}$-vector space V and with right-hand side

$$\begin{aligned}\widehat{F}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V} \\ (t, x) &\mapsto A(t)(x)\end{aligned}$$

for $A\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$. Suppose that V has an inner product $\langle \cdot, \cdot \rangle$. A ***Lyapunov pair*** for F is a pair (P, Q) where

(i) $P, Q\colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ are such that P is of class C^1, Q is continuous, and $P(t)$ and $Q(t)$ are symmetric, and

(ii) $\dot{P}(t) + P(t) \circ A(t) + A^T(t) \circ P(t) = -Q(t)$ for all $t \in \mathbb{T}$. •

Note that, with the notion of a Lyapunov pair, one can think of (1) P as being given, and part (ii) of the definition prescribing Q or (2) Q as being given, in which

case part (ii) prescribing a linear differential equation for P. Both ways of thinking about this will be useful.

At first encounter, such a definition seems to come from nowhere. However, the motivation for it is straightforward, as the following lemma shows, and its proof makes clear.

4.3.32 Lemma (Derivative of quadratic function along solutions of a linear ordinary differential equation) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *and with right-hand side*

$$\begin{aligned}\widehat{\mathrm{F}}\colon \mathbb{T}\times \mathsf{V} &\to \mathsf{V}\\ (\mathrm{t},\mathrm{x}) &\mapsto \mathrm{A}(\mathrm{t})(\mathrm{x})\end{aligned}$$

for $\mathrm{A}\colon \mathbb{T}\to \mathrm{L}(\mathsf{V};\mathsf{V})$. *Suppose that* V *has an inner product* $\langle\cdot,\cdot\rangle$. *Let* $\mathrm{P}\colon \mathbb{T}\to \mathrm{L}(\mathsf{V};\mathsf{V})$ *be of class* C^1 *and such that* P(t) *is symmetric for every* $\mathrm{t}\in\mathbb{T}$ *and let* $\mathrm{f_P}$ *be the corresponding time-varying quadratic function as in Definition 4.3.15. Then, for any solution* $\xi\colon \mathbb{T}\to \mathsf{V}$ *for* F*, we have*

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{f_P}(\mathrm{t},\xi(\mathrm{t})) = -\mathrm{f_Q}(\mathrm{t},\xi(\mathrm{t})),$$

where (P, Q) *is a Lyapunov pair for* F.

Proof We shall represent solutions using the state transition map as in Section 3.2.2.2. Thus, if $(t_0,x)\in\mathbb{T}\times\mathsf{V}$, the solution to the initial value problem

$$\dot\xi(t) = A(t)(\xi(t)),\quad \xi(t_0)=x,$$

is $\xi(t)=\Phi_A(t,t_0)(x)$. Now we directly compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f_P(t,\Phi_A(t,t_0)(x)) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &= \langle \dot P(t)\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle + \langle P(t)\circ\tfrac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &\quad + \langle P(t)\circ\Phi_A(t,x_0)(x),\tfrac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t,t_0)(x)\rangle\\ &= \langle \dot P(t)\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle + \langle P(t)\circ A(t)\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &\quad + \langle P(t)\circ\Phi_A(t,x_0)(x)),A(t)\circ\Phi_A(t,t_0)(x)\rangle\\ &= -\langle Q(t)\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle,\end{aligned}$$

where (P,Q) is a Lyapunov pair, i.e.,

$$Q(t) = -\dot P(t) - P(t)\circ A(t) - A^T(t)\circ P(t),\qquad t\in\mathbb{T}.$$ ■

The lemma allows us to provide the following connection to the Lie derivative characterisations of Lemmata 4.3.21 and 4.3.26.

4.3.33 Corollary (Lie derivative of quadratic function along a linear ordinary differential equation) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *and with right-hand side*

$$\begin{aligned}\widehat{F}\colon \mathbb{T} \times V &\to V\\ (t, x) &\mapsto A(t)(x)\end{aligned}$$

for $A\colon \mathbb{T} \to L(V; V)$*. Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$*. Let* $P\colon \mathbb{T} \to L(V; V)$ *be of class* C^1 *and such that* $P(t)$ *is symmetric for every* $t \in \mathbb{T}$ *and let* f_P *be the corresponding time-varying quadratic function as in Definition 4.3.15. Then,*

$$\mathscr{L}_F f_P(t, x) = -f_Q(t, x), \qquad (t, x) \in \mathbb{T} \times V.$$

Proof From the proof of the preceding lemma we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(t, \Phi_A(t, t_0)(x)) = -f_Q(t, \Phi_A(t, t_0)(x)).$$

Evaluating at $t = t_0$ gives the result. ■

We can now state and prove our main result concerning Lyapunov's Second Method for time-varying linear ordinary differential equations.

4.3.34 Theorem (Lyapunov's Second Method for linear time-varying ordinary differential equations) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *and with right-hand side*

$$\begin{aligned}\widehat{F}\colon \mathbb{T} \times V &\to V\\ (t, x) &\mapsto A(t)(x)\end{aligned}$$

for $A\colon \mathbb{T} \to L(V; V)$*. Suppose that* A *is continuous and that* $\sup \mathbb{T} = \infty$*. Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$*. Then the following statements hold.*

(i) The equation F *is stable if there exists* $P, Q\colon \mathbb{T} \to L(V; V)$ *with the following properties:*

  (a) P *is of class* C^1 *and* Q *is continuous;*

  (b) $P(t)$ *and* $Q(t)$ *are symmetric for every* $t \in \mathbb{T}$*;*

  (c) (P, Q) *is a Lyapunov pair for* F*;*

  (d) P *is positive-definite;*

  (e) Q *is positive-semidefinite.*

(ii) The equation F *is uniformly stable if there exists* $P, Q\colon \mathbb{T} \to L(V; V)$ *with the following properties:*

  (a) P *is of class* C^1 *and* Q *is continuous;*

  (b) $P(t)$ *and* $Q(t)$ *are symmetric for every* $t \in \mathbb{T}$*;*

(c) (P, Q) *is a Lyapunov pair for* F*;*
(d) P *is positive-definite;*
(e) P *is decrescent;*
(f) Q *is positive-semidefinite.*

(iii) *The equation* F *is asymptotically stable if there exists* $\mathrm{P}, \mathrm{Q} \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *with the following properties:*
(a) P *is of class* C^1 *and* Q *is continuous;*
(b) $\mathrm{P}(\mathrm{t})$ *and* $\mathrm{Q}(\mathrm{t})$ *are symmetric for every* $\mathrm{t} \in \mathbb{T}$*;*
(c) (P, Q) *is a Lyapunov pair for* F*;*
(d) P *is positive-definite;*
(e) Q *is positive-definite.*

(iv) *The equation* F *is uniformly asymptotically stable if there exists* $\mathrm{P}, \mathrm{Q} \colon \mathbb{T} \to \mathrm{L}(\mathsf{V}; \mathsf{V})$ *with the following properties:*
(a) P *is of class* C^1 *and* Q *is continuous;*
(b) $\mathrm{P}(\mathrm{t})$ *and* $\mathrm{Q}(\mathrm{t})$ *are symmetric for every* $\mathrm{t} \in \mathbb{T}$*;*
(c) (P, Q) *is a Lyapunov pair for* F*;*
(d) P *is positive-definite;*
(e) P *is decrescent;*
(f) Q *is positive-definite.*

We shall give two proofs of Theorem 4.3.34, one assuming Theorem 4.3.22 and the other an independent proof. The independent proof is interesting in and of itself because it makes use of methods particular to linear equations.

Proof of Theorem 4.3.34, assuming Theorem 4.3.22 If we collect together the conclusions of Lemma 4.3.17 and Corollary 4.3.33, we see that the hypotheses of parts (i)–(iv) of Theorem 4.3.34 imply those of the corresponding parts of Theorem 4.3.22, and thus the conclusions also correspond. ■

Independent proof of Theorem 4.3.34 (i) Let $t_0 \in \mathbb{T}$. Since P is positive-definite, by definition and by Lemma 4.3.14, there exists $C_1 \in \mathbb{R}_{>0}$ such that

$$C_1 \|x\|^2 \leq f_P(t, x), \qquad t \in \mathbb{T},\ x \in \mathsf{V}.$$

By Lemma 4.3.14, there exists $C_2 \in \mathbb{R}_{>0}$ such that

$$f_P(t_0, x) \leq C_1 \|x\|^2, \qquad x \in \mathsf{V}.$$

Since Q is positive-semidefinite, by Lemma 4.3.32 we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(t, \Phi_A(t, t_0)(x)) \leq 0$$

for all $x \in \mathsf{V}$ and $t \geq t_0$. Therefore, we have

$$C_1\|\Phi_A(t,t_0)(x)\|^2 \leq f_P(t,\Phi_A(t,t_0)(x)) \leq f_P(t_0,x) \leq C_2\|x\|^2$$

for every $x \in \mathsf{V}$ and $t \geq t_0$. Thus

$$\|\Phi_A(t,t_0)(x)\| \leq \sqrt{C_2/C_1}\|x\|,$$

which gives stability.

(ii) Here, since P is positive-definite and decrescent, by definition and by Lemma 4.3.14, we have $C_1, C_2 \in \mathbb{R}_{>0}$ such that

$$C_1\|x\|^2 \leq f_P(t,x) \leq C_2\|x\|^2, \qquad t \in \mathbb{T}.$$

As in the proof of part (i),

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(t,\Phi_A(t,t_0)(x)) \leq 0$$

for all $(t_0,x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. Therefore,

$$C_1\|\Phi_A(t,t_0)(x)\|^2 \leq f_P(t,\Phi_A(t,t_0)(x)) \leq f_P(t_0,x) \leq C_2\|x\|^2$$

for all $(t_0,x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. Thus,

$$\|\Phi_A(t,t_0)(x)\| \leq \sqrt{C_2/C_1}\|x\|$$

for every $(t_0,x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. This gives uniform stability, as desired.

(iii) Let $t_0 \in \mathbb{T}$. Here we have stability from part (i). From that part of the proof we also have $C_1, C_2 \in \mathbb{R}_{>0}$ (with C_2 possibly depending on t_0) such that

$$C_1\|\Phi_A(t,t_0)(x)\|^2 \leq f_P(t,\Phi_A(t,t_0)(x)) \leq f_P(t_0,x) \leq C_2\|x\|^2$$

for every $x \in \mathsf{V}$ and $t \geq t_0$. Since Q is positive-definite, by definition and by Lemma 4.3.14, there exists $C_3 \in \mathbb{R}_{>0}$ such that

$$C_3\|x\|^2 \leq f_Q(t,x), \qquad (t,x) \in \mathbb{T} \times \mathsf{V}.$$

Thus, by Lemma 4.3.32, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(t,\Phi_A(t,t_0)(x)) = -f_Q(t,\Phi_A(t,t_0)) \leq -C_3\|\Phi_A(t,t_0)(x)\|^2.$$

for all $x \in \mathsf{V}$ and $t \geq t_0$. Therefore, there exists $\gamma \in \mathbb{R}_{\geq 0}$ such that

$$\lim_{t\to\infty} f_P(t,\Phi_A(t,t_0)(x)) = \gamma.$$

We claim that $\gamma = 0$. Suppose otherwise, and that $\gamma \in \mathbb{R}_{>0}$. We then have

$$\begin{aligned} f_P(t, \Phi_A(t, t_0)(x)) &= f_P(t_0, x) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} f_P(\tau, \Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\ &= f_P(t_0, x) - \int_{t_0}^{t} f_Q(\tau, \Phi_A(\tau, t_0)) \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - C_3 \int_{t_0}^{t} \|\Phi_A(\tau, t_0)(x)\|^2 \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - \frac{C_3}{C_1} \int_{t_0}^{t} f_P(\tau, \Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - \frac{C_3}{C_1} \gamma (t - t_0). \end{aligned}$$

This implies that $\lim_{t\to\infty} f_P(t, \Phi_A(t, t_0)(x)) = -\infty$. This contradiction leads us to conclude that $\gamma = 0$. Finally, we then have

$$\lim_{t\to\infty} \|\Phi_A(t, t_0)(x)\|^2 \leq \lim_{t\to\infty} C_1^{-1} f_P(t, \Phi_A(t, t_0)(x)) = 0,$$

which gives asymptotic stability.

(iv) Here we have uniform stability from part (i). From that part of the proof we also have $C_1, C_2 \in \mathbb{R}_{>0}$ such that

$$C_1 \|\Phi_A(t, t_0)(x)\|^2 \leq f_P(t, \Phi_A(t, t_0)(x)) \leq f_P(t_0, x) \leq C_2 \|x\|^2$$

for every $(t_0, x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. Since Q is positive-definite by definition and by Lemma 4.3.14, there exists $C_3 \in \mathbb{R}_{>0}$ such that

$$C_3 \|x\|^2 \leq f_Q(t, x), \qquad (t, x) \in \mathbb{T} \times \mathsf{V}.$$

Thus, by Lemma 4.3.32, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(t, \Phi_A(t, t_0)(x)) = -f_Q(t, \Phi_A(t, t_0)) \leq -C_3 \|\Phi_A(t, t_0)(x)\|^2.$$

for all $(t_0, x) \in \mathbb{T} \times \mathsf{V}$ and $t \geq t_0$. Therefore, there exists $\gamma \in \mathbb{R}_{\geq 0}$ such that

$$\lim_{t\to\infty} f_P(t, \Phi_A(t, t_0)(x)) = \gamma.$$

We claim that $\gamma = 0$. Suppose otherwise, and that $\gamma \in \mathbb{R}_{>0}$. We then have

$$\begin{aligned} f_P(t, \Phi_A(t, t_0)(x)) &= f_P(t_0, x) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} f_P(\tau, \Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\ &= f_P(t_0, x) - \int_{t_0}^{t} f_Q(\tau, \Phi_A(\tau, t_0)) \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - C_3 \int_{t_0}^{t} \|\Phi_A(\tau, t_0)(x)\|^2 \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - \frac{C_3}{C_1} \int_{t_0}^{t} f_P(\tau, \Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\ &\leq f_P(t_0, x) - \frac{C_3}{C_1} \gamma (t - t_0). \end{aligned}$$

This implies that $\lim_{t \to \infty} f_P(t, \Phi_A(t, t_0)(x)) = -\infty$. This contradiction leads us to conclude that $\gamma = 0$. Finally, we then have

$$\lim_{t \to \infty} \|\Phi_A(t, t_0)(x)\|^2 \leq \lim_{t \to \infty} C_1^{-1} f_P(t, \Phi_A(t, t_0)(x)) = 0,$$

which gives uniform asymptotic stability, since C_1, C_2, and C_3 are independent of t_0. ■

4.3.35 Remark (Automatic implications of Theorem 4.3.34) We recall from Theorem 4.2.3 that the conclusions of stability, uniform stability, asymptotic stability, and uniform asymptotic stability are actually of the global sort given in Definition 4.2.1. Moreover, from Proposition 4.2.6 we see that uniform stability and stability are equivalent for linear homogeneous equations with constant coefficients, and similarly that uniform asymptotic stability and asymptotic stability are equivalent. •

4.3.6 The Second Method for linear equations with constant coefficients

The final setting in which we consider conditions for stability using Lyapunov's Second Method is that for linear homogeneous ordinary differential equations with constant coefficients.

As in the time-varying setting of the preceding section, in this section we work with Lyapunov functions that are especially adapted to the linear structure of the equations, namely the quadratic functions of Sections 4.3.1.4. In working with such functions, the derivatives along solutions, called the "Lie derivative" in Definitions 4.3.20 and 4.3.25, take a particular form that leads to the following definition and associated following result. What we have, of course, is a specialisation Definition 4.3.31.

4.3.36 Definition (Lyapunov pair for linear ordinary differential equations with constant coefficients) Let F be a system of linear homogeneous ordinary differential equations in an n-dimensional $\mathbb{R}$-vector space V with constant coefficients and with right-hand side

$$\begin{aligned}\widehat{F}\colon \mathbb{T}\times\mathsf{V} &\to \mathsf{V}\\ (t,x) &\mapsto A(x)\end{aligned}$$

for $A \in \mathrm{L}(\mathsf{V};\mathsf{V})$. Suppose that V has an inner product $\langle\cdot,\cdot\rangle$. A ***Lyapunov pair*** for F is a pair (P,Q) where

(i) $P, Q \in \mathrm{L}(\mathsf{V};\mathsf{V})$ are symmetric, and

(ii) $P\circ A + A^T\circ P = -Q$. •

As in the time-varying case, one can think of (1) P as being given, and part (ii) of the definition prescribing Q or (2) Q as being given, and (ii) of the definition giving a linear algebraic equation for P. Both ways of thinking about this will be useful.

Let us indicate the significance of the notion of a Lyapunov pair in this context.

4.3.37 Lemma (Derivative of quadratic function along solutions of a linear ordinary differential equation with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned}\widehat{\mathrm{F}}\colon \mathbb{T}\times\mathsf{V} &\to \mathsf{V}\\ (\mathrm{t},\mathrm{x}) &\mapsto \mathrm{A}(\mathrm{x})\end{aligned}$$

for $\mathrm{A}\in\mathrm{L}(\mathsf{V};\mathsf{V})$. *Suppose that* V *has an inner product* $\langle\cdot,\cdot\rangle$. *Let* $\mathrm{P}\in\mathrm{L}(\mathsf{V};\mathsf{V})$ *be symmetric and let* $\mathrm{f_P}$ *be the corresponding quadratic function as in Definition 4.3.11. Then, for any solution* $\xi\colon\mathbb{T}\to\mathsf{V}$ *for* F*, we have*

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{f_P}(\xi(\mathrm{t})) = -\mathrm{f_Q}(\xi(\mathrm{t})),$$

where (P,Q) *is a Lyapunov pair for* F.

Proof We shall represent solutions using the state transition map as in Section 3.2.2.2. Thus, if $(t_0,x)\in\mathbb{T}\times\mathsf{V}$, the solution to the initial value problem

$$\dot\xi(t) = A(\xi(t)),\quad \xi(t_0)=x,$$

is $\xi(t)=\Phi_A(t,t_0)(x)$. Now we directly compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f_P(\Phi_A(t,t_0)(x)) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle P\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &= \langle P\circ\tfrac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &\quad+\langle P\circ\Phi_A(t,x_0)(x),\tfrac{\mathrm{d}}{\mathrm{d}t}\Phi_A(t,t_0)(x)\rangle\\ &= \langle P\circ A\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle\\ &\quad+\langle P\circ\Phi_A(t,x_0)(x)),A\circ\Phi_A(t,t_0)(x)\rangle\\ &= -\langle Q\circ\Phi_A(t,t_0)(x),\Phi_A(t,t_0)(x)\rangle,\end{aligned}$$

where (P, Q) is a Lyapunov pair, i.e.,

$$Q = -P \circ A - A^T \circ P. \qquad \blacksquare$$

The lemma allows us to provide the following connection to the Lie derivative characterisation Lemma 4.3.26.

4.3.38 Corollary (Lie derivative of quadratic function along a linear ordinary differential equation with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times V &\to V \\ (t, x) &\mapsto A(x) \end{aligned}$$

for $A \in L(V; V)$*. Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$*. Let* $P \in L(V; V)$ *be symmetric and let* f_P *be the corresponding quadratic function as in Definition 4.3.11. Then,*

$$\mathscr{L}_F f_P(x) = -f_Q(x), \qquad x \in V.$$

Proof From the proof of the preceding lemma we have

$$\frac{d}{dt} f_P(\Phi_A(t, t_0)(x)) = -f_Q(\Phi_A(t, t_0)(x)).$$

Evaluating at $t = t_0$ gives the result. $\blacksquare$

We may now state our first result.

4.3.39 Theorem (Lyapunov's Second Method for linear ordinary differential equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times V &\to V \\ (t, x) &\mapsto A(x) \end{aligned}$$

for $A \in L(V; V)$*. Suppose that* $\sup \mathbb{T} = \infty$*. Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$*. Then the following statements hold.*

(i) The equation F *is stable if there exists* $P, Q \in L(V; V)$ *with the following properties:*
  (a) P *and* Q *are symmetric;*
  (b) (P, Q) *is a Lyapunov pair for* F*;*
  (c) P *is positive-definite;*
  (d) Q *is positive-semidefinite.*

(ii) The equation F *is asymptotically stable if there exists* $P, Q \in L(V; V)$ *with the following properties:*
  (a) P *and* Q *are symmetric;*

(b) (P, Q) *is a Lyapunov pair for* F*;*

(c) P *is positive-definite;*

(d) Q *is positive-definite.*

We shall give two proofs of Theorem 4.3.39, one assuming Theorem 4.3.27 and the other an independent proof. The independent proof is interesting in and of itself because it makes use of methods particular to linear equations.

Proof of Theorem 4.3.39, assuming Theorem 4.3.27 If we collect together the conclusions of Lemma 4.3.13 and Corollary 4.3.38, we see that the hypotheses of parts (i) and (ii) of Theorem 4.3.39 imply those of the corresponding parts of Theorem 4.3.27, and thus the conclusions also correspond. ■

Independent proof of Theorem 4.3.39 (i) Let $t_0 \in \mathbb{T}$. Since P is positive-definite, by Lemma 4.3.14, there exists $C_1, C_2 \in \mathbb{R}_{>0}$ such that

$$C_1\|x\|^2 \leq f_P(x) \leq C_2\|x\|^2, \qquad x \in \mathsf{V}.$$

Since Q is positive-semidefinite, by Lemma 4.3.37 we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(\Phi_A(t, t_0)(x)) \leq 0$$

for all $x \in \mathsf{V}$ and $t \geq t_0$. Therefore, we have

$$C_1\|\Phi_A(t, t_0)(x)\|^2 \leq f_P(t, \Phi_A(t, t_0)(x)) \leq f_P(t_0, x) \leq C_2\|x\|^2$$

for every $x \in \mathsf{V}$ and $t \geq t_0$. Thus

$$\|\Phi_A(t, t_0)(x)\| \leq \sqrt{C_2/C_1}\|x\|,$$

which gives stability.

(ii) Let $t_0 \in \mathbb{T}$. Here we have stability from part (i). From that part of the proof we also have $C_1, C_2 \in \mathbb{R}_{>0}$ such that

$$C_1\|\Phi_A(t, t_0)(x)\|^2 \leq f_P(\Phi_A(t, t_0)(x)) \leq f_P(x) \leq C_2\|x\|^2$$

for every $x \in \mathsf{V}$ and $t \geq t_0$. Since Q is positive-definite, by Lemma 4.3.14, there exists $C_3 \in \mathbb{R}_{>0}$ such that

$$C_3\|x\|^2 \leq f_Q(x), \qquad x \in \mathsf{V}.$$

Thus, by Lemma 4.3.37, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f_P(\Phi_A(t, t_0)(x)) = -f_Q(\Phi_A(t, t_0)) \leq -C_3\|\Phi_A(t, t_0)(x)\|^2.$$

for all $x \in \mathsf{V}$ and $t \geq t_0$. Therefore, there exists $\gamma \in \mathbb{R}_{\geq 0}$ such that

$$\lim_{t \to \infty} f_P(\Phi_A(t, t_0)(x)) = \gamma.$$

We claim that $\gamma = 0$. Suppose otherwise, and that $\gamma \in \mathbb{R}_{>0}$. We then have

$$\begin{aligned}
f_P(\Phi_A(t, t_0)(x)) &= f_P(x) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} f_P(\Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\
&= f_P(x) - \int_{t_0}^{t} f_Q(\Phi_A(\tau, t_0)) \, \mathrm{d}\tau \\
&\leq f_P(x) - C_3 \int_{t_0}^{t} \|\Phi_A(\tau, t_0)(x)\|^2 \, \mathrm{d}\tau \\
&\leq f_P(x) - \frac{C_3}{C_1} \int_{t_0}^{t} f_P(\Phi_A(\tau, t_0)(x)) \, \mathrm{d}\tau \\
&\leq f_P(x) - \frac{C_3}{C_1} \gamma (t - t_0).
\end{aligned}$$

This implies that $\lim_{t \to \infty} f_P(\Phi_A(t, t_0)(x)) = -\infty$. This contradiction leads us to conclude that $\gamma = 0$. Finally, we then have

$$\lim_{t \to \infty} \|\Phi_A(t, t_0)(x)\|^2 \leq \lim_{t \to \infty} C_1^{-1} f_P(\Phi_A(t, t_0)(x)) = 0,$$

which gives asymptotic stability. ■

4.3.40 Remark (Automatic implications of Theorem 4.3.39) We recall from Theorem 4.2.3 that the conclusions of stability, uniform stability, asymptotic stability, and uniform asymptotic stability are actually of the global sort given in Definition 4.2.1. •

4.3.41 Example (Example 4.2.10 cont'd) We again look at the linear homogeneous ordinary differential equation $\boldsymbol{F}$ on $\mathsf{V} = \mathbb{R}^2$ defined by the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}.$$

The inner product we use is the standard one:

$$\langle (u_1, u_2), (v_1, v_2) \rangle_{\mathbb{R}^2} = u_1 v_1 + u_2 v_2.$$

In this case, the induced norm is the standard norm for $\mathbb{R}^2$. Note that, if $\boldsymbol{L} \in \mathrm{L}(\mathbb{R}^2; \mathbb{R}^2)$, then the transpose with respect to the standard inner product is just the usual matrix transpose.

For this example, there are various cases to consider, and we look at them separately in view of Theorem 4.3.39. In the following discussion, the reader should compare the conclusions with those of Example 4.2.10.

1. $a = 0$ and $b = 0$: In this case, we know the system is unstable. Thus we will certainly not be able to find a Lyapunov pair $(\boldsymbol{P}, \boldsymbol{Q})$ for $\boldsymbol{F}$ with $\boldsymbol{P}$ positive-definite and $\boldsymbol{Q}$ positive-semidefinite. Note, however, that without knowing more, just the lack of existence of such a $(\boldsymbol{P}, \boldsymbol{Q})$ does not allow us to conclude anything about stability in this case. We shall have more to say about this case in Example 4.3.53–1.
2. $a = 0$ and $b > 0$: The matrices
$$\boldsymbol{P} = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
have the property that $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. Since $\boldsymbol{P}$ is positive-definite and $\boldsymbol{Q}$ is positive-semidefinite, stability follows from part (i) of Theorem 4.3.39. Note that asymptotic stability cannot be concluded from this $\boldsymbol{P}$ and $\boldsymbol{Q}$ using part (ii) (and indeed asymptotic stability does not hold in this case).
3. $a = 0$ and $b < 0$: We shall consider this case in Example 4.3.53–2, where we will be able to use Lyapunov methods to conclude instability.
4. $a > 0$ and $b = 0$: Here we take
$$\boldsymbol{P} = \begin{bmatrix} a^2 & a \\ a & 2 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix}$$
and verify that $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. The eigenvalues of $\boldsymbol{P}$ are $\{\frac{1}{2}(a^2 + 2 \pm \sqrt{a^4 + 4})\}$. One may verify that $a^2 + 2 > \sqrt{a^4 + 4}$, thus $\boldsymbol{P}$ is positive-definite. Since $\boldsymbol{Q}$ is positive-semidefinite, we conclude stability of $\boldsymbol{F}$ from part (i) of Theorem 4.3.39. However, we cannot conclude asymptotic stability from part (ii); indeed, asymptotic stability does not hold.
5. $a > 0$ and $b > 0$: Here we take
$$\boldsymbol{P} = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix}$$
having the property that $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. Since $\boldsymbol{P}$ is positive-definite and $\boldsymbol{Q}$ is positive-semidefinite, from part (i) of Theorem 4.3.39 we can conclude stability for $\boldsymbol{F}$. However, we cannot conclude asymptotic stability using part (ii). However, we *do* have asymptotic stability in this case. We can rectify this in one of two ways.
   (a) By choosing a different $\boldsymbol{P}$ and $\boldsymbol{Q}$ with both positive-semidefinite, we can ensure asymptotic stability by part (ii) of Theorem 4.3.39. Theorem 4.3.52 guarantees that this is possible.
   (b) By resorting to an invariance principle, we can rescue things for this particular $\boldsymbol{P}$ and $\boldsymbol{Q}$. This is explained in Theorem 4.3.49, and in Example 4.3.50 for this example particularly.

6. $a > 0$ and $b < 0$: We shall consider this case in Example 4.3.53–3, where we will be able to use Lyapunov methods to conclude instability.
7. $a < 0$ and $b = 0$: This case is much like case 1 in that the system is unstable; thus we cannot find a Lyapunov pair $(\boldsymbol{P}, \boldsymbol{Q})$ for $\boldsymbol{F}$ with $\boldsymbol{P}$ positive-definite and $\boldsymbol{Q}$ positive-semidefinite. In Example 4.3.53–4 we shall have more to say about this case.
8. $a < 0$ and $b > 0$: We shall consider this case in Example 4.3.53–5, where we will be able to use Lyapunov methods to conclude instability.
9. $a < 0$ and $b < 0$: We shall consider this case in Example 4.3.53–6, where we will be able to use Lyapunov methods to conclude instability. •

The reader can see from this example that, even for a simple linear ordinary differential equation with constant coefficients, the sufficient conditions of Lyapunov's Second Method leave a great deal of room for improvement. In the subsequent sections we shall address this somewhat, although it is still the case that the method is one that is difficult to apply conclusively.

4.3.7 Invariance principles

We shall see in Section 4.3.9 that the sufficient conditions for asymptotic stability of some of the flavours of Lyapunov's Second Method are actually also necessary. However, in practice, one often produces a locally positive-definite function whose Lie derivative is merely negative-semidefinite, not negative-definite as one needs for asymptotic stability. In order to deal with this commonly encountered situation, we provide in this section a strategy that falls under a general umbrella of what are know of as "invariance principles." We prove two associated theorems, one for autonomous, not necessarily linear, ordinary differential equations and one for linear ordinary differential equations with constant coefficients.

4.3.7.1 Invariant sets and limit sets In order to prove our result for general autonomous ordinary differential equations, we need a collection of preliminary definitions and results.

4.3.42 Definition (Invariant set) Let $\boldsymbol{F}$ be an ordinary differential equation with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n.$$

A subset $A \subseteq U$ is:

(i) **F-*invariant*** if, for all $(t_0, \boldsymbol{x}) \in \mathbb{T} \times A$, the solution $\xi\colon \mathbb{T}' \to U$ to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}(t, \xi(t)), \quad \xi(t_0) = \boldsymbol{x},$$

satisfies $\xi(t) \in A$ for every $t \in \mathbb{T}'$;

(ii) ***positively* F-*invariant*** if, for all $(t_0, x) \in \mathbb{T} \times A$, the solution $\xi\colon \mathbb{T}' \to U$ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t, \xi(t)), \quad \xi(t_0) = x,$$

satisfies $\xi(t) \in A$ for every $t \geq t_0$. •

4.3.43 Definition (Positive limit set) Let F be an autonomous ordinary differential equation with right-hand side

$$\widehat{F}\colon \mathbb{T} \times U \to \mathbb{R}^n$$
$$(t, x) \mapsto \widehat{F}_0(x).$$

Suppose that $\sup \mathbb{T} = \infty$ and that $0 \in \mathbb{T}$. Let $x_0 \in U$ and let $\xi\colon \mathbb{T}' \to U$ be the solution to the initial value problem

$$\dot{\xi} = \widehat{F}_0(\xi(t)), \quad \xi(0) = x_0,$$

and suppose that $\sup \mathbb{T}' = \infty$.

(i) A point $x \in U$ is a ***positive limit point of* $\mathbf{x_0}$** if there exists a sequence $(t_j)_{j \in \mathbb{Z}_{>0}} \subseteq \mathbb{R}$ such that
   (a) $t_j < t_{j+1}$, $j \in \mathbb{Z}_{>0}$,
   (b) $\lim_{j\to\infty} t_j = \infty$, and
   (c) $\lim_{j\to\infty} \xi(t_j) = x$.

(ii) The ***positive limit set of* $\mathbf{x_0}$**, denoted by $\Omega(F, x_0)$, is the set of positive limit points of x_0. •

Positive limit sets have many interesting properties.

4.3.44 Lemma (Properties of the positive limit set) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$
$$(\mathrm{t}, \mathbf{x}) \mapsto \widehat{\mathbf{F}}_0(\mathbf{x}).$$

Let $\mathrm{A} \subseteq \mathrm{U}$ *be compact and positively* $\mathbf{F}$*-invariant. If* $\mathbf{x}_0 \in \mathrm{A}$*, then* $\Omega(\mathbf{F}, \mathbf{x}_0)$ *is a nonempty, compact, and positively* $\mathbf{F}$*-invariant subset of* A*. Furthermore, if* $\xi\colon \mathbb{T}' \to \mathrm{U}$ *is the solution to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}_0(\xi(\mathrm{t})), \quad \xi(0) = \mathbf{x}_0,$$

then

$$\lim_{\mathrm{t}\to\infty} \mathrm{d}_{\Omega(\mathbf{F},\mathbf{x}_0)}(\xi(\mathrm{t})) = 0.$$

Proof Let $(t_j)_{j\in\mathbb{Z}_{>0}} \subseteq \mathbb{R}_{>0}$ satisfy $t_j < t_{j+1}$, $j \in \mathbb{Z}_{>0}$, and $\lim_{j\to\infty} t_j = \infty$. The sequence $(\xi(t_j))_{j\in\mathbb{Z}_{>0}} \subseteq A$ has a convergence subsequence by the Bolzano–Weierstrass Theorem.***missing stuff*** By definition, the limit $\boldsymbol{x}$ will be in $\Omega(\boldsymbol{F}, \boldsymbol{x}_0)$. Since A is closed and positively-invariant, $\boldsymbol{x} \in A$. Thus $\Omega(\boldsymbol{F}, \boldsymbol{x}_0)$ is a nonempty subset of A.

If $\boldsymbol{x} \in A \setminus \Omega(\boldsymbol{F}, \boldsymbol{x}_0)$, then there exists $\epsilon \in \mathbb{R}_{>0}$ and $T \in \mathbb{R}_{>0}$ such that $\mathsf{B}(\epsilon, \boldsymbol{x}) \cap \{\xi(t) \mid t \geq T\} = \emptyset$. Therefore, $A \setminus \Omega(\boldsymbol{F}, \boldsymbol{x}_0)$ is open, and thus $\Omega(\boldsymbol{F}, \boldsymbol{x}_0)$ is closed and so compact since A is compact.***missing stuff***

Let $\boldsymbol{x} \in \Omega(\boldsymbol{F}, \boldsymbol{x}_0)$ and let $t \in \mathbb{R}_{\geq 0}$. There then exists a sequence $(t_j)_{j\in\mathbb{Z}_{>0}}$ such that

$$\lim_{j\to\infty} \xi(t_j) = \boldsymbol{x}.$$

Let η_j, $j \in \mathbb{Z}_{>0}$, be the solution to the initial value problem

$$\dot{\eta}_j(t) = \widehat{\boldsymbol{F}}_0(\eta_j(t)), \quad \eta(0) = \xi(t_j).$$

Then

$$\lim_{j\to\infty} \xi(t + t_j) = \lim_{j\to\infty} \eta_j(t) = \xi(t),$$

by continuity of solutions with respect to initial conditions. This shows that $\xi(t) \in \Omega(\boldsymbol{F}, \boldsymbol{x}_0)$, and so that $\Omega(\boldsymbol{F}, x)$ is positively X-invariant.

Lastly, suppose that there exists $\epsilon \in \mathbb{R}_{>0}$ and a sequence $(t_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ such that

1. $t_j < t_{j+1}$, $j \in \mathbb{Z}_{>0}$,
2. $\lim_{j\to\infty} t_j = \infty$, and
3. $\mathrm{d}_{\Omega(\boldsymbol{F},\boldsymbol{x}_0)}(\xi(t_j)) \geq \epsilon$, $j \in \mathbb{Z}_{>0}$.

By the Bolzano–Weierstrass Theorem, since A is compact there exists a convergent subsequence $(t_{j_k})_{k\in\mathbb{Z}_{>0}}$ such that $(\xi(t_{j_k}))_{k\in\mathbb{Z}_{>0}}$ converges to, say $\boldsymbol{x} \in A$. Note that $\boldsymbol{x} \in \Omega(\boldsymbol{F}, \boldsymbol{x}_0)$. However, we also have $\mathrm{d}_{\Omega(\boldsymbol{F},\boldsymbol{x}_0)}(\boldsymbol{x}) \geq \epsilon$. This contradiction means that we must have

$$\lim_{t\to\infty} \mathrm{d}_{\Omega(\boldsymbol{F},\boldsymbol{x}_0)}(\xi(t)) = 0,$$

as claimed. ■

4.3.7.2 Invariance principle for autonomous equations

We are now ready to present the LaSalle Invariance Principle on the asymptotic behavior of the integral curves of vector fields.

4.3.45 Theorem (LaSalle Invariance Principle) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$
$$(\mathrm{t}, \mathbf{x}) \mapsto \widehat{\mathbf{F}}_0(\mathbf{x}).$$

Suppose that $\sup \mathbb{T} = \infty$ *and that* $0 \in \mathbb{T}$. *Let* $\mathrm{A} \subseteq \mathrm{U}$ *be compact and positively* $\mathbf{F}$*-invariant. Let* $\mathrm{V}\colon \mathrm{U} \to \mathbb{R}$ *be continuously differentiable and satisfy* $\mathscr{L}_{\mathbf{F}_0}\mathrm{V}(\mathbf{x}) \leq 0$ *for all* $\mathbf{x} \in \mathrm{A}$, *and let* B *be the largest positively* $\mathbf{F}$*-invariant set contained in* $\{\mathbf{x} \in \mathrm{A} \mid \mathscr{L}_{\mathbf{F}_0}\mathrm{V}(\mathbf{x}) = 0\}$. *Then the following statements hold:*

(i) for every $\mathbf{x} \in \mathrm{A}$*, the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}_0(\xi(\mathrm{t})), \quad \xi(0) = \mathbf{x},$$

satisfies $\lim_{\mathrm{t}\to\infty} \mathrm{d}_{\mathrm{B}}(\xi(\mathrm{t})) = 0$*;*

(ii) if B *consists of a finite number of isolated points, then, for every* $\mathbf{x} \in \mathrm{A}$*, there exists* $\mathbf{y} \in \mathrm{B}$ *such that the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}_0(\xi(\mathrm{t})), \quad \xi(0) = \mathbf{x},$$

satisfies $\lim_{\mathrm{t}\to\infty} \xi(\mathrm{t}) = \mathbf{y}$.

Proof (i) The function $V|A$ is bounded from below, because it is continuous on the compact set A.***missing stuff*** For $\boldsymbol{x} \in A$, let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}_0(\xi(t)), \quad \xi(0) = \boldsymbol{x}.$$

The function $t \mapsto V \circ \xi(t)$ is nonincreasing and bounded from below. Therefore, $\lim_{t\to\infty} V \circ \xi(t)$ exists and is equal to, say, $\alpha \in \mathbb{R}$. Now, let $\boldsymbol{y} \in \Omega(\boldsymbol{F}, \boldsymbol{x})$ and let $(t_j)_{j\in\mathbb{Z}_{>0}}$ satisfy $\lim_{j\to\infty} \xi(t_k) = \boldsymbol{y}$. By continuity of V, $\alpha = \lim_{j\to\infty} V \circ \xi(t_k) = V(\boldsymbol{y})$. This proves that $V(\boldsymbol{y}) = \alpha$ for all $\boldsymbol{y} \in \Omega(\boldsymbol{F}, \boldsymbol{x})$. Because $\Omega(\boldsymbol{F}, \boldsymbol{x})$ is positively $\boldsymbol{F}$-invariant, if $\boldsymbol{y} \in \Omega(\boldsymbol{F}, \boldsymbol{x})$ and if $\boldsymbol{\eta}$ is the solution to the initial value problem

$$\dot{\boldsymbol{\eta}}(t) = \widehat{\boldsymbol{F}}_0(\boldsymbol{\eta}(t)), \quad \boldsymbol{\eta}(0) = \boldsymbol{y},$$

then $\boldsymbol{\eta}(t) \in \Omega(\boldsymbol{F}, \boldsymbol{x})$ for all $t \in \mathbb{R}_{>0}$. Therefore, $V \circ \boldsymbol{\eta}(t) = \alpha$ for all $t \in \mathbb{R}_{>0}$ and, therefore, by Lemma 4.3.26, $\mathscr{L}_{\boldsymbol{F}_0} V(y) = 0$. Now, because $\mathscr{L}_{\boldsymbol{F}_0} V(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} \in \Omega(\boldsymbol{F}, \boldsymbol{x})$, we know that

$$\Omega(\boldsymbol{F}, \boldsymbol{x}) \subseteq \{\boldsymbol{x} \in A \mid \mathscr{L}_{\boldsymbol{F}_0} V(x) = 0\}.$$

This implies that $\Omega(\boldsymbol{F}, \boldsymbol{x}) \subseteq B$, and this proves this part of the theorem.

(ii) Let $\boldsymbol{x} \in A$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}_0(\xi(t)), \quad \xi(0) = \boldsymbol{x}.$$

Since $B = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$ is comprised of isolated points, there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$\overline{\mathsf{B}}(2\epsilon, \boldsymbol{y}_{j_1}) \cap \overline{\mathsf{B}}(2\epsilon, \boldsymbol{y}_{j_2}) = \emptyset$$

for all $j_1, j_2 \in \{1, \ldots, k\}$. By assumption and by part (i), there exists $T \in \mathbb{R}_{>0}$ such that

$$\xi(t) \in \cup_{j=1}^{k} \mathsf{B}(\epsilon, \boldsymbol{y}_j), \qquad t \geq T.$$

Since ξ is continuous, $\xi([T, \infty))$ is connected by ***missing stuff***. This, however, implies that there must exist $\boldsymbol{y} \in B$ such that $\xi([T, \infty)) \subseteq \mathsf{B}(\epsilon, \boldsymbol{y})$, giving this part of the theorem. ■

The following more or less immediate corollary provides a common situation where the LaSalle Invariance Principle is used.

4.3.46 Corollary (Barbashin–Krasovskiĭ criterion) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathsf{U} &\to \mathbb{R}^n\\ (t,\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathbf{x}).\end{aligned}$$

Suppose that $\sup\mathbb{T}=\infty$ *and that* $0\in\mathbb{T}$. *Let* $\mathbf{x}_0\in\mathsf{U}$ *be an equilibrium point for* $\mathbf{F}$. *Assume tha there exists a function* $V\colon \mathsf{U}\to\mathbb{R}$ *with the following properties:*

(i) V *is of class* C^1;

(ii) $V\in \mathrm{LPD}(\mathbf{x}_0)$;

(iii) $-\mathscr{L}_{\mathbf{F}_0}V\in\mathrm{LPSD}(\mathbf{x}_0)$.

Let $C=\{\mathbf{x}\in\mathsf{U}\mid \mathscr{L}_{\mathbf{F}_0}V(x)=0\}$. *If there exists* $r\in\mathbb{R}_{>0}$ *such that the only positively* $\mathbf{F}$*-invariant subset of* $C\cap\mathsf{B}(r,\mathbf{x}_0)$ *is* $\{\mathbf{x}_0\}$, *then* $\mathbf{x}_0$ *is asymptotically stable.*

Proof As in the proof of Theorem 4.3.27(i), the fact that $V\in\mathrm{LPD}(\boldsymbol{x}_0)$ ensures that there is a closed subset of some ball about $\boldsymbol{x}_0$ that is $\boldsymbol{F}$-positively invariant. The corollary then follows from Theorem 4.3.45. ■

4.3.7.3 Invariance principle for linear equations with constant coefficients

Next we turn to an invariance principle specifically adapted to linear ordinary differential equations with constant coefficients. Unsurprisingly, the construction is linear algebraic in nature. The key to the construction is the following definition.

4.3.47 Definition (Observability operator, observable pair) Let V and W be finite-dimensional $\mathbb{R}$-vector spaces, and let $A\in\mathrm{L}(\mathsf{V};\mathsf{V})$ and $C\in\mathrm{L}(\mathsf{V};\mathsf{W})$.

(i) The ***observability operator*** for the pair (A,C) is the linear map

$$\begin{aligned}O(A,C)\colon \mathsf{V} &\to \mathsf{U}^{\dim_{\mathbb{R}}(\mathsf{V})}\\ v &\mapsto (C(v),C\circ A(v),\dots,C^{\dim_{\mathbb{R}}(\mathsf{V})-1}\circ A(v)).\end{aligned}$$

(ii) The pair (A,C) is ***observable*** if $\mathrm{rank}(O(A,C))=\dim_{\mathbb{R}}(\mathsf{V})$. •

This definition, while clear, does not capture the essence of the attribute of observability. The following result goes towards clarifying this.

4.3.48 Lemma (Characterisation of observability) *Let* V *and* W *be finite-dimensional* $\mathbb{R}$*-vector spaces, and let* $\mathrm{A}\in\mathrm{L}(\mathsf{V};\mathsf{V})$ *and* $\mathrm{C}\in\mathrm{L}(\mathsf{V};\mathsf{W})$. *Let* $\mathbb{T}\subseteq\mathbb{R}$ *be a time-domain for which* $0\in\mathbb{T}$ *and* $\mathrm{int}(\mathbb{T})\neq\emptyset$. *Then* (A,C) *is observable if and only if, given* $\mathrm{x}_1,\mathrm{x}_2\in\mathsf{V}$ *with* $\xi_1,\xi_2\colon\mathbb{T}\to\mathsf{V}$ *the solutions to the initial value problems*

$$\dot{\xi}_a(t)=\mathrm{A}(\xi_a(t)),\quad \xi_a(0)=\mathrm{x}_a,\qquad a\in\{1,2\},$$

we have $\mathrm{C}\circ\xi_1=\mathrm{C}\circ\xi_2$ *if and only if* $\mathrm{x}_1=\mathrm{x}_2$.

Moreover, $\ker(\mathrm{O}(\mathrm{A},\mathrm{C}))$ *is the largest* A*-invariant subspace contained in* $\ker(\mathrm{C})$.

Proof Let $n = \dim_{\mathbb{R}}(\mathsf{V})$.

First suppose that (A, C) is observable and that $C \circ \xi_1 = C \circ \xi_2$. Then, differentiating successively with respect to t,

$$\frac{\mathrm{d}^j(C \circ \xi_a)}{\mathrm{d}t^j}(0) = C \circ A^j(x_a), \qquad j \in \mathbb{Z}_{\geq 0},\ a \in \{1, 2\}.$$

Thus we have

$$C \circ A^j(x_1) = C \circ A^j(x_2), \qquad j \in \mathbb{Z}_{\geq 0}.$$

Thus $x_1 - x_2 \in \ker(O(A, C))$, and so $x_1 = x_2$ since $O(A, C)$ is observable.

Next suppose that (A, C) is not observable, and so $O(A, C)$ is not injective. Thus there exists a nonzero $x_0 \in \ker(O(A, C))$, meaning that $C \circ A^j(x_0) = 0$, $j \in \{0, 1, \ldots, n-1\}$. By the Cayley–Hamilton Theorem, this implies that $C \circ A^j(x_0) = 0$ for $j \in \mathbb{Z}_{\geq 0}$. Therefore, for any $t \geq 0$

$$\sum_{j=0}^{\infty} \frac{C \circ A^j}{j!}(x_0) = C \circ \mathrm{e}^{At}(x_0) = 0.$$

Therefore, taking $x_1 = x_0$ and $x_2 = 0$, $C \circ \xi_1 = C \circ \xi_2$ while $x_1 \neq x_2$.

Now we prove the final assertion of the lemma. First let us show that $\ker(O(A, C)) \subseteq \ker(C)$. If $x \in \ker(O(A, C))$, then $C \circ A^j(x) = 0$ for $j \in \{0, 1, \ldots, n-1\}$. This holds in particular for $j = 0$, giving the desired conclusion in this case.

Next we show that the kernel of $O(A, C)$ is A-invariant. Let $x \in \ker(O(A, C))$ and compute

$$O(A, C) \circ A(x) = (C \circ A(x), \ldots, C \circ A^n(x)).$$

Since $x \in \ker(O(A, C))$, we have

$$C \circ A(x) = 0, \ldots, C \circ A^{n-1}(x) = 0.$$

Also, by the Cayley–Hamilton Theorem, $C \circ A^n(x) = 0$. This shows that

$$O(A, c) \circ A(x) = 0,$$

or that $A(x) \in \ker(O(A, C))$.

Finally, we show that, if S is an A-invariant subspace contained in $\ker(C)$, then S is a subspace of $\ker(O(A, C))$. Given such an S and $x \in \mathsf{S}$, $C(x) = 0$. Since S is A-invariant, $A(x) \in V$, and since $\mathsf{S} \subseteq \ker(C)$, $C \circ A(x) = 0$. Proceeding in this way we see that

$$C \circ A^2(x) = \cdots = C \circ A^{n-1}(x) = 0.$$

But this means exactly that x is in $\ker(O(A, C))$. ■

The idea of observability is this. The linear map C we view as providing us with "measurements" in W of the states in V. The pair (A, C) is observable if we can deduce the state behaviour of the system merely by observing the measurements via C.

With this brief discussion of observability, we can now state a version of Theorem 4.3.45 adapted specially for linear differential equations.

4.3.49 Theorem (Invariance principle for linear ordinary differential equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned}\widehat{F}\colon \mathbb{T}\times V &\to V\\ (t,x) &\mapsto A(x)\end{aligned}$$

for $A \in L(V;V)$*. Suppose that* $\sup \mathbb{T} = \infty$*. Suppose that* V *has an inner product* $\langle\cdot,\cdot\rangle$*. Then* F *is asymptotically stable if there exists* $P, Q \in L(V;V)$ *with the following properties:*

(i) P *and* Q *are symmetric;*

(ii) (P, Q) *is a Lyapunov pair for* F*;*

(iii) P *is positive-definite;*

(iv) Q *is positive-semidefinite;*

(v) (A, Q) *is observable.*

We shall offer two proofs of the preceding theorem, one assuming the more general Theorem 4.3.45 and the other an independent proof.

Proof of Theorem 4.3.49, assuming Theorem 4.3.45 Under the hypotheses of Theorem 4.3.49, the function $V = f_P$ satisfies the hypotheses of Corollary 4.3.46. The subset C from the statement of Corollary 4.3.46 is then exactly the subspace $\ker(Q)$. Since (A, Q) is observable, by Lemma 4.3.48 $\{0\}$ is the largest A-invariant subspace of $\ker(Q)$. Since any invariant subset is contained in an invariant subspace—namely the subspace generated by the subset—it follows that the only F-invariant subset of C is $\{0\}$. Thus Theorem 4.3.49 follows from Theorem 4.3.45, specifically its Corollary 4.3.46. ■

Independent proof of Theorem 4.3.49 We suppose that P is positive-definite, Q is positive-semidefinite, (A, Q) is observable, and that F is not asymptotically stable. By Theorem 4.3.27(i) we know that F is stable, so it must be the case that A has at least one eigenvalue on the imaginary axis, and, therefore, a nontrivial periodic solution ξ. From our characterisation of the operator exponential in Procedures 3.2.45 and 3.2.48, we know that this periodic solution takes values in a two-dimensional subspace that we shall denote by L. What's more, every solution of F with initial condition in L is periodic and remains in L, i.e., L is F-invariant. Indeed, if $x \in L$, then

$$A(x) = \lim_{t\to 0} \frac{e^{At}(x) - x}{t} \in L$$

since $x, e^{At}(x) \in L$. We also claim that the subspace L is in $\ker(Q)$. To see this, suppose that the solutions in L have period T. We have, for any solution $\xi\colon [0,T] \to L$ for F, by Lemma 4.3.37,

$$0 = f_P \circ \xi(T) - f_P \circ \xi(0) = \int_0^T \frac{d f_P \circ \xi}{dt}(t)\, dt = -\int_0^T f_Q \circ \xi(t)\, dt.$$

Since Q is positive-semidefinite, this implies that $f_Q \circ \xi(t) = 0$ for $t \in [0, T]$. Thus $\mathsf{L} \subseteq \ker(Q)$, as claimed. Thus, with our initial assumptions, we have shown the existence of a nontrivial A-invariant subspace of $\ker(Q)$. This is a contradiction, however, since (A, Q) is observable. It follows, therefore, that F is asymptotically stable. ■

Let us resume our Example 4.3.41 to conclude asymptotic stability in the case where this is possible.

4.3.50 Example (Example 4.3.41 cont'd) We continue with the linear homogeneous ordinary differential equation $\boldsymbol{F}$ on $\mathsf{V} = \mathbb{R}^2$ defined by the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}.$$

Again, we use the standard inner product.

We consider the case where $a > 0$ and $b > 0$, since we know that $\boldsymbol{A}$ is Hurwitz in this case. We take

$$\boldsymbol{P} = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix},$$

noting that $\boldsymbol{P}$ is positive-definite, $\boldsymbol{Q}$ is positive-semidefinite, and $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. Using Theorem 4.3.39, we can only conclude stability, and not asymptotic stability. But we can compute

$$\boldsymbol{O}(\boldsymbol{A}, \boldsymbol{Q}) = \begin{bmatrix} 0 & 0 \\ 0 & 2a \\ 0 & 0 \\ -2ab & -2a^2 \end{bmatrix},$$

implying that $(\boldsymbol{A}, \boldsymbol{Q})$ is observable. We can thus conclude from Theorem 4.3.49 that $\boldsymbol{F}$ is asymptotically stable. •

4.3.8 Instability theorems

In this section we provide two so-called instability theorems. While our results above in this section give sufficient conditions for various flavours of stability, instability theorems give sufficient conditions for instability. The instability results we give fit under the umbrella of Lyapunov's Second method since the characterisations we give involve functions having certain properties. While our sufficient conditions for stability using Lyapunov's Second Method in Sections 4.3.3, 4.3.4, 4.3.5, and 4.3.6 are quite comprehensive, we shall back off from this level of exhaustiveness here, and only give two theorems, both for autonomous ordinary differential equations, one in the linear case and one in the not necessarily linear case.

4.3.8.1 Instability theorem for autonomous equations Let us state the more general result first. Some notation is useful. Let $U \subseteq \mathbb{R}^n$ be open, let $f\colon U \to \mathbb{R}$ be continuous, and let $x_0 \in U$. We suppose that $r \in \mathbb{R}_{>0}$ is such that $\overline{\mathsf{B}}(r, x_0) \subseteq U$ and define, for $a \in \mathbb{R}$,

$$f^{-1}(r, > a) = \{x \in \overline{\mathsf{B}}(r, x_0) \mid \ f(x) > a\}.$$

With this simple piece of notation, we then have the following result.

4.3.51 Theorem (An instability for autonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$
$$(\mathrm{t}, \mathbf{x}) \mapsto \widehat{\mathbf{F}_0}(\mathbf{x}).$$

Suppose that $\sup \mathbb{T} = \infty$. *Then an equilibrium state* $\mathbf{x}_0 \in \mathrm{U}$ *is unstable if there exists a function* $\mathrm{V}\colon \mathrm{U} \to \mathbb{R}$ *and* $\mathrm{r} \in \mathbb{R}_{>0}$ *with the following properties:*

(i) V *is of class* C^1*;*

(ii) $\mathrm{V}(\mathbf{x}_0) = 0$*;*

(iii) $\overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0) \subseteq \mathrm{U}$*;*

(iv) $\mathrm{V}^{-1}(\mathrm{s}, > 0) \neq \emptyset$ *for every* $\mathrm{s} \in (0, \mathrm{r})$*;*

(v) $\mathscr{L}_{\mathbf{F}_0}\mathrm{V}(\mathbf{x}) \in \mathbb{R}_{>0}$ *for* $\mathbf{x} \in \mathsf{B}(\mathrm{r}, \mathbf{x}_0)$.

Proof Let $\epsilon = \frac{r}{2}$ and let $\delta \in \mathbb{R}_{>0}$. We show that there exists $(t_0, x) \in \mathbb{T} \times \mathsf{B}(\delta, x_0)$ such that the solution ξ to the initial value problem

$$\dot{\xi}(t) = \widehat{F}_0(\xi(t)), \quad \xi(t_0) = x$$

satisfies $\xi(T) \notin \mathsf{B}(\epsilon, x_0)$ for some $T \geq t_0$. Indeed, let $\delta \in \mathbb{R}_{>0}$ and choose $x \in V^{-1}(s, > 0)$ for $s \leq \min\{\epsilon, \delta\}$. We claim that $\xi(T) \notin \mathsf{B}(\epsilon, x_0)$ for some $T \geq t_0$. Suppose otherwise and let

$$\beta = \inf\{\mathscr{L}_{F_0} V(x') \mid \ x' \in \overline{\mathsf{B}}(\epsilon, x_0), V(x') \geq V(x)\}.$$

Note that $\beta \in \mathbb{R}_{>0}$ since it is the infimum of a positive-valued function over the compact set

$$\overline{\mathsf{B}}(\epsilon, x_0) \cap \{x' \in \overline{\mathsf{B}}(\epsilon, x_0) \mid \ V(x') \geq V(x)\}.$$

missing stuff Now we calculate, using Lemma 4.3.26,

$$\begin{aligned} V(\xi(t)) &= V(\xi(t_0)) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\xi(\tau))\, \mathrm{d}\tau \\ &= V(x) + \int_{t_0}^{t} \mathscr{L}_{F_0} V(\xi(\tau))\, \mathrm{d}\tau \\ &\geq V(x) + \beta(t - t_0). \end{aligned}$$

Thus $t \mapsto V(\xi(t))$ is unbounded as $t \to \infty$, which is a contradiction since $x \mapsto V(x)$ is bounded on $\overline{\mathsf{B}}(\epsilon, x_0)$. Thus we conclude that $\xi(T) \notin \overline{\mathsf{B}}(\epsilon, x_0)$ for some $T \geq t_0$. This gives the desired instability. ■

4.3.8.2 Instability theorem for linear equations with constant coefficients

Next we consider an instability theorem for linear homogeneous ordinary differential equations with constant coefficients. The result we give is one that makes use of very particular attributes of linear ordinary differential equations, and, in particular, makes use of the notion of observability introduced in Definition 4.3.47.

4.3.52 Theorem (An instability theorem for linear ordinary differential equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned}\widehat{\mathrm{F}}\colon \mathbb{T}\times \mathsf{V} &\to \mathsf{V}\\ (\mathrm{t},\mathrm{x}) &\mapsto \mathrm{A}(\mathrm{x})\end{aligned}$$

for $\mathrm{A}\in \mathrm{L}(\mathsf{V};\mathsf{V})$*. Suppose that* $\sup \mathbb{T} = \infty$*. Suppose that* V *has an inner product* $\langle\cdot,\cdot\rangle$*. Then* F *is unstable if there exists* $\mathrm{P},\mathrm{Q}\in \mathrm{L}(\mathsf{V};\mathsf{V})$ *with the following properties:*

(i) P *and* Q *are symmetric;*

(ii) (P, Q) *is a Lyapunov pair for* F*;*

(iii) P *is not positive-semidefinite;*

(iv) Q *is positive-semidefinite;*

(v) (A, Q) *is observable.*

Proof Since Q is positive-semidefinite and (A, Q) is observable, the argument from the proof of Theorem 4.3.49 shows that there are no nontrivial periodic solutions for F. Thus this part of the theorem will follow if we can show that F is not asymptotically stable. By hypothesis, there exists $x_0 \in \mathsf{V}$ so that $f_P(x_0) < 0$. Let $\xi(t) = \mathrm{e}^{At}(x_0)$ be the solution of F with initial condition x_0 at $t = 0$. As in the proof of Theorem 4.3.39(i), we have $f_P \circ \xi(t) \leq f_P(x_0) < 0$ for all $t \geq 0$ since Q is positive-semidefinite. Denote

$$r = \inf\{\|x\| \mid f_P(x) \leq f_P(x_0)\},$$

and observe that $r \in \mathbb{R}_{>0}$. We have shown that $\|\xi(t)\| \geq r$ for all $t \geq 0$. This prohibits internal asymptotic stability, and in this case, internal stability. ■

Let us use this theorem to fill in a few gaps left by our treatment of Example 4.3.41.

4.3.53 Example (Example 4.3.41 (cont'd)) We continue with the linear homogeneous ordinary differential equation $\boldsymbol{F}$ on $\mathsf{V} = \mathbb{R}^2$ defined by the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1\\ -b & -a\end{bmatrix}.$$

Again, we use the standard inner product.

We consider here the unstable cases.

1. $a = 0$ and $b = 0$: In this case, by Exercise 4.3.5, if $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$ with $\boldsymbol{Q}$ positive-semidefinite, then $(\boldsymbol{A}, \boldsymbol{Q})$ is not observable. This means that we cannot conclude instability using Theorem 4.3.52.
2. $a = 0$ and $b < 0$: If we define
$$\boldsymbol{P} = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} -b & 0 \\ 0 & 1 \end{bmatrix},$$
then one verifies that $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. However, $\boldsymbol{P}$ is not positive-semidefinite (its eigenvalues are $\{\pm\frac{1}{2}\}$), while $\boldsymbol{Q}$ is positive-definite. Since $\boldsymbol{Q}$ is invertible, one can immediately conclude observability, and, therefore, conclude instability from Theorem 4.3.52.
3. $a > 0$ and $b < 0$: We use the Lyapunov pair $(\boldsymbol{P}, \boldsymbol{Q})$ with
$$\boldsymbol{P} = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix}.$$
Here we compute
$$\boldsymbol{O}(\boldsymbol{A}, \boldsymbol{Q}) = \begin{bmatrix} 0 & 0 \\ 0 & 2a \\ 0 & 0 \\ -2ab & -2a^2 \end{bmatrix}.$$
Since $\boldsymbol{P}$ is not positive-semidefinite, since $\boldsymbol{Q}$ is positive-semidefinite, and since $(\boldsymbol{A}, \boldsymbol{Q})$ is observable we conclude from Theorem 4.3.52 that $\boldsymbol{F}$ is unstable.
4. $a < 0$ and $b = 0$: In this case, as in case 1, if $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$ with $\boldsymbol{Q}$ positive-semidefinite, then $(\boldsymbol{A}, \boldsymbol{Q})$ is not observable. Thus the instability that holds in this case cannot be determined from Theorem 4.3.52.
5. $a < 0$ and $b > 0$: We note that, if
$$\boldsymbol{P} = \begin{bmatrix} -b & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & -2a \end{bmatrix},$$
then $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. We also have
$$\boldsymbol{O}(\boldsymbol{A}, \boldsymbol{Q}) = \begin{bmatrix} 0 & 0 \\ 0 & -2a \\ 0 & 0 \\ 2ab & 2a^2 \end{bmatrix}.$$
Thus $(\boldsymbol{A}, \boldsymbol{Q})$ is observable. Since $\boldsymbol{P}$ is not positive-definite and since $\boldsymbol{Q}$ is positive-semidefinite, we conclude from Theorem 4.3.52 that $\boldsymbol{F}$ is unstable.
6. Here we again take
$$\boldsymbol{P} = \begin{bmatrix} -b & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & -2a \end{bmatrix}.$$
The same argument as in the previous case will tell us that $\boldsymbol{F}$ is unstable. •

4.3.9 Converse theorems

The results of Sections 4.3.3, 4.3.4, 4.3.5, and 4.3.6 provide useful sufficient conditions for stability and asymptotic stability of equilibria. However, if there are lots of examples of ordinary differential equations that are stable, but for which the hypotheses of these theorems do not hold, then this reduces their potential effectiveness in practice. For this reason, in this section we give six so-called "converse theorems," i.e., theorems that assert the manner in which the converses of conditions like those in the preceding sections also hold. One is for general, nonautonomous, not necessarily linear ordinary differential equations. The next is for exponential stability for nonautonomous ordinary differential equations. Both of these results are mirrored for autonomous systems, with self-contained proof for readers wishing to sidestep time dependence. The other two are results for linear homogeneous ordinary differential equations, one a result for time-varying equations and the other a result for equations with constant coefficients.

4.3.9.1 Converse theorems for nonautonomous equations We begin with the most general result.

4.3.54 Theorem (A converse theorem for nonautonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^n$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$, $\mathrm{T}_- \triangleq \inf \mathbb{T} > -\infty$, *and that* $\mathbf{F}$ *satisfies Assumption 4.1.1. If* $\mathbf{x}_0$ *is uniformly asymptotically stable, then there exists* $\mathrm{V}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}$ *such that*

(i) V *is of class* C^1,

(ii) $\mathrm{V} \in \mathrm{TVLPD}_{s_0}(\mathbf{x}_0)$,

(iii) $\mathrm{V} \in \mathrm{TVLD}_{s_0}(\mathbf{x}_0)$,

(iv) $(t, \mathbf{x}) \mapsto \frac{\partial \mathrm{V}}{\partial x_j}(t, \mathbf{x})$ *is in* $\mathrm{TVLD}_{s_0}(\mathbf{x}_0)$, *and*

(v) $-\mathscr{L}_{\mathbf{F}}\mathrm{V} \in \mathrm{TVLPD}_{s_0}(\mathbf{x}_0)$.

Proof ■

Next we specialise the preceding result to exponential stability, not just asymptotic stability.

4.3.55 Theorem (A converse theorem for exponential stability of nonautonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^n$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$. *Assume that* $\sup \mathbb{T} = \infty$, $\mathrm{T}_- \triangleq \inf \mathbb{T} > -\infty$, *and that there exists* $\mathrm{M}, \mathrm{r} \in \mathbb{R}_{>0}$ *such that*

$$\left| \frac{\partial \widehat{\mathrm{F}_\mathrm{j}}}{\partial \mathrm{x}_\mathrm{k}}(\mathrm{t}, \mathbf{x}) \right| \leq \mathrm{M}, \qquad \mathrm{j}, \mathrm{k} \in \{1, \dots, \mathrm{n}\}, \ (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0).$$

If there exist $\mathrm{L}, \delta, \sigma \in \mathbb{R}_{>0}$ *such that, if* $\mathbf{x} \in \mathrm{U}$ *satisfies* $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, *then* $\mathrm{t} \mapsto \Phi^{\mathbf{F}}(\mathrm{t}, \mathrm{t}_0, \mathbf{x}_0)$ *is defined on* $[\mathrm{t}_0, \infty)$ *and satisfies*

$$\|\Phi^{\mathbf{F}}(\mathrm{t}, \mathrm{t}_0, \mathbf{x}) - \mathbf{x}_0\| \leq \mathrm{L} \mathrm{e}^{-\sigma(\mathrm{t}-\mathrm{t}_0)} \|\mathbf{x} - \mathbf{x}_0\|,$$

then there exist $\mathrm{V} \colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}$ *and* $\mathrm{r}_0 \in \mathbb{R}_{>0}$ *such that*

(i) V *is of class* C^1;

(ii) there exists $\mathrm{C}_1 \in \mathbb{R}_{>0}$ *such that*

$$\left\| \frac{\partial \mathrm{V}}{\partial \mathrm{x}_\mathrm{j}}(\mathrm{t}, \mathbf{x}) \right\| \leq \mathrm{C}_1 \|\mathbf{x} - \mathbf{x}_0\|, \qquad \mathrm{j} \in \{1, \dots, \mathrm{n}\}, \ (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \mathsf{B}(\mathrm{r}_0, \mathbf{x}_0);$$

(iii) there exists $\mathrm{C}_2 \in \mathbb{R}_{>0}$ *such that*

$$\mathrm{C}_2 \|\mathbf{x} - \mathbf{x}_0\|^2 \leq \mathrm{V}(\mathrm{t}, \mathbf{x}) \leq \mathrm{C}_2^{-1} \|\mathbf{x} - \mathbf{x}_0\|^2, \qquad (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \mathsf{B}(\mathrm{r}_0, \mathbf{x}_0);$$

(iv) there exists $\mathrm{C}_3 \in \mathbb{R}_{>0}$ *such that*

$$\mathscr{L}_{\mathbf{F}} \mathrm{V}(\mathrm{t}, \mathbf{x}) \leq -\mathrm{C}_3 \|\mathbf{x} - \mathbf{x}_0\|^2, \qquad (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \mathsf{B}(\mathrm{r}_0, \mathbf{x}_0).$$

Proof We start with a few technical lemmata.

1 Lemma ***missing stuff*** *If* $\mathbb{T}$ *is an interval and if* $\boldsymbol{\gamma} \colon \mathbb{T} \to \mathbb{R}^\mathrm{n}$ *is of class* C^1, *then*

$$\frac{\mathrm{d}}{\mathrm{dt}} \|\boldsymbol{\gamma}(\mathrm{t})\| \leq \left\| \frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{dt}}(\mathrm{t}) \right\|.$$

Proof The first thing we need to do is understand what we mean by $\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{\gamma}(t)\|$, since it may be that $t \mapsto \|\boldsymbol{\gamma}(t)\|$ is not differentiable. We shall use the notion of weak differentiability from ***missing stuff***. First let us suppose that $\boldsymbol{\gamma}(t) \neq 0$. Then, by continuity, $\boldsymbol{\gamma}(\tau) \neq 0$ for τ nearby t. Then,

$$2\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\|\boldsymbol{\gamma}(\tau)\|\right)\|\boldsymbol{\gamma}(t)\| = \frac{\mathrm{d}}{\mathrm{d}\tau}\|\boldsymbol{\gamma}(\tau)\|^2 = 2\left\langle \frac{\mathrm{d}}{\mathrm{d}\tau}\boldsymbol{\gamma}(\tau), \boldsymbol{\gamma}(\tau) \right\rangle_{\mathbb{R}^n}.$$

Then, by the Cauchy–Bunyakovsky–Schwarz inequality,

$$2\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\|\boldsymbol{\gamma}(\tau)\|\right)\|\boldsymbol{\gamma}(t)\| = 2\left\langle \frac{\mathrm{d}}{\mathrm{d}\tau}\boldsymbol{\gamma}(\tau), \boldsymbol{\gamma}(\tau) \right\rangle_{\mathbb{R}^n} \leq 2\left\| \frac{\mathrm{d}}{\mathrm{d}\tau}\boldsymbol{\gamma}(\tau) \right\| \|\boldsymbol{\gamma}(\tau)\|.$$

Thus, when $\gamma(t) \neq 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma(t)\| \leq \left\|\frac{\mathrm{d}}{\mathrm{d}\tau}\gamma(\tau)\right\|.$$

We need to account for the possibility that $\gamma(t)$ may be zero. Note that

$$\Gamma \triangleq \{t \in \mathbb{T} \mid \|\gamma(t)\| > 0\}$$

is open. Thus, by ***missing stuff***, there exists a finite or countable set J and a collection I_j, $j \in J$, of open intervals such that $\Gamma = \cup_{j=1}^{\infty} I_j$. Let $\phi \in \mathcal{D}(\mathbb{T}; \mathbb{R})$. Then

$$\begin{aligned}\int_{\mathbb{T}} \|\gamma(t)\|\dot{\phi}(t)\,\mathrm{d}t &= \sum_{j\in J} \int_{I_j} \|\gamma(t)\|\dot{\phi}(t)\,\mathrm{d}t \\ &= -\sum_{j\in J} \int_{I_j} \frac{\langle \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t), \gamma(t)\rangle}{\|\gamma(t)\|}\phi(t)\,\mathrm{d}t.\end{aligned}$$

using integration by parts. Thus $t \mapsto \|\gamma(t)\|$ is differentiable in the sense of distributions, and its derivative in this sense is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma(t)\| = \begin{cases} \frac{\langle \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t), \gamma(t)\rangle}{\|\gamma(t)\|}, & \gamma(t) \neq \mathbf{0}, \\ 0, & \gamma(t) = 0. \end{cases}$$

The lemma now follows from our estimates above. ▼

2 Lemma *Let $\mathbb{T}$ be an interval and let $\alpha, \beta, \xi \colon \mathbb{T} \to \mathbb{R}$ be such that*

(i) *α and β are continuous,*

(ii) *ξ is continuously differentiable, and*

(iii) *$\alpha(\mathrm{t})\xi(\mathrm{t}) \leq \dot{\xi}(\mathrm{t}) \leq \beta(\mathrm{t})\xi(\mathrm{t})$, $\mathrm{t} \in \mathbb{T}$.*

Then, for any $\mathrm{t}_0 \in \mathbb{T}$,

$$\xi(\mathrm{t}_0)\mathrm{e}^{\int_{\mathrm{t}_0}^{\mathrm{t}} \alpha(\tau)\,\mathrm{d}\tau} \leq \xi(\mathrm{t}) \leq \xi(\mathrm{t}_0)\mathrm{e}^{\int_{\mathrm{t}_0}^{\mathrm{t}} \beta(\tau)\,\mathrm{d}\tau}, \qquad \mathrm{t} \geq \mathrm{t}_0.$$

Proof Denote $\eta \colon \mathbb{T} \to \mathbb{R}$ by

$$\eta(t) = \exp^{\int_{t_0}^{t} \beta(\tau)\,\mathrm{d}\tau}.$$

A direct computation gives

$$\frac{\mathrm{d}\eta}{\mathrm{d}t}(t) = \beta(t)\eta(t), \qquad t \in \mathbb{T}.$$

Noting that $\eta(t) > 0$ for every $t \in \mathbb{T}$, we then have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\xi(t)}{\eta(t)}\right) = \frac{\eta(t)\frac{\mathrm{d}}{\mathrm{d}t}\xi(t) - \xi(t)\frac{\mathrm{d}}{\mathrm{d}t}\eta(t)}{\eta(t)^2} = \frac{1}{\eta(t)}\left(\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) - \beta(t)\xi(t)\right) \leq 0$$

for $t \geq t_0$. Thus we have

$$\frac{\xi(t)}{\eta(t)} \leq \frac{\xi(t_0)}{\eta(t_0)} = \xi(t_0), \qquad t \geq t_0.$$

This gives the rightmost inequality in the statement of the lemma. The leftmost inequality follows from this by replacing "ξ" with "$-\xi$" and "β" with "α." ▼

3 Lemma *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$

and let $\mathbf{x}_0 \in \mathrm{U}$. *If there exists* $\mathrm{L} \in \mathbb{R}_{>0}$ *such that*

$$\|\widehat{\mathbf{F}}(\mathrm{t}, \mathbf{x})\| \leq \mathrm{L}\|\mathbf{x} - \mathbf{x}_0\|, \qquad (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \mathrm{U},$$

then, for $(\mathrm{t}_0, \mathbf{x}) \in \mathbb{T} \times \mathrm{U}$, *the solution* ξ *to the initial value problem*

$$\dot{\xi}(\mathrm{t}) = \widehat{\mathbf{F}}(\mathrm{t}, \xi(\mathrm{t})), \quad \xi(\mathrm{t}_0) = \mathbf{x},$$

satisfies

(i) $|\frac{\mathrm{d}}{\mathrm{dt}}\|\xi(\mathrm{t}) - \mathbf{x}_0\|^2| \leq 2\mathrm{L}\|\xi(\mathrm{t}) - \mathbf{x}_0\|^2$, $\mathrm{t} \geq \mathrm{t}_0$, *and*

(ii) $\|\mathbf{x} - \mathbf{x}_0\|\mathrm{e}^{-\mathrm{L}(\mathrm{t}-\mathrm{t}_0)} \leq \|\xi(\mathrm{t}) - \mathbf{x}_0\| \leq \|\mathbf{x} - \mathbf{x}_0\|\mathrm{e}^{\mathrm{L}(\mathrm{t}-\mathrm{t}_0)}$, $\mathrm{t} \geq \mathrm{t}_0$.

Proof We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t) - x_0\|^2 = 2\left\langle \frac{\mathrm{d}}{\mathrm{d}t}\xi(t), \xi(t) - x_0 \right\rangle_{\mathbb{R}^n}.$$

Thus, by the Cauchy–Bunyakovsky–Schwarz inequality,

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t) - x_0\|^2\right| &\leq 2\left\|\frac{\mathrm{d}}{\mathrm{d}t}\xi(t)\right\| \|\xi(t) - x_0\| \\ &= 2\|\widehat{F}(t, \xi(t))\|\|\xi(t) - x_0\| \\ &\leq 2L\|\xi(t) - x_0\|^2,\end{aligned}$$

giving the first part of the result.

For the second part, we first note that, from the first part of the lemma,

$$-2L\|\xi(t) - x_0\|^2 \leq \frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t) - x_0\|^2 \leq 2L\|\xi(t) - x_0\|^2.$$

The second part of the current lemma follows from Lemma 2. ▼

Let $r_0 < \min\{\delta, \frac{r}{L}\}$ and let $h = \frac{\ln(2k^2)}{2\sigma}$. Define

$$V(t,x) = \int_t^{t+h} \|\Phi^F(\tau,t,x) - x_0\|^2 \,\mathrm{d}\tau.$$

Then we have

$$V(t,x) \le L^2\|x - x_0\|^2 \int_t^{t+h} \mathrm{e}^{-2\sigma(\tau-t)} \,\mathrm{d}\tau = \frac{L^2\|x-x_0\|^2(1-\mathrm{e}^{-2\sigma h})}{2\sigma}.$$

By Lemma 3 we have

$$\|\Phi^F(\tau,t,x) - x_0\|^2 \ge \mathrm{e}^{-2L(t-\tau)}\|x - x_0\|^2,$$

from which we conclude that

$$V(t,x) \ge \|x - x_0\|^2 \int_t^{t+h} \mathrm{e}^{-2L(t-\tau)} \,\mathrm{d}\tau = \frac{\|x-x_0\|^2(1-\mathrm{e}^{-2Lh})}{2L}.$$

Taking

$$C_2 = \min\left\{\frac{L^2(1-\mathrm{e}^{-2\sigma h})}{2\sigma}, \frac{1-\mathrm{e}^{-2Lh}}{2L}\right\}$$

gives condition (iii).

By Theorem 3.1.8, solutions depend continuously differentiably on initial condition and time. Therefore, by ***missing stuff***, we can differentiate V under the integral sign:

$$\frac{\partial V}{\partial t}(t,x) = \|\Phi^F(t+h,t,x) - x_0\|^2 - \|\Phi^F(t,t,x)\|^2 + 2\int_t^{t+h} \left\langle \Phi^F(\tau,t,x) - x_0, \frac{\mathrm{d}}{\mathrm{d}t}\Phi^F(\tau,t,x)\right\rangle \mathrm{d}\tau$$

and

$$\frac{\partial V}{\partial x_j}(t,x) = 2\int_t^{t_h} \left\langle \Phi^F(\tau,t,x) - x_0, \frac{\partial}{\partial x_j}\Phi^F(\tau,t,x)\right\rangle \mathrm{d}\tau.$$

By Exercise 1.4.5 and the preceding two equations we then deduce that

$$\begin{aligned}\mathscr{L}_F V(t,x) &= \|\Phi^F(t+h,t,x) - x_0\|^2 - \|x - x_0\|^2 \\ &\le -(1 - L^2\mathrm{e}^{-2\sigma h})\|x - x_0\|^2 \le -\frac{1}{2}\|x - x_0\|^2,\end{aligned}$$

giving condition (iv).

Now we note, by the Chain Rule, that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \Phi_j^F}{\partial x_k}(\tau,t,x)\right) = \sum_{l=1}^n \frac{\partial \widehat{F}_j}{\partial x_l}(\tau,t,x)\frac{\partial \Phi_l^F}{\partial x_k}(\tau,t,x), \qquad j \in \{1,\dots,n\},$$

and that

$$\frac{\partial \Phi_j^{\boldsymbol{F}}}{\partial x_k}(t,t,\boldsymbol{x}) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

That is to say, the Jacobian matrix of $\Phi^{\boldsymbol{F}}$ satisfies a linear ordinary differential equation with initial condition being the identity matrix. We wish to use Lemma 3 with $\boldsymbol{x}_0$ being the zero matrix and $\boldsymbol{x} = \boldsymbol{I}_n$. To do so, we need to estimate the right-hand side of the preceding equation:

$$\begin{aligned}
\left(\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\frac{\partial \widehat{F}_j}{\partial x_l}(\tau,t,\boldsymbol{x})\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right)^2\right)^{1/2} &\leq \left(\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\left|\frac{\partial \widehat{F}_j}{\partial x_l}(\tau,t,\boldsymbol{x})\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|\right)^2\right)^{1/2} \\
&\leq \left(\sum_{j,k=1}^{n}\left(M\sum_{l=1}^{n}\left|\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|\right)^2\right)^{1/2} \\
&\leq \left(M^2\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\left|\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|\right)^2\right)^{1/2} \\
&\leq \left(M^2\sum_{j,k=1}^{n}\sum_{l=1}^{n}\left|\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|^2\right)^{1/2} \\
&\leq \left(M^2 n\sum_{k,l=1}^{n}\left|\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|^2\right)^{1/2} \\
&\leq M\sqrt{n}\left(\sum_{k,l=1}^{n}\left|\frac{\partial \Phi_l^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|^2\right)^{1/2}.
\end{aligned}$$

Here we have used the hypotheses on $\widehat{\boldsymbol{F}}$. Now we can use Lemma 3 to conclude that

$$\left(\sum_{j,k=1}^{n}\left|\frac{\partial \Phi_j^{\boldsymbol{F}}}{\partial x_k}(\tau,t,\boldsymbol{x})\right|^2\right)^{1/2} \leq \sqrt{n}\mathrm{e}^{M\sqrt{n}(\tau-t)}.$$

Therefore,

$$\left(\sum_{k=1}^{n}\left(\frac{\partial \Phi_k^{\boldsymbol{F}}}{\partial x_j}(\tau,t,\boldsymbol{x})\right)^2\right)^{1/2} \leq \sqrt{n}\mathrm{e}^{M\sqrt{n}(\tau-t)}.$$

Thus, using the Cauchy–Bunyakovsky–Schwarz inequality,

$$\begin{aligned}
\left|\frac{\partial V}{\partial x_j}(t,\boldsymbol{x})\right| &\leq 2\int_t^{t+h}\left|\left\langle \Phi^{\boldsymbol{F}}(\tau,t,\boldsymbol{x}) - \boldsymbol{x}_0, \frac{\partial}{\partial x_j}\Phi^{\boldsymbol{F}}(\tau,t,\boldsymbol{x})\right\rangle\right| \mathrm{d}\tau \\
&\leq 2L\sqrt{n}\|\boldsymbol{x}-\boldsymbol{x}_0\|\int_t^{t+h}\mathrm{e}^{-\sigma(\tau-t)}\mathrm{e}^{M\sqrt{n}(\tau-t)}\,\mathrm{d}\tau,
\end{aligned}$$

giving condition (ii). ■

4.3.9.2 Converse theorems for autonomous equations We now consider converse theorems for autonomous ordinary differential equations. The results essentially follow from those of the preceding section, but here we state and prove them independently for readers not needing to deal with time-varying equations.

4.3.56 Theorem (A converse theorem for autonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} &\to \mathbb{R}^n\\ (t,\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathbf{x}),\end{aligned}$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$, $T_- \triangleq \inf \mathbb{T} > -\infty$, *and that* $\mathbf{F}$ *satisfies Assumption 4.1.1. If* $\mathbf{x}_0$ *is asymptotically stable, then there exists* $\mathrm{V}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}$ *such that*

(i) V *is of class* C^1,

(ii) $\mathrm{V} \in \mathrm{LPD}_{s_0}(\mathbf{x}_0)$,

(iii) $\mathrm{V} \in \mathrm{LD}_{s_0}(\mathbf{x}_0)$,

(iv) $(t,\mathbf{x}) \mapsto \frac{\partial \mathrm{V}}{\partial x_j}(t,\mathbf{x})$ *is in* $\mathrm{LD}_{s_0}(\mathbf{x}_0)$, *and*

(v) $-\mathscr{L}_{\mathbf{F}}\mathrm{V} \in \mathrm{LPD}_{s_0}(\mathbf{x}_0)$.

4.3.57 Theorem (A converse theorem for exponential stability of autonomous ordinary differential equations) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}^n$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$, $T_- \triangleq \inf \mathbb{T} > -\infty$, *and that there exists* $M, r \in \mathbb{R}_{>0}$ *such that*

$$\left|\frac{\partial \widehat{\mathrm{F}}_{0,j}}{\partial x_k}(\mathbf{x})\right| \le M, \qquad j,k \in \{1,\dots,n\},\ \mathbf{x} \in \mathsf{B}(r,\mathbf{x}_0).$$

If there exist $L, \delta, \sigma \in \mathbb{R}_{>0}$ *such that, if* $\mathbf{x} \in \mathrm{U}$ *satisfies* $\|\mathbf{x}-\mathbf{x}_0\| < \delta$, *then* $t \mapsto \Phi^{\mathbf{F}}(t,t_0,\mathbf{x}_0)$ *is defined on* $[t_0,\infty)$ *and satisfies*

$$\|\Phi^{\mathbf{F}}(t,t_0,\mathbf{x}) - \mathbf{x}_0\| \le L e^{-\sigma(t-t_0)}\|\mathbf{x}-\mathbf{x}_0\|,$$

then there exist $\mathrm{V}\colon \mathrm{U} \to \mathbb{R}$ *and* $r_0 \in \mathbb{R}_{>0}$ *such that*

(i) V *is of class* C^1;

(ii) there exists $C_1 \in \mathbb{R}_{>0}$ *such that*

$$\left\|\frac{\partial \mathrm{V}}{\partial x_j}(\mathbf{x})\right\| \le C_1\|\mathbf{x}-\mathbf{x}_0\|, \qquad j \in \{1,\dots,n\},\ \mathbf{x} \in \mathsf{B}(r_0,\mathbf{x}_0);$$

(iii) *there exists* $C_2 \in \mathbb{R}_{>0}$ *such that*

$$C_2\|x - x_0\|^2 \leq V(x) \leq C_2^{-1}\|x - x_0\|^2, \qquad x \in B(r_0, x_0);$$

(iv) *there exists* $C_3 \in \mathbb{R}_{>0}$ *such that*

$$\mathscr{L}_F V(x) \leq -C_3\|x - x_0\|^2, \qquad x \in B(r_0, x_0).$$

Proof We start with a few technical lemmata.

1 Lemma *Let* F *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned} \widehat{F}\colon \mathbb{T} \times U &\to \mathbb{R}^n \\ (t, x) &\mapsto F_0(x) \end{aligned}$$

and let $x_0 \in U$. *If there exists* $L \in \mathbb{R}_{>0}$ *such that*

$$\|\widehat{F}_0(x)\| \leq L\|x - x_0\|, \qquad x \in U,$$

then, for $t_0, x) \in \mathbb{T} \times U$,

(i) $|\frac{d}{dt}\|\Phi^F(t, t_0, x) - x_0\|^2| \leq 2L\|\Phi^F(t, t_0, x) - x_0\|^2$, $t \geq t_0$, *and*

(ii) $\|x - x_0\|e^{-L(t-t_0)} \leq \|\Phi^F(t, t_0, x) - x_0\| \leq \|x - x_0\|e^{L(t-t_0)}$, $t \geq t_0$.

Proof We compute

$$\frac{d}{dt}\|\Phi^F(t, t_0, x) - x_0\|^2 = 2\left\langle \frac{d}{dt}\Phi^F(t, t_0, x), \Phi^F(t, t_0, x) - x_0 \right\rangle_{\mathbb{R}^n}.$$

Thus, by the Cauchy–Bunyakovsky–Schwarz inequality,

$$\begin{aligned} \left|\frac{d}{dt}\|\Phi^F(t, t_0, x) - x_0\|^2\right| &\leq 2\left\|\frac{d}{dt}\Phi^F(t, t_0, x)\right\| \|\Phi^F(t, t_0, x) - x_0\| \\ &= 2\|\widehat{F}(t, \Phi^F(t, t_0, x))\|\|\Phi^F(t, t_0, x) - x_0\| \\ &\leq 2L\|\Phi^F(t, t_0, x) - x_0\|^2, \end{aligned}$$

giving the first part of the result.

For the second part, we first note that, from the first part of the lemma,

$$-2L\|\Phi^F(t, t_0, x) - x_0\|^2 \leq \frac{d}{dt}\|\Phi^F(t, t_0, x) - x_0\|^2 \leq 2L\|\Phi^F(t, t_0, x) - x_0\|^2.$$

The second part of the current lemma follows from Lemma 2. ▼

Let $r_0 < \min\{\delta, \frac{r}{L}\}$ and let $h = \frac{\ln(2k^2)}{2\sigma}$. For some $t \in \mathbb{T}$, define

$$V(x) = \int_t^{t+h} \|\Phi^F(\tau, t, x) - x_0\|^2 \, d\tau.$$

Note that $V(\boldsymbol{x})$ is independent of t by Exercise 1.3.19. Then we have

$$V(\boldsymbol{x}) \leq L^2\|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \int_t^{t+h} \mathrm{e}^{-2\sigma(\tau-t)}\,\mathrm{d}\tau = \frac{L^2\|\boldsymbol{x}-\boldsymbol{x}_0\|^2(1-\mathrm{e}^{-2\sigma h})}{2\sigma}.$$

By Lemma 1 we have

$$\|\Phi^F(\tau,t,\boldsymbol{x})-\boldsymbol{x}_0\|^2 \geq \mathrm{e}^{-2L(t-\tau)}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2,$$

from which we conclude that

$$V(\boldsymbol{x}) \geq \|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \int_t^{t+h} \mathrm{e}^{-2L(t-\tau)}\,\mathrm{d}\tau = \frac{\|\boldsymbol{x}-\boldsymbol{x}_0\|^2(1-\mathrm{e}^{-2Lh})}{2L}.$$

Taking

$$C_2 = \min\left\{\frac{L^2(1-\mathrm{e}^{-2\sigma h})}{2\sigma}, \frac{1-\mathrm{e}^{-2Lh}}{2L}\right\}$$

gives condition (iii).

By Theorem 3.1.8, solutions depend continuously differentiably on initial condition and time. Therefore, by ***missing stuff***, we can differentiate V under the integral sign:

$$\frac{\partial V}{\partial x_j}(\boldsymbol{x}) = 2\int_t^{t+h} \left\langle \Phi^F(\tau,t,\boldsymbol{x})-\boldsymbol{x}_0, \frac{\partial}{\partial x_j}\Phi^F(\tau,t,\boldsymbol{x})\right\rangle \mathrm{d}\tau.$$

By Exercise 1.4.5 and the preceding two equations we then deduce that

$$\begin{aligned}\mathscr{L}_F V(\boldsymbol{x}) &= \|\Phi^F(t+h,t,\boldsymbol{x})-\boldsymbol{x}_0\|^2 - \|\boldsymbol{x}-\boldsymbol{x}_0\|^2\\ &\leq -(1-L^2\mathrm{e}^{-2\sigma h})\|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \leq -\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2,\end{aligned}$$

giving condition (iv).

Now we note, by the Chain Rule, that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \Phi_j^F}{\partial x_k}(\tau,t,\boldsymbol{x})\right) = \sum_{l=1}^n \frac{\partial \widehat{F}_j}{\partial x_l}(\tau,t,\boldsymbol{x})\frac{\partial \Phi_l^F}{\partial x_k}(\tau,t,\boldsymbol{x}), \qquad j\in\{1,\dots,n\},$$

and that

$$\frac{\partial \Phi_j^F}{\partial x_k}(t,t,\boldsymbol{x}) = \begin{cases}1, & j=k,\\ 0, & j\neq k.\end{cases}$$

That is to say, the Jacobian matrix of Φ^F satisfies a linear ordinary differential equation with initial condition being the identity matrix. We wish to use Lemma 1

with x_0 being the zero matrix and $x = I_n$. To do so, we need to estimate the right-hand side of the preceding equation:

$$\begin{aligned}
\left(\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\frac{\partial\widehat{F}_j}{\partial x_l}(\tau,t,x)\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right)^2\right)^{1/2} &\le \left(\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\left|\frac{\partial\widehat{F}_j}{\partial x_l}(\tau,t,x)\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|\right)^2\right)^{1/2}\\
&\le \left(\sum_{j,k=1}^{n}\left(M\sum_{l=1}^{n}\left|\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|\right)^2\right)^{1/2}\\
&\le \left(M^2\sum_{j,k=1}^{n}\left(\sum_{l=1}^{n}\left|\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|\right)^2\right)^{1/2}\\
&\le \left(M^2\sum_{j,k=1}^{n}\sum_{l=1}^{n}\left|\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|^2\right)^{1/2}\\
&\le \left(M^2 n\sum_{k,l=1}^{n}\left|\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|^2\right)^{1/2}\\
&\le M\sqrt{n}\left(\sum_{k,l=1}^{n}\left|\frac{\partial\Phi_l^F}{\partial x_k}(\tau,t,x)\right|^2\right)^{1/2}.
\end{aligned}$$

Here we have used the hypotheses on $\widehat{F}$. Now we can use Lemma 1 to conclude that

$$\left(\sum_{j,k=1}^{n}\left|\frac{\partial\Phi_j^F}{\partial x_k}(\tau,t,x)\right|^2\right)^{1/2} \le \sqrt{n}\mathrm{e}^{M\sqrt{n}(\tau-t)}.$$

Therefore,

$$\left(\sum_{k=1}^{n}\left(\frac{\partial\Phi_k^F}{\partial x_j}(\tau,t,x)\right)^2\right)^{1/2} \le \sqrt{n}\mathrm{e}^{M\sqrt{n}(\tau-t)}.$$

Thus, using the Cauchy–Bunyakovsky–Schwarz inequality,

$$\begin{aligned}
\left|\frac{\partial V}{\partial x_j}(t,x)\right| &\le 2\int_t^{t+h}\left|\left\langle\Phi^F(\tau,t,x)-x_0,\frac{\partial}{\partial x_j}\Phi^F(\tau,t,x)\right\rangle\right|\,\mathrm{d}\tau\\
&\le 2L\sqrt{n}\|x-x_0\|\int_t^{t+h}\mathrm{e}^{-\sigma(\tau-t)}\mathrm{e}^{M\sqrt{n}(\tau-t)}\,\mathrm{d}\tau,
\end{aligned}$$

giving condition (ii). ■

4.3.9.3 Converse theorem for time-varying linear equations Next we turn to converse results for linear ordinary differential equations. The first is for time-varying equations.

4.3.58 Theorem (A converse theorem for time-varying linear ordinary differential equations) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned}\widehat{F}\colon \mathbb{T} \times V &\to V\\ (t, x) &\mapsto A(t)(x)\end{aligned}$$

for $A\colon \mathbb{T} \to L(V; V)$ *continuous and bounded. Suppose that* $\sup \mathbb{T} = \infty$. *Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$. *Let* $Q\colon \mathbb{T} \to L(V; V)$ *have the following properties:*

(i) Q *is continuous;*

(ii) $Q(t)$ *is symmetric for every* $t \in \mathbb{T}$*;*

(iii) Q *is positive-definite;*

(iv) Q *is decrescent.*

Then there exists $P\colon \mathbb{T} \to L(V; V)$ *with the following properties:*

(i) P *is of class* C^1*;*

(ii) $P(t)$ *is symmetric for every* $t \in \mathbb{T}$*;*

(iii) (P, Q) *is a Lyapunov pair for* F*;*

(iv) P *is positive-definite;*

(v) P *is decrescent.*

Proof By Exercise 4.2.2(f), let $C_1, \sigma \in \mathbb{R}_{>0}$ be such that

$$|||\Phi_A(t, t_0)||| \leq C_1 e^{-\sigma(t-t_0)}, \qquad t \in \mathbb{T},\ t \geq t_0. \tag{4.30}$$

By Lemma 4.3.18, there exists $C_2 \in \mathbb{R}_{>0}$ such that

$$C_2\langle x, x\rangle \leq f_Q(t, x) \leq C_2^{-1}\langle x, x\rangle, \qquad (t, x) \in \mathbb{T} \times V. \tag{4.31}$$

We define

$$P(t) = \int_t^\infty \Phi_A(\tau, t)^T \circ Q(\tau) \circ \Phi_A(\tau, t)\, d\tau.$$

The integral exists by the inequalities (4.30) and (4.31).

For $(t, x) \in \mathbb{T} \times V$ we compute

$$\begin{aligned} f_P(t, x) &= \int_t^\infty f_Q(\tau, \Phi_A(\tau, t)(x))\, d\tau\\ &\leq C_2^{-1} \int_t^\infty \|\Phi_A(\tau, t)(x)\|^2\, d\tau\\ &\leq C_2^{-1}\|x\|^2 \int_t^\infty |||\Phi_A(\tau, t)|||^2\, d\tau\\ &\leq \frac{C_1}{C_2}\|x\|^2 \int_t^\infty e^{-\sigma(\tau-t)}\, d\tau = \frac{C_1}{C_2\sigma}\|x\|^2. \end{aligned}$$

Since A is bounded, there exists $M \in \mathbb{R}_{>0}$ such that $|||A(t)||| \leq M$ for each $t \in \mathbb{T}$, by Lemma 1 from the proof of Theorem 4.3.55 we have

$$\|\Phi_A(\tau, t)(x)\|^2 \geq \|x\|^2 \mathrm{e}^{-2M(\tau - t)}, \qquad \tau \geq t.$$

Therefore,

$$\begin{aligned} f_P(t, x) &= \int_t^\infty f_Q(\tau, \Phi_A(\tau, t)(x)) \, \mathrm{d}\tau \\ &\geq C_2 \int_t^\infty \|\Phi_A(\tau, t)(x)\|^2 \, \mathrm{d}\tau \\ &\geq C_2 \|x\|^2 \int_t^\infty \mathrm{e}^{-2M(\tau - t)} \, \mathrm{d}\tau = \frac{C_2}{2M} \|x\|^2. \end{aligned}$$

Letting $C = \min\{\frac{C_2}{2M}, \frac{C_2 \sigma}{C_1}\}$, we thus have

$$C\langle x, x \rangle \leq f_P(t, x) \leq C^{-1} \langle x, x \rangle,$$

showing that P is positive-definite and decrescent, by Lemma 4.3.18.

By the Fundamental Theorem of Calculus, P is continuously differentiable. By (3.8) we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi_A(\tau, t) = -\Phi_A(\tau, t) \circ A(t).$$

Thus

$$\begin{aligned} \dot{P}(t) = & -Q(t) + \int_t^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t} \Phi_A(\tau, t)^T \right) \circ Q(\tau) \circ \Phi_A(\tau, t) \, \mathrm{d}\tau \\ & + \int_t^\infty \Phi_A(\tau, t)^T \circ Q(\tau) \circ \left(\frac{\mathrm{d}}{\mathrm{d}t} \Phi_A(\tau, t) \right) \mathrm{d}\tau \\ = & -Q(t) - \int_t^\infty A(t)^T \circ \Phi_A(\tau, t)^T \circ Q(\tau) \circ \Phi_A(\tau, t) \, \mathrm{d}\tau \\ & - \int_t^\infty \Phi_A(\tau, t)^T \circ Q(\tau) \circ \Phi_A(\tau, t) \circ A(t) \, \mathrm{d}\tau \\ = & -Q(t) - A(t)^T \circ P(t) - P(t) \circ A(t), \end{aligned}$$

which shows that (P, Q) is a Lyapunov pair for F, as desired. ■

4.3.9.4 Converse theorem for linear equations with constant coefficients

Finally, we give a result for linear ordinary differential equations with constant coefficients. Here the results we give are quite detailed, in keeping with our detailed knowledge of such equations.

4.3.59 Theorem (A converse theorem for linear ordinary differential equations with constant coefficients) *Let* F *be a system of linear homogeneous ordinary differential equations in an* n*-dimensional* $\mathbb{R}$*-vector space* V *with constant coefficients and with right-hand side*

$$\begin{aligned}\widehat{\mathrm{F}}\colon \mathbb{T} \times \mathsf{V} &\to \mathsf{V}\\ (\mathrm{t}, \mathrm{x}) &\mapsto \mathrm{A}(\mathrm{x})\end{aligned}$$

for $\mathrm{A} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$*. Suppose that* $\sup \mathbb{T} = \infty$*. Suppose that* V *has an inner product* $\langle \cdot, \cdot \rangle$*. If* A *is Hurwitz, then the following statements hold:*

(i) for any symmetric $\mathrm{Q} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$*, there exists a unique symmetric* $\mathrm{P} \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ *so that* (P, Q) *is a Lyapunov pair for* F*;*

(ii) if Q *is positive-semidefinite with* P *the unique symmetric linear map for which* (P, Q) *is a Lyapunov pair for* F*, then* P *is positive-semidefinite;*

(iii) if Q *is positive-semidefinite with* P *the unique symmetric linear map for which* (P, Q) *is a Lyapunov pair for* F*, then* P *is positive-definite if and only if* (A, Q) *is observable.*

Proof (i) We claim that, if we define

$$P = \int_0^\infty \mathrm{e}^{A^T t} \circ Q \circ \mathrm{e}^{At} \, \mathrm{d}t, \tag{4.32}$$

then (P, Q) is a Lyapunov pair for F. First note that since A is Hurwitz, the integral does indeed converge by ***missing stuff***. We also have

$$\begin{aligned}A^T \circ P + P \circ A &= A^T \circ \left(\int_0^\infty \mathrm{e}^{A^T t} \circ Q \circ \mathrm{e}^{At} \, \mathrm{d}t \right) + \left(\int_0^\infty \mathrm{e}^{A^T t} \circ Q \circ \mathrm{e}^{At} \, \mathrm{d}t \right) \circ A\\ &= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathrm{e}^{A^T t} \circ Q \circ \mathrm{e}^{At} \right) \mathrm{d}t\\ &= \mathrm{e}^{A^T t} \circ Q \circ \mathrm{e}^{At} \Big|_0^\infty = -Q,\end{aligned}$$

as desired. We now show that P as defined is the *only* symmetric linear map for which (P, Q) is a Lyapunov pair for F. Suppose that $\hat{P}$ also has the property that $(\hat{P}, Q)$ is a Lyapunov pair for F, and let $\Delta = \hat{P} - P$. Then one sees that

$$A^T \circ \Delta + \Delta \circ A = 0.$$

If we let

$$\Lambda(t) = \mathrm{e}^{A^T t} \circ \Delta \circ \mathrm{e}^{At},$$

then

$$\frac{\mathrm{d}\Lambda}{\mathrm{d}t}(t) = \mathrm{e}^{A^T t} \circ \left(A^T \circ \Delta + \Delta \circ A \right) \circ \mathrm{e}^{At} = 0.$$

Therefore, Λ is constant, and since $\Lambda(0) = \Delta$, it follows that $\Lambda(t) = \Delta$ for all t. However, since A is Hurwitz, it also follows that $\lim_{t \to \infty} \Lambda(t) = 0$. Thus $\Delta = 0$, so that $\hat{P} = P$.

(ii) If P is defined by (4.32), then we have

$$f_P(x) = \int_0^\infty \langle Q \circ \mathrm{e}^{At}(x), \mathrm{e}^{At}(x) \rangle \, \mathrm{d}t.$$

Therefore, if Q is positive-semidefinite, it follows that P is positive-semidefinite.

(iii) Here we employ a lemma.

1 Lemma *If* Q *is positive-semidefinite then* (A, Q) *is observable if and only if the linear map* P *defined by (4.32) is invertible.*

Proof First suppose that (A, Q) is observable and let $x \in \ker(P)$. Then

$$\int_0^\infty \langle Q \circ \mathrm{e}^{At}(x), \mathrm{e}^{At}(x) \rangle \, \mathrm{d}t = 0.$$

Since Q is positive-semidefinite, this implies that $\mathrm{e}^{At}(x) \in \ker(Q)$ for all t. Differentiating this inclusion k times with respect to t gives $A^k \circ \mathrm{e}^{At}(x) \in \ker(Q)$ for any $k \in \mathbb{Z}_{>0}$. Evaluating at $t = 0$ shows that $x \in \ker(O(A, C))$. Since (A, Q) is observable, this implies that $x = 0$. Thus we have shown that $\ker(P) = \{0\}$, or equivalently that P is invertible.

Now suppose that P is invertible. Then the expression

$$\int_0^\infty \langle Q \circ \mathrm{e}^{At}(x), \mathrm{e}^{At}(x) \rangle \, \mathrm{d}t$$

is zero if and only if $x = 0$. Since Q is positive-semidefinite, this means that the expression

$$\langle Q \circ \mathrm{e}^{At}(x), \mathrm{e}^{At}(x) \rangle$$

is zero if and only if $x = 0$. Since e^{At} is invertible, this implies that Q must be positive-definite, and in particular, invertible. In this case, (A, Q) is clearly observable. ▼

With the lemma at hand, the remainder of the proof is straightforward. Indeed, from part (ii), we know that P is positive-semidefinite. The lemma now says that P is positive-definite if and only if (A, Q) is observable, as desired. ■

Let us resume our example started as Example 4.3.41.

4.3.60 Example (Example 4.3.41 cont'd) We resume looking at the case where

$$A = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}.$$

Let us look at a few cases to flesh out some aspects of Theorem 4.3.59.

1. $a > 0$ and $b > 0$: This is exactly the case when $\boldsymbol{A}$ is Hurwitz, so that part (i) of Theorem 4.3.59 implies that, for any symmetric $\boldsymbol{Q}$, there is a unique symmetric $\boldsymbol{P}$ so that $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. As we saw in the proof of Theorem 4.3.59, one can determine $\boldsymbol{P}$ with the formula

$$\boldsymbol{P} = \int_0^\infty \mathrm{e}^{\boldsymbol{A}^T t} \boldsymbol{Q} \mathrm{e}^{\boldsymbol{A}t} \, \mathrm{d}t. \tag{4.33}$$

However, to do this in this example is a bit tedious since we would have to deal with the various cases of a and b to cover all the various forms taken by $\mathrm{e}^{\boldsymbol{A}t}$. For example, suppose we take

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and let $a = 2$ and $b = 2$. Then we have

$$\mathrm{e}^{\boldsymbol{t}} = \mathrm{e}^{-t} \begin{bmatrix} \cos t + \sin t & \sin t \\ -2 \sin t & \cos t - \sin t \end{bmatrix}$$

In this case one can directly apply (4.33) with some effort to get

$$\boldsymbol{P} = \begin{bmatrix} \frac{5}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{8} \end{bmatrix}.$$

If we let $a = 2$ and $b = 1$ then we compute

$$\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{-t} \begin{bmatrix} 1 + t & t \\ -t & 1 - t \end{bmatrix}.$$

Again, a direct computation using (4.33) gives

$$\boldsymbol{P} = \begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Note that our choice of $\boldsymbol{Q}$ is positive-definite and that $(\boldsymbol{A}, \boldsymbol{Q})$ is, therefore, observable. Therefore, part (iii) of Theorem 4.3.59 implies that $\boldsymbol{P}$ is positive-definite. It may be verified that the $\boldsymbol{P}$'s computed above are indeed positive-definite.

However, it is not necessary to make such hard work of this. After all, the equation

$$\boldsymbol{A}^T \boldsymbol{P} + \boldsymbol{P} \boldsymbol{A} = -\boldsymbol{Q}$$

is nothing but a linear equation for $\boldsymbol{P}$. That $\boldsymbol{A}$ is Hurwitz merely ensures a unique solution for any symmetric $\boldsymbol{Q}$. If we denote

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}$$

and continue to use

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

then we must solve the linear equations

$$\begin{bmatrix} 0 & -b \\ 1 & -a \end{bmatrix}\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} + \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

subject to $a, b > 0$. One can then determine $\boldsymbol{P}$ for general (at least nonzero) a and b to be

$$\boldsymbol{P} = \begin{bmatrix} \frac{a^2+b+b^2}{2ab} & \frac{1}{2b} \\ \frac{1}{2b} & \frac{b+1}{2ab} \end{bmatrix}.$$

In this case, we are guaranteed that this is the unique $\boldsymbol{P}$ that does the job.

2. $a \leq 0$ and $b = 0$: As we have seen, in this case there is not always a solution to the equation

$$\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} = -\boldsymbol{Q}. \tag{4.34}$$

Indeed, when $\boldsymbol{Q}$ is positive-semidefinite and $(\boldsymbol{A}, \boldsymbol{Q})$ is observable, this equation is guaranteed to *not* have a solution (see Exercise 4.3.5). This demonstrates that when $\boldsymbol{A}$ is not Hurwitz, part (i) of Theorem 4.3.59 can fail in the matter of existence.

3. $a > 0$ and $b = 0$: In this case we note that, for any $C \in \mathbb{R}$, the matrix

$$\boldsymbol{P}_0 = C\begin{bmatrix} a^2 & a \\ a & 1 \end{bmatrix}$$

satisfies $\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} = \boldsymbol{0}$. Thus, if $\boldsymbol{P}$ is any solution to (4.34), then $\boldsymbol{P} + \boldsymbol{P}_0$ is also a solution. If we take

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix},$$

then, as we saw in Theorem 4.3.39, if

$$\boldsymbol{P} = \begin{bmatrix} a^2 & a \\ a & 2 \end{bmatrix},$$

then $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$. What we have shown is that $(\boldsymbol{P} + \boldsymbol{P}_0, \boldsymbol{Q})$ is also a Lyapunov pair for $\boldsymbol{F}$. Thus part (i) of Theorem 4.3.59 can fail in the matter of uniqueness when $\boldsymbol{A}$ is not Hurwitz. •

Notes and references

[Liapunov 1893]
[Bacciotti and Rosier 2005] for Lyapunov's Second Method.
[Kellett 2014] for comparison functions.
The original reference for this work is [LaSalle 1968].
[Barbashin and Krasovskiĭ 1952]
Theorem 4.3.51 is due to Chetaev.

Exercises

4.3.1 Determine whether the following functions are or are not of class $\mathcal{K}$:

(a) $[0,\infty) \ni x \mapsto \tan^{-1}(x) \in \mathbb{R}_{\geq 0}$;

(b) $[0,b) \ni x \mapsto x^\alpha \in \mathbb{R}_{\geq 0}$ for $b \in \mathbb{R}_{>0} \cup \{\infty\}$ and $\alpha \in \mathbb{R}_{>0}$;

(c) $[0,b) \ni x \mapsto \min\{\phi_1(x), \phi_2(x)\} \in \mathbb{R}_{\geq 0}$ for $b \in \mathbb{R}_{>0} \cup \{\infty\}$ and $\phi_1, \phi_2 \colon [0,a) \to \mathbb{R}_{\geq 0}$ of class $\mathcal{K}$;

(d) $[0,\pi) \ni x \mapsto \cos(x - \frac{\pi}{2}) + 1 \in \mathbb{R}_{\geq 0}$;

(e) $[0,2\pi) \ni x \mapsto \cos(x - \frac{\pi}{2}) + 1 \in \mathbb{R}_{\geq 0}$;

(f) $[0,b) \ni x \mapsto \begin{cases} \ln(x), & x > 0, \\ 0, & x = 0 \end{cases}$ for $b \in \mathbb{R}_{>0} \cup \{\infty\}$.

4.3.2 Prove Lemma 4.3.3.

4.3.3 Determine whether the following functions are or are not of class $\mathcal{L}$:

(a) $[a,\infty) \ni y \mapsto \mathrm{e}^{-\sigma y} \in \mathbb{R}_{\geq 0}$ for $a \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$;

(b) $[a,\infty) \ni y \mapsto y^\alpha$ for $a \in \mathbb{R}$ and α 1;

(c) $(-\frac{\pi}{2}, \frac{\pi}{2}) \ni y \mapsto \tan^{-1}(y)$;

(d) $[a,\infty) \ni y \mapsto -\ln(y)$ for $a \in \mathbb{R}$.

4.3.4 Determine whether the following functions are or are not of class $\mathcal{KL}$:

(a) $[0,b) \times [a,\infty) \ni (x,y) \mapsto \phi(x)\psi(y)$, where ϕ is one of the functions from Exercise 4.3.1 and ψ is one of the functions from Exercise 4.3.3;

(b) $[0,b) \times [0,\infty) \ni (x,y) \mapsto \dfrac{x}{\alpha x y + 1}$ for $b \in \mathbb{R}_{>0} \cup \{\infty\}$ and $\alpha \in \mathbb{R}_{>0}$;

(c) $[0,b) \times [0,\infty) \ni (x,y) \mapsto \dfrac{x}{\sqrt{2x^2 y + 1}}$.

4.3.5 Let $\boldsymbol{F}$ be the system of linear ordinary differential equations in $\mathbb{R}^2$ defined by the 2×2-matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & a \end{bmatrix},$$

for $a \geq 0$. Show that if $(\boldsymbol{P}, \boldsymbol{Q})$ is a Lyapunov pair for $\boldsymbol{F}$ for which $\boldsymbol{Q}$ is positive-semidefinite, then $(\boldsymbol{A}, \boldsymbol{Q})$ is not observable.

Section 4.4

Lyapunov's First (or Indirect) Method

The First Method of Lyapunov relates the stability of an equilibrium point to the stability of the linearisation about this equilibrium point. Therefore, in this section we provide a concrete impetus for the process of linearisation developed in Section 3.1. We shall discuss separately the First Method of Lyapunov in the nonautonomous and autonomous situation, since the autonomous case is much easier.

Let us briefly recall here the process of the linearisation of an ordinary differential equation $\boldsymbol{F}$ about an equilibrium state $\boldsymbol{x}_0$. We suppose that the right-hand side $\widehat{\boldsymbol{F}}$ is differentiable with respect to $\boldsymbol{x}$. Then the linearisation is the linear ordinary differential equation $\boldsymbol{F}_{L,\boldsymbol{x}_0}$ on $\mathbb{R}^n$ whose right-hand side is

$$\begin{aligned}\widehat{\boldsymbol{F}}_{L,\boldsymbol{x}_0} \colon \mathbb{T} \times \mathbb{R}^n &\to \mathbb{R}^n \\ (t, \boldsymbol{v}) &\mapsto \boldsymbol{D}\widehat{\boldsymbol{F}}_t(\boldsymbol{x}_0) \cdot \boldsymbol{v}.\end{aligned}$$

4.4.1 The First Method for nonautonomous equations

We shall work with a system of first-order ordinary differential equations $\boldsymbol{F}$ with right-hand side

$$\widehat{\boldsymbol{F}} \colon \mathbb{T} \times U \to \mathbb{R}^n,$$

where $U \subseteq \mathbb{R}^n$ is the state space, i.e., an open subset of $\mathbb{R}^n$. We shall consider an equilibrium point $\boldsymbol{x}_0 \in U$; thus, by Proposition 3.1.5, $\widehat{\boldsymbol{F}}(t, \boldsymbol{x}_0) = \boldsymbol{0}$ for all $t \in \mathbb{T}$.

The main theorem for this setting is then the following.

4.4.1 Theorem (Uniform asymptotic stability for linearisation implies uniform exponential stability for equilibria I) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}} \colon \mathbb{T} \times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$. *Assume that* $\sup \mathbb{T} = \infty$, *that* $\widehat{\mathbf{F}}$ *is continuously differentiable, and that there exist* $\mathrm{r}, \mathrm{L}, \mathrm{M} \in \mathbb{R}_{>0}$ *such that*

$$\left| \frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x}) \right| \leq \mathrm{M}, \qquad (\mathrm{t}, \mathbf{x}) \in \mathbb{T} \times \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0),\ \mathrm{j}, \mathrm{k} \in \{1, \dots, \mathrm{n}\}, \tag{4.35}$$

and

$$\left| \frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x}_1) - \frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x}_2) \right| \leq \mathrm{L} \|\mathbf{x}_1 - \mathbf{x}_2\|, \qquad \mathrm{t} \in \mathbb{T},\ \mathbf{x}_1, \mathbf{x}_2 \in \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0),\ \mathrm{j}, \mathrm{k} \in \{1, \dots, \mathrm{n}\}. \tag{4.36}$$

Then $\mathbf{x}_0$ is uniformly exponentially stable if its linearisation is uniformly asymptotically stable.

Proof First let us deduce some consequences of $\boldsymbol{F}$ satisfying the hypotheses of the theorem statement.

1 Lemma *If* $\mathbf{F}$ *is an ordinary differential equation whose right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathsf{U}\to\mathbb{R}^{\mathrm{n}}$$

satisfies:

(i) $\widehat{\mathbf{F}}$ *is continuously differentiable;*

(ii) there exist $\mathrm{r},\mathrm{L},\mathrm{M}\in\mathbb{R}_{>0}$ *such that (4.35) and (4.36) hold.*

Then there exists $\widehat{\mathbf{G}}\colon \mathbb{T}\times\mathsf{B}(\mathrm{r},\mathbf{x}_0)\to\mathbb{R}^{\mathrm{n}}$ *and* $\mathrm{C}\in\mathbb{R}_{>0}$ *such that*

$$\widehat{\mathrm{F}}_{\mathrm{j}}(\mathrm{t},\mathbf{x})=\sum_{\mathrm{k}=1}^{\mathrm{n}}\frac{\partial\widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t},\mathbf{x}_0)(\mathrm{x}_{\mathrm{k}}-\mathrm{x}_{0,\mathrm{k}})+\widehat{\mathrm{G}}_{\mathrm{j}}(\mathrm{t},\mathbf{x}),\qquad (\mathrm{t},\mathbf{x})\in\mathbb{T}\times\mathsf{B}(\mathrm{r},\mathbf{x}_0),$$

where

$$\|\widehat{\mathbf{G}}(\mathrm{t},\mathbf{x})\|\le \mathrm{C}\|\mathbf{x}-\mathbf{x}_0\|^2,\qquad (\mathrm{t},\mathbf{x})\in\mathbb{T}\times\mathsf{B}(\mathrm{r},\mathbf{x}_0) \tag{4.37}$$

Proof By the Mean Value Theorem, ***missing stuff***, we can write

$$\widehat{F}_j(t,\boldsymbol{x})=\widehat{F}_j(t,\boldsymbol{x}_0)+\sum_{k=1}^{n}\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{y})(x_k-x_{0,k})$$

for some $\boldsymbol{y}=s\boldsymbol{x}_0+(1-s)\boldsymbol{x}$, $s\in[0,1]$. Since $\boldsymbol{x}_0$ is an equilibrium point, we rewrite this as

$$\widehat{F}_j(t,\boldsymbol{x})=\sum_{k=1}^{n}\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{x}_0)(x_k-x_{0,k})+\sum_{k=1}^{n}\left(\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{y})-\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{x}_0)\right)(x_k-x_{0,k}).$$

If we define

$$\widehat{G}_j=\sum_{k=1}^{n}\left(\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{y})-\frac{\partial\widehat{F}_j}{\partial x_k}(t,\boldsymbol{x}_0)\right)(x_k-x_{0,k}),$$

it only remains to verify the estimate (4.37) for a suitable $C\in\mathbb{R}_{>0}$. By the

Cauchy–Bunyakovsky–Schwarz inequality, we have

$$\begin{aligned}\|\widehat{G}(t,x)\| &= \left(\sum_{j=1}^n\left(\sum_{k=1}^n\left(\frac{\partial\widehat{F}_j}{\partial x_k}(t,y)-\frac{\partial\widehat{F}_j}{\partial x_k}(t,x_0)\right)(x_k-x_{0,k})\right)^2\right)^{1/2}\\ &\leq \left(\sum_{j=1}^n\left(\sum_{k=1}^n\left(\frac{\partial\widehat{F}_j}{\partial x_k}(t,y)-\frac{\partial\widehat{F}_j}{\partial x_k}(t,x_0)\right)^2\|x-x_0\|^2\right)\right)^{1/2}\\ &\leq \left(\sum_{j=1}^n L^2\|y-x_0\|^2\|x-x_0\|^2\right)^{1/2}\\ &= \sqrt{n}L(1-s)\|x-x_0\|^2 \leq \sqrt{n}L\|x-x_0\|^2,\end{aligned}$$

and the lemma follows taking $C = \sqrt{n}L$. ▼

For brevity, let us denote $A(t) = D\widehat{F}(t,x_0)$. The assumptions of the theorem ensure that A satisfies the hypotheses of Theorem 4.3.55. Thus, since the linearisation is uniformly asymptotically stable, there exists $P\colon \mathbb{T}\to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ such that (P,I_n) is a Lyapunov pair for F_{L,x_0}. We define

$$\begin{aligned}V\colon \mathbb{T}\times U &\to \mathbb{R}\\ (t,x) &\mapsto f_P(t,x-x_0).\end{aligned}$$

Let $(t_0,x)\in\mathbb{T}\times \mathsf{B}(r,x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{F}(t,\xi(t)),\quad \xi(t_0)=x.$$

Then calculate, using the lemma above,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V(t,\xi(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)(\xi(t)-x_0),\xi(t)-x_0\rangle_{\mathbb{R}^n}\\ &= \langle\dot{P}(t)(\xi(t)),\xi(t)-x_0\rangle_{\mathbb{R}^n}+\langle P(t)(\widehat{F}(t,\xi(t))),\xi(t)-x_0\rangle_{\mathbb{R}^n}\\ &\quad+\langle P(t)(\xi(t)-x_0),\widehat{F}(t,\xi(t))\rangle_{\mathbb{R}^n}\\ &= \langle(\dot{P}(t)+P(t)A(t)+A^T(t)P(t))(\xi(t)-x_0),\xi(t)-x_0\rangle_{\mathbb{R}^n}\\ &\quad+2\langle P(t)(\xi(t)-x_0),\widehat{G}(t,\xi(t))\rangle_{\mathbb{R}^n}\\ &= -\|\xi(t)-x_0\|^2+2\langle P(t)(\xi(t)-x_0),\widehat{G}(t,\xi(t))\rangle_{\mathbb{R}^n}.\end{aligned}$$

Evaluating at $t=t_0$ and using Lemma 4.3.21, this shows that

$$\mathscr{L}_F V(t_0,x) = -\|x-x_0\|^2+2\langle P(t_0)(x-x_0),\widehat{G}(t_0,x)\rangle_{\mathbb{R}^n}$$

for $(t_0,x)\in\mathbb{T}\times\mathsf{B}(r,x_0)$. By Lemma 4.3.18, let $B\in\mathbb{R}_{>0}$ be such that

$$B\|v\|^2\leq\|P(t)(v)\|^2\leq B^{-1}\|v\|^2,\qquad (t,v)\in\mathbb{T}\times\mathbb{R}^n.$$

We have

$$\begin{aligned}|\langle \boldsymbol{P}(t)(\boldsymbol{x}-\boldsymbol{x}_0), \widehat{\boldsymbol{G}}(t,\boldsymbol{x})\rangle_{\mathbb{R}^n}| &\leq \|\boldsymbol{P}(t)(\boldsymbol{x}-\boldsymbol{x}_0)\|\|\widehat{\boldsymbol{G}}(t,\boldsymbol{x})\| \\ &\leq C\sqrt{B^{-1}}\|\boldsymbol{x}-\boldsymbol{x}_0\|^3 \leq C\sqrt{B^{-1}}r\|\boldsymbol{x}-\boldsymbol{x}_0\|^2,\end{aligned}$$

where C is as in the lemma. Therefore, if we shrink r sufficiently that $1-2C\sqrt{B^{-1}}r > \frac{1}{2}$, then

$$\mathscr{L}_{\boldsymbol{F}}V(t,\boldsymbol{x}) \leq -\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2, \qquad (t,\boldsymbol{x}) \in \mathbb{T}\times \mathsf{B}(r,\boldsymbol{x}_0).$$

Since we also have

$$B\|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \leq V(t,\boldsymbol{x}) \leq B^{-1}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2, \qquad (t,\boldsymbol{x}) \in \mathbb{T}\times \mathsf{B}(r,\boldsymbol{x}_0),$$

by Theorem 4.3.55, the theorem follows from Theorem 4.3.24. ■

4.4.2 The First Method for autonomous equations

Next we turn to Lyapunov's First Method for determining the stability of equilibria for nonautonomous ordinary differential equations.

4.4.2 Theorem (Asymptotic stability for linearisation implies exponential stability for equilibria II) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} &\to \mathbb{R}^{\mathrm{n}} \\ (\mathrm{t},\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathbf{x})\end{aligned}$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}_0$. *Assume that* $\sup\mathbb{T} = \infty$, *that* $\widehat{\mathbf{F}}$ *is continuously differentiable, and that there exist* $\mathrm{r}, \mathrm{L} \in \mathbb{R}_{>0}$ *such that*

$$\left|\frac{\partial \widehat{\mathrm{F}}_{0,\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathbf{x}_1) - \frac{\partial \widehat{\mathrm{F}}_{0,\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathbf{x}_2)\right| \leq \mathrm{L}\|\mathbf{x}_1-\mathbf{x}_2\|, \qquad \mathbf{x}_1,\mathbf{x}_2 \in \overline{\mathsf{B}}(\mathrm{r},\mathbf{x}_0),\ \mathrm{j},\mathrm{k} \in \{1,\dots,\mathrm{n}\}. \tag{4.38}$$

Then $\mathbf{x}_0$ *is exponentially stable if its linearisation is asymptotically stable.*

We offer two proofs of this theorem, one assuming Theorem 4.4.1 and the other an independent proof.

Proof of Theorem 4.4.2, assuming Theorem 4.4.1 The hypotheses of Theorem 4.4.2 clearly imply those of Theorem 4.4.1 since, in Theorem 4.4.2, $\widehat{F}$ is independent of t. Therefore, the hypotheses of Theorem 4.4.2 imply the conclusions of Theorem 4.4.1, i.e., that uniform asymptotic stability of the linearisation implies uniform exponential stability of the equilibrium. The proof in this case is concluded by recalling from Proposition 4.1.5 that the various flavours of uniform stability are equivalent to the corresponding flavours of stability for autonomous equations. ■

Independent proof of Theorem 4.4.2 First let us deduce some consequences of $\boldsymbol{F}$ satisfying the hypotheses of the theorem statement.

1 Lemma *If* $\mathbf{F}$ *is an autonomous ordinary differential equation whose right-hand side*

$$\begin{aligned}\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} &\to \mathbb{R}^{\mathrm{n}}\\ (\mathrm{t},\mathbf{x}) &\mapsto \widehat{\mathbf{F}}_0(\mathbf{x})\end{aligned}$$

satisfies:

(i) $\widehat{\mathbf{F}}_0$ *is continuously differentiable;*

(ii) there exist $\mathrm{r},\mathrm{L}\in\mathbb{R}_{>0}$ *such that (4.38) holds.*

Then there exists $\widehat{\mathbf{G}}_0\colon \mathsf{B}(\mathrm{r},\mathbf{x}_0)\to\mathbb{R}^{\mathrm{n}}$ *and* $\mathrm{C}\in\mathbb{R}_{>0}$ *such that*

$$\widehat{\mathrm{F}}_{0,\mathrm{j}}(\mathrm{t},\mathbf{x}) = \sum_{\mathrm{k}=1}^{\mathrm{n}} \frac{\partial \widehat{\mathrm{F}}_{0,\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathbf{x}_0)(\mathrm{x}_{\mathrm{k}}-\mathrm{x}_{0,\mathrm{k}}) + \widehat{\mathrm{G}}_{0,\mathrm{j}}(\mathbf{x}), \qquad \mathbf{x}\in\mathsf{B}(\mathrm{r},\mathbf{x}_0),$$

where

$$\|\widehat{\mathbf{G}}_0(\mathbf{x})\| \le \mathrm{C}\|\mathbf{x}-\mathbf{x}_0\|^2, \qquad \mathbf{x}\in\mathsf{B}(\mathrm{r},\mathbf{x}_0) \tag{4.39}$$

Proof By the Mean Value Theorem, ***missing stuff***, we can write

$$\widehat{F}_{0,j}(\boldsymbol{x}) = \widehat{F}_{0,j}(\boldsymbol{x}_0) + \sum_{k=1}^{n} \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{y})(x_k - x_{0,k})$$

for some $\boldsymbol{y} = s\boldsymbol{x}_0 + (1-s)\boldsymbol{x}$, $s\in[0,1]$. Since $\boldsymbol{x}_0$ is an equilibrium point, we rewrite this as

$$\widehat{F}_{0,j}(\boldsymbol{x}) = \sum_{k=1}^{n} \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{x}_0)(x_k - x_{0,k}) + \sum_{k=1}^{n}\left(\frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{y}) - \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{x}_0)\right)(x_k - x_{0,k}).$$

If we define

$$\widehat{G}_{0,j} = \sum_{k=1}^{n}\left(\frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{y}) - \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{x}_0)\right)(x_k - x_{0,k}),$$

it only remains to verify the estimate (4.39) for a suitable $C\in\mathbb{R}_{>0}$. By the Cauchy–Bunyakovsky–Schwarz inequality, we have

$$\begin{aligned}\|\widehat{G}_0(\boldsymbol{x})\| &= \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{y}) - \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{x}_0)\right)(x_k - x_{0,k})\right)^2\right)^{1/2}\\ &\le \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{y}) - \frac{\partial \widehat{F}_{0,j}}{\partial x_k}(\boldsymbol{x}_0)\right)^2 \|\boldsymbol{x}-\boldsymbol{x}_0\|^2\right)\right)^{1/2}\\ &\le \left(\sum_{j=1}^{n} L^2\|\boldsymbol{y}-\boldsymbol{x}_0\|^2\|\boldsymbol{x}-\boldsymbol{x}_0\|^2\right)^{1/2}\\ &= \sqrt{n}L(1-s)\|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \le \sqrt{n}L\|\boldsymbol{x}-\boldsymbol{x}_0\|^2,\end{aligned}$$

and the lemma follows taking $C = \sqrt{n}L$. ▼

For brevity, let us denote $\boldsymbol{A} = \boldsymbol{D}\widehat{\boldsymbol{F}}(x_0)$. Since the linearisation is asymptotically stable, by Theorem 4.3.56 there exists $\boldsymbol{P} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ such that $(\boldsymbol{P}, \boldsymbol{I}_n)$ is a Lyapunov pair for $\boldsymbol{F}_{L,x_0}$. We define

$$V \colon U \to \mathbb{R}, \quad x \mapsto f_{\boldsymbol{P}}(x - x_0).$$

Let $x \in \mathsf{B}(r, x_0)$ and let ξ be the solution to the initial value problem

$$\dot{\xi}(t) = \widehat{\boldsymbol{F}}_0(\xi(t)), \quad \xi(0) = x.$$

Then calculate, using the lemma above,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} V(\xi(t)) &= \frac{\mathrm{d}}{\mathrm{d}t} \langle \boldsymbol{P}(\xi(t) - x_0), \xi(t) - x_0 \rangle_{\mathbb{R}^n} \\
&= \langle \boldsymbol{P}(\widehat{\boldsymbol{F}}_0(\xi(t))), \xi(t) - x_0 \rangle_{\mathbb{R}^n} + \langle \boldsymbol{P}(\xi(t) - x_0), \widehat{\boldsymbol{F}}_0(\xi(t)) \rangle_{\mathbb{R}^n} \\
&= \langle (\boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P})(\xi(t) - x_0), \xi(t) - x_0 \rangle_{\mathbb{R}^n} \\
&\quad + 2\langle \boldsymbol{P}(\xi(t) - x_0), \widehat{\boldsymbol{G}}_0(\xi(t)) \rangle_{\mathbb{R}^n} \\
&= -\|\xi(t) - x_0\|^2 + 2\langle \boldsymbol{P}(\xi(t) - x_0), \widehat{\boldsymbol{G}}_0(\xi(t)) \rangle_{\mathbb{R}^n}.
\end{aligned}$$

Evaluating at $t = 0$ and using Lemma 4.3.21, this shows that

$$\mathscr{L}_{\boldsymbol{F}} V(x) = -\|x - x_0\|^2 + 2\langle \boldsymbol{P}(x - x_0), \widehat{\boldsymbol{G}}_0(x) \rangle_{\mathbb{R}^n}$$

for $x \in \mathsf{B}(r, x_0)$. By Lemma 4.3.14, let $B \in \mathbb{R}_{>0}$ be such that

$$B\|v\|^2 \leq \|\boldsymbol{P}(v)\|^2 \leq B^{-1}\|v\|^2, \qquad v \in \mathbb{R}^n.$$

We have

$$\begin{aligned}
|\langle \boldsymbol{P}(x - x_0), \widehat{\boldsymbol{G}}_0(x) \rangle_{\mathbb{R}^n}| &\leq \|\boldsymbol{P}(x - x_0)\| \|\widehat{\boldsymbol{G}}_0(x)\| \\
&\leq C\sqrt{B^{-1}}\|x - x_0\|^3 \leq C\sqrt{B^{-1}} r \|x - x_0\|^2,
\end{aligned}$$

where C is as in the lemma. Therefore, if we shrink r sufficiently that $1 - 2C\sqrt{B^{-1}}r > \frac{1}{2}$, then

$$\mathscr{L}_{\boldsymbol{F}} V(x) \leq -\frac{1}{2}\|x - x_0\|^2, \qquad x \in \mathsf{B}(r, x_0).$$

Since we also have

$$B\|x - x_0\|^2 \leq V(x) \leq B^{-1}\|x - x_0\|^2, \qquad x \in \mathsf{B}(r, x_0),$$

by Theorem 4.3.56, the theorem follows from Theorem 4.3.30. ■

4.4.3 An instability theorem

In this section we give a result that allows one to determine *instability* of an equilibrium from the linearisation. We shall work here only with autonomous ordinary differential equations.

4.4.3 Theorem (Spectral instability of linearisation implies instability for equilibria) *Let* $\mathbf{F}$ *be an autonomous ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$
$$(\mathrm{t}, \mathbf{x}) \mapsto \widehat{\mathbf{F}}_0(\mathbf{x})$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$*, that* $\widehat{\mathbf{F}}_0$ *is continuously differentiable. Then* $\mathbf{x}_0$ *is unstable if* $\operatorname{spec}(\widehat{\mathbf{F}}_{\mathrm{L},\mathbf{x}_0}) \cap \mathbb{C}_+ \neq \emptyset$*.*

Proof For brevity, let us denote $A = \widehat{F}_{L,x_0}$. First let us suppose that $\operatorname{spec}(A) \cap \mathrm{i}\mathbb{R} = \emptyset$. Then, according to Remark 3.2.36–5 ■

4.4.4 A converse theorem

In this section we consider the extent to which stability of the linearisation exactly characterises stability of an equilibrium point. As we know from the results and examples above, it is definitely *not* the case that stability of an equilibrium point necessitates stability of the linearisation. The following result shows that this necessity holds when the type of stability we are discussing is exponential stability.

4.4.4 Theorem (Exponential stability of an equilibrium implies exponential stability of linearisation) *Let* $\mathbf{F}$ *be an ordinary differential equation with right-hand side*

$$\widehat{\mathbf{F}}\colon \mathbb{T}\times \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$$

and let $\mathbf{x}_0 \in \mathrm{U}$ *be an equilibrium point for* $\mathbf{F}$*. Assume that* $\sup \mathbb{T} = \infty$*, that* $\widehat{\mathbf{F}}$ *is continuously differentiable, and that there exist* $\mathrm{r}, \mathrm{L}, \mathrm{M} \in \mathbb{R}_{>0}$ *such that*

$$\left|\frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x})\right| \leq \mathrm{M}, \qquad (\mathrm{t}, \mathbf{x}) \in \mathbb{T}\times \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0),\ \mathrm{j}, \mathrm{k} \in \{1, \dots, \mathrm{n}\}, \tag{4.40}$$

and

$$\left|\frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x}_1) - \frac{\partial \widehat{\mathrm{F}}_{\mathrm{j}}}{\partial \mathrm{x}_{\mathrm{k}}}(\mathrm{t}, \mathbf{x}_2)\right| \leq \mathrm{L}\|\mathbf{x}_1 - \mathbf{x}_2\|, \qquad \mathrm{t} \in \mathbb{T},\ \mathbf{x}_1, \mathbf{x}_2 \in \overline{\mathsf{B}}(\mathrm{r}, \mathbf{x}_0),\ \mathrm{j}, \mathrm{k} \in \{1, \dots, \mathrm{n}\}. \tag{4.41}$$

Then $\widehat{\mathbf{F}}_{\mathrm{L},\mathbf{x}_0}$ *is globally exponentially stable if* $\mathbf{x}_0$ *is exponentially stable.*

Proof Let us abbreviate $A(t) = \widehat{F}_{L,x_0}(t)$. Let us write

$$A(t)x = \widehat{F}(t,x) - \underbrace{(\widehat{F}(t,x) - A(t)x)}_{\widehat{G}(t,x)}.$$

According to Lemma 1 from the proof of Theorem 4.4.1, there exists $C, r \in \mathbb{R}_{>0}$ such that

$$\|\widehat{G}(t,x)\| \leq C\|x - x_0\|^2, \qquad (t,x) \in \mathbb{T}\times \mathsf{B}(r, x_0).$$

Since ■

Chapter 5

Transform methods for differential equations

In this section we give a very brief and not very rigorous introduction to "transform methods" for differential equations. We shall see a number of different transforms and use them in a number of different contexts. The basic idea, in all cases we shall consider, is that one applies a "transform" to a differential equation to convert it, perhaps only partially, from a differential equation into an algebraic equation. In all cases, this is a consequence of the transform converting the derivative of a function with respect to some independent variable into an algebraic expression in a new independent variable, which one might call the "transform variable." This algebraic equation, one hopes, is easier to solve than the original differential equation. What one has then is a solution of the equation in the transform variable. Then one applies an "inverse transform" to retrieve the solution in the original independent variables. It is this last step that is typically the sticking point in terms of obtaining a solution in closed form. However, even if one cannot obtain a closed-form solution in the manner one might like, often the use of transform methods is useful for arriving at forms for solutions, or for proving existence of solutions, in cases where "direct" methods may not be as useful.

Let us outline what we present in this chapter. We shall discuss three transforms in this text: the Laplace transform; the continuous-discrete Fourier transform; and the continuous-continuous Fourier transform. These transforms are presented in Section 5.1. We focus, in all cases, on giving the essential features of the transforms that we shall use in treating differential equations. This primarily means two things: (1) considering how the transforms interact with derivatives; (2) considering how transforms interact with convolution. The former will seem obvious, since we are dealing with differential equations. We shall see subsequently how convolution arises when transform methods are used. After we present the transforms and their properties, we see how they might be used to study differential equations. We start in Section 5.2 by looking at Laplace transform methods for differential equations. Among the methods we discuss are venerable methods for ordinary differential equations, and are a rich source of tedious computational exercises for students. We try to sidestep this facet of the techniques, focussing instead on some principles that underlie transform methods in general. In Section 5.3 we consider

Fourier transform methods for differential equations. While Laplace transform methods are "standard" for ordinary differential equations, for partial differential equations, the matter of what transform to use is often not as cut and dried as it is for ordinary differential equations. We shall, therefore, see that one should treat each problem separately and be open to what it requires.

As a closing comment, we note that transform methods are primarily useful for linear differential equations with constant coefficients, be they ordinary differential equations or partial differential equations. As we hope we have made clear in our presentation of ordinary differential equations thus far, while linear equations are important, they do not comprise anything like the entirety of all differential equations. As a consequence of this, the transform methods we describe here, while important, are about as broadly applicable as the methods we have seen thus far for solving linear ordinary differential equations with constant coefficients.

Contents

Section 5.1

The Fourier and Laplace transforms

In this section we shall introduce three transforms that we shall subsequently use to study differential equations. The presentation of these transforms is made with no discussion of differential equations themselves, in order to emphasise that these transforms have life outside their being applicable to differential equations. Unfortunately, our presentation is also very brief and not completely rigorous on all points. A complete and rigorous presentation can be found in many places, particularly as concerns the Fourier transforms we discuss.***missing stuff***

5.1.1 The continuous-discrete Fourier transform

The first transform we present applies to functions that are defined on a closed and bounded interval, which we assume to be $[0, L]$ for concreteness. The natural domain for the transform we consider is the set

$$\mathrm{L}^1([0, L]; \mathbb{C}) = \left\{ f \colon [0, L] \to \mathbb{C} \;\middle|\; \int_0^L |f(x)| \, \mathrm{d}x < \infty \right\}$$

of functions whose modulus is integrable.

5.1.1.1 The transform Let us give the definition. Note that the definition we give is for $\mathbb{C}$-valued functions, as this is most natural. We shall subsequently see how this specialises for $\mathbb{R}$-valued functions.

5.1.1 Definition (Continuous-discrete Fourier transform I) The ***continuous-discrete Fourier transform*** (***CDFT***) of $f \in \mathrm{L}^1([0, L]; \mathbb{C})$ is the function $\mathscr{F}_{\mathrm{CD}}(f) \colon \mathbb{Z} \to \mathbb{C}$ defined by

$$\mathscr{F}_{\mathrm{CD}}(f)(n) = \int_0^L f(x) \mathrm{e}^{-2\pi \mathrm{i} n \frac{x}{L}} \, \mathrm{d}x. \qquad \bullet$$

One can see why this is called the "continuous-discrete" Fourier transform: it takes a function of the continuous variable $x \in [0, L]$ and returns a function of the discrete variable $n \in \mathbb{Z}$. Before we discuss specific features of the transform, let us illustrate that this is something that can, in principle, be computed.

5.1.2 Examples (Continuous-discrete Fourier transform)

1. Let us take $f(x) = \cos(2\pi m \frac{x}{L})$ for $m \in \mathbb{Z}_{\geq 0}$. We then have, for $n \in \mathbb{Z}$,

$$\begin{aligned} \mathscr{F}_{\mathrm{CD}}(f)(n) &= \int_0^L \cos(2\pi m \tfrac{x}{L}) \mathrm{e}^{-2\pi \mathrm{i} n \frac{x}{L}} \, \mathrm{d}x \\ &= \int_0^L \cos(2\pi m \tfrac{x}{L}) \cos(2\pi n \tfrac{x}{L}) \, \mathrm{d}x - \mathrm{i} \int_0^L \cos(2\pi m \tfrac{x}{L}) \sin(2\pi n \tfrac{x}{L}) \, \mathrm{d}x. \end{aligned}$$

Note that

$$\begin{aligned}\cos(\alpha)\cos(\beta) &= \tfrac{1}{2}(\cos(\alpha-\beta)+\cos(\alpha+\beta))\\ \cos(\alpha)\sin(\beta) &= \tfrac{1}{2}(\sin(\alpha-\beta)+\sin(\alpha+\beta)),\end{aligned} \tag{5.1}$$

using some trigonometric identities you can look up. Now, since

$$\int_0^L \sin(2\pi k\tfrac{x}{L})\,\mathrm{d}x = \int_0^L \cos(2\pi k\tfrac{x}{L})\,\mathrm{d}x = 0, \qquad k \in \mathbb{Z}_{>0},$$

and

$$\int_0^L \cos(2\pi 0\tfrac{x}{L})\,\mathrm{d}x = L,$$

we have

$$\mathscr{F}_{\mathrm{CD}}(f)(n) = \begin{cases}\frac{L}{2}, & n = m,\\ 0, & n \neq m.\end{cases}$$

2. Here we take $f(x) = \sin(2\pi m\frac{x}{L})$ for $m \in \mathbb{Z}_{>0}$, and compute

$$\begin{aligned}\mathscr{F}_{\mathrm{CD}}(f)(n) &= \int_0^L \sin(2\pi m\tfrac{x}{L})\mathrm{e}^{-2\pi \mathrm{i} n\frac{x}{L}}\,\mathrm{d}x\\ &= \int_0^L \sin(2\pi m\tfrac{x}{L})\cos(2\pi n\tfrac{x}{L})\,\mathrm{d}x - \mathrm{i}\int_0^L \sin(2\pi m\tfrac{x}{L})\sin(2\pi n\tfrac{x}{L})\,\mathrm{d}x.\end{aligned}$$

We now use the trigonometric identities

$$\begin{aligned}\sin(2\pi m\tfrac{x}{L})\cos(2\pi n\tfrac{x}{L}) &= \tfrac{1}{2}(\sin(2\pi(m-n)\tfrac{x}{L})+\sin(2\pi(m+n)\tfrac{x}{L}))\\ \sin(2\pi m\tfrac{x}{L})\sin(2\pi n\tfrac{x}{L}) &= \tfrac{1}{2}(\cos(2\pi(m-n)\tfrac{x}{L})-\cos(2\pi(m+n)\tfrac{x}{L})).\end{aligned}$$

As in the preceding example, this then gives

$$\mathscr{F}_{\mathrm{CD}}(f)(n) = \begin{cases}-\mathrm{i}\frac{L}{2}, & n = m,\\ 0, & n \neq m.\end{cases}$$

3. Next we consider the function

$$f(t) = \begin{cases}1, & x \in [0, \frac{L}{2}],\\ -1, & x \in (\frac{L}{2}, L].\end{cases}$$

We have

$$\mathscr{F}_{\mathrm{CD}}(f)(0) = \int_0^L f(x)\,\mathrm{d}x = 0$$

and, for $n \neq 0$,

$$\begin{aligned}\mathscr{F}_{\mathrm{CD}}(f)(n) &= \int_0^L f(x)\mathrm{e}^{-2\pi n\frac{x}{L}}\,\mathrm{d}x\\ &= \int_0^{L/2} \mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\,\mathrm{d}x - \int_{L/2}^{L} \mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\,\mathrm{d}x\\ &= -\frac{L\mathrm{e}^{2\pi\mathrm{i}n\frac{x}{L}}}{2\pi\mathrm{i}n}\Big|_0^{L/2} + \frac{L\mathrm{e}^{2\pi\mathrm{i}n\frac{x}{L}}}{2\pi\mathrm{i}n}\Big|_{L/2}^{L}\\ &= \mathrm{i}\frac{L}{2\pi n}\left(\mathrm{e}^{-\pi\mathrm{i}n}-1\right) - \mathrm{i}\frac{L}{2\pi n}\left(1-\mathrm{e}^{-\pi\mathrm{i}n}\right).\end{aligned}$$

Now note that

$$\mathrm{e}^{-\pi\mathrm{i}n} = \cos(n\pi) - \mathrm{i}\sin(n\pi) = (-1)^n.$$

Thus

$$\mathscr{F}_{\mathrm{CD}}(f)(n) = \begin{cases} 0, & n = 0,\\ \mathrm{i}\frac{L}{n\pi}((-1)^n - 1), & n \neq 0.\end{cases}$$

•

Thus we can see that the CDFT is something that we might be able to calculate. However, this does not explain our interest in the CDFT. Indeed, our interest in the CDFT is a consequence of a few of its properties that we now enumerate.

First we demonstrate the linearity of the CDFT.

5.1.3 Proposition (Linearity of the CDFT) *The CDFT is a linear map:*

$$\mathscr{F}_{\mathrm{CD}}(\mathrm{f}_1 + \mathrm{f}_2)(\mathrm{n}) = \mathscr{F}_{\mathrm{CD}}(\mathrm{f}_1)(\mathrm{n}) + \mathscr{F}_{\mathrm{CD}}(\mathrm{f}_2)(\mathrm{n}), \quad \mathscr{F}_{\mathrm{CD}}(\mathrm{af})(\mathrm{n}) = \mathrm{a}\mathscr{F}_{\mathrm{CD}}(\mathrm{f})(\mathrm{n})$$

for every $\mathrm{f}, \mathrm{f}_1, \mathrm{f}_2 \in \mathrm{L}^1([0,\mathrm{L}];\mathbb{C})$ *and* $\mathrm{a} \in \mathbb{C}$.

Proof This follows by linearity of the integral. ■

Next we consider how the CDFT interacts with differentiation, since this will be an important part of how we use this transform with differential equations.

5.1.4 Proposition (CDFT and differentiation) *Let* $\mathrm{f}\colon [0,\mathrm{L}] \to \mathbb{C}$ *be continuously differentiable and suppose that* $\mathrm{f}(0) = \mathrm{f}(\mathrm{L})$. *Then*

$$\mathscr{F}_{\mathrm{CD}}(\tfrac{\mathrm{df}}{\mathrm{dx}})(\mathrm{n}) = \frac{2\pi\mathrm{in}}{\mathrm{L}}\mathscr{F}_{\mathrm{CD}}(\mathrm{f})(\mathrm{n}), \qquad \mathrm{n} \in \mathbb{Z}.$$

Proof Using integration by parts we compute

$$\begin{aligned}\mathscr{F}_{\mathrm{CD}}(\tfrac{\mathrm{d}f}{\mathrm{d}x})(n) &= \int_0^L \tfrac{\mathrm{d}f}{\mathrm{d}x}(x)\mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\,\mathrm{d}x\\ &= f(x)\mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\Big|_0^L + \frac{2\pi\mathrm{i}n}{L}\int_0^L f(x)\mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\,\mathrm{d}x\\ &= \frac{2\pi\mathrm{i}n}{L}\mathscr{F}_{\mathrm{CD}}(f)(n),\end{aligned}$$

as required. ■

Of course, the proposition can be applied recursively for higher-order derivatives.

5.1.5 Corollary (CDFT and higher-order differentiation) *Let* $\mathrm{f}\colon [0,\mathrm{L}] \to \mathbb{C}$ *be* k*-times continuously differentiable and suppose that* $\frac{\mathrm{d}^{\mathrm{j}}\mathrm{f}}{\mathrm{dx}^{\mathrm{j}}}(0) = \frac{\mathrm{d}^{\mathrm{j}}\mathrm{f}}{\mathrm{dx}^{\mathrm{j}}}(\mathrm{L})$, $\mathrm{j} \in \{0, 1, \ldots, \mathrm{k}-1\}$. *Then*

$$\mathscr{F}_{\mathrm{CD}}(\tfrac{\mathrm{d}^{\mathrm{k}}\mathrm{f}}{\mathrm{dx}^{\mathrm{k}}})(\mathrm{n}) = \left(\frac{2\pi\mathrm{in}}{\mathrm{L}}\right)^{\mathrm{k}} \mathscr{F}_{\mathrm{CD}}(\mathrm{f})(\mathrm{n}), \qquad \mathrm{n} \in \mathbb{Z}.$$

5.1.1.2 The inverse transform A crucial ingredient to the transform approach to differential equations is the inversion step, wherein one goes from the transform domain back to the original domain for the equation. A complete theory of inversion for the transforms we consider is difficult in any generality. Thus we shall present the theory in only a superficial (and not entirely correct) way. However, what is true is that, after a full development of the theory, the main ideas we present are correct in their essence, and, under suitable hypotheses, correct.

We motivate our constructions with an heuristic discussion. Suppose that one wishes to represent a function $f \in \mathrm{L}^1([0,L];\mathbb{R})$ as a linear combination of sine's and cosine's:

$$f(x) = \frac{a_0(f)}{2} + \sum_{n=1}^{\infty} (a_n(f)\cos(2\pi n\tfrac{x}{L}) + b_n(f)\sin(2\pi n\tfrac{x}{L})), \tag{5.2}$$

for real coefficients $a_n(f)$, $n \in \mathbb{Z}_{\geq 0}$, and $b_n(f)$, $n \in \mathbb{Z}_{>0}$. There seems to be no really good reason to expect such a representation to be meaningful. However, this was the hypothesis of Fourier in his efforts to understand heat flow, and it was an hypothesis that was, in precise ways, validated by various mathematicians over the years. We shall not do much to justify this hypothesis, and simply accept it as valid. Now, having done so, let us convert from sine's and cosine's to complex exponential functions by virtue of Euler's formula: $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$. That is, let us suppose that f is $\mathbb{C}$-valued and instead seek to write

$$f(x) = \sum_{n\in\mathbb{Z}} c_n(f)\mathrm{e}^{2\pi n\frac{x}{L}},$$

for complex coefficients $c_n(f)$, $n \in \mathbb{Z}$. Let us determine the coefficients $c_n(f)$ by performing the following calculation, done for $m \in \mathbb{Z}$:

$$\begin{aligned}
& f(x) = \lim_{N\to\infty} \sum_{n=-N}^{N} c_n(f)\mathrm{e}^{2\pi\mathrm{i}n\frac{x}{L}} \\
\implies & \int_0^L f(x)\mathrm{e}^{-2\pi\mathrm{i}m\frac{x}{L}}\,\mathrm{d}t = \lim_{N\to\infty} \int_0^L \left(\sum_{n=-N}^{N} c_n(f)\mathrm{e}^{2\pi\mathrm{i}n\frac{x}{L}}\right)\mathrm{e}^{-2\pi\mathrm{i}m\frac{x}{L}}\,\mathrm{d}x \\
\implies & \int_0^L f(x)\mathrm{e}^{-2\pi\mathrm{i}m\frac{x}{L}}\,\mathrm{d}t = \lim_{N\to\infty} \sum_{n=-N}^{N} c_n(f) \int_0^L \mathrm{e}^{2\pi\mathrm{i}m\frac{x}{L}}\mathrm{e}^{-2\pi\mathrm{i}n\frac{x}{L}}\,\mathrm{d}x.
\end{aligned}$$

Now, for $k \neq 0$, we have

$$\int_0^L \mathrm{e}^{2\pi \mathrm{i} k \frac{x}{L}} \, \mathrm{d}x = \frac{L}{2\pi \mathrm{i} k} \mathrm{e}^{2\pi \mathrm{i} k \frac{x}{L}} \Big|_0^L = \frac{L}{2\pi \mathrm{i} k}(1-1) = 0.$$

For $k = 0$ we have

$$\int_0^L \mathrm{e}^{2\pi \mathrm{i} 0 \frac{x}{L}} \, \mathrm{d}x = L.$$

Therefore,

$$\lim_{N \to \infty} \sum_{n=-N}^{N} c_n(f) \int_0^L \mathrm{e}^{2\pi \mathrm{i}(m-n)\frac{x}{L}} \, \mathrm{d}x = L c_m(f)$$

and so

$$\int_0^L f(x) \mathrm{e}^{-2\pi \mathrm{i} m \frac{x}{L}} \, \mathrm{d}t = L c_m(f).$$

Thus, finally,

$$c_m(f) = \frac{1}{L} \int_0^L f(x) \mathrm{e}^{-2\pi \mathrm{i} m \frac{x}{L}} \, \mathrm{d}t = \frac{\mathscr{F}_{\mathrm{CD}}(f)(m)}{L}.$$

Therefore, we have the formula

$$f(x) = \lim_{N \to \infty} \frac{1}{L} \sum_{n=-N}^{N} \mathscr{F}_{\mathrm{CD}}(f)(n) \mathrm{e}^{2\pi \mathrm{i} n \frac{x}{L}}.$$

This formula, and the derivation we give of it, is not valid. The derivation, for example, involves swapping an infinite sum and na integral, a swapping which is generally not valid. Nonetheless, the infinite sum on the right-hand side of this formula,

$$\frac{1}{L} \sum_{n \in \mathbb{Z}} \mathscr{F}_{\mathrm{CD}}(f)(n) \mathrm{e}^{2\pi \mathrm{i} n \frac{x}{L}},$$

is an interesting thing and is called the ***Fourier series*** for f.

5.1.6 Remarks (Convergence of Fourier series) Let us make a few comments on the convergence of the Fourier series, i.e., the existence of the limit

$$\lim_{N \to \infty} \frac{1}{L} \sum_{n=-N}^{N} \mathscr{F}_{\mathrm{CD}}(f)(n) \mathrm{e}^{2\pi \mathrm{i} n \frac{x}{L}}.$$

1. If $f \in \mathrm{L}^1([0, L]; \mathbb{C})$, then generally the series does not converge. Indeed, it can be shown that there exists $f \in \mathrm{L}^1([0, L]; \mathbb{C})$ such that the Fourier series diverges for each $x \in [0, L]$.
2. Even if f is continuous, the Fourier series will not generally converge, although it will converge at "most" $x \in [0, L]$.

3. If f is continuously differentiable and $f(0) = f(L)$, then the Fourier series for f converges (indeed uniformly) to f. •

The above computations, regardless of when they are precisely valid, do nonetheless give the following inverse for the CDFT.

5.1.7 "Definition" (Inverse of the CDFT) The *inverse* of the CDFT is the mapping that assigns to a map $F\colon \mathbb{Z} \to \mathbb{C}$ the function $\mathcal{F}_{\text{CD}}^{-1}(F)\colon [0, L] \to \mathbb{C}$ defined by

$$\mathcal{F}_{\text{CD}}^{-1}(F)(x) = \frac{1}{L} \sum_{n \in \mathbb{Z}} F(n) \mathrm{e}^{2\pi \mathrm{i} n \frac{x}{L}}.$$ •

Note that the "definition" is really quite senseless since the sum defining $\mathcal{F}_{\text{CD}}^{-1}(F)$ will definitely not converge, in general. However, what *is* true is that, under suitable hypotheses, $\mathcal{F}_{\text{CD}}^{-1} \circ \mathcal{F}_{\text{CD}}(f) = f$. That is to say, under suitable hypotheses, one can use the Fourier series to recover f from $\mathcal{F}_{\text{CD}}(f)$; it is in this sense that we mean that $\mathcal{F}_{\text{CD}}^{-1}$ is an inverse for $\mathcal{F}_{\text{CD}}$.

5.1.1.3 Convolution and the continuous-discrete Fourier transform As we have mentioned in our discussions of the philosophy of transform methods, one of the consequences of their use is that a differential equation is converted, possibly only partially, into an algebraic equation. As a consequence of solving the resulting algebraic equations, one often ends up needing to interpret the product of transformed functions. The question that arises is: What operation in the original variables corresponds to multiplication of functions in the transformed variables? The answer, as we shall see in three different contexts is: convolution.

We consider this first in the case of the CDFT. The transformed variables in this case reside in $\mathbb{Z}$, and so transformed functions are functions from $\mathbb{Z}$ to $\mathbb{C}$. Thus, in this case, we have $F, G\colon \mathbb{Z} \to \mathbb{C}$ and so the product of F and G is the function

$$\begin{aligned} FG\colon \mathbb{Z} &\to \mathbb{C} \\ n &\mapsto F(n)G(n). \end{aligned}$$

What we want to know is, if $F = \mathcal{F}_{\text{CD}}(f)$ and $G = \mathcal{F}_{\text{CD}}(g)$, is there a function $h\colon [0, L] \to \mathbb{C}$ for which $\mathcal{F}_{\text{CD}}(h) = FG$?

To answer this question, we make the following definition.

5.1.8 Definition (Periodic convolution) If $f, g \in \mathrm{L}^1([0, L]; \mathbb{C})$, the ***periodic convolution*** of f and g is the function

$$\begin{aligned} f * g\colon [0, L] &\to \mathbb{C} \\ x &\mapsto \int_0^L f(x - y) g(y) \, \mathrm{d}y. \end{aligned}$$ •

The operation of convolution is a interesting one, and has many properties that merit further exploration. For our purposes, we merely point out the following result.

5.1.9 Proposition (CDFT and convolution) *If* $\mathrm{f}, \mathrm{g} \in \mathrm{L}^1([0, \mathrm{L}]; \mathbb{C})$, *then*

$$\mathscr{F}_{\mathrm{CD}}(\mathrm{f} * \mathrm{g}) = \mathscr{F}_{\mathrm{CD}}(\mathrm{f})\mathscr{F}_{\mathrm{CD}}(\mathrm{g}).$$

Proof Let us extend f and g to be defined on $\mathbb{R}$ by requiring that they have period L. This is a fairly straightforward application of Fubini's Theorem, the change of variables theorem, and periodicity of f:

$$\begin{aligned}
\mathscr{F}_{\mathrm{CD}}(f * g)(n) &= \int_0^L f * g(x) \mathrm{e}^{-2\pi \mathrm{i} n \frac{x}{L}} \, \mathrm{d}x = \int_0^L \left(\int_0^L f(x-y) g(y) \, \mathrm{d}y \right) \mathrm{e}^{-2\pi \mathrm{i} n \frac{x}{L}} \, \mathrm{d}x \\
&= \int_0^L g(y) \left(\int_0^L f(x-y) \mathrm{e}^{-2\pi \mathrm{i} n \frac{x}{L}} \, \mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^L g(\sigma) \left(\int_{-\sigma}^{L-\sigma} f(\tau) \mathrm{e}^{-2\pi \mathrm{i} n \frac{\sigma+\tau}{L}} \, \mathrm{d}\tau \right) \mathrm{d}\sigma \\
&= \left(\int_0^L g(\sigma) \mathrm{e}^{-2\pi \mathrm{i} n \frac{\sigma}{L}} \, \mathrm{d}\sigma \right) \left(\int_0^L f(\tau) \mathrm{e}^{-2\pi \mathrm{i} n \frac{\tau}{L}} \, \mathrm{d}\tau \right) \\
&= \mathscr{F}_{\mathrm{CD}}(f)(n) \mathscr{F}_{\mathrm{CD}}(g)(n),
\end{aligned}$$

as claimed. ■

5.1.1.4 Extension to higher-dimensions In this section we briefly consider two things. First we consider the extension of the CDFT to functions whose domain is not $\mathbb{C}$, but $\mathbb{C}^n$. This is quite straightforward, since one simply applies the existing constructions component-wise. The second thing we do is extend the definition of the CDFT to functions whose domain has more than one variable. In this case the extension is still not that difficult, but does requite a tiny bit of thinking.

First we extend the CDFT to functions with values in $\mathbb{C}^n$. To do so, we note that, if $\boldsymbol{f} \colon [0, L] \to \mathbb{C}^n$, then we can write

$$\boldsymbol{f}(x) = (f_1(x), \dots, f_n(x))$$

for functions $f_1, \dots, f_n \colon [0, L] \to \mathbb{C}$. We then denote

$$\mathrm{L}^1([0, L]; \mathbb{C}^n) = \left\{ \boldsymbol{f} \colon [0, L] \to \mathbb{C}^n \;\middle|\; f_1, \dots, f_n \in \mathrm{L}^1([0, L]; \mathbb{C}) \right\}.$$

We can then make the following more or less obvious definition.

5.1.10 Definition (Continuous-discrete Fourier transform II) The ***continuous-discrete Fourier transform*** (***CDFT***) of $\boldsymbol{f} \in \mathrm{L}^1([0, L]; \mathbb{C}^n)$ is the function $\mathscr{F}_{\mathrm{CD}}(\boldsymbol{f}) \colon \mathbb{Z} \to \mathbb{C}^n$ defined by

$$\mathscr{F}_{\mathrm{CD}}(\boldsymbol{f})(n) = (\mathscr{F}_{\mathrm{CD}}(f_1)(n), \dots, \mathscr{F}_{\mathrm{CD}}(f_n)(n)).$$ •

The inverse of the CDFT in this case is also made component-wise. Thus if $\boldsymbol{F}\colon \mathbb{Z} \to \mathbb{C}^n$, we denote

$$\mathscr{F}_{\mathrm{CD}}^{-1}(\boldsymbol{F})(x) = (\mathscr{F}_{\mathrm{CD}}^{-1}(F_1)(x), \dots, \mathscr{F}_{\mathrm{CD}}^{-1}(F_n)(x)).$$

Of course, all of the caveats we made in Section 5.1.1.2 apply to the inverse in this case as well.

Next we consider the case when we have a function of multiple variables. For $L_1, \dots, L_n \in \mathbb{R}_{>0}$, we denote

$$C(\boldsymbol{L}) = [0, L_1] \times \cdots \times [0, L_n];$$

thus $C(\boldsymbol{L})$ is an n-dimensional cube. We also need to integrate functions of multiple variables. To do this, if $f\colon C(\boldsymbol{L}) \to \mathbb{C}$, then we denote

$$\int_{C(\boldsymbol{L})} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_0^{L_n} \cdots \int_0^{L_1} f(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

We also denote

$$\mathrm{L}^1(C(\boldsymbol{L}); \mathbb{C}) = \left\{ f\colon C(\boldsymbol{L}) \to \mathbb{C} \;\middle|\; \int_{C(\boldsymbol{L})} |f(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} < \infty \right\}.$$

With this notation, we make the following definition.

5.1.11 Definition (Continuous-discrete Fourier transform III) The *continuous-discrete Fourier transform* (*CDFT*) of $f \in \mathrm{L}^1(C(\boldsymbol{L}); \mathbb{C})$ is the function $\mathscr{F}_{\mathrm{CD}}(f)\colon \mathbb{Z}^n \to \mathbb{C}$ defined by

$$\mathscr{F}_{\mathrm{CD}}(f)(\boldsymbol{k}) = \int_{C(\boldsymbol{L})} f(x_1, \dots, x_n) \mathrm{e}^{-2\pi \mathrm{i}(k_1 \frac{x_1}{L_1} + \cdots + k_n \frac{x_n}{L_n})}\,\mathrm{d}\boldsymbol{x} \qquad \bullet$$

The inverse of the multivariable CDFT is then determined by analogy with the single-variable case. Thus, if $F\colon \mathbb{Z}^n \to \mathbb{C}$, we denote

$$\mathscr{F}_{\mathrm{CD}}^{-1}(F)(\boldsymbol{x}) = \frac{1}{L_1 \cdots L_n} \sum_{k_1 \in \mathbb{Z}} \cdots \sum_{k_n \in \mathbb{Z}} F(k_1, \dots, k_n) \mathrm{e}^{2\pi \mathrm{i} k_1 \frac{x_1}{L_1}} \cdots \mathrm{e}^{2\pi \mathrm{i} k_n \frac{x_n}{L_n}}.$$

Of course, care must be taken with interpreting this multiple infinite sum, just as in the single-variable case.

5.1.2 The continuous-continuous Fourier transform

The next transform we consider is one that will be applied to functions whose domain is unbounded, and we shall take the domain to be $\mathbb{R}$. In this case, the natural domain for the transform is the set

$$\mathrm{L}^1(\mathbb{R}; \mathbb{C}) = \left\{ f\colon \mathbb{R} \to \mathbb{C} \;\middle|\; \int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x < \infty \right\}$$

of functions whose modulus is integrable.

5.1.2.1 The transform

As with the CDFT, we consider $\mathbb{C}$-valued functions.

5.1.12 Definition (Continuous-continuous Fourier transform I) The *continuous-continuous Fourier transform* (*CCFT*) of $f \in \mathrm{L}^1(\mathbb{R};\mathbb{C})$ is the function $\mathscr{F}_{\mathrm{CC}}(f)\colon \mathbb{R} \to \mathbb{C}$ defined by

$$\mathscr{F}_{\mathrm{CC}}(f)(\nu) = \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x. \qquad \bullet$$

Let us compute the CCFT of a few functions so as to see how it works.

5.1.13 Examples (Continuous-continuous Fourier transform)

1. We let $L \in \mathbb{R}_{>0}$, $m \in \mathbb{Z}_{\geq 0}$, and take

$$f(x) = \begin{cases} \cos(2\pi m\frac{x}{L}), & |x| \leq \frac{L}{2}, \\ 0, & |x| > \frac{L}{2}. \end{cases}$$

We then calculate, for $\nu \neq \pm\frac{m}{L}$,

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f)(\nu) &= \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x \\
&= \int_{-L/2}^{L/2} \cos(2\pi m\tfrac{x}{L})\cos(2\pi\nu x)\,\mathrm{d}x - \mathrm{i}\int_{-L/2}^{L/2} \cos(2\pi m\tfrac{x}{L})\sin(2\pi\nu x)\,\mathrm{d}x \\
&= \int_{0}^{L/2} (\cos(2\pi(\tfrac{m}{L}-\nu)x) + \cos(2\pi(\tfrac{m}{L}+\nu)x))\,\mathrm{d}x \\
&= \left(\frac{\sin(2\pi(\frac{m}{L}-\nu)x)}{2\pi(\frac{m}{L}-\nu)} + \frac{\sin(2\pi(\frac{m}{L}+\nu)x)}{2\pi(\frac{m}{L}+\nu)}\right)\Bigg|_0^{L/2} \\
&= \left(\frac{\sin(\pi(m-L\nu))}{2\pi(\frac{m}{L}-\nu)} + \frac{\sin(\pi(m+L\nu))}{2\pi(\frac{m}{L}+\nu)}\right) \\
&= \left(\frac{(-1)^{m+1}\sin(\pi\nu L)}{2\pi(\frac{m}{L}-\nu)} + \frac{(-1)^{m+1}\sin(\pi\nu L)}{2\pi(\frac{m}{L}+\nu)}\right) \\
&= \frac{(-1)^{m+1}mL\sin(\pi\nu L)}{2\pi(m^2-L^2\nu^2)},
\end{aligned}$$

using the fact that sin is odd and cos is even, using (5.1), and using as well as the trigonometric identities

$$\begin{aligned}
\sin(\alpha+\beta) &= \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta), \\
\sin(\alpha-\beta) &= \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta).
\end{aligned} \tag{5.3}$$

For $\nu = \pm\frac{m}{L}$ we have

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f)(\pm\tfrac{m}{L}) &= \int_{-L/2}^{L/2} \cos^2(2\pi m\tfrac{x}{L})\,\mathrm{d}x \mp \mathrm{i}\int_{-L/2}^{L/2} \cos(2\pi m\tfrac{x}{L})\sin(2\pi m\tfrac{x}{L})\,\mathrm{d}x \\
&= \int_{0}^{L/2} (1+\cos(4\pi m\tfrac{x}{L}))\,\mathrm{d}x = \frac{L}{2},
\end{aligned}$$

using the trigonometric identity

$$\cos^2\alpha = \tfrac{1}{2}(1+\cos(2\alpha)).$$

In summary,

$$\mathscr{F}_{\mathrm{CC}}(f)(\nu) = \begin{cases} \frac{(-1)^{m+1}mL\sin(\pi\nu L)}{2\pi(m^2-L^2\nu^2)}, & \nu \neq \pm\frac{m}{L}, \\ \frac{L}{2}, & \nu = \pm\frac{m}{L}. \end{cases}$$

2. Next we take $f(x) = \sin(2\pi m\frac{x}{L})$ for $m \in \mathbb{Z}_{>0}$ and $L \in \mathbb{R}_{>0}$. We then have, for $\nu \neq \pm\frac{m}{L}$,

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f)(\nu) &= \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x \\
&= \int_{-L/2}^{L/2} \sin(2\pi m\tfrac{x}{L})\cos(2\pi\nu x)\,\mathrm{d}x + \mathrm{i}\int_{-L/2}^{L/2} \sin(2\pi m\tfrac{x}{L})\sin(2\pi\nu x)\,\mathrm{d}x \\
&= \mathrm{i}\int_{0}^{L/2} (\cos(2\pi(\tfrac{m}{L}-\nu)x) - \cos(2\pi(\tfrac{m}{L}+\nu)x)\,\mathrm{d}x \\
&= \mathrm{i}\left(\frac{\sin(2\pi(\frac{m}{L}-\nu)x)}{2\pi(\frac{m}{L}-\nu)} - \frac{\sin(2\pi(\frac{m}{L}+\nu)x)}{2\pi(\frac{m}{L}+\nu)}\right)\Bigg|_0^{L/2} \\
&= \mathrm{i}\left(\frac{(-1)^{m+1}\sin(\pi\nu L)}{2\pi(\frac{m}{L}-\nu)} + \frac{(-1)^{m+1}\sin(\pi\nu L)}{2\pi(\frac{m}{L}+\nu)}\right) \\
&= \mathrm{i}\frac{(-1)^{m+1}mL\sin(\pi\nu L)}{2\pi(m^2-L^2\nu^2)},
\end{aligned}$$

using the fact that cos is even and sin is odd, and using (5.1) and (5.3). For $\nu = \pm\frac{m}{L}$,

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f)(\pm\tfrac{m}{L}) &= \int_{-L/2}^{L/2} \sin(2\pi m\tfrac{x}{L})\cos(2\pi\nu x)\,\mathrm{d}x \pm \mathrm{i}\int_{-L/2}^{L/2} \sin^2(2\pi m\tfrac{x}{L})\,\mathrm{d}x \\
&= \mathrm{i}\int_0^{L/2} (1-\cos(4\pi m\tfrac{x}{L}))\,\mathrm{d}x = \pm\mathrm{i}\frac{L}{2}.
\end{aligned}$$

In summary,

$$\mathscr{F}_{\mathrm{CC}}(f)(\nu) = \begin{cases} \mathrm{i}\frac{(-1)^{m+1}mL\sin(\pi\nu L)}{2\pi(m^2-L^2\nu^2)}, & \nu \neq \pm\frac{m}{L}, \\ \mathrm{i}\frac{L}{2}, & \nu = \frac{m}{L}, \\ -\mathrm{i}\frac{L}{2}, & \nu = -\frac{m}{L}. \end{cases}$$

3. As a final example, let us take, for $L \in \mathbb{R}_{>0}$,

$$f(x) = \begin{cases} -1, & x \in [-\frac{L}{2}, 0], \\ 1, & x \in (0, \frac{L}{2}], \\ 0, & \text{otherwise}. \end{cases}$$

Here we compute

$$\mathscr{F}_{\mathrm{CC}}(f)(0) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 0$$

and, for $\nu \neq 0$,

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f)(\nu) &= \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x \\
&= -\int_{-L/2}^{0} \mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x + \int_{0}^{L/2} \mathrm{e}^{-2\pi\mathrm{i}\nu x}\,\mathrm{d}x \\
&= \frac{\mathrm{e}^{-2\pi\mathrm{i}\nu x}}{2\pi\mathrm{i}\nu}\Big|_{-L/2}^{0} - \frac{\mathrm{e}^{-2\pi\mathrm{i}\nu x}}{2\pi\mathrm{i}\nu}\Big|_{0}^{L/2} \\
&= \frac{1-\mathrm{e}^{\pi\mathrm{i}\nu L}}{2\pi\mathrm{i}\nu} - \frac{\mathrm{e}^{-\pi\mathrm{i}\nu L}-1}{2\pi\mathrm{i}\nu} \\
&= -\mathrm{i}\frac{2-(\mathrm{e}^{\pi\mathrm{i}\nu L}+\mathrm{e}^{-\pi\mathrm{i}\nu L})}{2\pi\nu} = \mathrm{i}\frac{\cos(\pi\nu L)-1}{\pi\nu}.
\end{aligned}$$

Thus, in summary,

$$\mathscr{F}_{\mathrm{CC}}(f)(\nu) = \begin{cases} 0, & \nu = 0, \\ \mathrm{i}\frac{\cos(\pi\nu L)-1}{\pi\nu}, & \nu \neq 0. \end{cases}$$

•

While the CCFT may indeed, be something calculable, one does not often want to calculate it. Instead, one is interested in its basic properties, to whose consideration we now turn.

First we show the linearity of the CCFT.

5.1.14 Proposition (Linearity of the CCFT) *The CCFT is a linear map:*

$$\mathscr{F}_{\mathrm{CC}}(\mathrm{f}_1 + \mathrm{f}_2)(\nu) = \mathscr{F}_{\mathrm{CC}}(\mathrm{f}_1)(\nu) + \mathscr{F}_{\mathrm{CC}}(\mathrm{f}_2)(\nu), \quad \mathscr{F}_{\mathrm{CC}}(\mathrm{af}) = \mathrm{a}\mathscr{F}_{\mathrm{CC}}(\mathrm{f}),$$

for every $\mathrm{f}, \mathrm{f}_1, \mathrm{f}_2 \in \mathrm{L}^1(\mathbb{R};\mathbb{C})$ *and* $\mathrm{a} \in \mathbb{C}$.

Proof This follows by linearity of the integral. ■

As with all of our transform, an essential matter to understand is how they interact with differentiation. For the CCFT this is recorded by the following proposition.

5.1.15 Proposition (CCFT and differentiation) *Let* $\mathrm{f} \in \mathrm{L}^1(\mathbb{R};\mathbb{C})$ *be continuously differentiable and such that* $\frac{\mathrm{df}}{\mathrm{dx}} \in \mathrm{L}^1(\mathbb{R};\mathbb{C})$. *Then*

$$\mathscr{F}_{\mathrm{CC}}(\tfrac{\mathrm{df}}{\mathrm{dx}})(\mathrm{x}) = 2\pi\mathrm{i}\nu\mathscr{F}_{\mathrm{CC}}(\mathrm{f})(\mathrm{x}).$$

Proof First we claim that $\lim_{|x|\to\infty} f(x) = 0$. Indeed, note that

$$f(x) = \int_0^x \tfrac{\mathrm{d}f}{\mathrm{d}x}(y)\,\mathrm{d}y.$$

Since $\frac{df}{dx} \in L^1(\mathbb{R};\mathbb{C})$, the limits

$$\lim_{x\to\infty} \int_0^x \frac{df}{dx}(y)\,dy, \quad \lim_{x\to-\infty} \int_0^x \frac{df}{dx}(y)\,dy$$

exist. Moreover, since $f \in L^1(\mathbb{R};\mathbb{C})$,

$$\lim_{x\to\infty} \int_0^x \frac{df}{dx}(y)\,dy = 0, \quad \lim_{x\to-\infty} \int_0^x \frac{df}{dx}(y)\,dy = 0.$$

Now, using integration by parts we compute

$$\begin{aligned}\mathcal{F}_{CC}(\tfrac{df}{dx})(\nu) &= \int_{-\infty}^{\infty} \frac{df}{dx}(x)e^{-2\pi i\nu x}\,dx \\ &= f(x)e^{-2\pi i\nu x}\Big|_{-\infty}^{\infty} + 2\pi i\nu \int_{-\infty}^{\infty} f(x)e^{-2\pi i\nu x}\,dx \\ &= 2\pi i\nu \mathcal{F}_{CC}(f)(\nu),\end{aligned}$$

as claimed. ■

The proposition can be applied recursively to higher-order derivatives.

5.1.16 Corollary (CCFT and higher-order derivatives) *Let* $f \in L^1(\mathbb{R};\mathbb{C})$ *be* k*-times continuously differentiable and suppose that* $\frac{d^k f}{dx^k} \in L^1(\mathbb{R};\mathbb{C})$. *Then*

$$\mathcal{F}_{CC}(\tfrac{d^k f}{dx^k})(\nu) = (2\pi i\nu)^k \mathcal{F}_{CC}(f)(\nu).$$

5.1.2.2 The inverse transform The inverse transform for the CCFT is even more difficult to motivate than that for the CDFT. However, we shall outline a process that is as valid as that we used for the CDFT in Section 5.1.1.2, which is to say it is not valid at all.

We hypothesise that we can write

$$f(x) = \int_{-\infty} c_\nu(f)e^{2\pi i\nu x}\,d\nu,$$

for some function $\nu \mapsto c_\nu$, analogously to (5.2). Flatly assuming this, we then make the following computations:

$$\begin{aligned} & f(x) = \lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} c_\nu(f)e^{2\pi i\nu x}\,d\nu \\ \implies & \int_{-\infty}^{\infty} f(x)e^{-2\pi i\mu x}\,dx = \lim_{\Omega\to\infty} \int_{-\infty}^{\infty} \left(\int_{-\Omega}^{\Omega} c_\nu(f)e^{2\pi i\nu x}\,d\nu \right) e^{-2\pi i\mu x}\,dx \\ \implies & \int_{-\infty}^{\infty} f(x)e^{-2\pi i\mu x}\,dx = \lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} c_\nu(f) \underline{\left(\int_{-\infty}^{\infty} e^{2\pi i\nu x}e^{-2\pi i\mu x}\,dx \right)} d\nu. \end{aligned}$$

In the case of the CDFT, the integral underlined in the preceding formula makes sense and can be given explicit form. This is not so for the CCFT, since the underlined integral simply does not exist in the standard way. It does exist in a certain sense, the precise characterisation of which resides somewhat beyond the scope of our presentation. We can, however, sketch how this works. To do this, we fix $M \in \mathbb{R}_{>0}$ and consider the computation for $\nu \neq \mu$:

$$\int_{-M}^{M} \mathrm{e}^{2\pi \mathrm{i} y x} \, \mathrm{d}x = \frac{\mathrm{e}^{2\pi \mathrm{i} y x}}{2\pi \mathrm{i} y x}\Big|_{-M}^{M} = \frac{\sin(2\pi M y)}{\pi y},$$

using the identity

$$\sin\theta = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\theta} - \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta}).$$

Let us make a few observations without proof:

1. $\displaystyle\int_{-\infty}^{\infty} \frac{\sin(2\pi M y)}{\pi y} \, \mathrm{d}y = 1;$
2. as $M \to \infty$, the function $y \mapsto \frac{\sin(2\pi M y)}{\pi y}$ gets "focussed" about $y = 0$, as shown in Figure 5.1.

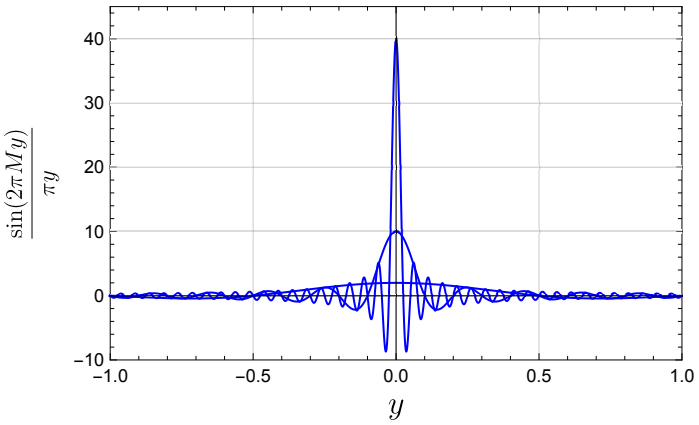


Figure 5.1 The focussing of $\frac{\sin(2\pi M y)}{\pi y}$ as $M \to \infty$

Now we compute

$$\lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} c_\nu(f) \left(\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}(\nu-\mu)x} \, \mathrm{d}x \right) \mathrm{d}\nu = \lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} c_{\mu+y}(f) \left(\lim_{M\to\infty} \int_{-M}^{M} \mathrm{e}^{2\pi \mathrm{i} y x} \, \mathrm{d}x \right) \mathrm{d}\nu$$
$$= \lim_{M,\Omega\to\infty} \int_{-\Omega}^{\Omega} c_{\mu+y} \frac{\sin(2\pi M y)}{\pi y} \, \mathrm{d}y.$$

As $M \to \infty$, because of the behaviour we see in Figure 5.1, the integral outside a small interval around $y = 0$ goes to zero. Moreover, because

$$\int_{-\infty}^{\infty} \frac{\sin(2\pi M y)}{\pi y} \, \mathrm{d}y = 1$$

and (assuming that $\nu \mapsto c_\nu(f)$ is continuous) since continuous functions are approximately constant about any point, we have

$$\lim_{M\to\infty}\left(\lim_{\Omega\to\infty}\int_{-\Omega}^{\Omega} c_\nu(f)\left(\lim_{M\to\infty}\int_{-M}^{M} e^{2\pi i(\nu-\mu)x}\,dx\right)d\nu\right) = c_\mu(f).$$

Putting this all together,

$$c_\nu(f) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i\nu x}\,dx = \mathscr{F}_{CC}(f)(\nu).$$

Thus we conclude that

$$f(x) = \lim_{\Omega\to\infty}\int_{-\Omega}^{\Omega} \mathscr{F}_{CC}(f)(\nu)e^{2\pi i\nu x}\,d\nu.$$

The preceding conclusion, and its derivation, are complete nonsense. For example, we cavalierly (and incorrectly) swapped limits and integrals multiple times without justification and we made some unwarranted conclusions based on asserted (and not entirely correct) properties of $\frac{\sin(2\pi My)}{\pi y}$. Nonetheless, the integral on the right-hand side of this formula,

$$\int_{-\infty}^{\infty} \mathscr{F}_{CC}(f)(\nu)e^{2\pi i\nu x}\,d\nu,$$

is important and is called the ***Fourier integral*** for f.

5.1.17 Remarks (Convergence of Fourier integrals) Let us make a few comments about the existence of the limit

$$\lim_{\Omega\to\infty}\int_{-\Omega}^{\Omega} \mathscr{F}_{CC}(f)(\nu)e^{2\pi i\nu x}\,d\nu.$$

1. If $f \in L^1(\mathbb{R};\mathbb{C})$, then generally the integral does not converge. Indeed, as with Fourier series, it can be shown that there exists $f \in L^1(\mathbb{R};\mathbb{C})$ such that the integral does not converge for any $x \in \mathbb{R}$.
2. Even if $f \in L^1(\mathbb{R};\mathbb{C})$ is continuous, the integral may diverge, although, sometimes, it will converge for "most" $x \in \mathbb{R}$.
3. If $f \in L^1(\mathbb{R};\mathbb{C})$ is continuously differentiable, if its derivative is in $L^1(\mathbb{R};\mathbb{C})$, and if, additionally,
$$\int_{-\infty}^{\infty} |f(x)|^2\,dx < \infty,$$
then the Fourier integral converges to x for each $x \in \mathbb{R}$. •

Although the derivation we give for the inverse of the CCFT is not always valid, we make the following "definition."

5.1.18 "Definition" (Inverse of the CCFT) The *inverse* of the CCFT is the mapping that assigns to a map $F\colon \mathbb{R} \to \mathbb{C}$ the function $\mathscr{F}_{CC}^{-1}(F)\colon \mathbb{R} \to \mathbb{C}$ defined by

$$\mathscr{F}_{CC}^{-1}(F)(x) = \int_{-\infty}^{\infty} F(\nu)e^{2\pi i\nu x}\,d\nu. \qquad \bullet$$

Of course, the "definition" of the inverse makes no sense at all, since the integral defining it will generally not exist. As with the inverse of the CDFT, what *is* true is that, under suitable hypotheses on f, $\mathscr{F}_{CC}^{-1} \circ \mathscr{F}_{CC}(f) = f$. In other words, the given inverse allows us to sometimes recover a function f from its CCFT.

5.1.2.3 Convolution and the continuous-continuous Fourier transform In this section we consider the relationship of convolution with products for the CCFT. The transformed variables in this case reside in $\mathbb{R}$, and so transformed functions are functions from $\mathbb{R}$ to $\mathbb{C}$. Thus, in this case, we have $F, G\colon \mathbb{R} \to \mathbb{C}$ and so the product of F and G is the function

$$\begin{aligned} FG\colon \mathbb{R} &\to \mathbb{C} \\ \nu &\mapsto F(\nu)G(\nu). \end{aligned}$$

What we want to know is, if $F = \mathscr{F}_{CC}(f)$ and $G = \mathscr{F}_{CC}(g)$, is there a function $h\colon \mathbb{R} \to \mathbb{C}$ for which $\mathscr{F}_{CC}(h) = FG$?

To answer this question, we make the following definition.

5.1.19 Definition (Convolution) If $f, g \in L^1(\mathbb{R};\mathbb{C})$, the ***convolution*** of f and g is the function

$$\begin{aligned} f * g\colon \mathbb{R} &\to \mathbb{C} \\ x &\mapsto \int_{-\infty}^{\infty} f(x-y)g(y)\,dy. \end{aligned} \qquad \bullet$$

The operation of convolution is a interesting one, and has many properties that merit further exploration. For our purposes, we merely point out the following result.

5.1.20 Proposition (CCFT and convolution) *If* $\mathrm{f}, \mathrm{g} \in L^1(\mathbb{R};\mathbb{C})$, *then*

$$\mathscr{F}_{CC}(\mathrm{f} * \mathrm{g}) = \mathscr{F}_{CD}(\mathrm{f})\mathscr{F}_{CD}(\mathrm{g}).$$

Proof This is a fairly straightforward application of Fubini's Theorem and the

change of variables theorem:

$$\begin{aligned}
\mathscr{F}_{\mathrm{CC}}(f * g)(\nu) &= \int_{-\infty}^{\infty} f * g(x) \mathrm{e}^{-2\pi \mathrm{i} \nu x} \, \mathrm{d}x = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x-y) g(y) \, \mathrm{d}y \right) \mathrm{e}^{-2\pi \mathrm{i} \nu x} \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} f(x-y) \mathrm{e}^{-2\pi \mathrm{i} \nu x} \, \mathrm{d}x \right) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} g(\sigma) \left(\int_{-\infty}^{\infty} f(\tau) \mathrm{e}^{-2\pi \mathrm{i} \nu (\sigma + \tau)} \, \mathrm{d}\tau \right) \mathrm{d}\sigma \\
&= \left(\int_{-\infty}^{\infty} g(\sigma) \mathrm{e}^{-2\pi \mathrm{i} \nu \sigma} \, \mathrm{d}\sigma \right) \left(\int_{-\infty}^{\infty} f(\tau) \mathrm{e}^{-2\pi \mathrm{i} \nu \tau} \, \mathrm{d}\tau \right) \\
&= \mathscr{F}_{\mathrm{CC}}(f)(\nu) \mathscr{F}_{\mathrm{CC}}(g)(\nu),
\end{aligned}$$

as claimed. ■

5.1.2.4 Extension to higher-dimensions In this section we extend the definition of the CCFT to (1) functions with values in $\mathbb{C}^n$ and (2) functions with multiple independent variables.

First we extend the CCFT to functions with values in $\mathbb{C}^n$. To do so, we note that, if $\boldsymbol{f} \colon \mathbb{R} \to \mathbb{C}^n$, then we can write

$$\boldsymbol{f}(x) = (f_1(x), \dots, f_n(x))$$

for functions $f_1, \dots, f_n \colon \mathbb{R} \to \mathbb{C}$. We then denote

$$\mathrm{L}^1(\mathbb{R}; \mathbb{C}^n) = \left\{ \boldsymbol{f} \colon \mathbb{R} \to \mathbb{C}^n \;\middle|\; f_1, \dots, f_n \in \mathrm{L}^1(\mathbb{R}; \mathbb{C}) \right\}.$$

We can then make the following more or less obvious definition.

5.1.21 Definition (Continuous-continuous Fourier transform II) The *continuous-continuous Fourier transform* (*CCFT*) of $\boldsymbol{f} \in \mathrm{L}^1(\mathbb{R}; \mathbb{C}^n)$ is the function $\mathscr{F}_{\mathrm{CC}}(\boldsymbol{f}) \colon \mathbb{R} \to \mathbb{C}^n$ defined by

$$\mathscr{F}_{\mathrm{CC}}(\boldsymbol{f})(\nu) = (\mathscr{F}_{\mathrm{CC}}(f_1)(\nu), \dots, \mathscr{F}_{\mathrm{CC}}(f_n)(\nu)).$$ •

The inverse of the CCFT in this case is also made component-wise. Thus if $\boldsymbol{F} \colon \mathbb{R} \to \mathbb{C}^n$, we denote

$$\mathscr{F}_{\mathrm{CC}}^{-1}(\boldsymbol{F})(x) = (\mathscr{F}_{\mathrm{CC}}^{-1}(F_1)(x), \dots, \mathscr{F}_{\mathrm{CC}}^{-1}(F_n)(x)).$$

Of course, all of the caveats we made in Section 5.1.2.2 apply to the inverse in this case as well.

Next we consider the case when we have a function of multiple variables. To do this, if $f \colon \mathbb{R}^n \to \mathbb{C}$, then we denote

$$\int_{\mathbb{R}^n} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

We also denote

$$\mathrm{L}^1(\mathbb{R}^n;\mathbb{C}) = \left\{ f\colon \mathbb{R}^n \to \mathbb{C} \;\middle|\; \int_{\mathbb{R}^n} |f(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} < \infty \right\}.$$

With this notation, we make the following definition, recalling the notation $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ for the Euclidean inner product.

5.1.22 Definition (Continuous-continuous Fourier transform III) The *continuous-continuous Fourier transform* (*CCFT*) of $f \in \mathrm{L}^1(\mathbb{R}^n;\mathbb{C})$ is the function $\mathscr{F}_{\mathrm{CC}}(f)\colon \mathbb{R}^n \to \mathbb{C}$ defined by

$$\mathscr{F}_{\mathrm{CC}}(f)(\boldsymbol{\nu}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{e}^{-2\pi \mathrm{i} \langle \boldsymbol{\nu}, \boldsymbol{x} \rangle_{\mathbb{R}^n}}\,\mathrm{d}\boldsymbol{x} \qquad \bullet$$

The inverse of the multivariable CCFT is then determined by analogy with the single-variable case. Thus, if $F\colon \mathbb{R}^n \to \mathbb{C}$, we denote

$$\mathscr{F}_{\mathrm{CC}}^{-1}(F)(\boldsymbol{x}) = \int_{\mathbb{R}^n} F(\boldsymbol{\nu}) \mathrm{e}^{2\pi \mathrm{i} \langle \boldsymbol{\nu}, \boldsymbol{x} \rangle_{\mathbb{R}^n}}\,\mathrm{d}\boldsymbol{x}.$$

Of course, care must be taken with interpreting this multiple integral, just as in the single-variable case.

5.1.3 The Laplace transform

Next we consider a transform that is a little different in flavour than the two Fourier transforms we just discussed. The domain of this transform consists of the following class of functions.

$$\mathscr{E}(\mathbb{R}_{\geq 0};\mathbb{C}) = \left\{ f\colon \mathbb{R}_{\geq 0} \to \mathbb{C} \;\middle|\; |f(x)| \leq M\mathrm{e}^{\sigma x},\ x \in \mathbb{R}_{\geq 0},\ \text{for some } M \in \mathbb{R}_{>0},\ \sigma \in \mathbb{R} \right\}.$$

For $f \in \mathscr{E}(\mathbb{R}_{\geq 0};\mathbb{C})$ we denote

$$\sigma(f) = \inf\{a \in \mathbb{R} \mid\ |f(x)| \leq M\mathrm{e}^{ax} \text{ for some } M \in \mathbb{R}_{>0}\}.$$

Also, for $a \in \mathbb{R}$, we denote

$$\mathbb{C}_{>a} = \{z \in \mathbb{C} \mid\ \mathrm{Re}(z) > a\}.$$

We shall say a function in $\mathscr{E}(\mathbb{R}_{\geq 0};\mathbb{C})$ is of ***exponential class***. This is a convenient class of functions to work with, for reasons that will be clear as we proceed.

5.1.3.1 The transform

The transform we consider is the following.

5.1.23 Definition (Laplace transform I) The *Laplace transform* of $f \in \mathscr{E}(\mathbb{R}_{\geq 0};\mathbb{C})$ is the function $\mathscr{L}(f)\colon \mathbb{C}_{>\sigma(f)} \to \mathbb{C}$ defined by

$$\mathscr{L}(f)(z) = \int_0^\infty f(x)\mathrm{e}^{-zx}\,\mathrm{d}x. \qquad \bullet$$

Let us compute explicitly a few Laplace transforms.

5.1.24 Examples (Laplace transform) We will consider Laplace transforms of the pretty uninteresting functions considered in Definition 2.3.12. In all cases, functions are defined on $\mathbb{R}_{\geq 0}$.

1. First let us consider $f(x) = x^k$ for $k \in \mathbb{Z}_{\geq 0}$. We note that $f \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ with $\sigma(f) = 0$. To see this, note that, if $\sigma \in \mathbb{R}_{>0}$, then $\lim_{x\to\infty} x^k e^{-\sigma x} = 0$. Since $x \mapsto x^k e^{-\sigma x}$ is continuous on $\mathbb{R}_{\geq 0}$, it is bounded. Therefore, there exists $M \in \mathbb{R}_{>0}$ such that
$$|x^k| e^{-\sigma x} \leq M, \qquad x \in \mathbb{R}_{\geq 0}.$$
Thus
$$|f(x)| = |x^k| \leq M e^{\sigma x}.$$
Also, if $\sigma \in \mathbb{R}_{\leq 0}$, then $\lim_{x\to\infty} x^k e^{-\sigma x} = \infty$. Thus, for any $M \in \mathbb{R}_{>0}$, for any sufficiently large x we have
$$|f(x)| e^{-\sigma x} = |x^k| e^{-\sigma x} \geq M \quad \Longrightarrow \quad |f(x)| \geq M e^{\sigma x}.$$
Thus the greatest lower bound of all σ's for which $|f(x)| \leq M e^{\sigma x}$ for some $M \in \mathbb{R}_{>0}$ is 0, i.e., $\sigma(f) = 0$.
Next, we claim that
$$\mathscr{L}(f)(z) = \frac{k!}{z^{k+1}}.$$
We can prove this by induction on k. For $k = 0$ we have
$$\int_0^\infty e^{-zx}\, dx = -\frac{e^{-zx}}{z}\Big|_0^\infty = \frac{1}{z},$$
and so our claim is true when $k = 0$. So suppose the claim is true for $k = m$ and let $k = m + 1$. We then have, using integration by parts,
$$\begin{aligned}\int_0^\infty x^{m+1} e^{-zx}\, dx &= -\frac{x^{m+1} e^{-zx}}{z}\Big|_0^\infty + \frac{m+1}{z}\int_0^\infty x^m e^{-zx}\, dz \\ &= \frac{m+1}{z}\frac{m!}{z^{m+1}} = \frac{(m+1)!}{z^{m+2}},\end{aligned}$$
as claimed.
2. Next we consider $f(x) = e^{ax}$ for $a \in \mathbb{R}$, noting that $f \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ with $\sigma(f) = a$.[1] To see this, note that, if $\sigma \geq a$, then $\lim_{x\to\infty} e^{ax} e^{-\sigma x} = 0$. Since $x \mapsto e^{ax} e^{-\sigma x}$ is continuous on $\mathbb{R}_{\geq 0}$, it is bounded. Therefore, there exists $M \in \mathbb{R}_{>0}$ such that
$$|e^{ax}| e^{-\sigma x} \leq M, \qquad x \in \mathbb{R}_{\geq 0}.$$
Thus
$$|f(x)| = |e^{ax}| \leq M e^{\sigma x}.$$

[1] In fact, we can consider $a \in \mathbb{C}$, in which case $\sigma(f) = \mathrm{Re}(z)$.

Also, if $\sigma \leq a$, then $\lim_{x\to\infty} e^{ax}e^{-\sigma x} = \infty$. Thus, for any $M \in \mathbb{R}_{>0}$, for any sufficiently large x we have

$$|f(x)|e^{-\sigma x} = |e^{ax}|e^{-\sigma x} \geq M \quad \Longrightarrow \quad |f(x)| \geq Me^{\sigma x}.$$

Thus the greatest lower bound of all σ's for which $|f(x)| \leq Me^{\sigma x}$ for some $M \in \mathbb{R}_{>0}$ is a, i.e., $\sigma(f) = a$.

In this case we can calculate

$$\mathscr{L}(f)(z) = \int_0^\infty e^{ax}e^{-zx}\,dx = \frac{e^{(a-z)x}}{a-z}\Big|_0^\infty = \frac{1}{z-a}.$$

3. Let us "combine" the preceding two examples and consider $f(x) = t^k e^{ax}$ for $k \in \mathbb{Z}_{\geq 0}$ and $a \in \mathbb{R}$.[2] Here we again have $\sigma(f) = a$, by an argument rather like that in part 2 above. In this case we have, by a change of variable $\zeta = z - a$,

$$\mathscr{L}(f)(z) = \int_0^\infty x^k e^{ax}e^{-zx}\,dx = \int_0^\infty x^k e^{-\zeta x}\,dx = \frac{k!}{\zeta^{k+1}} = \frac{k!}{(z-a)^{k+1}}.$$

4. Consider $f(x) = \sin(\omega x)$ for $\omega \in \mathbb{R}$. Then $f \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ and $\sigma(f) = 0$. We have, using integration by parts,

$$\begin{aligned}
\mathscr{L}(f)(z) &= \int_0^\infty \sin(\omega x)e^{-zx}\,dz \\
&= -\frac{\sin(\omega x)e^{-zx}}{z}\Big|_0^\infty + \frac{\omega}{z}\int_0^\infty \cos(\omega x)e^{-zx}\,dx \\
&= -\frac{\omega}{z}\frac{\cos(\omega x)e^{-zx}}{z}\Big|_0^\infty - \frac{\omega^2}{z^2}\int_0^\infty \sin(\omega x)e^{-zx}\,dx \\
&= \frac{\omega}{z^2}\left(1 - \omega\mathscr{L}(f)(z)\right).
\end{aligned}$$

Thus we can solve for $\mathscr{L}(f)(z)$ as

$$\mathscr{L}(f)(z) = \frac{\omega}{z^2 + \omega^2}.$$

5. We can perform a similar computation for $f(x) = \cos(\omega x)$. Here again, we have $\sigma(f) = 0$. We also compute

$$\begin{aligned}
\mathscr{L}(f)(z) &= \int_0^\infty \cos(\omega x)e^{-zx}\,dx \\
&= -\frac{\cos(\omega x)e^{-zx}}{z}\Big|_0^\infty - \frac{\omega}{z}\int_0^\infty \sin(\omega x)e^{-zx}\,dx \\
&= \frac{1}{z} + \frac{\omega}{z}\frac{\sin(\omega x)e^{-zx}}{z}\Big|_0^\infty - \frac{\omega^2}{z^2}\int_0^\infty \cos(\omega x)e^{-zx}\,dx \\
&= \frac{1}{z} + \frac{\omega^2}{z^2}\mathscr{L}(f)(z).
\end{aligned}$$

[2] As previous, we can consider $a \in \mathbb{C}$.

Solving for $\mathscr{L}(f)$ gives

$$\mathscr{L}(f)(z) = \frac{z}{z^2 + \omega^2}.$$

6. Now we combine all of the preceding computations to derive the Laplace transform of a general pretty uninteresting function. To do this, we first consider the $\mathbb{C}$-valued function $f(x) = x^k e^{(\sigma+i\omega)x}$. Here $\sigma(f) = \sigma$ (forgiving the abuse of notation). From 3 we have

$$\begin{aligned}\mathscr{L}(f)(z) &= \frac{k!}{((z-\sigma)+i\omega)^{k+1}} = \frac{k!((z-\sigma)-i\omega)^{k+1}}{((z-\sigma)^2+\omega^2)^{k+1}} \\ &= \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{k}{2j} \frac{(-1)^j k!(z-\sigma)^{k-2j}\omega^{2j}}{((z-\sigma)^2+\omega^2)^{k+1}} \\ &\quad + i \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{k}{2j+1} \frac{(-1)^{j+1} k!(z-\sigma)^{k-2j-1}\omega^{2j+1}}{((z-\sigma)^2+\omega^2)^{k+1}},\end{aligned}$$

using the Binomial Formula and where $\lfloor x \rfloor$ is the largest integer less than or equal to x.
Since

$$e^{(\sigma+i\omega)x} = e^{\sigma x}(\cos(\omega x) + i \sin(\omega x)),$$

we conclude that, if

$$f(x) = x^k e^{\sigma x} \cos(\omega x), \quad g(x) = x^k e^{\sigma x} \sin(\omega x),$$

then

$$\mathscr{L}(f)(z) = \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{k}{2j} \frac{(-1)^j (z-\sigma)^{k-2j}\omega^{2j}}{((z-\sigma)^2+\omega^2)^{k+1}}$$

and

$$\mathscr{L}(g)(z) = \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{k}{2j+1} \frac{(-1)^{j+1} (z-\sigma)^{k-2j-1}\omega^{2j+1}}{((z-\sigma)^2+\omega^2)^{k+1}}.$$ •

Let us prove some useful results concerning the Laplace transform. We start with the property of linearity, which we already used in some of the examples above.

5.1.25 Proposition (Linearity of the Laplace transform) *The space $\mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ is a $\mathbb{C}$-vector space, and the Laplace transform is a linear map:*

$$\mathscr{L}(f_1 + f_2)(z) = \mathscr{L}(f_1)(z) + \mathscr{L}(f_2)(z), \quad \mathscr{L}(af)(z) = a\mathscr{L}(f)(z)$$

for every $f, f_1, f_2 \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ and $a \in \mathbb{C}$.

Proof We claim that, if $f_1, f_2 \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$, then $f_1 + f_2 \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ and

$$\sigma(f_1 + f_2) \leq \max\{\sigma(f_1), \sigma(f_2)\}.$$

Indeed, suppose that $a > \max\{\sigma(f_1), \sigma(f_2)\}$. Then there exists $M_1, M_2 \in \mathbb{R}_{>0}$ such that

$$|f_j(x)| \leq M_j \mathrm{e}^{ax}, \qquad x \in \mathbb{R}_{\geq 0},\ j \in \{1, 2\}.$$

Then

$$|f_1(x) + f_2(x)|\mathrm{e}^{-ax} \leq |f_1(x)|\mathrm{e}^{-ax} + |f_2(x)|\mathrm{e}^{-ax} < (M_1 + M_2)\mathrm{e}^{ax}.$$

Thus $\sigma(f_1 + f_2) \leq \max\{\sigma(f_1), \sigma(f_2)\}$.

Now, the linearity of $\mathscr{L}$ follows from linearity of integration. ■

Next we illustrate how the Laplace transform interacts with differentiation. We see in the next result the sometimes useful consequences of the Laplace transform being defined for functions defined on $\mathbb{R}_{\geq 0}$.

5.1.26 Proposition (Laplace transform and differentiation) *Let* $\mathrm{f} \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ *be continuously differentiable and suppose that* $\frac{\mathrm{df}}{\mathrm{dx}} \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$. *Then*

$$\mathscr{L}\left(\frac{\mathrm{df}}{\mathrm{dx}}\right)(\mathrm{z}) = \mathrm{z}\mathscr{L}(\mathrm{f})(\mathrm{z}) - \mathrm{f}(0).$$

Proof Using integration by parts,

$$\int_0^\infty \frac{\mathrm{d}f}{\mathrm{d}x}(x)\mathrm{e}^{-zx}\,\mathrm{d}x = f(x)\mathrm{e}^{-zx} + z\int_0^\infty f(x)\mathrm{e}^{-zx}\,\mathrm{d}x = z\mathscr{L}(f)(z) - f(0),$$

as desired. ■

The proposition can be applied recursively to obtain the following result.

5.1.27 Corollary (Laplace transform and higher-order derivatives) *Let* $\mathrm{f} \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ *be* k*-times continuously differentiable and suppose that* $\frac{\mathrm{d}^{\mathrm{j}}\mathrm{f}}{\mathrm{dx}^{\mathrm{j}}} \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ *for* $\mathrm{j} \in \{1, \ldots, \mathrm{k}\}$. *Then*

$$\mathscr{L}\left(\frac{\mathrm{d}^{\mathrm{k}}\mathrm{f}}{\mathrm{dx}^{\mathrm{k}}}\right)(\mathrm{z}) = \mathrm{z}^{\mathrm{k}}\mathscr{L}(\mathrm{f})(\mathrm{z}) - \mathrm{f}(0)\mathrm{z}^{\mathrm{k}-1} - \cdots - \frac{\mathrm{d}^{\mathrm{k}-2}\mathrm{f}}{\mathrm{dx}^{\mathrm{k}-2}}(0)\mathrm{z} - \frac{\mathrm{d}^{\mathrm{k}-1}\mathrm{f}}{\mathrm{dx}^{\mathrm{k}-1}}(0).$$

5.1.3.2 The inverse transform To construct the inverse of the Laplace transform, we first make the following connection with the CCFT. If $f \in \mathcal{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$ and if $z = \sigma + \mathrm{i}\omega \in \mathbb{C}_{>\sigma(f)}$, then, denoting $\nu = \frac{\omega}{2\pi}$,

$$\begin{aligned}\mathscr{L}(f)(\sigma + \mathrm{i}\omega) &= \int_0^\infty f(x)\mathrm{e}^{-(\sigma+\mathrm{i}\omega)x}\,\mathrm{d}x \\ &= \int_0^\infty f(x)\mathrm{e}^{-\sigma\xi}\mathrm{e}^{-2\pi\mathrm{i}\nu\xi}\,\mathrm{d}\xi \\ &= \mathscr{F}_{\mathrm{CC}}(f\mathsf{E}_{-\sigma})(\nu) = \mathscr{F}_{\mathrm{CC}}(f\mathsf{E}_{-\sigma})\left(\frac{\omega}{2\pi}\right),\end{aligned}$$

where, for $a \in \mathbb{C}$, $\mathsf{E}_a \colon \mathbb{R} \to \mathbb{C}$ is the function $\mathsf{E}_a(x) = \mathrm{e}^{ax}$. (Note that, when computing the CCFT of $f\mathsf{E}_{-\sigma}$, we extend f to be defined on all of $\mathbb{R}$ by taking $f(x) = 0$ for $x \in \mathbb{R}_{<0}$.) Now we can use our "inverse" for the CCFT from Section 5.1.2.2 to deduce

$$\begin{aligned} f(x)\mathrm{e}^{-\sigma x} &= \int_{-\infty}^{\infty} \mathscr{F}_{\mathrm{CC}}(f\mathsf{E}_{-\sigma})(\nu)\mathrm{e}^{2\pi \mathrm{i}\nu x}\,\mathrm{d}\nu \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathscr{F}_{\mathrm{CC}}(f\mathsf{E}_{-\sigma})\Big(\frac{\omega}{2\pi}\Big)\mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathscr{L}(f)(\sigma + \mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}\omega. \end{aligned}$$

Note that this derivation is subject to all of the limitations of those for the inverse of the CCFT, so this should be kept in mind, and we refer to Remark 5.1.17 for a discussion of some things that are not true and some things that are true.

With these computations and caveats, we can now suggest what we mean by the inverse of the Laplace transform.

5.1.28 "Definition" (Inverse of the Laplace transform) The *inverse* of the Laplace transform is the mapping that assigns to a map $F \colon \mathbb{C}_{>a} \to \mathbb{C}$ the function $\mathscr{L}^{-1}(F) \colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ defined by

$$\mathscr{L}^{-1}(F)(x) = \frac{\mathrm{e}^{\sigma x}}{2\pi}\int_{-\infty}^{\infty} F(\sigma + \mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}\omega$$

for $\sigma > a$. •

As with our "definitions" of $\mathscr{F}_{\mathrm{CD}}^{-1}$ and $\mathscr{F}_{\mathrm{CC}}^{-1}$, this "definition" of the inverse of the Laplace transform makes no sense at all. It does make more sense when $F = \mathscr{L}(f)$, in which case the formula $\mathscr{L}^{-1} \circ \mathscr{L}(f) = f$ is sometimes literally true, and other times still useful. We also point out that there is another potential problem that must be addressed, and that is the dependence of $\mathscr{L}^{-1}$ on $\sigma > \sigma(f)$. It is a fact—a non-obvious one—that $\mathscr{L}^{-1}$ *does not* depend on $\sigma > \sigma(f)$. The reason is connected with the fact that, for $f \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$, $\mathscr{L}(f)$ is not just any $\mathbb{C}$-valued function on $\mathbb{C}_{\sigma(f)}$, but is an holomorphic function. It is the particular properties of holomorphic functions that leads to the independence on $\sigma > \sigma(f)$.

In cases where the inverse transform can be actually computed, one seldom computes it using the explicit inversion formula. Rather, there are tables of Laplace transforms and inverse Laplace transforms, and one's first move should be for such a table. In Example 5.1.24 we give some important examples of forward Laplace transforms. Let us now produce some related examples for inverse Laplace transforms.

5.1.29 Examples (Inverse Laplace transform)

1. Let us consider the function
$$F\colon \mathbb{C}_{>0} \to \mathbb{C}$$
$$z \mapsto \frac{1}{z^k}$$
for $k \in \mathbb{Z}_{>0}$. By Example 5.1.24–1 and linearity of the Laplace transform, we have
$$\mathscr{L}^{-1}(F)(x) = \frac{x^{k-1}}{(k-1)!}.$$

2. Next we consider the function
$$F\colon \mathbb{C}_{>a} \to \mathbb{C}$$
$$z \mapsto \frac{1}{(z-a)^k}$$
for $k \in \mathbb{Z}_{>0}$ and $a \in \mathbb{R}$.[3] By Example 5.1.24–3 and linearity of the Laplace transform we have
$$\mathscr{L}^{-1}(F)(x) = \frac{x^{k-1}\mathrm{e}^{ax}}{(k-1)!}.$$

3. The next function we consider is
$$F\colon \mathbb{C}_{>a} \to \mathbb{C}$$
$$z \mapsto \frac{z}{(z-a)^k}$$
for $k \geq 2$ and $a \in \mathbb{R}$.[4] Here we take $G(z) = \frac{1}{(z-a)^k}$ and note from our previous example that $G(z) = \mathscr{L}(g)(z)$, where $g(x) = \frac{x^{k-1}\mathrm{e}^{ax}}{(k-1)!}$. Now, by Proposition 5.1.26, we have
$$\mathscr{L}\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)(z) = z\mathscr{L}(g)(z) - g(0) = F(z).$$
Thus $F = \mathscr{L}(f)$, where $f = \frac{\mathrm{d}g}{\mathrm{d}x}$. Thus, wrapping all this up,
$$\mathscr{L}^{-1}(F) = \frac{x^{k-2}\mathrm{e}^{ax}}{(k-2)!} + \frac{ax^{k-1}\mathrm{e}^{ax}}{(k-1)!}.$$

4. Next we consider
$$F(z) = \frac{1}{(z-\sigma)^2 + \omega^2}$$
for $\sigma \in \mathbb{R}$ and $\omega \in \mathbb{R}_{>0}$. As per Example 5.1.24–6, the inverse Laplace transform is
$$\mathscr{L}^{-1}(F)(x) = \frac{1}{\omega}\mathrm{e}^{\sigma x}\sin(\omega x).$$

[3] We can take $a \in \mathbb{C}$, in which case the domain of F would be $\mathbb{C}_{>\mathrm{Re}(a)}$.
[4] As previously, we can take $a \in \mathbb{C}$.

In similar fashion, if

$$G(z) = \frac{z}{(z-\sigma)^2 + \omega^2},$$

then

$$\mathscr{L}^{-1}(G)(x) = \mathrm{e}^{\sigma x}\cos(\omega x).$$

5. Now we generalise the preceding example, considering

$$F(z) = \frac{1}{((z-\sigma)^2 + \omega^2)^k},$$

for $k \geq 2$, $\sigma \in \mathbb{R}$, and $\omega \in \mathbb{R}_{>0}$. Here we note that

$$F(z) = \underbrace{\frac{1}{(z-(\sigma+\mathrm{i}\omega))^k}}_{F_+(z)} \underbrace{\frac{1}{(z-(\sigma-\mathrm{i}\omega))^k}}_{F_-(z)}.$$

Let $F = \mathscr{L}(f)$, $F_+ = \mathscr{L}(f_+)$, and $F_- = \mathscr{L}(f_-)$. By 2 above,

$$f_+(x) = \frac{x^{k-1}\mathrm{e}^{(\sigma+\mathrm{i}\omega)x}}{(k-1)!}, \quad f_-(x) = \frac{x^{k-1}\mathrm{e}^{(\sigma-\mathrm{i}\omega)x}}{(k-1)!}.$$

By Proposition 5.1.31, we have

$$\begin{aligned} f(x) = f_+ * f_-(x) &= \int_0^x f_+(y)f_-(x-y)\,\mathrm{d}y \\ &= \frac{1}{((k-1)!)^2}\int_0^x y^{k-1}(x-y)^{k-1}\mathrm{e}^{(\sigma+\mathrm{i}\omega)y}\mathrm{e}^{(\sigma-\mathrm{i}\omega)(x-y)}\,\mathrm{d}y \\ &= \frac{1}{((k-1)!)^2}\sum_{j=0}^{k-1}\binom{k-1}{j}x^{k-j-1}\mathrm{e}^{(\sigma-\mathrm{i}\omega)x}\int_0^x y^{k+j-1}\mathrm{e}^{2\mathrm{i}\omega y}\,\mathrm{d}y. \end{aligned}$$

A simple inductive (on k) computation gives

$$\int_0^x y^k \mathrm{e}^{ay}\,\mathrm{d}y = \mathrm{e}^{ax}\sum_{j=0}^{k}\frac{(-1)^{k-j}k!}{j!a^{k-j+1}}x^j - \frac{(-1)^k k!}{a^{k+1}}.$$

Thus

$$\begin{aligned} f(x) &= \frac{1}{((k-1)!)^2}\sum_{j=0}^{k-1}\binom{k-1}{j}x^{k-j-1}\mathrm{e}^{(\sigma-\mathrm{i}\omega)x}\int_0^x y^{k+j-1}\mathrm{e}^{2\mathrm{i}\omega y}\,\mathrm{d}y \\ &= \frac{1}{((k-1)!)^2}\sum_{j=0}^{k-1}\binom{k-1}{j}x^{k-j-1}\mathrm{e}^{(\sigma-\mathrm{i}\omega)x} \\ &\quad \times\left(\mathrm{e}^{2\mathrm{i}\omega x}\sum_{l=0}^{k+j-1}\frac{(-1)^{k+j-l-1}(k-j-1)!}{l!(2\mathrm{i}\omega)^{k+j-l}}x^l + \frac{(-1)^{k+j}(k-1)!}{(2\mathrm{i}\omega)^k}\right) \\ &= \end{aligned}$$

5.1.3.3 Convolution and the Laplace transform In this section we consider the relationship of convolution with products for the Laplace transform. The transformed variables in this case reside in $\mathbb{C}_{>a}$ for some $a \in \mathbb{R}$, and so transformed functions are functions from $\mathbb{C}_{>a}$ to $\mathbb{C}$. Thus, in this case, we have $F, G\colon \mathbb{C}_{>a} \to \mathbb{C}$ and so the product of F and G is the function

$$\begin{aligned} FG\colon \mathbb{C}_{>a} &\to \mathbb{C} \\ z &\mapsto F(z)G(z). \end{aligned}$$

What we want to know is, if $F = \mathscr{L}(f)$ and $G = \mathscr{L}(g)$, is there a function $h\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ for which $\mathscr{L}(h) = FG$?

To answer this question, we make the following definition.

5.1.30 Definition (Causal convolution) If $f, g \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$, the ***causal convolution*** of f and g is the function

$$\begin{aligned} f * g\colon \mathbb{R}_{\geq 0} &\to \mathbb{C} \\ x &\mapsto \int_0^x f(x-y)g(y)\,\mathrm{d}y. \end{aligned}$$

•

Note that, for $f, g\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$, we have $g(y) = 0$ for $y < 0$ and $f(x-y) = 0$ for $y > x$. Thus

$$\int_0^x f(x-y)g(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} f(x-y)g(y)\,\mathrm{d}y,$$

and so this establishes the relationship between convolution and causal convolution. Note, however, that we do not need to require that f and g be in $\mathrm{L}^1(\mathbb{R}_{\geq 0}; \mathbb{C})$ for the definition to make sense.

The operation of convolution is a interesting one, and has many properties that merit further exploration. For our purposes, we merely point out the following result.

5.1.31 Proposition (Laplace transform and convolution) *If* $\mathrm{f}, \mathrm{g} \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$, *then* $\mathrm{f} * \mathrm{g} \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})$, $\sigma(\mathrm{f} * \mathrm{g}) \leq \max\{\sigma(\mathrm{f}), \sigma(\mathrm{g})\}$, *and*

$$\mathscr{L}(\mathrm{f} * \mathrm{g})(\mathrm{z}) = \mathscr{L}(\mathrm{f})(\mathrm{z})\mathscr{L}(\mathrm{g})(\mathrm{z})$$

for $\mathrm{z} \in \mathbb{C}_{>\mathrm{a}}$, *and for any* $\mathrm{a} > \max\{\sigma(\mathrm{f}), \sigma(\mathrm{g}), \sigma(\mathrm{f} * \mathrm{g})\}$.

Proof Let $a > \max\{\sigma(f), \sigma(g)\}$ and let $b \in \mathbb{R}$ be such that

$$a > b > \max\{\sigma(f), \sigma(g)\}.$$

Let $M \in \mathbb{R}_{>0}$ be such that $|f(x)|, |g(x)| \leq M\mathrm{e}^{bx}$ for $x \in \mathbb{R}_{\geq 0}$. Then

$$|f * g(x)| \leq \int_0^x |f(x-y)g(y)|\,\mathrm{d}y \leq M^2 \int_0^x \mathrm{e}^{b(x-y)}\mathrm{e}^{by}\,\mathrm{d}y \leq M^2 x\mathrm{e}^{bx}.$$

Since $\lim_{x\to\infty} xe^{(b-a)x} = 0$, there exists $R \in \mathbb{R}_{>0}$ such that

$$xe^{(b-a)x} \leq 1, \qquad x \geq R.$$

Next let

$$C = \sup\{xe^{(b-a)x} \mid x \in [0, R]\}.$$

Then, for $x \in \mathbb{R}_{\geq 0}$, we have $xe^{(b-a)x} \leq \max\{1, C\}$. Thus

$$|f * g(x)| \leq M^2 xe^{bx} \leq M^2 \max\{1, C\}e^{ax}, \qquad x \in \mathbb{R}_{\geq 0}.$$

This shows that $\sigma(f * g) \leq \max\{\sigma(f), \sigma(g)\}$.

The remainder of the proof is a fairly straightforward application of Fubini's Theorem and the change of variables theorem:

$$\begin{aligned}
\mathscr{L}(f * g)(z) &= \int_0^\infty f * g(x)e^{-zx}\,dx = \int_0^\infty \left(\int_0^x f(x-y)g(y)\,dy\right)e^{-zx}\,dx \\
&= \int_0^\infty g(y)\left(\int_y^\infty f(x-y)e^{-zx}\,dx\right)dy \\
&= \int_0^\infty g(\sigma)\left(\int_0^\infty f(\tau)e^{-z(\sigma+\tau)}\,d\tau\right)d\sigma \\
&= \left(\int_0^\infty g(\sigma)e^{-z\sigma}\,d\sigma\right)\left(\int_{-\infty}^\infty f(\tau)e^{-z\tau}\,d\tau\right) \\
&= \mathscr{L}(f)(z)\mathscr{L}(g)(z),
\end{aligned}$$

as claimed. ■

5.1.3.4 Extension to higher-dimensions In this section we extend the definition of the Laplace transform to (1) functions with values in $\mathbb{C}^n$ and (2) functions with multiple independent variables.

First we extend the Laplace transform to functions with values in $\mathbb{C}^n$. To do so, we note that, if $\boldsymbol{f}\colon \mathbb{R}_{\geq 0} \to \mathbb{C}^n$, then we can write

$$\boldsymbol{f}(x) = (f_1(x), \ldots, f_n(x))$$

for functions $f_1, \ldots, f_n\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$. We then denote

$$\mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C}^n) = \left\{\boldsymbol{f}\colon \mathbb{R}_{\geq 0} \to \mathbb{C}^n \;\middle|\; f_1, \ldots, f_n \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C})\right\}.$$

We also denote $\sigma(\boldsymbol{f}) = \max\{\sigma(f_1), \ldots, \sigma(f_n)\}$. We can then make the following more or less obvious definition.

5.1.32 Definition (Laplace transform II) The *Laplace transform* of $\boldsymbol{f} \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{C}^n)$ is the function $\mathscr{L}(\boldsymbol{f})\colon \mathbb{C}_{>\sigma(\boldsymbol{f})} \to \mathbb{C}^n$ defined by

$$\mathscr{L}(\boldsymbol{f})(z) = (\mathscr{L}(f_1)(z), \dots, \mathscr{L}(f_n)(z)). \qquad \bullet$$

The inverse of the Laplace transform in this case is also made component-wise. Thus if $\boldsymbol{F}\colon \mathbb{C}_{>a} \to \mathbb{C}^n$, we denote

$$\mathscr{L}^{-1}(\boldsymbol{F})(x) = (\mathscr{L}^{-1}(F_1)(x), \dots, \mathscr{L}^{-1}(F_n)(x)).$$

Of course, all of the caveats we made in Section 5.1.2.2, which apply to the inversion of the Laplace transform, apply to the inverse in this case as well.

Next we consider the case when we have a function of multiple variables. To do this, if $f\colon \mathbb{R}^n_{\geq 0} \to \mathbb{C}$, then we denote

$$\int_{\mathbb{R}^n_{\geq 0}} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_0^\infty \cdots \int_0^\infty f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

We also denote

$$\mathscr{C}(\mathbb{R}^n_{\geq 0}; \mathbb{C}) = \left\{ f\colon \mathbb{R}^n_{\geq 0} \to \mathbb{C} \;\middle|\; |f(\boldsymbol{x})| \leq M \mathrm{e}^{a\|\boldsymbol{x}\|} \text{ for some } M \in \mathbb{R}_{>0},\ a \in \mathbb{R} \right\}.$$

For $f \in \mathscr{C}(\mathbb{R}^n_{\geq 0}; \mathbb{C})$, denote

$$\sigma(f) = \inf \left\{ a \in \mathbb{R} \mid |f(x)| \leq M \mathrm{e}^{a\|\boldsymbol{x}\|} \text{ for some } M \in \mathbb{R}_{>0} \right\}.$$

With this notation, we make the following definition.

5.1.33 Definition (Laplace transform III) The *Laplace transform* of $f \in \mathscr{C}(\mathbb{R}^n_{\geq 0}; \mathbb{C})$ is the function $\mathscr{L}(f)\colon \mathbb{C}^n_{>\sigma(f)} \to \mathbb{C}$ defined by

$$\mathscr{L}(f)(\boldsymbol{z}) = \int_{\mathbb{R}^n_{\geq 0}} f(\boldsymbol{x}) \mathrm{e}^{-\langle \boldsymbol{z}, \boldsymbol{x} \rangle_{\mathbb{R}^n}}\, \mathrm{d}\boldsymbol{x} \qquad \bullet$$

The inverse of the multivariable Laplace transform is then determined by analogy with the single-variable case. Thus, if $F\colon \mathbb{C}_{>a} \to \mathbb{C}$, we denote

$$\mathscr{L}^{-1}(F)(\boldsymbol{x}) = \int_{\mathbb{R}^n} F(\boldsymbol{\sigma} + \mathrm{i}\boldsymbol{\omega}) \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle_{\mathbb{R}^n}}\, \mathrm{d}\boldsymbol{\omega}.$$

Of course, care must be taken with interpreting this multiple integral, just as in the single-variable case.

Exercises

5.1.1 Determine whether the following functions $f\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ are of exponential class, and, if they are, determine $\sigma(f)$.

(a) $f(x) = \begin{cases} \frac{1}{x}, & x \in \mathbb{R}_{>0}, \\ 0, & x = 0. \end{cases}$

(b) $f(x) = \mathrm{e}^{x^2}$.

(c) $f(x) = \mathrm{e}^{-x^2}$.

(d) $f(x) = a_k x^k + \cdots + a_1 x + a_0$, $a_0, a_1, \ldots, a_k \in \mathbb{R}$.

5.1.2 Answer the following questions.

(a) Given that the Laplace transform for $f(x) = \cos(\omega x)$ is $\mathscr{L}(f)(z) = \frac{z}{z^2+\omega^2}$, use Proposition 5.1.26 to determine $\mathscr{L}(g)$, where $g(x) = \sin(\omega x)$.

(b) Given that the Laplace transform for $f(x) = \sin(\omega x)$ is $\mathscr{L}(f)(z) = \frac{\omega}{z^2+\omega^2}$, use Proposition 5.1.26 to determine $\mathscr{L}(g)$, where $g(x) = \cos(\omega x)$.

Section 5.2

Laplace transform methods for linear ordinary differential equations with constant coefficients

Laplace transforms can be used to study various sorts of differential equations, both partial and ordinary. In this section, we will stick to considering the application of Laplace transform techniques to the study of linear ordinary differential equations with constant coefficients. This can be thought of as the prototypical application of transform methods in the theory of differential equations and, moreover, is one of the more elementary applications of transform theory. Thus this section can be seen as having a twofold purpose: (1) to demonstrate the basic philosophy of transform analysis in the study of differential equations; (2) to develop fully an application of the Laplace transform to ordinary differential equations. To both ends, the emphasis will be on seeing how transforms can be helpful in understanding differential equations, rather than in solving differential equations (although we shall see that the latter is a part of the story).

We shall apply the Laplace transform to the full spectrum of linear ordinary differential equations with constant coefficients, homogeneous and inhomogeneous, and scalar and multiple dependent variables.

5.2.1 Scalar homogeneous equations

We begin our discussion with scalar linear homogeneous ordinary differential equations with constant coefficients, first considered in detail in Section 2.2.2. Thus, as in that section we are working with differential equations

$$F\colon \mathbb{T} \times \mathbb{R} \oplus \mathrm{L}^{\leq k}_{\mathrm{sym}}(\mathbb{R}; \mathbb{R}) \to \mathbb{R}$$

with right-hand side

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1 x^{(1)} - a_0 x \tag{5.4}$$

for $a_0, a_1, \dots, a_{k-1} \in \mathbb{R}$. Given Corollary 5.1.27, the Laplace transform is particularly well suited for working with ordinary differential equations with initial conditions. Thus we shall consider the initial value problem

$$\begin{aligned} \frac{\mathrm{d}^k \xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1 \frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0 \xi(t) = 0, \\ \xi(0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(0) = x_0^{(1)},\ \cdots,\ \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(0) = x_0^{(k-1)}. \end{aligned} \tag{5.5}$$

We shall now take the Laplace transform of this initial value problem. To do so, it is, of course, tacitly assumed that all members of $\mathrm{Sol}(F)$ are of exponential class.

This is true, however, since all members of Sol(F) are also pretty uninteresting functions, and so are of exponential class when restricted to the domain $\mathbb{R}_{\geq 0}$ as we saw in Example 5.1.24. Another way to think of taking the Laplace transform of the equation, were one to not know *a priori* that solutions were of exponential class, would be to go ahead and take the transform assuming this is so, and then see if the assumption is valid by seeing if the equation can be solved (or by some other means). In any case, the following result records what happens when we take the Laplace transform of the initial value problem.

5.2.1 Proposition (Laplace transform of scalar homogeneous equation) *If $\hat{\xi}$ is the Laplace transform of the initial value problem (5.5), then*

$$\hat{\xi}(z) = \frac{\sum_{j=0}^{k} \sum_{l=0}^{j-1} a_j z^l \xi^{(j-l-1)}(0)}{z^k + a_{k-1} z^{k-1} + \cdots + a_1 z + a_0},$$

with the convention that $a_k = 1$.

Proof By Corollary 5.1.27 we have

$$\mathscr{L}\left(\frac{d^j \xi}{dt^j}\right)(z) = z^j \hat{\xi}(z) - \sum_{l=0}^{j-1} z^l \xi^{(j-l-1)}(0), \qquad j \in \{0, 1, \ldots, k\}.$$

Therefore, with the stated convention that $a_k = 1$,

$$\mathscr{L}\left(\sum_{j=0}^{k} a_j \frac{d^j \xi}{dt^j}\right) = \sum_{j=0}^{k} a_j \left(z^j \hat{\xi}(z) - \sum_{l=0}^{j-1} z^l \xi^{(j-l-1)}(0)\right),$$

and solving this equation for $\hat{\xi}(z)$ gives the asserted conclusion. ■

To obtain the solution to the initial value problem in the time-domain, we should apply the inverse transform to the expression from the proposition. To do this, one could, in principle, apply the definition of the inverse Laplace transform, Definition 5.1.28. However, in cases where one can actually compute the inverse transform, it is not typically done in this way. Indeed, typically one "looks up" the answer. However, to do this requires a manipulation of the form of the expression from the proposition, and we outline this in the following procedure.

5.2.2 Procedure (Partial fraction expansion) While we shall apply the procedure to a $\mathbb{C}$-valued function of a complex variable (namely, the Laplace transform of something), the construction is best explained in algebraic terms, so we present it in this way. Algebraically, the problem we are considering is a way of expressing a rational function, i.e., a quotient $R_{N,D} = \frac{N}{D}$ of polynomials N and D, in a manner where the roots of D and their multiplicities are accounted for properly.

Given two polynomials $N, D \in \mathbb{R}[X]$ with real coefficients, with D monic, with no common roots, and with $\deg(N) < \deg(D)$, do the following.

1. Find all roots of D and their multiplicities. Let the real roots be denoted by $r_1, \dots, r_l$ and let $m(r_j)$, $j \in \{1, \dots, l\}$, be the multiplicity of the root r_j. Let the complex roots be denote by $\rho_j = \sigma_j + \mathrm{i}\omega_j$, $\sigma_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}_{>0}$, $j \in \{1, \dots, p\}$ (along with the complex conjugate roots $\sigma_j - \mathrm{i}\omega_j$) and let $m(\rho_j)$, $j \in \{1, \dots, p\}$, be the multiplicity of the root ρ_j.
2. Write
$$R_{N,D} = \sum_{j=1}^{l} \sum_{k=1}^{m(r_j)} \frac{a_{j,k}}{(X - r_j)^k} + \sum_{j=1}^{p} \sum_{k=1}^{m(\rho_j)} \frac{\alpha_{j,k} X + \beta_{j,k}}{((X - \sigma_j)^2 + \omega_j^2)^k} \tag{5.6}$$
for constants $a_{j,k} \in \mathbb{R}$, $j \in \{1, \dots, l\}$, $k \in \{1, \dots, m(r_j)\}$, and $\alpha_{j,k}, \beta_{j,k} \in \mathbb{R}$, $j \in \{1, \dots, p\}$, $k \in \{1, \dots, m(\rho_j)\}$, that are to be determined.
3. Express the right-hand side of the preceding expression in the form
$$\frac{P}{(X - r_1)^{m(r_1)} \cdots (X - r_l)^{m(r_l)} ((X - \sigma_1)^2 + \omega_1^2)^{m(\rho_1)} \cdots ((X - \sigma_p)^2 + \omega_p^2)^{m(\rho_p)}},$$
for some polynomial $P \in \mathbb{R}[X]$.
4. By matching coefficients of powers of the indeterminate X, arrive at a set of linear algebraic equations for the constants $a_{j,k} \in \mathbb{R}$, $j \in \{1, \dots, l\}$, $k \in \{1, \dots, m(r_j)\}$, and $\alpha_{j,k}, \beta_{j,k} \in \mathbb{R}$, $j \in \{1, \dots, p\}$, $k \in \{1, \dots, m(\rho_j)\}$. It is a fact that these linear algebraic equations have a unique solution.
5. The ***partial fraction expansion*** of $R_{N,D}$ is then the right-hand side of the expression (5.6) with the constants as computed in the previous step. •

The idea of a partial fraction expansion in practice is straightforward, albeit quite tedious.

5.2.3 Examples (Partial fraction expansion)

1. We take $N = 5X + 4$ and $D = X^2 + X - 2$ so that
$$R_{N,D} = \frac{5X + 4}{X^2 + X - 2}.$$
We determine the roots of D to be $r_1 = 1$ and $r_2 = -2$, with $m(r_1) = m(r_2) = 1$. We then write
$$\frac{5X + 4}{X^2 + X - 2} = \frac{a_{1,1}}{X - 1} + \frac{a_{2,1}}{X + 2} = \frac{(a_{1,1} + a_{2,1})X + 2a_{1,1} - a_{2,1}}{(X - 1)(X + 2)}.$$
Thus, matching coefficients of powers of X in the numerator, we must have
$$a_{1,1} + a_{2,1} = 5, \quad 2a_{1,1} - a_{2,1} = 4 \quad \Longrightarrow \quad a_{1,1} = 3, \ a_{2,1} = 2.$$
Thus the partial fraction expansion is
$$R_{N,D} = \frac{3}{X - 1} + \frac{2}{X + 2}.$$

2. We take $N = -3X^2 + 5X + 2$ and $D = X^3 - 3X^2 + X - 3$, so that

$$R = \frac{-3X^2 + 5X + 2}{X^3 - 3X^2 + X - 3}.$$

The roots of the denominator polynomial are $r_1 = 3$, $\rho_1 = \mathrm{i}$, and $\bar{\rho}_1 = -\mathrm{i}$. We then write

$$\begin{aligned}\frac{-3X^2 + 5X + 2}{X^3 - 3X^2 + X - 3} &= \frac{a_{1,1}}{X - 3} + \frac{\alpha_{1,1}X + \beta_{1,1}}{(X - 0)^2 + 1} \\ &= \frac{(a_{1,1} + \alpha_{1,1})X^2 + (\beta_{1,1} - 3\alpha_{1,1})X + a_{1,1} - 3\beta_{1,1}}{(X - 3)(X^2 + 1)}.\end{aligned}$$

Matching coefficients of powers of X in the numerator, we must have

$$\begin{aligned}a_{1,1} + \alpha_{1,1} = -3,\ \ \beta_{1,1} - 3\alpha_{1,1} = 5,\ \ a_{1,1} - 3\beta_{1,1} = 2 \\ \Longrightarrow \quad a_{1,1} = -1,\ \alpha_{1,1} = -2,\ \beta_{1,1} = -1.\end{aligned}$$

The partial fraction expansion is

$$R_{N,D} = -\frac{1}{X - 3} - \frac{2X + 1}{X^2 + 1}.$$

3. We take $N = 2X^2 + 1$ and $D = X^3 + 3X^2 + 3X + 1$ so that

$$R_{N,D} = \frac{2X^2 + 1}{X^3 + 3X^2 + 3X + 1}.$$

The denominator polynomial has a single root $r_1 = -1$ which has multiplicity $m(r_1) = 3$. We write

$$\begin{aligned}\frac{2X^2 + 1}{X^3 + 3X^2 + 3X + 1} &= \frac{a_{1,1}}{X + 1} + \frac{a_{1,2}}{(X + 1)^2} + \frac{a_{1,3}}{(X + 1)^3} \\ &= \frac{a_{1,1}X^2 + (2a_{1,1} + a_{1,2})X + a_{1,1} + a_{1,2} + a_{1,3}}{(X + 1)^3}.\end{aligned}$$

Thus, matching coefficients of powers of X in the numerator,

$$\begin{aligned}a_{1,1} = 2,\ \ 2a_{1,1} + a_{1,2} = 0,\ \ a_{1,1} + a_{1,2} + a_{1,3} = 1 \\ \Longrightarrow \quad a_{1,1} = 2,\ a_{1,2} = -4,\ a_{1,3} = 3.\end{aligned}$$

Thus the partial fraction expansion is

$$R_{N,D} = \frac{2}{X + 1} - \frac{4}{(X + 1)^2} + \frac{3}{(X + 1)^3}.$$

•

There are complex function methods for computing the coefficients in a partial fraction decomposition, but we shall not present this here, mainly because this method for solving initial value problems offers very little in terms of insight, and nothing over the methods we learned in Procedure 2.2.18 for solving scalar linear homogeneous ordinary differential equations with constant coefficients. So presenting multiple methods for computing partial fraction expansions seems a little silly.

Now let us see how one uses the partial fraction expansion to compute the inverse Laplace transform of the expression from Proposition 5.2.1. This is most easily done via examples.

5.2.4 Examples (Solving scalar homogeneous equations using the Laplace transform)

1. Consider the initial value problem

$$\ddot{\xi}(t) + \dot{\xi}(t) - 2\xi(t) = 0, \quad \xi(0) = 5,\ \dot{\xi}(0) = -1.$$

Taking the Laplace transform of the initial value problem, with $\hat{\xi}$ denoting the Laplace transform of ξ, gives

$$z^2\hat{\xi}(z) - z\xi(0) - \dot{\xi}(0) + z\hat{\xi}(z) - \xi(0) - 2\hat{\xi}(z) = 0 \quad \Longrightarrow \quad \hat{\xi}(z) = \frac{5z+4}{z^2+z-2}.$$

Borrowing our partial fraction expansion from Example 5.2.3–1 we have

$$\hat{\xi}(z) = \frac{3}{z-1} + \frac{2}{z+2}.$$

Thus, referring to Example 5.1.29–2,

$$\xi(t) = 3\mathrm{e}^{t} + 2\mathrm{e}^{-2t}.$$

2. Consider the initial value problem

$$\dddot{\xi}(t) - 3\ddot{\xi}(t) + \dot{\xi}(t) - 3\xi(t) = 0, \quad \xi(0) = -3,\ \dot{\xi}(0) = -4,\ \ddot{\xi}(0) = -7.$$

Taking the Laplace transform of the initial value problem gives

$$z^3\hat{\xi}(z) - z^2\xi(0) - z\dot{\xi}(0) - \ddot{\xi}(0) - 3z^2\hat{\xi}(z) + 3z\xi(0) + 3\dot{\xi}(0) + z\hat{\xi}(z) - \xi(0) - 3\hat{\xi}(z) = 0$$
$$\Longrightarrow \quad \hat{\xi}(z) = \frac{-3z^2+5z+2}{z^3-3z^2+z-3}.$$

Borrowing our partial fraction expansion from Example 5.2.3–1 we have

$$\hat{\xi}(z) = -\frac{1}{z-3} - \frac{2z+1}{z^1+1}.$$

Thus, referring to Example 5.1.29–2 and Example 5.1.29–4,

$$\xi(t) = -\mathrm{e}^{3t} - 2\cos(t) - \sin(t).$$

3. Consider the initial value problem

$$\dddot{\xi}(t) + 3\ddot{\xi}(t) + 3\dot{\xi}(t) + \xi(t) = 0, \quad \xi(0) = 2,\ \dot{\xi}(0) = -6,\ \ddot{\xi}(0) = 13.$$

Taking the Laplace transform of the initial value problem gives

$$z^3\hat{\xi}(z) - z^2\xi(0) - z\dot{\xi}(0) - \ddot{\xi}(0) + 3z^2\hat{\xi}(z) - 3z\xi(0) - 3\dot{\xi}(0) + 3z\hat{\xi}(z) - 3\xi(0) + \hat{\xi}(z) = 0$$
$$\implies \quad \hat{\xi}(z) = \frac{2z^2 + 1}{z^3 + 3z^2 + 3z + 1}.$$

Borrowing our partial fraction expansion from Example 5.2.3–1 we have

$$\hat{\xi}(z) = \frac{2}{z+1} - \frac{4}{(z+1)^2} + \frac{3}{(z+1)^3}.$$

Thus, referring to Example 5.1.29–2,

$$\xi(t) = 2\mathrm{e}^{-t} - 4t\mathrm{e}^{-t} + \frac{3}{2}t^2\mathrm{e}^{-t}. \qquad \bullet$$

The above business about partial fraction expansions gives a reader who likes doing algorithmic computations a venue to exercise this skill. However, it is not really the point of the Laplace transform. The really useful feature of the Laplace transform for linear differential equations, and not just those equations that are scalar and homogeneous, is that initial value problems are converted into algebraic expressions. The use of partial fraction expansions to determine the inverse Laplace transform of these algebraic expressions is something of a novelty act.

5.2.2 Scalar inhomogeneous equations

We next consider scalar linear inhomogeneous ordinary differential equations, first considered in Section 2.3.2. Thus we are working with scalar ordinary differential equations with right-hand sides given by

$$\widehat{F}(t, x, x^{(1)}, \dots, x^{(k-1)}) = -a_{k-1}x^{(k-1)} - \cdots - a_1x^{(1)} - a_0x + b(t) \tag{5.7}$$

for $a_0, a_1, \dots, a_{k-1} \in \mathbb{R}$ and $b\colon \mathbb{T} \to \mathbb{R}$. The initial value problem we consider is then

$$\frac{\mathrm{d}^k\xi(t)}{\mathrm{d}t^k}(t) + a_{k-1}\frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(t) + \cdots + a_1\frac{\mathrm{d}\xi}{\mathrm{d}t}(t) + a_0\xi(t) = b(t),$$
$$\xi(0) = x_0,\ \frac{\mathrm{d}\xi}{\mathrm{d}t}(0) = x_0^{(1)},\ \cdots,\ \frac{\mathrm{d}^{k-1}\xi}{\mathrm{d}t^{k-1}}(0) = x_0^{(k-1)}. \tag{5.8}$$

As with inhomogeneous equations above, we take the Laplace transform of this equation. However, unlike in the homogeneous case, here taking the transform is not generally valid; indeed, it is valid if and only if $b \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{R})$.

5.2.5 Proposition (Laplace transform of scalar inhomogeneous equation) *Consider the scalar ordinary differential equation with right-hand side (5.7), and suppose that* b *is continuous and satisfies* $b \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{R})$. *If* $\hat{b}$ *is the Laplace transform of* b *and if* $\hat{\xi}$ *is the Laplace transform of the initial value problem (5.8), then*

$$\hat{\xi}(z) = \frac{\sum_{j=0}^{k} \sum_{l=0}^{j-1} a_j z^l \xi^{(j-l-1)}(0) + \hat{b}(z)}{z^k + a_{k-1} z^{k-1} + \cdots + a_1 z + a_0},$$

with the convention that $a_k = 1$.

Proof This follows immediately from the computations of Proposition 5.2.1. ■

There are two ways in which the proposition has value. One is theoretical and one is that it provides another tedious algorithmic procedure—augmenting the "method of undetermined coefficients"—for computing solutions when the inhomogeneous term is an also pretty uninteresting function. Let us consider these in turn.

First let us give an interpretation of Proposition 5.2.5 in terms of the Green's function from Section 2.3.1.3.

5.2.6 Proposition (Laplace transforms and the Green's function) *Consider the scalar linear homogeneous ordinary differential equation* F *with right-hand side (5.4). Then the following statements hold:*

(i) the Laplace transform of the Green's function $G_{F,0} \colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ *is given by* $G_{F,0}(t, \tau) = H_F(t - \tau)$ *where*

$$\mathscr{L}(H_F)(z) = \frac{1}{z^k + a_{k-1} z^{k-1} + \cdots + a_z z + a_0};$$

(ii) if $b \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ *is continuous, then the solution to the initial value problem (5.8) is given by* $\xi(t) = \xi_h(t) + H_F * b(t)$, *where* ξ_h *satisfies the homogeneous initial value problem*

$$\frac{d^k \xi_h(t)}{dt^k}(t) + a_{k-1} \frac{d^{k-1} \xi_h}{dt^{k-1}}(t) + \cdots + a_1 \frac{d\xi_h}{dt}(t) + a_0 \xi_h(t) = 0,$$
$$\xi_h(0) = x_0, \ \frac{d\xi_h}{dt}(0) = x_0^{(1)}, \ \cdots, \ \frac{d^{k-1}\xi_h}{dt^{k-1}}(0) = x_0^{(k-1)}.$$

Proof (i) According to Remark 2.3.11, $G_{F,0}(t, \tau) = H_F(t - \tau)$, where H_F satisfies the initial value problem

$$\frac{d^k H_F}{dt^k}(t) + a_{k-1} \frac{d^{k-1} H_F}{dt^{k-1}}(t) + \cdots + a_1 \frac{dH_F}{dt}(t) + a_0 H_F(t) = 0,$$
$$H_F(0) = 0, \ \frac{dH_K}{dt}(0) = 0, \ldots, \ \frac{d^{k-2} H_F}{dt^{k-2}}(0) = 0, \ \frac{d^{k-1} H_F}{dt^{k-1}}(0) = 1.$$

Therefore, according to Proposition 5.2.1,

$$\mathscr{L}(H_F)(z) = \frac{1}{z^k + a_{k-1}z^{k-1} + \cdots + a_z z + a_0},$$

as claimed.

(ii) This follows from Remark 2.3.11 and Exercise 2.3.2. However, let us also see how it follows from Proposition 5.2.5 when $b \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{R})$. Indeed, from Proposition 5.2.5 and part (i), we have

$$\hat{\xi}(z) = \hat{\xi}_{\mathrm{h}}(z) + \hat{H}_F(z)\hat{b}(z).$$

Now this part of the result follows from Proposition 5.1.31. ■

The preceding result provides one of the most compelling reasons to work with the Laplace transform, and additionally adds insight into the meaning of the Green's function that we saw in generality in Section 2.3.1.3.

Next let us turn to a less interesting but somehow more concrete application of the Laplace transform in the study of scalar linear inhomogeneous ordinary differential equations. Specifically, we consider such an equation F with right-hand side (5.7), and where b is an also pretty uninteresting function. In this case, as we see from Example 5.1.24, the Laplace transform $\hat{b}$ of b will be a rational function of the complex variable z whose numerator polynomial has degree strictly less than that of the denominator polynomial. Therefore, as per Proposition 5.2.5, the Laplace transform $\hat{\xi}$ of the solution ξ of the initial value problem (5.8) will itself be such a rational function of z. Thus we can perform a partial fraction expansion of $\hat{\xi}$ as per Procedure 5.2.2, and then perform the inversion of the Laplace transform as per Example 5.2.4 to obtain the solution. This is not something to be belaboured—not least because we already have the often easier "method of undetermined coefficients" for such situations–and we content ourselves with an illustration via a example.

5.2.7 Example (Solving scalar inhomogeneous equations using the Laplace transform) We consider the initial value problem

$$\ddot{\xi}(t) + \omega^2 \xi(t) = \sin(\omega t), \quad \xi(0) = x_0, \ \dot{\xi}(0) = x_0^{(1)},$$

for $\omega \in \mathbb{R}_{>0}$. Using Example 5.1.24–4 we compute the Laplace transform of this initial value problem:

$$z^2\hat{\xi}(z) - zx_0 - x_0^{(1)} + \omega^2\hat{\xi}(z) = \frac{\omega}{z^2 + \omega^2}$$
$$\implies \hat{\xi}(z) = \frac{\omega}{(z^2 + \omega^2)^2} + \frac{zx_0 + x_0^{(1)}}{z^2 + \omega^2}.$$

Using Example 5.1.29–4 and Example 5.1.29–5 we have

$$\xi(t) = x_0 \cos(\omega t) + \frac{x_0^{(1)}}{\omega}\sin(\omega t) - \frac{t}{2\omega}\cos(\omega t) + \frac{1}{2\omega^2}\sin(\omega t).$$

•

5.2.3 Systems of homogeneous equations

Now we turn to studying systems of equations using the Laplace transform, starting with the homogeneous case. As we did in Section 3.2, we shall work with systems whose state space is a finite-dimensional $\mathbb{R}$-vector space V. We should indicate what we mean by the Laplace transform of a function taking values in such a space.

5.2.8 Definition (Laplace transform for vector space-valued functions) Let V be an n-dimensional $\mathbb{R}$-vector space and let $\{e_1, \dots, e_n\}$ be a basis for V. For a function $\xi \colon \mathbb{R}_{\geq 0} \to \mathsf{V}$, write

$$\xi(t) = \xi_1(t)e_1 + \cdots + \xi_n(t)e_n$$

for $\xi_j \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, $j \in \{1, \dots, n\}$.

(i) Denote

$$\mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V}) = \{\xi \colon \mathbb{R}_{\geq 0} \to \mathsf{V} \mid \ \xi_1, \dots, \xi_n \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathbb{R})\}.$$

(ii) For $\xi \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$, denote

$$\sigma(f) = \max\{\sigma(\xi_1), \dots, \sigma(\xi_n)\}.$$

(iii) A function in $\mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$ will be said to be of ***exponential class***.

(iv) For $\xi \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$, the ***Laplace transform*** of ξ is

$$\begin{aligned} \mathscr{L}(\xi) \colon \mathbb{C}_{>\sigma(\xi)} &\to \mathsf{V}^{\mathbb{C}} \\ z &\mapsto \mathscr{L}(\xi_1)(z)e_1 + \cdots + \mathscr{L}(\xi_n)(z)e_n. \end{aligned}$$

•

Of course, one must verify that the preceding definitions are independent of the choice of basis, and we leave this to the reader as Exercise 5.2.1.

Now we proceed with the principal constructions. We consider a system of linear ordinary differential equations F with constant coefficients in an n-dimensional $\mathbb{R}$-vector space V, and with right-hand side

$$\begin{aligned} \widehat{F} \colon \mathbb{R}_{\geq 0} \times \mathsf{V} &\to \mathsf{V} \\ x &\mapsto A(x) \end{aligned}$$

for $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$. The associated initial value problem we study is then

$$\dot{\xi}(t) = A(\xi(t)), \quad \xi(0) = x_0. \tag{5.9}$$

Let us take the Laplace transform of this initial value problem.

5.2.9 Proposition (Laplace transform of system of homogeneous equations) *If $\hat{\xi}$ is the Laplace transform of the initial value problem (5.9), then*

$$\hat{\xi}(\mathrm{z}) = (\mathrm{z}\,\mathrm{id}_{\mathsf{V}} - \mathrm{A})^{-1}\mathrm{x}_0,$$

and $\hat{\xi}$ is defined on

$$\{\mathrm{z} \in \mathbb{C} \mid \ \mathrm{Re}(\mathrm{z}) > \mathrm{Re}(\lambda) \text{ for all } \lambda \in \mathrm{spec}(\mathrm{A})\}.$$

Proof This is a direct computation using Proposition 5.1.26:

$$z\hat{\xi}(z) - \xi(0) = A\hat{\xi}(z),$$

from which the result follows immediately after noting that $z\,\mathrm{id}_{\mathsf{V}} - A$ is invertible if the real part of z exceeds the real part of any eigenvalue of A. ■

As with scalar equations, the application of the Laplace transform permits a solution for systems of linear homogeneous equations with constant coefficients using just algebraic computations in the transformed variables. In order to understand the inverse $(z\,\mathrm{id}_{\mathsf{V}} - A)^{-1}$, let us think about how one may compute this inverse. We shall suppose that we have a basis $\{e_1, \dots, e_n\}$ for V and let $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ be the matrix representative for A. Then the matrix representative for $(z\,\mathrm{id}_{\mathsf{V}} - A)^{-1}$ is $(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}$. For $\boldsymbol{B} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$, let us denote by $\mathrm{Cof}(\boldsymbol{B})$ the $n \times n$-matrix whose (j,k)th entry is $(-1)^{j+k} \det \hat{\boldsymbol{B}}(j,k)$, where $\hat{\boldsymbol{B}}(j,k)$ is the $(n-1) \times (n-1)$-matrix obtained by deleting the jth row and kth column from $\boldsymbol{B}$. Then, by ***missing stuff***,

$$\mathrm{Cof}(\boldsymbol{B})^T\boldsymbol{B} = \boldsymbol{B}\,\mathrm{Cof}(\boldsymbol{B})^T = (\det \boldsymbol{B})\boldsymbol{I}_n.$$

Therefore,

$$(z\boldsymbol{I}_n - \boldsymbol{A})^{-1} = \frac{(z\boldsymbol{I}_n - \boldsymbol{A})^T}{\det(z\boldsymbol{I}_n - \boldsymbol{A})}.$$

Note that the entries of $\mathrm{Cof}(z\boldsymbol{I} - \boldsymbol{A})$ are determinants of $(n-1)\times(n-1)$-matrices whose entries are polynomials of degree at most 1 in z. Thus the entries of $\mathrm{Cof}(z\boldsymbol{I}_n - \boldsymbol{A})$ are polynomials of degree at most $n-1$. Thus, since $\det(z\boldsymbol{I}_n - \boldsymbol{A})$ is a monic polynomial of degree n in z, the entries of $(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}$ are rational functions in z whose numerator polynomial has degree strictly less than that of the denominator polynomial. Therefore, the inverse Laplace transform of $(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}$ can be computed by performing a partial fraction expansion on each of its entries, and then applying the inverse Laplace transforms of Example 5.1.29.

However, the inverse Laplace transform of $(z\,\mathrm{id}_{\mathsf{V}} - A)^{-1}$ is known to us already.

5.2.10 Proposition (Laplace transform of operator exponential) *For an* n*-dimensional $\mathbb{R}$-vector space V and for $\mathrm{A} \in \mathrm{L}(\mathsf{V};\mathsf{V})$, denote*

$$\begin{aligned} \exp_{\mathrm{A}} \colon \mathbb{R}_{\geq 0} &\to \mathrm{L}(\mathsf{V};\mathsf{V}) \\ \mathrm{t} &\mapsto \mathrm{e}^{\mathrm{At}}. \end{aligned}$$

Then $\mathscr{L}(\exp_{\mathrm{A}})(\mathrm{z}) = (\mathrm{z}\,\mathrm{id}_{\mathsf{V}} - \mathrm{A})^{-1}$.

Proof By Theorem 3.2.9(i) and since $\exp_A(t) = \Phi_A(t, 0)$, we note that $\exp_A$ satisfies the initial value problem

$$\frac{\mathrm{d}\exp_A}{\mathrm{d}t}(t) = A \circ \exp_A(t), \quad \exp_A(0) = \mathrm{id}_V .$$

Taking the Laplace transform of this initial value problem gives

$$z\,\widehat{\exp}_A(z) - \mathrm{id}_V = A \circ \widehat{\exp}_A(z) \quad \Longrightarrow \quad \widehat{\exp}_A(z) = (z\,\mathrm{id}_V - A)^{-1},$$

as claimed. ■

Let's illustrate this in a simple example.

5.2.11 Example (Operator exponential via the Laplace transform) We consider the linear map $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^2; \mathbb{R}^2)$ considered in Example 3.2.49:

$$\boldsymbol{A} = \begin{bmatrix} -7 & 4 \\ -6 & 3 \end{bmatrix}.$$

We compute

$$(z\boldsymbol{I}_2 - \boldsymbol{A})^{-1} = \begin{bmatrix} \frac{z-3}{z^2+4z+3} & \frac{4}{z^2+4z+3} \\ -\frac{6}{z^2+4z+3} & \frac{z+7}{z^2+4z+3} \end{bmatrix}.$$

We then use partial fraction expansions:

$$\begin{aligned} \frac{z-3}{z^2+4z+3} &= -\frac{2}{z+1} + \frac{3}{z+3}, \\ \frac{4}{z^2+4z+3} &= \frac{2}{z+1} - \frac{2}{z+3}, \\ -\frac{6}{z^2+4z+3} &= -\frac{3}{z+1} + \frac{3}{z+3}, \\ \frac{z+7}{z^2+4z+3} &= \frac{3}{z+1} - \frac{2}{z+3}. \end{aligned}$$

Using Example 5.1.29–2, we apply the inverse transform to get

$$\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} 3\mathrm{e}^{-3t} - 2\mathrm{e}^{-t} & -2\mathrm{e}^{-3t} + 2\mathrm{e}^{-t} \\ 3\mathrm{e}^{-3t} - 3\mathrm{e}^{-t} & -2\mathrm{e}^{-3t} + 3\mathrm{e}^{-t} \end{bmatrix},$$

just as in Example 3.2.49. •

It is a matter of taste whether one thinks that using Laplace transforms to compute the operator exponential is preferable to Procedure 3.2.48. It is, however, not such an important matter to resolve in favour of one method or the other; actually computing the operator exponential is seldom of interest *per se*. What is certainly true is that with Laplace transforms one loses the insight offered by invariant subspaces in Procedure 3.2.48. The benefits of the Laplace transform in this context arises in system theory, where complex function techniques offer some genuine insights. However, these topics are out of our scope here.

5.2.4 Systems of inhomogeneous equations

Next we consider systems of homogeneous equations. Thus we have an ordinary differential equation with state space V and with right-hand side

$$\begin{aligned} \widehat{F}\colon \mathbb{R}_{\geq 0} \times \mathsf{V} &\to \mathsf{V} \\ x &\mapsto A(x) + b(t), \end{aligned} \tag{5.10}$$

for $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ and for $b\colon \mathbb{R}_{\geq 0} \to \mathsf{V}$. The associated initial value problem we consider is

$$\dot{\xi}(t) = A(\xi(t)) + b(t), \quad \xi(0) = x_0. \tag{5.11}$$

We can, of course, easily take the Laplace transform of this initial value problem to get the following.

5.2.12 Proposition (Laplace transform of system of inhomogeneous equations) *Consider the system of scalar ordinary differential equations with right-hand side (5.10), and suppose that* b *is continuous and satisfies* $\mathsf{b} \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$. *If* $\hat{\mathsf{b}}$ *is the Laplace transform of* b *and if* $\hat{\xi}$ *is the Laplace transform of the initial value problem (5.11), then*

$$\hat{\xi}(\mathrm{z}) = (\mathrm{z}\,\mathrm{id}_{\mathsf{V}} - \mathrm{A})^{-1}(\mathrm{x}_0 + \hat{\mathrm{b}}(\mathrm{z})).$$

Proof The proof is an easy adaptation of that of Proposition 5.2.12. ■

As was the case with our discussion of scalar inhomogeneous equations in Section 5.2.2, the preceding result can be interpreted in two ways, one having theoretical value and the other as a means of computing solutions. We shall explore both.

The first result makes a connection with the formula given in Corollary 3.3.3 for solutions to systems of linear inhomogeneous equations, in the general setting of time-varying systems.

5.2.13 Proposition (Laplace transforms and convolutions for solutions of linear inhomogeneous equations) *Consider the system of scalar ordinary differential equations with right-hand side (5.10), and suppose that* b *is continuous. Then the solution to the initial value problem (5.11) is*

$$\xi(\mathrm{t}) = \mathrm{e}^{\mathrm{At}}(\mathrm{x}_0) + \exp_{\mathrm{A}} * \mathrm{b}(\mathrm{t}).$$

Proof This follows immediately from Corollary 3.3.3, after understanding that

$$\exp_A * b(t) = \int_0^t \mathrm{e}^{A(t-\tau)}(b(\tau))\,\mathrm{d}\tau.$$

However, here we shall give a proof using Laplace transforms, valid when $b \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$.

From Proposition 5.2.12 we have

$$\hat{\xi}(z) = (z\,\mathrm{id}_{\mathsf{V}} - A)^{-1}(x_0) + (z\,\mathrm{id}_{\mathsf{V}} - A)^{-1}\hat{b}(z).$$

By Proposition 5.2.10 we have

$$(z\,\mathrm{id}_{\mathsf{V}} - A)^{-1} = \mathscr{L}(\exp_A).$$

For $x \in \mathsf{V}$, let us denote

$$\begin{aligned} \mathrm{ev}_x \colon \mathrm{L}(\mathsf{V};\mathsf{V}) &\to \mathsf{V} \\ L &\mapsto L(x). \end{aligned}$$

We then have, noting that ev_{x_0} is a linear map,

$$\mathscr{L}(\mathrm{ev}_{x_0} \circ \exp_A)(z) = \mathrm{ev}_{x_0} \circ \mathscr{L}(\exp_A)(z) = (z\,\mathrm{id}_{\mathsf{V}} - A)(x_0).$$

Also, by Proposition 5.1.31,

$$\mathscr{L}(\exp_A * b)(z) = \mathscr{L}(\exp_A)(z)\hat{b}(z) = (z\,\mathrm{id}_{\mathsf{V}} - A)\hat{b}(z).$$

Therefore,

$$\hat{\xi}(z) = \mathrm{ev}_{x_0} \circ \mathscr{L}(\exp_A) + \mathscr{L}(\exp_A * b)(z).$$

Taking the inverse Laplace transform gives

$$\xi(t) = \mathrm{ev}_{x_0} \circ \mathrm{e}^{At} + \exp_A * b(t) = \mathrm{e}^{At}(x_0) + \exp_A * b(t),$$

as claimed. ■

Finally, in the base when b is an also pretty interesting function (meaning that, in a basis for V, the components of b are also pretty uninteresting functions), we can use Proposition 5.2.12, and partial fraction expansions, to compute solutions. We only validate this by a simple example since, in reality, this is not something one ever does.

5.2.14 Example (Solving systems of inhomogeneous equations using the Laplace transform) We take $\mathsf{V} = \mathbb{R}^2$ and

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & 0 \end{bmatrix}, \qquad \boldsymbol{b}(t) = \begin{bmatrix} 0 \\ \sin(\omega t) \end{bmatrix}.$$

We then calculate

$$(z\boldsymbol{I}_2 - \boldsymbol{A})^{-1} = \begin{bmatrix} \frac{z}{z^2+\omega^2} & \frac{1}{z^2+\omega^2} \\ -\frac{\omega^2}{z^2+\omega^2} & \frac{z}{z^2+\omega^2} \end{bmatrix}, \qquad \hat{b}(z) = \begin{bmatrix} 0 \\ \frac{\omega}{z^2+\omega^2} \end{bmatrix}.$$

Thus, by Proposition 5.2.12,

$$\begin{aligned}\hat{\xi}(z) &= \begin{bmatrix} \frac{z}{z^2+\omega^2} & \frac{1}{z^2+\omega^2} \\ -\frac{\omega^2}{z^2+\omega^2} & \frac{z}{z^2+\omega^2} \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} + \begin{bmatrix} \frac{z}{z^2+\omega^2} & \frac{1}{z^2+\omega^2} \\ -\frac{\omega^2}{z^2+\omega^2} & \frac{z}{z^2+\omega^2} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{\omega}{z^2+\omega^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\omega}{(z^2+\omega^2)^2} + \frac{x_{01}z+x_{02}}{z^2+\omega^2} \\ \frac{\omega z}{(z^2+\omega^2)^2} + \frac{x_{02}z-x_{01}\omega^2}{z^2+\omega^2} \end{bmatrix}.\end{aligned}$$

The last line was arrived at by performing the matrix multiplication, then performing a partial fraction expansion of the entries of the resulting vector. This, then, is a bit of effort that we do not fully illustrate. In any case, one can apply the conclusions of Example 5.1.29–4 and Example 5.1.29–5 to arrive at

$$\xi(t) = \begin{bmatrix} \frac{1}{2\omega^2}\sin(\omega t) - \frac{t}{2\omega}\cos(\omega t) + x_{01}\cos(\omega t) + \frac{x_{02}}{\omega}\sin(\omega t) \\ \frac{t}{2}\sin(\omega t) + -\omega x_{01}\sin(\omega t) + x_{02}\cos(\omega t) \end{bmatrix}.$$

We encourage the reader to understand the relationship between this answer and the one from Example 5.2.7. •

As with systems of homogeneous equations, the use of the Laplace transform to solve inhomogeneous equations does not have a lot to recommend it from a computational point of view. The advantages it has come more from exploiting the algebraic structure of the differential equation as a function of the transformed independent variable z.

Exercises

5.2.1 Let V be a finite-dimensional $\mathbb{R}$-vector space. Answer the following questions regarding Definition 5.2.8.

(a) Show that the definition of $\mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$ is independent of choice of basis.

(b) For $\xi \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$, show that the definition of $\sigma(\xi)$ is independent of choice of basis.

(c) For $\xi \in \mathscr{C}(\mathbb{R}_{\geq 0}; \mathsf{V})$, show that the definition of $\mathscr{L}(\xi)$ is independent of choice of basis.

Hint: *Use the change of basis formula (1.24).*

5.2.2 Determine the Laplace transform of the solution of the following initial value problems:

(a) $\dot{\xi}(t) + 3\xi(t) = 0$, $\xi(0) = 4$;

(b) $\ddot{\xi}(t) - 4\dot{\xi}(t) + 4\xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 1$;

(c) $\ddot{\xi}(t) - 4\dot{\xi}(t) - 4\xi(t) = 0$, $\xi(0) = 1$, $\dot{\xi}(0) = 1$;

(d) $\dddot{\xi}(t) - 7\ddot{\xi}(t) + 15\dot{\xi}(t) - 9\xi(t) = 0$, $\xi(0) = 1$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 1$;

(e) $\dddot{\xi}(t) + 3\ddot{\xi}(t) + 4\dot{\xi}(t) + 2\xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 2$;

(f) $\ddddot{\xi}(t) + \dddot{\xi}(t) + \ddot{\xi}(t) + \dot{\xi}(t) + \xi(t) = 0$, $\xi(0) = 0$, $\dot{\xi}(0) = 0$, $\ddot{\xi}(0) = 0$, $\dddot{\xi}(0) = 0$.

NB. These are the same initial value problems you worked out in Exercise 2.2.10.

5.2.3 Using partial fraction expansion, determine the solution to the initial value problems from Exercise 5.2.2.

5.2.4 Determine the Laplace transform of the solution for the following scalar linear inhomogeneous differential equations F with the stated initial conditions:

(a) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 2x^{(1)} + x - 3\mathrm{e}^t$, and $\xi(0) = 1$, $\dot{\xi}(0) = 1$;

(b) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 5x^{(1)} + 6x - 2\mathrm{e}^{3t} - \cos(t)$, and $\xi(0) = 0$, $\dot{\xi}(0) = 1$;

(c) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} - 2x^{(1)} + 5x - t\mathrm{e}^t \sin(2t)$, and $\xi(0) = 1$, $\dot{\xi}(0) = 0$;

(d) $F(t, x, x^{(1)}, x^{(2)}) = x^{(2)} + 4x - t\cos(2t) + \sin(2t)$, and $\xi(0) = 2$, $\dot{\xi}(0) = 1$;

(e) $F(t, x^{(1)}, x^{(2)}, x^{(3)}) = x^{(3)} - x - t\mathrm{e}^t$, and $\xi(0) = 1$, $\dot{\xi}(0) = 1$, $\ddot{\xi}(0) = 1$;

(f) $F(t, x, x^{(1)}, \dots, x^{(4)}) = x^{(4)} + 4x^{(2)} + 4x - \cos(2t) - \sin(2t)$, and $\xi(0) = 0$, $\dot{\xi}(0) = 0$, $\ddot{\xi}(t) = 0$, $\dddot{\xi}(t) = 0$.

NB. These are the same initial value problems you worked out in Exercise 2.3.5.

5.2.5 Using partial fraction expansion, determine the solution to the initial value problems from Exercise 5.2.4.

5.2.6 Determine the Laplace transform of the solution of the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \boldsymbol{A}\boldsymbol{\xi}(t), \quad \boldsymbol{\xi}(0) = \boldsymbol{x}_0,$$

for the following choices of $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$:

(a) $\boldsymbol{A} = \begin{bmatrix} 2 & -5 \\ 0 & 3 \end{bmatrix}$, $\boldsymbol{x}_0 = (0, 1)$;

(b) $\boldsymbol{A} = \begin{bmatrix} -1 & -2 \\ 1 & -3 \end{bmatrix}$, $\boldsymbol{x}_0 = (2, -3)$;

(c) $\boldsymbol{A} = \begin{bmatrix} 4 & -1 \\ 4 & 0 \end{bmatrix}$, $\boldsymbol{x}_0 = (1, 1)$;

(d) $\boldsymbol{A} = \begin{bmatrix} 5 & 0 & -6 \\ 0 & 2 & 0 \\ 3 & 0 & -4 \end{bmatrix}$, $\boldsymbol{x}_0 = (-3, -1, 0)$;

(e) $\boldsymbol{A} = \begin{bmatrix} 5 & 0 & -6 \\ 1 & 2 & -1 \\ 3 & 0 & -4 \end{bmatrix}$, $\boldsymbol{x}_0 = (1, 0, 1)$;

(f) $\boldsymbol{A} = \begin{bmatrix} 4 & 2 & -4 \\ 2 & 0 & -4 \\ 2 & 2 & -2 \end{bmatrix}$, $\boldsymbol{x}_0 = (4, 1, 2)$;

(g) $\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 3 & -1 & 3 \\ 0 & 1 & 2 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$, $\boldsymbol{x}_0 = (1, -1, 0, 1)$;

(h) $\boldsymbol{A} = \begin{bmatrix} -7 & 0 & 0 & -4 \\ -13 & -2 & -1 & -8 \\ 6 & 1 & 0 & 4 \\ 15 & 1 & 0 & 9 \end{bmatrix}$, $\boldsymbol{x}_0 = (-1, -1, 3, -2)$;

(i) $\boldsymbol{A} = \begin{bmatrix} 1 & 4 & -2 & 0 & 9 \\ 0 & -2 & 1 & 2 & -6 \\ -2 & 4 & -1 & 3 & 0 \\ -9 & 4 & 1 & 0 & 2 \\ 4 & 0 & 3 & -1 & 3 \end{bmatrix}$, $\boldsymbol{x}_0 = (0, 0, 0, 0, 0)$.

NB. These are the same initial value problems you worked out in Exercise 3.2.15.

5.2.7 Using partial fraction expansion, compute $\mathrm{e}^{\boldsymbol{A}t}$ for the linear transformations $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ from Exercise 5.2.6.

5.2.8 Determine the Laplace transform of the solution of the initial value problem

$$\dot{\boldsymbol{\xi}}(t) = \boldsymbol{A}\boldsymbol{\xi}(t) + \boldsymbol{b}(t), \quad \boldsymbol{\xi}(0) = \boldsymbol{0},$$

for the choices of $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ from Exercise 5.2.6 and for the following $\boldsymbol{b}$:

(a) $\boldsymbol{b}(t) = (0, 1)$;

(b) $\boldsymbol{b}(t) = (\cos(t), 0)$;

(c) $\boldsymbol{b}(t) = (\mathrm{e}^{2t}, 0)$;

(d) $\boldsymbol{b}(t) = (\sin(t), 0, 1)$;

(e) $\boldsymbol{b}(t) = (0, \mathrm{e}^{-t}, 0)$;

(f) $\boldsymbol{b}(t) = (\sin(2t), 0, 1)$;

(g) $\boldsymbol{b}(t) = (1, 0, 0, 1)$;

(h) $\boldsymbol{b}(t) = (\sin(t), 0, 0, \cos(t))$;

(i) $\boldsymbol{b}(t) = (0, 0, 0, 0, 0)$.

NB. These are the same initial value problems you worked out in Exercise 3.3.3.

5.2.9 Using partial fraction expansion, determine the solution to the initial value problems from Exercise 5.2.8.

Section 5.3

Fourier transform methods for differential equations

Chapter 6

An introduction to partial differential equations

In this chapter we introduce the subject of partial differential equations. We make no attempt whatsoever to provide a comprehensive treatment of partial differential equations. Rather, we attempt to introduce the subject by touching on some important ideas that arise in the theory, primarily through the use of targeted examples. Our focus is on three facets of the theory of partial differential equations: (1) characteristics of partial differential equations; (2) properties of elliptic, hyperbolic, and parabolic partial differential equations; (3) the notion of a weak solution. We shall also consider a few partial differential equations that can be solved. The only such equations we consider are those whose solution can be obtained by solving ordinary differential equations.

We begin our discussion in Section 6.1 by considering the notion of a "characteristic." This notion is sometimes revealing about the general characteristics of solutions of partial differential equations. It is also useful in understanding how one prescribes for partial differential equations the analogue of initial conditions for ordinary differential equations. For first-order partial differential equations studied in Section 6.2, the characteristics of the equation often allow a reduction of the finding of solutions to the finding of solutions for ordinary differential equations. We also work with first-order partial differential equations that come from conservation laws, since these allow a useful geometric understanding of solutions.

Next we turn to three of the important second-order partial differential equations that arise in many applications: the heat equation, the wave equation, and the potential equation. We shall apply the general ideas connected with transform methods from Chapter 5 to study some concrete partial differential equations with associated boundary value problems. As we shall see, there are three parts of this analysis. First of all, we reduce the problem to an ordinary differential equation connected with an eigenvalue/eigenvector problem. The solution of this eigenvalue/eigenvector analysis motivates transform analysis to reduce the partial differential equations to ordinary differential equations in the transformed variables. After obtaining a solution in the transformed variables, to get the solution in the original variables, we must apply the inverse transform. A naïve application of the inverse gives a "formal" solution to the boundary value problem. But, as

we commented upon at length for each of our transforms in Chapter 5, the matter of just when and how the inverse transform works is a matter of some subtlety. We shall provide some results in this chapter that show that, in fact, the "formal" solution is often an actual solution. As we shall see, this typically requires some difficult analysis.

The final main topic of the chapter is the important notion of a weak solution. This idea arises since it is often advantageous to relax what is meant by a solution, so that existence of solutions becomes easier. One than hopes that the existence of these weak solutions can be used to infer solutions in the normal sense.

The reader will observe that the scope of this chapter is a little different than that of our preceding chapters concerning the analysis of ordinary differential equations. More precisely, we work quite hard to solve a few rather specific problems. This is rather a feature of partial differential equations, in general. Indeed, a treatment of partial differential equations at the level with which we have treated ordinary differential equations thus far—and we should emphasise that this treatment of ordinary differential equations is not comprehensive—is simply not possible, and even a comprehensive treatment of such special cases as can be solved requires a substantial effort, and appears quite fragmented by comparison with what one can do with ordinary differential equations. Thus, the way in which this chapter should be viewed is as a superficial introduction to certain aspects of the theory of partial differential equations.

Contents

Section 6.1

Characteristics of partial differential equations

In this section we engage in a general discussion of so-called characteristics of partial differential equations. As we shall see, these permit a geometric understanding of some aspects of solutions of partial differential equations. Particularly, discussions of characteristics sometimes allow a natural discussion of the sorts of boundary conditions that are permitted by a differential equation.

6.1.1 Characteristic for linear partial differential equations

6.1.2 Characteristics for quasilinear partial differential equations

6.1.3 Characteristics for nonlinear partial differential equations

6.1.4 The Cauchy–Kovalevskaya Theorem

Section 6.2

First-order partial differential equations

We begin our detailed discussion of partial differential equations by considering first-order equations. We concentrate on the

6.2.1 The Method of Characteristics for first-order equations

6.2.2 First-order conservation laws

Section 6.3

The heat equation

The first partial differential equation we look at is an example of what is known of as a "parabolic" equation; see Section 1.3.4.3. The specific parabolic equation we work with is known as the "heat equation" or the "diffusion equation." We first considered this differential equation in Section 1.1.11, and there we derived it as a model for heat flow in a one-dimensional medium, and also gave higher-dimensional analogues. In this section we study the heat equation in a few different ways. First we examine the characteristics of the heat equation. By doing this we can understand some things about what "parabolic" means. We then consider particular instances of the equation, applying transform methods to obtain a "formal" solution for the equation. We shall prove some results that indicate when and how this "formal" solution is an actual solution. We shall work with two versions of the heat equation, one where the spatial variable is restricted to a finite interval, and the other where the spatial variable is unbounded.

6.3.1 Characteristics for the heat equation

6.3.2 The heat equation for a finite length rod

In this section we consider a particular physically meaningful instantiation of the heat equation, namely the case in which we are modelling the temperature distribution in a rod of finite length ℓ. For ordinary differential equations, one must specify an appropriate number of conditions, normally (but not always) all at the same time in order to determine the solution. These are called "initial conditions." Similar circumstances arise in partial differential equations. On silly considerations, one might speculate that three conditions are required for the heat equation since the equation has three derivatives. Well, the heat equation is a simple enough partial differential equation that the silly consideration suffices. The conditions can come in many forms, and we seek to develop familiarity through example. Let us suppose that the rod has finite length ℓ with the left end of the rod being at $x = 0$. Perhaps it is the case that we know the temperature at the ends of the rod:

$$u(0, t) = T_0, \quad u(\ell, t) = T_1.$$

Let us reduce this to a simple special case. By defining

$$v(x, t) = u(x, t) + \frac{T_0 - T_1}{\ell} x - T_0$$

We then see that

$$\frac{\partial v}{\partial t} = k \frac{\partial^2 v}{\partial x},$$

and that

$$v(0,t) = 0, \quad v(\ell,t) = 0.$$

Thus we may as well assume that $T_0 = T_1 = 0$. This gives us two boundary conditions. To provide another we may specify the temperature distribution in the rod at $t = 0$. Thus we may set

$$u(x,0) = f(x),$$

for some function f. Thus we have arrived at the following ***boundary value problem***:

$$\begin{aligned} \frac{\partial u}{\partial t} &= k\frac{\partial^2 u}{\partial x^2} & u(0,t) &= 0, \quad u(\ell,t) = 0 \\ & & u(x,0) &= f(x). \end{aligned} \tag{6.1}$$

The exact nature of the function f we leave undetermined for the moment. Also, we mention that other types of boundary conditions are possible. The reader may explore one of these in Exercise 6.3.2.

6.3.2.1 Formal solution Let us now set about obtaining a solution for the problem (6.1). We shall reduce the partial differential equation to solving a bunch of ordinary differential equations. Since this is the first time we are doing this, we shall motivate the idea in several different ways. In all cases, the starting point for the motivation is the *assumption* that we seek separable solutions for the heat equation, by which we mean solutions of the form $u(x,t) = \xi(x)\tau(t)$, i.e., we "separate" the solution into a part depending on time and a part depending on displacement. This is, of course, an assumption, and it is a good assumption if and only if it works. Since we are talking about it, apparently it works.

Motivation using eigenvalues and eigenfunctions

In Section 3.2.3, we have carefully studied ordinary differential equations whose solutions satisfy

$$\dot{\xi}(t) = A(\xi(t)), \quad \xi(t_0) = x_0,$$

where $\xi(t) \in \mathsf{V}$ and $A \in \mathrm{L}(\mathsf{V};\mathsf{V})$ for an n-dimensional $\mathbb{R}$-vector space V. Moreover, we saw that, except for issues concerning the need to work with generalised eigenvectors and complex eigenvalues, solutions were of the form

$$\xi(t) = \sum_{j=1}^{n} c_j \mathrm{e}^{\lambda_j t} v_j,$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues and $v_1, \dots, v_n$ are corresponding eigenvectors. The coefficients $c_1, \dots, c_n$ are determined from the initial conditions, e.g.,

$$\xi(0) = \sum_{j=1}^{n} c_j v_j = x_0,$$

and these linear equations can be solved for $c_1, \dots, c_n$.

For the heat equation, we proceed in a similar spirit. Here, however, we replace V with functions defined on the interval $[0, \ell]$. Additionally, because of the boundary conditions, we only consider functions defined on the interval $[0, \ell]$ that are zero at the endpoints. That is to say, let us consider

$$\mathsf{V} = \{\xi \colon [0, \ell] \to \mathbb{C} \mid \xi \text{ is infinitely differentiable, and } \xi(0) = \xi(\ell) = 0\}.$$

The linear map A is then $A(\xi) = \frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}$, and then the heat equation takes the form

$$\frac{\mathrm{d}u}{\mathrm{d}t}(t) = A(u(t)),$$

where we think of u as being a map $u \colon \mathbb{R}_{\geq 0} \to \mathsf{V}$, i.e., for each time t, $u(t) \in \mathsf{V}$ is an infinitely differentiable function on $[0, \ell]$ vanishing at 0 and ℓ. Now, motivated by our work in Section 3.2.3, we look for eigenvalues and eigenvectors for A; we shall call the eigenvectors ***eigenfunctions***, since they are indeed functions. Eigenvalues are then numbers λ satisfying

$$A(\xi) = \frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}(x) = \lambda\xi(x) \tag{6.2}$$

for some nonzero $\xi \in \mathsf{V}$. Note that this is an ordinary differential equation! It is an ordinary differential equation where the unknown parameter λ are the eigenvalues of A.

To determine λ, we solve the differential equation and apply the conditions that solutions must vanish at $x = 0$ and $x = \ell$. We do this according to three cases.

1. $\lambda \in \mathbb{R}_{>0}$: In this case the differential equation (6.2) has the solution

   $$\xi(x) = A_1 \sinh(\sqrt{\lambda}x) + A_2 \cosh(\sqrt{\lambda}x).$$

   Here, for those who for some reason have not seen them, sinh and cosh refer to the hyperbolic sine and cosine functions defined by

   $$\sinh(x) = \tfrac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x}), \quad \cosh(x) = \tfrac{1}{2}(\mathrm{e}^x + \mathrm{e}^{-x}).$$

   Let us apply the boundary conditions. The condition $\xi(0) = 0$ gives $A_2 = 0$. The condition $\xi(\ell) = 0$ then gives

   $$A_1 \sinh(\sqrt{\lambda}\ell) = 0.$$

   This can only hold when $A_1 = 0$. Thus the only way λ can be positive is if the resulting solution is identically zero. Thus $\lambda \in \mathbb{R}_{>0}$ cannot be an eigenvalue, since eigenfunctions are necessarily nonzero.

2. $\lambda = 0$: In this case the solution for (6.2) is

$$\xi(x) = A_1 x + A_2.$$

Again, an application of the boundary conditions gives $A_1 = A_2 = 0$, a situation which precludes $\lambda = 0$ from being an eigenvalue.

3. $\lambda \in \mathbb{R}_{<0}$: You are either beginning to worry because we are running out of options, or you think that I have cagily left the good case to the end... The solution of the ordinary differential equation (6.2) is

$$\xi(x) = A_1 \sin(\sqrt{-\lambda}x) + A_2 \cos(\sqrt{-\lambda}x).$$

The boundary condition $\xi(0) = 0$ gives $A_2 = 0$. The boundary condition $\xi(\ell) = 0$ gives

$$A_1 \sin(\sqrt{-\lambda}\ell) = 0.$$

Here we have an option other than $A_1 = 0$. By a propitious choice of λ it can be arranged that

$$\sin(\sqrt{-\lambda}\ell) = 0.$$

Indeed, if

$$\sqrt{-\lambda}\ell = n\pi$$

for some $n \in \mathbb{Z}_{>0}$, then we are set to go.

The above arguments indicate that any of the numbers

$$\lambda_n = -\frac{n^2\pi^2}{\ell^2}, \quad n \in \mathbb{Z}_{>0},$$

are eigenvalues for A and that the corresponding eigenfunctions are the elements of V given by

$$\xi_n(x) = \sin(n\pi\tfrac{x}{\ell}), \quad n \in \mathbb{Z}_{>0}.$$

We then follow through with the method by analogy to what we did in Section 3.2.3 when V is a finite-dimensional vector space, and seek a solution of the form

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{k\lambda_n t}\xi_n(x) = \sum_{n=1}^{\infty} c_n \underbrace{e^{-\frac{kn^2\pi^2}{\ell^2}t}}_{\tau_n(t)} \underbrace{\sin(n\pi\tfrac{x}{\ell})}_{\xi_n(x)}.$$

Note that u is an infinite sum of separated functions, i.e., products of a function of x with a function of t. To determine the coefficients c_n, $n \in \mathbb{Z}_{>0}$, we use the initial condition, again mimicking the finite-dimensional case. For our situation we have

$$u(x,0) = \sum_{n=1}^{\infty} c_n \sin(n\pi\tfrac{x}{\ell}) = f(x).$$

Motivated by our work with the inverse of the CDFT in Section 5.1.1.2, we attempt to solve for the coefficients c_n, $n \in \mathbb{Z}_{>0}$, by integration:

$$\begin{aligned}
&\sum_{n=1}^{\infty} c_n \sin(n\pi\tfrac{x}{\ell}) = f(x)\\
\implies &\sum_{n=1}^{\infty} c_n \int_0^{\ell} \sin(n\pi\tfrac{x}{\ell}) \sin(m\pi\tfrac{x}{\ell}) \,\mathrm{d}x = \int_0^{\ell} f(x) \sin(m\pi\tfrac{x}{\ell}) \,\mathrm{d}x\\
\implies &c_m = \frac{2}{\ell} \int_0^{\ell} f(x) \sin(m\pi\tfrac{x}{\ell}) \,\mathrm{d}x,
\end{aligned}$$

using the fact that

$$\int_0^{\ell} \sin(n\pi\tfrac{x}{\ell}) \sin(m\pi\tfrac{x}{\ell}) \,\mathrm{d}x = \begin{cases} 0, & m \neq n,\\ \frac{\ell}{2}, & m = n. \end{cases}$$

Summarising, we have obtained the formula

$$u(x,t) = \sum_{n=1}^{\infty} \left(\frac{2}{\ell} \int_0^{\ell} f(x) \sin(n\pi\tfrac{x}{\ell}) \,\mathrm{d}x \right) \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t} \sin(n\pi\tfrac{x}{\ell}) \tag{6.3}$$

for a solution. This is what we call a ***formal solution***, since (1) its derivation relies on some operations like swapping sums and integrals that are not *a priori* valid and (2) it comes to us as an infinite sum whose convergence is not necessarily known to us.

Motivation using the CDFT

Next we consider applying an appropriate transform to the heat equation to convert it into an algebraic equation. The idea here is motivated by the discussion of the CDFT in Section 5.1.1, but we have to make some modifications particular to the specific problem.

First let us see what is wrong with a "verbatim" application of the CDFT to the problem. Let us, as in our eigenvalue method above, consider the space

$$\mathsf{V} = \{\xi \colon [0,\ell] \to \mathbb{C} \mid \ \xi \text{ is infinitely differentiable, and } \xi(0) = \xi(\ell) = 0\}.$$

Because the problem is defined on $[0,\ell]$, given $\xi \in \mathsf{V}$, we can consider $\mathscr{F}_{\mathrm{CD}}(\xi) \colon \mathbb{Z} \to \mathbb{C}$ defined by

$$\mathscr{F}_{\mathrm{CD}}(\xi)(n) = \int_0^{\ell} \xi(x) \mathrm{e}^{-2\pi i n \frac{x}{\ell}} \,\mathrm{d}x, \qquad n \in \mathbb{Z}.$$

This is a perfectly valid thing to do. However, let us now begin to modify this, given the specific nature of elements of V. Taking this into account, we may want to

only consider applying the CDFT for the functions $e^{2\pi i n\frac{x}{\ell}}$ that satisfy the boundary conditions. Note, however, that

$$e^{2\pi i n\frac{0}{\ell}} = e^{2\pi i n\frac{\ell}{\ell}} = 1,$$

so *none* of these functions satisfy the boundary conditions at $x = 0$ and $x = \ell$. However, the failure also suggests a kludge: we can consider appropriate linear combinations of these functions. For example, the functions

$$e_n(x) = e^{2\pi i n\frac{x}{\ell}} - e^{-2\pi i n\frac{x}{\ell}}, \qquad n \in \mathbb{Z}_{>0},$$

do satisfy the boundary conditions. Note that $e_n(x) = 2i\sin(2n\pi\frac{x}{\ell})$. That is to say, our modified guess at an appropriate transform, which we denote by $\widehat{\mathscr{F}_{\text{CD}}}$, to apply is $\widehat{\mathscr{F}_{\text{CD}}}(\xi)\colon \mathbb{Z}_{>0} \to \mathbb{C}$ defined by

$$\widehat{\mathscr{F}_{\text{CD}}}(\xi)(n) = \int_0^\ell \xi(x)\sin(2\pi n\tfrac{x}{\ell})\,\mathrm{d}x, \qquad n \in \mathbb{Z}_{>0}.$$

This, however, still has a defect. The defect is the formula

$$\sin(2\pi n\tfrac{x-\ell/2}{\ell}) = -\sin(2\pi n\tfrac{x+\ell/2}{\ell}),$$

reflecting the fact that the sine functions we are using have a symmetry when reflected about $x = \frac{\ell}{2}$. Thus these functions will not be capable of representing any function that does not share this symmetry. Again, the failure suggests a kludge. Keeping in mind this symmetry of the sine functions, we fictitiously extend our functions to be defined on $[0, 2\ell]$ by taking, for $\xi \in \mathsf{V}$, the extension to satisfy

$$\xi(x) = -\xi(2\ell - x), \qquad x \in [\ell, 2\ell],$$

see Figure 6.1. Note that the extended function vanished at 0 and 2ℓ. We now apply our modified CDFT to these extended functions on $[0, 2\ell]$:

$$\begin{aligned}
\widehat{\widehat{\mathscr{F}_{\text{CD}}}}(\xi)(n) &= \int_0^{2\ell} \xi(x)\sin(2n\pi\tfrac{x}{2\ell})\,\mathrm{d}x \\
&= \int_0^{\ell} \xi(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x + \int_\ell^{2\ell} \xi(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x \\
&= \int_0^{\ell} \xi(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x + \int_0^{\ell} \xi(2\ell - x)\sin(n\pi\tfrac{2\ell-x}{\ell})\,\mathrm{d}x \\
&= 2\int_0^{\ell} \xi(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x.
\end{aligned}$$

Next we claim that

$$\widehat{\widehat{\mathscr{F}_{\text{CD}}}}\left(\frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}\right)(n) = -\frac{n^2\pi^2}{\ell^2}\widehat{\widehat{\mathscr{F}_{\text{CD}}}}(\xi)(n), \qquad n \in \mathbb{Z}_{>0},$$

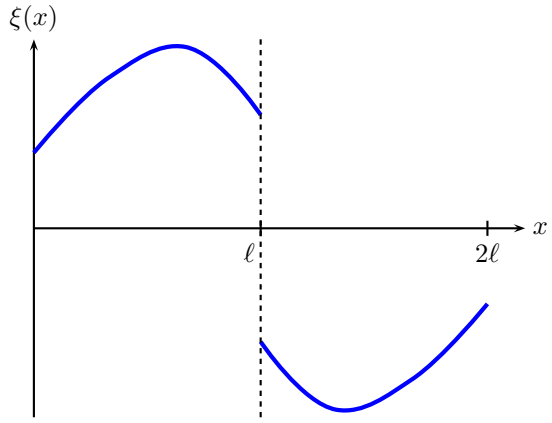


Figure 6.1 "Oddly" extending a function from $[0, \ell]$ to $[\ell, 2\ell]$

for every $\xi \in \mathsf{V}$. Indeed,

$$\begin{aligned}
\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}\left(\frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}\right)(n) &= 2\int_0^\ell \frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x \\
&= 2\frac{\mathrm{d}\xi}{\mathrm{d}x}(x)\sin(n\pi\tfrac{x}{\ell})\Big|_0^\ell - 2\frac{n\pi}{\ell}\int_0^\ell \frac{\mathrm{d}\xi}{\mathrm{d}x}(x)\cos(n\pi\tfrac{x}{\ell})\,\mathrm{d}x \\
&= -2\frac{n\pi}{\ell}\int_0^\ell \frac{\mathrm{d}\xi}{\mathrm{d}x}(x)\cos(n\pi\tfrac{x}{\ell})\,\mathrm{d}x \\
&= -2\frac{n\pi}{\ell}\xi(x)\cos(n\pi\tfrac{x}{\ell})\Big|_0^\ell - 2\frac{n^2\pi^2}{\ell^2}\int_0^\ell \xi(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x \\
&= -\frac{n^2\pi^2}{\ell^2}\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}(\xi)(n),
\end{aligned}$$

using integration by parts twice.

Now, after some somewhat crazy machinations, we have arrived at an hopefully appropriate transform to apply to the heat equation in this case. Let us go ahead and do this:

$$\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}\left(\frac{\partial u}{\partial t}\right)(n) = k\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}\left(\frac{\partial^2 u}{\partial x^2}\right)(n) = -\frac{kn^2\pi^2}{\ell^2}\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}(u)(n) \quad \Longrightarrow \quad \frac{\mathrm{d}\hat{u}_n}{\mathrm{d}t} = -\frac{kn^2\pi^2}{\ell^2}\hat{u}_n,$$

where we abbreviate $\widehat{\widehat{\mathscr{F}_{\mathrm{CD}}}}(u)(n) = \hat{u}_n$. We note now that this is an elementary first-order ordinary differential equation for $\hat{u}_n$ which we can solve:

$$\hat{u}_n(t) = c_n \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t},$$

for some constant c_n to be determined. Note that the preceding expression gives us the solution of the equation in the transformed variables, i.e., as a function of

(n, t). To get this as a function of (x, t), we apply the inverse transform, following Section 5.1.1.2. This means that we write

$$u(x,t) = \sum_{n=1}^{\infty} \widehat{\mathscr{F}_{\mathrm{CD}}}(u)(n) \sin(n\pi \tfrac{x}{\ell}) = \sum_{n=1}^{\infty} c_n \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t} \sin(n\pi \tfrac{x}{\ell}).$$

Now we are in the same situation as with our eigenvalue/eigenfunction method, and we use the initial condition $u(x, 0) = f(x)$ to obtain the unknown coefficients c_n, $n \in \mathbb{Z}_{>0}$:

$$c_n = \frac{2}{\ell} \int_0^{\ell} f(x) \sin(n\pi \tfrac{x}{\ell}) \, \mathrm{d}x.$$

This then gives the same ***formal solution***

$$u(x,t) = \sum_{n=1}^{\infty} \left(\frac{2}{\ell} \int_0^{\ell} f(x) \sin(n\pi \tfrac{x}{\ell}) \, \mathrm{d}x \right) \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t} \sin(n\pi \tfrac{x}{\ell})$$

as with our eigenvalue/eigenfunction method. It has the same limitations as well, of course.

A combination of the previous heuristics

We now have two methods of producing a formal solution to the problem (6.1). Each method has some benefits and drawbacks. Let us discuss these.

The eigenvalue/eigenfunction method provides us with a well motivated (based on our techniques of Section 3.2.3) problem that we can solve in some systematic way. It does, however, have the drawback of converting the problem into one whose connection with what we know is a bit tenuous. That is to say, we have moved from working with a linear ordinary differential equation in a finite-dimensional vector space to a "linear ordinary differential equation" in the space

$$\{\xi \colon [0, \ell] \to \mathbb{C} \mid \ \xi \text{ is infinitely differentiable, and } \xi(0) = \xi(\ell) = 0\}$$

that is not finite-dimensional. The validity of extending our methods of Section 3.2.3 is something one must question.

The method that adapts the CDFT to arrive at a formal solution has the benefit of being a part of a general idea that seems compelling, and which we have seen previously in Section 5.2: (1) apply a transform to a differential equation; (2) solve the equation in the transformed variables; (3) apply the inverse transform to get the solution in the original variables. However, the method is complex in that finding the transform that one must apply is complicated. Indeed, it is difficult to imagine how our machinations might be adapted to any general setting.

There is, however, a reasonable way of combining the two methods, and let us outline how to do this. What we do is describe a procedure that is vague enough that it has a chance of being applied in general settings, and then indicating how

each step is applied in the solution of the specific boundary value problem (6.1). In Sections 6.4 and 6.5 we apply the procedure to the wave equation and the potential equation. The hope is that the vague procedure and a few illustrations of it will provide the reader with enough background to attempt problems that are amenable to this method.

Here is the "procedure."

6.3.1 "Procedure" (A method for obtaining a formal solution to some boundary value problems) We suppose that we are given a linear partial differential equation with independent variables $(x_1, \dots, x_n) \in D \subseteq \mathbb{R}^n$ and with a single state $u \in \mathbb{R}$.

1. (a) *General strategy:* Make a decision about which variable(s) you will use to create an eigenvalue/eigenfunction problem, and define the space of functions you will use that satisfies the boundary conditions. Let us suppose, without loss of generality, that these variables are $(x_1, \dots, x_m)$ for $m \in \{1, \dots, n\}$. We assume that these variables reside in

$$D_0 = [0, \ell_1] \times \cdots \times [0, \ell_m],$$

and we denote by V the space of smooth functions on D_0 which satisfy the boundary conditions is each variable, when all others are fixed.

(b) *Application to the heat equation:* For the heat equation, it is natural to select the x variable as the candidate for the eigenvalue/eigenfunction problem, since the boundary conditions at $x = 0$ and $x = \ell$ are well adapted to defining the space

$$\mathsf{V} = \{\xi \colon [0, \ell] \to \mathbb{C} \mid \xi \text{ is infinitely differentiable, and } \xi(0) = \xi(\ell) = 0\}.$$

Note that the initial condition in the t variable does not work well in this respect, since the initial condition $u(x, 0) = f(x)$ is difficult to translate into a nice space of functions.

2. (a) *General strategy:* Associated with each of the chosen variables $(x_1, \dots, x_m)$ one will hopefully have an eigenvalue/eigenfunction problem in the form of an ordinary differential equation and associated boundary conditions. In order to make sense of the method, these equations must be decoupled, i.e., can be solved independently. With any luck, for each of the variables x_j, $j \in \{1, \dots, m\}$, one has eigenvalues $\lambda_{j,n}$, $n \in \mathbb{Z}_{>0}$, and associated eigenfunctions $\xi_{j,n}$, $n \in \mathbb{Z}_{>0}$.

(b) *Application to the heat equation:* This is the step where one has the ordinary differential equation with boundary values

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}(x) = \lambda\xi(x), \quad \xi(0) = \xi(\ell) = 0.$$

It is this equation with boundary values that leads to the eigenvalues $\lambda_n = -\frac{n^2\pi^2}{\ell^2}$ and eigenfunctions $\xi_n(x) = \sin(n\pi\frac{x}{\ell})$, $n \in \mathbb{Z}_{>0}$.

3. (a) *General strategy:* Associated to the eigenfunctions $\xi_{j,n}$, $j \in \{1, \dots, m\}$, $n \in \mathbb{Z}_{>0}$, we have an ***eigenfunction transform***. This we denote by $\mathscr{E} \colon \mathsf{V} \to \mathbb{Z}_{>0}^m$, and define by

$$\mathscr{E}(f)(n_1, \dots, n_m) = \int_0^{\ell_1} \cdots \int_0^{\ell_m} f(x_1, \dots, x_m) \xi_{1,n_1}(x_1) \cdots \xi_{m,n_m}(x_m) \, \mathrm{d}x_m \cdots \mathrm{d}x_1.$$

This transform should satisfy some rules with respect to differantiation.

(b) *Application to the heat equation:* The transform here is

$$\mathscr{E}(\xi)(n) = \int_0^{\ell} \xi(x) \sin(n\pi \tfrac{x}{\ell}) \, \mathrm{d}x.$$

It satisfies the condition

$$\mathscr{E}\left(\frac{\mathrm{d}^2\xi}{\mathrm{d}x^2}\right)(n) = -\frac{n^2\pi^2}{\ell^2} \mathscr{E}(\xi)(n)$$

for the second derivative.

4. (a) *General strategy:* Apply the eigenfunction transform to the partial differential equation to get differential equations (hopefully an ordinary differential equation) for the transformed variables

$$\hat{u}_{n_1 \cdots n_m} \triangleq \mathscr{E}(u)(n_1, \dots, n_m), \qquad n_1, \dots, n_m \in \mathbb{Z}_{>0}.$$

Note $\hat{u}_{n_1 \cdots n_m}$ are functions of $x_{m+1}, \dots, x_n$.

(b) *Application to the heat equation:* Here we have

$$\hat{u}_n(t) = \int_0^{\ell} u(x, t) \sin(n\pi \tfrac{x}{\ell}) \, \mathrm{d}x,$$

and so

$$\mathscr{E}\left(\frac{\partial u}{\partial t}\right)(n) = \mathscr{E}\left(k\frac{\partial^2 u}{\partial x^2}\right)(n) = -\frac{kn^2\pi^2}{\ell^2} \mathscr{E}(u)(n)$$

$$\implies \qquad \frac{\mathrm{d}\hat{u}_n}{\mathrm{d}t} = -\frac{kn^2\pi^2}{\ell^2} \hat{u}_n, \qquad n \in \mathbb{Z}_{>0}.$$

5. (a) *General strategy:* Hopefully solve the differential equation(s) for $\hat{u}_{n_1 \cdots n_m}$, with some unknown coefficients that will be determined using initial conditions. This will give $\hat{u}_{n_1 \cdots n_m}$ as functions of $x_{m+1}, \dots, x_n$.

(b) *Application to the heat equation:* We have $\hat{u}_n(t) = c_n \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2} t}$.

6. (a) *General strategy:* Write the formal solution, with unknown coefficients, as

$$u(x_1, \dots, x_m, x_{m+1}, x_n) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_m=1}^{\infty} \hat{u}_{n_1 \cdots n_m}(x_{m+1}, \dots, x_n) \xi_{n_1}(x_1) \cdots \xi_{n_m}(x_m).$$

(b) *Application to the heat equation:* We have

$$u(x,t) = \sum_{n=1}^{\infty} c_n \mathrm{e}^{-\frac{n^2\pi^2}{\ell^2}t} \sin(n\pi\tfrac{x}{\ell}).$$

7. (a) *General strategy:* Use the remaining boundary conditions to determine the unknown coefficients in the functions $\hat{u}_{n_1\cdots n_m}$. This produces the required formal solution.

(b) *Application to heat equation:* Using the computations above, we have

$$c_n = \frac{2}{\ell}\int_0^{\ell} f(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x,$$

giving the formal solution (6.3). •

There is a lot of "hopefullies" in the preceding procedure. However, there are a variety of problems where the method works out more or less like it does for the heat equation. One of the crucial steps is the determination of the eigenvalues and eigenfunctions in Step 2. For this step there is a well-developed (and difficult) theory that we present in Chapter 7.

6.3.2.2 Rigorous establishment of solutions In the previous section, we developed at length a methodology for arriving at a formal solution for (6.1). In carrying out the procedure, we made some steps that are certainly open to question. In this section we take the *result* of the manipulations, and show that it does solve the problem. In this way, even though some of the steps in the method of separation of variables are not strictly legit, we do not worry about it as the output of the procedure is a solution to the initial boundary value problem.

The main result in this section is the following which tells us that the situation is pretty good for the formal solution of the heat equation, even for very general boundary functions f. Indeed, we allow such boundary functions in the space

$$\mathrm{L}^2([0,\ell];\mathbb{C}) = \left\{ f\colon [0,\ell] \to \mathbb{C} \;\middle|\; \int_0^{\ell} |f(x)|^2\,\mathrm{d}x < \infty \right\},$$

which is equipped with the norm

$$\|f\|_2 = \left(\int_0^{\ell} |f(x)|^2\,\mathrm{d}x \right)^{1/2}.$$

Just why this is a good space to work with is not something that ought to be clear at this point, but will be developed in Chapter 7. In the statement of the result we denote by $f_{\mathrm{per}}\colon \mathbb{R} \to \mathbb{C}$ the ℓ-periodic extension of $f\colon [0,\ell] \to \mathbb{C}$ defined by $f_{\mathrm{per}}(x) = f(x - k\ell)$ if $x \in [k\ell, (k+1)\ell)$.

6.3.2 Theorem (Solutions for the heat equation on an interval) *Consider the boundary value problem*

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2} \qquad \begin{aligned} &u(0,t) = 0, \quad u(\ell,t) = 0\\ &u(x,0) = f(x). \end{aligned}$$

If $f \in L^2([0,\ell];\mathbb{C})$*, then*

(i) the series (6.3) converges uniformly on

$$\{(x,t) \in [0,\ell] \times \mathbb{R}_{\geq 0} \mid t \geq t_0\}$$

for each $t_0 \in \mathbb{R}_{>0}$*.*

Moreover, $u\colon [0,\ell] \times \mathbb{R}_{>0} \to \mathbb{C}$ *as defined by (6.3) has the following properties:*

(ii) u *satisfies the heat equation and the first two of the boundary conditions in (6.1) on* $[0,\ell] \times \mathbb{R}_{>0}$*;*

(iii) u *is infinitely differentiable on* $(0,\ell) \times \mathbb{R}_{>0}$*;*

(iv) $\lim_{t\to 0} \|u_t - f\|_2 = 0$ *where* $u_t\colon [0,\ell] \to \mathbb{C}$ *is defined by* $u_t(x) = u(x,t)$*.*

Furthermore,

(v) if f_{per} *is continuous and if* f' *is piecewise continuous, then the convergence of* u_t *to* f *is part (iv) is uniform;*

(vi) if f *is arbitrary, with only the property that, for* $x \in [0,\ell]$*,*

$$\sum_{n=1}^{\infty} c_n \sin(n\pi \tfrac{x}{\ell}) = f(x)$$

i.e., the series converges pointwise at x*, then* $\lim_{t\to 0} u(x,t) = f(x)$*.*

Proof (i) Consider the series for u:

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-\frac{kn^2\pi^2}{\ell^2}t} \sin(n\pi \tfrac{x}{\ell}),$$

with

$$c_n = \frac{2}{\ell}\int_0^{\ell} f(x)\sin(n\pi \tfrac{x}{\ell})\,dx.$$

Since $f \in L^2([0,\ell];\mathbb{C}) \subseteq L^1([0,\ell];\mathbb{C})$, by definition of c_n there exists $M > 0$ so that $|c_n| < M$ for all $n \in \mathbb{Z}_{>0}$. For fixed $t_0 \in \mathbb{R}_{>0}$ we have

$$|c_n e^{-\frac{kn^2\pi^2}{\ell^2}t} \sin(n\pi \tfrac{x}{\ell})| \leq M e^{-\frac{kn^2\pi^2}{\ell^2}t_0}$$

for all $t \geq t_0$ and $n \in \mathbb{Z}_{>0}$. Thus uniform convergence in

$$\{(x,t) \in [0,\ell] \times \mathbb{R}_{\geq 0} \mid t \geq t_0\}$$

will follow from the Weierstrass M-test if we can show that the series of real numbers $\sum_{n=1}^{\infty} \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t_0}$ converges. This can be shown to be true using, for example, the ratio test.

(ii) The series is uniformly convergent, as we saw in the preceding part of the proof. Moreover, the series with terms differentiated with respect to x or t any finite number of times will give a series whose terms have the form

$$P(n)\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\sin(n\pi\tfrac{x}{\ell}) \text{ or } P(n)\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\sin(n\pi\tfrac{x}{\ell}),$$

for some polynomial P. Such a series will again be uniformly convergent, again by the Weierstrass M-test. Therefore, by ***missing stuff***, we can interchange any finite number of derivatives of u with respect to x and t with the summation. Therefore, we have

$$\begin{aligned}\frac{\partial u}{\partial t}(x,t) &= \frac{\partial}{\partial t}\left(\sum_{n=1}^{\infty} c_n \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\sin(n\pi\tfrac{x}{\ell})\right)\\ &= -\sum_{n=1}^{\infty}\frac{c_n k n^2\pi^2}{\ell^2}\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\sin(n\pi\tfrac{x}{\ell})\\ &= k\frac{\partial^2}{\partial x^2}\left(\sum_{n=1}^{\infty} c_n \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\sin(n\pi\tfrac{x}{\ell})\right) = k\frac{\partial^2 u}{\partial x^2}(x,t),\end{aligned}$$

i.e., u satisfies the heat equation.

(iii) Our arguments from part (ii) allows us to conclude that u is infinitely differentiable on $(0,\ell)\times\mathbb{R}_{>0}$ by successive applications of ***missing stuff***.

(iv) We first note that

$$\|u_t - f\|_2^2 = \sum_{n=1}^{\infty}|c_n|^2\left(1-\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\right)^2$$

by Parseval's equality. Thus the result will follow if we can show that the series on the right converges uniformly as a function of t. For, if this is the case, then the function of t given by

$$g(t) = \sum_{n=1}^{\infty}|c_n|^2\left(1-\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\right)^2$$

is continuous, and thus the limit $\lim_{t\to 0} g(t)$ exists, and is equal to zero. To prove uniform convergence of this series we use Abel's test, ***missing stuff***, with $x = t$, $g_n(t) = 1-\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}$, and $f_n = |c_n|^2$. One directly verifies that, with these substitutions, the hypotheses of Abel's test are satisfied, and so the series converges uniformly.

(v) With the stated hypotheses, we saw in ***missing stuff*** that the series $\sum_{n=1}^{\infty}|c_n|$ converges. For fixed x we then have

$$|f(x)-u(x,t)| \le \sum_{n=1}^{\infty}|c_n|\left(1-\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\right). \tag{6.4}$$

By an application of Abel's test, following the same lines as in the proof of part (iv), it follows that $\lim_{t\to 0} u_t(x) = f(x)$. What's more, since the right-hand side of (6.4) is independent of x, this convergence is uniform.

(vi) Suppose that, for some $x \in [0, \ell]$, we have

$$f(x) = \sum_{j=1}^{\infty} c_n \sin(n\pi \tfrac{x}{\ell}).$$

Then, using (6.4) and the Dominated Convergence Theorem,

$$\lim_{t\to 0} |f(x) - u(x,t)| \leq \lim_{t\to 0} \sum_{n=1}^{\infty} |c_n| \left(1 - \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\right) = \sum_{n=1}^{\infty} |c_n| \lim_{t\to 0} \left(1 - \mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}\right) = 0,$$

as desired. ■

Let us make some comments on the character of the solution to the heat equation.

6.3.3 Remarks (Solutions for the heat equation)

1. The heat equation does a remarkable thing. It will take an extremely general class of functions, those in $\mathrm{L}^2([0,\ell];\mathbb{C})$, and instantaneously "smooth" them. This is a consequence of part (iii) of Theorem 6.3.2. This is due to the presence of the term $\mathrm{e}^{-\frac{kn^2\pi^2}{\ell^2}t}$ in the series which decays to zero very quickly with n, provided that $t > 0$.
2. Note also that the solution to the heat equation is continuous at $t = 0$, provided that f is continuously differentiable, since in this case the series for $x \mapsto u(x, 0)$ converges pointwise to f by ***missing stuff***. Thus, what the equation does is turns a continuously differentiable function f into an infinitely differentiable function, and it does this in a continuous way.
3. The reason the heat equation is sometimes called the "diffusion equation" is left for the reader to explore in Exercise 6.3.2. •

6.3.3 The heat equation for an infinite length rod

6.3.3.1 Formal solution

6.3.3.2 Rigorous establishment of solutions

Exercises

6.3.1 Suppose that a rod of length ℓ generates heat internally at a rate per unit length specified by a function $g\colon [0,\ell] \to \mathbb{R}$. Derive the partial differential equation governing the temperature in the rod as a function of time and position along the rod.

6.3.2 Consider the boundary value problem

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2} \qquad \frac{\partial u}{\partial x}(0,t) = 0, \quad \frac{\partial u}{\partial x}(\ell,t) = 0 \qquad u(x,0) = f(x).$$

Answer the following questions.

(a) Use Procedure 6.3.1 to obtain a formal solution to the boundary value problem.

(b) Do you think that the convergence results stated in Section 6.3.2.2 will apply in this case?

(c) What is the behaviour of the temperature distribution in the rod as $t \to \infty$?

(d) Contrast your answer from part (c) with the answer to the same question for the boundary value problem (6.1). Explain why each case makes sense based upon physical arguments.
Hint: *What do the boundary conditions mean in each case?*

(e) Why do you think the heat equation is sometimes called the diffusion equation?

In the next exercise, you will consider an alternative to Procedure 6.3.1.

6.3.3 Consider the boundary value problem (6.1).

(a) Justify why, for fixed $t > 0$, it makes sense to expect that one can write

$$u(x,t) = \sum_{n=1}^{\infty} c_n(t) \sin(n\pi \tfrac{x}{\ell}).$$

(b) Obtain a differential equation for the coefficients $c_n(t)$, and solve the differential equation.

(c) What parts of the above procedure are in need of justification? How might this justification be provided?

6.3.4 Answer the following questions.

(a) Show that the boundary value problem (6.1) is equivalent (part of the problem is to determine the nature of this equivalence) to the boundary value problem

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2} + f \qquad u(0,t) = 0, \quad u(\ell,t) = 0 \qquad u(x,0) = 0,$$

where f is a function of x.

In Exercise 6.3.1 you showed that the partial differential equation governing the temperature distribution in a rod with a heat source is of the form given in part (a).

(b) What is the behaviour of the temperature distribution in a rod subject to zero temperature at the endpoints, zero initial temperature, and heat generation determined by a function g as in Exercise 6.3.1? Does this make sense to you? Is it consistent with the heat equation being also known as the diffusion equation?

(c) Consider the special case when $f(x) = 2\alpha$ for a constant α. Plot the steady-state temperature distribution and make sure it makes intuitive sense to you.

6.3.5 The drying of lumber can be described by the heat equation. We suppose that we have a very long and very wide piece of lumber so that the moisture content essentially varies as a function of the smallest cross-sectional dimension of the wood, denoted by $0 \le x \le \ell$. If u is the moisture content of the wood, then u is a function of x and $t > 0$. Assume that for $t > 0$ the outer edge of the wood is "dry," and that at $t = 0$ the piece of wood is uniformly "wet" with "wetness" W.

(a) Write the boundary value problem with the boundary conditions determined by the above description.

(b) Show that the moisture content in the lumber is given by

$$u(x,t) = \frac{4W}{\pi} \sum_{n=1}^{\infty} e^{-\frac{k(2n+1)^2\pi^2}{\ell^2}t} \frac{\sin\left((2n+1)\pi\frac{x}{\ell}\right)}{2n+1},$$

where k is the diffusion constant appearing in the heat equation.

(c) Show that for $t > 0$ we have

$$|u(x,t)| \le \frac{4W}{\pi} \frac{1}{e^{\frac{k\pi^2}{\ell^2}t} - 1}.$$

Hint: *Use the following facts:*

1. *If the series $\sum_{n=1}^{\infty} s_n$ is convergent then we have*

$$\left| \sum_{n=1}^{\infty} s_n \right| \le \sum_{n=1}^{\infty} |s_n|;$$

2. *we have*

$$\sum_{n=0}^{\infty} \alpha^n = \frac{1}{1-\alpha}$$

provided that $|\alpha| < 1$.

(d) Determine at expression for a time beyond which the wood is guaranteed to be 99% dry.

Section 6.4

The wave equation

In this section, we perform the manipulations of the previous section for the "hyperbolic" representative of our three partial differential equations, the "wave equation." We first considered the wave equation in Section 1.1.12, deriving it for the vibrations of a string and presenting higher-dimensional analogues of it. We begin our discussion in this section by looking at characteristics for the wave equation. Unlike the situation for the heat equation, the characteristics for the wave equation are crucial to understanding the behaviour of solutions to the equation. We then turn to solving the wave equation, first by obtaining a "formal" solution. Since many of the moves here mirror those for the heat equation, we are somewhat more brief with our treatment of how to obtain a formal solution to the wave equation. Then we consider how to verify that the formal solution is a *bone fide* solution. As we did for the heat equation, we shall consider both finite length and infinite length versions of the wave equation.

6.4.1 Characteristics for the wave equation

6.4.2 The wave equation for a finite length string

We first consider the wave equation on an interval of length ℓ. As we saw in Section 1.1.12, this is the sort of model that comes up when considering the transverse vibrations of a taut string. As with the heat equation, one needs some boundary values in order to specify a solution to the wave equation. By applying the silly, but apparently correct, argument involving the number of derivatives, we deduce that one needs four boundary conditions. As we did for the heat equation, we shall give an example of a set of boundary values, and leave others to the exercises. We shall ask that the two ends of the string have a specified displacement. By using the same argument as was used for the heat equation, we may as well suppose that this displacement is zero at each end. At $t = 0$ we also specify the initial displacement of the string, as well as its initial velocity. In the usual scenario you have in mind, the initial velocity is zero, but it could be nonzero. Putting this into the form of equations, we arrive at the boundary value problem

$$\begin{aligned} &\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} && u(0,t) = 0, \quad u(\ell,t) = 0 \\ & && u(x,0) = f(x), \quad \frac{\partial u}{\partial t}(x,0) = g(x). \end{aligned} \tag{6.5}$$

We shall see as we go along what restrictions are required for the functions f and g.

6.4.2.1 Formal solution In arriving at a formal solution for the wave equation, we shall follow a strategy quite similar to that used for the heat equation. That is to say, we shall follow Procedure 6.3.1. Indeed, we shall number the steps in this procedure, just to further illustrate how it is applied.

1. We must first decide which of the independent variables, x or t, will be the subject of our search for eigenvalues and eigenfunctions. It is pretty evident, given the boundary/initial conditions of (6.5), that we ought to use the x-variable as the one to which we will associate eigenvalues and eigenfunctions. Given the boundary conditions at $x = 0$ and $x = \ell$, we work with the same space

$$\mathsf{V} = \{\xi \colon [0, \ell] \to \mathbb{C} \mid \xi \text{ is infinitely differentiable, and } \xi(0) = \xi(\ell) = 0\}$$

   as with the heat equation.
2. Associated with the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2},$$

   we have the "ordinary differential equation"

$$\frac{\mathrm{d}^2 u}{\mathrm{d}t^2} = A(u(t)),$$

   where $A(u) = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2}$. Thus it is for $A \in \mathrm{L}(\mathsf{V}; \mathsf{V})$ that we find eigenvalues and eigenfunctions.
   The eigenvalues λ satisfy

$$\frac{\mathrm{d}^2 \xi}{\mathrm{d}x^2} = \lambda \xi(x), \quad \xi(0) = \xi(\ell) = 0.$$

   This is exactly the same eigenvalue/eigenfunction problem as for the heat equation, and so has eigenvalues $\lambda_n = -\frac{n^2\pi^2}{\ell^2}$ and eigenfunctions $\xi_n(x) = \sin(n\pi\frac{x}{\ell})$, $n \in \mathbb{Z}_{>0}$.
3. The eigenfunction transform is the same as we had for the heat equation:

$$\mathscr{E}(\xi)(n) = \int_0^\ell \xi(x) \sin(n\pi\tfrac{x}{\ell})\, \mathrm{d}x.$$

   It is also the case that

$$\mathscr{E}\left(\frac{\mathrm{d}^2 \xi}{\mathrm{d}x^2}\right)(n) = -\frac{n^2\pi^2}{\ell^2} \mathscr{E}(\xi)(n),$$

   just as for the heat equation.

4. We have
$$\hat{u}_n(t) = \int_0^\ell u(x,t)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x.$$
Thus
$$\mathscr{E}\left(\frac{\partial^2 u}{\partial t^2}\right)(n) = \mathscr{E}\left(c^2\frac{\partial^2 u}{\partial x^2}\right)(n) = -\frac{c^2n^2\pi^2}{\ell^2}\mathscr{E}(u)(n)$$
$$\implies \quad \frac{\mathrm{d}^2\hat{u}_n}{\mathrm{d}t^2} = -\frac{c^2n^2\pi^2}{\ell^2}\hat{u}_n, \qquad n \in \mathbb{Z}_{>0}.$$
5. We solve the preceding ordinary differential equation to get
$$\hat{u}_n(t) = c_n\cos(\tfrac{cn\pi}{\ell}t) + d_n\sin(\tfrac{cn\pi}{\ell}t).$$
6. We then have
$$u(x,t) = \sum_{n=1}^{\infty}(c_n\cos(\tfrac{cn\pi}{\ell}t) + d_n\sin(\tfrac{cn\pi}{\ell}t))\sin(n\pi\tfrac{x}{\ell}).$$
7. The initial conditions are then
$$u(x,0) = \sum_{n=1}^{\infty} c_n\sin(n\pi\tfrac{x}{\ell}) = f(x),$$
$$\frac{\partial u}{\partial t}(0,t) = \sum_{n=0}^{\infty} d_n\frac{cn\pi}{\ell}\sin(n\pi\tfrac{x}{\ell}) = g(x).$$
Therefore,
$$c_n = \frac{2}{\ell}\int_0^\ell f(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x,$$
$$d_n = \frac{4}{cn\pi}\int_0^\ell g(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x,$$
for $n \in \mathbb{Z}_{\geq 0}$. Thus we have the formal solution
$$u(x,t) = \sum_{n=1}^{\infty}\left(\left(\frac{2}{\ell}\int_0^\ell f(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x\right)\cos(\tfrac{cn\pi}{\ell}t) + \left(\frac{4}{cn\pi}\int_0^\ell g(x)\sin(n\pi\tfrac{x}{\ell})\,\mathrm{d}x\right)\sin(\tfrac{cn\pi}{\ell}t)\right)\sin(n\pi\tfrac{x}{\ell}). \quad (6.6)$$

6.4.2.2 Rigorous establishment of solutions Now let us examine the nature of the formal solution we obtained in the preceding section, and see how it functions as a solution to the boundary value problem (6.5). As we shall see, the nature of the result is a little different from the corresponding situation with the heat equation.

The main result is the following, which only deals with an initial displacement of the string. The reader is asked to consider the case where the initial velocity is nonzero in Exercise 6.4.7.

6.4.1 Theorem (Solutions for the wave equation on an interval) *Consider the boundary value problem*

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \qquad \begin{aligned} &u(0,t) = 0, \quad u(\ell,t) = 0 \\ &u(x,0) = f(x), \quad \frac{\partial u}{\partial t}(x,0) = 0. \end{aligned}$$

Let u *be the function defined by (6.6). If* $f\colon [0,\ell] \to \mathbb{R}$ *has the property that* $f_{\text{odd}}\colon \mathbb{R} \to \mathbb{R}$ *is twice continuously differentiable, then* u *is the unique solution satisfying the wave equation and all boundary conditions. Furthermore,*

$$u(x,t) = \frac{1}{2}(f_{\text{odd}}(x+ct) + f_{\text{odd}}(x-ct)).$$

Proof By ***missing stuff***, the series (6.6) for u_t converges uniformly in x for each t. Using the trigonometric identity

$$2\sin a \cos b = \sin(a+b) + \sin(a-b),$$

we note that

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} c_n \sin(n\pi \tfrac{x}{\ell}) \cos(\tfrac{cn\pi}{\ell} t) \\ &= \sum_{n=1}^{\infty} c_n \frac{1}{2}(\sin(n\pi \tfrac{x+ct}{\ell}) + \sin(n\pi \tfrac{x-ct}{\ell})) \\ &= \frac{1}{2}(f_{\text{odd}}(x+ct) + f_{\text{odd}}(x-ct)), \end{aligned}$$

as stated in the result. We now show that this function satisfies the wave equation and the boundary conditions. We compute

$$\frac{\partial^2 u}{\partial t^2} = \frac{1}{2}c^2(f''_{\text{odd}}(x+ct) + f''_{\text{odd}}(x-ct)), \quad \frac{\partial^2 u}{\partial t^2} = \frac{1}{2}(f''_{\text{odd}}(x+ct) + f''_{\text{odd}}(x-ct)),$$

so the wave equation is satisfied. We also have

$$u(0,t) = \frac{1}{2}(f_{\text{odd}}(ct) + f_{\text{odd}}(-ct)) = 0$$

since f_{odd} is odd. We also have

$$\begin{aligned} u(\ell,t) &= \frac{1}{2}(f_{\text{odd}}(\ell+ct) + f_{\text{odd}}(\ell-ct)) \\ &= \frac{1}{2}(f_{\text{odd}}(\ell+ct) - f_{\text{odd}}(-\ell+ct)) \\ &= \frac{1}{2}(f_{\text{odd}}(\ell+ct) - f_{\text{odd}}(\ell+ct)) \\ &= 0 \end{aligned}$$

since f_{odd} is periodic with period 2ℓ. Clearly $u(x,0) = f(x)$ and we also have

$$\frac{\partial u}{\partial t}(x,0) = \frac{1}{2}c(f'_{\text{odd}}(x) - f'_{\text{odd}}(x)) = 0,$$

verifying the zero velocity initial condition. ■

6.4.2 Remarks (Solutions for the wave equation)

1. The name "wave equation" comes from the characterisation of the solution as

$$u(x,t) = \frac{1}{2}(f_{\text{odd}}(x+ct) + f_{\text{odd}}(x-ct)).$$

With this characterisation, the solution is a superposition of two "travelling waves" moving with velocity c, one moving in the positive x-direction, and the other in the negative x-direction. Thus, unlike the heat equation where the initial condition is smoothed, the wave equation tends to simply propagate the initial condition.

2. Note that, unlike the heat equation, the smoothness of the solution of the wave equation is inherited from the initial condition f.

3. Given that the solution is defined explicitly in terms of the initial condition f, one is inclined to try to define solutions for initial condition functions f that are not twice continuously differentiable. Indeed, this is often done, and one by convention denotes the solution by

$$u(x,t) = \frac{1}{2}(f_{\text{odd}}(x+ct) + f_{\text{odd}}(x-ct)),$$

regardless of whether f is smooth enough to actually allow u to satisfy the wave equation itself. •

6.4.3 The wave equation for an infinite length string

6.4.3.1 Formal solution

6.4.3.2 Rigorous establishment of solutions

Exercises

6.4.1 Suppose that the string used in the derivation of the wave equation has a density that varies along the length of the string. Determine the partial differential equation governing the vertical displacement of the string. Ignore the effects of gravity and assume constant tension in the string.

6.4.2 In the derivation of the wave equation in describing the vertical displacement of a vibrating string, the effects of gravity are ignored.

(a) Derive the equations governing the vertical displacement of the string when gravity is considered. Assume that the density of the string and the tension in the string are independent of the displacement along the string.

(b) What is the steady-state displacement of the string in this case?

6.4.3 Suppose that a cable of length ℓ dangles vertically and that we wish to measure the horizontal displacement of the cable after an initial displacement. Let x denote the distance along the cable, with $x = 0$ denoting the bottom end of the cable.

(a) What is the tension in the cable as a function of x?

(b) Derive the partial differential equation governing the horizontal displacement of the cable, and setup a boundary value problem that describes the physical system.

6.4.4 In the derivation of the wave equation for the vibrating string, the effects of energy dissipation were neglected. A simple model for energy dissipation gives the equation

$$\frac{\mathrm{d}^2u}{\mathrm{d}t^2} + 2\delta\frac{\partial u}{\partial t} = c^2\frac{\partial^2 u}{\partial x^2}$$

describing the vertical displacement of the string, with $\delta > 0$ a constant.

(a) What is the appropriate boundary value problem given the same physical boundary conditions utilised in Section 6.4?

(b) Determine a formal solution for the boundary value problem of part (a). For simplicity, assume that $g = 0$, i.e., that the initial velocity of the string is zero. (Note that there are annoying complications that make the form of the solution depend on the size of δ.)

(c) How does the solution differ from the "travelling wave" character described by Theorem 6.4.1?

In the next exercise, you will show that the small longitudinal vibrations in a rod are governed by the wave equation. To do this, you need the following physical law.

Hooke's Law The stress, i.e., the pressure exerted by the rod's displacement, is proportional to the strain, the latter being given by $\frac{\partial u}{\partial x}$, with u the longitudinal displacement. •

6.4.5 Use Hooke's Law in combination with Newton's First Law of Motion to ascertain that the longitudinal vibrations in a rod satisfy the wave equation.

6.4.6 Consider the function $f\colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1, & x \in [\frac{3}{8}, \frac{5}{8}] \\ 0, & \text{otherwise.} \end{cases}$$

Following Remark 6.4.2–3, we let u denote the solution to the boundary value problem (6.5) with $c = 1$, with f as given, and with $g = 0$, despite the fact that f is not twice continuously differentiable.

(a) If $u_t\colon [0,1] \to \mathbb{R}$ is defined by $u_t(x) = u(x,t)$, plot u_t for $t \in \{0, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}, 1\}$.

(b) In what sense are there "travelling waves" in the solution?

6.4.7 Consider the wave equation boundary value problem (6.5), with zero initial displacement f and nonzero initial velocity g.

(a) Show that the displacement of the string satisfies the equation

$$u(x,t) = \frac{1}{2c}\left(\int_0^{x+ct} g(s)\,\mathrm{d}s - \int_0^{x-ct} g(s)\,\mathrm{d}s\right).$$

(b) What are the conditions on g which ensure that this will, in fact, be a solution of the boundary value problem?

(c) What is the solution of the boundary value problem when f and g are both nonzero?

6.4.8 Why do we require the initial displacement function for the wave equation to be twice continuously differentiable, whereas for the heat equation and the potential equation we can obtain a solution for boundary functions that are merely square integrable? (The idea of this question is that you understand the proofs of Theorems 6.3.2, 6.4.1, and 6.5.1 sufficiently well that you can extract the salient feature that answers the question.)

Section 6.5

The potential equation

The final equation we look at is the potential equation, the "elliptic" of our three representatives. We first considered this equation, illustrating various places where it arises, in Section 1.1.13. As with the heat and wave equations, we begin our discussion of the potential equation by discussing its relationship with characteristics. We then turn to solving the potential equation for a few different sorts of regions in the plane. As we did with the heat and wave equations, we obtain formal solutions using transform methods, and then prove that these are actually solutions to the problem.

6.5.1 Characteristics for the potential equation

6.5.2 The potential equation for a bounded rectangle

The first setting in which we consider the potential equation is a bounded rectangle

$$D = \{(x, y) \in \mathbb{R}^2 \mid x \in [0, a],\ y \in [0, b]\}.$$

The boundary conditions we specify will be on the boundary of D (generally, this need not be the case, although it often is). Thus we will consider boundary conditions of the form

$$u(0, y) = g_1(y), \quad u(a, y) = g_2(y), \quad u(x, 0) = f_1(x), \quad u(x, b) = f_2(x).$$

That is, we specify the value of u along the boundary of D. This gives the following boundary value problem:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad \begin{aligned} u(0, y) &= g_1(y), \quad u(a, y) = g_2(y) \\ u(x, 0) &= f_1(x), \quad u(x, b) = f_2(x). \end{aligned} \tag{6.7}$$

The boundary conditions we give are known as ***Dirichlet boundary conditions***. A specification of the value of the derivative of u along the boundary gives what are called ***Neumann boundary conditions***. Mixed conditions are also allowable, and those that specify that a linear combination of the displacement and the derivative be zero are called ***Robin boundary conditions***. Thus one can have a large set of possible boundary conditions for the simple potential equation defined on a rectangular domain. Although we stick with the Dirichlet boundary conditions, the reader may explore some alternatives in the exercises.

6.5.2.1 Formal solution We employ the by now venerable method of separation of variables. One can just go ahead and have at it, but it turns out that the easier

thing to do is to break the problem down into four boundary value problems:

$$\frac{\partial^2 u_1}{\partial x^2} + \frac{\partial^2 u_1}{\partial y^2} = 0 \qquad \begin{aligned} &u_1(0,y) = g_1(y), \quad u_1(a,y) = 0, \\ &u_1(x,0) = 0, \quad u_1(x,b) = 0; \end{aligned} \tag{6.8}$$

$$\frac{\partial^2 u_2}{\partial x^2} + \frac{\partial^2 u_2}{\partial y^2} = 0 \qquad \begin{aligned} &u_2(0,y) = 0, \quad u_2(a,y) = g_2(y), \\ &u_2(x,0) = 0, \quad u_2(x,b) = 0; \end{aligned} \tag{6.9}$$

$$\frac{\partial^2 u_3}{\partial x^2} + \frac{\partial^2 u_3}{\partial y^2} = 0 \qquad \begin{aligned} &u_3(0,y) = 0, \quad u_3(a,y) = 0, \\ &u_3(x,0) = f_1(x), \quad u_3(x,b) = 0; \end{aligned} \tag{6.10}$$

$$\frac{\partial^2 u_4}{\partial x^2} + \frac{\partial^2 u_4}{\partial y^2} = 0 \qquad \begin{aligned} &u_4(0,y) = 0, \quad u_4(a,y) = 0, \\ &u_4(x,0) = 0, \quad u_4(x,b) = f_2(x). \end{aligned} \tag{6.11}$$

Due to the linearity of the problem, if u_1, u_2, u_3, and u_4 satisfy the above boundary value problems, then it is quite clear that $u = u_1 + u_2 + u_3 + u_4$ satisfies (6.7). Thus we have traded one possibly annoying boundary value problem with four that we hope are simpler.

Let us obtain the solutions to the four problems, starting with that for u_1. We apply Procedure 6.3.1, indexing everything with the subscript "1" to denote that we are working with the first of the four boundary value problems.

1. For the problem (6.8) for u_1, the natural choice for of the "eigenfunction variable" is y, since the boundary condition at $x = 0$ involves the function g_1. Thus, given the boundary conditions at $y = 0$ and $y = b$, we take

$$\mathsf{V}_1 = \{\eta_1 \colon [0,b] \to \mathbb{C} \mid \eta_1 \text{ is infinitely differentiable, and } \eta_1(0) = \eta_1(b) = 0\}.$$

2. Associated to the potential equation

$$\frac{\partial^2 u_1}{\partial y^2} + \frac{\partial^2 u_1}{\partial x^2} = 0,$$

we have the "ordinary differential equation"

$$-\frac{\partial^2 u_1}{\partial x^2} = A(u_1(x)),$$

where $A(u_1) = \frac{\mathrm{d}^2 u_1}{\mathrm{d}y^2}$. Thus, as with the heat and wave equations, we find eigenvalues and eigenfunctions for $A \in \mathrm{L}(\mathsf{V}_1; \mathsf{V}_1)$.

Note that there are some different choices here, in terms of sign. There were also for the heat and wave equations; it's just that for the potential equation, there is a minus sign that one has to put somewhere. We could, for example, have used instead the "ordinary differential equation"

$$\frac{\partial^2 u_1}{\partial x^2} = A_1'(u_1(x)),$$

where $A_1'(u_1) = -\frac{\mathrm{d}^2 u_1}{\mathrm{d}y^2}$. Nothing would change with the final answer, of course. Just the place where the minus sign is handled would change.
In any case, working with A_1 as we have defined it, we have the same eigenvalues and eigenfunctions as for the heat and wave equations:

$$\lambda_{1,n} = -\frac{n^2\pi^2}{b^2}, \quad \eta_{1,n}(x) = \sin(n\pi\frac{x}{b}), \qquad n \in \mathbb{Z}_{>0}.$$

3. The eigenfunction transform is the same as we had for the heat and wave equations:

$$\mathscr{E}(\eta_1)(n) = \int_0^b \eta_1(y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y.$$

It is also the case that

$$\mathscr{E}\left(\frac{\mathrm{d}^2\eta_1}{\mathrm{d}y^2}\right)(n) = -\frac{n^2\pi^2}{b^2}\mathscr{E}(\eta_1)(n),$$

just as for the heat and wave equations.

4. We have

$$\hat{u}_{1,n}(x) = \int_0^b u_1(x,y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y.$$

Thus

$$-\mathscr{E}\left(\frac{\partial^2 u_1}{\partial x^2}\right)(n) = \mathscr{E}\left(\frac{\partial^2 u_1}{\partial y^2}\right)(n) = -\frac{n^2\pi^2}{b^2}\mathscr{E}(u_1)(n)$$

$$\implies \qquad \frac{\mathrm{d}^2\hat{u}_{1,n}}{\mathrm{d}x^2} = \frac{n^2\pi^2}{b^2}\hat{u}_{1,n}, \qquad n \in \mathbb{Z}_{>0}.$$

5. We solve the preceding ordinary differential equation to get

$$\hat{u}_{1,n}(x) = c_{1,n}\cosh(n\pi\tfrac{x}{b}) + d_{1,n}\sinh(n\pi\tfrac{x}{b}).$$

6. We then have

$$u_1(x,y) = \sum_{n=1}^{\infty}(c_{1,n}\cosh(n\pi\tfrac{x}{b}) + d_{1,n}\sinh(n\pi\tfrac{x}{b}))\sin(n\pi\tfrac{y}{b}).$$

7. The boundary conditions at $x = 0$ and $x = a$ are now employed:

$$u_1(0,y) = \sum_{n=1}^{\infty} c_{1,n}\sin(n\pi\tfrac{y}{b}) = g_1(y),$$

$$u_1(a,y) = \sum_{n=1}^{\infty}(c_{1,n}\cosh(n\pi\tfrac{a}{b}) + d_{1,n}\sinh(n\pi\tfrac{a}{b}))\sin(n\pi\tfrac{y}{b}) = 0.$$

Let us work with the second condition first. If we apply the eigenfunction transform to this condition, using the fact that

$$\int_0^b \sin(m\pi\tfrac{y}{b})\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y = \begin{cases} \frac{b}{2}, & n = m, \\ 0, & n \neq m. \end{cases}$$

we obtain that

$$c_{1,n}\cosh(n\pi\tfrac{a}{b}) + d_{1,n}\sinh(n\pi\tfrac{a}{b}) = 0, \qquad n \in \mathbb{Z}_{>0}.$$

This will be satisfied if we select

$$c_{1,n} = a_{1,n}\sinh\Big(n\pi\tfrac{a}{b}\Big), \quad d_{1,n} = -a_{1,n}\cosh\Big(n\pi\tfrac{a}{b}\Big),$$

for some as yet undetermined $a_{1,n} \in \mathbb{R}$. Now you dig into your bag of tricks for hyperbolic functions, and observe that

$$\sinh(\alpha-\beta) = \sinh(\alpha)\cosh(\beta) - \cosh(\alpha)\sinh(\beta)$$

(this can simply be verified directly). This gives

$$\begin{aligned} c_{1,n}\cosh(n\pi\tfrac{x}{b}) + d_{1,n}\sinh(n\pi\tfrac{x}{b}) & \\ = a_{1,n}\sinh\Big(n\pi\tfrac{a}{b}\Big)\cosh(n\pi\tfrac{x}{b}) - \cosh\Big(n\pi\tfrac{a}{b}\Big)\sinh(n\pi\tfrac{x}{b}) & \\ = a_{1,n}\sinh(n\pi\tfrac{a-x}{b}). & \end{aligned}$$

This gives

$$u_1(x,y) = \sum_{n=1}^{\infty} a_{1,n}\sinh(n\pi\tfrac{a-x}{b})\sin(n\pi\tfrac{y}{b}).$$

We now apply the boundary condition at $x = 0$:

$$a_{1,n}\sinh(n\pi\tfrac{a}{b}) = \frac{2}{b}\int_0^b g_1(y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y,$$

and finally obtain the formal solution for u_1:

$$u_1(x,y) = \sum_{n=1}^{\infty}\left(\frac{2}{b\sinh(n\pi\frac{a}{b})}\int_0^b g_1(y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y\right)\sinh(n\pi\tfrac{a-x}{b})\sin(n\pi\tfrac{y}{b}). \quad (6.12)$$

Now we apply Procedure 6.3.1 to obtain the solution u_2 for the boundary value problem (6.9).

1. Here, by the same reasoning as above for the boundary value problem associated with u_1, we work with an eigenvalue problem associated with the y-variable, and work with the vector space

$$\mathsf{V} = \{\eta_2 \colon [0,b] \to \mathbb{C} \mid \ \eta_2 \text{ is infinitely differentiable, and } \eta_2(0) = \eta_2(b) = 0\}.$$

2. Again, as with the u_1 boundary value problem, we have

$$-\frac{\partial^2 u_2}{\partial x^2} = A(u_2),$$

where $A(u_2) = \frac{\mathrm{d}^2 u_2}{\mathrm{d}x^2}$. We find the same eigenvalues and eigenvectors, of course:

$$\lambda_{2,n} = -\frac{n^2\pi^2}{b^2}, \quad \eta_{2,n}(x) = \sin(n\pi\frac{x}{b}), \qquad n \in \mathbb{Z}_{>0}.$$

3. The eigenfunction transform is the same as we had for the heat and wave equations:

$$\mathscr{E}(\eta_2)(n) = \int_0^b \eta_2(y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y.$$

It is also the case that

$$\mathscr{E}\left(\frac{\mathrm{d}^2\eta_2}{\mathrm{d}y^2}\right)(n) = -\frac{n^2\pi^2}{b^2}\mathscr{E}(\eta_2)(n),$$

just as for the heat and wave equations.

4. We have

$$\hat{u}_{2,n}(x) = \int_0^b u_2(x,y)\sin(n\pi\tfrac{y}{b})\,\mathrm{d}y.$$

Thus

$$-\mathscr{E}\left(\frac{\partial^2 u_2}{\partial x^2}\right)(n) = \mathscr{E}\left(\frac{\partial^2 u_2}{\partial y^2}\right)(n) = -\frac{n^2\pi^2}{b^2}\mathscr{E}(u_2)(n)$$

$$\implies \quad \frac{\mathrm{d}^2\hat{u}_{2,n}}{\mathrm{d}x^2} = \frac{n^2\pi^2}{b^2}\hat{u}_{2,n}, \qquad n \in \mathbb{Z}_{>0}.$$

5. We solve the preceding ordinary differential equation to get

$$\hat{u}_{2,n}(x) = c_{2,n}\cosh(n\pi\tfrac{x}{b}) + d_{2,n}\sinh(n\pi\tfrac{x}{b}).$$

6. We then have

$$u_2(x,y) = \sum_{n=1}^{\infty}(c_{2,n}\cosh(n\pi\tfrac{x}{b}) + d_{2,n}\sinh(n\pi\tfrac{x}{b}))\sin(n\pi\tfrac{y}{b}).$$

7. The boundary conditions at $x = 0$ and $x = a$ are now employed:

$$u_2(0,y) = \sum_{n=1}^{\infty} c_{2,n}\sin(n\pi\tfrac{y}{b}) = 0,$$

$$u_2(a,y) = \sum_{n=1}^{\infty}(c_{2,n}\cosh(n\pi\tfrac{a}{b}) + d_{2,n}\sinh(n\pi\tfrac{a}{b}))\sin(n\pi\tfrac{y}{b}) = g_2(y).$$

Let us work with the first condition first. If we apply the eigenfunction transform to this condition we immediately get $c_{2,n} = 0$, $n \in \mathbb{Z}_{>0}$. Then we apply the eigenfunction transform to the boundary condition at $x = a$ to get

$$d_{2,n} \sinh(n\pi \tfrac{a}{b}) = \frac{2}{b} \int_0^b g_2(y) \sin(n\pi \tfrac{y}{b}) \, \mathrm{d}y, \qquad n \in \mathbb{Z}_{>0}.$$

This gives the formal solution for u_2:

$$u_2(x, y) = \sum_{n=1}^{\infty} \left(\frac{2}{b \sinh(n\pi \frac{a}{b})} \int_0^b g_2(y) \sin(n\pi \tfrac{y}{b}) \, \mathrm{d}y \right) \sinh(n\pi \tfrac{x}{b}) \sin(n\pi \tfrac{y}{b}). \tag{6.13}$$

That is two of the four boundary value problems solved. The remaining two solutions can be obtained by a simple plausibility argument. Indeed, by symmetry in x and y, one can obtain the solutions for u_3 and u_4. Let us just write the answers. For u_3 we have

$$u_3(x, y) = \sum_{n=1}^{\infty} \left(\frac{2}{\sinh(n\pi \frac{b}{a}) a} \int_0^a f_1(x) \sin\left(n\pi \tfrac{x}{a}\right) \mathrm{d}x \right) \sinh\left(n\pi \tfrac{b-y}{a}\right) \sin\left(n\pi \tfrac{x}{a}\right). \tag{6.14}$$

and for u_4 we have

$$u_4(x, y) = \sum_{n=1}^{\infty} \left(\frac{2}{\sinh(n\pi \frac{b}{a}) a} \int_0^a f_2(x) \sin\left(n\pi \tfrac{x}{a}\right) \mathrm{d}x \right) c_{4,n} \sinh\left(n\pi \tfrac{y}{a}\right) \sin\left(n\pi \tfrac{x}{a}\right). \tag{6.15}$$

The formal solution to the boundary value problem (6.7) is then $u = u_1 + u_2 + u_3 + u_4$. Clearly, this will be a tedious solution to obtain in practise. If one adds to this the fact that one can additionally have boundary conditions that involve the derivative of u, one can see that there are myriad possibilities for solutions.

6.5.2.2 Rigorous establishment of solutions

For the Dirichlet problem, (6.7), we have the following result concerning the nature of the formal solution over which we laboured.

6.5.1 Theorem (Solutions for the potential equation on a bounded rectangle) *Consider the boundary value problem (6.7), and let* $\mathrm{u} = \mathrm{u}_1 + \mathrm{u}_2 + \mathrm{u}_3 + \mathrm{u}_4$ *be the series defined by equations (6.12), (6.13), (6.14), and (6.15). The following statements hold.*

(i) If $\mathrm{f}_1, \mathrm{f}_2 \in \mathrm{L}^2([0, \mathrm{a}]; \mathbb{C})$ *and* $\mathrm{g}_1, \mathrm{g}_2 \in \mathrm{L}^2([0, \mathrm{b}]; \mathbb{C})$*, then*

  (a) the series expression for u *converges uniformly on* $[\epsilon, \mathrm{a} - \epsilon] \times [\epsilon, \mathrm{b} - \epsilon]$ *for any* $\epsilon \in \mathbb{R}_{>0}$*,*

  (b) u *is infinitely differentiable on* $(0, \mathrm{a}) \times (0, \mathrm{b})$*,*

  (c) u *satisfies the potential equation in* $(0, \mathrm{a}) \times (0, \mathrm{b})$*, and*

(d) $\lim_{y\to 0}\|u_y - f_1\|_2 = 0$ $\lim_{y\to b}\|u_y - f_2\|_2 = 0$
$\lim_{x\to 0}\|u_x - g_1\|_2 = 0$ $\lim_{x\to a}\|u_x - f_2\|_2 = 0$,
where $u_x\colon [0,b] \to \mathbb{R}$ *and* $u_y\colon [0,a] \to \mathbb{R}$ *are defined by* $u_x(y) = u_y(x) = u(x,y)$.

(ii) *If* f_1, f_2, g_1, *and* g_2 *are twice continuously differentiable and satisfy*

$$f_1(0) = f_1(a) = f_2(0) = f_2(a) = g_1(0) = g_1(b) = g_2(0) = g_2(b) = 0,$$

then

(a) *the series for* u *converges uniformly,*
(b) u *in continuous on* $[0,a]\times[0,b]$,
(c) u *is infinitely differentiable on* $(0,a)\times(0,b)$,
(d) u *satisfies the potential equation on* $(0,a)\times(0,b)$, *and*
(e) u *satisfies the boundary conditions.*

Proof For simplicity, we assume that $a = b = \pi$ and that $g_2 = f_1 = f_2 = 0$. The first assumption we can make by a change of independent variable, if necessary. The second assumption we can make by linearity, as if part (ii) holds for all of u_1, u_2, u_3, and u_4, it will hold for their sum.

(i) Consider the series solution for u:

$$u(x,t) = \sum_{n=1}^{\infty} \frac{c_n}{\sinh(n\pi)} \sinh(n(\pi - x))\sin(ny),$$

where

$$c_n = \frac{2}{\pi}\int_0^{\pi} g_1(y)\sin(ny)\,dy.$$

We will show that this series converges uniformly on $[\epsilon,\pi]\times[0,\pi]$ for any $\epsilon \in \mathbb{R}_{>0}$. One easily sees that

$$\sinh(n(\pi - x)) \le \sinh(n(\pi - \epsilon)), \qquad n \in \mathbb{Z}_{>0},\ x \in [\epsilon,\pi].$$

An easy application of the definition of sinh gives the estimates

$$2\sinh(n(\pi-\epsilon)) < e^{n(\pi-\epsilon)}, \quad 2\sinh(n\pi) \ge e^{n\pi}(1 - e^{-n\pi})$$

for any $\epsilon > 0$. Thus

$$\frac{\sinh(n(\pi - x))}{\sinh(n\pi)} \le \frac{\sinh(n(\pi-\epsilon))}{\sinh(n\pi)} \le \frac{e^{-n\epsilon}}{1 - e^{-2\pi}}$$

for $n \in \mathbb{Z}_{>0}$ and $x \in [\epsilon,\pi]$. By definition of c_n and since $g_1 \in L^2([0,\pi];\mathbb{C})$, there exists $M \in \mathbb{R}_{>0}$ such that $|c_n| \le M$ for $n \in \mathbb{Z}_{>0}$. Therefore,

$$\left|\frac{c_n}{\sinh(n\pi)}\sinh(n(\pi - x))\sin(ny)\right| \le \frac{M}{1 - e^{-2\pi}}e^{-n\epsilon}, \qquad n \in \mathbb{Z}_{>0},\ (x,y) \in [\epsilon,\pi]\times[0,\pi].$$

One can now use the same arguments involving the Weierstrass M-test as in the proof of Theorem 6.3.2(i) to prove uniform convergence of the series definition u on $[\epsilon, \pi] \times [0, \pi]$. Moreover, the arguments from the proof of Theorem 6.3.2(iii) also apply here to show that u is infinitely differentiable on $(0, \pi) \times (0, \pi)$. Also, one can swap derivatives and sums, just as in the proof of Theorem 6.3.2(ii) to show that u satisfies the potential equation on $(0, \pi) \times (0, \pi)$ and satisfies the boundary conditions at $y = 0$, $y = \pi$, and $x = \pi$. Finally, the argument from Theorem 6.3.2(iv) involving uniform convergence (as a function of x) of the series for $\|u_x - g_1\|_2$ implies that $\lim_{x\to 0} \|u_x - g_1\|_2 = 0$.

(ii) The parts that do not follow from part (i) are parts (ii a), (ii b), and (ii e). By ***missing stuff*** (applied to g_1 and g_1'), the two series

$$\sum_{n=1} c_n \sin(ny), \quad \sum_{n=1} nc_n \sin(nx)$$

converge uniformly to g_1 and g_1', respectively. We prove the uniform convergence of the series for u using Abel's Test with $f_n(x, y) = c_n \sin(ny)$ and $g_n(x, y) = \frac{\sinh(n(\pi - x))}{\sinh(n\pi)}$, $n \in \mathbb{Z}_{>0}$. To show that the hypotheses of Abel's Test apply, we must show that $g_{n+1}(x, y) \leq g_n(x, y)$, $n \in \mathbb{Z}_{>0}$. This follows from the following lemma.

1 Lemma *If $\beta > 0$ and $\beta \geq \alpha$ and*

$$\psi(\mathrm{x}) = \frac{\sinh(\alpha \mathrm{x})}{\sinh(\beta \mathrm{x})},$$

then $\psi'(\mathrm{x}) \leq 0$ for $\mathrm{x} \geq 0$.

Proof This is pure trickery. We compute

$$\begin{aligned}\sinh^2(\beta x)\psi'(x) &= \alpha \sinh(\beta x)\cosh(\alpha x) - \beta \cosh(\beta x)\sinh(\alpha x)\\ &= -\frac{\beta^2 - \alpha^2}{2}\Big(\frac{\sinh((\alpha + \beta)x)}{\alpha + \beta} - \frac{\sinh((\beta - \alpha)x)}{\beta - \alpha}\Big),\end{aligned}$$

where we have used $\sinh(\xi + \eta) = \sinh(\xi)\cosh(\eta) + \cosh(\xi)\sinh(\eta)$, and have skipped some steps. Now note that if we define

$$\rho(x) = \frac{\sinh((\alpha + \beta)x)}{\alpha + \beta} - \frac{\sinh((\beta - \alpha)x)}{\beta - \alpha},$$

then $\rho(0) = 0$ and

$$\rho'(x) = \sinh((\alpha + \beta)x) - \sinh((\beta - \alpha)x),$$

which is positive since $\beta \geq \alpha$ and since sinh is an increasing function. Thus $\rho(x) \geq 0$ for $x \geq 0$. This, along with the fact that $\beta \geq \alpha$, ensures that $\psi'(x) \leq 0$ for $x \geq 0$ as claimed. ▼

Now, since $g_1(x, y) \leq 1$, it follows that the sequence $(g_n)_{n\in\mathbb{Z}_{>0}}$ is uniformly bounded. This shows that the series for u converges uniformly by Abel's test. Thus gives parts (ii a) and (ii b). The continuity of u at the boundaries follows as per Remark 6.3.3–2. That u satisfies the boundary conditions is trivial. ■

6.5.2 Remarks (Solutions for the potential equation)

1. The solution to the Dirichlet problem shares many of the features of the heat equation in terms of smoothing the boundary functions. The reason this works is the presence in the series solutions of the hyperbolic sine function, which near the boundary behaves like a negative exponential. This gives the smoothing property of the coefficients that we use to employ the Weierstrass M-test infinitely often.
2. As interesting as is the infinite smoothness of the solutions to the potential equation on the interior of the domain, they are further analytic. This property of the potential equation contributes to (or arises from, depending on your point of view) the value of the potential equation in the nontrivial subject of complex potential theory. •

6.5.3 The potential equation for a semi-unbounded rectangle

6.5.3.1 Formal solution

6.5.3.2 Rigorous establishment of solutions

6.5.4 The potential equation for an unbounded rectangle

6.5.4.1 Formal solution

6.5.4.2 Rigorous establishment of solutions

Exercises

6.5.1 The steady-state heat distribution in a disk of radius R is governed by the potential equation, and due to the geometry of the problem, it is convenient to use polar coordinates to describe the problem.

(a) Write the potential equation in polar coordinates (r, θ) defined by $x = r\cos\theta$ and $y = r\sin\theta$.

(b) Suppose that the heat flow from the outer edge of the disk is specified by a function $f(\theta)$. Mathematically express the boundary condition determined by this physical description.

(c) Why should f satisfy the constraint

$$\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta = 0?$$

(d) If $f \in \mathrm{L}_2([0, 2\pi]; \mathbb{R})$, describe the nature of the temperature distribution on the interior of the disk, given what you know about the behaviour of the solution to the potential equation.

6.5.2 Consider a heat exchanger tube with inner radius R_0 and outer radius R_1. Suppose that at steady state the temperature of the fluid inside the tube is T_0 and outside the tube is T_1. The temperature distribution in the tube as it varies from the inner wall to the outer wall is governed by the potential equation.

(a) Write the potential equation in polar coordinates (r, θ) defined by $x = r\cos\theta$ and $y = r\sin\theta$.

(b) Using the above description of the problem, setup a boundary value problem describing the temperature distribution in the heat exchanger tube as it varies from the inner wall to the outer wall.

(c) Argue that the solution will be independent of θ, and use this conclusion to obtain the desired distribution of temperature.

Section 6.6

Weak solutions of partial differential equations

Chapter 7

Second-order boundary value problems

Contents

Section 7.1

Linear maps on Banach and Hilbert spaces

It is expected that students using this text will have at least one course in linear algebra, and so will be acquainted with basic concepts from the subject. However, we will push quite far beyond the *basic* ideas in linear algebra, as this is necessitated by the subject matter. While some of the basic concepts of linear algebra in finite-dimensions will provide some valuable intuition, it will be important to remember that things can get significantly more complicated in infinite-dimensions. Furthermore, as we shall see in Sections 7.2 and 7.3, these differences have meaning in terms of physical applications.

7.1.1 Linear maps on normed vector spaces

In the preceding introductory discussion, the properties of linear maps were purely algebraic. In applications, to ensure that one is doing something meaningful, one also needs to pay attention to matters of convergence and continuity, thus necessitating a discussion of the rôle played by norms in discussing linear maps. As we shall see, the basic dichotomy is that between continuous and discontinuous linear maps, with the discontinuous version being the one of most interest.

7.1.1.1 Continuous linear maps If $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ are normed vector spaces, then one can certainly have the usual notion for linear maps from U to V. However, as U and V have norms defined on them, there are additional notions one can define. In order to put these notions into context, it is useful to talk about maps of a general nature from U to V, or perhaps more generally, from subsets of U to V. To this end, if $A \subseteq \mathsf{U}$ is an open set, a map $\phi \colon A \to \mathsf{V}$ is ***continuous at*** $\mathbf{u_0}$ if for each $\epsilon > 0$ there exists $\delta > 0$ so that $\|u - u_0\|_{\mathsf{U}} < \delta$ implies that $\|\phi(u) - \phi(u_0)\|_{\mathsf{V}} < \epsilon$. The map ϕ is ***continuous*** if it is continuous at each point $u \in A$. Note that this generalises the usual notion of continuity for functions on $\mathbb{R}$, or more generally for functions on $\mathbb{R}^n$. For linear maps, we say that $L \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ is ***bounded*** if there exists $M > 0$ so that

$$\|L(u)\|_{\mathsf{V}} \leq M\|u\|_{\mathsf{U}}$$

for every $u \in \mathsf{U}$. In finite-dimensions, it is quite easy to show that *all* linear maps are bounded (Example 7.1.3). However, in infinite-dimensions, this is not so, as exhibited by the following example.

7.1.1 Example Let $C^1([0,1];\mathbb{R})$ denote the set of continuously differentiable functions on the interval $[0,1]$, and define a linear map $L \colon C^1([0,1];\mathbb{R}) \to C^0([0,1];\mathbb{R})$ by

$L(f) = f'$. On each vector space we use the norm[1]

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|.$$

We claim that L is not bounded. We do so by showing that for any $M > 0$ there exists a function f_M with the property that $\|f_M'\|_\infty > M\|f_M\|$. We proceed as follows. For $M > 0$ let $\epsilon > 0$ have the property that

$$|\frac{\cos\frac{1}{\epsilon}}{\epsilon^2}| > M. \tag{7.1}$$

Since $|\cos\frac{1}{x}| < 1$, it follows that there is some ϵ sufficiently small that (7.1) holds. Now define

$$f_M(x) = \begin{cases} -\epsilon^{-3}\Big(\cos\frac{1}{\epsilon} + \epsilon\sin\frac{1}{\epsilon}\Big)x^2 + \epsilon^{-2}\Big(\cos\frac{1}{\epsilon} + 2\epsilon\sin\frac{1}{\epsilon}\Big)x, & x \in [0,\epsilon) \\ \sin\frac{1}{x}, & x \in [\epsilon, 1]. \end{cases}$$

One may verify that $f_M \in C^1([0,1];\mathbb{R})$. Indeed, f_M has been designed so that its graph on $[0,\epsilon]$ is a parabola connecting the point $(0,0)$ with the point $(\epsilon, \sin\frac{1}{\epsilon})$, and does so in a way that the derivative agrees with that of $\sin\frac{1}{x}$ at $x = \epsilon$. In any event, the essential fact is that $\|f_M\|_\infty = 1$ and that $\|f_M'\|_\infty > M$. This shows that L is not bounded. •

Thus not all linear maps are bounded, and our example shows that some not very exotic linear maps are actually unbounded. This is why the study of unbounded linear operators is a useful subject, and is undertaken in Section 7.1.1.2. This notwithstanding, the following result gives an interesting characterisation of bounded linear maps.

7.1.2 Theorem *If* $(\mathsf{U}, \|\cdot\|_\mathsf{U})$ *and* $(\mathsf{V}, \|\cdot\|_\mathsf{V})$ *are normed vector spaces then* $\mathrm{L} \in \mathrm{L}(\mathsf{U};\mathsf{V})$ *is bounded if and only if it is continuous.*

Proof First let us show that L is continuous if and only if it is continuous at 0. This means that we need to show that L is continuous if and only if for every $\epsilon > 0$ there exists $\delta > 0$ so that $\|L(u) - L(0)\|_\mathsf{V} = \|L(u)\|_\mathsf{V} < \epsilon$ provided that $\|u - 0\|_\mathsf{U} = \|u\|_\mathsf{U} < \delta$. Clearly, only the "if" part of this statement has nontrivial content. Thus assume that L is continuous at 0, so it is desired to show that L is continuous at any $u_0 \in \mathsf{U}$. The definition of continuity at u_0 says that for every $\epsilon > 0$ there exists $\delta > 0$ so that

$$\|u - u_0\|_\mathsf{U} < \delta \quad \Longrightarrow \quad \|L(u) - L(u_0)\|_\mathsf{V} < \epsilon.$$

[1]Note that this is indeed a norm on both $C^0([0,1];\mathbb{R})$ and $C^1([0,1];\mathbb{R})$, but that with this norm, $C^1([0,1];\mathbb{R})$ is not a Banach space as $C^0([0,1];\mathbb{R})$ is. To make $C^1([0,1];\mathbb{R})$ a Banach space, one could use the norm defined by

$$\|f\|_\infty^1 = \sup_{x\in[0,1]} f(x) + \sup_{x\in[0,1]} f'(x).$$

However, for the purposes of this example, we do not really care that $C^1([0,1];\mathbb{R})$ is a Banach space.

This simply means that $L(\mathsf{B}(\delta, u_0)) \subseteq \mathsf{B}(\epsilon, L(u_0))$. We have

$$\begin{aligned}\mathsf{B}(\delta, u_0) &= \{u \in \mathsf{U} \mid \ \|u - u_0\|_{\mathsf{U}} < \epsilon\}\\ &= \{u_0 + u \mid \ \|u\|_{\mathsf{U}} < \delta\}\\ &= \{u_0 + u \mid \ u \in \mathsf{B}(\delta, 0)\},\end{aligned}$$

and similarly

$$\mathsf{B}(\epsilon, L(u_0)) = \{L(u_0) + v \mid \ v \in \mathsf{B}(\epsilon, 0)\}.$$

Now note that

$$\begin{aligned}L(\mathsf{B}(\delta, u_0)) &= \{L(u_0 + u) \mid \ u \in \mathsf{B}(\delta, 0)\}\\ &= \{L(u_0) + L(u) \mid \ u \in \mathsf{B}(\delta, 0)\}.\end{aligned}$$

Therefore, if $\epsilon > 0$ and we choose $\delta > 0$ so that $L(\mathsf{B}(\delta, 0)) \subseteq \mathsf{B}(\epsilon, 0)$ then we have $L(\mathsf{B}(\delta, u_0)) \subseteq \mathsf{B}(\epsilon, L(u_0))$, showing that L is continuous at u_0, as desired.

First suppose that L is bounded with $M > 0$ having the property that $\|L(u)\|_{\mathsf{V}} \leq M\|u\|_{\mathsf{U}}$. For $\epsilon > 0$ take $\delta = \frac{\epsilon}{M}$. For $\|u\|_{\mathsf{U}} < \delta$ we then have

$$\|L(u)\| \leq M\|u\|_{\mathsf{U}} = \epsilon,$$

showing that L is continuous at 0, and hence continuous.

Now suppose that L is continuous. Then L is continuous at $0 \in \mathsf{U}$. Fix $\delta > 0$ satisfying

$$\|u\|_{\mathsf{U}} \leq \delta \quad \Longrightarrow \quad \|L(u)\|_{\mathsf{V}} \leq 1.$$

Then for any $u \in \mathsf{U} \setminus \{0\}$ we have

$$\begin{aligned}\|L(u)\|_{\mathsf{V}} &= \Big\|\tfrac{\|u\|_{\mathsf{U}}}{\delta} L\Big(\tfrac{\delta}{\|u\|_{\mathsf{U}}} u\Big)\Big\|\\ &= \frac{\|u\|_{\mathsf{U}}}{\delta}\Big\|L\Big(\tfrac{\delta}{\|u\|_{\mathsf{U}}} u\Big)\Big\|\\ &\leq \frac{\|u\|_{\mathsf{U}}}{\delta},\end{aligned}$$

since $\Big\|\frac{\delta}{\|u\|_{\mathsf{U}}} u\Big\|_{\mathsf{U}} = \delta$. This shows that L is bounded with $M = \frac{1}{\delta}$. ■

Let us give some examples of continuous linear maps.

7.1.3 Examples

1. Let us show that a linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ is continuous. Suppose that the linear map is represented by the $n \times m$ matrix $\boldsymbol{L}$. Let $C > 0$ have the property that $|L_{ij}| < C$ for $i, j \in \{1, \ldots, n\}$. For $\boldsymbol{x} \in \mathbb{R}^n$ let $j_x \in \{1, \ldots, m\}$ have the property that

$$|x^{j_x}| = \sup_{j \in \{1, \ldots, m\}} |x^j|.$$

Then we have

$$\Big(\sup_{j\in\{1,\dots,m\}} |x^j|\Big)^2 = (x^{j_x})^2 \leq \|\boldsymbol{x}\|^2.$$

We then have

$$\begin{aligned}\|\boldsymbol{L}\boldsymbol{x}\|^2 &= \sum_{i=1}^m \Big(\sum_{j=1}^n L_{ij}x^j\Big)^2\\ &\leq \sum_{i=1}^m \Big(\sum_{j=1}^n C|x^j|\Big)^2\\ &\leq \sum_{i=1}^m C^2 n\Big(\sup_{j\in\{1,\dots,n\}} |x^j|\Big)^2\\ &\leq \sum_{i=1}^m C^2 n\|\boldsymbol{x}\|^2.\end{aligned}$$

Thus we have $\|\boldsymbol{L}\boldsymbol{x}\| \leq M\boldsymbol{x}$ if we take $M = C\sqrt{n}$.

2. The linear map L from $C^0([0,1];\mathbb{R})$ to $C^0([0,1];\mathbb{R})$ defined by

$$L(f)(x) = \int_0^x f(\xi)\,\mathrm{d}\xi$$

is, we claim, continuous in the $\|\cdot\|_\infty$ norm. Since $f \in C^0([0,1];\mathbb{R})$ is continuous, it is bounded. Thus there exists $M > 0$ so that $|f(x)| \leq M$ for each $x \in [0,1]$. Then we have

$$|L(f)(x)| = \Big|\int_0^x f(\xi)\,\mathrm{d}\xi\Big| \leq \int_0^x |f(\xi)\,\mathrm{d}\xi| \leq Mx.$$

Therefore we have $\|L(f)\|_\infty \leq \|f\|_\infty$, showing that L is bounded and so continuous.

3. Let us show that the preceding linear map is also bounded on the normed vector space $(\mathrm{L}_2([0,1];\mathbb{F}, \|\cdot\|_2)$. For $f \in \mathrm{L}_2([0,1];\mathbb{F})$ we compute

$$\begin{aligned}|L(f)(x)|^2 &= \Big|\int_0^x f(\xi)\,\mathrm{d}\xi\Big|^2\\ &\leq \Big|\int_0^x \mathrm{d}\xi\Big|\Big|\int_0^x |f(\xi)|^2\,\mathrm{d}\xi\Big|\\ &\leq x\|f\|_2^2,\end{aligned}$$

where we have used the Cauchy-Bunyakovsky-Schwartz inequality. From this we compute

$$\|L(f)\|_2^2 = \int_0^1 |L(f)(\xi)|^2\,\mathrm{d}\xi \leq \tfrac{1}{2}\|f\|_2^2,$$

thus showing that L is bounded, and so continuous. •

A useful property of continuous linear maps is that the set of all such things is itself a normed vector space. We have already seen in the preamble of this chapter that it is a vector space, so we need only provide a norm for the set of continuous linear maps from $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ to $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$. This is given to us by the following result.

7.1.4 Theorem *If* $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ *to* $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ *are normed vector spaces, and* $\mathrm{L}(\mathsf{U}; \mathsf{V}) \subseteq \mathrm{L}(\mathsf{U}; \mathsf{V})$ *denotes the set of continuous linear maps, then, for* $\mathrm{L} \in \mathrm{L}(\mathsf{U}; \mathsf{V})$*, the object* $\|\mathrm{L}\|_{\mathsf{U},\mathsf{V}}$ *defined by*

$$\|\mathrm{L}\|_{\mathsf{U},\mathsf{V}} = \sup_{\|u\|_{\mathsf{U}}=1} \|\mathrm{L}(u)\|_{\mathsf{V}}$$

defines a norm on $\mathrm{L}(\mathsf{U}; \mathsf{V})$*. Furthermore, this norm is complete if* $\|\cdot\|_{\mathsf{V}}$ *is complete.*

Proof The norm properties ***missing stuff*** and ***missing stuff*** are easily seen. Also, for any u of norm 1 we have $\|L(u)\|_{\mathsf{V}} \leq \|L\|_{\mathsf{U},\mathsf{V}}$ giving for any $u \in \mathsf{U}$, $\|L(u)\|_{\mathsf{V}} \leq \|L\|_{\mathsf{U},\mathsf{V}}\|u\|_{\mathsf{U}}$. Therefore, if $\|L\|_{\mathsf{U},\mathsf{V}} = 0$ it follows that $\|L(u)\|_{\mathsf{V}} = 0$ for all $u \in \mathsf{U}$, so that $L = 0$, thus verifying ***missing stuff***. For the triangle inequality we have

$$\begin{aligned}\|L_1 + L_2\|_{\mathsf{U},\mathsf{V}} &= \sup_{\|u\|=1} \|L_1(u) + L_2(v)\|_{\mathsf{V}} \\ &\leq \sup_{\|u\|=1} \|L_1(u)\| + \sup_{\|u\|=1} \|L_2(u)\| \\ &= \|L_1\| + \|L_2\|.\end{aligned}$$

This verifies that $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ is indeed a norm.

Now suppose that V is a Banach space and let $\{L_j\}_{j\in\mathbb{N}}$ be a Cauchy sequence in $\mathrm{L}(\mathsf{U}; \mathsf{V})$ with respect to the norm $\|\cdot\|_{\mathsf{U},\mathsf{V}}$. We claim that the sequence $\{L_j(u)\}_{j\in\mathbb{N}}$ is a Cauchy sequence for each $u \in \mathsf{U}$. Indeed, we have $\|L_j(u) - L_k(u)\|_{\mathsf{V}} \leq \|L_j - L_k\|_{\mathsf{U},\mathsf{V}}\|u\|$ for all $j, k \in \mathbb{N}$, from which the claim follows. Therefore, for each $u \in \mathsf{U}$ the sequence $\{L_j(u)\}_{j\in\mathbb{N}}$ converges. Let us denote by $L(u) \in \mathsf{V}$ the vector to which the sequence converges, thus defining a map $L\colon \mathsf{U} \to \mathsf{V}$. It is easy to see that the map L is linear.

We still must show that L is continuous and that the sequence $\{L_j(u)\}_{j\in\mathbb{N}}$ converges to L. Let $\epsilon > 0$ and let $N \in \mathbb{N}$ have the property that $\|L_j - L_k\|_{\mathsf{U},\mathsf{V}} < \epsilon$ for all $j, k \geq N$. If $\|u\| = 1$ then we have $\|L_j(u) - L_k(u)\|_{\mathsf{V}} < \epsilon$ giving

$$\lim_{k\to\infty} \|L_j(u) - L_k(u)\|_{\mathsf{V}} = \|L_j(u) - L(u)\|_{\mathsf{V}} < \epsilon,$$

this latter holding for all $\|u\| = 1$. This shows that the linear map $L_j - L$ maps open balls around 0 to bounded balls around 0, and as we saw in Theorem 7.1.2, this means exactly that $L_j - L$ is continuous. Since L_j is continuous, this implies that L is itself continuous. That the sequence $\{L_j\}_{j\in\mathbb{N}}$ converges to L also follows from this computation. ■

The norm $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ is sometimes called the ***operator norm***.

For the linear maps of Example 7.1.3, we may explicitly compute the resulting operator norms.

7.1.5 Example (Example 7.1.3 cont'd)

1. We first consider the case of a linear map $\boldsymbol{L}\colon \mathbb{R}^m \to \mathbb{R}^n$. We claim that $\|\boldsymbol{L}\|_{\mathbb{R}^m,\mathbb{R}^n}$ is equal to the largest eigenvalue of the matrix $\boldsymbol{L}^T\boldsymbol{L}$. First note that $\boldsymbol{L}^T\boldsymbol{L}$ is a symmetric matrix, so its eigenvalues are all real. Furthermore, its eigenvalues are nonnegative since $\boldsymbol{L}^T\boldsymbol{L}$ is positive-definite (we will be discussing such notions in greater generality in Section 7.1.3.2). Let $\boldsymbol{x} \in \mathbb{R}^m$ be an eigenvector for the largest eigenvalue λ of $\boldsymbol{L}^T\boldsymbol{L}$. We then compute

$$\|\boldsymbol{L}\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{x} = \lambda^2\boldsymbol{x}^T\boldsymbol{x} = \lambda^2\|\boldsymbol{x}\|^2.$$

   This shows that $\|\boldsymbol{L}\|_{\mathbb{R}^m,\mathbb{R}^n} \geq \lambda$. Now note that since $\boldsymbol{L}^T\boldsymbol{L}$ is symmetric we may find an orthonormal basis $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ for $\mathbb{R}^m$ comprised of eigenvectors of $\boldsymbol{L}^T\boldsymbol{L}$. We may then write

$$\boldsymbol{x} = (\boldsymbol{x}\cdot\boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{x}\cdot\boldsymbol{v}_m)\boldsymbol{v}_m$$

   for any $\boldsymbol{x} \in \mathbb{R}^m$. We then have

$$\begin{aligned}
\|\boldsymbol{L}\boldsymbol{x}\|^2 &= \boldsymbol{x}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{x}\\
&= \sum_{i,j=1}^m (\boldsymbol{x}\cdot\boldsymbol{v}_i)(\boldsymbol{x}\cdot\boldsymbol{v}_j)\boldsymbol{v}_i^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{v}_j\\
&= \sum_{i,j=1}^m (\boldsymbol{x}\cdot\boldsymbol{v}_i)(\boldsymbol{x}\cdot\boldsymbol{v}_j)\lambda_i\lambda_j\boldsymbol{v}_i^T\boldsymbol{v}_i\\
&= \sum_{i=1}^m \lambda_i^2(\boldsymbol{x}\cdot\boldsymbol{v}_i)^2 \leq \lambda^2\sum_{i=1}^m (\boldsymbol{x}\cdot\boldsymbol{v}_i)^2 = \lambda_2\|\boldsymbol{x}\|^2,
\end{aligned}$$

   thus showing that $\|\boldsymbol{L}\|_{\mathbb{R}^m,\mathbb{R}^n} \leq \lambda$.

2. For the linear map $L(f)(x) = \int_0^x f(\xi)\,\mathrm{d}\xi$ defined on $(C^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$ we claim that the operator norm is 1. In Example 7.1.3 we showed that the operator norm is at least 1. To show that it is at most 1, consider the function $f(x) = c$ for some nonzero constant c. We then have $\|L(f)\|_\infty = c$, giving our assertion. •

7.1.1.2 Linear operators In our applications of discontinuous linear maps, it will be useful to consider linear maps defined not on an entire vector space, but only on a subspace. Thus, given a normed vector space $(\mathsf{V}, \|\cdot\|)$, a ***linear operator on*** V is a pair $(L, \mathrm{dom}(L))$ where $\mathrm{dom}(L) \subseteq \mathsf{V}$ is a subspace and $L \in \mathrm{L}(\mathrm{dom}(L); \mathsf{V})$. Two linear operators $(L_1, \mathrm{dom}(L_1))$ and $(L_2, \mathrm{dom}(L_2))$ on V are ***equal*** if $\mathrm{dom}(L_1) = \mathrm{dom}(L_2)$ and $L_1 = L_2$. We write $(L_1, \mathrm{dom}(L_1)) = (L_2, \mathrm{dom}(L_2))$ if the linear operators are equal. A linear operator $(L_1, \mathrm{dom}(L_1))$ is an ***extension*** of a linear operator $(L_2, \mathrm{dom}(L_2))$ if $\mathrm{dom}(L_2) \subseteq \mathrm{dom}(L_1)$ and if $L_1|\,\mathrm{dom}(L_2) = L_2$. We write $(L_2, \mathrm{dom}(L_2)) \subseteq (L_1, \mathrm{dom}(L_1))$ is $(L_1, \mathrm{dom}(L_1))$ is an extension of $(L_2, \mathrm{dom}(L_2))$.

We are generally interested in the case when L is discontinuous and when $\mathrm{dom}(L)$ is dense in V. In this case there are additional refinements one can impose

on an linear operator that may not be continuous. A useful weaker notion is that of closedness.

7.1.6 Definition Let $(L, \mathrm{dom}(L))$ be a linear operator on V. A Cauchy sequence $\{v_j\}_{j\in\mathbb{N}} \subseteq \mathrm{dom}(L)$ is ***compatible*** with L if the sequence $\{A(v_j)\}_{j\in\mathbb{N}}$ converges. The linear operator $(L, \mathrm{dom}(L))$ is ***closed*** if for every sequence $\{v_j\}_{j\in\mathbb{N}}$ compatible with L we have

(i) $v_0 = \lim_{j\to\infty} v_j \in \mathrm{dom}(L)$ and

(ii) $L(v_0) = \lim_{j\to\infty} L(v_j)$. •

The distinction between continuity and closedness is subtle. First let us show that certain continuous linear operators are closed.

7.1.7 Proposition *A linear operator* (L, dom(L)) *on* V *with* dom(L) *closed and with* L *continuous is closed.*

Proof We note that since $\mathrm{dom}(L)$ is closed, *every* convergent sequence in $\mathrm{dom}(L)$ converges to a point in $\mathrm{dom}(L)$. Thus condition (i) in Definition 7.1.6 is always satisfied. What's more, since the image of a convergent sequence under a continuous function is convergent (Exercise 7.1.5), it also follows that condition (ii) of Definition 7.1.6 holds under the hypotheses of the proposition. ■

The way in which a closed linear operator may not be continuous is this: while for a continuous linear operator we know that if $\{v_j\}_{j\in\mathbb{N}}$ converges to v_0 then $\{L(v_j)\}_{j\in\mathbb{N}}$ converges to $L(v_0)$, all we know for closed linear operators is that different sequences converging to the same point in $\mathrm{dom}(L)$ will have images under L converging to the same point in V.

It turns out that we will naturally encounter discontinuous linear operators that, while they are not closed, they are nearly so. Let us describe these "nearly closed" linear operators. Call a sequence $\{v_j\}_{j\in\mathbb{N}}$ in V a ***null sequence*** if it converges to $0 \in \mathsf{V}$. Since the image of a convergent sequence under a continuous function is convergent (see Exercise 7.1.5), it follows that there is a null sequence $\{v_j\}_{j\in\mathbb{N}}$ for which the sequence $\{L(v_j)\}_{j\in\mathbb{N}}$ does not converge. There are then two possibilities:

1. the image of any null sequence under L either converges to 0 or does not converge;
2. there exists a null sequence whose image under L converges to $u_0 \in \mathsf{V} \setminus \{0\}$.

We wish to disallow the second of these possibilities. Indeed, if a null sequence $\{v_j\}_{j\in\mathbb{N}}$ has the property that $\{L(v_j)\}_{j\in\mathbb{N}}$ converges to u_0, then the null sequence $\{av_j\}_{j\in\mathbb{N}}$ has the property that $\{L(av_j)\}_{j\in\mathbb{N}}$ converges to au_0. Thus sequences converging to the same point in V, when mapped under L may converge to different points in V. This is an odd circumstance, and thankfully applications do not normally involve linear maps of this sort. Let us then formally classify the sort of discontinuous linear map that is of interest to us.

7.1.8 Definition A linear operator $(L, \mathrm{dom}(L))$ is ***closable*** if for every null sequence $\{v_j\}_{j\in\mathbb{N}}$ in $\mathrm{dom}(L)$ the sequence $\{L(v_j)\}_{j\in\mathbb{N}}$ satisfies either of the following two criterion:

(i) it converges to 0 or

(ii) it does not converge. •

The very term "closable" implies that there should be a way to go from a closable linear operator to one that is closed. This is indeed the case, and let us describe how this works by means of the following theorem.

7.1.9 Theorem *Let* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *be a closable linear operator on a Banach space* V. *There then exists a closed linear operator* $(\overline{\mathrm{L}}, \mathrm{dom}(\overline{\mathrm{L}}))$ *which is an extension of* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$.

Proof We first define $\mathrm{dom}(\overline{L})$. We take $C(L)$ to be the collection of Cauchy sequences $\{v_j\}_{j\in\mathbb{N}}$ in $\mathrm{dom}(L)$ for which $\{L(v_j)\}_{j\in\mathbb{N}}$ converges. Let us say that two elements $\{u_j\}_{j\in\mathbb{N}}$ and $\{v_j\}_{j\in\mathbb{N}}$ of $C(L)$ are ***equivalent*** if $\{u_j - v_j\}_{j\in\mathbb{N}}$ is a null sequence. We take $\mathrm{dom}(\overline{L})$ to be the set of equivalence classes under this equivalence relation. We must show first that $\mathrm{dom}(\overline{L})$ is a subspace of V. To see this, first note that if $\{v_j\}_{j\in\mathbb{N}}$ converges to $v_0 \in \mathsf{V}$ and $\{L(v_j)\}_{j\in\mathbb{N}}$ converges to $u_0 \in \mathsf{V}$, then $\{av_j\}_{j\in\mathbb{N}}$ converges to $av_0 \in \mathsf{U}$ and $\{L(av_j)\}_{j\in\mathbb{N}}$ converges to au_0 for $a \in \mathbb{R}$. Thus the equivalence class containing $\{av_j\}_{j\in\mathbb{N}}$ is also in $\mathrm{dom}(\overline{L})$. In like manner, one shows that if the equivalence classes containing $\{u_j\}_{j\in\mathbb{N}}$ and $\{v_j\}_{j\in\mathbb{N}}$ are in $\mathrm{dom}(\overline{L})$, then so too is the equivalence class containing $\{u_j + v_j\}_{j\in\mathbb{N}}$. Thus $\mathrm{dom}(\overline{L})$ is indeed a vector space.

Next we need to define $\overline{L}$. We take $\overline{L}(v) = L(v)$ for $v \in \mathrm{dom}(L)$. For $\overline{v} \in \mathrm{dom}(\overline{L})$ we define $\overline{L}(\overline{v}) = \lim_{j\to\infty} L(v_j)$ where $\lim_{j\to\infty} v_j = \overline{v}$. The only possible problem with this definition is that it may depend on the choice of the sequence $\{v_j\}_{j\in\mathbb{N}}$ that approaches $\overline{u}$. However, using the fact that $(L, \mathrm{dom}(L))$ is closable, one can show that this is not the case. ■

7.1.10 Remarks

1. The subspace $\mathrm{dom}(\overline{L})$ need not be closed. Indeed, if $\mathrm{dom}(\overline{L})$ is closed and $(\overline{L}, \mathrm{dom}(\overline{L}))$ is closed then one can show that it must be the case that $\overline{L}$ is in fact continuous. Our interest is decidedly in discontinuous linear operators, so we should not expect that $\mathrm{dom}(\overline{L})$ be closed.

2. Note that the preceding result does not indicate whether $(\overline{L}, \mathrm{dom}(\overline{L}))$ is unique. Indeed, there will generally be more than one closed linear operator satisfying the hypotheses of the theorem. However, if we restrict to the closed linear operator with the smallest domain, then this fixes the linear operator to an linear operator we call the ***closure*** of $(L, \mathrm{dom}(L))$. It is this linear operator that is explicitly constructed in the proof of Theorem 7.1.9. •

There is a delicacy to the above considerations that is a bit bewildering to a newcomer to analysis in infinite-dimensions. It is therefore useful to ground these considerations with a simple example.

7.1.11 Example Let us take $\mathsf{V} = \mathrm{L}_2([0,1];\mathbb{F})$. We wish to consider differentiation as a linear operator. Clearly this cannot be defined on all of V since $\mathrm{L}_2([0,1];\mathbb{F})$ contains functions that are certainly not differentiable. We therefore take $\mathrm{dom}(L)$ to be the collection of functions $f\colon [0,1] \to \mathbb{R}$ of the form

$$f(x) = \int_0^x f'(\xi)\,\mathrm{d}\xi, \tag{7.2}$$

for some function $f' \in \mathrm{L}_2([0,1];\mathbb{F})$. The "prime" on f' should be interpreted carefully; it does not necessarily mean the derivative in the usual sense, but defines what is known as the **L_2-*derivative*** of f.[2] In any event, we define a map $L\colon \mathrm{dom}(L) \to \mathrm{L}_2([0,1];\mathbb{F})$ by assigning to f the function f' given by (7.2).

We claim that L is not continuous but is closed. To see that L is not continuous consider the sequence of functions $\{f_j\}_{j\in\mathbb{N}}$ defined by $f_j(x) = (1 + j2)x^j$. We compute $\|f_j\|_2 = 1$, so that $\{f_j\}_{j\in\mathbb{N}}$ is a sequence of bounded functions. Were L to be continuous, the image of this sequence under L should also be bounded. However, we compute $\|L(f_j)\|_2 = (1+j2)^{1/2}$. Therefore we have $\lim_{j\to\infty}\|L(f_j)\|_2 = \infty$, so L cannot be bounded, and so cannot be continuous. In Figure 7.1 we show the situation, illustrating the

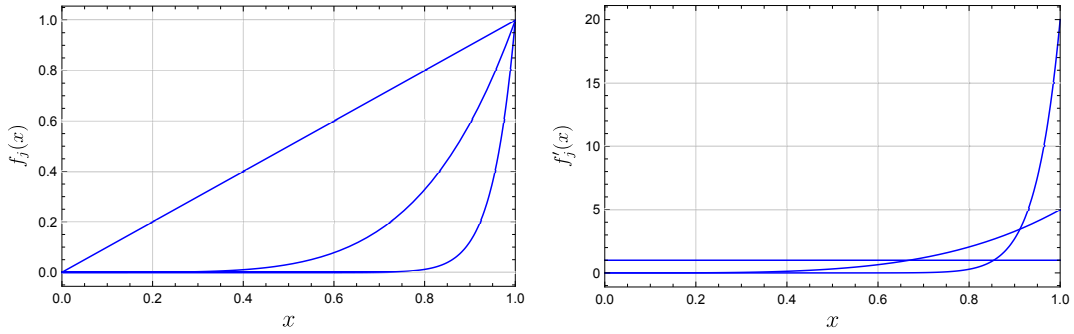


Figure 7.1 A bounded sequence (left) whose image under L (right) is not bounded

cases when $j \in \{1, 5, 20\}$.

To show that L is closed, let $\{f_j\}_{j\in\mathbb{N}}$ be a Cauchy sequence of functions in $\mathrm{dom}(L)$ having the property that $\{L(f_j)\}_{j\in\mathbb{N}}$ converges. We must show that $\lim_{j\to\infty} f_j \in \mathrm{dom}(L)$ and that $L(\lim_{j\to\infty} f_j) = \lim_{j\to\infty} L(f_j) = L(f_0)$. Since $\mathrm{L}_2([0,1];\mathbb{F})$ is complete we know that $g_0 = \lim_{j\to\infty} L(f_j) \in \mathrm{L}_2([0,1];\mathbb{F})$. Define

$$f_0(x) = \int_0^x g_0(\xi)\,\mathrm{d}\xi \in \mathrm{dom}(L).$$

[2] More generally, one may define the L_p-derivative in an analogous manner. With this notion of differentiability, a function may be L_p-differentiable for some p but not L_q-differentiable for $q > p$.

We claim that $\lim_{j\to\infty} f_j = f_0$. Since $\mathrm{L}_2([0,1];\mathbb{F})$ is complete we know that $\tilde{f}_0 = \lim_{j\to\infty} f_j \in \mathrm{L}_2([0,1];\mathbb{F})$. Thus we have

$$f_0(x) = \int_0^x \lim_{j\to\infty} f_j'(\xi)\,\mathrm{d}\xi$$

and

$$\tilde{f}_0(x) = \lim_{j\to\infty} \int_0^x f_j'(\xi)\,\mathrm{d}\xi,$$

where $f_j' = L(f_j)$. That L is closed now follows immediately from the Dominated Convergence Theorem. •

The linear operator in the preceding example was given to us as being closed. It is possible, however, to define the same operator as being the closure of a closable operator. Let us indicate how this might arise.

7.1.12 Example We define a linear operator $(\tilde{L}, \mathrm{dom}(\tilde{L}))$ as follows. We let $\mathrm{dom}(\tilde{L})$ be the collection of differentiable functions on $[0,1]$ whose derivative lies in $\mathrm{L}_2([0,1];\mathbb{F})$. If $(L, \mathrm{dom}(L))$ denotes the linear operator of the preceding example, then $\mathrm{dom}(\tilde{L}) \subset \mathrm{dom}(L)$. To see that the inclusion is strict, note that the function

$$f(x) = \begin{cases} x, & x \in [0, \frac{1}{2}] \\ 1 - x, & x \in (\frac{1}{2}, 1] \end{cases} \tag{7.3}$$

is in $\mathrm{dom}(L)$ but is not in $\mathrm{dom}(\tilde{L})$ since f is not differentiable at $x = \frac{1}{2}$. We then define $\tilde{L}\colon \mathrm{dom}(\tilde{L}) \to \mathrm{L}_2([0,1];\mathbb{F})$ by $L(f) = f'$, where now f' *really* means the derivative in the usual sense.

Let us show that $(\tilde{L}, \mathrm{dom}(\tilde{L}))$ is closable. We let $\{f_j\}_{j\in\mathbb{N}}$ be a null sequence in $\mathrm{dom}(\tilde{L})$ for which $\{\tilde{L}(f_j)\}_{j\in\mathbb{N}}$ is convergent and converges to $g \in \mathrm{L}_2([0,1];\mathbb{F})$. For $h \in \mathrm{L}_2([0,1];\mathbb{F})$ we have

$$\lim_{j\to\infty} \langle h, f_j' \rangle = \langle h, g \rangle$$

since the inner product is continuous (Exercise 7.1.2). Now further suppose that h is continuously differentiable and that $h(0) = h(1) = 0$. The collection of all such functions is dense in $\mathrm{L}_2([0,1];\mathbb{F})$. To see this, note that any such function possesses a uniformly convergent Fourier series, and since the Fourier basis functions are dense, so too must be the collection of such functions. In any event, with h so restricted we have

$$\langle h, f_j' \rangle = \int_0^1 h(x)\overline{f_j'(x)}\,\mathrm{d}x = -\int_0^1 \overline{f_j(x)}h'(x)\,\mathrm{d}x = -\langle h', f_j \rangle,$$

by an integration by parts. Again using continuity of the inner product we infer that $\lim_{j\to\infty}\langle h, f_j' \rangle = -\lim_{j\to\infty}\langle h', f_j \rangle = 0$. Thus g is orthogonal to a dense subset of $\mathrm{L}_2([0,1];\mathbb{F})$, and so must be zero. This shows that $(\tilde{L}, \mathrm{dom}(\tilde{L}))$ is closable.

It is not true, however, that $(\tilde{L}, \mathrm{dom}(\tilde{L}))$ is closed. To see this, recall that the function f defined in (7.3) has a uniformly convergent Fourier series. Thus for this function there exists a sequence $\{f_j\}_{j\in\mathbb{N}}$ in $\mathrm{dom}(\tilde{L})$ so that $\lim_{j\to\infty} f_j = f$ in $\mathrm{L}_2([0,1];\mathbb{F})$. It is also the case that the sequence $\{L(f_j)\}_{j\in\mathbb{N}}$ converges in this case. However, $f \notin \mathrm{dom}(\tilde{L})$, so the linear operator is not closed. This shows that in practice the difference between a linear operator that is merely closable and one that is closed is often a minor discrepancy that can be redressed by adding to the domain of the closable operator a suitable collection of limit functions. •

7.1.1.3 Invertibility of linear operators The notion of invertibility of a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^n$ is well understood, and is equivalent to the condition that if we think of L as an $n \times n$ matrix then $\det(L) \neq 0$. As expected, for linear operators defined on infinite-dimensional normed vector spaces, the issues are more complicated. Indeed, as we shall see, there are various ways in which a linear operator can be singular, and only some of the possibilities will be of interest to us.

Let us first consider injective linear operators. In the following discussion we let $(L, \mathrm{dom}(L))$ be a linear operator on V. The following result has likely been encountered in a basic linear algebra course.

7.1.13 Lemma *A linear operator* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *on* V *is injective if and only if* $\ker(\mathrm{L}) = \{0\}$.

Proof First suppose that L is injective and that $L(v) = 0$. Since $L(0) = 0$ this implies that $v = 0$. Next suppose that $\ker(L) = \{0\}$ and that $L(v_1) = L(v_2)$. Then $L(v_1 - v_2) = 0$ by linearity, implying that $v_1 = v_2$. ■

If $(L, \mathrm{dom}(L))$ is injective then $L\colon \mathrm{dom}(L) \to \mathrm{image}(L)$ is necessarily an isomorphism. In this case we define a linear operator $(L^{-1}, \mathrm{image}(L))$ where $L\colon \mathrm{image}(L) \subseteq \mathsf{V} \to \mathrm{dom}(L) \subseteq \mathsf{U}$. Note that L^{-1} defined in this manner is *not* defined on all of V, only on $\mathrm{image}(L)$. We shall say that $(L^{-1}, \mathrm{image}(L))$ is the ***inverse*** of L, and so say that $(L, \mathrm{dom}(L))$ is ***invertible***.

7.1.14 Definition Let $(\mathsf{V}, \|\cdot\|)$ be a normed vector space and let $(L, \mathrm{dom}(L))$ be a linear operator on V.

(i) $(L, \mathrm{dom}(L))$ is ***essentially regular*** if
  (a) L is injective,
  (b) $(L^{-1}, \mathrm{image}(L))$ is continuous, and
  (c) $\mathrm{cl}(\mathrm{image}(L)) = \mathsf{V}$.

(ii) $(L, \mathrm{dom}(L))$ is ***regular*** if it is essentially regular and if $\mathrm{image}(L) = \mathsf{V}$.

(iii) $(L, \mathrm{dom}(L))$ is ***singular*** if it is neither regular nor essentially regular. •

In a manner resembling closed and closable linear operators, one can go from an essentially regular linear operator to a regular linear operator in a natural way.

7.1.15 Proposition *Let* $(\mathsf{V}, \|\cdot\|)$ *be a Banach space and let* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *be a linear operator on* V. *If* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is essentially regular then there exists a regular linear operator* $(\overline{\mathrm{L}}, \mathrm{dom}(\overline{\mathrm{L}}))$ *on* V *which is an extension of* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$. $(\overline{\mathrm{L}}, \mathrm{dom}(\overline{\mathrm{L}}))$ *is called the* ***regularisation*** *of* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$.

Proof We proceed by defining $\overline{L}^{-1}$. For any $v_0 \in \mathsf{V}$ there exists a Cauchy sequence $\{v_j\}_{j\in\mathbb{N}}$ in image(L) and converging to v_0. Since L^{-1} is continuous the sequence $\{L^{-1}(v_j)\}_{j\in\mathbb{N}}$ converges to $u_0 \in \mathsf{V}$. We define $\overline{L}^{-1}(v_0) = u_0$. One shows that the collection of $u \in \mathsf{V}$ that are images under L^{-1} of Cauchy sequences in image(L) form a subspace of V, and we denote this subspace by $\mathrm{dom}(\overline{L})$. One then defines $\overline{L} = (\overline{L}^{-1})^{-1}$. ■

In Sections 7.2 and 7.3 we shall be interested in solutions of equations of the form $L(v) = u$. To ensure the existence to such an equation, one wants $u \in$ image(L); to guarantee uniqueness of the solution, one wants L to be injective; and to ensure that the solutions of the equation do not vary wildly as one varies u, one wants the inverse to be continuous. This motivates our interest in regular linear operators. However, we shall also be very interested in singular linear operators for reasons that will not be clear at this time (although we shall see some reason for this in Section 7.1.3). Nevertheless, let us say a few words about singular linear operators. Our interest is in closed operators, and the following result gives some important features of closed operators as concerns their invertibility.

7.1.16 Proposition *Let* $(\mathsf{V}, \|\cdot\|)$ *be a normed vector space and suppose that* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is a closed linear operator on* V. *The following statements hold:*

(i) if $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is invertible then its inverse is closed;*

(ii) if $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is invertible and if its inverse is continuous then* image(L) *is closed in* V.

Proof (i) Suppose that the sequence $\{u_j\}_{j\in\mathbb{N}}$ converges to u_0 and that $\{L^{-1}(u_j)\}_{j\in\mathbb{N}}$ converges to v_0. Note that $\{L \circ L^{-1}(u_j)\}_{j\in\mathbb{N}}$ converges to u_0. Therefore, since $(L, \mathrm{dom}(L))$ is closed it follows that $v_0 \in \mathrm{dom}(L)$ and that $L(v_0) = u_0$. From this we see that $v_0 = L^{-1}(u_0)$ as desired.

(ii) Let $\{u_j\}_{j\in\mathbb{N}}$ be a sequence in image(L) converging to $u_0 \in \mathsf{V}$. There then exists a sequence $\{v_j\}_{j\in\mathbb{N}}$ in $\mathrm{dom}(L)$ for which $L(v_j) = u_j$, $j \in \mathbb{N}$. Since the sequence $\{u_j\}_{j\in\mathbb{N}}$ is Cauchy and since L^{-1} is continuous the sequence $\{v_j\}_{j\in\mathbb{N}}$ must be Cauchy. Therefore, the sequences $\{v_j\}_{j\in\mathbb{N}}$ and $\{L(v_j) = u_j\}_{j\in\mathbb{N}}$ converge. Since L is closed it follows that $v_0 = \lim_{j\to\infty} v_j \in \mathrm{dom}(L)$ and $L(v_0) = \lim_{j\to\infty} u_j$. Thus $u_0 \in$ image(L) as desired. ■

The preceding result allows the following classification of closed operators.

7.1.17 Theorem *Let* $(\mathsf{V}, \|\cdot\|)$ *be a normed vector space and suppose that* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is a closed linear operator on* V. *Then* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *falls into one of the following mutually exclusive classes:*

(i) $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is regular;*

(ii) $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is not invertible;*

(iii) $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is invertible,* $(\mathrm{L}^{-1}, \mathrm{dom}(\mathrm{L}))$ *is unbounded,* $\mathrm{image}(\mathrm{L}) \subset \mathsf{V}$, *and* $\mathrm{cl}(\mathrm{image}(\mathrm{L})) = \mathsf{V}$;

(iv) $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is invertible and* $\mathrm{cl}(\mathrm{image}(\mathrm{L})) \subset \mathsf{V}$.

Proof Clearly $(L, \mathrm{dom}(L))$ must be either regular, essentially regular, or singular. If $(L, \mathrm{dom}(L))$ is singular then it falls into exactly one of the last three classes. Thus to prove the theorem, we need only assert that $(L, \mathrm{dom}(L))$ cannot be essentially regular if it is closed. This, however, follows from part (ii) of Proposition 7.1.16. ■

Let us exhibit simple examples that fall into the classes enumerated in Theorem 7.1.17.

7.1.18 Examples

1. On any normed vector space $(\mathsf{V}, \|\cdot\|)$ the linear operator $(\mathrm{id}_{\mathsf{V}}, \mathsf{V})$ is continuous and invertible. Furthermore, $\mathrm{image}(\mathrm{id}_{\mathsf{V}}) = \mathsf{V}$, so this linear operator is regular.
2. On any normed vector space $(\mathsf{V}, \|\cdot\|)$ the linear operator (L, V) defined by $L(v) = 0$ is continuous. It is certainly not invertible, however, so it represents an example of case (ii) of the theorem.
3. On $\mathsf{V} = (\mathrm{L}_2([0,1]; \mathbb{F})$ consider the linear operator $(L, \mathrm{dom}(L))$ given by $\mathrm{dom}(L) = \mathsf{V}$ and $L(f)(x) = xf(x)$. It is clear that L is continuous since we have

$$\|L(f)\|_2^2 = \int_0^1 |xf(x)|^2 \, \mathrm{d}x \leq \int_0^1 |f(x)|^2 \, \mathrm{d}x = \|f\|_2^2.$$

   It is also evident that L is invertible since $L(f) = 0$ obviously implies that $f = 0$ a.e. We note that $\mathrm{image}(L) \subset \mathsf{V}$ since the function $f(x) = 1$ is not in the image of L. Indeed, if this function *were* in $\mathrm{image}(L)$ then there would be a function $f \in \mathrm{L}_2([0,1]; \mathbb{F})$ so that $xf(x) = 1$. Thus $f(x) = \frac{1}{x}$, but this function is not in $\mathrm{L}_2([0,1]; \mathbb{F})$. However, if we define

$$S = \{f \in \mathrm{L}_2([0,1]; \mathbb{F}) \mid \text{ there exists a neighbourhood of 0 on which } f \text{ vanishes}\},$$

   then clearly $S \subseteq \mathrm{image}(L)$. Furthermore, one easily sees that S is dense in V, showing that $\mathrm{image}(L)$ is dense in V. This shows that $(L, \mathrm{dom}(L))$ belongs to the functions of case (iii) of the theorem.
4. Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an inner product space with $\{e_j\}_{j\in\mathbb{N}}$ a complete orthonormal family in V. For $v \in \mathsf{V}$ let $a_j(v)$ be the components of v in the complete orthonormal family so that

$$v = \sum_{j=1}^{\infty} a_j(v) e_j.$$

We define the ***shift operator*** on V to be the linear operator $(L, \operatorname{dom}(L) = \mathsf{V})$ defined by

$$L(v) = \sum_{j=1}^{\infty} a_j(v) e_{j+1}.$$

By Parseval's inequality we have $\|L(v)\| \leq \|v\|$, thus L is continuous.
It is also evident that $(L, \operatorname{dom}(L))$ is invertible. Indeed, suppose that $L(v) = 0$. Then

$$0 = \sum_{j=1}^{\infty} a_j(v) e_{j+1} = 0e_1 + \sum_{j=1}^{\infty} a_j(v) e_{j+1}.$$

By ***missing stuff*** it follows that $a_j(v) = 0$, $j \in \mathbb{N}$, or that $v = 0$.
We also claim that $\operatorname{image}(L)$ is not dense in V. Indeed, it is clear that the function e_1 is orthogonal to $\operatorname{image}(L)$, which prohibits $\operatorname{image}(L)$ from being dense.
All this shows that L belongs to the class of linear operator described by case (iv) of the theorem. •

7.1.1.4 Linear functions For a normed vector space $(\mathsf{V}, \|\cdot\|)$ over $\mathbb{F}$, the ***dual*** of V is the collection of continuous $\mathbb{F}$-valued linear maps in V. Thus the dual of V is $\mathrm{L}(\mathsf{V}; \mathbb{F})$, which we abbreviate as V^*. Of some interest to us will be the dual space of a Hilbert space. The following result characterises such duals.

7.1.19 Theorem (Riesz representation theorem) *Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a Hilbert space. For each $\alpha \in \mathsf{V}^*$ there exists a unique $\mathrm{v}_\alpha \in \mathsf{V}$ so that $\alpha(\mathrm{u}) = \langle \mathrm{u}, \mathrm{v}_\alpha \rangle$ for each $\mathrm{u} \in \mathsf{V}$.*

Proof If $\alpha = 0$ then we can take $v_\alpha = 0$. So let $\alpha \in \mathsf{V}^* \setminus \{0\}$. We claim that $\ker(\alpha)$ is a closed subspace of V. It is certainly a subspace. To show that it is closed, let $\{v_j\}_{j\in\mathbb{N}}$ be a Cauchy sequence in $\ker(\alpha)$. Then the sequence $\{\alpha(v_j)\}_{j\in\mathbb{N}}$ is certainly convergent. Since α is continuous and V is complete, it follows that $\{v_j\}_{j\in\mathbb{N}}$ is convergent. Since $\alpha \neq 0$, $\ker(\alpha) \neq \mathsf{V}$. Therefore, since $\ker(\alpha)$ is closed, we can choose a nonzero vector $v_0 \in \ker(\alpha)^\perp$, supposing this vector to further have length 1. Define $\bar{\alpha} \in \mathsf{V}^*$ by $\bar{\alpha}(v) = \overline{\alpha(v)}$. We claim that we can take $v_\alpha = \bar{\alpha}(v_0)v_0$. Indeed note that for $u \in \mathsf{V}$ the vector $\alpha(u)v_0 - \alpha(v_0)u$ is in $\ker(\alpha)$. Therefore

$$0 = \langle \alpha(u)v_0 - \alpha(v_0)u, v_0 \rangle = \alpha(u) - \alpha(v_0)\langle u, v_0 \rangle.$$

Thus $\alpha(u) = \langle u, \bar{\alpha}(v_0)v_0 \rangle$. Thus v_α as defined meets the desired criterion. Let us show that this is the only vector satisfying the conditions of the theorem. Suppose that $v_1, v_2 \in \mathsf{V}$ have the property that $\alpha(u) = \langle u, v_1 \rangle = \langle u, v_2 \rangle$ for all $u \in \mathsf{V}$. Then $\langle u, v_1 - v_2 \rangle = 0$ for all $u \in \mathsf{V}$. In particular, taking $u = v_1 - v_2$ we have $\|v_1 - v_2\|^2 = 0$, giving $v_1 = v_2$. ■

Said in more sophisticated language, the Riesz representation theorem says that V^* is isomorphic to V, the isomorphism being given as $\alpha \mapsto v_\alpha$.

7.1.2 Linear maps on inner product spaces

The structure of certain linear maps on inner product spaces will be of great interest to us. In the case of continuous linear maps, the Riesz representation theorem makes this discussion quite simple. For discontinuous linear maps—our concern will be with linear operators—one must be careful as the types of possible behaviour are various, differing in sometimes subtle ways.

7.1.2.1 The adjoint of a continuous linear map Consider two $\mathbb{F}$-inner product spaces $(\mathsf{U}, \langle\cdot,\cdot\rangle_{\mathsf{U}})$ and $(\mathsf{V}, \langle\cdot,\cdot\rangle_{\mathsf{V}})$. For fixed $v \in \mathsf{V}$ consider the map from U to $\mathbb{F}$ given by $u \mapsto \langle L(u), v\rangle$. Since the inner product is continuous (Exercise 7.1.2) this map is an element of U^*. In this way we assign to each $v \in \mathsf{V}$ an element $\alpha_v \in \mathsf{U}^*$. By the Riesz representation theorem this therefore defines an element $u_v \in \mathsf{U}$. In other words, we have defined a map $L^* \colon \mathsf{V} \to \mathsf{U}$. It is a straightforward exercise, given as Exercise 7.1.8, to show that L^* is linear. We call L^* the ***adjoint*** of L in this case.

Let us consider an example of an adjoint defined on an infinite-dimensional vector space.

7.1.20 Example On $\mathrm{L}_2([0,1];\mathbb{F})$ we consider the linear transformation defined by $L(f)(x) = xf(x)$, as in Example 7.1.18–3. We showed in that preceding example that L is continuous, so it certainly possesses an adjoint as we describe here. Let $f, g \in \mathsf{V}$ and compute

$$\langle L(f), g\rangle = \int_0^1 xf(x)\overline{g(x)}\,\mathrm{d}x = \int_0^1 f(x)\overline{xg(x)}\,\mathrm{d}x = \langle f, L^*(g)\rangle$$

where $L^*(g)(x) = xg(x)$. Thus we see in this case that $L^* = L$. •

The following results might help in understanding the adjoint, telling us what it looks like in $\mathbb{F}^n$ with the inner product being the dot product.

7.1.21 Proposition *Consider the inner product on $\mathbb{F}^{\mathrm{n}}$ given by the dot product:*

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\bar{\mathbf{y}} = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{x}_{\mathrm{i}}\bar{\mathrm{y}}_{\mathrm{i}}.$$

If $\mathbf{L} \in \mathrm{L}(\mathbb{F}^{\mathrm{n}};\mathbb{F}^{\mathrm{m}})$ is a linear map (i.e., an $\mathrm{m}\times\mathrm{n}$ matrix with entries in $\mathbb{F}$), then $\mathbf{L}^* = \bar{\mathbf{L}}^{\mathrm{T}}$. That is, the matrix corresponding to the linear map $\mathbf{L}^*$ is obtained by taking the conjugate of all entries in the transpose $\mathbf{L}^{\mathrm{T}}$.

Proof One may write the dot product in terms of matrix multiplication like this:

$$\boldsymbol{x}\cdot\bar{\boldsymbol{y}} = \boldsymbol{x}^T\bar{\boldsymbol{y}}.$$

The definition of adjoint is then as follows. For $\boldsymbol{x} \in \mathbb{F}^m$, $\boldsymbol{L}^*\boldsymbol{x} \in \mathbb{F}^n$ satisfies

$$(\boldsymbol{L}^*\boldsymbol{x})\cdot\bar{\boldsymbol{y}} = \boldsymbol{x}\cdot\left(\overline{\boldsymbol{L}\boldsymbol{y}}\right)$$

for every $\boldsymbol{y} \in \mathbb{F}^n$. Using the matrix multiplication characterisation of the dot product, this gives, for every $\boldsymbol{y} \in \mathbb{F}^n$,

$$\begin{aligned} & (\boldsymbol{L}^*\boldsymbol{x})^T\bar{\boldsymbol{y}} = \boldsymbol{x}^T\big(\overline{\boldsymbol{L}\boldsymbol{y}}\big) \\ \implies \quad & \big(\boldsymbol{x}^T(\boldsymbol{L}^*)^T\big)\bar{\boldsymbol{y}} = \boldsymbol{x}^T\big(\bar{\boldsymbol{L}}\bar{\boldsymbol{y}}\big) \\ \implies \quad & \big(\boldsymbol{x}^T(\boldsymbol{L}^*)^T\big)\bar{\boldsymbol{y}} = \big(\boldsymbol{x}^T\bar{\boldsymbol{L}}\big)\bar{\boldsymbol{y}}. \end{aligned}$$

Since this must be true for every $\boldsymbol{y} \in \mathbb{F}^n$ we can assert that

$$\begin{aligned} & \boldsymbol{x}^T(\boldsymbol{L}^*)^T = \boldsymbol{x}^T\bar{\boldsymbol{L}} \\ \implies \quad & \boldsymbol{L}^*\boldsymbol{x} = \bar{\boldsymbol{L}}^T\boldsymbol{x}. \end{aligned}$$

Thus we have shown that $\boldsymbol{L}^* = \bar{\boldsymbol{L}}^T$, as desired. ■

Thus, if $\mathbb{F} = \mathbb{R}$, a self-adjoint linear on $\mathbb{R}^n$ is simply a symmetric matrix. However, our principal interest is in understanding self-adjoint maps in the case when V is infinite-dimensional.

7.1.2.2 The adjoint of a linear operator The discussion in the preceding section was facilitated by the ability to use the Riesz representation theorem for continuous $\mathbb{F}$-valued linear maps. However, since much of our attention will be focused on the investigation of discontinuous linear maps, we will also benefit from trying to extend the notion of adjoint to this case. Thus in this section we let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space and we consider a linear operator $(L, \mathrm{dom}(L))$ defined on V. The problem arises that given $u \in \mathsf{V}$, the map $v \mapsto \langle L(v), u\rangle$ may not be continuous. We do, however, have the following result.

7.1.22 Lemma *Let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be a Hilbert space and let* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *be a linear operator on* V. *The following statements hold:*

(i) the set

$$\mathrm{C_L} = \{\mathrm{u} \in \mathsf{V} \mid \exists\, \mathrm{w_u} \in \mathsf{V} \text{ so that } \langle \mathrm{L(v)}, \mathrm{u}\rangle = \langle \mathrm{v}, \mathrm{w_u}\rangle \ \forall\ \mathrm{v} \in \mathrm{dom(L)}\}$$

is a nonempty subspace of V*;*

(ii) if $\mathrm{dom(L)}$ *is dense in* V *and if* $\mathrm{u} \in \mathrm{C_L}$ *then the set*

$$\mathrm{C_{L,u}} = \{\mathrm{w_u} \in \mathsf{V} \mid \langle \mathrm{L(v)}, \mathrm{u}\rangle = \langle \mathrm{v}, \mathrm{w_u}\rangle,\ \mathrm{v} \in \mathrm{dom(L)}\}$$

consists of a single element;

(iii) if $\mathrm{dom(L)}$ *is dense in* V *then the map sending* $\mathrm{u} \in \mathrm{C_L}$ *to the unique element* $\mathrm{w_u} \in \mathrm{C_{L,u}}$ *is linear.*

Proof (i) That C_L is empty follows since 0 is obviously in C_L. Suppose that $u \in C_L$ and $a \in \mathbb{F}$. Then there exists $w_u \in \mathsf{V}$ so that $\langle L(v), u\rangle = \langle v, w_u\rangle$ for all $v \in \mathrm{dom}(L)$. Therefore $\langle L(v), au\rangle = \langle v, aw_u\rangle$ for all $v \in \mathrm{dom}(L)$, showing that $au \in C_L$. In like manner one shows that if $u_1, u_2 \in C_L$ then $u_1 + u_2 \in C_L$.

(ii) Let $w_1, w_2 \in C_{L,u}$ so that $\langle v, w_1\rangle = \langle v, w_2\rangle$ for all $v \in \mathrm{dom}(L)$. Thus $\langle v, w_1 - w_2\rangle = 0$ for all $v \in \mathrm{dom}(L)$, showing that $w_1 = w_2$ since $\mathrm{dom}(L)$ is dense in V.

(iii) We must show that if $u, u_1, u_2 \in C_L$ and $a \in \mathbb{R}$ then we have

$$C_{L,au} = \{aw_u\}, \quad C_{L,u_1+u_2} = \{w_{u_1} + w_{u_2}\}.$$

We have $\langle L(v), u\rangle = \langle v, w_u\rangle$ for all $v \in \mathrm{dom}(L)$ from which we assert that $\langle L(v), au\rangle = \langle v, aw_u\rangle$ for all $v \in \mathrm{dom}(L)$. This shows that $aw_u \in C_{L,au}$, showing that $C_{L,au} = \{aw_u\}$ by (ii). In like manner one shows that $C_{L,u_1+u_2} = \{w_{u_1} + w_{u_2}\}$. ■

The above lemma is, in actuality not difficult, but it does require a moments thought to understand the notation and the consequences. The important consequence is the following result.

7.1.23 Corollary *If* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is a Hilbert space with* $(\mathrm{L}, \mathrm{dom(L)})$ *be a linear operator on* V *with* $\mathrm{dom(L)}$ *dense in* V*, then there exists a linear operator* $(\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$ *on* V *satisfying* $\langle \mathrm{L(v)}, \mathrm{u}\rangle = \langle \mathrm{v}, \mathrm{L}^*(\mathrm{u})\rangle$ *for all* $\mathrm{v} \in \mathrm{dom(L)}$ *and* $\mathrm{u} \in \mathrm{dom}(\mathrm{L}^*)$*. The operator* $(\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$ *is the* ***adjoint*** *of* $(\mathrm{L}, \mathrm{dom(L)})$*.*

Proof The result follows by transcribing the notation of Lemma 7.1.22. Indeed, we take $\mathrm{dom}(L^*) = C_L$ and L^* to be the map that assigns to $u \in \mathrm{dom}(L^*)$ the unique element $w_u \in C_{L,u}$. ■

Let us enumerate some of the useful properties of the adjoint.

7.1.24 Proposition *Let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be a Hilbert space with* $(\mathrm{L}, \mathrm{dom(L)})$ *and* $(\mathrm{M}, \mathrm{dom(M)})$ *linear operators on* V *for which* $\mathrm{dom(L)}$ *and* $\mathrm{dom(M)}$ *are dense in* V*. If* $(\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$ *is the adjoint of* $(\mathrm{L}, \mathrm{dom(L)})$ *then the following statements hold:*

(i) $(\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$ *is closed;*

(ii) if $(\mathrm{L}, \mathrm{dom(L)}) \subseteq (\mathrm{M}, \mathrm{dom(M)})$ *then* $(\mathrm{M}^*, \mathrm{dom}(\mathrm{M}^*)) \subseteq (\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$*;*

(iii) if $\mathrm{dom(L)} = \mathsf{V}$ *and* L *is continuous then* $\mathrm{dom}(\mathrm{L}^*) = \mathsf{V}$ *and* L^* *is continuous;*

(iv) if $\mathrm{dom}(\mathrm{L}^*)$ *is dense in* V *then* $(\mathrm{L}, \mathrm{dom(L)}) \subseteq (\mathrm{L}^{**}, \mathrm{dom}(\mathrm{L}^{**}))$*;*

(v) if $(\mathrm{L}, \mathrm{dom(L)})$ *is closable with closure* $(\overline{\mathrm{L}}, \mathrm{dom}(\overline{\mathrm{L}}))$ *then* $(\overline{\mathrm{L}}^*, \mathrm{dom}(\overline{\mathrm{L}}^*)) = (\mathrm{L}^*, \mathrm{dom}(\mathrm{L}^*))$*;*

(vi) if $(\mathrm{L}, \mathrm{dom(L)})$ *is closed then* $\mathrm{dom}(\mathrm{L}^*)$ *is dense in* V *and* $(\mathrm{L}, \mathrm{dom(L)}) = (\mathrm{L}^{**}, \mathrm{dom}(\mathrm{L}^{**}))$*.*

Proof (i) Let $\{u_j\}_{j\in\mathbb{N}}$ be a sequence in $\mathrm{dom}(L^*)$ converging to $u_0 \in \mathsf{V}$ and suppose that $\{L(u_j)\}_{j\in\mathbb{N}}$ converges to $v_0 \in \mathsf{V}$. Since the inner product is continuous (Exercise 7.1.2) we have

$$\lim_{j\to\infty} \langle L(v), u_j\rangle = \langle L(v), u_0\rangle, \quad \lim_{j\to\infty} \langle v, L^*(u_j)\rangle = \langle v, v_0\rangle$$

for all $v \in \mathrm{dom}(L)$. Since $\langle L(v), u_j\rangle = \langle v, L^*(u_j)\rangle$, $j \in \mathbb{N}$ we ascertain that $\langle L(v), u_0\rangle = \langle v, v_0\rangle$ for all $v \in \mathrm{dom}(L)$. Therefore, by Lemma 7.1.22, $u_0 \in \mathrm{dom}(L^*)$ and $v_0 = L^*(u_0)$. This shows that $(L^*, \mathrm{dom}(L^*))$ is closed.

(ii) Let $u_0 \in \mathrm{dom}(M)$ and let $v_0 = M^*(u_0)$. Thus we have $\langle M(v), u_0\rangle = \langle v, v_0\rangle$ for all $v \in \mathrm{dom}(M)$. Since $(L, \mathrm{dom}(L)) \subseteq (M, \mathrm{dom}(M))$ this implies that $\langle L(v), u_0\rangle = \langle v, v_0\rangle$ for all $v \in \mathrm{dom}(L)$. This shows that $u_0 \in \mathrm{dom}(L^*)$ and that $L^*(u_0) = v_0$ by Lemma 7.1.22.

(iii) This follows directly from the construction of Section 7.1.2.2.

(iv) First note that if $\mathrm{dom}(L^*)$ is dense in V then the construction of Corollary 7.1.23 applies and we may actually define $(L^{**}, \mathrm{dom}(L^{**}))$. Let $v_0 \in \mathrm{dom}(L^{**})$. Thus we have $\langle L^*(u), v_0\rangle = \langle u, L^{**}(v_0)\rangle$ for all $u \in \mathrm{dom}(L^*)$. If $u \in \mathrm{dom}(L^*)$ then $\langle L(v), u\rangle = \langle v, L^*(u)\rangle$ for each $v \in \mathrm{dom}(L^*)$. ***missing stuff***

(v) From part (ii) we have $(\overline{L}^*, \mathrm{dom}(\overline{L}^*)) \subseteq (L^*, \mathrm{dom}(L^*))$, thus it is the opposite "inclusion" we must show. Thus let $u_0 \in \mathrm{dom}(L^*)$ and let $v_0 = L^*(u_0)$. Thus $\langle L(v), u_0\rangle = \langle v, v_0\rangle$ for all $v \in \mathrm{dom}(L)$. Let $\{v_j\}_{j\in\mathbb{N}}$ be a sequence in $\mathrm{dom}(L)$ converging to $\overline{v}$, and suppose that $\{L(v_j)\}_{j\in\mathbb{N}}$ converges to $\overline{u}$. Since $(L, \mathrm{dom}(L))$ is closable we have $\overline{v} \in \mathrm{dom}(\overline{L})$ and $\overline{L}(\overline{v}) = \overline{u}$. By continuity of the inner product we also have

$$\lim_{j\to\infty}\langle L(v_j), u_0\rangle = \langle \overline{u}, u_0\rangle, \quad \lim_{j\to\infty}\langle v_j, v_0\rangle = \langle \overline{v}, v_0\rangle,$$

giving $\langle \overline{L}(\overline{v}), u_0\rangle = \langle \overline{v}, v_0\rangle$. Since this can be done for each $\overline{v} \in \mathrm{dom}(\overline{L})$ this shows that $u_0 \in \mathrm{dom}(\overline{L}^*)$ and that $v_0 = \overline{L}^*(u_0)$, as desired.

(vi) ***missing stuff*** ■

Let us give some examples of adjoints so that we may appreciate that the subtleties in the definition do arise in simple situations.

7.1.25 Examples In each of the next three examples we consider the Hilbert space $(\mathrm{L}_2([0,1];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$. The subspace of this Hilbert space that will be basis for the domain of all linear operators we consider is the subset S of functions $f \in \mathrm{L}_2([0,1];\mathbb{F})$ for which there exists a function $f' \in \mathrm{L}_2([0,1];\mathbb{F})$ so that

$$f(x) = \int_0^x f'(\xi)\,\mathrm{d}\xi.$$

The examples we consider differ by their imposing on functions in S various boundary conditions. In all cases the operator is the L_2-differentiation operator which assigns to $f \in S$ the function $f' \in \mathrm{L}_2([0,1];\mathbb{F})$.

1. Here we take $\mathrm{dom}(L_1) = S$. As we saw in Example 7.1.11, the linear operator assigning to $f \in S$ the L_2-derivative f' is a closed linear operator. Since, the differentiable functions on $[0,1]$ are dense in $\mathrm{L}_2([0,1];\mathbb{F})$ (by part ***missing stuff*** of ***missing stuff***), it follows that $S = \mathrm{dom}(L_1)$ is dense in $\mathrm{L}_2([0,1];\mathbb{F})$. Let us determine the adjoint of the linear operator $(L_1, \mathrm{dom}(L_1))$.

Lemma *If* $(\mathrm{L}_1^*, \mathrm{dom}(\mathrm{L}_1^*))$ *is the adjoint of* $(\mathrm{L}_1, \mathrm{dom}(\mathrm{L}_1))$ *then*

$$\mathrm{dom}(\mathrm{L}_1^*) = \{\mathrm{g} \in \mathrm{S} \mid \ \mathrm{g}(0) = \mathrm{g}(1) = 0\}$$

and $\mathrm{L}_1^*(\mathrm{g}) = -\mathrm{g}'$.

Proof Let us denote by $(L, \mathrm{dom}(L))$ the linear operator defined by

$$\mathrm{dom}(L) = \{g \in S \mid \ g(0) = g(1) = 0\}, \quad L(f) = -f'.$$

Now let $f \in \mathrm{dom}(L_1)$

$$\begin{aligned}\langle L_1(f), g\rangle_2 &= \int_0^1 f'(x)\overline{g(x)}\,\mathrm{d}x \\ &= f(x)\overline{g(x)}\big|_0^1 - \int_0^1 f(x)\overline{g'(x)}\,\mathrm{d}x \\ &= \langle f, L(g)\rangle_2,\end{aligned}$$

showing that $(L, \mathrm{dom}(L)) \subseteq (L_1^*, \mathrm{dom}(L_1^*))$.
To show the converse "inclusion" we proceed as follows. Let $g, h \in \mathrm{L}_2([0,1]; \mathbb{F})$ satisfy $\langle L(f), g\rangle_2 = \langle f, h\rangle_2$ for each $f \in \mathrm{dom}(L_1)$. We compute

$$\begin{aligned}\langle f, h\rangle_2 &= \int_0^1 f(x)\overline{h(x)}\,\mathrm{d}x \\ &= f(x)\Big(\int_0^x \overline{h(\xi)}\,\mathrm{d}\xi\Big)\Big|_0^1 - \int_0^1 f'(x)\Big(\int_0^x \overline{h(\xi)}\,\mathrm{d}\xi\Big)\,\mathrm{d}x. \qquad (7.4)\end{aligned}$$

Define $f_1(x) = 1$ so that $f_1 \in \mathrm{dom}(L_1)$ and $L_1(f_1) = 0$. Then we have

$$0 = \langle L_1(f_1), g\rangle_2 = \langle f_1, h\rangle_2 = \int_0^1 \overline{h(x)}\,\mathrm{d}x. \qquad (7.5)$$

Thus the first term in (7.4) vanishes and we have

$$\langle f, h\rangle_2 = -\int_0^1 f'(x)\Big(\int_0^x \overline{h(\xi)}\,\mathrm{d}\xi\Big)\,\mathrm{d}x,$$

this holding for all $f \in \mathrm{dom}(L_1)$. Since $\langle L(f), g\rangle = \langle f, h\rangle$ for all $f \in \mathrm{dom}(L_1)$ we see that

$$\int_0^1 f'(x)\Big(\overline{g(x)} + \int_0^x \overline{h(\xi)}\,\mathrm{d}\xi\Big)\,\mathrm{d}x = 0$$

for all $f \in \mathrm{dom}(L_1)$. Taking

$$f(x) = \int_0^x \Big(g(\xi) + \int_0^\xi h(s)\,\mathrm{d}x\Big)\,\mathrm{d}\xi$$

we readily compute

$$\int_0^1 f'(x)\Big(\overline{g(x)} + \int_0^x \overline{h(\xi)}\,\mathrm{d}\xi\Big)\,\mathrm{d}x = \int_0^1 \Big|g(x) + \int_0^x h(\xi)\,\mathrm{d}\xi\Big|^2\,\mathrm{d}x = 0.$$

This implies that

$$g(x) = -\int_0^x h(\xi)\,\mathrm{d}\xi,$$

giving $h = g'$ and $g(0) = 0$. From (7.5) we also have $g(1) = 0$, giving the desired "inclusion." ■

2. Next we take

$$\mathrm{dom}(L_2) = \{f \in S \mid\ f(0) = 0\}.$$

Although it is not completely obvious, one can readily show that $\mathrm{dom}(L_2)$ is dense in S, and therefore that $\mathrm{dom}(L_2)$ is dense in V. (This is done formally and more generally as part of the proof of Theorem 7.2.18, and involves showing that small changes at the endpoint to satisfy the endpoint can be chosen to affect the L_2-norm in an arbitrarily small way.) The following lemma records the adjoint in this case.

Lemma *If* $(\mathrm{L}_2^*, \mathrm{dom}(\mathrm{L}_2^*))$ *is the adjoint of* $(\mathrm{L}_2, \mathrm{dom}(\mathrm{L}_2))$ *then*

$$\mathrm{dom}(\mathrm{L}_2^*) = \{\mathrm{g} \in \mathrm{S} \mid\ \mathrm{g}(1) = 0\}$$

and $\mathrm{L}_2^*(\mathrm{g}) = -\mathrm{g}'$.

Proof This is Exercise 7.1.18. ■

3. Finally we consider

$$\mathrm{dom}(L_3) = \{f \in S \mid\ f(0) = f(1)\}.$$

The adjoint in this case turns out to be the same as the linear operator itself.

Lemma *If* $(\mathrm{L}_3^*, \mathrm{dom}(\mathrm{L}_3^*))$ *is the adjoint of* $(\mathrm{L}_3, \mathrm{dom}(\mathrm{L}_3))$ *then* $(\mathrm{L}_3^*, \mathrm{dom}(\mathrm{L}_3^*)) = (-\mathrm{L}_3, \mathrm{dom}(\mathrm{L}_3))$.

Proof This is Exercise 7.1.19. ■

The examples illustrate how the definition of the adjoint depends on the exact nature of not only the operator, but the domain on which the operator is defined. •

Of significant interest to us will be operators that are "self-adjoint." There are various notions related with this idea, so we state these notions formally.

7.1.26 Definition Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a Hilbert space with $(L, \mathrm{dom}(L))$ a linear operator for which $\mathrm{dom}(L)$ is dense in V.

(i) $(L, \mathrm{dom}(L))$ is ***self-adjoint*** if $(L, \mathrm{dom}(L)) = (L^*, \mathrm{dom}(L^*))$.

(ii) $(L, \mathrm{dom}(L))$ is ***symmetric*** if $\langle L(v), u\rangle = \langle v, L(u)\rangle$ for each $u, v \in \mathrm{dom}(L)$. •

7.1.2.3 Alternative theorems

7.1.3 Spectral properties of linear operators

While eigenvalues and eigenvectors for linear maps $L\colon \mathbb{R}^n \to \mathbb{R}^n$ are easy to characterise, eigenvalues, or more generally the spectrum, for a linear operator takes significantly more care. We should mention at this time that this careful consideration lies at the very heart of what culminates in Sections 7.2 and 7.3. Indeed, students wishing to truly understand that the subjects of Fourier series and Fourier transforms are united should know that this understanding starts with the content of this section.

7.1.3.1 Spectral properties for operators on Banach spaces We begin with a general discussion of the spectrum of a closed linear operator $(L, \mathrm{dom}(L))$ defined on a Banach space $(\mathsf{V}, \|\cdot\|)$, assuming that $\mathrm{dom}(L)$ is dense in V. We assume that V is a $\mathbb{C}$-vector space since we shall be dealing with complex scalar multiplication. Indeed, if $\lambda \in \mathbb{C}$ then denote $(L_\lambda, \mathrm{dom}(L_\lambda))$ as the linear operator defined by $\mathrm{dom}(L_\lambda) = \mathrm{dom}(L)$ and $L_\lambda(v) = L(v) - \lambda\,\mathrm{id}_{\mathsf{V}}(v)$. Recall that if L is closed then L is either regular or singular (i.e., it cannot be essentially regular). With this in mind, one defines the spectral properties of L as follows.

7.1.27 Definition Let $(\mathsf{V}, \|\cdot\|)$ be a Banach space with $(L, \mathrm{dom}(L))$ a linear operator for which $\mathrm{dom}(L)$ is dense in V.

(i) $\lambda \in \mathbb{C}$ is in the ***resolvent set*** for $(L, \mathrm{dom}(L))$ if $(L_\lambda, \mathrm{dom}(L_\lambda))$ is regular.

(ii) $\lambda \in \mathbb{C}$ is in the ***spectrum*** for $(L, \mathrm{dom}(L))$ if $(L_\lambda, \mathrm{dom}(L_\lambda))$ is singular. We denote the spectrum of $(L, \mathrm{dom}(L))$ by $\mathrm{spec}(L)$.

(iii) If $(L_\lambda, \mathrm{dom}(L_\lambda))$ is not invertible then λ is an ***eigenvalue*** for $(L, \mathrm{dom}(L))$ and nonzero vectors in $\ker(L_\lambda)$ are ***eigenvectors*** for $(L, \mathrm{dom}(L))$ corresponding to the eigenvalue λ. The dimension of $\ker(L_\lambda)$ is the ***multiplicity*** of λ. The collection of eigenvalues is the ***point spectrum*** of $(L, \mathrm{dom}(L))$ which we denote by $\mathrm{spec}_0(L)$.

(iv) If $(L_\lambda, \mathrm{dom}(L_\lambda))$ is invertible but

  (a) L_λ^{-1} is unbounded,

  (b) $\mathrm{image}(L_\lambda) \subset \mathsf{V}$, and

  (c) $\mathrm{cl}(\mathrm{image}(L_\lambda)) = \mathsf{V}$

then λ is in the ***continuous spectrum*** of $(L, \mathrm{dom}(L))$. The continuous spectrum of $(L, \mathrm{dom}(L))$ is denoted $\mathrm{spec}_1(L)$.

(v) If $(L_\lambda, \mathrm{dom}(L_\lambda))$ is invertible but $\mathrm{cl}(\mathrm{image}(L_\lambda)) \subset \mathsf{V}$ then λ is in the ***residual spectrum*** of $(L, \mathrm{dom}(L))$. The residual spectrum of $(L, \mathrm{dom}(L))$ is denoted $\mathrm{spec}_{-1}(L)$. The dimension of $\mathsf{V}/\mathrm{cl}(\mathrm{image}(L_\lambda))$ is the ***deficiency*** of λ. •

Note that our definition of deficiency in part (v) requires the notion of a quotient V/U of a vector space V by a subspace U. Readers unfamiliar with the notion of a quotient space need not despair since, as we shall shortly see, the linear operators of interest to us have empty residual spectrum. Let us give an example to illustrate our notions of spectrum.

7.1.28 Example On $(\mathsf{V} = \mathrm{L}_2([0,1];\mathbb{F}), \|\cdot\|_2)$ we consider the linear operator $(L, \mathrm{dom}(L) = \mathsf{V})$ defined by

$$L(f)(x) = \int_0^x f(\xi)\,\mathrm{d}\xi. \tag{7.6}$$

To examine the spectrum of L we need to consider the operator $L_\lambda = L - \lambda\,\mathrm{id}_{\mathsf{V}}$. First we take $\lambda = 0$ where $L_\lambda = L$. We note that image(L) consists of functions which vanish at $x = 0$ and which possess an L_2-derivative. The collection of all such functions is dense in $\mathrm{L}_2([0,1];\mathbb{F})$. Indeed, note that the differentiable functions vanishing at $x = 0$ are dense in image(L) by ***missing stuff***. One can also easily see that the differentiable functions vanishing at $x = 0$ are dense in the set of all differentiable functions. Thus image(L) is dense in V by ***missing stuff***. We claim that $(L, \mathrm{dom}(L))$ is invertible and that $(L^{-1}, \mathrm{image}(L))$ is unbounded. That L is invertible is clear since

$$\int_0^x f(\xi)\,\mathrm{d}\xi = 0 \quad \Longrightarrow \quad f(\xi) = 0 \text{ a.e.}$$

The unboundedness of L^{-1} follows since if f possess an L_2-derivative and vanishes at $x = 0$ then $L^{-1}(f) = f'$. We have seen in Example 7.1.11 that this map in unbounded. This shows that image$(L_0) \subset \mathsf{V}$, $\mathrm{cl}(\mathrm{image}(L_0)) = \mathsf{V}$, and that L_0^{-1} is unbounded. Thus $0 \in \mathrm{spec}_1(L)$.

Now we consider $\lambda \neq 0$. Here we claim that $(L_\lambda, \mathrm{dom}(L_\lambda))$ is regular. First let us show that it is invertible. Let $L_\lambda(f) = 0$. Thus

$$\int_0^x f(\xi)\,\mathrm{d}\xi - \lambda f(x) = 0 \quad \Longrightarrow \quad f'(x) - \frac{1}{\lambda} f(x) = 0.$$

The solution to this ordinary differential equation is $f(x) = Ce^{x/\lambda}$. Using the initial condition $f(0) = 0$ we see that $C = 0$, thus showing that L_λ is invertible. Next we show that image$(L_\lambda) = \mathsf{V}$. For $f \in \mathsf{V}$ we must find $g \in \mathrm{dom}(L)$ so that $L_\lambda(g) = f$, or equivalently

$$\int_0^x g(\xi)\,\mathrm{d}\xi - \lambda g(x) = f(x).$$

Define $h(x) = f(x) + \lambda g(x)$ so that

$$h(x) = \int_0^x g(\xi)\,\mathrm{d}\xi.$$

In particular, it follows that h possesses an L_2-derivative and that $h(0) = 0$. Therefore

$$g(x) = h'(r) \qquad \Longrightarrow \qquad h'(x) - \frac{1}{\lambda}h(x) = -\frac{1}{\lambda}f(x).$$

This equations can now be solved using an integrating factor, as you learned when you were a child, and the solution is

$$h(x) = -\frac{f(x)}{\lambda} - \frac{e^{x/\lambda}}{\lambda}\int_0^x f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi,$$

using the fact that $h(0) = 0$. Differentiating this then gives a function g satisfying $L_\lambda(g) = f$:

$$g(x) = -\frac{f(x)}{\lambda} - \frac{e^{x/\lambda}}{\lambda^2}\int_0^x f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi. \tag{7.7}$$

Thus $\mathrm{image}(L_\lambda) = \mathsf{V}$. Finally, we show that L_λ^{-1} is continuous. We compute, using (7.7),

$$\begin{aligned}
|L_\lambda^{-1}(f)(x)|^2 &= \Big|\frac{f(x)}{\lambda} + \frac{e^{x/\lambda}}{\lambda^2}\int_0^x f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\Big|^2 \\
&\le \frac{1}{|\lambda|^2}|f(x)|^2 + \frac{2M}{|\lambda^3|}|f(x)|\Big|\int_0^x f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\Big| \\
&\quad + \frac{M}{|\lambda|^4}\Big|\int_0^x f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\Big|^2
\end{aligned}$$

where

$$M = \sup_{x\in[0,1]}\{|e^{x/\lambda}|\}.$$

Now we use the Cauchy-Schwartz-Bunyakovsky inequality to further compute

$$\begin{aligned}
|L_\lambda^{-1}(f)(x)|^2 &\le \frac{1}{|\lambda|^2}|f(x)|^2 + \frac{2M}{|\lambda^3|}|f(x)|\Big(\int_0^x |f(\xi)|^2\,\mathrm{d}\xi\Big)\Big(\int_0^x |e^{-\xi/\lambda}|\,\mathrm{d}\xi\Big) \\
&\quad + \frac{M}{|\lambda|^4}\Big(\int_0^x |f(\xi)|^2\,\mathrm{d}\xi\Big)^2\Big(\int_0^x |e^{-\xi/\lambda}|\,\mathrm{d}\xi\Big)^2 \\
&\le a|f(x)|^2 + b|f(x)|\|f\| + c\|f\|^2,
\end{aligned}$$

where $a, b, c > 0$ are messy constants that are independent of f. Then we compute, again using the Cauchy-Schwartz-Bunyakovsky inequality,

$$\begin{aligned}\|L_\lambda^{-1}(f)\|^2 &= \int_0^1 |L_\lambda^{-1}(f)(x)|^2 \,\mathrm{d}x \\ &\leq \int_0^1 \Big(a|f(x)|^2 + b|f(x)|\|f\| + c\|f\|^2\Big)\,\mathrm{d}x \\ &= a\|f\|^2 + b\|f\|\Big(\int_0^1 \mathrm{d}x\Big)^{1/2}\Big(\int_0^1 |f(x)|^2\,\mathrm{d}x\Big)^{1/2} + c\|f\|^2 \\ &= (a+b+c)\|f\|_2.\end{aligned}$$

This shows that L_λ^{-1} is continuous for $\lambda \neq 0$.

Thus, all of the above shows the following: On $(\mathrm{L}_2([0,1];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ the linear L given in (7.6) satisfies

1. $\mathrm{spec}_0(L) = \emptyset$,
2. $\mathrm{spec}_1(L) = \{0\}$, and
3. $\mathrm{spec}_{-1}(L) = \emptyset$.

While some of the computations used to deduce these conclusions may be tedious, they are not essentially difficult. •

7.1.3.2 Spectral properties for operators on Hilbert spaces The eigenvalues and eigenvectors of a self-adjoint or symmetric linear operator have some useful properties. Let us first consider eigenvalues for symmetric linear operators. Note that the following result does *not* say that a symmetric linear *has* eigenvalues.

7.1.29 Theorem *Let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be an* $\mathbb{F}$*-inner product space and let* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *be a symmetric linear transformation on* V*. The following statements hold:*

(i) $\langle \mathrm{L}(\mathrm{v}), \mathrm{v}\rangle$ *is real for each* $\mathrm{v} \in \mathrm{dom}(\mathrm{L})$*;*

(ii) $\mathrm{spec}_0(\mathrm{L}) \subseteq \mathbb{R}$*;*

(iii) $\mathrm{spec}_1(\mathrm{L}) \subseteq \mathbb{R}$*;*

(iv) if λ_1 *and* λ_2 *are distinct eigenvalues for* L*, and if* v_i *is an eigenvector for* λ_i*,* $\mathrm{i} = 1, 2$*, then* $\langle \mathrm{v}_1, \mathrm{v}_2\rangle = 0$*.*

Proof (i) We have

$$\langle L(v), v\rangle = \langle v, L(v)\rangle = \overline{\langle L(v), v\rangle},$$

using the fact that L is symmetric.

(ii) Suppose that $\lambda \in \mathrm{spec}_0(L)$ and that $\lambda \neq 0$, otherwise the result is trivial. Let v be an eigenvector for λ and note that

$$\langle L(v), v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.$$

We also have, using the properties of the inner product,

$$\begin{aligned}\langle L(v), v\rangle &= \langle v, L(v)\rangle\\ &= \overline{\langle L(v), v\rangle}\\ &= \overline{\lambda\langle v, v\rangle}\\ &= \bar{\lambda}\langle v, v\rangle,\end{aligned}$$

since $\langle v, v\rangle$ is real. This shows that

$$\lambda\langle v, v\rangle = \bar{\lambda}\langle v, v\rangle,$$

giving $\bar{\lambda} = \lambda$ as $\langle v, v\rangle \neq 0$.

(iii) This is the most difficult part of the theorem, and to prove it we use two technical lemmas.

1 Lemma *If* $(\mathrm{L}, \mathrm{dom(L)})$ *is an invertible linear operator on a normed vector space* $(\mathsf{V}, \|\cdot\|)$ *for which* $(\mathrm{L}^{-1}, \mathrm{image(L)})$ *is unbounded, then there exists a sequence* $\{\mathrm{v_j}\}_{\mathrm{j}\in\mathbb{N}}$ *with the following properties:*

(i) $\|\mathrm{v_j}\| = 1$, $\mathrm{j} \in \mathbb{N}$;

(ii) $\|\mathrm{L(v_j)}\| < \frac{1}{\mathrm{j}}$, $\mathrm{j} \in \mathbb{N}$.

Proof Let $\mathbb{S} = \{v \in \mathrm{image}(L) \mid \|v\| = 1\}$. We claim that $L^{-1}(\mathbb{S}) \subseteq \mathsf{V}$ is unbounded. To see this, suppose that $L^{-1}(\mathbb{S})$ is bounded. Denote by $\overline{\mathsf{B}}(r, 0) = \{v \in \mathsf{V} \mid \|v\| \leq r\}$ the closed ball of radius r centred at $0 \in \mathsf{V}$. Since $L^{-1}(\mathbb{S})$ is bounded there exists $M > 0$ with the property that $L^{-1}(\overline{\mathsf{B}}(1, 0)) \subseteq \overline{\mathsf{B}}(M, 0)$. Now let $\epsilon > 0$. Choosing $\delta = \frac{\epsilon}{M}$ we see that $L^{-1}(\overline{\mathsf{B}}(\delta, 0)) \subseteq \overline{\mathsf{B}}(\epsilon, 0)$ by linearity of L^{-1}. This shows that if $L^{-1}(\mathbb{S})$ is bounded then L^{-1} is bounded.

Now, since $L^{-1}(\mathbb{S})$ is unbounded there exists a sequence $\{u_k\}_{k\in\mathbb{N}}$ in $\mathbb{S}$ so that $\lim_{k\to\infty}\|L^{-1}(u_k)\| = \infty$. Since $u_j \in \mathrm{image}(L)$ there exists a sequence $\{\tilde{v}_j\}_{j\in\mathbb{N}}$ in $\mathrm{dom}(L)$ so that $L(\tilde{v}_j) = u_j$, $j \in \mathbb{N}$. Therefore $L^{-1} \circ L(\tilde{v}_j) = \tilde{v}_j = L^{-1}(u_j)$. Thus $\lim_{j\to\infty}\|\tilde{v}_j\| = \infty$. Defining $\left\{v_j = \frac{\tilde{v}_j}{\|v_j\|}\right\}_{j\in\mathbb{N}}$ we see that $\lim_{j\to\infty}\|L(v_j)\| = \lim_{j\to\infty}\frac{\|u_j\|}{\|\tilde{v}_j\|} = 0$. Thus there exists a subsequence $\{v_{j_k}\}_{k\in\mathbb{N}}$ of $\{v_j\}_{j\in\mathbb{N}}$ having the property as asserted in the lemma. ▼

2 Lemma *If* $(\mathrm{L}, \mathrm{dom(L)})$ *is a symmetric linear operator on an inner product space* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *and if* $\lambda = \xi + \mathrm{i}\eta \in \mathbb{C}$ *then* $\|(\mathrm{L} - \lambda\,\mathrm{id_V})(\mathrm{v})\|^2 \geq \eta^2\|\mathrm{v}\|^2$ *for each* $\mathrm{v} \in \mathrm{dom(L)}$.

Proof We compute

$$\begin{aligned}\|(L - \lambda\,\mathrm{id}_\mathsf{V})(v)\|^2 &= \langle (L - \lambda\,\mathrm{id}_\mathsf{V})(v), (L - \lambda\,\mathrm{id}_\mathsf{V})(v)\rangle\\ &= \|L(v)\|^2 - \langle L(v), \lambda v\rangle - \langle \lambda v, L(v)\rangle + \|\lambda v\|^2\\ &= \|L(v)\|^2 - \overline{\lambda}\langle L(v), v\rangle - \lambda\langle L(v), v\rangle + (\xi^2 + \eta^2)\|v\|^2\\ &= \|L(v)\|^2 - 2\xi\langle L(v), v\rangle + \xi^2\|v\|^2 + \eta^2\|v\|^2\\ &= \|(L(v) - \xi\,\mathrm{id}_\mathsf{V})(v)\|^2 + \eta^2\|v\|^2\\ &\geq \eta^2\|v\|^2,\end{aligned}$$

as desired. ▼

We now proceed with the proof by showing that if $\mathrm{Im}(\lambda) \neq 0$ then $L - \lambda \mathrm{id}_{\mathsf{V}}$ is bounded. Let us write $\lambda = \xi + i\eta$. For any sequence $\{v_j\}_{j\in\mathbb{N}}$ with the property that $\|v_j\| = 1$, $j \in \mathbb{N}$, by Lemma 2 we have $\|(L - \lambda \mathrm{id}_{\mathsf{V}})(v_j)\| \geq |\eta|$ for $j \in \mathbb{N}$. Thus there exists $N \in \mathbb{N}$ so that $\|(L - \lambda \mathrm{id}_{\mathsf{V}})(v_j)\| > \frac{1}{j}$ provided that $j \geq N$. By Lemma 1 this means that $(L - \lambda \mathrm{id}_{\mathsf{V}})^{-1}$ must be bounded if $\mathrm{Im}(\lambda) \neq 0$, meaning that no such λ can lie in the continuous spectrum of L.

(iv) Let $\lambda_1, \lambda_2 \in \mathbb{R}$ and $v_1, v_2 \in \mathrm{dom}(L)$ be as specified. Then we compute

$$\begin{aligned}(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle &= \langle \lambda_1 v_1, v_2\rangle - \langle v_1, \lambda_2 v_2\rangle \\ &= \langle L(v_1), v_2\rangle - \langle v_1, L(v_2)\rangle \\ &= 0,\end{aligned}$$

using properties of the inner product and self-adjointness of L. Since $\lambda_1 \neq \lambda_2$, it follows that v_1 and v_2 are orthogonal as stated. ■

With part (i) of the theorem at hand, the following definition makes sense.

7.1.30 Definition Suppose that $(L, \mathrm{dom}(L))$ is a symmetric linear operator on $(\mathsf{V}, \langle\cdot,\cdot\rangle)$.

(i) $(L, \mathrm{dom}(L))$ is ***positive-definite*** if $\langle L(v), v\rangle \geq 0$ for each $v \in \mathsf{V}$ and $\langle L(v), v\rangle = 0$ only if $v = 0$.

(ii) $(L, \mathrm{dom}(L))$ is ***negative-definite*** if $(-L, \mathrm{dom}(L))$ is positive-definite. •

Now let us consider a further refinement that can be made for linear operators that are not only symmetric, but self-adjoint.

7.1.31 Theorem *If* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *is a self-adjoint linear operator on a Hilbert space* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *then* $\mathrm{spec}(\mathrm{L}) \subseteq \mathbb{R}$ *and* $\mathrm{spec}_{-1}(\mathrm{L}) = \emptyset$.

Proof Since a self-adjoint linear operator is symmetric, from Theorem 7.1.29 we need only show that $\mathrm{spec}_{-1}(L) = \emptyset$. We begin with a lemma that is of interest in its own right.

1 Lemma *Let* $(\mathrm{L}, \mathrm{dom}(\mathrm{L}))$ *be a linear operator, not necessarily self-adjoint, on a Hilbert space* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *with* $\mathrm{dom}(\mathrm{L})$ *dense in* V, *and let* $\lambda \in \mathrm{spec}_{-1}(\mathrm{L})$ *have deficiency* m. *Then* $\bar\lambda$ *is an eigenvalue of* L^* *with multiplicity* m.

Proof Note that $\dim(\mathsf{V}/\,\mathrm{cl}(\mathrm{image}(L_\lambda))) = \dim(\mathrm{image}(L_\lambda)^\perp)$. Indeed, for those familiar with the notation involved with quotient spaces, the map sending $v + \mathrm{cl}(\mathrm{image}(L_\lambda)) \in \mathsf{V}/\,\mathrm{cl}(\mathrm{image}(L_\lambda))$ to the orthogonal projection of v onto $\mathrm{image}(L_\lambda)^\perp$ is an isomorphism of $\mathsf{V}/\,\mathrm{cl}(\mathrm{image}(L_\lambda))$ with $\mathrm{image}(L_\lambda)^\perp$. For $v \in \mathrm{dom}(L)$ and $u \in \mathrm{image}(L_\lambda)^\perp$ we have $\langle (L - \lambda \mathrm{id}_{\mathsf{V}})(v), u\rangle = 0 = \langle v, 0\rangle$. This shows that $0 \in \mathrm{dom}((L - \lambda \mathrm{id}_{\mathsf{V}})^*)$ and that $(L - \lambda \mathrm{id}_{\mathsf{V}})^*(u) = 0$. Now note that $(L - \lambda \mathrm{id}_{\mathsf{V}})^* = L^* - \bar\lambda \mathrm{id}_{\mathsf{V}}$ (this is Exercise 7.1.9). This shows that $u \in \mathrm{image}(L_\lambda)^\perp$ is an eigenvector for L^* with eigenvalue $\bar\lambda$, as desired. ▼

Now we proceed with the proof. If $\lambda \in \operatorname{spec}_{-1}(L)$ then, since $(L, \operatorname{dom}(L))$ is self-adjoint and by Lemma 1, we know that $\lambda \in \operatorname{spec}_0(L)$. Thus $\operatorname{spec}_{-1}(L) \subseteq \mathbb{R}$. However, if $\lambda \in \mathbb{R}$ is in $\operatorname{spec}_{-1}(L)$ then Lemma 1 implies that λ is an eigenvalue of L^* and so an eigenvalue of L. However, points in $\operatorname{spec}_{-1}(L)$ cannot be eigenvalues, so the result follows. ■

This theorem is an important one, and we shall make use of it in Section 7.2 when talking about boundary value problems. The reader might also recall that if V is a *finite-dimensional* inner product space, then there is always a basis of orthogonal eigenvectors for a self-adjoint linear transformation. The reader is led through a proof of this in Exercise 7.1.12. In infinite-dimensions, things are more subtle. Indeed, in infinite dimensions it is possible that there be no eigenvalues, that there be finitely many eigenvalues, or that there be infinitely many eigenvalues. The first two of these possibilities is exhibited in Exercises 7.1.14 and 7.1.15.

Exercises

7.1.1 For $(\mathsf{V}, \|\cdot\|)$ a normed vector space, show that the function $\mathsf{V} \ni v \mapsto \|v\| \in \mathbb{R}$ is continuous.

7.1.2 For $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ an inner product space, show that the function $(v_1, v_2) \ni \mathsf{V}\times\mathsf{V} \mapsto \langle v_1, v_2\rangle \in \mathbb{F}$ is continuous.

7.1.3 On $C^0([0,1];\mathbb{R})$ consider the linear transformation L defined by

$$L(f)(x) = \int_0^x \xi f(\xi)\,\mathrm{d}\xi.$$

Answer the following questions, using as a norm $\|\cdot\|_\infty$.

(a) Show that L is continuous.

(b) Show that $\|L\| = 1$, with $\|\cdot\|$ the operator norm.

7.1.4 Show that the operator norm $\|\cdot\|_{\mathbb{R}^m,\mathbb{R}^n}$ defined in Example 7.1.5–1 is not derived from an inner product on $\mathrm{L}(\mathbb{R}^m;\mathbb{R}^n)$.

7.1.5 Let $(\mathsf{U}, \|\cdot\|_\mathsf{U})$ and $(\mathsf{V}, \|\cdot\|_\mathsf{V})$ be normed vector spaces. Show that if $\phi\colon A \subseteq \mathsf{U} \to \mathsf{V}$ is a continuous map, then for every convergent sequence $\{u_j\}_{j\in\mathbb{N}}$ in A, the sequence $\{L(u_j)\}_{j\in\mathbb{N}}$ is also convergent.

It is *a priori* not clear why a closed linear operator is referred to as "closed," particularly in light of the fact that we already have in mind a notion of closedness. In the next exercise you provide motivation for this terminology. You will need the following definition. For a map $f\colon S \to T$ between sets S and T, the ***graph*** if f is the set $\operatorname{graph} f = \{(x, f(x)) \mid x \in S\}$.

7.1.6 Show that a linear operator $(L, \operatorname{dom}(L))$ on a normed vector space $(\mathsf{V}, \|\cdot\|)$ is closed if and only if $\operatorname{graph} L \subseteq \operatorname{dom}(L) \times \mathsf{V}$ is closed.***missing stuff***

7.1.7 Show that the kernel of a closed linear operator is a closed subspace.

7.1.8 Let L be a continuous linear transformation of an inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$. Show that the resulting map L^* is linear.

7.1.9 Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $\lambda \in \mathbb{F}$. What is the adjoint of the linear operator (I_λ, V) defined by $I_\lambda(v) = \lambda v$?

7.1.10 On $(\mathsf{V} = \mathrm{L}_2([0,1];\mathbb{F}), \langle\cdot,\cdot\rangle)$ consider the linear operator $(L, \mathrm{dom}(L) = \mathsf{V})$ defined by

$$L(f)(x) = \int_0^x f(\xi)\,\mathrm{d}\xi.$$

Show that $\mathrm{dom}(L^*) = \mathsf{V}$ and that

$$L^*(f)(x) = \int_x^1 f(\xi)\,\mathrm{d}\xi.$$

7.1.11 If $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is an inner product space and if $L \in \mathrm{L}(\mathsf{V};\mathsf{V})$ is a self-adjoint linear map, continuous with respect to the norm defined by the inner product, show that the operator norm, which we denote by $|||\cdot|||$, satisfies

$$|||L||| = \sup_{\|v\|=1} \langle L(v), v\rangle.$$

In the following exercise you will be led through an unconventional proof that a self-adjoint linear transformation on a finite-dimensional vector space possesses a basis of eigenvectors. The proof we lead you through is designed to assist you when we come to the technical material in Section 7.2.

7.1.12 Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a finite-dimensional $\mathbb{R}$-inner product space and let $L\colon \mathsf{V} \to \mathsf{V}$ be a self-adjoint linear map. Define

$$\mathbb{S}_1 = \{v \in v \mid \|v\| = 1\}$$

to be the sphere of radius r in V in the norm $\|\cdot\|$ defined by the inner product $\langle\cdot,\cdot\rangle$.

(a) Show that $\mathbb{S}$ is closed and bounded.

Consider the function $\phi_1\colon \mathsf{V} \to \mathbb{R}$ defined by $\phi_1(v) = |\langle L(v), v\rangle|$.

(b) Argue that the restriction of ϕ_1 to $\mathbb{S}_1$ attains its maximum on $\mathbb{S}_1$.

Now recall the Lagrange multiplier theorem.

Lagrange multiplier theorem *On the finite-dimensional $\mathbb{R}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ let $\mathrm{f}, \mathrm{g}\colon \mathsf{V} \to \mathbb{R}$ be functions with g having the property that for every $\mathrm{v} \in \mathrm{g}^{-1}(0)$, $\mathrm{g}'(\mathrm{v}) \neq 0$. Then the derivative of the restriction of f to $\mathrm{g}^{-1}(0)$ vanishes at a point $\mathrm{v}_0 \in \mathrm{g}^{-1}(0)$ if and only if there exists $\lambda \in \mathbb{R}$ so that the derivative of the function*

$$\mathsf{V} \ni \mathrm{v} \mapsto \mathrm{f}(\mathrm{v}) + \lambda \mathrm{g}(\mathrm{v}) \in \mathbb{R}$$

vanishes at v_0.

Motivated by this, define $\psi_1 \colon \mathsf{V} \to \mathbb{R}$ by $\psi_1(v) = \langle v, v \rangle - 1$ so that $\mathbb{S}_1 = \psi_1^{-1}(0)$.

(c) Show that $\psi_1'(v) \neq 0$ for all $v \in \psi_1^{-1}(0)$.
Hint: *To differentiate a function on* V*, use an orthonormal basis for* V *to write the function in terms of the components of a point* $\mathrm{v} \in \mathsf{V}$*, and then differentiate in the usual manner (you may have seen before the derivative of a function on a vector space as the "gradient" of the function).*

Now note that by (b), the restriction of the function ϕ_1 to $\mathbb{S}_1$ attains its maximum on $\mathbb{S}_1$. Thus, at the point $v_1 \in \mathbb{S}_1$ where the restriction of ϕ_1 attains its maximum, the derivative of the restriction must vanish.

(d) Show that the point $v_1 \in \mathbb{S}_1$ where the restriction of ϕ_1 attains its maximum is an eigenvector for L.
Hint: *Use the Lagrange multiplier theorem, this being valid by (c).*
Hint: *There are two cases to consider: (1)* $\langle \mathrm{L}(\mathrm{v}_1), \mathrm{v}_1 \rangle > 0$ *and* $\langle \mathrm{L}(\mathrm{v}_1), \mathrm{v}_1 \rangle < 0$.

Let v_1 be as in part (d) and consider the subspace $\mathsf{V}_2 = v_1^\perp$ which is the orthogonal complement to $\operatorname{span}(v_1)$. Define $\phi_2 \colon \mathsf{V} \to \mathbb{R}$ by

$$\phi_2(v) = \phi_1(v - \langle v, v_1 \rangle v_1),$$

let $\mathbb{S}_2 = \mathbb{S}_1 \cap \mathsf{V}_2$, and define $\psi_2 \colon \mathsf{V}_2 \to \mathbb{R}$ by

$$\psi_2(v) = \langle v, v \rangle - 1.$$

(e) Show that there exists a linear map $L_2 \colon \mathsf{V}_2 \to \mathbb{R}$ so that $\phi_2(v) = |\langle L_2(v), v \rangle|$ for $v \in \mathsf{V}_2$.

(f) Using part (e), argue that the above procedure can be emulated to show that the point at which ϕ_2 attains its maximum on $\mathbb{S}_2$ is an eigenvector v_2 for L_2.

(g) Show that $v_2 \in \mathsf{V}_2$ is an eigenvector for L, as well as being an eigenvector for L_2.

(h) Show that this process terminates after at most $n = \dim(\mathsf{V})$ applications of the above procedure.
Hint: *Determine what causes the process to terminate?*

(i) Show that the procedure produces a collection, $\{v_1, \dots, v_n\}$ of orthonormal eigenvectors for L. Be careful that you handle properly the case when the above process terminates *before* n steps.

7.1.13 Come to grips with Exercise 7.1.12 in the case when $\mathsf{V} = \mathbb{R}^2$, $\langle \cdot, \cdot \rangle$ is the "dot product," and for each of the following three self-adjoint linear maps.

(a) $L(x, y) = (2x, y)$.

(b) $L(x, y) = (-x, 2y)$.

(c) $L(x, y) = (x, 0)$.

Thus you should in each case identify the maps ϕ_1 and ϕ_2, and show geometrically why maximising these functions picks off the eigenvectors as stated in Exercise 7.1.12.

7.1.14 Consider again the inner product space $(\mathrm{L}_2([0,1],\mathbb{R}),\langle\cdot,\cdot\rangle_2)$, and define a function $k\colon [0,1]\to\mathbb{R}$ by $k(x)=x$. Now define a linear transformation L_k by $(L_k(f))(x)=k(x)f(x)$. Show that L_k is self-adjoint, but has no eigenvalues.

7.1.15 Consider the inner product space $(\mathrm{L}_2([0,1],\mathbb{R}),\langle\cdot,\cdot\rangle_2)$, and define a function $k\colon [0,1]\to\mathbb{R}$ by

$$k(x)=\begin{cases}0, & x\in[0,\frac{1}{2})\\ 1, & x\in[\frac{1}{2},1].\end{cases}$$

Now define a linear transformation L_k as in Exercise 7.1.14. Show that the only eigenvalues for L_k are $\lambda_1=0$ and $\lambda_2=1$, and characterise all eigenvectors for each eigenvalue.

7.1.16 Let $(L,\mathrm{dom}(L))$ be an invertible linear operator on a normed vector space $(\mathsf{V},\|\cdot\|)$. Show that $\lambda\in\mathbb{F}$ is an eigenvalue for L with eigenvector v if and only if λ^{-1} is an eigenvalue for L^{-1} with eigenvector v.

7.1.17 Let L be a continuous linear transformation of an inner product space $(\mathsf{V},\langle\cdot,\cdot\rangle)$. Show that L is self-adjoint if and only if it is symmetric.

7.1.18 Prove that the adjoint in Example 7.1.25–2 is as stated in the lemma of that example.

7.1.19 Prove that the adjoint in Example 7.1.25–3 is as stated in the lemma of that example.

Section 7.2

Second-order regular boundary value problems

The discussion to this point has revolved around trigonometric polynomials. A curious student will wonder, however, whether other expansions are possible, or desirable. Indeed, we know that the set of L_2-integrable functions on an interval forms an separable Hilbert space, and as such admits a basis of orthogonal functions. The only such functions so far considered are trigonometric functions. The notion of separability does not suggest that there is a *distinguished* basis of orthogonal functions. Well, the trigonometric polynomials are, in fact, not the unique polynomials with their properties of orthogonality and of forming a basis. In this chapter we introduce other classes of polynomials, and some applications where they arise.

While in the title of this chapter we describe the problems we deal with as "regular," we will not actually say what this means until Section 7.3. This, however, will not be an impediment to understanding the content of the chapter.

7.2.1 Introductory examples

Before we launch off into fun generalities, it is worth looking at two simple examples. With each example, we attempt to accomplish something different. In Section 7.2.1.1 we look at a simple boundary value problem, one that we have seen before, and reveal some additional structure in this simple problem. The problem we look at in Section 7.2.1.2 is a simple problem along the lines of those in Chapter 6, but that exhibits some odd behaviour. This latter problem suggests that the problems we encountered in Chapter 6 are merely simple examples of a class of problems, and that a study of this class may be worth undertaking. (Worth it or not, we spend the remainder of the chapter, after this section, in this endeavour.)

7.2.1.1 Some structure for a simple boundary value problem In each of the partial differential equations of Chapter 6 we encountered a differential equation of the form

$$y''(x) = \lambda y(x),$$

with the boundary conditions $y(0) = y(\ell) = 0$. This is an example of a "second-order boundary value problem." We shall give a general definition for these as part of our general development that is to follow; our intention here is to provide a glimpse into the nature of such problems. Our first manoeuvre is to pose the problem as an eigenvalue problem. To do this, we define a linear map L from a subset of $L_2([0,\ell];\mathbb{R})$ into $L_2([0,1];\mathbb{R})$. The subset on which L is defined is called the ***domain*** of L and is denoted $\operatorname{dom}(L)$. The definition of $\operatorname{dom}(L)$ is an integral part of the definition of L, and this is one area where the development is somewhat

different than you are used to when dealing with linear maps in finite-dimensions. In any event, we define $\mathrm{dom}(L)$ to be the set of those functions $f \in \mathrm{L}_2([0,\ell];\mathbb{R})$ which satisfy

1. f is differentiable, i.e., $f \in C^1([0,\ell];\mathbb{R})$;
2. there exists a function $f'' \in \mathrm{L}_2([0,\ell],\mathbb{R})$ so that
$$f'(x) = f'(0) + \int_0^x f''(\xi)\,\mathrm{d}\xi; \tag{7.8}$$
3. $f(0) = f(\ell) = 0$.

Note that $C^2([0,\ell];\mathbb{R}) \subseteq \mathrm{dom}(L)$, but there are technical reasons, not discussed in detail here, for using the more general definition for $\mathrm{dom}(L)$.[3] With this definition for $\mathrm{dom}(L)$, we define $L\colon\ \mathrm{dom}(L) \to \mathrm{L}_2([0,\ell];\mathbb{R})$ by $L(f) = f''$ with f'' the function (not necessarily the second derivative of f!) satisfying (7.8). It is clear that $\mathrm{dom}(L)$ is a subspace of $\mathrm{L}_2([0,\ell];\mathbb{R})$, and that L is a surjective linear map.

The linear map L has some interesting properties with respect to the inner product $\langle\cdot,\cdot\rangle_2$ on $\mathrm{L}_2([0,\ell];\mathbb{R})$. For example, if $g \in \mathrm{dom}(L)$ then we compute

$$\begin{aligned}
\langle L(f), g\rangle_2 &= \int_0^\ell L(f)(x)g(x)\,\mathrm{d}x \\
&= \int_0^\ell f''(x)g(x)\,\mathrm{d}x \\
&= f'(x)g(x)\big|_0^\ell - \int_0^\ell f'(x)g'(x)\,\mathrm{d}x \\
&= -\int_0^\ell f'(x)g'(x)\,\mathrm{d}x \\
&= -f(x)g'(x)\big|_0^\ell + \int_0^\ell f(x)g''(x)\,\mathrm{d}x \\
&= \langle f, L(g)\rangle_2,
\end{aligned}$$

where we have twice used integration by parts. This implies that, when restricted to $\mathrm{dom}(L)$, L is self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_2$. Therefore, ***missing stuff*** suggests that the eigenvalues for L, if there be any at all, will be real, and that eigenvectors for distinct eigenvalues will be orthogonal. However, it is perhaps not clear that L has eigenvalues, although if one thinks about it for a moment, it clearly does. Indeed, if λ is an eigenvalue with eigenvector f then we have

$$f'' = \lambda f, \quad f(0) = f(\ell) = 0.$$

[3] If we were to take $\mathrm{dom}(L) = C^2([0,\ell];\mathbb{R})$, then L is not a closed operator, meaning that L does not map closed sets to closed sets. Also, if we take this restricted definition for $\mathrm{dom}(L)$, then L as we define it is not surjective. It turns out that either of these provide adequate reason for taking a more general definition for $\mathrm{dom}(L)$.

But we have seen this already with the partial differential equations of Chapter 6! Indeed, there we determined that there were infinitely many such eigenvalues, and these were of the form

$$\lambda_n = -\frac{n^2\pi^2}{\ell^2}, \quad n \in \mathbb{N}.$$

Corresponding to these eigenvalues were the eigenvectors

$$f_n(x) = \sqrt{\frac{2}{\ell}} \sin\Big(n\pi\tfrac{x}{\ell}\Big), \quad n \in \mathbb{N}.$$

Here we add a normalisation factor of $\sqrt{\frac{2}{\ell}}$ to ensure that these eigenvectors have norm 1 with respect to the inner product $\langle \cdot, \cdot \rangle_2$. Although we glossed over this in Chapter 6, let us show that these eigenvectors are complete in $\mathrm{L}_2([0,\ell];\mathbb{R})$.

7.2.1 Proposition *The functions $\{\mathrm{f_n}\}_{n\in\mathbb{N}}$ are a complete orthonormal family in $\mathrm{L}_2([0,\ell];\mathbb{R})$.*

Proof By ***missing stuff*** it suffices to show that if $f \in \mathrm{L}_2([0,\ell];\mathbb{R})$ is orthogonal to all of the functions f_n, $n \in \mathbb{N}$, then $f = 0$. Let $f_{n,\mathrm{odd}}$ be the odd extension of f_n and let f_{odd} be the odd periodic extension of an arbitrary function f. Also define

$$e_0 = \frac{1}{\sqrt{\ell}}, \quad e_n(x) = \sqrt{\frac{2}{\ell}} \cos\Big(n\pi\tfrac{x}{\ell}\Big), \quad n \in \mathbb{N},$$

with $e_{n,\mathrm{even}}$ the even extension. By ***missing stuff*** the family of functions $\{f_{n,\mathrm{odd}}\}_{n\in\mathbb{N}} \cup \{e_{n,\mathrm{even}}\}_{n\in\mathbb{N}_0}$ is a complete orthonormal family in $\mathrm{L}^{\mathrm{per}}_2([0,2\ell];\mathbb{R})$. Therefore, if f_{odd} is orthogonal to all of the functions in the family $\{f_{n,\mathrm{odd}}\}_{n\in\mathbb{N}} \cup \{e_{n,\mathrm{even}}\}_{n\in\mathbb{N}_0}$, it must be the zero function. However, since f_{odd} is odd, it is orthogonal to $e_{n,\mathrm{even}}$, $n \in \mathbb{N}_0$ by ***missing stuff***. Thus we may conclude that if f_{odd} is orthogonal to $f_{n,\mathrm{odd}}$, $n \in \mathbb{N}$, then f_{odd} is zero. However, since

$$\int_0^{2\ell} f_{\mathrm{odd}}(x) f_{n,\mathrm{odd}}(x)\,\mathrm{d}x = 2\int_0^{\ell} f(x) f_n(x)\,\mathrm{d}x$$

by oddness of f_{odd} and $f_{n,\mathrm{odd}}$, it then follows that if f is orthogonal to f_n, $n \in \mathbb{N}$, then it must be zero. This completes the proof. ■

Thus we see that in this case, the self-adjoint linear transformation L has a property like self-adjoint transformations on finite-dimensional vector spaces: it possess a basis (in the appropriate sense) of eigenvectors. However, as we saw in Exercises 7.1.14 and 7.1.15, this property cannot be expected on the basis of L being merely self-adjoint. There must be some additional structure in the definition of L that guarantees its possessing a basis of eigenvectors. Let us begin to unlock some of this structure by characterising $\mathrm{dom}(L)$ in terms of Fourier series. The following characterisation of $\mathrm{dom}(L)$ is also useful.

7.2.2 Proposition dom(L) *consists exactly of those functions* $\mathrm{f} \in \mathrm{L}^2([0,\ell];\mathbb{R})$ *which satisfy*

$$\sum_{\mathrm{n}=1}^{\infty} \mathrm{n}^4 |\langle \mathrm{f}, \mathrm{f}_\mathrm{n} \rangle|^2 < \infty.$$

Idea of Proof This result follows in the mold of the statements of***missing stuff***. ■

With this result at hand, if we write $f \in \mathrm{dom}(L)$ as

$$f(x) = \sum_{n=1}^{\infty} c_n f_n,$$

where $c_n = \langle f, f_n \rangle$, then we can differentiate this expression term-by-term to get

$$L(f) = f'' = -\sum_{n=1}^{\infty} \frac{n^2\pi^2}{\ell^2} c_n f_n. \tag{7.9}$$

Thus, in the basis $\{f_n\}_{n\in\mathbb{N}}$ for $\mathrm{L}_2([0,\ell];\mathbb{R})$, L works "diagonally." This is hardly surprising since the functions f_n, $n \in \mathbb{N}$, are eigenvectors for L.

A further key to understanding why L should possess a basis of eigenvectors comes from looking at the inverse of L. First we should show that L is indeed invertible.

7.2.3 Proposition *The map* $\mathrm{L}\colon\ \mathrm{dom}(\mathrm{L}) \to \mathrm{L}_2([0,\ell];\mathbb{R})$ *as defined is invertible.*

Proof By the very construction of L, it is surjective, as the function f'' in (7.8) is arbitrary in $\mathrm{L}_2([0,\ell];\mathbb{R})$. To show that L is injective we need only show that $\ker(L) = \{0\}$ (why?). But for this we note that if $L(f) = 0$ then $f'' = 0$, which gives f' as a constant function by (7.8). This means that $f = ax + b$ for some $a, b \in \mathbb{R}$. However, the only such function satisfying the boundary conditions is $f = 0$, thus showing that L is injective, and so invertible. ■

Note that there is something going on here that cannot happen in finite-dimensional vector spaces. The situation we have is a linear mapping defined on a proper subspace U of a vector space V, mapping into V itself: $L\colon \mathsf{U} \subset \mathsf{V} \to \mathsf{V}$. In finite-dimensions it is not possible for L to be invertible as $\dim(\mathsf{U}) < \dim(\mathsf{V})$. However, apparently this scenario *is* possible in infinite-dimensions. This is perhaps counterintuitive. Nonetheless, let us press on. An essential part of why the linear map L should possess a basis of eigenvectors is not only the existence of an inverse to L, but the nature of this inverse. For reasons of convention, let us denote $\mathcal{G} = -L^{-1}\colon \mathrm{L}_2([0,\ell] \to \mathrm{dom}(L)$. The following result gives the form of $\mathcal{G}$.

7.2.4 Proposition *Let* $f \in L_2([0,\ell];\mathbb{R})$ *and suppose that* $u \in \mathrm{dom}(L)$ *satisfies* $u = \mathcal{G}(f)$. *Then*

$$u(x) = \int_0^\ell G(x,y)f(y)\,dy,$$

where $G\colon [0,\ell]\times[0,\ell] \to \mathbb{R}$ *is the function defined by*

$$G(x,y) = \begin{cases} (\ell - x)y, & y < x \\ x(\ell - y), & y \geq x. \end{cases}$$

Furthermore,

$$G(x,y) = \sum_{n=1}^{\infty} \frac{2\ell^2}{n^2\pi^2} \sin\Big(n\pi\tfrac{x}{\ell}\Big)\sin\Big(n\pi\tfrac{y}{\ell}\Big).$$

Proof Note that u satisfies

$$L(u) = L \circ \mathcal{G}(f) = -f.$$

Thus $u'' = -f$ and $u(0) = u(\ell) = 0$. We may determine u by integrating. First we have

$$u'(x) = u'(0) - \int_0^x f(y)\,dy,$$

giving***missing stuff***

$$\begin{aligned} u(x) &= u(0) + u'(0)x - \int_0^x \Big(\int_0^y f(z)\,dz\Big)dy \\ &= u'(0)x - \int_0^x (x-y)f(y)\,dy. \end{aligned}$$

Substituting $x = \ell$ and noting that $u(\ell) = 0$ this gives

$$u'(0) = \int_0^\ell (\ell - y)f(y)\,dy.$$

Now we substitute this into our expression for $u(x)$ to get

$$\begin{aligned} u(x) &= u'(0)x - \int_0^x (x-y)f(y)\,dy \\ &= \int_0^\ell x(\ell - y)f(y)\,dy - \int_0^x (x-y)f(y)\,dy \\ &= \int_0^x \Big(x(\ell - y) - (x-y)\Big)f(y)\,dy + \int_x^\ell x(\ell - y)f(y)\,dy \\ &= \int_0^x (\ell - x)yf(y)\,dy + \int_x^\ell x(\ell - y)f(y)\,dy \\ &= \int_0^\ell G(x,y)f(y)\,dy, \end{aligned}$$

giving the first part of the result.

For the second part of the result we use (7.9), and submit to glossing over details of convergence. Suffice it to say that all the operations we perform are legal. In any case, if we look at (7.9), we can simply read off $\mathcal{G} = -L^{-1}$ in terms of the basis functions $\{f_n\}_{n\in\mathbb{N}}$:

$$\mathcal{G}(f) = \sum_{n=1}^{\infty} \frac{\ell^2}{n^2\pi^2} c_n f_n.$$

Now we compute

$$\begin{aligned}\mathcal{G}(f)(x) &= \sum_{n=1}^{\infty} \frac{\ell^2}{n^2\pi^2} \langle f, f_n\rangle f_n(x)\\ &= \sum_{n=1}^{\infty} \frac{\ell^2}{n^2\pi^2} f_n(x) \int_0^\ell f(y) f_n(y)\,\mathrm{d}y\\ &= \int_0^\ell \Big(\sum_{n=1}^{\infty} \frac{2\ell^2}{n^2\pi^2} \sin\Big(n\pi\tfrac{x}{\ell}\Big)\sin\Big(n\pi\tfrac{y}{\ell}\Big)\Big) f(y)\,\mathrm{d}y.\end{aligned}$$

Our result now follows by comparing this last expression with that derived in the first part of the proof. ■

Let us summarise what we have done in this section, as we shall encounter these ideas in general in Section 7.2.2, and there you will probably want to refer back to this simple example to ground yourself.

1. We have constructed a linear mapping L from a domain $\mathrm{dom}(L)$ into $\mathrm{L}_2([0,\ell];\mathbb{R})$, with the domain being specified by the character of the operator, as well as by boundary conditions.
2. L is shown to possess a countable set of eigenvalues, tending to infinity.
3. The corresponding eigenvectors are shown to form a complete orthonormal family.
4. Corresponding to L is its (essentially) inverse $\mathcal{G}$. This is defined by the use of the function G which is known as the ***Green function*** for the mapping L, after George Green (1793–1841). The Green function is represented in two ways, one coming from a more or less direct computation, and the other from a representation of L is the basis of its own eigenvectors. At this point, the significance of the Green function should be lost on you. However, in our general development, it will play a key rôle.

7.2.1.2 A boundary value problem with peculiar eigenvalues The example we give arises from a partial differential equation with boundary conditions. We shall not get into the details of the physical setup, as that is peripheral to our

intentions at the moment. Let us simply produce the problem:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \qquad \begin{aligned} &u(0,t) = 0, \quad -u(1,t) = \frac{\partial u}{\partial x}(1,t) \\ &u(x,0) = f(x), \quad \frac{\partial u}{\partial t}(x,0) = g(x). \end{aligned}$$

Thus the equation is a wave equation with a rather peculiar boundary condition at $x = 1$. Physically the boundary condition arises when the string at $x = 1$ is not fixed, but is attached to a spring that allows vertical movement of the string. Obviously, for simplicity we have set some of the physical constants to 1. Let us not go through all the details of the separation of variables, but simply produce that part of it that is relevant to our discussion here. If we take $u(x,t) = X(x)T(t)$ in the usual manner, then the problem for X reduces to

$$X''(x) = \lambda X(x), \quad X(0) = 0, \; -X(1) = X'(1).$$

To solve this boundary value problem for X, we go through the various possibilities for λ.

1. $\lambda > 0$. Here the solutions to the differential equation have the form

   $$X(x) = A \sinh(\sqrt{\lambda}x) + B \cosh(\sqrt{\lambda}x).$$

   The boundary condition $X(0) = 0$ gives $B = 0$. The boundary condition $-X(1) = X'(1)$ gives

   $$\begin{aligned} &-A \sinh \sqrt{\lambda} = A\sqrt{\lambda} \cosh \sqrt{\lambda} \\ \Longrightarrow \quad &\sqrt{\lambda} = -\tanh \sqrt{\lambda}, \end{aligned}$$

   assuming that we do not allow $A = 0$. Note that for $x > 0$ the function $\tanh x$ is positive. Thus we are allowed to disavow λ from being positive.
2. $\lambda = 0$: In this case the differential equation for X has solutions $X(x) = Ax + B$. The boundary condition $X(0) = 0$ gives $B = 0$. The boundary condition for $-X(1) = X'(1)$ gives $A = -A$, which is clearly nonsense. Thus λ is nonzero.
3. $\lambda < 0$: We seek resort in the last case; funny how that always works out. In any case, the solutions to the differential equation for X are

   $$X(x) = A \sin(\sqrt{-\lambda}x) + B \cos(\sqrt{-\lambda}x).$$

   The boundary condition $X(0) = 0$ gives $B = 0$ and the boundary condition $-X(1) = X'(1)$ reads

   $$\begin{aligned} &-A \sin \sqrt{-\lambda} = A\sqrt{-\lambda} \cos \sqrt{-\lambda} \\ \Longrightarrow \quad &-\sqrt{-\lambda} = \tan \sqrt{-\lambda}, \end{aligned}$$

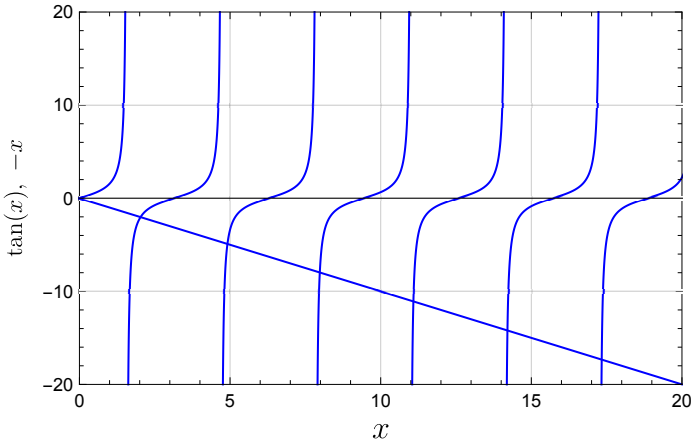


Figure 7.2 Roots for $\tan x = -x$

if we assume that $A \neq 0$. This equation has roots, although we cannot give a useful closed-form expression for them. Indeed, in Figure 7.2 we graph $\tan x$ and x on the same axes, and one readily sees that there are an infinite number of solutions to the equation $\tan x = -x$ for $x > 0$.

If we adopt the notation of Section 7.2.1.1 we define L in the same way, but now we take $\mathrm{dom}(L)$ to be the set of those functions $f \in \mathrm{L}_2([0,1];\mathbb{R})$ which satisfy

1. f is differentiable, i.e., $f \in C^1([0,1];\mathbb{R})$;
2. there exists a function $f'' \in \mathrm{L}_2([0,1],\mathbb{R})$ so that

$$f'(x) = f'(0) + \int_0^x f''(\xi)\,\mathrm{d}\xi;$$

3. $f(0) = 0$ and $-f(1) = f'(1)$.

What we determined above is that the eigenvalues are the solutions $\{\lambda_n\}_{n\in\mathbb{N}}$ which satisfy $-\sqrt{-\lambda_n} = \tan\sqrt{-\lambda_n}$. Again, we have no closed-form expression for these. The eigenvectors corresponding to these eigenvalues are the functions $\{f_n\}_{n\in\mathbb{N}}$ in $\mathrm{dom}(L)$ of the form

$$f_n = A_n \sin(\sqrt{-\lambda_n}x), \quad n \in \mathbb{N},$$

where the constant A_n is defined so that $\langle f_n, f_n\rangle = 1$. One can compute

$$A_n = \Big(\frac{1}{2} - \frac{\sin(2\sqrt{-\lambda_n})}{4\sqrt{-\lambda_n}}\Big)^{-1/2}, \quad n \in \mathbb{N}.$$

Thus the *form* of the eigenvectors resembles those for the simple boundary value problem of Section 7.2.1.1, but the frequencies of the sinusoids are no longer as nice as $\frac{n\pi}{\ell}$. It is now not so easy to argue that the analogue of Proposition 7.2.1 is true. In fact, it is downright difficult. What we will now do is formulate a general

problem, of which the problem in this section is an example, and show that the situation of Section 7.2.1.1 is repeated for all problems in this class. This is a rather remarkable turn of events, as I hope you can realise by thinking about the examples in this section.

7.2.2 Sturm-Liouville problems

We shall formulate in this section a class of second-order boundary value problems known as ***Sturm-Liouville problems***, after the two French mathematicians Jacques Charles François Sturm (1803–1855) and Joseph Liouville (1809–1882). It is possible to formulate more general boundary value problems with, for example, higher-order differential equations. Such problems do come up in applications, and are common in optimal control, for example. However, our focus on second-order problems is motivated by the frequency of their appearance in describing problems such as those encountered in Chapter 6. This also allows us to be slightly more concrete, and the more ambitious reader will have no problem imaging the generalisations, then going to references.

7.2.2.1 Second-order boundary value problems In this section we formulate general boundary value problems, then reduce these to problems that are self-adjoint, as these will be most interesting for us. The differential equations we consider are of the form

$$p_2(x)y''(x) + p_1(x)y'(x) + p_0(x)y(x) = 0, \tag{7.10}$$

and are defined on the interval $[a, b]$. We suppose that $p_k\colon [a, b] \to \mathbb{R}$, $k \in \{0, 1, 2\}$, is k times continuously differentiable, and that $p_2(x) \neq 0$ for $x \in [a, b]$. In Section 7.3 we shall consider cases where this latter assumption is dropped, as these do arise in practise. As a second-order differential equation, (7.10) is entitled to two subsidiary conditions to fully determine its solutions. The conditions we consider are of the form

$$\alpha_1 y'(a) + \alpha_0 y(a) = 0, \quad \beta_1 y'(b) + \beta_0 y(b) = 0, \quad \alpha_1^2 + \alpha_0^2, \beta_1^2 + \beta_0^2 \neq 0. \tag{7.11}$$

Thus we allow the specification of any nontrivial linear combination of the function and its derivatives at each endpoint of the interval $[a, b]$. This certainly includes all examples and exercises of Chapter 6, although more general possibilities can be introduced which permit, for example, mixing the conditions from the two boundary points. However, we shall stick to the conditions (7.11). The differential equation (7.10) and the boundary conditions (7.11) arise from a linear mapping, just as we saw in Section 7.2.1.1. We denote this mapping by L, and as expected, we first need to define its domain $\operatorname{dom}(L)$. We take $\operatorname{dom}(L)$ to be the set of functions $f\colon [a, b] \to \mathbb{R}$ having the properties

1. f is differentiable;

2. there exists a function $f'' \in \mathrm{L}_2([0,\ell],\mathbb{C})$ so that

$$f'(x) = f'(0) + \int_0^x f''(\xi)\,\mathrm{d}\xi; \tag{7.12}$$

3. $\alpha_1 f'(a) + \alpha_2 f(a) = 0$ and $\beta_1 f'(b) + \beta_2 f(b) = 0$.

With $\mathrm{dom}(L)$ thus defined, we define $L\colon\ \mathrm{dom}(L) \to \mathrm{L}_2([a,b];\mathbb{R})$ by

$$L(f)(x) = p_2(x)f''(x) + p_1(x)f'(x) + p_0 f(x),$$

where $f'' \in \mathrm{L}_2([a,b];\mathbb{R})$ is given by (7.12).

In the preceding discussion, no mention has been made of self-adjointness, a notion that figured prominently in Section 7.2.1.1. In fact, the linear map L that we just defined will not generally be self-adjoint. In order to ensure that it *is* self-adjoint, some conditions are needed on the coefficients p_0, p_1, and p_2. A complete classification of self-adjointness does not seem to be easy to come by, and is not necessary for our purposes in any case. What we do is give a class of self-adjoint problems. First comes the definition of a new linear map $L^+\colon C^2([a,b];\mathbb{R}) \to \mathrm{L}_2([a,b];\mathbb{R})$ given by

$$L^+(f) = (p_2 f)'' - (p_1 f)' + (p_0 f).$$

It is not the case that L^+ is the adjoint of L. However, it is a mapping that is related to the adjoint in a manner given to us by the following result.

7.2.5 Proposition *For* $\mathrm{f}, \mathrm{g} \in \mathrm{C}^2([\mathrm{a},\mathrm{b}];\mathbb{R})$ *we have*

$$\mathrm{gL(f)} - \mathrm{fL}^+(\mathrm{g}) = [\mathrm{fg}]'$$

where

$$[\mathrm{fg}] = \mathrm{p}_1\mathrm{fg} + \mathrm{p}_2\mathrm{f}'\mathrm{g} - \mathrm{f}(\mathrm{p}_2\mathrm{g})'.$$

Proof This is a direct, if tedious, computation. ■

The upshot of this that is of value for us is the following result, which follows from integrating the conclusion of the proposition.

7.2.6 Corollary (Green's formula) $\langle \mathrm{f}, \mathrm{L(g)}\rangle_2 - \langle \mathrm{g}, \mathrm{L}^+(\mathrm{f})\rangle_2 = [\mathrm{fg}](\mathrm{b}) - [\mathrm{fg}](\mathrm{a})$.

Thus this provides a somewhat easy to understand class of boundary value problems that are self-adjoint. Indeed, if $L = L^+$[4] and if $f, g \in \mathrm{dom}(L)$ ensures that $[fg](a) = [fg](b) = 0$, then L is self-adjoint. The following result gives the form of L in such cases.

[4]Here we abuse notation a little. When we write $L = L^+$ we mean that both expressions give the same result when applied to functions in $C^2([a,b];\mathbb{R})$. The abuse of notation is that L, as defined, only takes as argument functions in $\mathrm{dom}(L)$.

7.2.7 Proposition $\mathrm{L} = \mathrm{L}^+$ *if and only if* $\mathrm{p}_1 = \mathrm{p}_2'$. *Furthermore, if* $\mathrm{L} = \mathrm{L}^+$ *then the boundary conditions (7.11) ensure that* L *is self-adjoint.*

Proof $L = L^+$ if and only if

$$\begin{aligned} & p_2 f'' + 2p_2' f' + p_2'' f - p_1 f' - p_1' f + p_0 f = p_2 f'' + p_1 f' + p_0 f \\ \iff \quad & 2p_2' f' + p_2'' f - 2p_1 f' - p_1' f = 0. \end{aligned}$$

If this is to be true for all $f \in C^2([a,b];\mathbb{R})$ then we should have $2p_2' = 2p_1$ and $p_2'' = p_1'$, thus giving the first assertion of the proposition. For the second assertion we use the fact that $p_1 = p_2'$ so that

$$[fg] = p_2' fg + p_2 f' g - p_2 f g' - p_2' f g' = p_2 f' g - p_2 f g'.$$

We then have, for $f, g \in \operatorname{dom}(L)$,

$$[fg](a) = p_2(a)\Big(f'(a)g(a) - f(a)g'(a)\Big).$$

Now, since $f, g \in \operatorname{dom}(L)$ we have

$$\begin{aligned} & \alpha_1 f'(a) + \alpha_0 f(a) = 0, \quad \alpha_1 g'(a) + \alpha_0 g(a) = 0 \\ \implies \quad & \begin{bmatrix} f'(a) & f(a) \\ g'(a) & g(a) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \implies \quad & \det \begin{bmatrix} f'(a) & f(a) \\ g'(a) & g(a) \end{bmatrix} = 0 \\ \implies \quad & f'(a)g(a) - f(a)g'(a) = 0, \end{aligned}$$

since $\alpha_1^2 + \alpha_0^2 \neq 0$. From this we conclude that $[fg](a) = 0$. We similarly conclude that $[fg](b) = 0$, and so from Corollary 7.2.6 and from the first part of the proposition that L is self-adjoint. ■

7.2.8 Remark It is possible to arrive at the consequences of the preceding proposition via "direct" means, involving integration by parts. While this seems to involve less trickery than the above slick derivation, the above derivation is preferable for a few reasons, including that it is easily generalised to provide conditions for self-adjointness of a general class of boundary value problems, that it is less yucky than successive applications of integration by parts, and because it introduces the generally interesting linear mapping L^+. •

Tradition dictates that we write $p_2 = p$, $p_1 = p_2'$ (this by Proposition 7.2.7), and $p_0 = -q$. We have thus arrived at a class of self-adjoint second-order boundary value problems. Let us restate this, using the notation $L_{p,q}$ for L, reflecting the form for the linear mapping arrived at above:

$$L_{p,q}(y) = (py')' - qy = 0 \qquad \begin{aligned} & \alpha_1 y'(a) + \alpha_0 y(a) = 0 \\ & \beta_1 y'(b) + \beta_0 y(b) = 0. \end{aligned}$$

We denote by $\operatorname{dom}(L_{p,q})$ the domain as defined above. This problem forms the basis for the discussion of this section.

Before we launch into the details of the properties of the above self-adjoint boundary value problem, or more properly the slight generalisation of it formulated in Section 7.2.2.2, let us explore a way in which a problem not in the form of $L_{p,q} = 0$ can be transformed into a problem of that type. That is to say, we wish to ascertain if there are differential equations of the form

$$p_2(x)y''(x) + p_1(x)y'(x) + p_0(x)y(x) = 0 \tag{7.13}$$

which are *not* of the form $L_{p,q}(y) = 0$ (in other words, we do not *a priori* have $p_1 = p_2'$) but which may be transformed into an equation of that type. The following result records an instance when this is possible.

7.2.9 Proposition *Consider the differential equation (7.13) defined on the interval* $\mathrm{I} \subseteq \mathbb{R}$ *and suppose that* $\mathrm{p}_2(\mathrm{x}) > 0$ *for all* $\mathrm{x} \in \mathrm{I}$. *If* $\frac{\mathrm{p}_1}{\mathrm{p}_2}$ *is integrable on* I, *then* y *is a solution to this differential equation if and only if* $\mathrm{L}_{\mathrm{p,q}}(\mathrm{y}) = 0$ *where*

$$\mathrm{p}(\mathrm{x}) = \exp\Big(\int_{\mathrm{x}_0}^{\mathrm{x}} \frac{\mathrm{p}_1(\xi)}{\mathrm{p}_2(\xi)}\,\mathrm{d}\xi\Big), \quad \mathrm{q}(\mathrm{x}) = -\frac{\mathrm{p}_0(\mathrm{x})}{\mathrm{p}_2(\mathrm{x})}\mathrm{p}(\mathrm{x}).$$

Proof Suppose that y is a solution to (7.13), and let p be as specified in the statement of the proposition. Then

$$\begin{aligned}
& p_2(x)y''(x) + p_1(x)y'(x) + p_0(x)y(x) = 0 \\
\iff \quad & y''(x) + \frac{p_1(x)}{p_2(x)}y'(x) + \frac{p_0(x)}{p_2(x)}y(x) = 0 \\
\iff \quad & p(x)\Big(y''(x) + \frac{p_1(x)}{p_2(x)}y'(x) + \frac{p_0(x)}{p_2(x)}y(x)\Big) = 0 \\
\iff \quad & \frac{\mathrm{d}}{\mathrm{d}x}\Big(p(x)y'(x)\Big) - q(x)y(x) = 0,
\end{aligned}$$

using the fact that

$$p'(x) = \frac{p_1(x)}{p_2(x)}p(x).$$

■

This tells us that a large number of systems can be put into the form of $L_{p,q} = 0$, thereby improving the applicability of the techniques we discuss.

7.2.2.2 A general eigenvalue problem The optimist would think that the eigenvalue problem of interest would simply be that of finding eigenvalues for the linear mapping $L_{p,q}$ discussed in the preceding section. But nooo, this is far too easy. More to the point, however, is that it would omit a class of problems of physical relevance. The good news is that the linear mapping $L_{p,q}\colon \operatorname{dom}(L_{p,q}) \to \mathrm{L}_2([a,b];\mathbb{R})$ forms the starting point for our slightly more general problem.

We let $r\colon [a,b] \to \mathbb{R}$ be a continuous function that is nowhere zero on $[a,b]$. We may as well suppose, therefore, that $r(x) > 0$ for all $x \in [a,b]$. The problem we now consider is the following:

$$\boxed{\begin{aligned} L_{p,q,r}(y) = r^{-1}\Big((py')' - qy\Big) = 0 \qquad & \alpha_1 y'(a) + \alpha_0 y(a) = 0 \\ & \beta_1 y'(b) + \beta_0 y(b) = 0. \end{aligned}} \tag{7.14}$$

Note that the solutions of the differential equation $L_{p,q,r}(y) = 0$ are the same as the solutions of the differential equation $L_{p,q}(y) = 0$. Thus we are justified in denoting $\operatorname{dom}(L_{p,q,r}) = \operatorname{dom}(L_{p,q})$. However, what is different is the eigenvalues and eigenvectors for $L_{p,q,r}$ and $L_{p,q}$. First we should verify that $L_{p,q,r}$ is self-adjoint. To see this, we need to define a special inner product by

$$\langle f, g\rangle_r = \int_a^b f(x)g(x)r(x)\,\mathrm{d}x.$$

This does indeed define an inner product on $\mathrm{L}_2([a,b];\mathbb{R})$, as was essentially verified in Exercise 7.2.1. Let $\|\cdot\|_r$ denote the norm defined by this inner product. One readily verifies that $\|f\|_r < \infty$ if and only if $\|f\|_2 < \infty$. Thus the set of functions bounded in the norm of $\langle\cdot,\cdot\rangle_2$ is the same as the set of functions bounded in the norm $\langle\cdot,\cdot\rangle_r$. With respect to the inner product $\langle\cdot,\cdot\rangle_r$, $L_{p,q,r}$ is self-adjoint.

7.2.10 Proposition *For all* $\mathrm{f}, \mathrm{g} \in \operatorname{dom}(\mathrm{L}_{\mathrm{p,q,r}})$, $\langle \mathrm{L}_{\mathrm{p,q,r}}(\mathrm{f}), \mathrm{g}\rangle_{\mathrm{r}} = \langle \mathrm{f}, \mathrm{L}_{\mathrm{p,q,r}}(\mathrm{g})\rangle_{\mathrm{r}}$.

Proof This is a simple computation:

$$\begin{aligned} \langle L_{p,q,r}(f), g\rangle_r &= \int_a^b L_{p,q,r}(f)(x)g(x)r(x)\,\mathrm{d}x \\ &= \int_a^b r^{-1}(x)L_{p,q}(f)(x)g(x)r(x)\,\mathrm{d}x \\ &= \int_a^b L_{p,q}(f)(x)g(x)\,\mathrm{d}x \\ &= \int_a^b f(x)L_{p,q}(g)(x)\,\mathrm{d}x \\ &= \int_a^b r^{-1}(x)f(x)L_{p,q}(g)(x)r(x)\,\mathrm{d}x \\ &= \int_a^b f(x)L_{p,q,r}(g)(x)r(x)\,\mathrm{d}x \\ &= \langle f, L_{p,q,r}(g)\rangle_r, \end{aligned}$$

as desired. ■

Thus, even though $L_{p,q,r}$ is not self-adjoint with respect to the usual inner product $\langle\cdot,\cdot\rangle_2$, it is nonetheless self-adjoint, and so we can apply ***missing stuff***, and deduce that all eigenvalues are real, and that eigenvectors for distinct eigenvalues are orthogonal with respect to $\langle\cdot,\cdot\rangle_r$. Let us denote the eigenvalue problem by P so that P is the problem defined by

$$\boxed{\begin{aligned}(py')' - qy = \lambda ry \qquad & \alpha_1 y'(a) + \alpha_0 y(a) = 0\\ & \beta_1 y'(b) + \beta_0 y(b) = 0.\end{aligned}} \tag{7.15}$$

We denote by $\operatorname{spec}_0(P)$ the set of all $\lambda \in \mathbb{R}$ so that (7.15) has a nontrivial solution y. Such a solution y is an eigenvector, but let us adopt the traditional terminology by calling it an ***eigenfunction***. However, we are still not guaranteed that $L_{p,q,r}$ even has eigenvalues, since the matter of existence of eigenvalues for linear maps on Hilbert spaces is nontrivial (cf. Exercises 7.1.14 and 7.1.15). We shall deal with the matter of existence of eigenvalues in Section 7.2.3.2. However, let us here state some properties of the collection of eigenvalues, should they exist. We recall that if $\{x_j\}_{j\in\mathbb{N}}$ is a sequence in $\mathbb{R}$, a ***cluster point*** for the sequence is a point x for which there is a subsequence $\{x_{j_k}\}_{k\in\mathbb{N}}$ converging to x.

7.2.11 Theorem $\operatorname{spec}_0(P)$ *is either a finite or a countable set. If it is a countable set, then it has no finite cluster point.*

Proof For fixed $l \in \mathbb{R}$, not necessarily an eigenvalue for $L_{p,q,r}$, we consider the differential equation

$$L_{p,q}(y) - lry = 0. \tag{7.16}$$

This, being a second-order linear equation, possesses two linearly independent solutions, $y_1(l,x)$ and $y_2(l,x)$, satisfying the initial conditions

$$y_1(l,a) = 1,\ y_1'(l,a) = 0, \quad y_2(l,a) = 0,\ y_2'(l,a) = 1.$$

Note that y_1 and y_2 are functions of l, so we explicitly denote this dependence. What's more, in a sufficiently interesting course in differential equations, you will learn that since the differential equation depends on l in an analytic manner, so too will the solutions of the differential equation depend on l in an analytic manner.[5]

Since any solution of the differential equation (7.16) is a linear combination of y_1 and y_2, it follows that $l \in \mathbb{R}$ is an eigenvalue for $L_{p,q,r}$ with an eigenfunction in $\operatorname{dom}(L_{p,q,r})$ if and only if there exists $c_1, c_2 \in \mathbb{R}$, not both zero, so that $c_1y_1(l,x) + c_2y_2(l,x)$ is a solution of (7.16), and so that the boundary conditions for $\operatorname{dom}(L_{p,q,r})$ are satisfied. The boundary conditions when applied to such a function take the form

$$\begin{bmatrix}\alpha_1 y_1'(l,a) + \alpha_0 y_1(l,a) & \alpha_1 y_2'(l,a) + \alpha_0 y_2(l,a)\\ \beta_1 y_1'(l,a) + \beta_0 y_1(l,a) & \beta_1 y_2'(l,a) + \beta_0 y_2(l,a)\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

[5]We will assume this fact without proof, hoping that it is at least believable. Readers interested in a proof are referred to the references. Let us proceed with the proof, with this fact assumed.

For this to have a nontrivial solution for $c_1, c_2 \in \mathbb{R}$ then the coefficient matrix should have zero determinant. Let us denote the determinant by Δ. Since $\alpha_1^2+\alpha_0^2, \beta_1^2+\beta_0^2 \neq 0$, Δ is not identically zero.***missing stuff*** Thus Δ will be a nonzero analytic function for l, and the zeroes of Δ correspond exactly to eigenvalues for P. Now recall that a nonzero analytic function, if it vanishes at $l \in \mathbb{R}$, must be nonzero on some open subset of $\mathbb{R}$ containing l. From this we conclude that $\operatorname{spec}_0(P)$ is either finite or countable. Now we show that if $\operatorname{spec}_0(P)$ is countable then it has no finite cluster point. Suppose that there is a finite cluster point l_0. Then this implies that there is a sequence $\{l_j\}_{j\in\mathbb{N}} \subseteq \operatorname{spec}_0(P)$ converging to l_0. Since Δ is analytic, this implies that Δ is identically zero.[6] This is a contradiction, showing that $\operatorname{spec}_0(P)$ has no finite cluster point, as claimed. ■

Let us make sure that we appreciate the impact of this result. It does *not* tell us that $\operatorname{spec}_0(P) \neq \emptyset$. But it does tell us something about the nature of $\operatorname{spec}_0(P)$. One of the advantages to the form of $\operatorname{spec}_0(P)$ is that we can assume without loss of generality that $0 \notin \operatorname{spec}_0(P)$. Indeed, suppose that $0 \in \operatorname{spec}_0(P)$, and consider the modified problem $P(c)$ defined by

$$(py')' - (q - cr)y = \lambda ry \qquad \begin{aligned} \alpha_1 y'(a) + \alpha_0 y(a) &= 0 \\ \beta_1 y'(b) + \beta_0 y(b) &= 0 \end{aligned} \tag{7.17}$$

for some $c \in \mathbb{R}$. The following trivial result relates the eigenvalues for P and $P(c)$.

7.2.12 Lemma *Let* $(\lambda, \mathrm{f}) \in \mathbb{R} \times \operatorname{dom}(\mathrm{L}_{\mathrm{p,q,r}})$. *Then* λ *is an eigenvalue for* P *with eigenfunction* f *if and only if* $\lambda - \mathrm{c}$ *is an eigenvalue for* P(c) *with eigenfunction* f.

The significance of the lemma is that there is a simple correspondence between the eigenvalues and eigenfunctions for P and $P(c)$. In particular, if $0 \in \operatorname{spec}_0(P)$, then by Theorem 7.2.11 we may choose $c \in \mathbb{R}$ so that $0 \notin \operatorname{spec}_0(P(c))$. Once one has the eigenvalues and eigenfunctions for $P(c)$, those for P are readily recovered. What's more, and this is the punchline as far as we are concerned, the number of eigenvalues are the same for $\operatorname{spec}_0(P)$ and $\operatorname{spec}_0(P(c))$, and the eigenfunctions for P are dense in $\mathrm{L}_2([a,b];\mathbb{R})$ if and only if those for $P(c)$ are dense in $\mathrm{L}_2([a,b];\mathbb{R})$. This is all relevant as in the next assumption we shall assume that $0 \notin \operatorname{spec}_0(P)$, and we should realise that this is not at all a restrictive assumption.

7.2.3 The Green function and completeness of eigenfunctions

Throughout this subsection, we assume that $0 \notin \operatorname{spec}_0(P)$. *As we saw in the closing remarks of the preceding subsection, this is not a substantive restriction.*

To actually *prove* things about the eigenvalue problem P in equation (7.15) is, as one may imagine, nontrivial. One way to get to such proofs is via the Green function, which is, of course, a generalisation of the Green function introduced in

[6] Here we use another fact that we do not prove, namely that the value of an analytic function is determined by its value on a convergent sequence of points.

the example of Section 7.2.1.1. The Green function is an extraordinarily useful feature of the boundary value problems we consider, and it appears in contexts more general than we present it here. Our principal interest in the Green function is due to its value in proving a completeness result for the eigenvalue problem P that mirrors ***missing stuff***. However, the Green function is of further utility in dealing with singular problems, as we shall see in Section 7.3.

7.2.3.1 The Green function We wish to pursue the possibility of generalising the discussion in Section 7.2.1.1 surrounding the nature of the inverse of $L_{p,q,r}$. Note that our tacit assumption in this section that $0 \notin \operatorname{spec}_0(P)$ at least allows the possibility of an inverse. That is to say, if $0 \in \operatorname{spec}_0(P)$, then $L_{p,q,r}$ is guaranteed to not be invertible (why?). The following nontrivial result gives the existence of the generalisation of the function G of Section 7.2.1.1.

7.2.13 Theorem *Consider the linear map* $\mathrm{L_{p,q,r}}\colon \operatorname{dom}(\mathrm{L_{p,q,r}}) \to \mathrm{L_2}([a,b];\mathbb{R})$ *defined by (7.14). There exists a unique function* $\mathrm{G_{p,q,r}}\colon [a,b]\times[a,b] \to \mathbb{R}$ *so that the function* $\mathrm{g}_\xi\colon [a,b] \to \mathbb{R}$ *defined by* $\mathrm{g}_\xi(x) = \mathrm{G_{p,q,r}}(x,\xi)$ *has the following properties:*

(i) $\mathrm{G_{p,q,r}}$ *is continuous;*

(ii) the partial derivatives $\frac{\partial^j}{\partial x^j}\mathrm{G_{p,q,r}}$, $j \in \{1,2\}$, *are continuous on the set*

$$\{(x,\xi) \in [a,b]\times[a,b] \mid x \neq \xi\};$$

(iii) $\mathrm{g}'_\xi(\xi+) - \mathrm{g}'_\xi(\xi-) = \frac{1}{\mathrm{p}(\xi)}$;

(iv) $\mathrm{L_{p,q,r}}(\mathrm{g}_\xi) = 0$ *for* $x \neq \xi$;

(v) the boundary conditions

$$\alpha_1 \mathrm{g}'_\xi(a) + \alpha_0 \mathrm{g}_\xi(a) = 0, \quad \beta_1 \mathrm{g}'_\xi(b) + \beta_0 \mathrm{g}_\xi(b) = 0$$

are satisfied for $\xi \in [a,b]$.

Furthermore, the function $\mathrm{G_{p,q,r}}$ *has the property that the mapping* $\mathcal{G}_{\mathrm{p,q,r}}\colon \mathrm{L_2}([a,b];\mathbb{R}) \to \operatorname{dom}(\mathrm{L_{p,q,r}})$ *defined by*

$$\mathcal{G}_{\mathrm{p,q,r}}(u)(x) = \int_a^b \mathrm{G_{p,q,r}}(x,\xi)u(\xi)\,d\xi$$

is exactly $-\mathrm{L}^{-1}_{\mathrm{p,q,r}}$.

Proof The proof of the theorem revolves around a description of the solutions of the differential equation $L_{p,q,r}(y) = -f$, where $f \in \mathrm{L_2}([a,b];\mathbb{R})$ is some fixed but arbitrary function. The following result records this solution. The result will be known to students having had a good introductory course in linear differential equations, except perhaps for the fact that f is merely integrable, and not something stronger like continuous. The proof is not difficult, but requires some buildup that constitutes a significant diversion.

1 Lemma *Consider the initial value problem*

$$y''(x) + a_1(x)y'(x) + a_0(x)y(x) = b(x), \quad y(a) = y_0,\ y'(a) = v_0. \tag{7.18}$$

Let y_1 *and* y_2 *be solutions to the homogeneous problem (i.e., that with* $b = 0$*) satisfying the initial conditions*

$$y_1(a) = 1,\ y_1'(a) = 0, \quad y_2(a) = 0,\ y_2'(a) = 1,$$

and define the corresponding ***Wronskian***[7] $W(y_1, y_2)\colon [a, b] \to \mathbb{R}$ *by*

$$W(y_1, y_2)(x) = \det \begin{bmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{bmatrix}.$$

Then the solution to (7.18) is

$$y(x) = y_0 y_1(x) + v_0 y_2(x) + \int_a^x \frac{1}{W(y_1, y_2)(\xi)} \det \begin{bmatrix} y_1(\xi) & y_2(\xi) \\ y_1(x) & y_2(x) \end{bmatrix} b(\xi)\, d\xi.$$

Proof Note that the proposed solution has the form

$$y(x) = y_h(x) + u_1(x)y_1(x) + u_2(x)y_2(x)$$

where y_h solves the homogeneous equation, and where

$$u_1(x) = -\int_a^x \frac{y_2(\xi)b(\xi)}{W(y_1, y_2)(\xi)}\, d\xi, \quad u_2(x) = \int_a^x \frac{y_1(\xi)b(\xi)}{W(y_1, y_2)(\xi)}\, d\xi. \tag{7.19}$$

Let us therefore determine the expressions for a general u_1 and u_2 so that the function $y(x) = u_1(x)y_1(x) + u_2(x)y_2(x)$ satisfies the equation

$$y''(x) + a_1(x)y'(x) + a_0(x)y(x) = b(x).$$

We compute

$$y' = (u_1'y_1 + u_2'y_2) + (u_1y_1' + u_1y_2').$$

Let us impose the condition that $u_1'y_1 + u_2'y_2 = 0$. Thus we seek u_1 and u_2 that satisfy this condition, and which also satisfy the differential equation. With this condition imposed we compute

$$y'' = u_1'y_1' + u_2'y_2' + u_1u_1'' + u_2y_2''.$$

Substituting y into the differential equation yields

$$u_1(y_1'' + a_1y_1' + a_0y_1) + u_2(y_2'' + a_1y_2' + a_0y_2) + u_1'y_1' + u_2'y_2' = b.$$

[7] After Josef Hoëné de Wronski (1778–1853). Wronski was a "philosopher mathematician," and as a consequence he (1) published a lot of rubbish and (2) had a high opinion of himself. Nevertheless, he apparently had a few good days, and the Wronskian, one supposes, must be a result of one of these.

By virtue of y_1 and y_2 satisfying the homogeneous equation, the first two terms vanish. Thus we have arrived at the two linear equations

$$\begin{aligned} u_1' y_1 + u_2' y_2 &= 0 \\ u_1' y_1' + u_2' y_2' &= b \end{aligned}$$

for u_1' and u_2'. If u_1 and u_2 satisfy these equations, then our proposed solution solves the equation as desired. However, solving the two linear equations shows that if we choose u_1 and u_2 so that

$$u_1'(x) = \frac{y_2(x)b(x)}{W(y_1, y_2)(x)}, \quad u_2'(x) = \frac{y_1(x)b(x)}{W(y_1, y_2)(x)},$$

then we will have suitable functions u_1 and u_2. This computation establishes our claim that with u_1 and u_2 as defined by (7.19), the function $y = u_1 y_1 + u_2 y_2$ solves the differential equation. Now we merely note that the solution proposed by the lemma is the sum of the solution we have just obtain and a solution of the homogeneous problem. This sum thus solves the equation. Also, the initial conditions may be checked immediately. ▼

Motivated by the lemma, let y_1 and y_2 be solutions to $L_{p,q,r}(y) = 0$ satisfying the initial conditions

$$y_1(a) = 1,\ y_1'(a) = 0, \quad y_2(a) = 0,\ y_2'(a) = 1.$$

Now define $K_{p,q,r} \colon [a, b] \times [a, b] \to \mathbb{R}$ by

$$K_{p,q,r}(x, \xi) = \frac{1}{p(\xi)W(y_1, y_2)(\xi)} \det \begin{bmatrix} y_1(\xi) & y_2(\xi) \\ y_1(x) & y_2(x) \end{bmatrix}$$

for $\xi \le x$ and let $K_{p,q,r}(x, \xi) = 0$ for $\xi > x$. Since $K_{p,q,r}(\xi+, \xi) = 0$ it follows that $K_{p,q,r}$ is continuous on $[a, b] \times [a, b]$. We also have

$$\frac{\partial K_{p,q,r}}{\partial x} = \frac{1}{p(\xi)W(y_1, y_2)(\xi)} \det \begin{bmatrix} y_1(\xi) & y_2(\xi) \\ y_1'(x) & y_2'(x) \end{bmatrix}$$

and

$$\frac{\partial^2 K_{p,q,r}}{\partial x^2} = \frac{1}{p(\xi)W(y_1, y_2)(\xi)} \det \begin{bmatrix} y_1(\xi) & y_2(\xi) \\ y_1''(x) & y_2''(x) \end{bmatrix},$$

as may be verified by a direct computation. This shows that $K_{p,q,r}$ is twice continuously differentiable on the stated domain. Also note that

$$\frac{\partial K_{p,q,r}}{\partial x}(\xi+, \xi) = \frac{1}{p(\xi)W(y_1, y_2)(\xi)} \det \begin{bmatrix} y_1(\xi) & y_2(\xi) \\ y_1'(\xi+) & y_2'(\xi+) \end{bmatrix} = \frac{1}{p(\xi)},$$

as per (iii). What's more, by Lemma 1, the function $u\colon [a,b] \to \mathbb{R}$ defined by

$$u(x) = -\int_a^b K_{p,q,r}(x,\xi) f(\xi)\,\mathrm{d}\xi$$

satisfies $L_{p,q,r}(u) = -f$.

We shall now modify $K_{p,q,r}$ so that it satisfies the boundary conditions. We seek functions $c_1, c_2\colon [a,b] \to \mathbb{R}$ so that the function

$$G_{p,q,r}(x,\xi) = -K_{p,q,r}(x,\xi) + c_1(\xi)y_1(x) + c_2(\xi)y_2(x)$$

satisfies (v). Since y_1 and y_2 are solutions to the homogeneous problem, it also follows that $G_{p,q,r}$ satisfies (iv). The boundary conditions applied to $G_{p,q,r}$ take the form

$$\begin{bmatrix} \alpha_1 \frac{\partial G_{p,q,r}}{\partial x}(a,\xi) + \alpha_0 G_{p,q,r}(a,\xi) \\ \beta_1 \frac{\partial G_{p,q,r}}{\partial x}(b,\xi) + \beta_0 G_{p,q,r}(b,\xi) \end{bmatrix} = \\ \begin{bmatrix} \alpha_1 \frac{\partial K_{p,q,r}}{\partial x}(a,\xi) + \alpha_0 K_{p,q,r}(a,\xi) \\ \beta_1 \frac{\partial K_{p,q,r}}{\partial x}(b,\xi) + \beta_0 K_{p,q,r}(b,\xi) \end{bmatrix} + \begin{bmatrix} \alpha_1 y_1'(a) + \alpha_0 y_1(a) & \alpha_1 y_2'(a) + \alpha_0 y_2(a) \\ \beta_1 y_1'(a) + \beta_0 y_1(a) & \beta_1 y_2'(a) + \beta_0 y_2(a) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

We desire to choose c_1 and c_2 so that the right-hand side is identically zero. This is possible since, referring to the proof of Theorem 7.2.11, the determinant of the matrix

$$\begin{bmatrix} \alpha_1 y_1'(a) + \alpha_0 y_1(a) & \alpha_1 y_2'(a) + \alpha_0 y_2(a) \\ \beta_1 y_1'(a) + \beta_0 y_1(a) & \beta_1 y_2'(a) + \beta_0 y_2(a) \end{bmatrix}$$

is nonzero by virtue of 0 not being an eigenvalue for P. Thus, taking

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = -\begin{bmatrix} \alpha_1 y_1'(a) + \alpha_0 y_1(a) & \alpha_1 y_2'(a) + \alpha_0 y_2(a) \\ \beta_1 y_1'(a) + \beta_0 y_1(a) & \beta_1 y_2'(a) + \beta_0 y_2(a) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \frac{\partial K_{p,q,r}}{\partial x}(a,\xi) + \alpha_0 K_{p,q,r}(a,\xi) \\ \beta_1 \frac{\partial K_{p,q,r}}{\partial x}(b,\xi) + \beta_0 K_{p,q,r}(b,\xi) \end{bmatrix}$$

gives $G_{p,q,r}$ the property (iv). One can readily verify that since y_1 and y_2 are twice continuously differentiable on $[a,b]$, the continuity properties (i) and (ii) for $G_{p,q,r}$ are inherited from $K_{p,q,r}$. Also, since y_1 and y_2 are continuously differentiable on $[a,b]$ and since c_1 and c_2 are continuous, the property (iii) for $G_{p,q,r}$ is also inherited from $K_{p,q,r}$. Also, it is clear that $\mathcal{G}_{p,q,r} = -L_{p,q,r}^{-1}$ by construction of $G_{p,q,r}$.

It only remains to show that $G_{p,q,r}$ is the unique function with the properties (i)–(v). Suppose that there is another such function $\tilde{G}_{p,q,r}$. Let $\tilde{g}_\xi$ be defined by $\tilde{g}_\xi = \tilde{G}_{p,q,r}(x,\xi)$. Since the derivatives of g_ξ and $\tilde{g}_\xi$ share the discontinuity at $x = \xi$ specified by (iii), it follows that their difference, $g_\xi - \tilde{g}_\xi$, will be continuously differentiable. It follows that $g_\xi - \tilde{g}_\xi$ satisfies the boundary conditions, Also, since $L_{p,q,r}(g_\xi) = 0$ and $L_{p,q,r}(\tilde{g}_\xi) = 0$, it follows that $L_{p,q,r}(g_\xi - \tilde{g}_\xi) = 0$. From this we infer that since $g_\xi' - \tilde{g}_\xi'$ is continuous on $[a,b]$ for all ξ, so too is $g_\xi'' - \tilde{g}_\xi''$. However, since 0 is not an eigenvalue for P this implies that $g_\xi - \tilde{g}_\xi = 0$. ■

This result is largely a technical one, although it is of some importance. It provides us with a characterisation of the ***Green function*** $G_{p,q,r}$ is terms of $L_{p,q,r}$. Also, note that the proof is sort of constructive, and is actually constructive provided that you can obtain all solutions of the homogeneous equation $L_{p,q,r}(y) = 0$. In Exercise 7.2.10 you can use the procedure in the proof to construct the Green function for the example of Section 7.2.1.1.

Let us now note an important property of the linear mapping $\mathcal{G}_{p,q,r}$, or more precisely its image in $\operatorname{dom}(L_{p,q,r})$. To state our result, we need some terminology. Let A be an arbitrary index set, not necessarily finite or countable, and let $\mathcal{F} = \{f_a\}_{a\in A}$ be a collection of $\mathbb{R}$-valued functions on $[a, b]$. The set of functions $\mathcal{F}$ is ***equicontinuous*** if for every $\epsilon > 0$ there exists a $\delta > 0$ so that $|f_a(x_1) - f_a(x_2)| < \epsilon$ for all $a \in A$, provided that $|x_1 - x_2| < \delta$. Similarly, the set of functions is ***uniformly bounded*** if there exists $M > 0$ so that $|f_a(x)| < M$ for all $x \in [a, b]$ and for all $a \in A$. The key property here is that δ and M may be chosen *independent of* $a \in A$.

7.2.14 Proposition *The set of functions*

$$\{\mathcal{G}_{\mathrm{p,q,r}}(\mathrm{u})\}_{\substack{\mathrm{u}\in \mathrm{L}_2([\mathrm{a,b}];\mathbb{R})\\ \|\mathrm{u}\|_{\mathrm{r}}\leq 1}}$$

is equicontinuous and uniformly bounded.

Proof Since $G_{p,q,r}$ is continuous on the closed and bounded domain $[a, b] \times [a, b]$ it is uniformly continuous***missing stuff*** and so, for any $\epsilon > 0$ there exists $\delta > 0$ so that, provided $|x_1 - x_2| < \delta$,

$$|G_{p,q,r}(x_1, \xi) - G_{p,q,r}(x_2, \xi)| < \epsilon,$$

for any $\xi \in [a, b]$. Since r is continuous and positive, and since $[a, b]$ is closed and bounded there exists $\underline{r} > 0$ so that $r(x) \geq \underline{r}$ for every $x \in [a, b]$. A simple computation then gives $\|f\|_2 \leq \frac{1}{\sqrt{\underline{r}}}\|f\|_2$ for any $f \in \mathrm{L}_2([a, b]; \mathbb{R})$. Thus for any $u \in \mathrm{L}_2([a, b]; \mathbb{R})$ and for any $\epsilon > 0$ there exists $\delta > 0$ so that if $|x_1 - x_2| < \delta$ we have

$$\begin{aligned}
|\mathcal{G}_{p,q,r}(u)(x_1) - \mathcal{G}_{p,q,r}(u)(x_2)| &= \Big| \int_a^b G_{p,q,r}(x_1, \xi)u(\xi)\,\mathrm{d}\xi - \int_a^b G_{p,q,r}(x_2, \xi)u(\xi)\,\mathrm{d}\xi \Big| \\
&= \Big| \int_a^b \Big(G_{p,q,r}(x_1, \xi) - G_{p,q,r}(x_2, \xi)\Big)u(\xi)\,\mathrm{d}\xi \Big| \\
&\leq \int_a^b \Big|G_{p,q,r}(x_1, \xi) - G_{p,q,r}(x_2, \xi)\Big||u(\xi)|\,\mathrm{d}\xi \\
&\leq \epsilon\sqrt{b-a}\|u\|_2 \\
&\leq \frac{\epsilon\sqrt{b-a}}{\sqrt{\underline{r}}}\|u\|_r.
\end{aligned}$$

where we have used the Cauchy-Bunyakovsky-Schwartz inequality in the penultimate step. This shows that the set of functions given in the statement of the result is

indeed equicontinuous. Since $G_{p,q,r}$ is bounded on $[a,b]\times[a,b]$ (being a continuous function on a closed bounded set) we can choose an $M>0$ so that $|G_{p,q,r}(x,\xi)|\leq M$ for all $(x,\xi)\in[a,b]\times[a,b]$. A computation much like that in the first part of the proof then gives

$$|\mathcal{G}_{p,q,r}(u)(x)|\leq M\sqrt{b-a}\|u\|_2.$$

This shows that the set of functions given in the statement of the result is bounded in the norm $\|\cdot\|_\infty$. That this set is also bounded in the norm $\|\cdot\|_2$, and hence the norm $\|\cdot\|_r$, now follows from***missing stuff***. ■

This result says that $\mathcal{G}_{p,q,r}$ is a bounded, and therefore continuous by***missing stuff***, linear map with respect to the norm $\|\cdot\|_r$. Thus we may define the operator norm of $\mathcal{G}_{p,q,r}$ as in***missing stuff***:

$$\|\mathcal{G}_{p,q,r}\|_{r\to r}=\sup_{\|u\|_r=1}\|\mathcal{G}_{p,q,r}(u)\|_r.$$

This, you should understand, is a nontrivial assertion. Indeed, the map $L_{p,q,r}$ is itself not continuous, so continuity of $\mathcal{G}_{p,q,r}$, which is essentially the inverse of $L_{p,q,r}$, does not follow in any easy way.

7.2.3.2 Completeness of eigenfunctions

We continue with the assumption that $0\notin\operatorname{spec}_0(L_{p,q,r})$ *so that we may define the Green function* $G_{p,q,r}$. *We again mention that this can be done without loss of generality by a simple modification of the problem.*

The Green function introduced in the preceding section now becomes a valuable tool for us in proving the existence of eigenvalues for $L_{p,q,r}$. The key fact is the continuity of $\mathcal{G}_{p,q,r}$ proved in Proposition 7.2.14. We first note that it is easy to show that $\mathcal{G}_{p,q,r}$ is self-adjoint, referring the reader to Exercise 7.2.12 to work this out. First we need to relate the eigenvalues and eigenfunctions for $L_{p,q,r}$ to those for $G_{p,q,r}$. The following result is Exercise 7.1.16, keeping in mind that $\mathcal{G}_{p,q,r}$ is not $L^{-1}_{p,q,r}$, but $-L^{-1}_{p,q,r}$.

7.2.15 Lemma *Let* $(\lambda,\mathrm{f})\in\mathbb{R}\times\operatorname{dom}(\mathrm{L_{p,q,r}})$. *Then* λ *is an eigenvalue for* P *with eigenfunction* f *if and only if* $-\lambda^{-1}$ *is an eigenvalue for* $\mathcal{G}_{\mathrm{p,q,r}}$ *with eigenfunction* f.

Thus the act of finding eigenvalues and eigenfunctions for the eigenvalue problem P is related in a simple way to funding eigenvalues and eigenfunctions for $\mathcal{G}_{p,q,r}$. The following theorem starts us off by providing a description of a single eigenvalue and eigenvector for $G_{p,q,r}$.

7.2.16 Theorem *One of the two numbers* $\|\mathcal{G}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}}$ *or* $-\|\mathcal{G}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}}$ *is an eigenvalue for* $\mathcal{G}_{\mathrm{p,q,r}}$. *Furthermore, define*

$$\mathbb{S}_{\mathrm{r}}=\{\mathrm{u}\in\mathrm{L}_2([\mathrm{a},\mathrm{b}];\mathbb{R})\mid\ \|\mathrm{u}\|_{\mathrm{r}}=1\}.$$

Then the collection of norm 1 *eigenfunctions for the above eigenvalue are the functions* $\mathrm{u}\in\mathbb{S}_{\mathrm{r}}$ *which maximise (if* $\|\mathcal{G}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}}$ *is an eigenvalue) or minimise (if* $-\|\mathcal{G}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}}$ *is an*

eigenvalue) the function $\mathrm{Q_{p,q,r}}\colon \mathbb{S}_\mathrm{r} \to \mathbb{R}$ *defined by*

$$\mathrm{Q_{p,q,r}(u)} = \int_\mathrm{a}^\mathrm{b} \int_\mathrm{a}^\mathrm{b} \mathrm{G_{p,q,r}(x,\xi)u(\xi)r(x)}\,\mathrm{d}\xi\,\mathrm{dx}.$$

Proof We refer to Exercise 7.1.11 to retrieve the fact that, since $\mathcal{G}_{p,q,r}$ is continuous and self-adjoint (for the latter, refer to Exercise 7.2.12),

$$\|\mathcal{G}_{p,q,r}\|_{r\to r} = \sup_{\|u\|=1} |\langle \mathcal{G}_{p,q,r}(u), u\rangle_r|. \tag{7.20}$$

This permits two possibilities:

1. $\|\mathcal{G}_{p,q,r}\|_{r\to r} = \sup_{\|u\|=1}\langle \mathcal{G}_{p,q,r}(u), u\rangle_r$;
2. $\|\mathcal{G}_{p,q,r}\|_{r\to r} = -\inf_{\|u\|=1}\langle \mathcal{G}_{p,q,r}(u), u\rangle_r$.

Let us first suppose the first of these cases. Then, by the definition of supremum there exists a sequence of functions $\{u_j\}_{j\in\mathbb{N}}$ with $\|u_j\|_r = 1$, $j \in \mathbb{N}$, so that

$$\lim_{j\to\infty} \langle \mathcal{G}_{p,q,r}(u_j), u_j\rangle_r = \|\mathcal{G}_{p,q,r}\|_{r\to r}.$$

By Proposition 7.2.14 the sequence of functions $\{\mathcal{G}_{p,q,r}(u_j)\}_{j\in\mathbb{N}}$ is equicontinuous and uniformly bounded. For such a collection of functions, the following quite nontrivial and non-obvious result applies.

1 Lemma (Arzela-Ascoli theorem) *If* $\mathcal{F} = \{\mathrm{f_a}\}_{\mathrm{a}\in\mathrm{A}}$ *is an equicontinuous, uniformly bounded collection of* $\mathbb{R}$*-valued functions on* $[\mathrm{a},\mathrm{b}]$ *then there is a mapping* $\phi\colon \mathbb{N} \to \mathrm{A}$ *so that the sequence* $\{\mathrm{f_{\phi(j)}}\}_{\mathrm{j}\in\mathbb{N}}$ *converges uniformly on* $[\mathrm{a},\mathrm{b}]$.

Proof Let $\{q_k\}_{k\in\mathbb{N}}$ be the collection of rational numbers in $[a,b]$, enumerated in some arbitrary manner. The collection of numbers $\{f_a(q_1)\}_{a\in A}$ forms a bounded subset of $\mathbb{R}$. Thus there exists a sequence of distinct functions $\{f_{k_1}\}_{k_1\in\mathbb{N}}$ in $\mathcal{F}$ with the property that $\{f_{k_1}(q_1)\}_{k_1\in\mathbb{N}}$ converges. Now consider the sequence $\{f_{k_1}(q_2)\}_{k_1\in\mathbb{N}}$. Again, this is a bounded subset of $\mathbb{R}$ so there exists a subsequence of functions $\{f_{k_2}\}_{k_2\in\mathbb{N}} \subseteq \{f_{k_1}\}_{k_1\in\mathbb{N}}$ with the property that the sequence $\{f_{k_2}(q_2)\}_{k_2\in\mathbb{N}}$ converges. One may continue in this manner defining nested sequences of functions

$$\{f_{k_1}\}_{k_1\in\mathbb{N}} \supseteq \{f_{k_2}\}_{k_2\in\mathbb{N}} \supseteq \cdots \supseteq \{f_{k_n}\}_{k_n\in\mathbb{N}} \supseteq \cdots$$

Now define a sequence of functions $\{f_n = f_{n_n}\}_{n\in\mathbb{N}}$. We claim that this is a uniformly convergent sequence. Let $\epsilon > 0$ and choose $\delta > 0$ so that $|f_a(x_1) - f_a(x_2)| < \frac{\epsilon}{3}$ for all $a \in A$, provided that $|x_1 - x_2| < \delta$. Now let $\{Q_1, \ldots, Q_K\}$ be a collection of rational numbers having the property that for any $x \in [a,b]$ there exists $k \in \{1,\ldots,K\}$ so that $|x - Q_k| < \delta$. Since the sequences

$$\{f_n(Q_1)\}_{n\in\mathbb{N}}, \ldots, \{f_n(Q_K)\}_{n\in\mathbb{N}}$$

converge, there exists $N \in \mathbb{N}$ so that $|f_m(Q_k) - f_n(Q_k)| < \frac{\epsilon}{3}$ for $k \in \{1, \dots, K\}$. Now, for $x \in [a, b]$ let $k \in \{1, \dots, K\}$ have the property that $|x - Q_k| < \delta$. Then we have

$$\begin{aligned} |f_m(x) - f_n(x)| &= |(f_m(x) - f_m(Q_k)) + (f_m(Q_k) - f_n(Q_k)) + (f_n(Q_k) - f_n(x))| \\ &\le |f_m(x) - f_m(Q_k)| + |f_m(Q_k) - f_n(Q_k)| + |f_n(Q_k) - f_n(x)| \\ &\le \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} = \epsilon, \end{aligned}$$

provided that $n, m > N$. As the choice of N does not depend on x, only on ϵ, uniform convergence follows. ▼

By virtue of the lemma we select a subsequence $\{\mathcal{G}_{p,q,r}(u_{j_k})\}_{k\in\mathbb{N}}$ of functions that converges uniformly on $[a, b]$. As we saw in ***missing stuff***, this implies L_2 convergence of this same sequence. It follows that the limit function which we denote by f_1, is in $L_2([a, b]; \mathbb{R})$. Let $\mu_1 = \|\mathcal{G}_{p,q,r}\|_{r\to r}$ and compute

$$\|\mathcal{G}_{p,q,r}(u_{j_k}) - \mu_1 u_{j_k}\|_r^2 = \|\mathcal{G}_{p,q,r}(u_{j_k})\|_r^2 + \mu_1^2\|u_{j_k}\|_r^2 - 2\mu_1\langle \mathcal{G}_{p,q,r}(u_{j_k}), u_{j_k}\rangle_r.$$

By construction of the sequence $\{u_{j_k}\}_{k\in\mathbb{N}}$ we have

$$\lim_{k\to\infty}\Big(\|\mathcal{G}_{p,q,r}(u_{j_k})\|_r^2 + \mu_1^2\|u_{j_k}\|_r^2 - 2\mu_1\langle \mathcal{G}_{p,q,r}(u_{j_k}), u_{j_k}\rangle_r\Big) = \|f_1\|_r^2 - \mu_1^2. \tag{7.21}$$

This allows us to conclude that the limit function f_1 does not identically vanish. Noting that $\|\mathcal{G}_{p,q,r}(u_{j_k})\|_r^2 \le \mu_1^2$, we can also conclude from (7.21) that

$$0 \le \|\mathcal{G}_{p,q,r}(u_{j_k}) - \mu_1 u_{j_k}\|_r^2 \le 2\mu_1^2 - 2\mu_1\langle \mathcal{G}_{p,q,r}(u_{j_k}), u_{j_k}\rangle_r.$$

As the limit as $k \to \infty$ of the term on the right tends to zero, so too does the term in the middle, thus giving

$$\lim_{k\to\infty}\|\mathcal{G}_{p,q,r}(u_{j_k}) - \mu_1 u_{j_k}\|_r^2 = 0. \tag{7.22}$$

An application of the triangle inequality and the relation $\|\mathcal{G}_{p,q,r}(u)\|_r \le \|G_{p,q,r}(u)\|_{r\to r}\|u\|_r$ then gives

$$\begin{aligned} 0 &\le \|\mathcal{G}_{p,q,r}(f_1) - \mu_1 f_1\|_r \\ &\le \|\mathcal{G}_{p,q,r}(f_1) - \mathcal{G}_{p,q,r}(\mathcal{G}_{p,q,r}(f_1))\|_r + \|\mathcal{G}_{p,q,r}(\mathcal{G}_{p,q,r}(f_1)) - \mu_1\mathcal{G}_{p,q,r}(u_{j_k})\|_r + \\ &\quad \|\mu_1\mathcal{G}_{p,q,r}(u_{j_k}) - \mu_1 f_1\|_r \\ &\le \|f_1 - \mathcal{G}_{p,q,r}(u_{j_k})\|_r + \|\mathcal{G}_{p,q,r}(u_{j_k}) - \mu_1 u_{j_k}\|_r + |\mu_1|\|\mathcal{G}_{p,q,r}(u_{j_k}) - f_1\|_r. \end{aligned}$$

As $k \to \infty$ this final expression tends to zero by (7.22) along with the definition of the sequence $\{u_{j_k}\}_{k\in\mathbb{N}}$. This gives $\mathcal{G}_{p,q,r}(f_1) = \mu_1 f_1$, showing that $\mu_1 = \|\mathcal{G}_{p,q,r}\|_{r\to r}$ is indeed an eigenvalue for $\mathcal{G}_{p,q,r}$. An entirely similar argument can be worked out for the case when $\|\mathcal{G}_{p,q,r}\|_{r\to r} = -\inf_{\|u\|=1}\langle \mathcal{G}_{p,q,r}(u), u\rangle_r$.

It now remains to exhibit the character of the eigenvalues stated in the theorem. Note that

$$Q_{p,q,r}(u) = \langle \mathcal{G}_{p,q,r}(u), u\rangle_r.$$

Let us consider the case where $\|\mathcal{G}_{p,q,r}\|_{r\to r}$ is an eigenvalue, the other case following like lines. Given the function f_1 defined above, denote $\phi_1 = \frac{f_1}{\|f_1\|_r}$ the corresponding normalised eigenvector. If then follows that

$$Q_{p,q,r}(\phi_1) = \langle \mathcal{G}_{p,q,r}(\phi_1), \phi_1 \rangle_r = \mu_1 \|\phi_1\|_r^2 = \mu_1 = \|\mathcal{G}_{p,q,r}\|_{r\to r}.$$

Thus, for any $u \in \mathbb{S}_r$ we have

$$Q_{p,q,r}(u) = \langle \mathcal{G}_{p,q,r}(u), u \rangle_r \leq \|\mathcal{G}_{p,q,r}\|_{r\to r},$$

giving the result, since this argument holds for any unit length eigenvector for the eigenvalue μ_1. ■

The above theorem is highly nontrivial, so it is worth separating out its essential character. The first main feature is that the existence of an eigenvalue for $\mathcal{G}_{p,q,r}$ is guaranteed, thus also ensuring at least one eigenvalue for P by Lemma 7.2.15. The importance of this is clear given the concerns raised in Exercise 7.1.14 about the existence of eigenvalues for linear maps on infinite-dimensional vector spaces. However, it is also of some importance to observe the simple character of this eigenvalue, and of the corresponding eigenvectors, by using the inner product $\langle \cdot, \cdot \rangle_r$. The reader may wish to refer to the Exercise 7.1.12 to remember how this is accomplished in finite-dimensions. The basic idea is the same, but due to the complications of function spaces, certain features that are easy in the finite-dimensional case are nontrivial in Theorem 7.2.16. Nevertheless, we may proceed essentially as in Exercise 7.1.12 and modify the Green function, using the existing eigenvalue, to a function which has the same features. The following result indicates how this is done.

7.2.17 Proposition *Let $\mu_1, \ldots, \mu_m$ be a finite collection of not necessarily distinct eigenvalues for $\mathcal{G}_{p,q,r}$ with $\phi_1, \ldots, \phi_m$ the corresponding eigenvectors which may be assumed to be orthonormal. If $G^m_{p,q,r} \colon [a,b] \times [a,b] \to \mathbb{R}$ is defined by*

$$G^m_{p,q,r}(x,\xi) = G_{p,q,r}(x,\xi) - r(\xi) \sum_{j=1}^{m} \mu_j \phi_j(x) \phi_j(\xi)$$

then the map $\mathcal{G}^m_{p,q,r} \colon L_2([a,b];\mathbb{R}) \to \mathrm{dom}(L_{p,q,r})$ defined by

$$\mathcal{G}^m_{p,q,r}(u)(x) = \int_a^b G^m_{p,q,r}(x,\xi) u(\xi)\, d\xi$$

has the following properties:

(i) the set of functions

$$\{\mathcal{G}^m_{p,q,r}(u)\}_{\substack{u \in L_2([a,b];\mathbb{R}) \\ \|u\|_r \leq 1}}$$

is equicontinuous and uniformly bounded;

(ii) $\|\mathcal{G}^{\mathrm{m}}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}} = \sup_{\|\mathrm{u}\|_{\mathrm{r}}} \left|\langle \mathcal{G}^{\mathrm{m}}_{\mathrm{p,q,r}}(\mathrm{u}), \mathrm{u}\rangle_{\mathrm{r}}\right|$;

(iii) $\|\mathcal{G}^{\mathrm{m}}_{\mathrm{p,q,r}}\|_{\mathrm{r}\to\mathrm{r}} \neq 0$.

Proof (i) Referring to the proof of Proposition 7.2.14, the essential feature of $G_{p,q,r}$ that allows one to prove that result is that the set of functions $\{g_\xi\}_{\xi\in[a,b]}$ is uniformly continuous, where $g_\xi(x) = G_{p,q,r}(x,\xi)$. That is to say, for each $\epsilon > 0$ there exists $\delta > 0$ so that, independent of ξ, $|g_\xi(x_1) - g_\xi(x_2)| < \epsilon$ provided that $|x_1 - x_2| < \delta$. This also holds for $G^m_{p,q,r}$ since it is a continuous function on $[a,b]\times[a,b]$. This, plus boundedness of $G^m_{p,q,r}$, allows the argument of Proposition 7.2.14 to be applied here.

(ii) This part of the result will follow from Exercise 7.1.11 if we can show that $\mathcal{G}^m_{p,q,r}$ is self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_r$. To verify this we compute

$$\begin{aligned}
\langle \mathcal{G}^m_{p,q,r}(f), g\rangle_r &= \int_a^b \mathcal{G}^m_{p,q,r}(f)(x)g(x)r(x)\,\mathrm{d}x\\
&= \int_a^b \mathcal{G}_{p,q,r}(f)(x)g(x)r(x)\,\mathrm{d}x -\\
&\quad \sum_{j=1}^m \int_a^b \Big(\int_a^b \mu_j\phi_j(x)\phi_j(\xi)f(\xi)r(\xi)\,\mathrm{d}\xi\Big)g(x)r(x)\,\mathrm{d}x\\
&= \int_a^b f(x)\mathcal{G}(g)(x)r(x)\,\mathrm{d}x -\\
&\quad \sum_{j=1}^m \int_a^b \Big(\mu_j\phi_j(\xi)\phi_j(x)g(x)r(x)\,\mathrm{d}x\Big)f(\xi)r(\xi)\,\mathrm{d}\xi\\
&= \int_a^b f(\xi)\mathcal{G}^m_{p,q,r}(g)(\xi)r(\xi)\\
&= \langle f, \mathcal{G}^m_{p,q,r}(g)\rangle_r,
\end{aligned}$$

where we have used the fact that $\mathcal{G}_{p,q,r}$ is self-adjoint.

(iii) For $f\in \mathrm{L}_2([a,b];\mathbb{R})$ we compute

$$\begin{aligned}
-L_{p,q,r}\circ\mathcal{G}^m_{p,q,r}(f)(x) &= -L_{p,q,r}\circ\mathcal{G}_{p,q,r}(f)(x) + L_{p,q,r}\Big(\sum_{j=1}^m\int_a^b \mu_j r(\xi)\phi_j(x)\phi_j(\xi)f(\xi)\,\mathrm{d}\xi\Big)\\
&= f(x) + \sum_{j=1}^m \mu_j\langle f,\phi_j\rangle_r L_{p,q,r}(\phi_j)(x)\\
&= f(x) - \sum_{j=1}^m \langle f,\phi_j\rangle_r\phi_j(x),
\end{aligned}$$

using the fact that $\mathcal{G}_{p,q,r} = -L^{-1}_{p,q,r}$, and Lemma 7.2.15. Now if $\|\mathcal{G}^m_{p,q,r}\|_{r\to r} = 0$ then we

have

$$f = \sum_{j=1}^{m} \langle f, \phi_j \rangle_r \phi_j, \tag{7.23}$$

which should hold for every $f \in \mathrm{L}_2([a,b];\mathbb{R})$. Since the eigenfunctions $\phi_1, \dots, \phi_m$ are also eigenfunctions for P by Lemma 7.2.15, it follows that these functions are continuously differentiable. Therefore, the right-hand side of (7.23) is continuously differentiable for every $f \in \mathrm{L}_2([a,b];\mathbb{R})$. In particular, since there are certainly functions in $\mathrm{L}_2([a,b];\mathbb{R})$ that are not continuously differentiable, (7.23) cannot hold. Thus $\|\mathcal{G}^m_{p,q,r}\|_{r\to r}$ cannot be zero. ■

The previous result indicates that we can iteratively apply Theorem 7.2.16 to produce an infinite sequence of eigenvalues $\{\mu_j\}_{j\in\mathbb{N}}$ for $\mathcal{G}_{p,q,r}$, and so by Lemma 7.2.15 an infinite sequence of eigenvalues $\{\lambda_j = -\mu_j^{-1}\}_{j\in\mathbb{N}}$ for P. By Theorem 7.2.11, the eigenvalues for P are unbounded in magnitude as $j \to \infty$. Thus we have successfully captured some of the features we saw in the simple boundary value problem of Section 7.2.1.1. Flush with our success, we may, for a function $f \in \mathrm{L}_2([a,b];\mathbb{R})$ write its ***generalised Fourier series*** as

$$\mathrm{FS}[f] = \sum_{n=1}^{\infty} \langle f, \phi_n \rangle \phi_n,$$

at least formally. What remains is to show that the normalised eigenfunctions form a complete orthonormal family. This is the content of the following result.

7.2.18 Theorem *Let* P *be the eigenvalue problem (7.15). A set* $\{\phi_n\}_{n\in\mathbb{N}}$ *of orthonormal eigenvectors for* P *is a complete orthonormal family. What's more, if* $f \in \mathrm{dom}(L_{p,q,r})$ *then* FS[f] *converges uniformly to* f.

Proof We first prove that any function $f \in \mathrm{dom}(L_{p,q,r})$ can be arbitrarily well approximated by its generalised Fourier series. To do this, we first work with $\mathcal{G}_{p,q,r}$. For $x \in [a,b]$ let us define $\tilde{g}_x \colon [a,b] \to \mathbb{R}$ be defined by $\tilde{g}_x(\xi) = G_{p,q,r}(x,\xi)$. If ϕ_n is the eigenfunction corresponding to μ_n, $n \in \mathbb{N}$, then we compute

$$\begin{aligned}
|\langle \tilde{g}_x, \phi_n \rangle_r| &= \Big| \int_a^b \tilde{g}_x(\xi)\phi_n(\xi) r(\xi)\, \mathrm{d}\xi \Big| \\
&\geq \underline{r} \Big| \int_a^b G(x,\xi)\phi_n(\xi)\, \mathrm{d}\xi \Big| \\
&= \underline{r} |\mathcal{G}_{p,q,r}(\phi_n)(x)| \\
&= \underline{r} |\mu_n \phi_n(x)|.
\end{aligned}$$

Here $\underline{r} > 0$ is defined so that $r(x) \geq \underline{r}$ for all $x \in [a,b]$. Bessel's inequality then gives

$$\sum_{n=1}^{\infty} \underline{r}^2 |\mu_n \phi_n(x)|^2 \leq \sum_{n=1}^{\infty} |\langle \tilde{g}_x, \phi_n \rangle_r|^2 \leq \|\tilde{g}_x\|_r^2, \tag{7.24}$$

this holding for all $x \in [a,b]$. Since $\tilde{g}_x$ is bounded, it follows that $\|\tilde{g}_x\|_r^2 < \infty$. Thus the series on the left in (7.24) converges pointwise to a limit function of x. We claim that the convergence is actually uniform. This simply follows from the continuity of the function $\|\tilde{g}_x\|_r^2$ (this itself following from continuity of $\tilde{g}_x$). Thus we may term-by-term integrate (7.24) with respect to x. The left series when integrated gives

$$\begin{aligned}\sum_{n=1}^{\infty}\int_a^b \bar{r}^2|\mu_n\phi_n(x)|^2\mathrm{d}x &= \sum_{n=1}^{\infty}\bar{r}^2\mu_n^2\int_a^b \phi_n^2(x)\,\mathrm{d}x\\ &\geq \frac{\bar{r}^2}{R}\sum_{n=1}^{\infty}\mu_n^2\int_a^b \phi_n^2(x)r(x)\,\mathrm{d}x\\ &= \frac{\bar{r}^2}{R}\sum_{n=1}^{\infty}\mu_n^2,\end{aligned}$$

where $\bar{r} = \sup_{x\in[a,b]}\{r(x)\}$. Integration of the expression on the right in (7.24) gives

$$\begin{aligned}\int_a^b \|\tilde{g}_x\|_r^2\,\mathrm{d}x &= \int_a^b\int_a^b G_{p,q,r}(x,\xi)^2 r(\xi)\,\mathrm{d}x\,\mathrm{d}\xi\\ &\leq \int_a^b\int_a^b M^2R\,\mathrm{d}x\,\mathrm{d}\xi\\ &= M^2R(b-a)^2 < \infty,\end{aligned}$$

where $M = \sup_{x,\xi\in[a,b]}\{G_{p,q,r}(x,\xi)\}$. This allows us to conclude that

$$\bar{r}^2\sum_{n=1}^{\infty}\mu_n^2 < \infty.$$

In particular, we may conclude that the eigenvalues for $\mathcal{G}_{p,q,r}$ satisfy $\lim_{n\to\infty}\mu_n = 0$. Now we adopt the notation $G^m_{p,q,r}$ and $\mathcal{G}^m_{p,q,r}$ of Proposition 7.2.17. Since $\|\mathcal{G}^m_{p,q,r}\|_{r\to r} = |\mu_m|$, it follows that for any $u \in \mathrm{L}_2([a,b];\mathbb{R})$ we have

$$\|\mathcal{G}^m_{p,q,r}(u)\|_r = \Big\|\mathcal{G}_{p,q,r}(u) - \sum_{j=1}^{m}\mu_j\langle u,\phi_j\rangle_r\phi_j\Big\| \leq |\mu_m|\|u\|_r.$$

By Theorem 7.2.11 and Lemma 7.2.15, $\lim_{m\to\infty}|\mu_m| = 0$ so that

$$\lim_{m\to\infty}\Big\|\mathcal{G}_{p,q,r}(u) - \sum_{j=1}^{m}\mu_j\langle u,\phi_j\rangle_r\phi_j\Big\| = 0. \tag{7.25}$$

Now, for $n > m$ we directly have

$$\sum_{j=m}^{n}\mu_j\langle u,\phi_j\rangle\phi_j = \mathcal{G}_{p,q,r}\Big(\sum_{j=m}^{n}\langle u,\phi_j\rangle_r\phi_j\Big).$$

This gives

$$\begin{aligned}\Big|\sum_{j=m}^{n}\mu_j\langle u,\phi_j\rangle\phi_j(x)\Big| &= \Big|\mathcal{G}_{p,q,r}\Big(\sum_{j=m}^{n}\langle u,\phi_j\rangle_r\phi_j\Big)(x)\Big| \\ &= \Big|\int_a^b\Big(\sum_{j=m}^{n}\langle u,\phi_j\rangle_r G_{p,q,r}(x,\xi)\phi_j(\xi)\Big)\,\mathrm{d}\xi\Big| \\ &\le \frac{M}{\bar r}\Big|\int_a^b\Big(\sum_{j=m}^{n}\langle u,\phi_j\rangle_r\phi_j(\xi)r(\xi)\Big)\,\mathrm{d}\xi\Big| \\ &\le \frac{M\sqrt{b-a}}{\bar r}\Big(\sum_{j=m}^{n}|\langle u,\phi_j\rangle_r|^2\Big)^{1/2},\end{aligned}$$

where $M = \sup_{(x,\xi)\in[a,b]\times[a,b]}\{G(x,\xi)\}$, $\bar r = \sup_{x\in[a,b]}\{r(x)\}$, and where we have used the Cauchy-Bunyakovsky-Schwartz inequality in the penultimate step. From Bessel's inequality, as $n, m \to \infty$ the term in the last line goes to zero. This implies that the series

$$\sum_{j=1}^{\infty}\mu_j\langle u,\phi_j\rangle\phi_j$$

converges uniformly, and therefore converges to a continuous function. Since $\mathcal{G}_{p,q,r}(u)$ is also continuous, from (7.25) we deduce that

$$\mathcal{G}_{p,q,r}(u) = \sum_{j=1}^{m}\mu_j\langle u,\phi_j\rangle_r\phi_j \tag{7.26}$$

for any $u \in \mathrm{L}_2([a,b];\mathbb{R})$ with convergence being, again, uniform. If $f \in \mathrm{dom}(L_{p,q,r})$ then we may write $f = \mathcal{G}_{p,q,r}(u)$ for some $u \in \mathrm{L}_2([a,b];\mathbb{R})$. In this case we have

$$\mu_j\langle u,\phi_j\rangle_r = \langle u,\mathcal{G}_{p,q,r}(\phi_j)\rangle_r = \langle\mathcal{G}_{p,q,r}(u),\phi_j\rangle_r = \langle f,\phi_j\rangle_r$$

for any $j \in \mathbb{N}$. The equation (7.26) then gives

$$f = \sum_{j=1}^{\infty}\langle f,\phi_j\rangle\phi_j,$$

with convergence being uniform. This proves the final assertion of the theorem.

The first assertion, that $\{\phi_n\}_{n\in\mathbb{N}}$ is dense in $\mathrm{L}_2([a,b];\mathbb{R})$, will follow if we can show that $\mathrm{dom}(L_{p,q,r})$ is dense in $\mathrm{L}_2([a,b];\mathbb{R})$. From ***missing stuff*** we know that twice continuously differentiable functions on $[a,b]$, $C^2([a,b];\mathbb{R})$, are dense in $\mathrm{L}_2([a,b];\mathbb{R})$. We claim that the twice continuously differentiable functions satisfying the boundary

conditions of $\mathrm{dom}(L_{p,q,r})$ are dense in $C^2([a,b];\mathbb{R})$. We do this as follows. Note that the function $B \in C^2([a,b];\mathbb{R})$ given by

$$B(x) = (\alpha_0\beta_1 + \alpha_0 b\beta_0 - \alpha_1\beta_0 - a\alpha_0\beta_0)x^2 + (a^2\alpha_0\beta_0 + 2a\alpha_1\beta_0 - \alpha_0 b^2\beta_0 - 2\alpha_0 b\beta_1)x + (2\alpha_1 b\beta_1 + 2a\alpha_0 b\beta_1 - 2a\alpha_1\beta_1 - a^2\alpha_0\beta_1 + \alpha_1 b^2\beta_0 + a\alpha_0 b^2\beta_0 - 2a\alpha_1 b\beta_0 - a^2\alpha_0 b\beta_0)$$

satisfies the boundary conditions (it has been constructed to be a function quadratic in x whose coefficients are obtained by asking that the boundary conditions be satisfied). Given $f \in C^2([a,b];\mathbb{R})$ we "blend" f with B to get a function that is close to f, but which satisfies the boundary conditions. We do this as follows. For $\delta > 0$ define $\chi_\delta \colon \mathbb{R} \to \mathbb{R}$ by

$$\chi_\delta(x) = \begin{cases} 0, & x < 0 \\ 15(\frac{x}{\delta})^4 - 24(\frac{x}{\delta})^5 + 10(\frac{x}{\delta})^6, & x \in [0,\delta] \\ 1, & x > \delta. \end{cases}$$

One may verify that χ_δ is twice continuously differentiable, and that $\chi_\delta(0) = \chi_\delta'(0) = 0$ and that $\chi_\delta(\delta) = 1$ and $\chi_\delta'(\delta) = 0$. In Figure 7.3 Now, for a given $f \in C^2([a,b];\mathbb{R})$

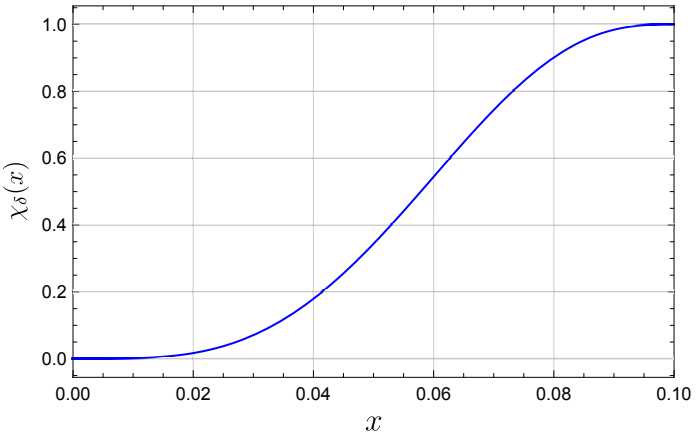


Figure 7.3 The graph of χ_δ for $\delta = \frac{1}{10}$

and small $\delta > 0$ define $f_\delta \colon [a,b] \to \mathbb{R}$ by

$$f_\delta(x) = f(x)\Big(1 - \chi_\delta(a + 2\delta - x) - \chi_\delta(x - b + 2\delta)\Big) + B(x)\Big(\chi_\delta(a + 2\delta - x) + \chi_\delta(x - b + 2\delta)\Big).$$

The function f_δ is designed to have the following properties:

1. $f_\delta \in C^2([a,b];\mathbb{R})$;
2. $f_\delta(x) = f(x)$ for $x \in [a + 2\delta, b - 2\delta]$;

3. $f_\delta(x) = B(x)$ for $x \in [a, a+\delta]$ and for $x \in [b-\delta, b]$;
4. for each $x \in [a+\delta, a+2\delta]$ and $x \in [b-2\delta, b-\delta]$ we have

$$f_\delta(x) = (1-\alpha(x))f(x) + \alpha(x)B(x),$$

where $\alpha(x) \in [0,1]$.

Thus f_δ "looks like" B near the boundaries of the interval, and equals f in the middle of the interval. Also, property 3 ensures that f_δ satisfies the boundary conditions of $\mathrm{dom}(L_{p,q,r})$. The definition of f_δ gives

$$\begin{aligned}
|f_\delta(x) - f(x)| &= \Big| f(x)\Big(-\chi_\delta(a+2\delta-x) - \chi_\delta(x-b+2\delta)\Big) + \\
&\qquad B(x)\Big(\chi_\delta(a+2\delta-x) + \chi_\delta(x-b+2\delta)\Big)\Big| \\
&\leq |f(x)\Big(\chi_\delta(a+2\delta-x) + \chi_\delta(x-b+2\delta)\Big)| + \\
&\qquad |B(x)\Big(\chi_\delta(a+2\delta-x) + \chi_\delta(x-b+2\delta)\Big)| \\
&\leq 2\Big(|f(x)| + |B(x)|\Big).
\end{aligned}$$

Since f and B are continuous on $[a,b]$, it follows that there exists $M > 0$ so that

$$|f_\delta(x) - f(x)| \leq M$$

for all $x \in [a,b]$. If r is bounded by $\bar{r} > 0$ one then computes

$$\begin{aligned}
\|f - f_\delta\|_r^2 &= \int_a^b (f(x) - f_\delta(x))^2 r(x)\,\mathrm{d}x \\
&= \int_a^{a+2\delta} (f(x) - f_\delta(x))^2 r(x)\,\mathrm{d}x + \int_{b-2\delta}^b (f(x) - f_\delta(x))^2 r(x)\,\mathrm{d}x \\
&\leq \int_a^{a+2\delta} M^2 \bar{r}\,\mathrm{d}x + \int_{b-2\delta}^b M^2 \bar{r}\,\mathrm{d}x \\
&= 4\delta M^2 \bar{r}.
\end{aligned}$$

This shows that for any $\epsilon > 0$ there exists a function $f_\delta \in \mathrm{dom}(L_{p,q,r}) \cap C^2([a,b];\mathbb{R})$ for which $\|f - f_\delta\|_r < \epsilon$.

With the above setup in place, let $f \in \mathrm{L}_2([a,b];\mathbb{R})$ and let $\epsilon > 0$. Then there exists $f_1 \in C^2([a,b];\mathbb{R})$ so that $\|f - f_1\|_r < \frac{\epsilon}{3}$ and there exists $f_2 \in C^2([a,b];\mathbb{R}) \cap \mathrm{dom}(L_{p,q,r})$ so that $\|f_1 - f_2\|_r < \frac{\epsilon}{3}$. Also, since $f_2 \in \mathrm{dom}(L_{p,q,r})$, there exists an $N \in \mathbb{N}$ so that $\|f_2 - f_{2,n}\| < \frac{\epsilon}{3}$ if $n \geq N$, where $f_{2,n}$ is the nth partial sum for the generalised Fourier series of f_2. We then compute

$$\begin{aligned}
\|f - f_{2,n}\|_r &= \|(f - f_1) + (f_1 - f_2) + (f_2 - f_{2,n})\|_r \\
&\leq \|f - f_1\|_r + \|f_1 - f_2\|_r + \|f_2 - f_{2,n}\|_r \\
&< \epsilon,
\end{aligned}$$

provided that $n \geq N$. This shows that any function in $L_2([a,b];\mathbb{R})$ can be arbitrarily well approximated by a finite linear combination of the eigenfunctions $\{\phi_n\}_{n\in\mathbb{N}}$, and the result now follows from***missing stuff***. ■

During the course of the proof we showed that the eigenvalues of $\mathcal{G}_{p,q,r}$ tend to zero in magnitude. From this and Lemma 7.2.15 we have the following corollary.

7.2.19 Corollary *If* $\mathrm{spec}_0(\mathrm{P}) = \{\lambda_\mathrm{n}\}_{\mathrm{n}\in\mathbb{N}}$ *then* $\lim_{\mathrm{n}\to\infty}|\lambda_\mathrm{n}| = \infty$.

We shall improve this result in Theorem 7.2.20.

7.2.4 Approximate behaviour of eigenvalues and eigenfunctions

While Theorem 7.2.18 is one of the triumphs of applied mathematics, telling us that solutions to some types of differential equations may be used to approximate quite arbitrary functions, it does not tell us much about the character of the solutions to the differential equation. In this section we set about determining the character of the large eigenvalues and their corresponding eigenfunctions.

7.2.4.1 Eigenvalue properties First let us state a result of general utility, giving a more refined description of the eigenvalues than is provided by Corollary 7.2.19.

7.2.20 Theorem *If* $\mathrm{p}(\mathrm{x}) > 0$ *for all* $\mathrm{x} \in [\mathrm{a},\mathrm{b}]$ *(as we have been assuming all along), then there are at most a finite number of positive eigenvalues for* P. *Therefore, it is possible to index* $\mathrm{spec}_0(\mathrm{P}) = \{\lambda_\mathrm{n}\}_{\mathrm{n}\in\mathbb{N}}$ *so that*

$$\cdots < \lambda_\mathrm{n} < \cdots < \lambda_{\mathrm{k}+1} < 0 < \lambda_\mathrm{k} < \cdots < \lambda_1,$$

and so that $\lim_{\mathrm{n}\to\infty}\lambda_\mathrm{n} = -\infty$. *When ordered in this way,* $\mathrm{spec}_0(\mathrm{P})$*, and the corresponding eigenfunctions, is said to have* ***descending order***.

Proof Define $E_{p,q,r}\colon \mathrm{dom}(L_{p,q,r}) \times \mathrm{dom}(L_{p,q,r}) \to \mathbb{R}$ by

$$E_{p,q,r}(y_1,y_2) = \int_a^b \Big(p(x)y_1'(x)y_2'(x) + q(x)y_1(x)y_2(x)\Big)\,\mathrm{d}x - \Big(p(x)y_1'(x)y_2(x)\Big)\Big|_a^b.$$

Let us now consider the left boundary condition $\alpha_1 y'(a) + \alpha_0 y(a) = 0$. If $\alpha_1 = 0$ then $y(a) = 0$ from which we deduce that $p(a)y_1'(a)y_2(a) = 0$. If $\alpha_1 \neq 0$ then we have $y_1'(a) = -\frac{\alpha_0}{\alpha_1}y_1(a)$ so that $p(a)y_1'(a)y_2(a) = -\frac{\alpha_0}{\alpha_1}p(a)y_2(a)^2$. In all cases this gives $p(a)y_1'(a)y_2(a) = -Ay_2(a)^2$ for some constant A. A similar conclusion holds at the right endpoint where we will have $p(b)y_1'(b)y_2(b) = By_2(b)^2$. This gives

$$E_{p,q,r}(y_1,y_2) = \int_a^b \Big(p(x)y_1'(x)y_2'(x) + q(x)y_1(x)y_2(x)\Big)\,\mathrm{d}x + Ay_2(a)^2 + By_2(b)^2. \tag{7.27}$$

Now we use integration by parts to derive

$$\int_a^b \Big(p(x)y_1'(x))y_2'(x)\,\mathrm{d}x = \Big(p(x)y_1'(x)y_2(x)\Big)\Big|_a^b - \int_a^b \Big(p(x)y_1'(x)\Big)'y_2(x)\,\mathrm{d}x.$$

Plugging this latter expression into the definition of $E_{p,q,r}$ we get

$$E_{p,q,r}(y_1, y_2) = -\int_a^b L_{p,q}(y_1)(x)y_2(x)\,\mathrm{d}x. \tag{7.28}$$

In particular, if λ_n is an eigenvalue with normalised eigenfunction ϕ_n, $n \in \mathbb{N}$, we have

$$E_{p,q,r}(\phi_n, \phi_n) = -\int_a^b L_{p,q}(\phi_n)\phi_n\,\mathrm{d}x = -\int_a^b r(x)\phi_n^2(x)\,\mathrm{d}x = -\lambda_n. \tag{7.29}$$

Let us leave this relation aside for a moment and return to (7.27). If $\underline{q} = \inf_{x\in[a,b]}\{q(x)\}$ and $\bar{r} = \sup_{x\in[a,b]}\{r(x)\}$ then, for $\|y\|_r = 1$ we compute

$$\begin{aligned} E_{p,q,r}(y, y) &= \int_a^b \Big(p(x)y'(x)^2 + q(x)y(x)^2\Big)\,\mathrm{d}x + Ay(a)^2 + By(b)^2 \\ &\geq \|\sqrt{p}y'\|_2^2 + \frac{\underline{q}}{\bar{r}}\int_a^b r(x)y^2(x)\,\mathrm{d}x + Ay(a)^2 + By(b)^2 \\ &= \|\sqrt{p}y'\|_2^2 + \frac{\underline{q}}{\bar{r}} + Ay(a)^2 + By(b)^2. \end{aligned} \tag{7.30}$$

Now there are four cases to consider: (1) $A, B \geq 0$, (2) $A \geq 0$ and $B < 0$, (3) $B \geq 0$ and $A < 0$, and (4) $A, B < 0$. In the first case we immediately have

$$E_{p,q,r}(y, y) \geq \|\sqrt{p}y'\|_2^2 + \frac{\underline{q}}{\bar{r}} \geq \frac{\underline{q}}{\bar{r}}.$$

In this case, the result follows from (7.29) since in this case we have proven that $\lambda_n \leq -\frac{\underline{q}}{\bar{r}}$. When one of A or B is negative, however, further estimates are required.

To this end, let $y \in \mathrm{dom}(L_{p,q,r})$ and define $f_y\colon [a, b] \to \mathbb{R}$ by

$$f_y(x) = \int_a^x r(\xi)y^2(\xi)\,\mathrm{d}\xi.$$

By the mean value theorem, if $\|y\|_r = 1$ then we have

$$\begin{aligned} f_y'(c) &= \frac{f_y(b) - f_y(a)}{b - a} \\ \implies \quad (b - a)r(c)y(c)^2 &= \int_a^b r(x)y^2(x)\,\mathrm{d}x = 1 \end{aligned}$$

for some $c \in (a, b)$. Next, integration by parts gives

$$\begin{aligned} \int_a^c y(x)y'(x)\,\mathrm{d}x &= y^2(x)\Big|_a^c - \int_a^c y(x)y'(x)\,\mathrm{d}x \\ \implies \quad 2\int_a^c y(x)y'(x)\,\mathrm{d}x + y(a)^2 &= y(c)^2. \end{aligned}$$

Working on this last equality we derive, with $\underline{r} = \inf_{x\in[a,b]}\{r(x)\}$ and $\underline{p} = \inf_{x\in[a,b]}\{p(x)\}$,

$$\begin{aligned} y(a)^2 &= y(c)^2 - 2\int_a^c y(x)y'(x)\,\mathrm{d}x \\ &\leq y(c)^2 + 2\int_a^b |y(x)||y'(x)|\,\mathrm{d}x \\ &\leq \frac{1}{(b-a)p(c)} + \frac{2}{\underline{r}}\int_a^c r(x)|y(x)||y'(x)|\,\mathrm{d}x \\ &\leq \frac{1}{(b-a)p(c)} + \frac{2}{\underline{r}}\Big(\int_a^b r(x)y'(x)^2\,\mathrm{d}x\Big)^{1/2} \\ &\leq \frac{1}{(b-a)p(c)} + \frac{1}{\sqrt{\underline{r}\,\underline{p}}}\Big(\int_a^b p(x)y'(x)^2\,\mathrm{d}x\Big)^{1/2}, \end{aligned}$$

where, in the last step, we used the Cauchy-Bunyakovsky-Schwartz inequality, along with the fact that $\|y\|_r = 1$. This then gives

$$\begin{aligned} & y(a)^2 \leq \tilde{C} + \tilde{C}\|\sqrt{p}y'\|_r \\ \Longrightarrow \quad & Ay(a)^2 \geq A\tilde{C} + A\tilde{C}\|\sqrt{p}y'\|_2, \end{aligned}$$

if $A < 0$, and where $\tilde{C} = \max\{\frac{1}{(b-a)p(c)}, \frac{1}{\sqrt{\underline{r}\,\underline{p}}}\}$. Defining $C = -A\tilde{C}$ gives

$$Ay(a)^2 \geq -C\|\sqrt{p}y'\|_2 - C.$$

A similar analysis may be used to determine the estimate

$$By(b)^2 \geq -C\|\sqrt{p}y'\|_2 - C$$

when $B < 0$. Combining these estimates with (7.30) we obtain, provided that $A, B < 0$.

$$\begin{aligned} E_{p,q,r}(y,y) &\geq \|\sqrt{p}y'\|_2^2 + \frac{\underline{q}}{\bar{r}} + Ay(a)^2 + By(b)^2 \\ &\geq \|\sqrt{p}y'\|_2^2 + \frac{\underline{q}}{\bar{r}} - 2C - 2C\|\sqrt{p}y'\|_2 \\ &= (\|\sqrt{p}y'\|_2^2 - C)^2 - C^2 - 2C + \frac{\underline{q}}{\bar{r}} \\ &\geq -C^2 - 2C + \frac{\underline{q}}{\bar{r}}. \end{aligned}$$

By taking y to be the normalised eigenfunction ϕ_n for the eigenvalue λ_n, $n \in \mathbb{N}$, this shows that as long as $A, B < 0$ we have $\lambda_n \leq C^2 + 2C - \frac{\underline{q}}{\bar{r}}$. This proves the result when $A, B < 0$. The cases when $A \geq 0$ and $B < 0$ and when $A < 0$ and $B \geq 0$ follow in the same manner. ■

7.2.21 Remarks

1. An upshot of the result is that it is always possible to introduce a constant $c < 0$ so that the problem $P(c)$ defined by (7.17), i.e., the problem with eigenvalues shifted by c, has all negative eigenvalues.
2. Because of the seemingly endemic distaste for negative numbers, many authors change the sign of λ to ensure that there at most finitely many negative eigenvalues, and that the positive eigenvalues tend to ∞ as $n \to 0$. Other authors use $-L_{p,q}$ in place of $L_{p,q}$. However, negative numbers do not scare this author, so they are allows as eigenvalues.
3. The function $E_{p,q,r}\colon \operatorname{dom}(L_{p,q,r}) \times \operatorname{dom}(L_{p,q,r}) \to \mathbb{R}$ introduced in the proof of the theorem often can be interpreted physically as the energy of the system. •

Now let us turn to a basic result which gives a characterisation of the eigenvalues. As usual, we denote

$$\begin{aligned} \bar{p} &= \sup_{x\in[a,b]}\{p(x)\}, \quad \underline{p} = \inf_{x\in[a,b]}\{p(x)\},\\ \bar{q} &= \sup_{x\in[a,b]}\{q(x)\}, \quad \underline{q} = \inf_{x\in[a,b]}\{q(x)\},\\ \bar{r} &= \sup_{x\in[a,b]}\{r(x)\}, \quad \underline{r} = \inf_{x\in[a,b]}\{r(x)\}, \end{aligned}$$

and with this notation we have the following result.

7.2.22 Theorem *Consider the eigenvalue problem* P *of equation (7.15) and assume that the boundary conditions ensure that* $\mathrm{y(a)y'(a) = y(b)y'(b) = 0}$. *Let* $\mathrm{k} \in \{1, 2\}$ *be the number of endpoints at which eigenfunctions are specified to vanish by the boundary conditions. Then for each* $\mathrm{n} \in \mathbb{N}$

$$\Big(\frac{\mathrm{k_n}\pi}{\mathrm{b-a}}\Big)^2 \frac{\mathrm{p}}{\underline{\mathrm{r}}} + \frac{\mathrm{q}}{\underline{\mathrm{r}}} \le \lambda_{\mathrm{n}} \le \Big(\frac{\mathrm{k_n}\pi}{\mathrm{b-a}}\Big)^2 \frac{\bar{\mathrm{p}}}{\bar{\mathrm{r}}} + \frac{\bar{\mathrm{q}}}{\bar{\mathrm{r}}}.$$

Proof First we prove a technical result.

1 Lemma *Suppose that we have functions* $\mathrm{p, q, r, \tilde{p}, \tilde{q}, \tilde{r}\colon [a, b] \to \mathbb{R}}$ *satisfying*

$$\mathrm{\tilde{p}(x) \le p(x), \quad \tilde{q}(x) \le q(x), \quad \tilde{r}(x) \le r(x).}$$

for each $\mathrm{x} \in [\mathrm{a}, \mathrm{b}]$. *Consider the two eigenvalue problems* P *and* $\tilde{\mathrm{P}}$ *defined by*

$$\mathrm{(py')' - qy = \lambda ry} \qquad \begin{aligned} \alpha_1\mathrm{y'(a)} + \alpha_0\mathrm{y(a)} &= 0\\ \beta_1\mathrm{y'(b)} + \beta_0\mathrm{y(b)} &= 0 \end{aligned}$$

and

$$\mathrm{(\tilde{p}y')' - \tilde{q}y = \tilde{\lambda}\tilde{r}y} \qquad \begin{aligned} \alpha_1\mathrm{y'(a)} + \alpha_0\mathrm{y(a)} &= 0\\ \beta_1\mathrm{y'(b)} + \beta_0\mathrm{y(b)} &= 0, \end{aligned}$$

respectively. Suppose that the boundary conditions ensure that A *and* B *in (7.27) are nonnegative. If* $\operatorname{spec}_0(\mathrm{P}) = \{\lambda_\mathrm{n}\}_{\mathrm{n}\in\mathbb{N}}$ *and* $\operatorname{spec}_0(\tilde{\mathrm{P}})$ *have descending order, then we have* $\tilde{\lambda}_\mathrm{n} \le \lambda_\mathrm{n}$ *for each* $\mathrm{n} \in \mathbb{N}$.

Proof ***missing stuff*** ▼

■

A look at Figure 7.2 indicates that the eigenvalues for the problem of Section 7.2.1.2 become equally spaced as $n \to \infty$. Interestingly, this is generally true, and this is the content of the next result.

7.2.23 Theorem *For the eigenvalue problem* P *of equation (7.15) let* $\operatorname{spec}_0(\mathrm{P}) = \{\lambda_\mathrm{n}\}_{\mathrm{n}\in\mathbb{N}}$ *be given the descending order. Then*

$$\lim_{\mathrm{n}\to\infty} \frac{\lambda_\mathrm{n}}{\mathrm{n}^2} = -\frac{\pi^2}{(\mathrm{b}-\mathrm{a})^2}.$$

Proof First of all, by a suitable choice of c, let us transform the problem to a problem $P(c)$ as given by (7.17) having the property that all eigenvalues are negative. This is possible by Theorem 7.2.20. Note that the theorem holds for P if and only if it holds for $P(c)$ since the eigenvalues for $P(c)$ are merely those of P shifted by c. Thus we may write $\lambda_n = -\omega_n^2$ for some $\omega_n > 0$, $n \in \mathbb{N}$. We now transform[8] the problem via a change of dependent and independent variable as follows:

$$\xi = \int_a^x \sqrt{\frac{r(x)}{p(x)}}\,\mathrm{d}x, \quad \eta(\xi) = \Big(p(x(\xi))r(x(\xi))\Big)^{1/4} y(x(\xi)).$$

Now a direct computation shows that (7.15) is equivalent to

$$\frac{\mathrm{d}^2\eta(\xi)}{\mathrm{d}\xi^2} - \rho(\xi)\eta(\xi) = -\omega^2\eta(\xi)$$

where $\xi \in [0, b-a]$, and where

$$\rho(\xi) = \frac{\sigma''(\xi)}{\sigma(\xi)} + \frac{q(x(\xi))}{r(x(\xi))},$$

with $\sigma(\xi) = \Big(p(x(\xi))r(x(\xi))\Big)^{1/4}$. The punchline is that we may as well assume that $a = 0$, $b = \ell$, and that $p(x) = r(x) = 1$ for $x \in [0, \ell]$. Let us do this, and for simplicity return to our previous notation. Thus we have the problem

$$y'' - qy = -\omega^2 y \qquad \begin{aligned} \alpha_1 y'(0) + \alpha_0 y(0) &= 0 \\ \beta_1 y'(\ell) + \beta_0 y(\ell) &= 0. \end{aligned}$$

We shall thus prove the theorem for this simplified system.***missing stuff*** ■

[8] This transformation is known as the ***Liouville transformation***.

7.2.4.2 Eigenfunction properties Now let us turn our attention briefly to a discussion of eigenfunctions. We discuss the general eigenvalue problem P given by equation (7.15). Let us write $\operatorname{spec}_0(P) = \{\lambda_n\}_{n\in\mathbb{N}}$ in descending order, with $\{\phi_n\}_{n\in\mathbb{N}}$ the corresponding eigenfunctions. Motivated by the vibrating string as a boundary value problem, let us say that a point $x \in [a, b]$ is a ***node*** for the eigenfunction ϕ_n, $n \in \mathbb{N}$, if $\phi_n(x) = 0$.

7.2.24 Theorem *Let* $\mathrm{n} \in \mathbb{N}$ *and suppose that* $\mathrm{x}_1, \mathrm{x}_2$ *are nodes for* ϕ_n *with the properties that* $\mathrm{x}_1 < \mathrm{x}_2$ *and there are no nodes for* ϕ_n *in* $(\mathrm{x}_1, \mathrm{x}_2)$*. Then* $\phi_{\mathrm{n}+1}$ *has a node in* $(\mathrm{x}_1, \mathrm{x}_2)$*.*

Proof Suppose that ϕ_{n+1} has no node in (x_1, x_2). We may as well suppose that both ϕ_n and ϕ_{n+1} are strictly positive on (x_1, x_2). The eigenfunctions satisfy the equations

$$\begin{aligned}(p\phi_n')' - (q + \lambda_n r)\phi_n &= 0\\ (p\phi_{n+1}')' - (q + \lambda_{n+1} r)\phi_{n+1} &= 0.\end{aligned}$$

Now multiply the first equation by ϕ_{n+1} and the second by ϕ_n and subtract the resulting two equations to get

$$(p\phi_n')'\phi_{n+1} - (p\phi_{n+1}')'\phi_n = (\lambda_n - \lambda_{n+1})r\phi_n\phi_{n+1}$$

The expression on the right is by design positive on (x_1, x_2). Thus integrating we get

$$\int_{x_1}^{x_2} \Big((p\phi_n')'\phi_{n+1} - (p\phi_{n+1}')'\phi_n\Big)\,\mathrm{d}x > 0. \tag{7.31}$$

The integrand is

$$\Big((p\phi_n')'\phi_{n+1} - (p\phi_{n+1}')'\phi_n\Big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(p(\phi_n'\phi_{n+1} - \phi_n\phi_{n+1}')\Big).$$

Therefore, integrating the left-hand side of (7.31) gives

$$p(x_2)\phi_n'(x_2)\phi_{n+1}(x_2) - p(x_1)\phi_n'(x_1)\phi_{n+1}(x_1) > 0, \tag{7.32}$$

using the fact that $\phi_n(x_1) = \phi_n(x_2) = 0$. However, since $\phi_n(x_2) = 0$ and $\phi_n(x) > 0$ for $x \in (x_1, x_2)$, we have $\phi_n'(x_2 < 0$. Similarly we deduce that $\phi_n'(x_1) > 0$. This shows that the expression (7.32) cannot hold, so our original assumption that ϕ_{n+1} cannot vanish on (x_1, x_2) must not be true. ■

7.2.5 Summary

This chapter has been an almost entirely theoretical one. In the exercises you will be asked to explore some aspects of the theory, as well as see how it arises in some applications. The points one should take away from this chapter are the following.

1. The properties that we saw in ***missing stuff*** for trigonometric series generalise to a far more general class of problems, of which Fourier series are an example.

2. Part of the reason for our success in being able to say so much about the eigenvalue problem (7.15) is that it defines a self-adjoint mapping. This, at least, guarantees it having real eigenvalues and orthogonal eigenfunctions. However, much more than this is needed, as in infinite dimensions, self-adjointness is not even sufficient to ensure the existence of eigenvalues. Thus, in some sense, the results of this section are somewhat miraculous.
3. Key in the success of the development is the Green function. This being the inverse of $L_{p,q,r}$ and having such nice properties (as enumerated particularly in Proposition 7.2.14) allows us to formulate the problem of finding an eigenfunction essentially as that of minimising a function on the unit norm functions in $\mathrm{L}_2([a,b];\mathbb{R})$ (cf. Theorem 7.2.16). In this way, we are able to emulate the spirit of Exercise 7.1.12, even though the analysis is rather more complicated. In particular, we should point out that the use of the Arzela-Ascoli theorem in the proof of Theorem 7.2.16 makes this result one that qualifies as difficult.
4. Interestingly, one is able to prove some useful facts about the behaviour of the eigenvalues of general boundary value problems. This can be useful in some types of approximate analysis.
5. Similarly, one can understand the behaviour of eigenfunctions to some extent.

7.2.6 Notes

The classical text for this material, although it is rather advanced, is that of Coddington and Levinson [1984]. Another text, delivered at a somewhat more palatable pace is that of Troutman [1994]. For an introduction, we refer to [Stakgold 1979]. It is possible to attain such completeness results without using the Green function [e.g., Troutman 1994].

Exercises

7.2.1 Let $I \subseteq \mathbb{R}$ be an interval and let $r\colon I \to \mathbb{R}$ be Riemann integrable with the property that $r(x) > 0$ for every $x \in I$. Show that

$$\langle f, g\rangle_r = \int_I f\bar{g}r\,\mathrm{d}x$$

defines an inner product on $C^0(I;\mathbb{F})$.

7.2.2 For the boundary value problems below do the following:

1. find all values of λ for which the problem admits a nontrivial solution X;
2. for each λ give a nontrivial solution X;
3. for the three values of λ smallest in absolute value, plot the corresponding solution X.

(a) $X''(x) + \lambda X(x) = 0$, $X(0) = 0$, $X(l) = 0$.

(b) $X''(x) + \lambda X(x) = 0$, $X'(0) = 0$, $X(l) = 0$.
(c) $X''(x) = \lambda X(x)$, $X'(0) = 0$, $X'(l) = 0$.
(d) $X''(x) - \lambda X(x) = 0$, $X(0) = 0$, $X'(l) = 0$.

7.2.3 Show that the map L defined for the example of Section 7.2.1.1 is not continuous with respect to the norm defined by the inner product $\langle \cdot, \cdot \rangle_2$.

7.2.4 Consider the following potential equation with boundary conditions that are used to model the deflection of an airplane wing of length ℓ and width w:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad \begin{aligned} &u(0, y) = 0, \quad \frac{\partial u}{\partial x}(w, y) = K u(w, y) \\ &u(x, 0) = f_1(x), \quad u(x, \ell) = f(x). \end{aligned} \tag{7.33}$$

Here K is a design constant that governs the deflection of the wing, and f determines the shape of the cross-section of the wing at its tip.

(a) Using separation of variables, show that the above boundary value problem leads to following eigenvalue problem:

$$y''(x) = \lambda y(x), \quad x \in [0, w] \qquad \begin{aligned} &y(0) = 0 \\ &y'(w) = K y(w). \end{aligned}$$

(b) For the various cases for K (it can be any real number), determine the eigenvalues and eigenvectors for the problem from (a).
(c) Use the eigenvalues and eigenvectors from (b) to obtain a solution to the boundary value problem (7.33).

Throughout this chapter we have considered only real solutions to the problem P. However, as with Fourier series, it is possible to allow complex solutions as well. The following problem indicates how to do this, and why the reduction to the real case can be made without loss of generality.

7.2.5 Consider the problem P defined by (7.15).

(a) Show that if $y\colon [a, b] \to \mathbb{C}$ is a complex solution to P, then the complex conjugate $\bar{y}\colon [a, b] \to \mathbb{C}$ is also a solution to P. In a natural way, to each complex solution to P associate a real solution.
(b) If $y_1, y_2\colon [a, b] \to \mathbb{R}$ are two real solutions to P, show that $y = y_1 + iy_2$ is a complex solution to P.

7.2.6 Often a physical problem will have boundary conditions that specify the ratio of y' and y at the endpoints. Show that in this case there exists $\alpha, \beta \in [0, 2\pi)$ so that the boundary conditions of (7.15) have the form

$$\begin{aligned} \cos\alpha y(a) - \sin\alpha y'(a) &= 0 \\ \cos\beta y(b) + \sin\beta y'(b) &= 0. \end{aligned}$$

7.2.7 Show directly using integration by parts that the eigenvalue problem

$$L_{p,q}(y) = (py')' - qy = \lambda y \qquad \begin{aligned} \alpha_1 y'(a) + \alpha_0 y(a) &= 0 \\ \beta_1 y'(b) + \beta_0 y(b) &= 0 \end{aligned}$$

is symmetric. That is, show that

$$\int_a^b L_{p,q}(f)(x)g(x)\,\mathrm{d}x = \int_a^b f(x)L_{p,q}(g)(x)\,\mathrm{d}x$$

for $f, g \in \mathrm{dom}(L_{p,q})$.

In this chapter we have introduced the notion of a generalised Fourier series corresponding to functions satisfying a certain class of boundary value problem. Sadly, the Fourier series as considered in ***missing stuff*** is not quite of this form. In this exercise you will reconcile at least part of this by showing that the Fourier series of ***missing stuff*** are eigenfunctions for a self-adjoint boundary value problem.

7.2.8 Consider the boundary value problem defined on the interval $[0, \ell]$ by

$$L(y) = y''(x) = \lambda y(x) \qquad \begin{aligned} y(0) &= y(\ell) \\ y'(0) &= y'(\ell). \end{aligned}$$

Let $\mathrm{dom}(L)$ be those continuously differentiable functions possessing a second derivative in $\mathrm{L}_2([0, \ell], \mathbb{R})$ and which satisfy the boundary conditions.

(a) Following the proof of Proposition 7.2.7, show that if $f, g \in \mathrm{dom}(L)$ then $\langle L(f), g\rangle_2 = \langle f, L(g)\rangle_2$.

(b) Directly using integration by parts, show that if $f, g \in \mathrm{dom}(L)$ then $\langle L(f), g\rangle_2 = \langle f, L(g)\rangle_2$.

This shows (in two ways) that L is self-adjoint, although it is not of the form we have considered in this chapter. It also turns out that L has an infinite collection of eigenvalues, and the corresponding eigenfunctions form a complete orthonormal family. This is the content of the next question.

(c) Find the eigenvalues and eigenfunctions for L.

7.2.9 Consider the one-dimensional Schrödinger[9] equation

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\frac{\partial^2\psi}{\partial x^2}$$

for particle of mass m moving in one-dimension. The constant $\hbar$ is ***Planck's constant***. The dependent variable ψ is the ***wave function*** for the particle. Suppose that the particle moves on a ring so that the wave function can be taken to be 2π-periodic in x.

[9] After Erwin Rudolf Josef Alexander Schrödinger (1887–1961)

(a) Use the method of separation of variables to arrive at an eigenvalue problem for the Schrödinger equation.
(b) Find the eigenvalues and eigenfunctions for the eigenvalue problem of part (a).
(c) Use your eigenfunctions to obtain a general expression for the wave function.

7.2.10 Use the construction provided in the proof of Theorem 7.2.13 to compute the Green function for the example of Section 7.2.1.1.

7.2.11 Show that $G_{p,q,r}(x,\xi) = G_{p,q,r}(\xi,x)$ for all $x, \xi \in [a,b]$.

7.2.12 Show that $\mathcal{G}_{p,q,r}$ is self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_r$.

7.2.13 Verify that if $u \in \mathrm{L}_2([a,b];\mathbb{R})$ then $\mathcal{G}_{p,q,r}(u) \in \mathrm{dom}(L_{p,q,r})$.
Hint: *Use the definition of $\mathcal{G}_{\mathrm{p,q,r}}$ in terms of the Green function, and then use the properties of the Green function enumerated in Theorem 7.2.13.*

7.2.14 Let $p\colon [0,1] \to \mathbb{R}$ be positive and continuously differentiable and consider the eigenvalue problem

$$p(x)y''(x) = \lambda y(x) \qquad \begin{aligned} y(0) &= 0\\ y(1) &= 0. \end{aligned}$$

Let $\{\lambda_n\}_{n\in\mathbb{N}}$ denote the eigenvalues for this system, with $\{\phi_n\}_{n\in\mathbb{N}}$ the corresponding normalised eigenfunctions.

(a) Show that there exists functions $\tilde{p}$, $\tilde{q}$, and $\tilde{r}$ defined on $[0,1]$ with $\tilde{p}$ and $\tilde{r}$ positive, and constants α_0, α_1, β_0, and β_1 so that the eigenvalue problem

$$(\tilde{p}y')' - \tilde{q}y = \lambda\tilde{r}y \qquad \begin{aligned} \alpha_1 y'(0) + \alpha_0 y(0) &= 0\\ \beta_1 y'(1) + \beta_0 y(1) &= 0 \end{aligned}$$

has eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ and eigenfunctions $\{\phi'_n\}_{n\in\mathbb{N}}$.
(b) Show that the eigenfunctions from part (a) are orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle_{\tilde{r}}$.

7.2.15 In Exercise 6.5.2 you determined the temperature distribution across the walls of a heat exchanger tube. In part (c) of that problem, you simply solved a differential equation, and no eigenvalue problem was encountered.

(a) If one applies the method of separation of variables (forgetting the assumption that the temperature is independent of θ, what is the eigenvalue problem defined by the "r-part" of the problem, assuming that $T_0 = T_1 = 0$?
(b) Are any of the eigenvalues positive?
Hint: *Understand enough of the proof of Theorem 7.2.20 to answer the question.*
(c) Is zero an eigenvalue?

In the following exercise you will be introduced to ***Rayleigh's principle***, named after John William Strutt, or more commonly, Lord Rayleigh (1842–1919).

7.2.16 Consider the eigenvalue problem (7.15) and suppose that all eigenvalues are positive (which can always be done by shifting the eigenvalues if necessary). For a nonzero function $y \in \mathrm{dom}(L_{p,q,r})$ define the ***Rayleigh quotient*** by

$$R_{p,q,r}(y) = \frac{E_{p,q,r}(y)}{\|y\|_r^2}.$$

Denote the eigenvalues in descending order by $\{\lambda_n\}_{n\in\mathbb{N}}$ and the corresponding normalised eigenfunctions by $\{\phi_n\}_{n\in\mathbb{N}}$. For $N \in \mathbb{N}$ let $\mathcal{F}_N$ denote the collection of nonzero functions in $\mathrm{dom}(L_{p,q,r})$ that are orthogonal to the first $N-1$ eigenfunctions.

For $y \in \mathrm{dom}(L_{p,q,r})$ let

$$y_N = \sum_{n=1}^{N} \langle y, \phi_n \rangle \phi_n$$

denote the Nth partial sum in the generalised Fourier series.

(a) Show that $E_{p,q,r}(y_N, y) = \sum_{n=1}^{N} \lambda_n \langle y, \phi_n \rangle^2$.
Hint: *Use (7.28).*

(b) Use the result from (a) to show that for $y \in \mathrm{dom}(L_{p,q,r})$ we have

$$\sum_{n=1}^{\infty} \lambda_n c_n^2 \leq E_{p,q,r}(y, y).$$

(c) Prove the following result.

Proposition $\lambda_N = R_{p,q,r}(y_N) = \min_{y\in\mathcal{F}_N} R_{p,q,r}(y)$.

Hint: *Use Parseval's equality for functions in* $\mathcal{F}_N$.

Section 7.3

Second-order singular boundary value problems

The eigenvalue problem P studied in the previous chapter (i.e., the one defined by equation (7.15)) has some properties that simply do not hold in certain physical situations. These are characterised by the problem being defined on a closed interval $[a, b]$ of finite length, and by the fact that on this interval the functions p and r are strictly positive. In this chapter we look to relax these assumptions. Doing so, it turns out, opens a Pandora's box of complications that make the boundary value problems we study in this chapter much more difficult than those of Section 7.2. It turns out that there are many physical problems which retain the character of the easier problems, but there are others which do not. Our first order of business is to classify the more general class of boundary value problems we look at. This we do in Section 7.3.1.

7.3.1 Classification of boundary value problems

The problems we study in this chapter are so-called "singular problems." In this section we shall first distinguish singular problems from nonsingular problems, and then look in detail at the various types of singular problems, at least as these can be ascertained by simply "looking at" the problem. It turns out that there is a further classification in the singular case, that between the "limit-point" and "limit-circle" cases, that has no analogue in the nonsingular case. This is a crucial distinction, and we look into this in Section 7.3.1.2.

7.3.1.1 Regular and singular boundary value problems Our first order of business in to generalise in a fairly straightforward manner the boundary value problems of Section 7.2. One of the generalisations is to allow intervals that are open and/or unbounded. This requires some new notation for the inner product defined by a positive function r. Thus, if $I \subseteq \mathbb{R}$ is an interval and if $r\colon I \to \mathbb{R}$ is a positive continuous function, then we as usual define the inner product between two $\mathbb{C}$-valued functions on I by

$$\langle f, g\rangle_r = \int_I f(x)\bar{g}(x)r(x)\,\mathrm{d}x. \tag{7.34}$$

We shall denote by $\mathrm{L}^r_2(I;\mathbb{C})$ those functions $f\colon I \to \mathbb{C}$ for which $\|f\|_r = \sqrt{\langle f, f\rangle_r} < \infty$. If I is closed and bounded, then $\mathrm{L}^r_2(I;\mathbb{C}) = \mathrm{L}_2(I;\mathbb{C})$ (cf. Exercise 7.3.1). However, for intervals that are not closed and bounded, these sets of functions may not be the same, so we need to make a distinction here that was not necessary in Section 7.2.

Let us now try to be as precise as we can about what we mean by a boundary value problem.

7.3.1 Definition Let $I \subseteq \mathbb{R}$ be an interval. A ***boundary value problem*** on I consists of the following data:

(i) functions $p, q, r \colon I \to \mathbb{R}$ satisfying
  (a) p is twice continuously differentiable,
  (b) q and r are continuous, and
  (c) $p(x), r(x) > 0$ for all $x \in I$;

(ii) for each finite endpoint $e \in I$, a boundary condition of the form $\alpha_1 y'(e) + \alpha_0 y(e) = 0$;

(iii) for each finite endpoint $e \in \operatorname{cl}(I) \setminus I$ with the property that the limits

$$\lim_{x \to e} p(x), \quad \lim_{x \to e} r(x), \quad \lim_{x \to e} q(x)$$

exist with the first two limits being strictly positive, a boundary condition of the form $\alpha_1 y'(e) + \alpha_0 y(e) = 0$;

(iv) a subspace $\operatorname{dom}(L_{p,q,r})$ of functions from I to $\mathbb{C}$ consisting of those functions f with the properties that
  (a) f is differentiable,
  (b) there exists a function $f'' \in \mathrm{L}^r_2(I; \mathbb{C})$ so that

$$f'(x) = f'(x_0) + \int_{x_0}^{x} f''(\xi) \, \mathrm{d}\xi,$$

  and
  (c) f satisfies the boundary conditions (ii) and/or (iii) when these are applicable;

(v) the linear mapping $L_{p,q,r} \colon \operatorname{dom}(L_{p,q,r}) \to \mathrm{L}^r_2(I; \mathbb{C})$ defined by

$$L_{p,q,r}(y) = r^{-1}\Big((py')' - qy\Big).$$

An endpoint $e \in \operatorname{cl}(I)$ is ***regular*** if it is finite and if the limits

$$\lim_{x \to e} p(x), \quad \lim_{x \to e} r(x), \quad \lim_{x \to e} q(x) \tag{7.35}$$

exist with the first two limits being strictly positive. An endpoint that is not regular is ***singular***. A boundary value problem is ***regular*** if both endpoints are regular. A boundary value problem that is not regular is ***singular***. •

Let us give some examples of boundary value problems so that we can try to better understand the above lengthy definition.

7.3.2 Examples

1. We shall encounter in Section 7.3.4.1 the Bessel equation. Let us not physically motivate the equation at this point, but merely produce an eigenvalue problem that we will use to demonstrate our classification procedures. The eigenvalue problem we consider is

$$xy''(x) + y'(x) - \frac{\nu^2}{x}y = \lambda x y(x),$$

which is ***Bessel's equation of order $\nu \geq 0$***. Thus $p(x) = x$, $q(x) = -\frac{\nu^2}{x}$, and $r(x) = x$. The matter of classification is determined once the interval is specified. Let $\epsilon > 0$ and take $I = [\epsilon, 1]$, $p(x) = x$, $q(x) = 0$, and $r(x) = x$. Since I is closed and bounded, this defines a regular problem, and so requires the corresponding boundary conditions. Let us be concrete and choose boundary conditions $y(\epsilon) = 0$ and $y(1) = 0$. This, of course, is the sort of problem we looked at in detail in Chapter 6.

2. Let us use the same data as in Example 1, but now take $I = (\epsilon, 1]$. Thus the only difference is the exclusion of the left endpoint of the interval. However, since we are taking $\epsilon > 0$, the limits of part (iii) and equation (7.35) of the definition are met, so we still need to have the boundary conditions specified at these points. Thus we see that the problem is still regular.

3. Let us keep the same p, q, and r as in Example 1, but now take $I = (0, 1]$. Now the limits of part (iii) and equation (7.35) of the definition are *not* satisfied since both p and r assume the value of zero in the limit. Thus this problem is singular, and one specifies only the boundary condition at the right endpoint.

4. In Section 7.3.4.2 we will see how the Legendre equation[10] comes up in looking at the Laplacian in spherical coordinates. Here we merely reproduce the differential equation:

$$(1 - x^2)y''(x) - 2xy'(x) = \lambda y(x).$$

Thus $p(x) = 1 - x^2$, $q(x) = 0$, and $r(x) = 1$. The interval of definition for this differential equation is $(-1, 1)$. Since the limit for p at both endpoints is zero, the problem is singular. •

7.3.3 Assumption Note that in Example 2, the removal of the left endpoint does nothing to essentially change the problem from Example 1. This is generally the case, as can easily be seen. Thus we make the following blanket assumption from now on.

> *If a problem is regular, we shall assume that the interval of definition is closed and bounded.* •

[10] After Adrien-Marie Legendre (1752–1833).

7.3.4 Remark Note that in Examples 3 and 4, while p and r are indeed positive on I as they are required to be, they vanish in the limit at one or both of the endpoints. Some authors allow the problem of Example 3 to be defined on $[0, 1]$ and of Example 4 to be defined on $[-1, 1]$. In this case the problem arises because p and r then vanish on the physical domain of the system. However, such an approach has some disadvantages, and the wiser approach is to not allow p and r to vanish on the physical domain. •

One of the main features of the singular problem is its allowance of intervals that are not closed and bounded. The possible intervals on which a singular problem might be defined are

1. $I = (-\infty, b]$,
2. $I = [a, \infty)$,
3. $I = [a, b)$,
4. $I = (a, b]$,
5. $I = (a, b)$,
6. $I = (a, \infty)$,
7. $I = (-\infty, b)$,
8. $I = \mathbb{R}$.

We intend to reduce this to essentially two cases, at least as far as the development of the theory goes. Any given singular example may take any of the above forms. However, we wish to contend that there are essentially two cases to consider. We do this as follows.

7.3.5 Lemma *Any problem with an interval of the form 1–4 can be transformed into a problem with interval* $[0, \infty)$*, and any problem with an interval of the form 5–8 can be transformed into a problem with interval* $(-\infty, \infty)$*.*

Proof The lemma is proved by making a change of the independent variable for the problem. Thus we prove the lemma by merely listing the change of variable in the eight cases. In each case, it is easy to verify that the change of variable is a infinitely differentiable bijection with infinitely differentiable inverse, thus ensuring that the change of variable is valid.

1 We define the map $\xi : (-\infty, b] \to [0, \infty)$ by $\xi(x) = b - x$.
2 We define the map $\xi : [a, \infty) \to [0, \infty)$ by $\xi(x) = x - a$.
3 We define the map $\xi : [a, b) \to [0, \infty)$ by $\xi(x) = \tan(\frac{\pi}{2} \frac{x-a}{b-a})$.
4 We define the map $\xi : (a, b] \to [0, \infty)$ by $\xi(x) = \tan(\frac{\pi}{2} \frac{b-x}{b-a})$.
5 We define the map $\xi : (a, b) \to (-\infty, \infty)$ by $\xi(x) = \tan(\pi \frac{2x-(b+a)}{b-a})$.
6 We define the map $\xi : (a, \infty) \to (-\infty, \infty)$ by $\xi(x) = \ln(x - a)$.
7 We define the map $\xi : (-\infty, b) \to (-\infty, \infty)$ by $\xi(x) = \ln(b - x)$.
8 There is nothing to do in this case. ■

The lemma tells us that, up to a change of the independent variable, there are essentially two singular problems, one defined on the interval $[0, \infty)$ and the other defined on the interval $(-\infty, \infty)$. This is of some help in developing the general theory as it keeps us from having to deal with the eight cases separately. Let us see how this works in some examples.

7.3.6 Examples (Example 7.3.2 cont'd)

1. By Lemma 7.3.5, the singular Bessel boundary value problem Example 7.3.2–3 can be reduced so as to be defined on the interval $[0, \infty)$. Let us be explicit about this, in fact. As we see in the proof of Lemma 7.3.5, to define the problem on the new interval $[0, \infty)$ we should make the change of independent variable $\xi(x) = \tan(\frac{\pi}{2}(1 - x))$. If $\eta(\xi) = y(x(\xi))$ then we ascertain that η satisfies the differential equation

$$\tfrac{\pi}{4}(1 + \xi^2)^2(\pi - 2\arctan\xi)\eta''(\xi) + \tfrac{\pi}{4}\Big(2\xi(\pi - 2\arctan\xi) - 2(1 + \xi^2)\Big)\eta'(\xi) = \lambda(1 - \tfrac{2}{\pi}\arctan\xi)\eta(\xi) \qquad (7.36)$$

   for $\xi \in [0, \infty)$.

2. For the Legendre boundary value problem, Example 7.3.2–4, the interval can be transformed to $(-\infty, \infty)$. Indeed, as in Example 3, we can grab an explicit expression for the transformation of the independent variable from the proof of the lemma: $\xi(x) = \tan(\pi x)$. One may directly compute, with notation as in Example (7.36),

$$(1+\xi^2)^2(\pi^2-\arctan^2\xi)\eta''(\xi)+\Big(2\xi(\pi^2-\arctan^2\xi)-2\arctan\xi(1+\xi^2)\Big)\eta'(\xi) = \lambda\eta(\xi), \qquad (7.37)$$

   with $\xi \in (-\infty, \infty)$.

7.3.7 Remark In equations (7.36) and (7.37) we see explicitly how Lemma 7.3.5 works. Note, however, that one never does this transformation in practice, but retains the equations in the original coordinates. We merely produce the results of making the transformation to explicitly indicate the manner in which the eight different singular problems are reduced to two in Lemma 7.3.5. •

7.3.1.2 The limit-point and limit-circle cases In this section we discuss a means of distinguishing two fundamental classes of singular boundary value problems. The exact reasons why the distinction goes along the lines presented should not be obvious to a first time reader; it only follows from a detailed look into the theory of these singular problems. In the development in this section we make use of Lemma 7.3.5 to reduce the singular problems to those defined on the interval $[0, \infty)$ or on the interval $(-\infty, \infty)$.

7.3.8 Definition Consider a boundary value problem defined on an interval $I \in \{[0, \infty), (-\infty, \infty)\}$. Let $e \in \{-\infty, +\infty\}$ be an infinite endpoint of I, and define

$$I_e = \begin{cases} (-\infty, 0], & e = -\infty \\ [0, \infty), & e = +\infty. \end{cases}$$

The problem is in the ***limit-circle case at*** **e** if for a given $\lambda_0 \in \mathbb{C}$, every y satisfying

$$(py')' - qy = \lambda_0 ry \tag{7.38}$$

has the property that $y \in \mathrm{L}_2^r(I_e; \mathbb{C})$. If the problem is not in the limit-circle case on I_e, it is in the ***limit-point case at*** **e**. •

In order to make sure that the distinction between the limit-circle and limit-point cases depends only on the problem data I, p, q, and r, and not on a choice of a particular $\lambda_0 \in \mathbb{C}$ in (7.38), we prove the following result due to Hermann Klaus Hugo Weyl (1885–1955).

7.3.9 Theorem *Consider a boundary value problem defined on an interval* $\mathrm{I} \in \{[0, \infty), (-\infty, \infty)\}$. *Let* $\mathrm{e} \in \{-\infty, +\infty\}$ *be an infinite endpoint for* I *and let* $\mathrm{I_e}$ *be the corresponding interval as given in Definition 7.3.8. If there exists a* $\lambda_0 \in \mathbb{C}$ *so that every solution* y *of*

$$(\mathrm{py}')' - \mathrm{qy} = \lambda_0 \mathrm{ry}$$

belongs to $\mathrm{L}_2^{\mathrm{r}}(\mathrm{I_e}, \mathbb{C})$, *then for every* $\lambda \in \mathbb{C}$, *every solution* y *of*

$$(\mathrm{py}')' - \mathrm{qy} = \lambda \mathrm{ry}$$

belongs to $\mathrm{L}_2^{\mathrm{r}}(\mathrm{I_e}, \mathbb{C})$.

Proof ***missing stuff*** ■

The theorem ensures tells us that the notion of being in the limit-circle case is a problem dependent notion. Furthermore, it tells us that to determine whether a problem is in the limit-circle case, one need only check the solution of the differential equation

$$(py')' - qy = \lambda_0 ry$$

for a particular $\lambda_0 \in \mathbb{C}$, allowing one to choose an easy one, if it happens that this is possible. Let us illustrate this for a couple of examples.

7.3.10 Examples

1. We take the eigenvalue problem $y'' = \lambda y$, and consider two possible types of intervals.
   (a) If $I = [0, \infty)$ then choosing $\lambda = 0$ gives the solution $y(x) = Ax + B$ to the differential equation. This solution is generally not in $\mathrm{L}_2([0, \infty); \mathbb{R})$, and so we deduce that the problem is in the limit-point case at $+\infty$.
   (b) Next we take $I = (-\infty, \infty)$. Here there are two singular endpoints. The previous case allows us to conclude that the problem is in the limit-point case at $+\infty$. Also, since a function of the form $y(x) = Ax + B$ is not generally in $\mathrm{L}_2((-\infty, 0]; \mathbb{R})$, we see that the problem is also in the limit-point case at $-\infty$.

2. Let us now look at Bessel's equation of order ν, which gave the eigenvalue problem

$$xy''(x) + y'(x) - \frac{\nu^2}{x}y = \lambda x y(x).$$

We shall look at two singular problems associated with this equation.

(a) The first case we consider is $I = (0, b]$ for some $0 < b < \infty$. Since it suffices to investigate solutions to the equation for a particular λ, let us choose $\lambda = 0$. We may then see by direct computation that two linearly independent solutions to the equation are $y_1(x) = x^\nu$ and $y_2(x) = x^{-\nu}$ if $\nu > 0$. For $\nu = 0$ two linearly independent solutions are $y_1(x) = 1$ and $y_2(x) = \ln x$. For $\nu > 1$ we compute

$$\|y_2\|_r^2 = \int_0^b x x^{-2\nu}\,\mathrm{d}x = \frac{x^{2(1-\nu)}}{2(1-\nu)}\Big|_0^b. \tag{7.39}$$

Therefore, for $\nu > 1$, the function y_2 is not in $\mathrm{L}_2^r((0, b]; \mathbb{R})$, and we conclude that we are in the limit-point case at 0. For $\nu = 1$ we compute

$$\|y_2\|_r^2 = \int_0^b x x^{-2}\,\mathrm{d}x = \ln x\Big|_0^b.$$

This means that we are in the limit-point case when $\nu = 1$. For $0 < \nu < 1$ we see from (7.39) that we are in fact in the limit-circle case at $x = 0$. For $\nu = 0$ we have

$$\|y_2\|_r^2 = \int_0^b x \ln^2 x\,\mathrm{d}x = \Big(\frac{x^2}{4} - \frac{x^2 \ln x}{2} + \frac{x^2 \ln^2 x}{2}\Big)\Big|_0^b.$$

Thus $y_2 \notin \mathrm{L}_2^r((0, b]; \mathbb{R})$, and the zeroth-order Bessel equation is in the limit-point case at $x = 0$ when $\nu = 0$. The above is summarised in Table 7.1.

Table 7.1 Classification of Bessel's equation into limit-point or limit-circle cases for the interval $(0, b]$

ν	limit-point or limit-circle at 0
0	limit-point
$\nu \in (0, 1)$	limit-circle
$\nu \in [1, \infty)$	limit-point

(b) Next we take $[a, \infty)$ for some $a > 0$. Here the singular endpoint is at $+\infty$. The solutions we determined in the previous part of the problem still hold. In this case, the above integrals may be used, along with the appropriately modified endpoints, to produce the categorisation of Table 7.2.

Table 7.2 Classification of Bessel's equation into limit-point or limit-circle cases for the interval $[a, \infty)$

ν	limit-point or limit-circle at ∞
0	limit-point
$\nu \in (0, 1]$	limit-point
$\nu \in (1, \infty)$	limit-circle

3. Now we look at the Legendre equation

$$(1 - x^2)y''(x) - 2xy'(x) = \lambda y(x)$$

defined for $x \in (-1, 1)$. There are two singular endpoints. Here we may verify that for $\lambda = 0$ two linearly independent solutions are $y_1(x) = 1$ and $y_2(x) = \ln(1 + x) - \ln(1 - x)$. Clearly $y_1 \in \mathrm{L}^r_2((-1, 1); \mathbb{R})$. One also computes

$$\int_{-1}^{1} y_2^2(x)\,\mathrm{d}x = \frac{2\pi^2}{3},$$

so that $y_2 \in \mathrm{L}_2((-1, 1); \mathbb{R})$, thus allowing us to conclude that the problem is in the limit-circle case at each endpoint. •

It ought not be clear at this point what is the relevance of the distinction between the limit-circle and limit-point cases. One reason for our interest in this distinction will be elucidated in Theorem 7.3.13 below.

7.3.2 Eigenvalues and eigenfunctions for singular problems

Note that the problems covered in some detail in Section 7.2 are always regular. As was the case with these problems, the eigenvalue problem is of principle interest, even for singular problems. We denote the eigenvalue problem again as P, thus making a slight abuse of notation:

$$\boxed{(py')' - qy = \lambda ry, \quad y \in \mathrm{dom}(L_{p,q,r})} \tag{7.40}$$

with $\lambda \in \mathbb{C}$ an unknown parameter. As we did with the regular problems of Section 7.2, we shall denote by $\mathrm{spec}_0(P) \subseteq \mathbb{C}$ the set of complex numbers λ for which (7.40) admits a solution. Thus $\mathrm{spec}_0(P)$ is the collection of eigenvalues for the problem, as usual. The notion of spectrum for singular problems is more sophisticated than for regular problems. As we shall see in Section 7.3.3, the natural generalisation of the notion of spectrum allows for points in the spectrum that are not eigenvalues.

7.3.2.1 Basic properties In this section we shall show that the essential features of the eigenvalues and eigenfunctions for regular problems extend to singular problems. In particular, eigenvalues are real, and eigenfunctions for distinct eigenvalues are orthogonal. For the regular problems this followed from self-adjointness of $L_{p,q,r}$. However, a look at the proofs for this self-adjointness (see Propositions 7.2.7 and 7.2.10) reveals that they depend on the regular boundary conditions being present at each endpoint. As this is not necessarily the case for singular problems, we need to start from scratch.

Let us get right to this.

7.3.11 Proposition *Consider the problem* P *defined on the interval* I. *Suppose that for any two eigenfunctions* y_1 *and* y_2, *perhaps associated with distinct eigenvalues, for* P *we have the property that at each endpoint* e *for* I *we have*

$$\lim_{\mathrm{x}\to\mathrm{e}} \mathrm{p}(\mathrm{x})\mathrm{y}_1(\mathrm{x})\mathrm{y}_2'(\mathrm{x}) = \lim_{\mathrm{x}\to\mathrm{e}} \mathrm{p}(\mathrm{x})\mathrm{y}_2(\mathrm{x})\mathrm{y}_1'(\mathrm{x}). \tag{7.41}$$

Then $\mathrm{spec}(\mathrm{P}) \subseteq \mathbb{R} \subseteq \mathbb{C}$. *Moreover,*

(i) at a regular endpoint e, *(7.41) is satisfied at* e, *and*

(ii) if $\lim_{\mathrm{x}\to\mathrm{e}} \mathrm{p}(\mathrm{x}) = 0$ *and if all eigenfunctions are bounded, then (7.41) holds at* e.

Proof Let $\lambda \in \mathrm{spec}(P)$ have the eigenfunction $y\colon I \to \mathbb{C}$. Then

$$(py')' - qy = \lambda ry,$$

and since p, q, and r are $\mathbb{R}$-valued, we also have

$$(p\bar{y}')' - q\bar{y} = \bar{\lambda} r\bar{y}.$$

Multiply the first of these equations by $\bar{y}$ and the second by y and subtract the resulting equations to get

$$(py')'\bar{y} - (p\bar{y}')'y = (\lambda - \bar{\lambda})ry\bar{y}.$$

Now let $[a, b] \subseteq I$ and integrate over this interval:

$$\int_a^b \Big((py')'\bar{y} - (p\bar{y}')'y\Big)\,\mathrm{d}x = (\lambda - \bar{\lambda}) \int_a^b ry\bar{y}\,\mathrm{d}x. \tag{7.42}$$

The integral on the left in (7.42) may be evaluated by parts:

$$\begin{aligned}\int_a^b \Big((py')'\bar{y} - (p\bar{y}')'y\Big)\,\mathrm{d}x &= p(x)y'(x)\bar{y}(x)\Big|_a^b - p(x)\bar{y}'(x)y(x)\Big|_a^b - \int_a^b py'\bar{y}'\,\mathrm{d}x + \int_a^b p\bar{y}'y'\,\mathrm{d}x \\ &= p(x)y'(x)\bar{y}(x)\Big|_a^b - p(x)\bar{y}'(x)y(x)\Big|_a^b.\end{aligned}$$

Now we take the limit as a approaches the left endpoint e_1 and b the right endpoint e_2, and by the assumption (7.41), this gives

$$\lim_{\substack{a\to e_1\\ b\to e_2}} \int_a^b \Big((py')'\bar{y} - (p\bar{y}')'y \Big)\,\mathrm{d}x = 0.$$

Now, returning back to (7.42), we have

$$(\lambda - \bar{\lambda}) \lim_{\substack{a\to e_1\\ b\to e_2}} \int_a^b r(x)|y(x)|^2\,\mathrm{d}x = 0. \tag{7.43}$$

Since the integral is nonzero, this implies that $\bar{\lambda} = \lambda$, as desired.

Now suppose that e is a regular endpoint for I. Then we may as well suppose that $e \in I$, as per Assumption 7.3.3. Then we have $\alpha_1 y'(e) + \alpha_0 y(e) = 0$. If $\alpha_1 = 0$ then we immediately have $y(e) = 0$ and so (i) holds. If $\alpha_1 \neq 0$ then we have

$$\alpha_1 y_1'(e) + \alpha_0 y_1(e) = 0, \quad \alpha_1 y_2'(e) + \alpha_0 y_2(e) = 0.$$

Thus

$$y_1'(e)y_2(e) = -\frac{\alpha_0}{\alpha_1} y_1(e)y_2(e), \quad y_2'(e)y_1(e) = -\frac{\alpha_0}{\alpha_1} y_2(e)y_1(e),$$

giving $y_1'(e)y_2(e) = y_2'(e)y_1(e)$, and so (i) follows in this case as well.

The statement (ii) is clear as both limits in (7.41) are zero in this case. ■

Note that, as expected, our result indicates that the eigenvalues for a regular problem are real. It also shows that the eigenvalues for at least some singular problems are real. In fact, the condition (7.41) turns out to be satisfied in almost any boundary value problem. For example, in all of the problems of Example 7.3.2, both singular and regular, the condition (7.41) is satisfied, although in the singular cases, this is not obvious.

Now let us turn to orthogonality of eigenfunctions.

7.3.12 Proposition *If the problem* P *satisfies the condition (7.41) at each endpoint, and for each pair of eigenfunctions* y_1 *and* y_2*, then eigenfunctions for distinct eigenvectors of* P *are orthogonal with respect to the inner product* $\langle\cdot,\cdot\rangle_{\mathrm{r}}$.

Proof Let $\lambda_1, \lambda_2 \in \operatorname{spec}(P)$ be distinct with respective real eigenfunctions y_1 and y_2. We may then proceed exactly as in the proof of Proposition 7.3.11, replacing y with y_1 and $\bar{y}$ with y_2, to get an equation which is the analogue of (7.43) in this case:

$$(\lambda_1 - \lambda_2) \lim_{\substack{a\to e_1\\ b\to e_2}} \int_a^b r(x)y_1(x)\bar{y}_2(x)\,\mathrm{d}x = 0.$$

Since $\lambda_1 \neq \lambda_2$, the result follows. ■

7.3.2.2 Classification by spectral properties The classification above has been done along the lines of the interval of definition for the problem. This is essentially determined by the physics of the problem, and one does not have to do any analysis to decide whether a problem is regular or singular, and if it is singular, which of the two cases of Lemma 7.3.5 applies. Let us turn to a useful characterisation, albeit one that cannot be determined readily by simply looking at the physics. To this end, we shall say that the boundary value problem P defined by (7.40) is ***pseudo-regular*** if $\operatorname{spec}(P)$ is a countable set $\{\lambda_n\}_{n\in\mathbb{N}}$ which can be ordered so that

1. $\cdots < \lambda_n < \cdots < \lambda_{k+1} < 0 < \lambda_k < \cdots < \lambda_1$ and
2. $\lim_{n\to\infty} \lambda_n = -\infty$.

Thus a problem is pseudo-regular when its spectrum has the properties deduced in Theorem 7.2.20 for regular problems. In particular, of course, a regular problem is pseudo-regular. Some singular problems are also pseudo-regular, although it is generally non-trivial to ascertain whether a given singular problem is pseudo-regular. The valuable fact about pseudo-regular problems is that the eigenfunction properties provided for regular problems in Section 7.2 often extend easily to pseudo-regular problems. In particular, the eigenfunctions for a pseudo-regular boundary value problem form a complete orthonormal family. It can be shown that the singular problems of Example 7.3.2 are actually pseudo-regular, although this is not an entirely trivial deduction.

The following result makes a connection between spectral properties and the limit-circle/limit-point discussion of Section 7.3.1.2. This is a quite nontrivial result, and we refer to the references for a proof.

7.3.13 Theorem *If a boundary value problem defined on the interval* $\mathrm{I} \in \{[0,\infty), (-\infty,\infty)\}$ *is in the limit-circle case at all infinite endpoints, then it is pseudo-regular.*

The theorem does *not* say that if a problem is in the limit-point case at one or more endpoints then it is prohibited from being pseudo-regular. Indeed, there are examples with endpoints in the limit-point case for which the boundary value problem is pseudo-regular, some of these having physical importance. In such cases, one can often proceed "by hand" to determine pseudo-regularity.

7.3.3 The theory for singular boundary value problems

Let us give for now the briefest outline of how the general theory progresses for singular eigenvalue problems. It turns out that it is convenient in this development to use the Lebesgue-Stieltjes integral.

7.3.3.1 Problems defined on $[0,\infty)$ In this case we proceed by considering a sequence of regular problems defined on intervals $[0, b_n]$, where $\{b_j\}_{j\in\mathbb{N}}$ is an increasing sequence with $\lim_{j\to\infty} b_j = \infty$. At the boundary b_j we impose the regular

boundary condition $y(b_j) = 0$. Since each of the problems on the intervals $[0, b_j]$ is then regular, it possesses an countable eigenvalue sequence $\{\lambda_{n,j}\}_{n\in\mathbb{N}}$ which we assume to be in descending order. For each fixed $j \in \mathbb{N}$ we define a monotone increasing function $\rho_j \colon [0, \infty) \to \mathbb{R}$... no time!

7.3.3.2 Problems defined on $(-\infty, \infty)$

Tune in next year...

7.3.4 Applications that yield singular boundary value problems

In the preceding sections we presented a few singular eigenvalue problems as illustrations of certain parts of the general development. Interestingly, singular eigenvalue problems are very common in applications, so the attention devoted to them is merited, even though they require a level of sophistication a cut above that shown for the regular problems in Section 7.2. In this section we look at a small collection of physical problems that exhibit singular behaviour.

7.3.4.1 The vibrating of a drum

We consider a circular drum of radius b. One may determine that the partial differential equation governing the vertical deflection u of the drumhead is

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u,$$

where u is the Laplacian of Section 6.5. Since the drumhead is circular, it makes sense to work in polar coordinates (r, θ) defined in the usual manner by

$$x = r\cos\theta, \quad y = r\sin\theta.$$

In Exercises 6.5.1 and 6.5.2 the Laplacian in polar coordinates is derived to be

$$\nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}.$$

Therefore the partial differential equation governing the vertical displacements $u(r, \theta, t)$ of the drumhead is then given explicitly in polar coordinates by

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}.$$

If we suppose that the outer edge of the drumhead is glued down (or something), a natural boundary condition is $u(b, \theta, t) = 0$. We should also require that u be a 2π-periodic function of θ.

To this equation, we apply the venerable method of separation of variables. Let us simplify life by looking for radially symmetric vibrations of the drumhead. This means that we ask that u be independent of θ. Substituting $u(r, t) = R(r)T(t)$ into the differential equation gives

$$\frac{1}{c^2}R(r)\ddot{T}(t) = R''(r)T(t) + \frac{1}{r}R'(r)T(t).$$

Now, as usual, divide by $R(r)T(t)$ to get

$$\frac{1}{c^2}\frac{\ddot{T}(t)}{T(t)} = \frac{R''(r)}{R(r)} + \frac{1}{r}\frac{R'(r)}{R(r)}.$$

Carrying on as usual, we declare that both sides of this equation must be equal to a constant which we denote by λ, giving the two differential equations

$$\begin{aligned}\ddot{T}(t) &= c^2\lambda T(t)\\ rR''(r) + R'(r) &= \lambda r R(r).\end{aligned}$$

We recognise the second of these equations as the Bessel equation of order 0. Since $r \in (0, b]$, we are in the singular case as exemplified by Example 7.3.2–3. What's more, from Table 7.1 we see that the problem is in the limit-circle case at the singular endpoint $x = 0$. From Theorem 7.3.13 we see that this implies that the problem is pseudo-regular. Thus we the eigenvalue problem will have a collection of eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ that are, but for a finite number, negative, and $\lim_{n\to\infty}\lambda_n = -\infty$. The corresponding normalised eigenfunctions $\{\phi_n\}_{n\in\mathbb{N}}$ will comprise a complete orthonormal family relative to the inner product $\langle\cdot,\cdot\rangle_r$.

The above characterisation of the solution to the eigenvalue problem is of some interest as it tells us a great deal about the character of the problem. However, the Bessel equation is such a classic and common equation that public opinion dictates that we say something about the solution. One can verify by understanding the proof of Theorem 7.2.20 that the eigenvalues of the Bessel equation are nonpositive. Furthermore, one can verify that 0 is not an eigenvalue (see Exercise 7.3.3). This makes it valid to get rid of the λ in the solution by making the change of independent variable $x = \sqrt{-\lambda}r$, and then defining $y(x) = R(\frac{1}{\sqrt{-\lambda}}x)$. The differential equation for y is then

$$xy''(x) + y'(x) + xy(x) = 0, \tag{7.44}$$

which is the zeroth-order Bessel equation in standard form. In a course on differential equations, one will learn how to obtain a polynomial (essentially) series representation for the solutions of (7.44) about the singular point 0. We shall not provide this series representation. Let us merely say that the zeroth-order (or any order, for that matter) Bessel's equation is a linear, second-order differential equation, albeit one with non-constant coefficients. This entitles it to two linearly independent solutions. It is the form of these functions, at least as a series expansion, that one obtains in one's differential equation course. In that course, you will arrive at the two linear independent solutions that are typically denoted J_0 and Y_0. J_0 is called the ***zeroth-order Bessel function of the first kind*** and Y_0 is called the ***zeroth-order Bessel function of the second kind***. In Figure 7.4 are plotted J_0 and Y_0. Note that J_0 has a well-defined limit at $x = 0$, but that Y_0 is unbounded at $x = 0$. In fact, one can verify that the singularity of Y_0 at 0 is logarithmic. Note that *any* solution of (7.44) will be of the form $y(x) = c_1J_0(x) + c_2Y_0(x)$. Going back to the

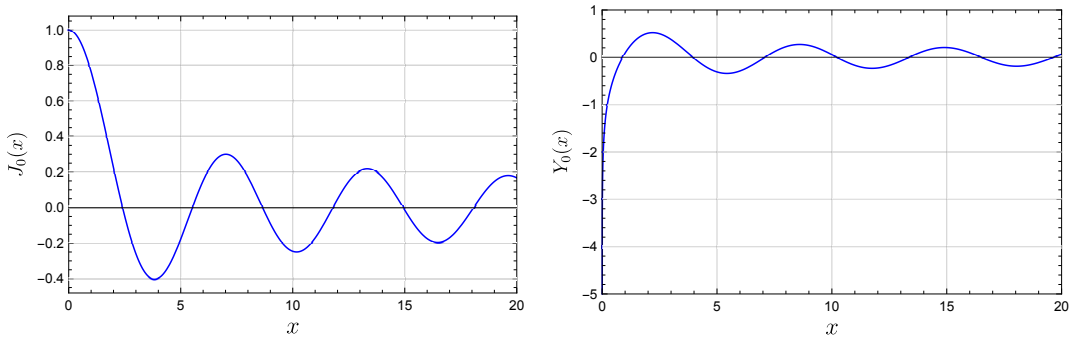


Figure 7.4 J_0 (left) and Y_0 (right)

original variables we have

$$R(r) = c_1 J_0(\sqrt{-\lambda} r) + c_2 Y_0(\sqrt{-\lambda} r).$$

Let us now see how we can use the solution of the equation (7.44) to obtain information about the eigenvalues and eigenfunctions for the singular eigenvalue problem

$$rR''(r) + R'(r) = \lambda r R(r), \qquad R(b) = 0.$$

On physical grounds (this also comes out of the theory of Section 7.3.3) we reject the presence of Y_0 in our solution as this would imply unbounded deflections of the drumhead. Thus the physically useful solutions for the eigenvalue problem will be multiples of J_0. The boundary condition $R(b) = 0$ then translates to

$$R(\sqrt{-\lambda} b) = c_1 J_0(\sqrt{-\lambda} b) = 0.$$

This gives the algebraic equation $J_0(\sqrt{-\lambda} b) = 0$ that must be satisfied by the eigenvalues. That is to say, for every root z of J_0, there corresponds an eigenvalue satisfying $\sqrt{-\lambda} b = z$. From Figure 7.4 we find it believable that it is possible to enumerate the roots $\{z_n\}_{n\in\mathbb{N}}$ of J_0 nicely in ascending order. This then gives the eigenvalues as $\{-\frac{z_n^2}{b^2}\}_{n\in\mathbb{N}}$. The eigenfunctions are then simply $R_n(r) = J_0(\sqrt{-\lambda_n} r) = J_0(\frac{z_n}{b} r)$. These functions are necessarily orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_r$ so that we have

$$\int_0^b r R_n(r) R_m(r)\, \mathrm{d}r = 0$$

if $m \neq n$. We may also normalise the eigenfunctions by defining

$$c_n = \Big(\int_0^b r R_n^2(r)\, \mathrm{d}r \Big)^{-1/2}$$

so that the functions $\{\phi_n = c_n R_n\}_{n\in\mathbb{N}}$ form a complete orthonormal family. This orthonormal family may be used to obtain an expression for the displacement $u(r, t)$ of the drumhead at all times (see Exercise 7.3.7).

Many other physical systems yield eigenfunctions in terms of Bessel functions, and some of these arise in the exercises in this section and in Section 7.2.

7.3.4.2 The Laplacian in spherical coordinates Given our definition in Section 6.5 of the Laplacian for planar domains, it is clear that one may define the Laplacian on a domain in $\mathbb{R}^n$ by

$$\nabla^2 f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

In this section we shall be interested in the case where $n = 3$. What's more, just as we used polar coordinates in the preceding section, in this section we shall use nonstandard coordinates, now the ***spherical coordinates*** (r, θ, ϕ) defined by

$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r \cos\theta.$$

These coordinates are illustrated in Figure 7.5 where we can see that r is, of course,

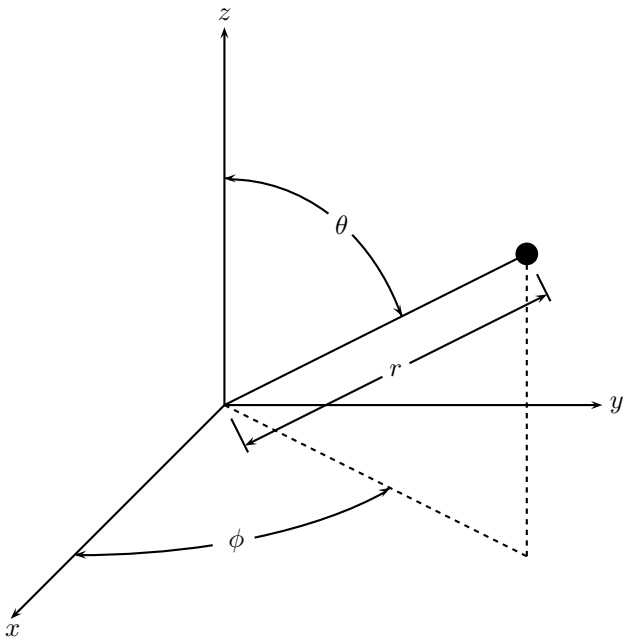


Figure 7.5 Spherical coordinates

the distance from the origin, and θ and ϕ are what one would normally think of as "latitude" and "longitude." Note that $\theta \in (0, \pi)$ and $\phi \in (-\pi, \pi)$. If we express the three-dimensional Laplacian,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

in spherical coordinates we have, after the requisite computations (see Exercise 7.3.8),

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial f}{\partial r}\Big) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial f}{\partial\theta}\Big) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}. \tag{7.45}$$

As with the Laplacian in two-dimensions, the Laplacian may be used to describe the steady-state temperature distribution in a uniform solid. That is to say, if the temperature is a function u on a domain in $\mathbb{R}^3$, the equation $\nabla^2 u = 0$ will describe the equilibrium temperature of the solid. So suppose that we are given a spherical solid of radius b with a known temperature distribution on its boundary. Further suppose that the distribution of temperature on the boundary is independent of ϕ in spherical coordinates. The temperature $u(r,\theta)$ in the interior of the body satisfies the boundary value problem

$$\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial u}{\partial r}\Big) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial u}{\partial\theta}\Big) = 0, \qquad u(b,\theta) = f(\theta).$$

To make this equation more tractable, we engage in trickery. We introduce the new variable $s = \cos\theta \in (-1,1)$, and then we compute (see Exercise 7.3.8) that the above boundary value problem becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial v}{\partial r}\Big) + \frac{1}{r^2}\frac{\partial}{\partial s}\Big((1-s^2)\frac{\partial v}{\partial s}\Big) = 0, \qquad v(b,s) = g(s), \tag{7.46}$$

where $v(r,s) = u(r,\arccos s)$ and $g(s) = f(\arccos s)$. In the usual manner, we seek a separable solution $u(r,\theta) = R(r)S(s)$. We substitute this into the partial differential equation to get

$$S(s)R''(r) + \frac{2}{r}S(s)R'(r) + \frac{1}{r^2}R(r)\Big((1-s^2)S'(s)\Big)' = 0.$$

Division by $R(r)S(s)/r^2$ gives

$$\frac{\Big((1-s^2)S'(s)\Big)'}{S(s)} = -r^2\frac{R''(r)}{R(r)} - 2r\frac{R'(r)}{R(r)}.$$

The usual argument of setting both sides of the equation equal to a constant λ gives the two ordinary differential equations

$$\begin{aligned}\Big((1-s^2)S'(s)\Big)' &= \lambda S(s)\\ r^2R''(r) + 2rR'(r) &= -\lambda R(r).\end{aligned}$$

The first of these equations we recognise as the eigenvalue problem associated with Legendre's equation as discussed in Example 7.3.2–4.

With this physical example as motivation, let us say a few things about the Legendre equation. As it is defined for $s \in (-1, 1)$ we see that it is singular, and is in the limit-circle case at both endpoints as seen in Example 7.3.10–3. Therefore, as with the zeroth-order Bessel equation, the Legendre eigenvalue problem is pseudo-regular, and so has all the properties of a regular eigenvalue problem. Thus the eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ for the problem are negative, except possibly for an at most finite number, and the corresponding normalised eigenfunctions $\{\phi_n\}_{n\in\mathbb{N}}$ form a complete orthonormal family. Unlike the situation with the Bessel equation, one can actually be explicit about the eigenvalues and eigenfunctions for the Legendre equation. To do this in a methodical manner, one should use the same theory for obtaining polynomial series expansions about singular points for differential equations as was alluded to in Section 7.3.4.1 in relation to the Bessel equation. We shall sidestep this, and merely present the eigenvalues and eigenfunctions. The eigenvalues are $\{\lambda_n = -n(n+1)\}_{n\in\mathbb{N}_0}$ and the corresponding bounded eigenfunctions are $\{P_n\}_{n\in\mathbb{N}}$ where P_n is given by ***Rodrigues' formula***:

$$P_n(s) = \frac{1}{n!2^n}\frac{\mathrm{d}^n}{\mathrm{d}s^n}(s^2-1)^n, \quad n \in \mathbb{N}_0.$$

These are called the ***Legendre polynomials***. There are also unbounded solutions to Legendre's equation, but we reject these, again on physical grounds, although one can make this rigorous. That the stated eigenvalues and eigenfunctions are indeed eigenvalues and eigenfunctions is easy to verify (see Exercise 7.3.9). However, one must also show that these are *all* of the eigenvalues and eigenfunctions, and this requires a little more work, although it is not extraordinarily difficult.

Our solution to the eigenvalue problem solves half of the boundary value problem. To fully determine the solution, one must also consider the R-equation. This is yet another singular equation, so let us just quit while we are ahead, and mention that it is possible to use series methods for ordinary differential equations to derive a solution for the R-equation upon substitution of the eigenvectors obtained from the Legendre equation.

7.3.4.3 The approximate age of the earth In 1820 Fourier proposed that his series methods could be extended to determine the age of the earth. William Thomson, more commonly known as Lord Kelvin, (1824–1907) picked this up, and used computations as we give here to deduce a figure of between 4×10^8 and 10^9 years, based upon some approximate data. While these estimates have been improved upon through the use of advanced equipment (not of the mathematical variety), it is interesting to go through Kelvin's argument.

The first approximation one makes is that the earth is flat (!). In doing so, the earth is parameterised by the depth from the surface, the other dimensions being assumed to be infinite. We let $x \in [0, \infty)$ denote the distance from the surface of the earth, and t denotes time, as usual. It is not unreasonable to expect that the

temperature distribution u in the earth should satisfy the heat equation

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}.$$

As boundary conditions we take $u(0,t) = 0$ (meaning that the temperature at the surface is roughly constant) and we also use the initial condition $u(x,0) = f(x)$ to specify an initial distribution of temperature away from the earth's surface. In the usual separation of variable way, we then arrive at the two differential equations

$$\begin{aligned} T'(t) &= k\lambda T(t) \\ X''(x) &= \lambda X(x). \end{aligned}$$

This gives

$$u(x,t) = X(x)T(0)e^{k\lambda t}.$$

On physical grounds, we reject the possibility that $\lambda > 0$; we do not expect the temperature to increase as a function of time. Now let us concentrate our attention on the X-equation with its eigenvalue problem

$$X''(x) = \lambda X(x), \quad X(0) = 0.$$

Note that this problem is singular as the interval of definition is infinite: $[0,\infty)$. In Example 7.3.10–1 we ascertained that the problem is in the limit-point case at the singular endpoint. Therefore, we cannot use Theorem 7.3.13 to conclude that the problem is pseudo-regular. In fact, the problem is *not* pseudo-regular. Let us therefore proceed directly and see what happens. First let us consider the case when $\lambda = 0$. This gives $X(x) = ax + b$, and the boundary condition gives $b = 0$, thus leaving us with $X(x) = ax$ in this case. We reject this possibility on the physical grounds that it gives an unbounded temperature as $x \to \infty$. Thus we are left with $\lambda < 0$. For convenience, let us write $\lambda = -\omega^2$. The solution for the X-equation is then $X(x) = a\cos(\omega x) + b\sin(\omega x)$, and the boundary condition $X(0) = 0$ gives $b = 0$. These arguments, some of them a little hokey, give a typical solution to the X-equation as $X(x) = a\sin(\omega x)$. The question now is how we can determine ω. For other problems of this sort, we always had another boundary condition to invoke, and this ensured that we had a countable number of possibilities,, $\{\omega_n\}_{n\in\mathbb{N}}$, for ω. Then we wrote

$$u(x,t) = \sum_{n=1}^{\infty} a_n e^{-k\omega_n^2 t}\sin(\omega_n x). \tag{7.47}$$

If we were to impose a boundary condition $u(\ell,t) = 0$ for some $\ell > 0$ this would give $\omega_n = \frac{n\pi}{\ell}$. We see that as $\ell \to \infty$, the frequencies become closer and closer together, making us think that perhaps we can change the sum in (7.47) to an integral:

$$u(x,t) = \int_0^{\infty} e^{-k\omega^2 t}a(\omega)\sin(\omega x)\,\mathrm{d}\omega. \tag{7.48}$$

Just as in (7.47) we use the initial condition $u(x,0) = f(x)$ to determine the coefficients a_n, $n \in \mathbb{N}$, we want to somehow determine the unknown function $a(\omega)$ using this same initial condition. Setting $t = 0$ in (7.48) gives

$$u(x,0) = \int_0^\infty a(\omega)\sin(\omega x)\,\mathrm{d}\omega = f(x). \tag{7.49}$$

To "solve" this equation for $a(\omega)$ is not an easy matter, and is the subject of the theory of Fourier transforms. We shall not cover this subject here, but merely state that if f satisfies some conditions, then the equation (7.49) implies that

$$a(\omega) = \frac{2}{\pi}\int_0^\infty f(x)\sin(\omega x)\,\mathrm{d}x.$$

The function $a(\omega)$ so defined is known as the ***Fourier sine transform*** of f. Other sorts of Fourier transforms are available, including the ***Fourier cosine transform***

$$b(\omega) = \frac{2}{\pi}\int_0^\infty f(x)\cos(\omega x)\,\mathrm{d}x$$

and the plain ol' ***Fourier transform***

$$\hat{f}(\omega) = \int_{-\infty}^\infty f(x)e^{-i\omega x}\,\mathrm{d}x.$$

There subject of transform theory is one which, when given even the remotest degree of serious attention, will take up at least as much space as we have devoted to this point in our treatment of Fourier series and eigenvalue problems. Thus we can be forgiven for not saying too much about it at this point.

But back to the problem at hand. Using the Fourier sine series expression which we have been handed for $a(\omega)$, the solution to the boundary value problem is now

$$\begin{aligned} u(x,t) &= \frac{2}{\pi}\int_0^\infty e^{-k\omega^2 t}\Big(\int_0^\infty f(\xi)\sin(\omega\xi)\,\mathrm{d}\xi\Big)\sin(\omega x)\,\mathrm{d}\omega \\ &= \frac{1}{\pi}\int_0^\infty e^{-k\omega^2 t}\Big(\int_{-\infty}^\infty f(\xi)\cos\omega(x-\xi)\,\mathrm{d}\xi\Big)\,\mathrm{d}\omega, \end{aligned}$$

if we extend f to be defined on $\mathbb{R}$ by $f(x) = -f(-x)$ for $x < 0$ (thus we extend f by requiring it to be odd) and using the fact that cos is an even function. Now we swap the order of integration to give

$$u(x,t) = \frac{1}{\pi}\int_{-\infty}^\infty f(\xi)\Big(\int_0^\infty e^{-k\omega^2 t}\cos\omega(x-\xi)\,\mathrm{d}\omega\Big)\,\mathrm{d}\xi.$$

The inner integral is one that is determined by the well-known (or easily looked up) integral

$$\int_0^\infty e^{-\omega^2}\cos(a\omega)\,\mathrm{d}\omega = \frac{\sqrt{\pi}}{2}e^{-a^2/4}.$$

Using the appropriate modifications to this integral we obtain

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \int_{-\infty}^{\infty} f(\xi) \exp\Big(-\frac{(x-\xi)^2}{4kt}\Big)\, \mathrm{d}\xi.$$

Now we define a new variable s by $\xi = x + 2\sqrt{kt}s$. The change of variable formula then gives

$$u(x,t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x + 2\sqrt{kt}s) e^{-s^2}\, \mathrm{d}s.$$

Now let's make a specific choice for f and proceed further. We take

$$f(\xi) = \begin{cases} U_0, & \xi > 0 \\ -U_0, & \xi < 0 \\ 0, & \xi = 0, \end{cases}$$

keeping in mind our requirement that f be odd when extended to $\mathbb{R}$. We thus see that

$$f(x + 2\sqrt{kt}s) = \begin{cases} U_0, & s > -\frac{x}{2\sqrt{kt}} \\ -U_0, & s < -\frac{x}{2\sqrt{kt}} \\ 0, & s = -\frac{x}{2\sqrt{kt}}. \end{cases}$$

We then have

$$\begin{aligned} u(x,t) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x + 2\sqrt{kt}s) e^{-s^2}\, \mathrm{d}s \\ &= \frac{U_0}{\sqrt{\pi}} \int_{-\frac{x}{2\sqrt{kt}}}^{\frac{x}{2\sqrt{kt}}} e^{-s^2}\, \mathrm{d}s + \frac{U_0}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{kt}}}^{\infty} e^{-s^2}\, \mathrm{d}s - \frac{U_0}{\sqrt{\pi}} \int_{-\infty}^{-\frac{x}{2\sqrt{kt}}} e^{-s^2}\, \mathrm{d}s \\ &= \frac{2U_0}{\sqrt{\pi}} \int_{0}^{\frac{x}{2\sqrt{kt}}} e^{-s^2}\, \mathrm{d}s + \frac{U_0}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{kt}}}^{\infty} e^{-s^2}\, \mathrm{d}s - \frac{U_0}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{kt}}}^{\infty} e^{-s^2}\, \mathrm{d}s \\ &= \frac{2U_0}{\sqrt{\pi}} \int_{0}^{\frac{x}{2\sqrt{kt}}} e^{-s^2}\, \mathrm{d}s. \end{aligned}$$

In this way we have arrived at a nice compact formula which expresses the temperature distribution in the earth, given that the initial temperature distribution was uniform. We have made some shady moves in deriving this formula. However, these moves were all actually justified. For example, one can verify by direct computation that the function $u(x,t)$ does actually satisfy the heat equation and the boundary conditions.

Now, let us see how we can use this formula for $u(x,t)$ to ascertain the approximate age of the earth. The value of k can be determined empirically, and the value of U_0 can be taken roughly to be the temperature of molten lava, as this should

take on the temperature of the interior of a very old earth. Another piece of data that we can measure is the temperature gradient at the earth's surface. That is to say, we can obtain estimates for $\frac{\partial u}{\partial x}(0,T)$ where T is the "present" time. We can also directly compute this from the solution $u(x,t)$:

$$\frac{\partial u}{\partial x}(0,T) = \frac{U_0}{\sqrt{\pi k t}}.$$

From this we obtain

$$T = \frac{1}{\pi k}\Big(\frac{U_0}{\frac{\partial u}{\partial x}(0,T)}\Big)^2.$$

It is from this formula that Lord Kelvin obtained his estimate of between 4×10^8 and 10^9 years for the age of the earth. While this estimate is made via a number of approximations and rough estimates, the significant gap between it and theological speculations concerning the age of the earth was a cause of some discussion.

7.3.5 Summary

Generally speaking, singular eigenvalue problems are quite complex. As such, the extent to which one wishes to become expert in these problems is a matter of negotiation. Here is a guide to such.

1. One should be able to determine whether a given eigenvalue problem is regular or singular, and if singular, one should be able to ascertain which endpoints are singular.
2. One should understand the grounds on which the distinction between the limit-point and limit-circle cases is made, and be prepared to make this distinction, at least for problems where one can solve the ensuing differential equations for *some* value of the parameter λ.
3. Problems in the limit-circle case at all singular endpoints are nice because they are pseudo-regular, so there eigenvalues and eigenfunctions behave like those for regular problems.
4. Problems in the limit-point case at at least one singular endpoint are not guaranteed to be pseudo-regular, although they may be.
5. The notion of spectrum for problems in the limit-point case is not what you expect. The generalisation of the idea of spectrum from the regular case leads to the possibility in the singular case of there being points in the spectrum that are not eigenvalues. Thus for some singular problems, the distinction between the spectrum $\operatorname{spec}(L)$ and the point spectrum $\operatorname{spec}_0(L)$ becomes real.
6. Singular problems do come up in applications, so their study is merited. Many interesting "special functions," examples being presented here being the Bessel functions and the Legendre polynomials, arise as eigenfunctions for pseudo-regular singular eigenvalue problems.

7.3.6 Notes

Books have been written about Bessel functions [Tranter 1969, Watson 1995], so obviously we can only scratch the surface as concerns a discussion of their properties.

Exercises

7.3.1 Let $I \subseteq \mathbb{R}$ be an interval with $r\colon I \to \mathbb{R}$ a continuous positive function. Define the inner product $\langle \cdot, \cdot \rangle_r$ as in (7.34) and let $\mathrm{L}^r_2(I;\mathbb{C})$ be those functions with finite norm with respect to this inner product.

(a) Show that if I is closed and bounded then $\mathrm{L}^r_s(I;\mathbb{C}) = \mathrm{L}_2(I;\mathbb{C})$.
Hint: *Use the fact that continuous functions on compact sets attain their maximum and minimum.*

(b) Let $I = (0, 1]$ and let $r(x) = x$. Show that $\mathrm{L}^r_2(I;\mathbb{C}) \neq \mathrm{L}_2(I;\mathbb{C})$.

7.3.2 Suppose that the I is bounded with endpoints $e_1 < e_2$, and suppose that the problem data is periodic so that

$$\lim_{x\to e_1} p(e_1) = \lim_{x\to e_2} p(e_2), \quad \lim_{x\to e_1} q(e_1) = \lim_{x\to e_2} q(e_2), \quad \lim_{x\to e_1} r(e_1) = \lim_{x\to e_2} r(e_2).$$

Show that if we impose periodic boundary conditions

$$y(a) = y(b), \quad y'(a) = y'(b),$$

then the eigenvalues for the boundary value problem are real.
Hint: *Suitably modify the proof of Proposition 7.3.11.*

7.3.3 Show that the eigenvalues of the problem

$$(xy')' = \lambda xy, \quad y(b) = 0$$

that yield bounded eigenfunctions are strictly positive.

7.3.4 In Exercise 6.4.3 you determined that the horizontal deflections of a string dangling vertically are governed by the partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = g\frac{\partial}{\partial x}\Big(x\frac{\partial u}{\partial x}\Big),$$

for $t \geq 0$ and $x \in [0, \ell]$. The problem has the single natural boundary condition $u(\ell, t) = 0$.

(a) Use separation of variables to arrive at an eigenvalue problem in x.

(b) Is the eigenvalue problem in part (a) regular or singular?

(c) Find an algebraic equation that governs the location of the eigenvalues, and provide the form of the eigenfunctions.
Hint: *Make a change of independent variable* $\xi = 2\sqrt{-\lambda}x$.

(d) Why is a second boundary condition not necessary?

7.3.5 Consider the eigenvalue problem you derived in part (a) of Exercise 7.2.15.

(a) Is the eigenvalue problem regular or singular?
(b) Provide an algebraic equation which is satisfied by the eigenvalues.
(c) Determine the form of the eigenfunctions.
(d) If $T\colon [R_0, R_1] \to \mathbb{R}$ is an arbitrary radially symmetric temperature distribution across the walls of the heat exchanger tube, express T as a linear combination of eigenfunctions.

7.3.6 Consider the ***Hermite equation***[11]

$$y'' - x^2 y = \lambda y$$

defined on $(-\infty, \infty)$.

(a) What are p, q, and r?
(b) Is the problem regular or singular? Why?
(c) Make the change of variable $z = e^{x^2/2} y$ and show that z satisfies the differential equation
$$z'' - 2xz' - z = \lambda z.$$
(d) Show that the differential equation from part (c) has two linearly independent solutions
$$z_1(x) = 1, \quad z_2(x) = \int_0^x e^{\xi^2}\, \mathrm{d}\xi$$
for some value of λ (part of the question is to determine *which* value of λ).
(e) Determine whether the Hermite equation is in the limit-point or the limit-circle case at its two endpoints.
(f) Can you deduce whether the problem is pseudo-regular?

Consider the vibrating drumhead problem discussed in Section 7.3.4.1. The problem was only partly solved in that section. Here we will finish the problem.

7.3.7 Denote by $\{\lambda_n\}_{n\in\mathbb{N}}$ and $\{\phi_n\}_{n\in\mathbb{N}}$ the eigenvectors and normalised eigenfunctions as determined from the Bessel eigenvalue problem yielded by the "R-equation" in the analysis of Section 7.3.4.1. Suppose that the drumhead has an initial vertical deflection given by

$$u(r, 0) = f(r), \quad \frac{\partial u}{\partial t}(r, 0) = 0.$$

Give an expression for the displacement $u(r, t)$ of the drumhead for all $r \in (0, b]$ and for all $t > 0$.

[11] After Charles Hermite (1822–1901).

7.3.8 Consider the Laplacian in three-dimensions.

(a) Verify that the Laplacian in spherical coordinates is as given in (7.45).

(b) Verify that for a function that is independent of "longitude" ϕ, the introduction of the new coordinate $s = \cos\theta$ gives the partial differential equation of (7.46)

7.3.9 Let $\{\lambda_n\}_{n\in\mathbb{N}_0}$ be the eigenvalues for the Legendre equation with $\{P_n\}_{n\in\mathbb{N}_0}$ the eigenfunctions as defined in Section 7.3.4.2.

(a) Show that for each $n \in \mathbb{N}_0$, P_n is an eigenfunction for the Legendre equation with eigenvalue λ_n.

(b) Show that the polynomials P_n, $n \in \mathbb{N}_0$, satisfy the recursion relation

$$(n+1)P_{n+1}(s) = (2n+1)sP_n(s) - nP_{n-1}(s), \quad n \in \mathbb{N}.$$

Hint: *Show that*

$$\mathrm{P_{n+1}(s) - P_{n-1}(s) = \frac{2n+1}{2^n n!}\frac{d^{n-1}}{ds^{n-1}}(s^2-1)^n}$$
$$\mathrm{P_{n+1}(s) - sP_n(s) = \frac{n}{2^n n!}\frac{d^{n-1}}{ds^{n-1}}(s^2-1)^n},$$

using Rodrigues' formula.

(c) Show that

$$P'_{n+1}(s) - P'_{n-1}(s) = (2n+1)P_n(s), \quad n \in \mathbb{N}.$$

(d) Use part (c) to deduce

$$(2n+1)\int_s^1 P_n(\sigma)\,d\sigma = P_{n-1}(s) - P_{n+1}(s).$$

(e) Now conclude that

$$(2n+1)\int_{-1}^1 P_n^2(s)\,ds = \int_{-1}^1 P_n(s)P'_{n+1}(s)\,ds = P_n(s)P_{n+1}(s)\Big|_{-1}^1 = 2.$$

Hint: *Use the fact that* $\mathrm{P_n}$ *is orthogonal to* $\mathrm{P_m}$ *for any* $\mathrm{m < n}$.

(f) Determine the constants c_n, $n \in \mathbb{N}_0$, so that the eigenfunctions $\{\phi_n = c_nP_n\}_{n\in\mathbb{N}_0}$ are orthonormal.

7.3.10 Change standard problem on $[0,\infty)$ to one on $(0,1]$.***missing stuff***

Bibliography

Anderson, B. D. O. [1972] *The reduced Hermite criterion with application to proof of the Liénard–Chipart criterion*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **17**(5), pages 669–672, ISSN: 0018-9286, DOI: 10.1109/TAC.1972.1100142.

Anderson, B. D. O., Jury, E. I., and Mansour, M. [1987] *On robust Hurwitz polynomials*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **32**(10), pages 909–913, ISSN: 0018-9286, DOI: 10.1109/TAC.1987.1104459.

Bacciotti, A. and Rosier, L. [2005] *Liapunov Functions and Stability in Control Theory*, ISBN: 978-3-540-21332-1.

Barbashin, E. A. and Krasovskiĭ, N. N. [1952] *On global stability of motion*, Rossiĭskaya Akademiya Nauk. Doklady Akademii Nauk, **86**(3), pages 453–456, ISSN: 0869-5652.

Brown, C. M. [2007] *Differential Equations, A Modeling Approach*, number 150 in Quantitative Applications in the Social Sciences, SAGE Publications: Los Angeles/London/New Delhi/Singapore, ISBN: 978-1-4129-4108-2.

Chapellat, H. and Bhattacharyya, S. P. [1989] *An alternative proof of Kharitonov's theorem*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **34**(4), pages 448–450, ISSN: 0018-9286, DOI: 10.1109/9.28021.

Coddington, E. E. and Levinson, N. [1955] *Theory of Ordinary Differential Equations*, McGraw-Hill: New York, NY, New edition: [Coddington and Levinson 1984].

— [1984] *Theory of Ordinary Differential Equations*, 8th edition, Robert E. Krieger Publishing Company: Huntington/New York, ISBN: 978-0-89874-755-3, First edition: [Coddington and Levinson 1955].

Dasgupta, S. [1988] *Kharitonov's theorem revisited*, Systems & Control Letters, **11**(5), pages 381–384, ISSN: 0167-6911, DOI: 10.1016/0167-6911(88)90096-5.

Fujiwara, M. [1915] *Über die Wurzeln der algebraischen Gleichungen*, The Tôhoku Mathematical Journal. Second Series, **8**, pages 78–85, ISSN: 0040-8735.

Gantmacher, F. R. [1959] *The Theory of Matrices*, translated by K. A. Hirsch, volume 2, Chelsea: New York, NY, Reprint: [Gantmacher 2000].

— [2000] *The Theory of Matrices*, translated by K. A. Hirsch, volume 2, American Mathematical Society: Providence, RI, ISBN: 978-0-8218-2664-5, Original: [Gantmacher 1959].

Hermite, C. [1854] *Sur le nombre des racines d'une équation algèbrique comprise entre des limites données*, Journal für die Reine und Angewandte Mathematik, **52**, pages 39–51, ISSN: 0075-4102.

Hurwitz, A. [1895] *Über di Bedingungen unter welchen eine Gleichung nur Wurzeln mit negativen reellen Teilen besitz*, Mathematische Annalen, **46**, pages 273–284, ISSN: 0025-5831, URL: https://eudml.org/doc/157760 (visited on 07/11/2014).

Kellett, C. M. [2014] *A compendium of comparison function results*, Mathematics of Control, Signals, and Systems, **26**(3), pages 339–374, ISSN: 0932-4194.

Kharitonov, V. L. [1978] *Asymptotic stability of an equilibrium position of a family of systems of linear differential equations*, Differentsial'nye Uravneniya, **14**, pages 2086–2088, ISSN: 0374-0641.

LaSalle, J. P. [1968] *Stability theory for ordinary differential equations*, Journal of Differential Equations, **4**(1), pages 57–65, ISSN: 0022-0396, DOI: 10.1016/0022-0396(68)90048-X.

Liapunov, A. M. [1893] *A special case of the problem of stability of motion*, Rossiĭskaya Akademiya Nauk. Matematicheskiĭ Sbornik, **17**, pages 252–333, ISSN: 0368-8666.

Liénard, A. and Chipart, M. [1914] *Sur la signe de la partie réelle des racines d'une équation algébrique*, Journal de Mathématiques Pures et Appliquées. Neuvième Série, **10**(6), pages 291–346, ISSN: 0021-7824.

Mansour, M. and Anderson, B. D. O. [1993] *Kharitonov's theorem and the second method of Lyapunov*, Systems & Control Letters, **20**(3), pages 39–47, ISSN: 0167-6911, DOI: 10.1016/0167-6911(93)90085-K.

Maxwell, J. C. [1868] *On governors*, Proceedings of the Royal Society. London. Series A. Mathematical and Physical Sciences, **16**, pages 270–283, ISSN: 1364-5021, URL: http://www.jstor.org/stable/112510 (visited on 07/10/2014).

Minnichelli, R. J., Anagnost, J. J., and Desoer, C. A. [1989] *An elementary proof of Kharitonov's stability theorem with extensions*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **34**(9), pages 995–998, ISSN: 0018-9286, DOI: 10.1109/9.35816.

Parks, P. C. [1962] *A new proof of the Routh–Hurwitz stability criterion using the second method of Liapunov*, Proceedings of the Cambridge Philosophical Society, **58**(4), pages 694–702, DOI: 10.1017/S030500410004072X.

Routh, E. J. [1877] *A Treatise on the Stability of a Given State of Motion*, Adam's Prize Essay, Cambrige University.

Stakgold, I. [1967] *Boundary Value Problems of Mathematical Physics*, 2 volumes, Macmillian: New York, NY, Reprint: [Stakgold 2000].

— [1979] *Green's Functions and Boundary Value Problems*, Pure and Applied Mathematics, Dekker Marcel Dekker: New York, NY, ISBN: 0-471-81967-0, New edition: [Stakgold 2000].

— [2000] *Boundary Value Problems of Mathematical Physics*, 2 volumes, 29 Classics in Applied Mathematics, Society for Industrial and Applied Mathematics: Philadelphia, PA, ISBN: 978-0-89871-456-2, Original: [Stakgold 1967].

Tranter, C. J. [1969] *Bessel Functions with Some Physical Applications*, Hart Publishing Company: New York, NY, ISBN: 978-0-340-04959-4.

Troutman, J. L. [1994] *Boundary Value Problems of Applied Mathematics*, PWS Publishing Company: Boston, MA, ISBN: 978-0-534-19116-0.

Watson, G. N. [1922] *A Treatise on the Theory of Bessel Functions*, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, New edition: [Watson 1995].

— [1995] *A Treatise on the Theory of Bessel Functions*, 2nd edition, Cambridge Mathematical Library, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, ISBN: 978-0-521-48391-9, First edition: [Watson 1922].