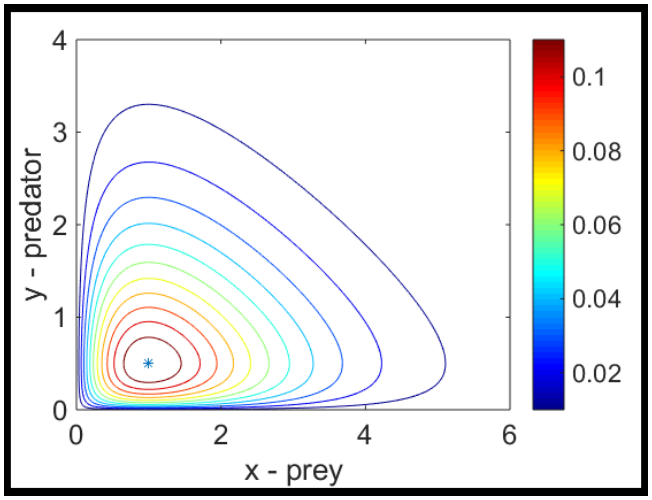
University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

School of Mathematics and Statistics
Scoil na Matamaitice agus na Staitisticí

Uncertainty Quantification (ACM 41000)



Dr Lennon Ó Náraigh – ODEs / PDEs
Dr Michelle Carey – Statistical Inference

Lecture notes in Uncertainty Quantification (ODEs / PDEs) January 2018

Mathematical Modelling in the Sciences (ACM10070)

- Subject: Applied and Computational Mathematics
- School: School of Mathematics and Statistics
- Module coordinator: Dr Lennon Ó Náraigh
- Lecturers: Dr Lennon Ó Náraigh (Differential Equations) and Dr Michelle Carey (Statistical Inference)
- Credits: 5
- Level: 4
- Semester: Second

This module is a synthesis of modern Applied Mathematics and Statistical Inference, and introduces students to statistical methods to determine model parameters in otherwise deterministic mathematical models. The module starts with a review of deterministic models - ordinary and partial differential equations (both linear and nonlinear). Included in this review is a topical summary of nonlinear ordinary differential equations, reaction-diffusion partial differential equations, and optimization of partial-differential equation models using an adjoint method. Students will learn how to use these equation systems to model physical systems. In doing so, various model parameters appear, which in turn need to be modelled. These parameters can be determined by reference to experimental data, which can then be used to make predictions. As such, in the second part of the module students will learn how to estimate and attain confidence limits for the parameters of the differential equations from noisy and often partially observed data.

What will I learn?

On completion of this module students should be able to

1. Solve a variety of linear and nonlinear ordinary differential equation systems
2. Do the same for partial differential equation models
3. Apply the adjoint method to optimization problems involving parabolic partial differential equations
4. Apply such systems of equations in modelling physical systems
5. Understand the origin of various model parameters when such equation systems are used as mathematical models of various physical systems

6. Attain statistical inference for the parameters of linear and nonlinear ordinary differential equation systems.

7. Do the same for partial differential equation models.

Editions

First edition: January 2018

Acknowledgements

- Chapter 1 is based on material from the reference text by Stephen Strogatz [Str01].
- Chapter 2 is from my head.
- Chapter 3 is now left blank.
- Chapter 4 is based on material by Dr Miguel Bustamante.
- Chapters 5-7 are from my head.
- Chapter 8 is again based on the book by Strogatz [Str01].
- Chapter 9 is from the reference text by Pavliotis and Stuart [PS08].
- Chapter 10 is from the reference text by Guenther and Lee [GL96].
- Chapter 11 is again from the reference text by Pavliotis and Stuart [PS08].
- The remaining chapters are from my head.

Contents

Chapter 0

Introduction

0.1 Outline

Here is an executive summary of the aims of this course. If you cannot remember the more detailed outline that follows, at least keep the following in mind:

- We will review the theory of ordinary and partial differential equations – both linear and nonlinear.
- We will apply this theory to model various physical scenarios.
- In so doing, we will encounter model parameters that are not known *a priori*.
- We will use statistical inference to estimate these parameters, thereby formulating complete models.

Throughout this module, and indeed, throughout the mathematical world, the term 'Ordinary Differential Equations' is abbreviated as 'ODEs', and similarly, 'Partial Differential Equations' is abbreviated as 'PDEs'. This convention is assumed hereafter.

0.2 The Structure of the Module

Learning and Assessment

- Twenty four classes, two per week.
- Assessment is by way of four mini-projects, each counting for 25% of the final grade:
  - Mini-project on theoretical and numerical aspects of ODEs

  - Mini-project on theoretical and numerical aspects of PDEs
  - Mini-project on parameter estimation for ODEs
  - Mini-project on parameter estimation for PDEs

  Each project shall be written in Latex, in the form of a short report, with proper structure, equation annotation, argumentation, and bibliography.

Two groups, one module

This module has a very diverse mix of students of different backgrounds. The module is designed with all students in mind. As such, in weeks 1-6 the class will split in two:

- One group will look at more elementary material, including an overview of ODEs and PDEs
- For the other group, this material will be assumed known, and more advanced topics will be covered – corresponding chapters herein are marked with an asterisk (*)

At the end of Week 6, the two groups will merge back into one for the second part of the module on Statistical Inference, at which point both groups will be will prepared for the remainder of the module. Students are encouraged to participate in the module at the level of their abilities and interests, and not just to revert to the elementary group because it is 'easier'. Also, as the module is assessed by project work alone, all students will be assessed equally.

Textbooks

- Lecture notes will be put on the web. These are self-contained.
- For the ODEs/PDEs part of the module, here are some books for extra reading:
  - *Nonlinear dynamics and chaos*, S. Strogatz, Westview Press (2000) – Reference [Str01].
  - *Multiscale Methods: Averaging and Homgenization*, G. A. Pavliotis and A. M. Stuart, Sprinter (2008) – Reference [PS08].
  - *Partial Differential Equations of Mathematical Physics and Integral Equations*, R. B. Guenther and J. W. Lee (1996) – Reference [GL96].
  - *Mathematical Methods for Physics and Engineering*, K. F. Riley, M. P. Hobson, S. J. Bence (2006).

0.3 Policies

Policy on extenuating circumstances

This module adheres to the offical UCD policy regarding extenuating circumstances:

- Serious issues (serious illness, hospitalization, etc.) are dealt with through a formal process administered by the College of Science;
- Minor issues (e.g. assignments missed due to minor illness) are dealt with through direct contact with the lecturer – via email, with supporting documentation as necessary. In order to keep track of such extenuating circumstances, please use the phase 'minor extenuating circumstances ACM41000 in the subject line in the email.

Plagiarism

Plagiarism is a **serious academic offence**. While plagiarism may be easy to commit unintentionally, it is defined by the act not the intention.

- All students are responsible for being familiar with the Universitys policy statement on plagiarism and are encouraged, if in doubt, to seek guidance from an academic member of staff.
- The University encourages students to adopt good academic practice by maintaining academic integrity in the presentation of all academic work.
- For more detailed information see:

  `http://www.ucd.ie/governance/resources/policypage-plagiarismpolicy/`

Chapter 1

Overview of First-Order ODEs

1.1 Outline

We review the theory of first-order ODEs. We first of all define an ODE. We then introduce a taxonomy of first-order ODEs – elementary, separable, autonomous, and linear. In this way, we can begin to develop a set of powerful tools to enable us to solve all of the first-order ODEs that are of practical importance in applications.

1.2 ODE – the definition

Definition 1.1 *Let $y = y(x)$ be a function. A* **first-order ordinary-differential equation** *is a relationship of the form*

$$F\left(x, y, \frac{dy}{dx}\right) = 0,$$

where $F(\cdot, \cdot, \cdot)$ is some function of three variables.

This is a very general definition, almost too general. The ODEs we encounter in this module are all of the form

$$\boxed{\frac{dy}{dx} = f(x, y),} \tag{1.1}$$

where $f(\cdot, \cdot)$ is a function of two variables.

Similarly, there are second-order ODEs:

Definition 1.2 *Let $y = y(x)$ be a function. A* **second-order ordinary-differential equation** *is a relationship of the form*

$$F\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0,$$

where $F(\cdot,\cdot,\cdot,\cdot)$ is some function of four variables.

Again, the second-order ODEs in this module are of the form

$$\boxed{\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right),} \tag{1.2}$$

where now $f(\cdot,\cdot,\cdot)$ is a function of three variables. Of course, it is possible to have an n^{th}-order ODE, but we will only go up to second-order in this module.

1.3 Elementary ODES

Definition 1.3 *An ODE of the form*

$$\frac{dy}{dx} = f(x) \tag{1.3}$$

is called **elementary**. *In other words, the RHS is a function of one variable only – x.*

Theorem 1.1 *All elementary ODEs have a solution.*

Proof: Let us integrate both sides of Eq. (1.3):

$$\int \frac{dy}{dx}\mathrm{d}x = \int f(x)\,\mathrm{d}x.$$

We can use a change-of-variables on the left-hand side:

$$y = y(x), \qquad \mathrm{d}y = \frac{dy}{dx}\mathrm{d}x.$$

Hence, the integral becomes

$$\int \mathrm{d}y = \int f(x)\,\mathrm{d}x,$$

or

$$y(x) = \int f(x)\,\mathrm{d}x + C,$$

where C is a **constant of integration**. ∎

Example: Solve the ODE

$$\frac{dy}{dx} = \mathrm{e}^x + 2.$$

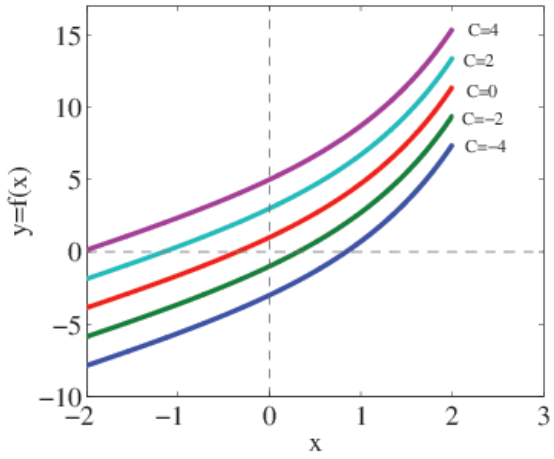


Figure 1.1: Some members of the family of solution curves for $dy/dx = \mathrm{e}^x + 2$

.

A quick shorthand for the change-of-variables argument is **formally** to multiply both sides by $\mathrm{d}x$. This yields the expression

$$\mathrm{d}y = (\mathrm{e}^x + 2)\,\mathrm{d}x.$$

Now, we integrate:

$$\int \mathrm{d}y = \int (\mathrm{e}^x + 2)\,\mathrm{d}x,$$

or

$$y = \mathrm{e}^x + 2x + C,$$

where C is a constant of integration.

In this example, we can replace the arbitrary constant C by any real number and the solution is still valid. Thus, $y = \mathrm{e}^x + 2x + C$ is called the **general solution**. Letting C range over all real numbers gives the **family of solution curves**. Geometrically, these are a set of curves, where each curve is the same, up to a constant value (Fig. 1.1).

In many applications, we will be given an extra piece of information of the form

$$y(x = a) = b.$$

This fixes a value of the constant of integration and brings us from the general solution to a **particular solution**. In general, for a first-order ODE, one such condition is needed, for a second-order ODE, two such conditions are needed, etc.

1.4 Separation of variables

Definition 1.4 *An ODE of the form $dy/dx = f(x, y)$ is called* **separable** *if it can be re-written in the form*

$$\frac{dy}{dx} = g(x)h(y). \tag{1.4}$$

Examples of separable RHSs:

1. $\cos x \sin y$
2. $\mathrm{e}^x y^2$

Examples of **non-separable** RHSs:

1. $x - y$
2. $\sin(xy)$

We now focus on solving Eq. (1.4). First, divide both sides by $h(y)$:

$$\frac{1}{h(y)}\frac{dy}{dx} = g(x).$$

Now, integrate both sides with respect to (wrt) x:

$$\int \frac{1}{h(y)}\frac{dy}{dx}\mathrm{d}x = \int g(x)\,\mathrm{d}x.$$

We can use a change-of-variables on the left-hand side:

$$y = y(x), \qquad \mathrm{d}y = \frac{dy}{dx}\mathrm{d}x.$$

Hence, the integral becomes

$$\boxed{\int \frac{\mathrm{d}y}{h(y)} = \int g(x)\,\mathrm{d}x.} \tag{1.5}$$

Thus, solving the ODE (1.4) (hard) is reduced to doing an integral (easy).

Example: Solve the ODE

$$\frac{dy}{dx} = \mathrm{e}^x y^2.$$

This is a separable problem. Again, as a shorthand for the change-of-variables argument, we **formally** multiply both sides by $\mathrm{d}x$:

$$\mathrm{d}y = \mathrm{e}^x y^2 \,\mathrm{d}x.$$

We divide both sides by y^2:

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

Now we integrate:

$$\int y^{-2}\mathrm{d}y = \int \mathrm{e}^x \,\mathrm{d}x.$$

The solution is

$$-\frac{1}{y} = \mathrm{e}^x + C,$$

where C is a constant. Finally, we solve for y:

$$y = -\frac{1}{C + \mathrm{e}^x}.$$

1.5 Autonomous ODEs – the definition

Definition 1.5 *A first-order ODE of the form*

$$\frac{dy}{dx} = f(y)$$

is called an **autonomous** *ODE.*

This is a useful case to consider, and is broader than what initially appears to be the case, as all separable ODEs can be brought into autonomous form, as the following theorem demonstrates:

Theorem 1.2 *A general separable ODE,*

$$\frac{dy}{dx} = g(x)h(y)$$

can always be made into an autonomous ODE by the change of variable $\eta = \int g(x)\mathrm{d}x$.

Proof: Divide the ODE $dy/dx = g(x)h(y)$ by $g(x)$ to obtain

$$\frac{1}{g(x)}\frac{dy}{dx} = h(y).$$

Define a new variable $\eta = \int g(x)\mathrm{d}x$ and let y be a function of η: $y \to y(\eta)$. Then,

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{d\eta}\frac{d\eta}{dx}, \\ \frac{dy}{dx} &= \frac{dy}{d\eta}g(x), \\ \frac{1}{g(x)}\frac{dy}{dx} &= \frac{dy}{d\eta}, \\ h(y) &= \frac{dy}{d\eta}. \end{aligned}$$ ∎

Thus, although this section deals with autonomous ODEs only, you should remind yourself that this discussion extends easily to separable ODEs.

1.6 Autonomous ODEs – qualitative behaviour

Autonomous ODEs are very special because their behaviour can be fully determined in the absence of an explicit solution – starting with the idea of fixed points:

Definition 1.6 (Fixed points of autonomous ODEs) *Let*

$$\frac{dy}{dx} = f(y)$$

be an autonomous ODE. The **fixed points** *of the ODE are those constant values of y for which*

$$f(y) = 0.$$

Example: Find the fixed points of the autonomous ODE

$$\frac{dy}{dx} = y^2 - y - 12.$$

The fixed points correspond to $y^2 - y - 12 = 0$. Using the quadratic formula, these are

$$y_1 = \frac{+1 - \sqrt{1 - 4(-12)}}{2} = -3, \qquad y_2 = \frac{+1 + \sqrt{1 - 4(-12)}}{2} = 4.$$

These fixed points are shown on Figure 1.2. We further use this information to construct the so-called **one-dimensional vector field** associated with the ODE:

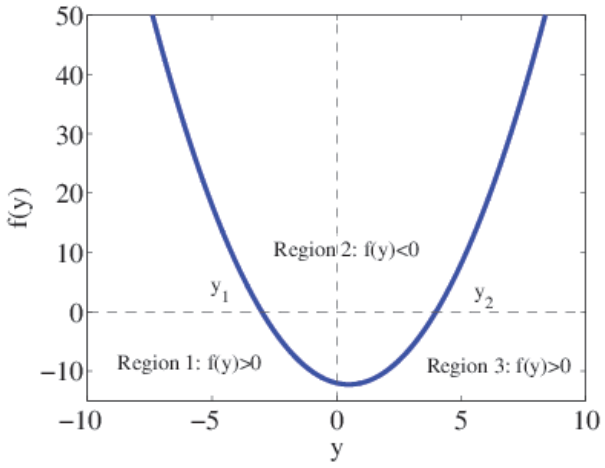


Figure 1.2: RHS of ODE $dy/dx = y^2 - y - 12$.

- If the system starts in region 1, then $dy/dx > 0$, and y will tend to increase. Therefore, the flow in this region is towards $y = y_1$.

- If a system starts at the fixed point $y = y_1$, then it will stay there for all time.

- If the system starts in region 2, then $dy/dx < 0$, and y will tend to decrease. Therefore, the flow in this region is towards $y = y_1$, the lower fixed point.

- If a system starts at the fixed point $y = y_2$, then it will stay there for all time.

- If the system starts in region 3, then $dy/dx > 0$, and y will tend to increase. Therefore, the flow in this region is towards $y = \infty$.

The vector field is a set of arrows on the real line showing the direction of the flow of the solution – e.g. Figure 1.3.

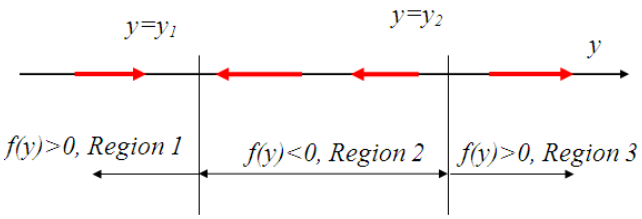


Figure 1.3: One-dimensional vector field for $dy/dx = y^2 - y - 12$.

1.7 Stability theory – qualitative

Consider again a generic autonomous ODE $dy/dx = f(y)$. Recall, the fixed points of the equation are those constant values of y for which

$$f(y) = 0.$$

In other words, the fixed points are the roots of the equation $f(y) = 0$.

Definition 1.7 (Characterization of fixed points as stable, unstable or as a node) *Let y_0 be a fixed point of $dy/dx = f(y)$.*

- *We call y_0 stable if all the arrows in the one-dimensional vector field associated with the ODE flow into y_0.*
- *We call y_0 unstable if all the arrows in the one-dimensional vector field flow away from y_0.*
- *A fixed point that is neither stable nor unstable is called a* **node**.

Example: We look again at the ODE

$$\frac{dy}{dx} = y^2 - y - 12.$$

Recall the fixed points:

$$y_1 = \frac{+1 - \sqrt{1 - 4(-12)}}{2} = -3, \qquad y_2 = \frac{+1 + \sqrt{1 - 4(-12)}}{2} = 4.$$

The one-dimensional vector field is shown in Figure 1.4. The fixed point y_1 is stable and the fixed

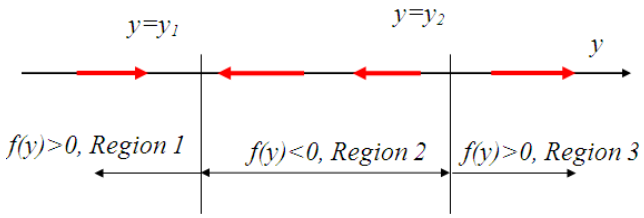


Figure 1.4: One-dimensional vector field for $dy/dx = y^2 - y - 12$.

point y_2 is unstable.

Stability theory

There is good reason for the nomenclature concerning stable and unstable equilibria. For, consider a solution $y(x) = y_* + \delta(x)$ to the ODE $dy/dx = f(y)$, where y_* is a fixed point and $f(y_*) = 0$. We have

$$\text{LHS} = \frac{dy}{dx} = \frac{d}{dx}\left[y_* + \delta(x)\right] = \frac{d}{dx}\delta(x)$$

since y_* is a constant. The RHS is approximated using the best straight-line approximation:

$$\text{RHS} = f(y) = f(y_* + \delta) = \underbrace{\cancel{f(y_*)}}_{f(y_*)=0} + f'(y_*)\delta + O(\delta^2). \tag{1.6}$$

Set LHS = RHS to obtain

$$\frac{d}{dx}\delta(x) = f'(y_*)\delta.$$

This is a separable ODE; we separate the variables to get

$$\frac{\mathrm{d}\delta}{\delta} = \underbrace{f'(y_*)}_{=\text{Const.}}\,\mathrm{d}x.$$

Integrate both sides:

$$\int \frac{\mathrm{d}\delta}{\delta} = \int f'(y_*)\mathrm{d}x + C = f'(y_*)x + C.$$

The LHS can be integrated using standard integration formulae:

$$\ln\delta = f'(y_*)x + C.$$

Exponentiate:

$$\delta(x) = \underbrace{D}_{=\mathrm{e}^C}\mathrm{e}^{f'(y_*)x}.$$

If $\delta(x = 0) = \delta_0$, then $D = \delta_0$ and the solution is

$$\delta(x) = \delta_0\mathrm{e}^{f'(y_*)x}.$$

- If the eqiulibrium is stable, then $f'(y_*) < 0$ and the small disturbance $\delta(x)$ decays exponentially. But $y(x) = y_* + \delta(x)$, hence
$$\lim_{x\to\infty} y(x) = y_*.$$
- If the equilibrium is unstable, then $f'(y_*) > 0$, and the small disturbance $\delta(x)$ increases exponentially. In practice, the small-amplitude approximation (1.6) would break down at some point; in any case, the solution $y(x)$ moves away from the fixed point as $x \to \infty$.

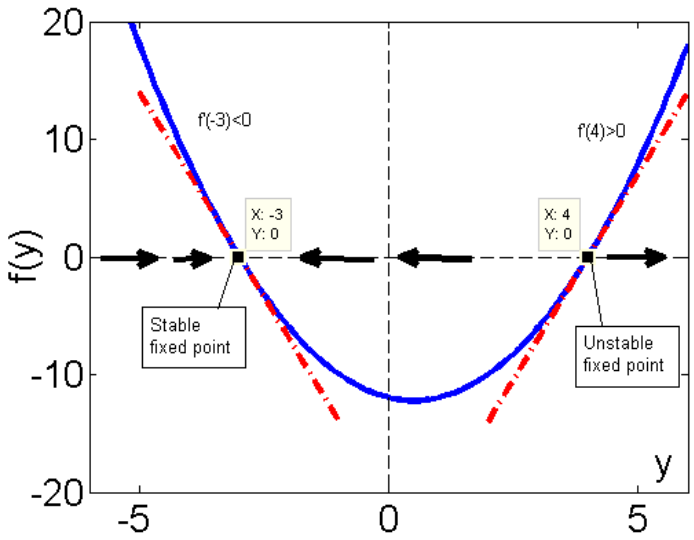


Figure 1.5: More detailed description of the fixed points of $dy/dx = y^2 - y - 12$ showing the stable (y_1) and unstable (y_2) fixed points and the corresponding signs of $f'(y_1)$ and $f'(y_2)$.

Example: Look again at the ODE

$$\frac{dy}{dx} = y^2 - y - 12$$

and comment on $f'(y_*)$ at the different fixed points $y_1 = -3$ and $y_2 = 4$.

Solution: We have $f'(y) = 2y - 1$, and

$$f'(y_1) = 2(-3) - 1 = -7, \qquad f'(y_2) = 2(4) - 1 = 7.$$

By inspection of the one-dimensional vector field (see Figure 1.5 for a more detailed description), y_1 is stable, corresponding to $f'(y_1) < 0$, while $f(y_2)$ is unstable, corresponding to $f'(y_2) > 0$.

Bifurcations

Bifurcation analysis is concerned with the autonomous ODEs of the form $dy/dx = f(y; k)$, where y is (as usual) the dependent variable but there is an added complication now – the apperance of a paramter k. The character of the fixed points can change as the parameter k is varied. As such, consider the autonomous ODE

$$\frac{dy}{dx} = y^2 - k,$$

where k is a parameter. The fixed points are given by

$$f(y;k) = 0, \text{ or } y^2 - k = 0.$$

Thus, the fixed points are

$$y_0 = \pm\sqrt{k}.$$

However, k is an adjustable parameter. If $k < 0$ there are no real fixed points. If $k = 0$ there is only one fixed point. If $k > 0$ there are two fixed points, and as k increases, these fixed points move further and further apart. We can plot the fixed points as a function of k on a graph – called a **bifurcation diagram**. This is shown in Figure 1.6.

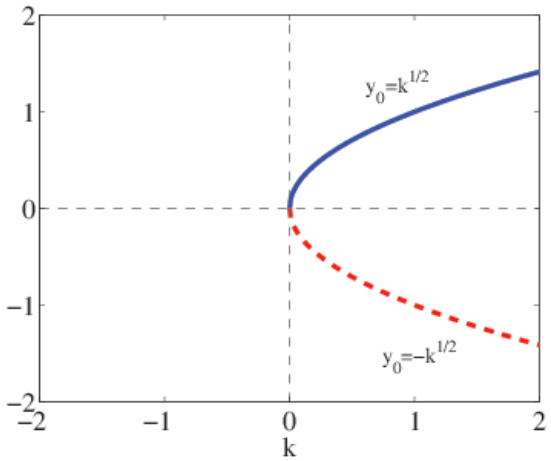


Figure 1.6: Bifurcation diagram for $dy/dx = y^2 - k$.

Example: Consider the ODE

$$\frac{dy}{dx} = y^2 - ky + 1, \qquad k \in \mathbb{R}.$$

Draw the bifurcation curve and say whether the fixed points are stable or unstable.

The fixed points are given by the roots of the quadratic equation

$$f(y;k) = 0, \qquad y^2 - ky + 1 = 0,$$

or

$$y_* = \frac{k \pm \sqrt{k^2 - 4}}{2}.$$

We gain information from the sign of $k^2 - 4$:

- For $k^2 > 4$ there are two fixed points.
- For $k^2 < 4$ there are no fixed points.

Caution is needed here, as the criterion for the existence of fixed points is $|k| > 2$, hence $k < -2$ or $k > 2$. The bifurcation curve is shown in Figure 1.7. The upper branch (solid line) corresponds to

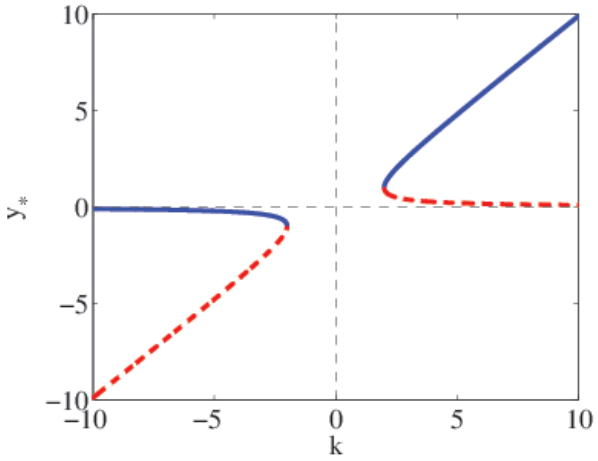


Figure 1.7: Bifurcation diagram for $dy/dx = y^2 - ky + 1$.

the positive sign in the quadratic equation for the fixed points. Next, we determine whether the fixed points are stable or unstable by taking two representative k-values and plotting the one-dimensional vector field for these values. The results show that the lower fixed point is stable and the upper one is unstable.

1.8 First-order linear ODEs

In this section we encounter a third and final functional form for which a simple analytical solution is available: the famous **linear first-order ODE**:

$$\frac{dy}{dx} + P(x)y = Q(x),$$

for which the **integrating-factor technique** is available.

Consider an ODE of the form

$$\boxed{\frac{dy}{dx} + P(x)y = Q(x),} \tag{1.7}$$

where $P(x)$ and $Q(x)$ are continuous functions. Then a solution $y(x)$ can be found, through the following algorithm:

1. Compute the **integrating factor**

$$\mu(x) = \mathrm{e}^{\int P(x)\mathrm{d}x}.$$

2. Multiply both sides of the basic equation (1.7) by $\mu(x)$ to obtain

$$\mu\left[\frac{dy}{dx} + P(x)y\right] = \mu Q.$$

3. Observe that the LHS of this new equation is a **perfect derivative** and can in fact be re-written as

$$\frac{d}{dx}\left[\mu y\right] = \mu Q.$$

4. Integrate once to obtain

$$\mu y = C + \int \mu(x)Q(x)\,\mathrm{d}x,$$

where C is an arbitrary constant of integration.

5. Re-arrange to find the general solution:

$$\boxed{y(x) = \frac{C}{\mu(x)} + \frac{1}{\mu(x)}\int \mu(x)Q(x)\,\mathrm{d}x.} \tag{1.8}$$

Example: Solve the ODE

$$\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{1}{x}y = \sin(x).$$

Solution: Identify $P = 1/x$ and $Q = \sin x$, hence

$$\mu = \mathrm{e}^{\int P(x)\,\mathrm{d}x} = \mathrm{e}^{\int (1/x)\mathrm{d}x} = \mathrm{e}^{\ln x} = x,$$

hence

$$\text{Integrating Factor} = \mu = x.$$

Multiply both sides of the ODE by the integrating factor to obtain

$$x\frac{\mathrm{d}y}{\mathrm{d}x} + y = x\sin(x). \tag{1.9}$$

But notice that

$$\frac{\mathrm{d}}{\mathrm{d}x}(xy) = x\frac{\mathrm{d}y}{\mathrm{d}x} + y \qquad \text{(Product Rule)} \quad = \quad x\sin(x) \qquad \text{(Equation (1.9))},$$

so

$$\frac{\mathrm{d}}{\mathrm{d}x}(xy) = x\sin(x).$$

Integrate both sides:

$$xy = C + \int x\sin(x)\,\mathrm{d}x.$$

We have to do integration by parts here. Use LIATE to indentify $u = x$:

$$\begin{aligned} x &= u \implies \mathrm{d}x = \mathrm{d}u \\ \mathrm{d}v = \sin(x)\,\mathrm{d}x &\implies v = -\cos(x). \end{aligned}$$

So

$$\begin{aligned} xy &= C + uv - \int v\mathrm{d}u, \\ &= C + x\left[-\cos(x)\right] + \int \cos(x)\,\mathrm{d}x, \\ &= C - x\cos(x) + \sin(x). \end{aligned}$$

Divide across by x:

$$y = \frac{C}{x} - \cos(x) + \frac{1}{x}\sin(x).$$

Worked example

An ODE to model hormone secretion is given by

$$\frac{dh}{dt} = a - b\cos(\omega t) - kh, \qquad t > 0, \tag{1.10}$$

where $h(t)$ is the amount of a certain hormone in the blood at any time t, and a, b, and k are **positive** constants related to the type of hormone. Solve for $h(t)$, given $h(0) = h_0$. Hence, explain what happens to the level of hormone in the blood as $t \to \infty$.

First of all, we re-arrange this equation as

$$\frac{dh}{dt} + kh = a - b\cos(\omega t).$$

The equation is now in the standard form for application of the integrating-factor technique. Here, the independent variable is t and the dependent variable is h. Thus, $P(t) = k$ and $Q(t) = a - b\cos(2\pi t)$. The integrating factor is therefore

$$\mu(t) = \mathrm{e}^{\int P(t)\,\mathrm{d}t} = \mathrm{e}^{\int k\mathrm{d}t} = \mathrm{e}^{kt},$$

and the ODE can be re-written as a perfect derivative:

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

or

$$\frac{d}{dt}(\mu h) = \mathrm{e}^{kt}\left[a - b\cos(\omega t)\right].$$

Integrate:

$$\mu h = C + \int \mathrm{e}^{kt}\left[a - b\cos(\omega t)\right]\mathrm{d}t.$$

There are two integrals to do here. The first one is

$$I = \int \mathrm{e}^{kt}\mathrm{d}t = \frac{1}{k}\mathrm{e}^{kt},$$

which is easy. The second one is

$$J = \int \mathrm{e}^{kt}\cos(\omega t)\,\mathrm{d}t.$$

This requires to integrations by parts:

$$\begin{aligned}
J &= \int \underbrace{\cos(\omega t)}_{u}\underbrace{\mathrm{e}^{kt}\mathrm{d}t}_{=\mathrm{d}v}, \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) - \int \frac{1}{k}\mathrm{e}^{kt}(-\omega\sin\omega t)\,\mathrm{d}t, \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k}\int \mathrm{e}^{kt}\sin\omega t\,\mathrm{d}t, \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k}\int \underbrace{\sin\omega t}_{=u}\underbrace{\mathrm{e}^{kt}\mathrm{d}t}_{=\mathrm{d}v}, \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k}\left[\frac{1}{k}\mathrm{e}^{kt}\sin\omega t - \int \frac{1}{k}\mathrm{e}^{kt}(+\omega\cos\omega t)\,\mathrm{d}t\right], \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k^2}\mathrm{e}^{kt}\sin\omega t - \frac{\omega^2}{k^2}\int \mathrm{e}^{kt}\cos\omega t\,\mathrm{d}t, \\
&= \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k^2}\mathrm{e}^{kt}\sin\omega t - \frac{\omega^2}{k^2}J.
\end{aligned}$$

Now this is a neat algebraic equation for J:

$$J\left(1 + \frac{\omega^2}{k^2}\right) = \frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k^2}\mathrm{e}^{kt}\sin\omega t,$$

or

$$J = \frac{\frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k^2}\mathrm{e}^{kt}\sin\omega t}{1 + \frac{\omega^2}{k^2}}.$$

The general solution is

$$\mu h = C + aI - bJ,$$

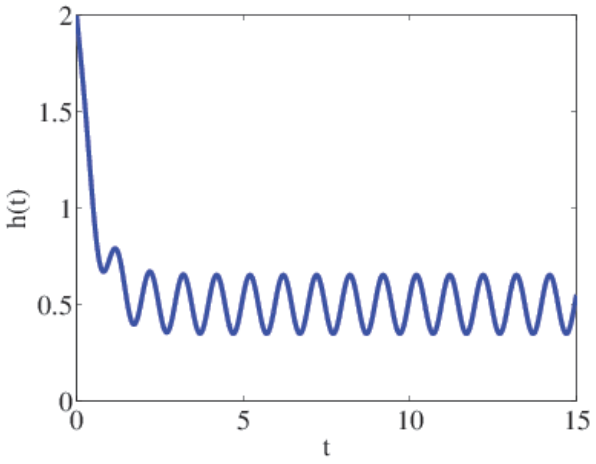


Figure 1.8: Particular solution of Eq. (1.10).

or

$$\mu h = C + \frac{a}{k}\mathrm{e}^{kt} - b\frac{\frac{1}{k}\mathrm{e}^{kt}\cos(\omega t) + \frac{\omega}{k^2}\mathrm{e}^{kt}\sin\omega t}{1+\frac{\omega^2}{k^2}}.$$

Divide out by the integrating factor $\mu = \mathrm{e}^{kt}$:

$$\boxed{h(t) = C\mathrm{e}^{-kt} + \frac{a}{k} - b\frac{\frac{1}{k}\cos(\omega t) + \frac{\omega}{k^2}\sin\omega t}{1+\frac{\omega^2}{k^2}}.} \tag{1.11}$$

This is the general solution. For the particular solution, we use $h(0) = 2$ and obtain

$$h(0) = h_0 = C + \frac{a}{k} - b\frac{\frac{1}{k}}{1+\frac{\omega^2}{k^2}} = C + \frac{a}{k} - \frac{bk}{k^2+\omega^2},$$

hence

$$C = h_0 - \frac{a}{k} + \frac{bk}{k^2+\omega^2}.$$

The particular solution is therefore

$$h(t) = \left(h_0 - \frac{a}{k} + \frac{bk}{k^2+\omega^2}\right)\mathrm{e}^{-kt} + \frac{a}{k} - b\frac{\frac{1}{k}\cos(\omega t) + \frac{\omega}{k^2}\sin\omega t}{1+\frac{\omega^2}{k^2}}.$$

The solution is plotted in Fig. 1.8 for some particular values of the various parameters. As $t \to \infty$, the exponential term tends to zero and only the other two terms survive. In other words,

$$\lim_{t\to\infty} h(t) = \frac{a}{k} - b\frac{\frac{1}{k}\cos(\omega t) + \frac{\omega}{k^2}\sin\omega t}{1+\frac{\omega^2}{k^2}}$$

This is called a **quasi-steady state**. We cannot call it a steady state, since it depends on time. However, the initial **transience** produced by the initial condition h_0 does not appear in this quasi-

steady state, and this is what makes it special. The hormone level 'loses memory' of the initial condition and the level is dictated only by the terms on the RHS of Equation (1.10).

1.9 Existence of solutions

Theorem 1.3 *Let $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ be a real-valued function of two real variables. Assume that f is continuous and that there exists a real number K with*

$$|f(x, y_2) - f(x, y_1)| \leq K|y_2 - y_1|, \tag{1.12}$$

for all $x \in [a, b]$, and for all $y_1, y_2 \in \mathbb{R}$. Then the ODE

$$\frac{dy}{dx} = f(x, y) \tag{1.13}$$

with initial condition

$$y(x = a) = y_0, \qquad y_0 \in \mathbb{R}$$

has a unique solution in some finite interval $[a, b_*]$, with $a < b_* \leq b$.

Remarks:

- The proof of this theorem requires advanced topics and is not covered in this module.
- Even so, it is a relatively weak statement: it is guaranteed that a unique solution exists only in some interval $[a, b_*]$ – the solution does not even necessarily exist out to $x = b$. For this reason, the theorem is called a **local-existence theorem**.
- Also, the condition (1.12) is rather restrictive: this is a stronger condition than simple continuity.
- Finally, note the precise wording of the theorem: the K-number ('Lipschitz constant') has to be independent of x!
- The K-number can, however, depend on the limits a and b of the x-set.
- The f-functions in this module always have the property (1.12); you would have to be very unfortunate to encounter a function without it!

A similar theorem exists for second-order ODEs of the form

$$\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right). \tag{1.14}$$

Again, you can think of the existence condition as being that the RHS of Eq. (1.14) be a 'sufficiently smooth function'.

Chapter 2

Introduction to Second-Order ODEs

2.1 Outline

We introduce second-order ODEs, starting with applications in Mechanics. This will enable us to formulate a more general theory in later sections.

2.2 Review of Linear ODEs, Constant-Coefficient ODEs

Definition 2.1 *A first-order ODE of the form*

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

is called **linear** *if f is a linear function of y:*

$$f(x, y) = a(x)y + b(x).$$

Similarly, a second-order ODE of the form

$$\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right)$$

is called **linear** *if f is a linear function of y and dy/dx:*

$$f\left(x, y, \frac{dy}{dx}\right) = a(x)y + b(x)\frac{dy}{dx} + c(x).$$

A second important defintion is the following:

Definition 2.2 *The above ODEs are called* **homogeneous** *if* $c(x) = 0$*; otherwise, they are called* **inhomogeneous**.

There are good reasons for this tedious classification:

- A first-order linear **homogeneous** ODE has one independent solution and one constant of integration;
- A second-order linear **homogeneous** ODE has two independent solutions; the general solution is obtained by adding the two independent solutions; the two coefficients in front of the two independent solutions are the two constants of integration.

The constants are fixed by the initial conditions: one condition for a first-order ODE, two for a second-order ODE. The fixing of the constants makes the solution unique.

Example: Consider the following linear first-order ODE with constant coefficients:

$$\frac{dy}{dx} = ay + b.$$

The general solution is available immediately through the integrating-factor technique:

$$y = C\mathrm{e}^{ax} - \frac{b}{a}.$$

Hence, there is one independent solution and one constant of integration.

This is not that exciting. However, let us consider a second-order ODE with constant coefficients:

$$\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right), \qquad f\left(x, y, \frac{dy}{dx}\right) = ay + b\frac{dy}{dx} + c.$$

In other words,

$$\frac{d^2y}{dx^2} - b\frac{dy}{dx} - ay = c.$$

We momentarily focus on the homogeneous case with $c = 0$; the case with $c \neq 0$ can be recovered after the following analysis. As such, we consider

$$\frac{d^2y}{dx^2} - b\frac{dy}{dx} - ay = 0.$$

Let us take it as given that a candidate solution is $y = A\mathrm{e}^{\lambda x}$. Substitute the following relations into the ODE:

$$\frac{dy}{dx} = A\lambda\mathrm{e}^{\lambda x}, \qquad \frac{d^2y}{dx^2} = A\lambda^2\mathrm{e}^{\lambda x};$$

the result is

$$A\lambda^2 \mathrm{e}^{\lambda x} - bA\lambda \mathrm{e}^{\lambda x} - a\left(A\mathrm{e}^{\lambda x}\right).$$

Tidy up:

$$A\lambda^2 \mathrm{e}^{\lambda x} - bA\lambda \mathrm{e}^{\lambda x} - aA\mathrm{e}^{\lambda x} = 0.$$

We can cancel the exponentials, and the A's (since $\mathrm{e}^{\lambda x}$ is never zero):

$$\lambda^2 - b\lambda - a = 0.$$

This is a quadratic equation with two roots:

$$\lambda_1 = \frac{b + \sqrt{b^2 + 4a}}{2}, \qquad \lambda_2 = \frac{b - \sqrt{b^2 + 4a}}{2}.$$

There are therefore two candidate solutions:

$$y_1(x) = A_1 \mathrm{e}^{\lambda_1 x}, \qquad y_2(x) = A_2 \mathrm{e}^{\lambda_2 x},$$

assuming $\lambda_1 \neq \lambda_2$. We can add the two equations together to obtain the **general solution**:

$$y(x) = y_1(x) + y_2(x) = A_1 \mathrm{e}^{\lambda_1 x} + A_2 \mathrm{e}^{\lambda_2 x}$$

As befitting a second-order linear ODE, there are two constants of integration associated with the two candidate solutions. The sum of candidate solutions is called a **linear superposition**.

By direct computatation, the solution of the corresponding inhomogeneous equation (with a constant right-hand side equal to c) is

$$y(x) = y_1(x) + y_2(x) = A_1 \mathrm{e}^{\lambda_1 x} + A_2 \mathrm{e}^{\lambda_2 x} - (c/a).$$

Concident roots

When $b^2 + 4a = 0$, $\lambda_1 = \lambda_2$ in the foregoing analysis, and both solutions collapse to

$$y_1 = A_1 \mathrm{e}^{(b/2)x}.$$

However, a second solution is still expected in this case – only the foregoing analysis has failed to find it. As such, we make the inspired guess

$$y_2 = A_2 x \mathrm{e}^{(b/2)x}.$$

We compute

$$\frac{dy_2}{dx} = A_2(b/2)x\mathrm{e}^{(b/2)x} + A_2\mathrm{e}^{(b/2)x}, \qquad \frac{d^2y_2}{dx^2} = A_2(b/2)^2x\mathrm{e}^{(b/2)x} + A_2b\mathrm{e}^{(b/2)x}.$$

We view y_2 as a **trial solution** and substitute into the ODE – as such we aim to verify that y_2 satisfies the ODE identically. We have:

$$\frac{d^2y_2}{dx^2} - b\frac{dy_2}{dx} - ay_2 = \left[A_2(b/2)^2x\mathrm{e}^{(b/2)x} + A_2b\mathrm{e}^{(b/2)x}\right] - b\left[A_2(b/2)x\mathrm{e}^{(b/2)x} + A_2\mathrm{e}^{(b/2)x}\right] - a\left[A_2x\mathrm{e}^{(b/2)x}\right].$$

Gather powers of x:

$$\begin{aligned}\frac{d^2y_2}{dx^2} - b\frac{dy_2}{dx} - ay_2 &= xA_2\left(\tfrac{1}{4}b^2 - \tfrac{1}{2}b^2 - a\right)\mathrm{e}^{(b/2)x} \\ &\qquad + A_2\left(b - b\right)\mathrm{e}^{(b/2)x}, \\ &= xA_2\left(\tfrac{1}{4}b^2 - \tfrac{1}{2}b^2 + \tfrac{1}{4}b^2\right)\mathrm{e}^{(b/2)x} \\ &= 0,\end{aligned}$$

hence

$$y_2 = xA_2\mathrm{e}^{(b/2)x}$$

is a second solution, and the general solution is

$$y(x) = y_1(x) + y_2(x) = A_1\mathrm{e}^{(b/2)x} + A_2x\mathrm{e}^{(b/2)x}.$$

Again, the solution of the corresponding inhomogeneous equation is

$$y(x) = y_1(x) + y_2(x) = A_1\mathrm{e}^{(b/2)x} + A_2x\mathrm{e}^{(b/2)x} - (c/a).$$

2.3 The harmonic oscillator

A particle of mass m experiences a spring force $F = -kx$, where $k > 0$ is the spring constant ('Hooke's Law'). Using Newton's Law,

$$\text{Force} = \text{mass} \times \text{acceleration},$$

and the following initial conditions:

$$x(t=0) = 0, \qquad \left.\frac{dx}{dt}\right|_{t=0} = \sqrt{k/m},$$

find the particle's position for all times, $x(t)$.

The acceleration of the particle is the second derivative of position with respect to time,

$$\text{acceleration} = \frac{d^2x}{dt^2}.$$

Substitute this formula into Newton's Law:

$$m\frac{d^2x}{dt^2} = -kx.$$

This is a second-order linear ODE. According to our classification, the f-function is

$$f\left(x, t, \frac{dx}{dt}\right) = ax + b\frac{dx}{dt} + c, \qquad a = -k/m, \qquad b = c = 0.$$

Therefore, we make a trial solution

$$y = A\mathrm{e}^{\lambda t}.$$

The derivatives are

$$\frac{dx}{dt} = \lambda A\mathrm{e}^{\lambda t}, \qquad \frac{d^2x}{dt^2} = \lambda^2 A\mathrm{e}^{\lambda t}.$$

Substitute these into the ODE $d^2x/dt^2 = -kx$:

$$\lambda^2 A\mathrm{e}^{\lambda t} = -\frac{k}{m}A\mathrm{e}^{\lambda t}.$$

Because the exponential $\mathrm{e}^{\lambda t}$ is never zero, there is some cancellation (the A's cancel too), and we are left with

$$\lambda^2 = -\frac{k}{m}.$$

We are perfectly within our rights to take square roots here – the only problem is that they are imaginary:

$$\lambda_1 = +\mathrm{i}\sqrt{k/m}, \qquad \lambda_2 = -\mathrm{i}\sqrt{k/m}.$$

To save ink/chalk/typing, we are going to call

$$\omega := \sqrt{k/m}.$$

Hence, the two candidate solutions are

$$y_1 = A_1\mathrm{e}^{\mathrm{i}\omega t}, \qquad y_1 = A_2\mathrm{e}^{-\mathrm{i}\omega t},$$

and the general solution is

$$y(t) = y_1 + y_2 = A_1\mathrm{e}^{\mathrm{i}\omega t} + A_2\mathrm{e}^{-\mathrm{i}\omega t}.$$

To fix the constants, we apply the initial conditions:

$$y(0) = 0, \qquad \left.\frac{dy}{dt}\right|_{t=0} = \omega,$$

or

$$A_1 + A_2 = 0, \qquad \mathrm{i}\omega\,(A_1 - A_2) = \omega.$$

Hence,

$$A_1 = -A_2, \quad 2\mathrm{i}A_1 = 1 \implies A_1 = \frac{\omega}{2\mathrm{i}}.$$

The particular solution is therefore

$$y(t) = \frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega t}\right).$$

However, this is nothing other than $\sin\omega t$!

$$y(t) = \sin\omega t,$$

as can be verified by substitution $y = \sin\omega t$ into the ODE.

2.4 Mechanical models with friction

Newton's law for a particle executing one-dimensional motion reads

$$[\text{mass}] \times [\text{acceleration}] = [\text{force}],$$

or

$$m\frac{d^2x}{dt^2} = f\left(t, x, \frac{dx}{dt}\right),$$

where f represents the net force on the particle. If we can solve this ODE and express $x = x(t; x_0)$, then we have **solved for the motion** – we can predict the location of the particle for all time.

For a particle executing motion in a viscous medium, the net force is the frictional force and the external force, summed together:

$$f = f_{\text{fric}} + f_{\text{ext}}.$$

Friction opposes motion, and the frictional force must therefore be proportional to velocity. A general model for the friction force is thus

$$f = -mDv|v|^n, \qquad v = \frac{dx}{dt},$$

where the minus sign indicates that this force acts in a direction **against** the velocity, and D is the so-called **drag coefficient**. In this module, we consider the simplest possible friction model, with $n = 0$. Hence, Newton's equation reads

$$m\frac{d^2x}{dt^2} = -mD\frac{dx}{dt} + f_{\text{ext}}.$$

A spring in a viscous medium

Consider a particle of mass m attached to a spring and experiencing the spring force $F_1 = -kx$. The particle moves in a viscous medium and experiences the frictional force $F_2 = -mDv$, and an external force $F_3 = F_0\cos(\omega_0 t)$. Derive a condition on the damping coefficient D for the particle's inertia to be negligible. Hence, compute the terminal velocity of the particle in this limiting case.

Newton's Law states that

$$m\frac{d^2x}{dt^2} = F_1 + F_2 + F_3.$$

In other words,

$$m\frac{d^2x}{dt^2} + mD\frac{dx}{dt} + kx = F_0\cos(\omega t).$$

Introduce a new variable $\tau = \omega t$:

$$\omega^2\frac{d^2x}{d\tau^2} + D\omega\frac{dx}{d\tau} + \omega_0^2 x = \frac{F_0}{m}\cos\tau, \qquad \omega_0^2 = k/m.$$

Introduce x_0, a 'typical' length scale, and divide across by $\omega^2 x_0$:

$$\frac{d^2}{d\tau^2}(x/x_0) + \frac{D}{\omega}\frac{d}{d\tau}(x/x_0) + \frac{\omega_0^2}{\omega^2}(x/x_0) = \frac{F_0}{mx_0\omega^2}\cos\tau.$$

Finally, multiply across by ω/D:

$$\frac{\omega}{D}\frac{d^2}{d\tau^2}(x/x_0) + \frac{d}{d\tau}(x/x_0) + \frac{\omega_0^2}{\omega D}(x/x_0) = \frac{F_0}{m\omega D x_0}\cos\tau. \tag{2.1}$$

Finally, let us introduce **dimensionless groups**

$$\epsilon := \frac{\omega}{D}, \qquad R = \frac{\omega_0^2}{\omega D}, \qquad F = \frac{F_0}{m\omega D x_0}.$$

Thus, Equation (2.1) becomes

$$\epsilon\frac{d^2}{d\tau^2}(x/x_0) + \frac{d}{d\tau}(x/x_0) + R(x/x_0) = F\cos\tau. \tag{2.2}$$

The beautiful thing about Equation (2.2) is that everything is non-dimensional. Thus, it is mean-

ingful to speak of D as being 'large': in this instances, what we mean is that the dimensionless group $\epsilon = \omega/D$ is small: $\epsilon \ll 1$. Indeed, in the limit as $\epsilon \to 0$, the equation becomes a first-order differential equation[1].

$$\frac{d}{d\tau}(x/x_0) + R(x/x_0) = F\cos\tau.$$

We now imagine taking the limit $\epsilon \to 0$ in the dimensionless equation (2.2), and then restoring the dimensional parameters. Then, the equation to solve is

$$D\frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m}\cos(\omega t)\,.$$

Thus, the limit $\epsilon \to 0$ amounts simply to throwing away the acceleration term. In other words, it is as though we have set $m = 0$. This is called **negligible inertia** ($m = 0$; mass is the ability of a body to resist changes in acceleration – inertia). In summary,

Inertia is negligible

if and only if Acceleration term can be thrown away

if and only if $D \to \infty$ with $R = \dfrac{\omega_0^2}{\omega D} = O(1)$ and $F = \dfrac{F_0}{m\omega D x_0} = O(1)$.

Therefore, in this limit, to solve for the terminal velocity, it suffices to set the net force to zero:

$$F_1 + F_2 + F_3 = 0 \implies -kx - mDv + F_0\cos(\omega t) = 0.$$

Unhappily, this is still an ODE that must be solved:

$$mDv + kx = F_0\cos(\omega_0 t)\,,$$

or

$$\frac{dx}{dt} + \frac{k}{mD}x = \frac{F_0}{mD}\cos(\omega t)\,.$$

However, this is precisely the equation of hormone secretion in Chapter 1,

$$\frac{dh}{dt} + \alpha h = \beta + \gamma\cos(\omega t),$$

with solution

$$h(t) = (\text{Const.})\,\mathrm{e}^{-\alpha t} + \frac{\beta}{\alpha} + \gamma\frac{\alpha\cos(\omega t) + \omega\sin(\omega t)}{\alpha^2 + \omega^2}.$$

Let us fill in what the parameters are, using $\omega_0 = \sqrt{k/m}$ to denote the **natural frequency** of the

[1] This is called a **singular limit**, and care must be taken here. The passage to this limit is treated in a rigorous fashion in the advanced sections of this module

spring: $\alpha = \omega_0^2/D$, $\beta = 0$, $\gamma = (F_0/mD)$:

$$x(t) = (\text{Const.})\,\mathrm{e}^{-(\omega_0/D)\omega_0 t} + \frac{F_0}{mD}\frac{\frac{\omega_0^2}{D}\cos(\omega t) + \omega\sin(\omega t)}{\left(\frac{\omega_0^2}{D}\right)^2 + \omega^2}$$

To make life easy here, we define

$$A(\omega) := \frac{F_0}{mD}\frac{1}{\left(\frac{\omega_0^2}{D}\right)^2 + \omega^2}.$$

The ODE therefore has solution

$$x(t) = (\text{Const.})\,\mathrm{e}^{-(\omega_0/D)\omega_0 t} + A(\omega)\left[\frac{\omega_0^2}{D}\cos(\omega t) + \omega\sin(\omega t)\right].$$

The velocity is therefore

$$v = \frac{dx}{dt} = -(\omega_0/D)\omega_0\,(\text{Const.})\,\mathrm{e}^{-(\omega_0/D)\omega_0 t} + A(\omega)\left[-\frac{\omega_0^2}{D}\omega\sin(\omega t) + \omega^2\cos(\omega t)\right].$$

The terminal velocity is given by the limit $t \to \infty$:

$$v_\infty = \lim_{t\to\infty} v(t) = \omega^2 A(\omega)\left[\cos(\omega t) - \frac{\omega_0^2}{\omega D}\sin(\omega t)\right].$$

Again, this is **independent** of initial conditions.

The skydiver

Consider a body falling in the earth's gravitational field $f_{\text{ext}} = -g$, subject to the linear viscous force just described. Solve for the motion, subject to $x(t=0) = H$, and $dx/dt(t=0) = 0$.

We have

$$m\frac{d^2x}{dt^2} = -mD\frac{dx}{dt} - mg.$$

Introduce $v = dx/dt$ and divide by m. The result is a first-order ODE:

$$\frac{dv}{dt} + Dv = -g.$$

This is an old favourite, and the integrating-factor method applies:

$$\frac{d}{dt}\left(v\mathrm{e}^{Dt}\right) = -g\mathrm{e}^{Dt}.$$

Hence,

$$v = A\mathrm{e}^{-Dt} - \frac{g}{D}.$$

We have,

$$v(t=0) = 0 = A - (g/D), \implies A = g/D,$$

and

$$v = \frac{g}{D}\left(\mathrm{e}^{-Dt} - 1\right).$$

But

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

hence

$$\frac{dx}{dt} = \frac{g}{D}\left(\mathrm{e}^{-Dt} - 1\right).$$

Integrate again:

$$x - x(t=0) = \frac{g}{D}\left(-\frac{1}{D}\mathrm{e}^{-Dt} - t\right),$$

or

$$x(t) = H - \frac{g}{D}\left(\frac{1}{D}\mathrm{e}^{-Dt} - t\right).$$

Consider again the formula for v, $v = (g/D)(\mathrm{e}^{-Dt} - 1)$. For large t, such that the skydiver has not yet hit the ground $x(t) = 0$, the exponential factor is negligible, and

$$v = v_\infty = -\frac{g}{D}.$$

This is the constant, **terminal velocity**. Note, also that the terminal velocity can be obtained by setting

$$f_{\text{friction}} + f_{\text{external}} = 0,$$

$$-mDv - mg = 0 \implies v = -\frac{g}{D}.$$

These two examples illustrate an essential concept for particles in a very viscous medium:

> A particle moving in a very viscous medium has negligible inertia. The acceleration term can be set to zero in Newton's equation. To find the terminal velocity, it suffices to set the sum of the friction force and the external and internal forces to zero. The terminal velocity is always independent of initial conditions.

2.5 A final word of caution

In general an n^{th}-order linear ODE

$$\frac{d^n y}{dx^n} = \sum_{j=0}^{n-1} a_j(x) \frac{d^j y}{dx^j}$$

has n independent solutions that are summed together to find the general solution, with n free parameters, C_j:

$$y(x) = \sum_{j=0}^{n} C_j y_j(x).$$

Thus, when you find n independent candidate solutions, you know you are done. **This is not necessarily true for nonlinear ODEs**. You might have a second-order nonlinear ODE, and you might find two independent solutions. It is NOT true that the most general solution is then a superposition of the two candidate solutions. Some people tend to forget this lesson – please remember it!

Chapter 3

Analysis of general linear Second-Order ODEs

This chapter was going to look at the method of variation of parameters for second-order linear inhomogeneous ODEs. We will now skip this and move onto other things. There is a good article on variation of parameters on Wikipedia, ro any readers disappointed by the sudden disappearance of this chapter.

Chapter 4

Systems of linear ODEs

4.1 Outline

We look at systems of two coupled linear ODEs.

4.2 Introduction

Consider the homogeneous linear system for two real functions $x(t), y(t)$, functions of the real "time" variable t:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a\,x + b\,y, \tag{4.1a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = c\,x + d\,y, \tag{4.1b}$$

where a, b, c, d are real constants such that $ad - bc \neq 0$.

The solution to this system depends on the eigenvectors and eigenvalues of the underlying 2×2 real matrix

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Let us re-write the system (4.1) in matrix form. Defining the column vector

$$\boldsymbol{v}(t) \equiv \begin{pmatrix} x(t) \\ y(t) \end{pmatrix},$$

we define its time derivative as

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}.$$

Hence, the system (4.1) is equivalent to

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{v}(t), \tag{4.2}$$

where the RHS is understood as the usual matrix operation:

$$\boldsymbol{A}\boldsymbol{v}(t) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} a\,x(t) + b\,y(t) \\ c\,x(t) + d\,y(t) \end{pmatrix}.$$

Notice that the condition $ad - bc \neq 0$ is simply $\det \boldsymbol{A} \neq 0$, i.e., $\boldsymbol{A}$ is invertible.

Example: Express the system of equations

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= y, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= x, \end{aligned}$$

in matrix form. Also, find a solution of the equations.

Solution: In matrix form, the system of equations reads

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \boldsymbol{v}(t) \qquad \boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

with $\det \boldsymbol{A} = -1 \neq 0$.

Let us first solve the system by hand. We take derivative of the first equation and get $\mathrm{d}^2x/\mathrm{d}t^2 = \mathrm{d}y/\mathrm{d}t$, and next, we use the second equation to replace $\mathrm{d}y/\mathrm{d}t$ so we get: $\mathrm{d}^2x/\mathrm{d}t^2 = x(t)$. The solution of this equation is $x(t) = a_1\,\exp(t) + a_2\,\exp(-t)$, where a_1, a_2 are arbitrary constants. To obtain $y(t)$, we can use the first equation: $y(t) = x'(t)$ so we get $y(t) = a_1\,\exp(t) - a_2\,\exp(-t)$. Combining these solutions to form the vector $\boldsymbol{v}(t)$ we get:

$$\boldsymbol{v}(t) = a_1 \exp(t) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2 \exp(-t) \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Notice that the vectors $(1, 1)^T$ and $(1, -1)^T$ appearing in this solution are precisely eigenvectors of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This matrix is known as a reflection matrix. It reflects points with respect to the line $y = x$. We

have

$$\boldsymbol{A}\begin{pmatrix} a_1 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_1 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_1 \end{pmatrix}$$

and

$$\boldsymbol{A}\begin{pmatrix} a_2 \\ -a_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_2 \\ -a_2 \end{pmatrix} = \begin{pmatrix} -a_2 \\ a_2 \end{pmatrix} = (-1)\begin{pmatrix} a_2 \\ -a_2 \end{pmatrix},$$

for any $a_1, a_2 \in \mathbb{R}$.

At the moment the relation between the solution of our system of ODEs and the eigenvectors of the matrix $\mathbb{A}$ is rather imprecise but it can be made rigorous, so that we can produce a solution $\boldsymbol{v}(t)$ of the system, given an arbitrary invertible matrix $\mathbb{A}$. As such, we recall here some facts / properties of eigenvalues and eigenvectors, which will be useful in what follow. Throughout, $\boldsymbol{A}$ be a 2×2 matrix with real entries, and let $\boldsymbol{u}$ be a non-zero column vector with 2 entries (so that not all of the entries of $\boldsymbol{u}$ are zero).

- The vector $\boldsymbol{u}$ is called an eigenvector for $\boldsymbol{A}$ if $\boldsymbol{A}\boldsymbol{u} = \lambda\,\boldsymbol{u}$, where λ is some number.

  Indeed, λ is called an eigenvalue for $\boldsymbol{A}$ and the eigenvector $\boldsymbol{u}$ is said to correspond to λ.

- The equation $\det(\mathbb{A} - \lambda\,\mathbb{I}_2) = 0$ is a quadratic equation for λ and is called the **characteristic equation**. The roots of the characteristic equation can be shown to be the eigenvalues of the matrix $\boldsymbol{A}$.

- The determinant of any 2×2 matrix $\boldsymbol{A}$ is equal to the product of its two eigenvalues. As such, since we assume that the matrix $\boldsymbol{A}$ is invertible throughout this chapter, this is equivalent to the condition that none of its eigenvalues is equal to zero.

4.3 Second-order linear ODEs fall into the present framework

We have already looked at second-order **linear homogeneous** ODEs (Chapter 2), the most general form of which was

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

(The inhomoegenoues counterpart of this ODE would have $c(t) \neq 0$ on the RHS). Such homogeneous ODEs can be incorporated into the present framework, by letting

$$y = \frac{\mathrm{d}x}{\mathrm{d}t}.$$

Then,

$$\begin{aligned}\frac{\mathrm{d}^2x}{\mathrm{d}t^2} &= \frac{\mathrm{d}y}{\mathrm{d}t},\\ &= b(t)\frac{dx}{dt} + a(t)x,\\ &= b(t)y + a(t)x.\end{aligned}$$

So we have a system of equations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} y \\ b(t)y + a(t)x \end{pmatrix},\\ &= \begin{pmatrix} 0 & 1 \\ b(t) & a(t) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.\end{aligned}$$

4.4 General solution of the homogeneous linear system

Let us consider the system (4.1) again:

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \mathbb{A}\boldsymbol{v},$$

where the matrix $\mathbb{A}$ is a general 2×2 real matrix with $\det \mathbb{A} \neq 0$.

4.4.1 The diagonalisable case

We first of all assume that $\boldsymbol{A}$ is diagonalizabile – i.e. linearly independent eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ exist. We start with some notation:

- **The Initial Value Problem** for this system, corresponds to the above evolution equation supplemented with an initial condition $\boldsymbol{v}(0) = \boldsymbol{v}_0$, a constant vector.

  Notice that the amount of information in this initial condition (two real numbers, one for each component of $\boldsymbol{v}_0$) is the same as in the case of 2^{nd}-order linear ODEs discussed in Chapter 2. This does not happen by chance – the reasons for this agreement are noted here in Section 4.3.

- **The General Solution** of the system of equations (4.1) corresponds therefore to the solution of the IVP when the two components of the vector $\boldsymbol{v}_0$ are arbitrary constants.

Then, we can say that the general solution of Equation (4.1) is given by

$$\boldsymbol{v}(t) = a_1 \exp(\lambda_1 t)\boldsymbol{u}_1 + a_2 \exp(\lambda_2 t)\boldsymbol{u}_2, \tag{4.3}$$

where a_1, a_2 are arbitrary constants.

This can be seen by direct computation as follows. Explicitly, let us compute the time derivative of $\boldsymbol{v}(t)$:

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= a_1\,\lambda_1\,\exp(\lambda_1\,t)\boldsymbol{u}_1 + a_2\,\lambda_2\,\exp(\lambda_2\,t)\boldsymbol{u}_2\\ &= a_1\,\exp(\lambda_1\,t)\mathbb{A}\boldsymbol{u}_1 + a_2\,\exp(\lambda_2\,t)\mathbb{A}\boldsymbol{u}_2\\ &= \mathbb{A}\left(a_1\,\exp(\lambda_1\,t)\boldsymbol{u}_1 + a_2\,\exp(\lambda_2\,t)\boldsymbol{u}_2\right)\\ &= \mathbb{A}\boldsymbol{v}(t).\end{aligned}$$

So $\boldsymbol{v}(t)$ satisfies the system of ODEs. Now, let us see if the constants a_1, a_2 can be chosen to generate any given initial condition $\boldsymbol{v}(0) = \boldsymbol{v}_0$. We have:

$$\boldsymbol{v}(0) = a_1\,\boldsymbol{u}_1 + a_2\,\boldsymbol{u}_2 = \boldsymbol{v}_0,$$

and this system has a unique solution for a_1, a_2, because the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are linearly independent.

Example: Solve the system of equations

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{v}, \qquad \boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

using the eigenvalue analysis.

Solution: The eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -1$, with corresponding eigenvectors $\boldsymbol{u}_1 = (1,1)^T$ and $\boldsymbol{u}_2 = (1,-1)^T$ respectively. Therefore the general solution of the ODE system is

$$\boldsymbol{v}(t) = a_1\,\exp(\lambda_1\,t)\boldsymbol{u}_1 + a_2\,\exp(\lambda_2\,t)\boldsymbol{u}_2 = a_1\exp(t)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2\exp(-t)\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

which coincides with the solution obtained previously with a direct method.

4.4.2 The non-diagonalisable case

In this case, the matrix $\boldsymbol{A}$ has only one linearly independent eigenvector. Notice that therefore, its two eigenvalues are equal.

Remark 4.1 *Warning: the converse is not true: if a matrix has two equal eigenvalues, then this does not imply that the matrix has only one eigenvector. Think of the identity matrix: its two eigenvalues are equal to 1, and it has two linearly independent eigenvectors).*

When a matrix $\boldsymbol{A}$ has only one eigenvector, the solution of the system $\boldsymbol{v}'(t) = \boldsymbol{A}\boldsymbol{v}(t)$ cannot be obtained from the previous analysis (e.g. Equation (4.3), and a new method needs to be introduced.

Say whether the following matrix is diagonalizable or not.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Solution: The characteristic equation for this matrix is

$$\det(\boldsymbol{A} - \lambda \mathbb{I}_2) = 0,$$

$$\implies \det\left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right] = 0,$$

$$\implies \det\begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} = 0,$$

so finally $(1-\lambda)^2 = 0$, which gives an eigenvalue $\lambda = 1$ with multiplicity two.

We now write the equation for the eigenvector $\boldsymbol{u}$ corresponding to $\lambda = 1$. We have, writing $\boldsymbol{u} = (w, z)^T$,

$$(\boldsymbol{A} - \lambda \mathbb{I}_2)\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

hence

$$\begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix}\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Putting $\lambda = 1$ we get

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives only one independent equation: $z = 0$ (The second equation, $0 = 0$, is an identity).

The result is then that $z = 0$, while w is arbitrary, so we get the eigenvector

$$\boldsymbol{u} = w\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

only one linearly-independent eigenvector, so the matrix $\boldsymbol{A}$ in this example is non-diagonalisable.

The Hamilton–Cayley Theorem

In this section, we will need to use the **Hamilton–Cayley Theorem** and its consequences. The Hamilton–Cayley Theorem says that any matrix $\mathbb{R}^{n\times n}$ satisfies its own characteristic equation. This can readily be verified in the 2×2 case by direct computation, where the characteristic equation $\det(\boldsymbol{A} - \lambda\mathbb{I}) = 0$ reduces to

$$\begin{aligned} p(\lambda) &= \lambda^2 - \lambda\mathrm{tr}(\boldsymbol{A}) + \det(\boldsymbol{A}), \\ p(\lambda) &= 0. \end{aligned}$$

Hence, the characteristic equation is a second-order polynomial, which can be re-written in the variable x as

$$p(x) = x^2 - x\mathrm{tr}(\boldsymbol{A}) + \det(\boldsymbol{A}).$$

The Hamilton–Cayley Theorem in this instance therefore states that $p(\boldsymbol{A}) = 0$, hence

$$\boldsymbol{A}^2 - \boldsymbol{A}\mathrm{tr}(\boldsymbol{A}) + \det(\boldsymbol{A})\mathbb{I} = 0. \tag{4.4}$$

But $\mathrm{tr}(\boldsymbol{A}) = \lambda_1 + \lambda_2$, and $\det(\boldsymbol{A}) = \lambda_1\lambda_2$, where λ_1 and λ_2 are the eigenvalues of $\boldsymbol{A}$. Hence, Equation (4.4) becomes:

$$\boldsymbol{A}^2 - \boldsymbol{A}(\lambda_1 + \lambda)2) + \lambda_1\lambda_2\mathbb{I} = 0. \tag{4.5}$$

We specialize to the non-diagonalizabile case, when $\boldsymbol{A}$ has only one eigenvalue, with $\lambda_1 = \lambda_1 = \lambda$. Then, Equation (4.5) becomes:

$$\boldsymbol{A}^2 - 2\lambda\boldsymbol{A} + \lambda^2\mathbb{I} = 0, \tag{4.6}$$

which can also be written as

$$(\boldsymbol{A} - \lambda\mathbb{I})^2 = 0. \tag{4.7}$$

Identifying non-diagonalizabile matrices

Back to the general case, there is a simple test to determine if a given 2×2 matrix $\boldsymbol{A}$ is non-diagonalisable. First, this matrix $\boldsymbol{A}$ needs to have only one eigenvalue, call it λ. Second, the matrix $(\boldsymbol{A} - \lambda\mathbb{I}_2)$ needs to be different from the zero matrix, which can be checked by direct inspection. As a consequence, the matrix $\boldsymbol{A}$ has only one eigenvector, call it $\boldsymbol{u}$.

In addition, by Equation (4.7), we have

$$(\boldsymbol{A} - \lambda\mathbb{I}_2)^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

In general terms, we now have a matrix $\boldsymbol{M} = \boldsymbol{A} - \lambda \mathbb{I}_2$, such that $\boldsymbol{M}^2 = 0$, hence $\boldsymbol{M}^2\boldsymbol{x} = 0$ for all $\boldsymbol{x} \in \mathbb{R}^2$. As such,

$$\boldsymbol{M}^2\boldsymbol{x} = 0 \implies \begin{cases} \text{either } \boldsymbol{x} \in \ker(\boldsymbol{M}), \\ \text{or } M\boldsymbol{x} \in \ker(\boldsymbol{M}). \end{cases}$$

The first case amounts to saying that $\boldsymbol{x} = \boldsymbol{u}$, i.e. $\boldsymbol{x}$ is equal to the eigenvector. The second case amounts to saying that $M\boldsymbol{x} = k\boldsymbol{u}$. Both cases are covered by writing

$$\boldsymbol{M}\boldsymbol{x} = k\boldsymbol{u}, \qquad \text{with} \qquad \begin{cases} k = 0, & \text{if } \boldsymbol{x} \text{ is proportional to } \boldsymbol{u}, \\ k \neq 0, & \text{otherwise.} \end{cases}$$

For example, taking $\boldsymbol{A}$ from the previous example, i.e.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

with eigenvector $\boldsymbol{u} = (1, 0)^T$ and eigenvalue $\lambda = 1$, let us define

$$\boldsymbol{x} = \begin{pmatrix} \ell \\ k \end{pmatrix},$$

where ℓ, k are arbitrary real numbers. Then we get

$$\begin{aligned} (\boldsymbol{A} - \lambda \mathbb{I}_2)\, \boldsymbol{x} &= \begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} \begin{pmatrix} \ell \\ k \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \ell \\ k \end{pmatrix} \\ &= \begin{pmatrix} k \\ 0 \end{pmatrix}, \end{aligned}$$

which is proportional to $\boldsymbol{u} = (1, 0)^T$ and equal to zero if and only if $k = 0$, i.e., if and only if $\boldsymbol{x} = \ell\boldsymbol{u}$.

The solution method for non-diagonalizabile matrices

There is a general formula that gives the solution of the ODE system (4.1) in the case when the matrix $\boldsymbol{A}$ is non-diagonalisable:

$$\boldsymbol{v}(t) = \exp(\lambda\, t)\boldsymbol{v}_0 + t \exp(\lambda\, t)\, (\boldsymbol{A} - \lambda \mathbb{I}_2)\, \boldsymbol{v}_0, \tag{4.8}$$

where $\boldsymbol{v}_0$ is the arbitrary initial condition for the vector $\boldsymbol{v}(0)$, consisting of two arbitrary real components.

The validity of Equation (4.8) as a solution to the ODE (4.1) is again checked by direct computation: we have

$$\begin{aligned}\frac{d\boldsymbol{v}}{dt} &= \mathrm{e}^{\lambda t}\left[\lambda\boldsymbol{v}_0 + (\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0\right] + \mathrm{e}^{\lambda t}\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0,\\ &= \boldsymbol{A}\left(\mathrm{e}^{\lambda t}\boldsymbol{v}_0\right) + t\mathrm{e}^{\lambda t}\underbrace{\left[\lambda\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\right]\boldsymbol{v}_0}.\end{aligned}$$

Use the Hamilton–Cayley theorem:

$$(\boldsymbol{A}-\lambda\mathbb{I}_2)^2\,\boldsymbol{v}_0 = 0$$

Hence,

$$(\boldsymbol{A}-\lambda\mathbb{I}_2)\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0 = 0,$$

and

$$\boldsymbol{A}\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0 = \lambda\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0.$$

So the term with the underbrace in the expression for $d\boldsymbol{v}/dt = \cdots$ is really just $\boldsymbol{A}\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0$, hence

$$\begin{aligned}\frac{d\boldsymbol{v}}{dt} &= \boldsymbol{A}\left(\mathrm{e}^{\lambda t}\boldsymbol{v}_0\right) + t\mathrm{e}^{\lambda t}\underbrace{\left[\lambda\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\right]\boldsymbol{v}_0},\\ &= \boldsymbol{A}\left(\mathrm{e}^{\lambda t}\boldsymbol{v}_0\right) + t\mathrm{e}^{\lambda t}\boldsymbol{A}\left[(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0\right],\\ &= \boldsymbol{A}[\left(\mathrm{e}^{\lambda t}\boldsymbol{v}_0\right) + t\mathrm{e}^{\lambda t}\,(\boldsymbol{A}-\lambda\mathbb{I}_2)\,\boldsymbol{v}_0,\\ &= \boldsymbol{A}\boldsymbol{v},\end{aligned}$$

as required.

4.5 Non-linear systmes of ODEs

A non-linear system of ODEs is defined as the system

$$\frac{dx}{dt} = F(x,y), \tag{4.9a}$$

$$\frac{dy}{dt} = G(x,y), \tag{4.9b}$$

where the functions $F(x,y)$ and $G(x,y)$ are not necessarily linear. As in the case of one-dimensional autonomous ODEs, we define a fixed point (x_*,y_*) as constant solution of Equation (4.9), such that

$$F(x_*,y_*) = 0, \qquad G(x_*,y_*) = 0.$$

We examine a solution that consists of a small disturbance away from the fixed point:

$$x(t) = x_* + \delta x(t), \qquad y_* + \delta y(t),$$

where δx and δy are small. We substitute this trial solution into Equation (4.9), and **linearize** the resulting set of equations, in the following sense:

$$F(x_* + \delta x, y_* + \delta y) = \cancel{F(x_*, y_*)} + F_x(x_*, y_*)\delta x + F_y(x_*, y_*)\delta y + \text{higher-order terms},$$

where $F_x = \partial F/\partial x$ and $F_y = \partial F/\partial y$, and where the higher-order terms have been neglected because δx and δy are assumed to be small. Similarly,

$$G(x_* + \delta x, y_* + \delta y) = \cancel{G(x_*, y_*)} + G_x(x_*, y_*)\delta x + G_y(x_*, y_*)\delta y + \text{higher-order terms},$$

Hence, Equation (4.9) becomes

$$\frac{d}{dt}\delta x = F_x(x_*, y_*)\delta x + F_y(x_*, y_*)\delta y + \text{higher-order terms}, \tag{4.10a}$$

$$\frac{d}{dt}\delta y = G_x(x_*, y_*)\delta x + G_y(x_*, y_*)\delta y + \text{higher-order terms}. \tag{4.10b}$$

This can be re-written in matrix form:

$$\frac{d}{dt}\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \begin{pmatrix} F_x(x_*, y_*) & F_y(x_*, y_*) \\ G_x(x_*, y_*) & G_y(x_*, y_*) \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$$

where reference to the higher-order terms is henceforth omitted. Thus, in a small neighbourhood of the fixed point (or 'critical point'), the nonlinear equations behave as though they were linear; the corresponding $\boldsymbol{A}$-matrix is now

$$\boldsymbol{A} = \begin{pmatrix} F_x(x_*, y_*) & F_y(x_*, y_*) \\ G_x(x_*, y_*) & G_y(x_*, y_*) \end{pmatrix}. \tag{4.11}$$

Definition 4.1 *The matrix in Equation* (4.11) *is called the* **Jacobian** *of the fixed point of the nonlinear system.*

In general, the eigenvalues λ of the Jacobian will be complex numbers, with $\lambda = \lambda_\mathrm{r} + \mathrm{i}\lambda_\mathrm{i}$. The fixed point is called

- **Stable** if $\lambda_\mathrm{r} < 0$ for both eigenvalues,
- **Unstable** if $\lambda_\mathrm{r} > 0$ for at least one eigenvalue,

- **Neutral** if $\lambda_{\rm r} = 0$ for both eigenvalues.

In the remaining part of this chapter we look at one particular nonlinear system that corresponds to **predator-prey dynamics**.

Worked Example – Predator-prey dynamics

Here, we take

$$F(x,y) = A\,x - B\,x\,y, \qquad G(x,y) = -C\,y + D\,x\,y,$$

where A, B, C, D are positive constants. Hence, x is the prey and y is the predator.

In order to find the fixed point(s) of this system, therefore, we need to solve the following algebraic system of equations:

$$A\,x - B\,x\,y = 0\,, \quad -C\,y + D\,x\,y = 0.$$

Clearly, we can factorise these equations to get

$$x\,(A - B\,y) = 0, \qquad y\,(D\,x - C) = 0.$$

We see that there is one solution given by $(x_* = 0, y_* = 0)$. This corresponds to the situation when there is no population of either species. Obviously an initial condition like that will remain so at subsequent times. There is another solution of this algebraic system, given by

$$x_* = \frac{C}{D}, \qquad y_* = \frac{A}{B}.$$

Clearly, this solution is relevant ecologically since both populations are positive.

We examine the stability of both fixed points. In the first case, with $x_* = y_* = 0$, the Jacobian is

$$\boldsymbol{A} = \begin{pmatrix} A & 0 \\ 0 & -C \end{pmatrix},$$

with eigenvalues $\lambda_1 = A$ and $\lambda_2 = -C$. Hence, one of the eigenvalues has positive real part, corresponding to instability. In the second case, with $x_* = C/D$ and $y_* = A/B$ we have

$$\begin{aligned} \boldsymbol{A} &= \begin{pmatrix} A - By_* & -Bx_* \\ Dy_* & -C + Dx_* \end{pmatrix}, \\ &= \begin{pmatrix} 0 & -BC/D \\ DA/B & 0 \end{pmatrix}, \end{aligned}$$

with eigenvalues

$$\lambda = \pm \mathrm{i}\sqrt{AC}.$$

In this case, the expected behaviour of solutions that remain near to the fixed point $(x_*, y_*) = (C/D, A/B)$ is periodic.

For simple two-dimensional non-linear systems, the full global behaviour (i.e. far from fixed points) can be obtained from the following arguments. First, take the quotient between $dy/dt$and dx/dt to obtain

$$\frac{dy/dt}{dx/dt} = \frac{dy}{dx}.$$

$$\frac{dy}{dx} = \frac{G(x,y)}{F(x,y)}.$$

In the case of the Rabbit/Fox system this leads to the slope equation

$$\frac{dy}{dx} = \frac{-C\,y + D\,x\,y}{A\,x - B\,x\,y}.$$

This is now an ODE in a single variable, which can be solved by the method of separation of variables.

First, rewrite the ODE as

$$\frac{dy}{dx} = \frac{y}{x}\left(\frac{-C + Dx}{A - By}\right).$$

Formally multiply up both sides by dx to obtain

$$(A - By)\,\frac{dy}{y} = (-C + Dx)\,\frac{dx}{x},$$

hence

$$A(dy/y) - B\,dy = -C(dx/x) + D\,dx.$$

Integrate:

$$A\ln(y) + C\ln(x) = By + Dx + \mathrm{Const}..$$

Hence,

$$\ln(y^A x^C) = By + Dx + \mathrm{Const}..$$

and so finally,

$$y^A x^C \mathrm{e}^{-By}\mathrm{e}^{-Dx} = k,$$

where k is a constant. By plotting the different level sets of of the function

$$\Phi(x,y) = y^A x^C \mathrm{e}^{-By}\mathrm{e}^{-Dx},$$

the trajectories of the predator-prey dynamics in the xy plane can be mapped out – e.g. Figure 4.1.

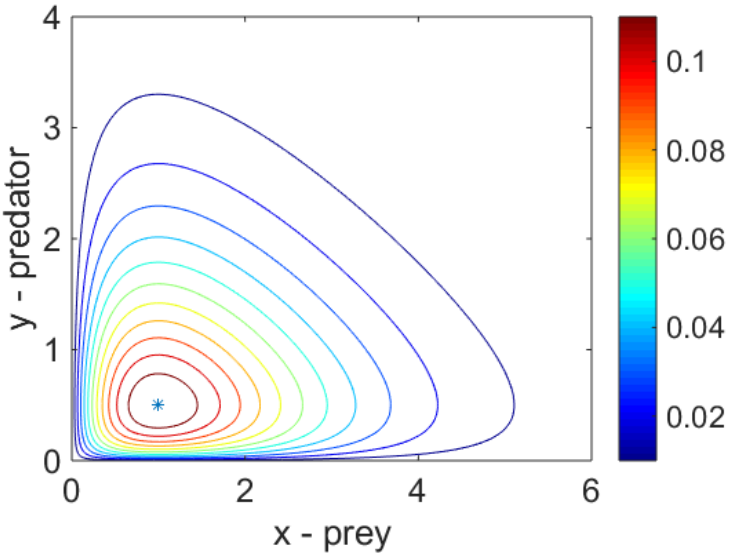


Figure 4.1: Constant of the motion (hence, trajectories) of a predator-prey system with $A = 2/3$, $B = 4/3$, $C = D = 1$.

The trajectories are closed. Notice in this figure that the nonzero fixed point is shown at $(x_*, y_*) = (1, 0.5)$. Nearby trajectories are almost circular.

When a function such as $\Phi(x, y)$ can be found such that $\Phi(x, y) = \text{const} = k$ on trajectories, k is called a **constant of the motion**. Obviously, these can be very useful, as they enable one to solve the nonlinear problem completely.

Chapter 5

Introduction to Fourier Series

5.1 Outline

5.2 Review of inner products

We start with the general definition of the inner product, valid for a general vector space V. You can think of V as being just $\mathbb{R}^n$, although the general definition here is good, because it will enable us later on to extend the notion of the inner product to spaces of functions.

Definition 5.1 *Let V be a finite-dimensional real vector space. A scalar product on V is a map*

$$V \times V \to \mathbb{R},$$
$$(\boldsymbol{x}, \boldsymbol{y}) \to \langle \boldsymbol{x}, \boldsymbol{y} \rangle,$$

that is bilinear:

1. $\langle \lambda \boldsymbol{x} + \mu \boldsymbol{y}, \boldsymbol{z} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \mu \langle \boldsymbol{y}, \boldsymbol{z} \rangle$,

2. $\langle \boldsymbol{x}, \lambda \boldsymbol{y} + \mu \boldsymbol{z} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \mu \langle \boldsymbol{x}, \boldsymbol{z} \rangle$,

for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and $\lambda, \mu \in \mathbb{R}$.

To make contact with familiar ideas, we consider the definition of a scalar product in the case for

$V = \mathbb{R}^n$. As such, we consider the usual basis on $\mathbb{R}^n$:

$$\begin{aligned} \boldsymbol{e}_1 &= (1, 0, \cdots, 0), \\ \boldsymbol{e}_2 &= (0, 1, \cdots, 0), \\ \vdots &= \vdots \\ \boldsymbol{e}_n &= (0, 0, \cdots, 1). \end{aligned}$$

We define the **dot product** of two basis vectors:

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij},$$

where δ_{ij} is the Kronecker delta. Extend this definition by linearity two arbitrary vectors in $\mathbb{R}^n$:

$$\begin{aligned} \boldsymbol{a} &= a_1\boldsymbol{e}_1 + \cdots + a_n\boldsymbol{e}_n, \\ \boldsymbol{b} &= b_1\boldsymbol{e}_1 + \cdots + b_n\boldsymbol{e}_n, \\ \boldsymbol{a} \cdot \boldsymbol{b} &= (a_1\boldsymbol{e}_1 + \cdots + a_n\boldsymbol{e}_n) \cdot (b_1\boldsymbol{e}_1 + \cdots + b_n\boldsymbol{e}_n), \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\delta_{ij}b_j, \\ &= a_1b_1 + \cdots a_nb_n. \end{aligned}$$

Note that the dot-product is **positive** because

$$\boldsymbol{a} \cdot \boldsymbol{a} = a_1^2 + a_2^2 + \cdots + a_n^2;$$

it is definite because the only way that $\boldsymbol{a} \cdot \boldsymbol{a} = 0$ can be satisfied is if $\boldsymbol{a} = 0$. Thus, the dot product is called **positive-definite**. For these reasons, we define the **length** of a vector $\boldsymbol{a} \in \mathbb{R}^n$ as follows:

$$\begin{aligned} \boldsymbol{a} &= a_1\boldsymbol{e}_1 + \cdots + a_n\boldsymbol{e}_n, \\ \text{Length of } \boldsymbol{a} \equiv |\boldsymbol{a}| := \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}} &= \sqrt{a_1^2 + \cdots + a_n^2}. \end{aligned}$$

Theorem 5.1 *The scalar product satisfies the Cauchy–Schwartz inequality:*

$$|\boldsymbol{a} \cdot \boldsymbol{b}| \leq |\boldsymbol{a}||\boldsymbol{b}|.$$

Proof: Consider

$$\phi(\lambda) := (\lambda\boldsymbol{a} + \boldsymbol{b})^2 = (\lambda\boldsymbol{a} + \boldsymbol{b}) \cdot (\lambda\boldsymbol{a} + \boldsymbol{b}).$$

We have $\phi(\lambda) \geq 0$ and

$$\phi(\lambda) = \lambda^2|\boldsymbol{a}|^2 + 2\lambda\boldsymbol{a} \cdot \boldsymbol{b} + |\boldsymbol{b}|^2.$$

This is a quadratic function in λ, with roots

$$\lambda_{\pm} = \frac{\boldsymbol{a}\cdot\boldsymbol{b} \pm \sqrt{(\boldsymbol{a}\cdot\boldsymbol{b})^2 - |\boldsymbol{a}|^2|\boldsymbol{b}|^2}}{|\boldsymbol{a}|^2}.$$

But $\phi(\lambda) \geq 0$, the quadratic function has at most one real root, so

$$(\boldsymbol{a}\cdot\boldsymbol{b})^2 - |\boldsymbol{a}|^2|\boldsymbol{b}|^2 \leq 0,$$

or

$$|\boldsymbol{a}\cdot\boldsymbol{b}| \leq |\boldsymbol{a}||\boldsymbol{b}|,$$

as required.

Remark 5.1 *The Cauchy–Schwartz inequality is true for any scalar product with the positive-definite property* $\langle \vec{,}\boldsymbol{x}\rangle \geq 0$.

Since $|\boldsymbol{a}\cdot\boldsymbol{b}| \leq |\boldsymbol{a}||\boldsymbol{b}|$, we define the **angle between vectors** $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\cos\theta = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}, \qquad 0 \leq \theta \leq \pi.$$

This in turn enables us to specify when two vectors in $\mathbb{R}^n$ are orthogonal:

Definition 5.2 *Two vectors are* **orthogonal** *if angle between them is* $\pi/2$ *(if their dot product is zero):*

$$\boldsymbol{a}\cdot\boldsymbol{b} = 0.$$

These definitions now enable us to define a very special kind of **basis** for $\mathbb{R}^n$:

Definition 5.3 *A basis* $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n$ *for* $\mathbb{R}^n$ *is called* **orthonormal** *if*

$$\boldsymbol{f}_i \cdot \boldsymbol{f}_j = \delta_{ij}.$$

The components of an arbitrary vector $\boldsymbol{a}$ *with respect to this basis are given by*

$$\boldsymbol{a} = \beta_1\boldsymbol{f}_1 + \cdots + \beta_n\boldsymbol{f}_n,$$

and

$$\beta_i = \boldsymbol{a}\cdot\boldsymbol{f}_i.$$

Remark 5.2 *The set of the usual basis vectors* $\{\boldsymbol{e}_i\}_{i=1}^n$ *for* $\mathbb{R}^n$ *is an orthonormal basis for* $\mathbb{R}^n$*. However, it is only one of many possible orthonormal bases for* $\mathbb{R}^n$*. Any two orthonormal bases for* $\mathbb{R}^n$ *are however connected by a* **rotation matrix***:*

$$\boldsymbol{f}_i = R\boldsymbol{e}_i, \qquad R \in \mathbb{R}^{n\times n}, \qquad R^T R = \mathbb{I}.$$

5.3 Spaces of functions

Recall, the set V_Ω of all real-valued functions,

$$V_\Omega = \{f | f : (\Omega \subset \mathbb{R}) \to \mathbb{R}\}$$

is a vector space, with pointwise operations of addition and scalar multiplication, and Ω is an interval (open or closed) on $\mathbb{R}$.

Definition 5.4 *The set*

$$L^2(\Omega) = \left\{ f \in V_\Omega \middle| \int_\Omega |f(x)|^2 \mathrm{d}x < \infty \right\}$$

is a vector subspace of V_Ω *called the space of* **square-integrable functions**.

Theorem 5.2 *The map*

$$\begin{aligned}(\cdot|\cdot) : L^2(\Omega) \times L^2(\Omega) &\to \mathbb{R},\\ (f, g) &\to \int_\Omega f(x)g(x)\mathrm{d}x\end{aligned}$$

is a scalar product on the vector space $L^2(\Omega)$*.*

The proof is easy: all you do is show bi-linearity.

Definition 5.5 *Let* $f \in L^2$*. Then the* **length of the function** f *is denoted by* $\|f\|_2$*, and is defined by*

$$\|f\|_2^2 := \langle f, f\rangle = \int_\Omega |f(x)|^2 \mathrm{d}x.$$

Definition 5.6 *Let* $f, g \in L^2$*. These functions are* **orthogonal** *if*

$$\langle f, g\rangle = \int_\Omega f(x)g(x)\mathrm{d}x = 0.$$

As an example of 'the length of a function', let $\Omega = [-\pi, \pi]$. The length of the function $\sin(x)$ is given by

$$\| \sin(x) \|_2^2 = \int_{-\pi}^{\pi} \sin^2(x) \mathrm{d}x = \pi.$$

The functions $\sin(x)$ and $\cos(x)$ are orthogonal because

$$\int_{-\pi}^{\pi} \sin(x) \cos(x) = 0.$$

We now focus on a further important example: We consider a vector space of functions on $\Omega = [-\pi, \pi]$ defined as follows:

$$F_n = \left\{ f(x) \in V_\Omega \middle| f(x) = a_0 + \sum_{i=1}^{n} \left[a_i \cos(ix) + b_i \sin(ix) \right] \right\},$$

where the a_i's and b_i's are ordinary real numbers. In other words,

$$F_n \subset V_\Omega, \qquad F_n = \mathcal{S}\Big(1, \cos(x), \cdots, \cos(nx), \sin(x), \cdots, \sin(nx) \Big)$$

Thus, a typical element in F_n is

$$f(x) = a_0 + \sum_{i=1}^{n} \left[a_i \cos(ix) + b_i \sin(ix) \right].$$

This is a square-integrable function, because

$$\langle f, f \rangle = \pi \left[2a_0^2 + \sum_{i=1}^{n} (a_i^2 + b_i^2) \right].$$

We prove this statement now:

$$\langle f, f \rangle = \int_{-\pi}^{\pi} \mathrm{d}x \left[a_0 + \sum_{j=1}^{n} \left[a_j \cos(jx) + b_j \sin(jx) \right] \right] \left[a_0 + \sum_{j=1}^{n} \left[a_j \cos(jx) + b_j \sin(jx) \right] \right],$$

$$\begin{aligned} \langle f, f \rangle = 2\pi a_0^2 + 2a_0 \int_{-\pi}^{\pi} \mathrm{d}x \sum_{j=1}^{n} \left[a_j \cos(jx) + b_j \sin(jx) \right] \\ + \int_{-\pi}^{\pi} \mathrm{d}x \left[\sum_{j=1}^{n} \left[a_j \cos(ix) + b_j \sin(jx) \right] \right] \left[\sum_{j=1}^{n} \left[a_j \cos(jx) + b_j \sin(jx) \right] \right], \end{aligned}$$

$$\langle f, f\rangle = 2\pi a_0^2 + 2a_0 \int_{-\pi}^{\pi} \mathrm{d}x \sum_{j=1}^{n} \left[a_j \cos(jx) + b_j \sin(jx)\right]$$
$$+\int_{-\pi}^{\pi} \mathrm{d}x \sum_{j=1}^{n}\sum_{k=1}^{n} \left[a_j a_k \cos(jx)\cos(kx) + a_j b_k \cos(jx)\sin(kx) + b_j a_k \cos(jx)\cos(kx) + b_j b_k \sin(jx)\sin(kx)\right],$$

Take the summation signs outside the integrals:

$$\langle f, f\rangle = 2\pi a_0^2 + 2a_0 \sum_{j=1}^{n} \int_{-\pi}^{\pi} \mathrm{d}x \left[a_j \cos(jx) + b_j \sin(jx)\right]$$
$$+\sum_{j=1}^{n}\sum_{k=1}^{n} \int_{-\pi}^{\pi} \mathrm{d}x \left[a_j a_k \cos(jx)\cos(kx) + a_j b_k \cos(jx)\sin(kx) + b_j a_k \sin(jx)\cos(kx) + b_j b_k \sin(jx)\sin(kx)\right].$$

But

$$\int_{-\pi}^{\pi} \sin(px) = \int_{-\pi}^{\pi} \cos(px) = 0, \qquad p = \{1, 2, \cdots\}$$

hence

$$\langle f, f\rangle = 2\pi a_0^2$$
$$+\sum_{j=1}^{n}\sum_{k=1}^{n} \int_{-\pi}^{\pi} \mathrm{d}x \left[a_j a_k \cos(jx)\cos(kx) + a_j b_k \cos(jx)\sin(kx) + b_j a_k \sin(jx)\cos(kx) + b_j b_k \cos(jx)\cos(kx)\right].$$

Let's tackle the other integrals:

$$\int_{-\pi}^{\pi} \sin(jx)\cos(kx) = \tfrac{1}{2}\int_{-\pi}^{\pi} \left[\sin((j+k)x) + \sin((j-k)x)\right] \mathrm{d}x.$$

If $j = k$, then this is

$$\tfrac{1}{2}\int_{-\pi}^{\pi} \sin(2kx)\mathrm{d}x = -\frac{1}{4k}\left[\cos(\pi k) - \cos(-\pi k)\right] = 0;$$

otherwise it is

$$-\tfrac{1}{2}\left[\frac{\cos((j+k)x)}{j+k} + \frac{\cos((j-k)x)}{j-k}\right]_{-\pi}^{\pi} = 0.$$

Thus, the cross terms vanish in the sum for $(f|f)$:

$$\langle f, \rangle = 2\pi a_0^2 + \sum_{j=1}^{n}\sum_{k=1}^{n} \int_{-\pi}^{\pi} \mathrm{d}x \left[a_j a_k \cos(jx)\cos(kx) + b_j b_k \cos(jx)\cos(kx)\right].$$

Let's tackle the first term:

$$\int_{-\pi}^{\pi} \cos(jx)\cos(kx) = \tfrac{1}{2}\int_{-\pi}^{\pi} [\cos((j-k)x) + \cos((j+k)x)]\,\mathrm{d}x.$$

If $j = k$ this is

$$\tfrac{1}{2}\int_{-\pi}^{\pi} [1 + \cos(2kx)] = \pi + \frac{1}{4k}\left[\sin(\pi k) - \sin(-\pi k)\right] = \pi + 0,$$

otherwise it is

$$\tfrac{1}{2}\left[\frac{\sin((j-k)x)}{j-k} + \frac{\sin((j+k)x)}{j+k}\right]_{-\pi}^{\pi} = 0.$$

In other words,

$$\int_{-\pi}^{\pi} \cos(jx)\cos(kx) = \pi\delta_{jk}.$$

The sine integral is identical. Thus,

$$\langle f, f\rangle = 2\pi a_0^2 + \sum_{j=1}^{n}\sum_{k=1}^{n} \pi\delta_{jk}\left[a_j a_k + b_j b_k\right] = 2\pi a_0^2 + \pi\sum_{j=1}^{n}[a_j^2 + b_j^2].$$

Summarizing, we have shown:

- $\int_{-\pi}^{\pi} \mathrm{d}x \sin(jx) = \int_{-\pi}^{\pi} \mathrm{d}x \cos(jx) = 0$,
- $\int_{-\pi}^{\pi} \mathrm{d}x \sin(jx)\cos(jx) = 0$,
- $\int_{-\pi}^{\pi} \mathrm{d}x \sin(jx)\sin(kx) = \int_{-\pi}^{\pi} \mathrm{d}x \cos(jx)\cos(kx) = \pi\delta_{jk}$.

Thus,

$$\begin{aligned}
f_0(x) &= \tfrac{1}{\sqrt{2\pi}},\\
f_1(x) &= \tfrac{1}{\sqrt{\pi}}\cos(x),\\
&\vdots \quad\quad \vdots\\
f_n(x) &= \tfrac{1}{\sqrt{\pi}}\cos(nx),\\
g_1(x) &= \tfrac{1}{\sqrt{\pi}}\sin(x),\\
&\vdots \quad\quad \vdots\\
g_n(x) &= \tfrac{1}{\sqrt{\pi}}\sin(nx)
\end{aligned}$$

are linearly independent, span $F_n(\Omega)$ and form an **orthonormal basis** for the space. By definition

/ result 5.3, an arbitrary element $f(x)$ in $F_n(\Omega)$ has the representation

$$f(x) = \sum_{j=0}^{n} \langle f, f_i \rangle f_j(x) + \sum_{i=1}^{n} \langle f, g_j \rangle g_j(x).$$

$F_n(\Omega)$ is therefore a $2n+1$-dimensional real vector space.

5.4 Fourier series: the limit $n \to \infty$

Letting $n \to \infty$ in the definition of $F_n(\Omega)$, we obtain the set of all functions

$$F_\infty(\Omega) = \left\{ f(x) \in V_\Omega \middle| \begin{array}{c} f(x) = a_0 + \sum_{j=1}^{\infty} \left[a_j \cos(jx) + b_j \sin(jx) \right]; \\ \text{the Fourier series converges to its generating function } f \end{array} \right\}.$$

The coefficients a_0, a_j, and b_j can be obtained by taking the scalar product of $f(x)$ with the basis functions

$$\begin{aligned} f_0(x) &= \tfrac{1}{\sqrt{2\pi}}, \\ f_j(x) &= \tfrac{1}{\sqrt{\pi}} \cos(jx), \\ g_j(x) &= \tfrac{1}{\sqrt{\pi}} \sin(jx), \end{aligned}$$

where now $j \in \{1, 2, \cdots\}$ ranges over all positive integers:

$$\begin{aligned} a_j &= \langle f, f_j \rangle = \int_{-\pi}^{\pi} f(x) f_j(x) \mathrm{d}x, \\ b_j &= \langle f, g_j \rangle = \int_{-\pi}^{\pi} f(x) g_j(x) \mathrm{d}x. \end{aligned}$$

A series of the form $a_0 + \sum_{j=1}^{\infty} \left[a_j \cos(jx) + b_j \sin(jx) \right]$ is called a **Fourier series** and the coefficients a_j and b_j are called **Fourier coefficients**.

Provided $f(x)$ is square integrable (i.e. in $f \in L^2$), its Fourier coefficients can be calculated. It does not follow, however, that the corresponding Fourier series converges to $f(x)$. That is, the following diagram is not commutative (Figure 5.4):

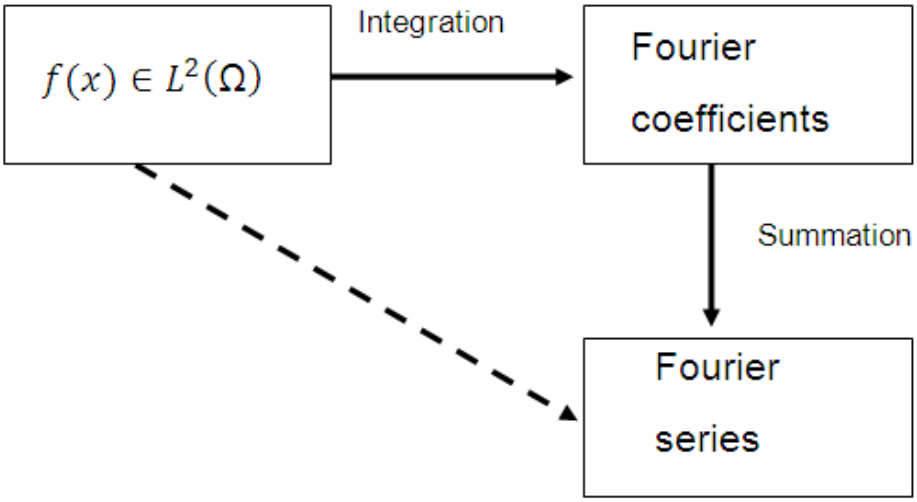


Figure 5.4: The creation of a Fourier series from a generating function and the creation of functions from Fourier series are not **always** inverses: this is not a commutative diagram.

To ensure that this diagram is commutative, that is, to ensure that the Fourier series generated by a function's Fourier coefficients converges to the function itself, we need some stronger conditions than square-integrability.

Definition 5.7 *A function $f(x) \in L^2$, $\Omega = [a, b]$ is called piecewise smooth if there is a partition of $[a, b]$,*

$$a = x_0 < x_1 < x_2 < \cdots x_n = b,$$

such that f has a continuous derivative (i.e. C^1) on each **closed** *subinterval $[x_m, x_{m+1}]$.*

Example:

1. A function that is C^1 on $[a, b]$ is piecewise smooth on $[a, b]$.
2. The function
$$f(x) = \begin{cases} 2x, & 0 \le x \le \frac{1}{2}, \\ \frac{1}{2}, & \frac{1}{2} < x \le 1 \end{cases}$$
is piecewise smooth on $[0, 1]$ but is not continuous on $[0, 1]$.
3. The function
$$f(x) = |x|$$

is both continuous and piecewise smooth on $[-1,1]$, despite $f'(x)$ not being defined at $x=0$. This is because we partition $[-1,1]$ into two subintervals $[-1,0]$ and $[0,1]$. When worrying about $f'(x)$ near $x=0$, note that on $[0,1]$ we only care about the right limit,

$$f'(0^+) = \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{f(0+\epsilon)-f(0)}{\epsilon} = 1,$$

while for $[-1,0]$ we care only about the left limit,

$$f'(0^-) = \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \frac{f(0-\epsilon)-f(0)}{\epsilon} = -1.$$

4. The function $f(x) = |x|^{1/2}$ is continuous on $[-1,1]$ but is not piecewise smooth on $[-1,1]$, since $f'(0^+)$ and $f'(0^-)$ do not exist.

Theorem 5.3 *Let $f \in L^2((-\pi,\pi))$ be piecewise smooth on the closed interval $[-\pi,\pi]$ and continuous on the open interval $(-\pi,\pi)$. Then, the Fourier series associated with f converges for all $x \in [-\pi,\pi]$ and converges to the generating function $f(x)$ for all $x \in (-\pi,\pi)$.*

We don't prove this result here – we instead use it in what follows.

Now suppose instead that $f \in L^2((-\pi,\pi))$ is piecewise smooth on the closed interval $[-\pi,\pi]$ and piecewise continuous on $(-\pi,\pi)$, with a single discontinuity at $x=a$. Then the Fourier series converges to $f(x)$ on $(-\pi,\pi)$, except at $x=a$, where it converges to

$$\frac{f(a-0)+f(a+0)}{2}.$$

5.5 A numerical example

Consider the function

$$f(x) = x^2$$

on the interval $[-\pi,\pi]$. This is a continuous function with continuous derivative. Thus, the function converges to its Fourier series on the interior of the interval, $(-\pi,\pi)$:

$$x^2 = a_0 f_0(x) + \sum_{i=1}^{\infty} \left[a_i f_i(x) + b_i g_i(x)\right],$$

where

$$\begin{aligned} a_0 &= \int_{-\pi}^{\pi} x^2 f_0(x), \qquad f_0(x) = \tfrac{1}{\sqrt{2\pi}}, \\ a_j &= \int_{-\pi}^{\pi} x^2 f_j(x), \qquad f_j(x) = \tfrac{1}{\sqrt{\pi}} \cos(jx), \\ b_j &= \int_{-\pi}^{\pi} x^2 g_j(x), \qquad g_j(x) = \tfrac{1}{\sqrt{\pi}} \sin(jx), \end{aligned}$$

Now

$$\begin{aligned} a_0 &= \tfrac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} x^2 \mathrm{d}x, \\ &= \tfrac{1}{\sqrt{2\pi}} \tfrac{2}{3} \pi^3, \\ a_j &= \tfrac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} x^2 \cos(jx) \mathrm{d}x, \\ &= \tfrac{1}{\sqrt{\pi}} \left[\frac{2x \cos(jx)}{j^2} + \frac{(-2 + j^2 x^2) \sin(jx)}{j^3} \right]_{-\pi}^{\pi}, \\ &= \tfrac{1}{\sqrt{\pi}} \frac{4\pi(-1)^j}{j^2}, \\ b_i &= \tfrac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} x^2 \sin(jx) \mathrm{d}x = 0. \end{aligned}$$

Hence,

$$\begin{aligned} x^2 &= \left(\tfrac{1}{\sqrt{2\pi}} \tfrac{2}{3} \pi^3 \right) \tfrac{1}{\sqrt{2\pi}} + \sum_{j=1}^{\infty} \left(\tfrac{1}{\sqrt{\pi}} \frac{4\pi(-1)^j}{j^2} \right) \left(\tfrac{1}{\sqrt{\pi}} \cos(jx) \right), \\ &= \tfrac{\pi^2}{3} + 4 \sum_{j=1}^{\infty} \tfrac{(-1)^j}{j^2} \cos(jx). \end{aligned}$$

Now, Theorem 5.3 guarantees that the Fourier series converges to its generating function on $(-\pi, \pi)$. The only reason why such a result cannot be extended **in general** to the closed interval $[-\pi, \pi]$ is because of possible discontinuities at the boundary points (here the comment below the theorem would apply). However, we can view $f(x) = x^2$ as a continuous function on all of $\mathbb{R}$, so long as we enforce the periodicity constraint (See Fig. 5.1). Thus, the Fourier series will converge to its generating function everywhere – including on the boundary points $x = \pm\pi$. We therefore set $x = \pi$

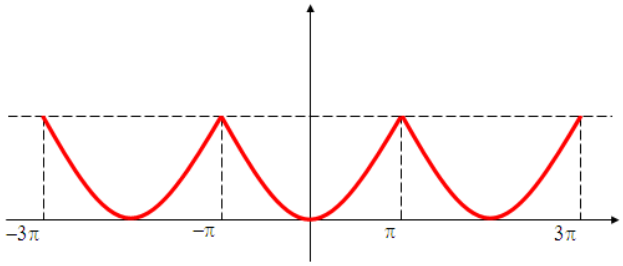


Figure 5.1: The function $f(x) = x^2$ viewed as a 2π-periodic function on all of $\mathbb{R}$.

in the Fourier series:

$$\begin{aligned} x^2 &= \tfrac{\pi^2}{3} + 4\sum_{j=1}^{\infty} \tfrac{(-1)^j}{j^2}\cos(jx), \\ \pi^2 &= \tfrac{\pi^2}{3} + 4\sum_{j=1}^{\infty} \tfrac{(-1)^j}{j^2}(-1)^j, \\ \tfrac{2}{3}\pi^2 &= 4\sum_{j=1}^{\infty} \tfrac{1}{j^2}, \\ \tfrac{1}{6}\pi^2 &= \sum_{j=1}^{\infty} \tfrac{1}{j^2}, \end{aligned}$$

a result first proved by Euler.

A second example

Here, we point out a situation where convergence of the Fourier series to its generating function cannot be extended to the boundary points: $f(x) = x$ – See Figure 5.2. At $x = \pi$, Theorem 5.3

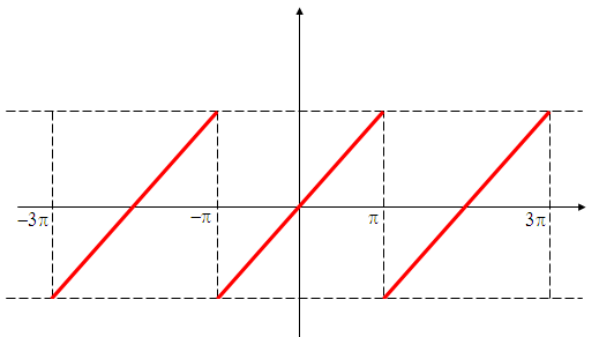


Figure 5.2: The function $f(x) = x$ viewed as a 2π-periodic function on all of $\mathbb{R}$: it possesses jump discontinuities at the boundary points.

tells us that the Fourier series of $f(x)$ will converge to

$$\frac{f(\pi+) + f(\pi-)}{2} = \frac{1-1}{2} = 0.$$

5.6 Term-by-term differentiation / integration of Fourier series

Suppose that $f(x)$ converges everywhere on the interval $[-\pi, \pi]$ to its Fourier series, and that $f(x)$ is differentiable on $(-\pi, \pi)$, such that

$$f(x) = a_0 + \sum_{j=1}^{\infty} [a_j f_j(x) + b_j g_j(x)].$$

We remark very briefly that operations such as

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}x} &= \sum_{j=1}^{\infty} [a_j \frac{\mathrm{d}f_j}{\mathrm{d}x} + b_j \frac{\mathrm{d}g_j}{\mathrm{d}x}], \\ \int f(x)\,\mathrm{d}x &= a_0 x + \sum_{j=1}^{\infty} [a_j \left(\int f_j\,\mathrm{d}x \right) + b_j \left(\int g_j\,\mathrm{d}x \right)] + \text{Const.} \end{aligned}$$

are not always legitimate. Indeed, there are very precise theorems that rule in/out such term-by-term integration and differentiation. There is not time to look into these theorems in this module – for our purposes, we work on the basis that such term-by-term operations are legitimate until we have reason to assume the contrary. This will be very important in the next chapter when we look at series solutions of various partial differential equations.

Chapter 6

Introduction to PDEs

Overview

We formulate the theory of diffusion / heat conduction. Using Fourier-like series, we construct the solution of this equation.

6.1 Physical background

In this module, the partial differential equation we focus on is the diffusion / heat equation. Physically, this describes two equivalent pheonomena: the diffusion of particles via Brownian motion, or the flow of heat from hotter to colder regions via conduction. For the present discussion we demonstrate how to set up the physical model for diffusion of particles. In this context, let $u(x,t)$ be the concentration of particles undergoing Brownian motion, or the diffusion of heat in a metal rod (say). It is an evolutionary equation, so it contains a partial derivative with respect to time, $\partial u/\partial t$. Because the total number of particles

$$\int_\Omega u(x,t)\mathrm{d}x$$

is conserved (here $\Omega = (a,b)$ is the spatial domain of interest), it must be a **flux-conservative equation**:

$$\frac{\partial u}{\partial t} + \frac{\partial J}{\partial x} = 0,$$

where J is the flux of particles (or the heat flux). Such an equation manifestly conserves the total particle number, for suitable boundary conditions:

$$\begin{aligned}\frac{d}{dt}\int_\Omega u(x,t)\mathrm{d}x &= \int_\Omega \mathrm{d}x\frac{\partial u}{\partial t},\\ &= -\int_\Omega \mathrm{d}x\frac{\partial J}{\partial x},\\ &= -\left[J(b,t)-J(a,t)\right].\end{aligned}$$

Thus, if $J(b,\cdot)-J(a,\cdot)=0$, the total particle number is conserved. Now, we focus on the flux. If the flow of particles is proportional to minus the concentration gradient, particles will flow from regions of high concentration to regions of low concentration:

$$J \propto -\frac{\partial u}{\partial x}.$$

This is called **Fick's Law of diffusion** (call the constant of proportionality D). Thus,

$$J=-D\frac{\partial u}{\partial x}, \qquad \frac{\partial u}{\partial t}=-\frac{\partial J}{\partial x},$$

or

$$\frac{\partial u}{\partial t}=D\frac{\partial^2 u}{\partial x^2}. \tag{6.1}$$

This is the celebrated **diffusion equation**. Finally, let's look at the physical units of the proportionality factor D. It is,

$$\frac{[\text{Concentration}]}{[\text{Time}]}=[D]\,\frac{[\text{Concentration}]}{[\text{Length}]^2},$$

hence

$$D=\frac{[\text{Length}]^2}{[\text{Time}]}.$$

Boundary conditions

We will solve the diffusion equation on the interval $\Omega=(0,L)$. We pose the following initial condition:

$$u(x,t=0)=f(x), \qquad 0\le x\le L.$$

Depending on the physical scenario, we will use one or the other of the following boundary conditions:

- Dirichlet condition: $u(x=0,t)=u(x=L,t)=0, \qquad t>0.$
- Neuman condition: $\partial_x u(x=0,t)=\partial_x u(x=L,t)=0, \qquad t>0.$

The first choice (the Dirichlet condition) corresponds to fixing u at the boundary. This corresponds to a model of temperature in a metal rod, whose temperature is fixed at the end points. The second choice (Neuman / no-flux condition) corresponds to fixing the derivative of $u(x,t)$ at the boundaries $x = 0$ and $x = L$. This corresponds to a model of Brownian particles with no flow of particles through the boundaries.

6.2 Separation of variables

For the rest of this section we work with Dirichlet boundary conditions, and we now proceed to solve the PDE. We start by noting that the diffusion equation is linear in the concentration $u(x,t)$, and the only coefficient in the equation is D, which is constant. When faced with a **linear, constant-coefficient** partial-differential equation, the following **separation of variables** procedure works: We make the following trial solution:

$$u(x,t) = X(x)T(t).$$

Substituting this ansatz into the PDE gives a set of ODEs – a much simpler problem:

$$X(x)\frac{dT}{dt} = DT(t)\frac{d^2X}{dx^2}.$$

Now divide out by $X(x)T(t)$. We obtain

$$\frac{1}{T}\frac{dT}{dt} = \frac{D}{X}\frac{d^2X}{dx^2} \quad (= -\lambda D).$$

But now the LHS is a function of t alone and the RHS is a function of x alone. The only way for this relation to be satisfied is if

$$\text{LHS} = \text{RHS} = \text{Const.} := -\lambda D.$$

Let us also substitute the trial solution into the boundary and initial conditions:

$$\begin{aligned} \text{Initial condition:} \qquad & u(x, t=0) = X(x)T(0) = f(x), \qquad 0 < x < L, \\ \text{Boundary condition:} \qquad & T(t)X(0) = T(t)X(L) = 0, \qquad t > 0. \end{aligned}$$

Solving for $X(x)$

Focusing on the $X(x)$-equations, we have:

$$\begin{aligned} \frac{1}{X}\frac{d^2X}{dx^2} &= -\lambda, \qquad 0 < x < L, \\ X(0) &= X(L) = 0. \end{aligned}$$

Equation in the **bulk** $0 < x < L$:

$$\frac{d^2X}{dx^2} + \lambda X = 0, \tag{6.2}$$

Different possibilities for λ:

1. $\lambda = 0$. Then, the solution is $X(x) = Ax + B$, with $dX/dx = A$. However, the BCs specify $X(0) = 0$, hence $B = 0$. They also specify $X(L) = 0$, hence $A = 0$. Thus, only the trivial solution remains, in which we have no interest.

2. $\lambda < 0$. Then, the solution is $X(x) = Ae^{\mu x} + Be^{-\mu x}$, where $\mu = \sqrt{-\lambda}$. The BCs give

   $$A + B = Ae^{\mu L} + Be^{-\mu L} = 0.$$

   Grouping the first two of these equations together gives

   $$A = -B\frac{1 - e^{-\mu L}}{1 - e^{\mu L}}.$$

   But $A + B = 0$, hence

   $$\begin{aligned} B\left[1 - \frac{1 - e^{-\mu L}}{1 - e^{\mu L}}\right] &= 0, \\ B\left[\frac{1 - e^{\mu L} - (1 - e^{-\mu L})}{1 - e^{\mu L}}\right] &= 0, \\ B\left[-e^{\mu L} + e^{-\mu L}\right] &= 0, \\ B\sinh(\mu L) &= 0, \end{aligned}$$

   which has only the trivial solution.

3. Thus, we are forced into the third option: $\lambda > 0$.

Solving Equation (6.2) with $\lambda > 0$ gives

$$X(x) = A\cos(\sqrt{\lambda}x) + B\sin(\sqrt{\lambda}x),$$

with boundary condition

$$A \cdot 1 + B \cdot 0 = A\cos(\sqrt{\lambda}L) + B\sin(\sqrt{\lambda}L) = 0.$$

Hence, $A = 0$. Grouping the second and third equations in this string together therefore gives

$$B\sin(\sqrt{\lambda}L) = 0.$$

Of course, $B = 0$ is a solution, but this is the trivial one. Therefore, we must try to solve

$$\sin(\sqrt{\lambda}L) = 0.$$

This is possible, provided

$$\sqrt{\lambda}L = n\pi, \qquad n \in \{1, 2, \cdots\}.$$

Thus,

$$\lambda = \lambda_n = \frac{n^2\pi^2}{L^2},$$

and

$$X(x) = B_n \sin\left(\frac{n\pi x}{L}\right),$$

where B_n labels the constant of integration.

Solving for $T(t)$

Now substitute $\lambda_n = n^2\pi^2/L^2$ back into the $T(t)$-equation:

$$\frac{1}{T}\frac{dT}{dt} = -\lambda D = -\lambda_n D.$$

Solving give

$$T(t) = T(0)e^{-\lambda_n Dt},$$

or

$$T(t) = T(0)e^{-n^2\pi^2 Dt/L^2}.$$

Putting it all together

Recall the ansatz:

$$u(x,t) = X(x)T(t).$$

Thus, we have a solution

$$X(x)T(t) = T(0)B_n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}.$$

Calling $T(0)B_n := C_n$, this is

$$X_n(x)T_n(t) = C_n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}.$$

The label n is just a label on the solution. However, each $n = 1, 2, \cdots$ produces a different solution, linearly independent of all the others. We can add all of these solutions together to obtain a **general solution** of the PDE:

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} X_n(x)T_n(t), \\ &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}. \end{aligned}$$

We are almost there. However, we still need to take care of the initial condition,

$$\begin{aligned} u(x,t=0) &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right), \\ &= f(x). \end{aligned}$$

But the functions

$$\left\{ \sin\left(\frac{n\pi x}{L}\right) \right\}_{n=1}^{\infty}$$

are orthogonal on $[0, L]$:

$$\begin{aligned} I_{n,m} &= \int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x, \\ &= \frac{L}{\pi} \int_0^{\pi} \sin(ny)\sin(my)\mathrm{d}y, \qquad y = (\pi/L)x. \end{aligned}$$

If $n \neq m$ this is

$$\frac{L}{\pi}\left[\frac{\sin((m-n)x)}{2(m-n)} - \frac{\sin((m+n)x)}{2(m+n)}\right]_0^{\pi} = 0;$$

if $n = m$ it is

$$\frac{L}{\pi}\frac{\pi}{2} = \frac{L}{2},$$

hence $I_{nm} = (L/2)\delta_{nm}$. Thus, consider the IC again:

$$\begin{aligned} u(x, t = 0) &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right), \\ &= f(x). \end{aligned}$$

Multiply both sides by $\sin(m\pi x/L)$ and integrate:

$$\begin{aligned} \int_0^L f(x) \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x &= \int_0^L \sum_{n=1}^{\infty} C_n \sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{n\pi x}{L}\right), \\ &= \sum_{n=1}^{\infty} C_n \int_0^L \sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{n\pi x}{L}\right), \\ &= \sum_{n=1}^{\infty} C_n \frac{L}{2} \delta_{m,n}, \\ &= \frac{C_n L}{2}. \end{aligned}$$

Hence,

$$C_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x,$$

and, substituting back into the general solution, we have

$$\begin{aligned} u(x, t) &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}, \\ &= \sum_{n=1}^{\infty} \left[\frac{2}{L} \int_0^L f(s) \sin\left(\frac{m\pi s}{L}\right) \mathrm{d}s\right] \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}, \end{aligned} \tag{6.3}$$

which is a solution to the diffusion equation that satisfies the boundary and initial conditions.

Remark 6.1 *We note that Equation* (6.3) *is a series solution, and the question can be asked if term-by-term differentiation of the series is legitimate, i.e. to conclude that* $u_t = Du_{xx}$. *The anwer in this case is that such term-by-term maniupatlions is legitimate. But this is topic is beyond the scope of this part of the module.*

6.3 Worked example: Cooling of a rod from a constant initial temperature

Suppose that we use the diffusion equation to model the temperature distribution in a metal rod. Suppose furthermore that the initial temperature distribution $f(x)$ in the rod is constant, i.e. $f(x) = u_0$. We wish to find the solution at later times. From Equation (6.3), we have to work out

$$C_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{m\pi x}{L}\right)\mathrm{d}x, \qquad f(x) = u_0.$$

That is,

$$\begin{aligned}
\frac{2u_0}{L}\int_0^L \sin\left(\frac{m\pi x}{L}\right)\mathrm{d}x &= \frac{2u_0}{\pi}\int_0^\pi \sin(my)\mathrm{d}y, \\
&= -\frac{2u_0}{\pi}\frac{\cos(my))}{m}\Big|_0^\pi, \\
&= -\frac{2u_0}{\pi}\frac{(\cos(m\pi)-\cos(0))}{m}, \\
&= \frac{2u_0}{\pi}\frac{1-(-1)^m}{m}, \\
&= \begin{cases} 0, & m \text{ even}, \\ \frac{4u_0}{m\pi}, & m \text{ odd}. \end{cases}
\end{aligned}$$

Thus,

$$u(x,t) = \frac{4u_0}{\pi}\sum_{m=0}^{\infty}\frac{\sin\left(\frac{(2m+1)\pi x}{L}\right)}{2m+1}e^{-(2m+1)^2\pi^2 Dt/L^2}$$

This solution is plotted in Figure 6.1. The solution is rather odd:

- At time $t = 0$ the solution is simply $u(x, t = 0) = 1$ everywhere in the domain.
- The solution then **instantaneously** adjusts so that $u(x, t > 0) \approx 1$ in the bulk of the domain, and $u(x, t > 0) = 0$ on the boundary.
- Subsequently, the solution decays so that $u(x, t \to \infty) =$ boundary value $= 0$.

6.4 Well-posed problems

The following reasonable conditions ought to be satisfied by a mathematical model of some physical system:

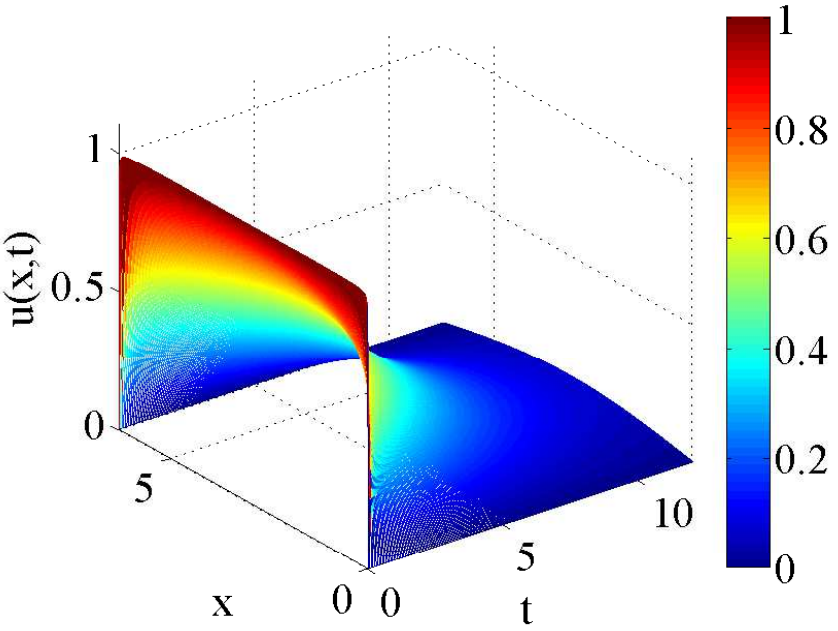


Figure 6.1: Solution for the diffusion equation with $u(x, t = 0) = 1$ and $u(0, t > 0) = u(2\pi, t > 0) = 0$. The model parameters are set $D = u_0 = 1$.

1. **Existence of solutions:** The mathematical model must possess at least one solution. Interpretation: the physical system exists over at least some finite time interval.

2. **Uniqueness of solutions:** For given boundary and initial conditions, the mathematical model has at most one solution. Interpretation: identical initial states of the system lead to the same outcome.

3. **Continuous dependence on parameters:** The solution of the mathematical model depends continuously on the initial conditions and parameters. Interpretation: Small changes in initial conditions or parameters lead to small changes in the outcome.

A model set of equations that satisfies these criteria is called **well posed**. The diffusion equation satisfies these criteria:

- Certainly at least one solution exists – we have constructed it using Fourier series. And, moreover this series can be differentiated term-by-term, to produce a continuously-differntiable solution.

- The continuous dependence on initial parameters is tricky and we do not consider it here. However, you may verify it numerically by looking at two numerical solutions whose initial conditions are 'similar' – the similarity between the two solutions will persist over time.

- Finally, there is the question of uniqueness, which we address below.

The diffusion equation: uniqueness of solutions

Let us define the sets

$$\Omega_t = \{(x,t)|x \in [0,L], t \in (0,\infty)\},$$
$$\overline{\Omega_t} = \{(x,t)|x \in [0,L], t \in [0,\infty)\},$$

and the function space

$$\mathcal{C}^2\left(\overline{\Omega_t}\right) = \left\{ u(x,t) \middle| u_{xx} \text{ is continuous in } \Omega_t \text{ and } u \text{ is continuous in } \overline{\Omega_t} \right\}.$$

The reason for looking at this space is that we want to do integrations over space for each point in time $t \geq 0$, so we need u and u_{xx} to be continuous on the closed set $[0,L]$, for all $t \geq 0$.

We have the following theorem:

Theorem 6.1 *The diffusion equation studied so far, namely*

$$\begin{aligned} &\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \qquad \text{on } (0,L), \\ &u(x,t=0) = f(x), \qquad 0 \leq x \leq L, \\ &u(x=0,t) = u(x=L,t) = 0, \qquad t > 0 \end{aligned}$$

has at most one solution in the space $\mathcal{C}^2\left(\overline{\Omega_t}\right)$.

Proof: Consider two solutions $u_1, u_2 \in \mathcal{C}^2\left(\overline{\Omega_t}\right)$ to the diffusion equation, with identical boundary and initial conditions. Form the difference $v = u_1 - u_2$. Then,

$$\begin{aligned} v_t &= (u_1 - u_2)_t, \\ &= u_{1t} - u_{2t}, \\ &= Du_{1xx} - Du_{2xx}, \\ &= (u_1 - u_2)_{xx}, \\ &= Dv_{xx}, \qquad x \in (0,L), \end{aligned}$$

and the difference of the solutions also satisfies the diffusion equation. Now consider the BCs and ICs:

$$\begin{aligned} \text{BC}: &\qquad v(0,t>0) = u_1(0,t>0) - u_2(0,t>0) = 0, \\ \text{BC}: &\qquad v(L,t>0) = u_1(L,t>0) - u_2(L,t>0) = 0, \\ \text{IC}: &\qquad v(x,t=0) = u_1(x,t=0) - u_2(x,t=0) = 0, \qquad 0 \leq x \leq L. \end{aligned}$$

Let us form the L^2 norm of v and then differentiate it with respect to time:

$$\|v\|_2^2 = \int_\Omega v^2(x,t)\,\mathrm{d}x,$$

where the integral exists because x-derivatives of v up to the second order are continuous on $[0, L]$. Thus,

$$\begin{aligned}
\tfrac{1}{2}\frac{d}{dt}\|v\|_2^2 &= \tfrac{1}{2}\frac{d}{dt}\int_\Omega v(x,t)^2\,\mathrm{d}x,\\
&= \tfrac{1}{2}\int_\Omega \frac{\partial}{\partial t}v^2\,\mathrm{d}x,\\
&= \int_\Omega v v_t\,\mathrm{d}x,\\
&= D\int_\Omega v v_{xx}\,\mathrm{d}x,\\
&= D\int_\Omega \left[\frac{\partial}{\partial x}\left(v v_x\right) - v_x^2\right]\,\mathrm{d}x,\\
&= D\left[v(L)v_x(L) - v(0)v_x(0)\right] - D\int_\Omega v_x^2\,\mathrm{d}x,\\
&= -D\int_\Omega v_x^2,\mathrm{d}x.
\end{aligned}$$

This is an ordinary differential equation in time and can be formally integrated:

$$\|v\|_2^2(t) = \|v\|_2^2(0) - 2D\int_0^t \|v_x\|_2^2(s)\,\mathrm{d}s,$$

hence

$$\|v\|_2^2(t) \le \|v\|_2^2(0) = 0.$$

Hence,

$$\|v\|_2^2(t) \le 0 \implies \|v\|_2^2(t) = 0.$$

The only way for this equation to be satisfied is if

$$v(x,t) = 0,$$

and the solutions agree, $u_1 = u_2$.

6.5 Inhomogeneous boundary conditions

So far, we have solved the diffusion equation

$$\begin{aligned} &\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \qquad \text{on } (0, L), \\ &u(x, t = 0) = f(x), \qquad 0 \leq x \leq L, \\ &u(x = 0, t) = u(x = L, t) = 0, \qquad t > 0. \end{aligned}$$

We have seen how these boundary conditions did not correspond very well to a model for particle diffusion, since there is a net flow of particles through the boundary. However, physical intuition does suggest it is a good model for the cooling of a metal rod, whose end points are held at a fixed temperature. Let us now discuss the different possibilities for boundary conditions in a more systematic way – including the possibility of inhomogeneous boundary conditions.

Dirichlet conditions

The function $u(x, t > 0)$ is specified on the boundaries:

$$\begin{aligned} u(0, t > 0) &= g_1(t), \\ u(L, t > 0) &= g_2(t). \end{aligned}$$

If the functions $g_1 = g_2 = 0$, then we have **homogeneous Dirichlet conditions**:

$$\begin{aligned} u(0, t > 0) &= 0, \\ u(L, t > 0) &= 0 \end{aligned}$$

This is the case we have considered so far.

Neumann conditions

The **derivative** $u_x(x, t > 0)$ is specified on the boundaries:

$$\begin{aligned} u_x(0, t > 0) &= g_1(t), \\ u_x(L, t > 0) &= g_2(t). \end{aligned}$$

If the functions $g_1 = g_2 = 0$, then we have **homogeneous Neumann conditions**, corresponding to **no flux through the boundaries**. This case has been discussed briefly already:

$$\begin{aligned} u_x(0, t > 0) &= 0, \\ u_x(L, t > 0) &= 0. \end{aligned}$$

Mixed conditions

As the name suggests, this set is a mixture of Dirichlet and Neumann conditions:

$$\begin{aligned} \alpha_1 u_x(0, t > 0) + \alpha_2 u(0, t > 0) &= g_1(t), \\ \alpha_3 u_x(L, t > 0) + \alpha_4 u(L, t > 0) &= g_2(t). \end{aligned}$$

Periodic boundary conditions

The function $u(x, t > 0)$ has the same value on either boundary point:

$$u(0, t) = u(L, t), \qquad t > 0.$$

In practice, these are not very realistic boundary conditions but they are used in numerical experiments because they are easy to implement. However, they can be used to mimic an infinite domain, if the periodic length L is made long enough.

Example: Consider the diffusion equation with inhomogeneous Dirichlet conditions

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \qquad \text{on } (0, L),$$

$$\begin{aligned} &u(x, t = 0) = 0, \qquad 0 \le x \le L, \\ &u(x = 0, t > 0) = 0, \\ &u(x = L, t > 0) = u_1, \end{aligned}$$

where u_1 is a constant. Solve for $u(x, t)$.

Separation of variables fails here. To see why, let us valiantly attempt the solution $u(x, t) = X(x)T(t)$. Then, at the boundary $x = L$, we would have $X(L)T(t) = u_1$, which implies that any coefficients in the $X(x)$-solution depend on time – impossible.

Instead, we break up the solution into a bit that solves the diffusion equation with the given BCs (**particular integral**), and a bit that solves that solves the diffusion equation with zero BCs (**homogeneous solution**).

Because the BCs are independent of time, we expect the PI to be independent of time also – call the PI $u_E(x)$:

$$\begin{aligned} u_E''(x) &= 0, \qquad 0 < x < L, \\ u_E(0) &= 0, \\ u_E(L) &= u_1. \end{aligned}$$

This solution to the ODE is $u_E(x) = Ax + B$, and the BCs give $B = 0$ and $AL = u_1$, hence

$$u_E(x) = u_1(x/L).$$

Now write

$$u(x,t) := u_0(x,t) + u_E(x) \implies u_0(x,t) = u(x,t) - u_E(x).$$

where $u(x,t)$ is the solution to the full equation and u_0 is to be determined. By linearity, u_0 solves the diffusion equation:

$$u_{0t} = Du_{0xx}, \qquad 0 < x < L,$$

and

$$\begin{aligned} &u_0(x = 0, t > 0) = 0, \\ &u_0(x = L, t > 0) = u(L,t) - u_E(L,t) = u_1 - u_1 = 0, \\ &u_0(x, t = 0) = u(0,t) - u_E(0,t) = 0 - u_1(x/L) := f(x), \qquad 0 \le x \le L. \end{aligned}$$

Thus, $u_0(x,t)$ satisfies the diffusion equation with homogeneous BCs – it is the **homogeneous solution**.

But we know how to solve such an equation: the solution is

$$u_0(x,t) = \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}.$$

Moreover,

$$\begin{aligned} u_0(x,t=0) &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right), \\ &= u(x,t=0) - u_E(x), \\ &= 0 - u_1(x/L). \end{aligned}$$

Multiply both sides by $\sin(m\pi x/L)$ and integrate over $[0, L]$:

$$\begin{aligned} -(u_1/L)\int_0^L x \sin\left(\frac{n\pi x}{L}\right) \mathrm{d}x &= \sum_{n=1}^{\infty} C_n \int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x, \\ &= \sum_{n=1}^{\infty} C_n (L/2)\delta_{nm}, \\ &= C_m L/2. \end{aligned}$$

In other words,

$$\begin{aligned} C_m &= -\frac{2u_1}{L^2}\int_0^L x \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x, \\ &= -\frac{2u_1}{L^2}\frac{L^2}{\pi^2}\int_0^{\pi} y \sin(my)\, \mathrm{d}y, \\ &= -\frac{2u_1}{\pi^2}\frac{1}{m^2}\left[\sin(my) - my\cos(my)\right]_{y=0}^{y=\pi}, \\ &= -\frac{2u_1}{\pi^2}\frac{1}{m}\left[-\pi\cos(m\pi)\right], \\ &= \frac{2u_1}{m\pi}(-1)^m, \end{aligned}$$

hence

$$C_m = \frac{2u_1}{m\pi}(-1)^m,$$

and

$$u_0(x,t) = \sum_{n=1}^{\infty} \frac{2u_1}{n\pi}(-1)^n \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}.$$

Now put it all together:

$$u(x,t) = u_0(x,t) + u_E(x),$$

hence

$$u(x,t) = \frac{u_1}{L}x + \frac{2u_1}{n\pi}\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \sin\left(\frac{n\pi x}{L}\right) e^{-n^2\pi^2 Dt/L^2}$$

The homogeneous part of the solution is called the **transient part** and it decays to zero, leaving only the particular integral, or equilibrium part.

6.6 Inhomogeneities in the bulk, or source terms

We introduced the diffusion equation as a model for the 'smoothing-out' of concentration gradients over time, where the concentration measures the density of particles:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \qquad x \in (0, L),$$

with suitable boundary and initial conditions. What happens if particles are, at the same time, being injected into the system? Then, we study the equation

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + q, \qquad x \in (0, L),$$

where $q(x,t)$ is a function called the **source**. Dimensionally, we must have

$$\frac{[\text{Concentration}]}{[\text{Time}]} = [q]\,,$$

thus, the source has as its interpretation the **rate** at which matter (or, in other applications, heat) is being injected into the system. In this chapter we solve such equations.

In this section we assume $q = q(x)$ only. For definiteness, we assume the following BCs and ICs:

$$\begin{aligned}
\text{BC} \qquad & u(0, t > 0) = b_1 = \text{Const.}, \\
\text{BC} \qquad & u(L, t > 0) = b_2 = \text{Const.}, \\
\text{IC} \qquad & u(x, t = 0) = f(x), \qquad 0 \le x \le L.
\end{aligned}$$

As in the case of inhomogeneous BCs, we introduce a **particular integral** to soak up the contribution from the source and the boundary conditions:

$$u(x,t) = u_0(x,t) + u_E(x),$$

where

$$\begin{aligned}
u_E''(x) + q(x) &= 0, \qquad 0 < x < L \\
u_E(0) &= b_1, \\
u_E(L) &= b_2,
\end{aligned}$$

This is a simple ODE, which we assume can be solved. Next, we solve for the homogeneous part

$u_0(x,t)$, where

$$u(x,t) = u_0(x,t) + u_E(x) = \text{full solution},$$

and

$$u_0(x,t) = u(x,t) - u_E(x).$$

Study $u_0(x,t)$:

$$\begin{aligned} u_{0t} &= u_t(x,t) - 0, \\ &= Du_{xx} + q(x), \\ &= D\left(u_0 + u_E\right)_{xx} + q(x), \\ &= Du_{0xx} + Du_{Exx} + q(x), \\ &= Du_{0xx}, \end{aligned}$$

with BCs

$$\begin{aligned} \text{BC} \quad & u_0(0, t > 0) = 0, \\ \text{BC} \quad & u_0(L, t > 0) = 0, \\ \text{IC} \quad & u_0(x, t = 0) = f(x) - u_E(x), \qquad 0 \le x \le L. \end{aligned}$$

This is the homogeneous diffusion equation with homogeneous Dirichlet BCs. But we know the solution then:

$$u_0(x,t) = \sum_{n=1}^{\infty} C_n \sin\left(\frac{\pi n x}{L}\right) \mathrm{e}^{-n^2\pi^2 Dt/L^2}.$$

Moreover,

$$\begin{aligned} u_0(x, t=0) &= \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right), \\ &= f(x) - u_E(x). \end{aligned}$$

As before, multiply both sides by $\sin(m\pi x/L)$ and integrate over $[0, L]$:

$$\begin{aligned} \int_0^L \left[f(x) - u_E(x)\right] \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x &= \int_0^L \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x, \\ &= \sum_{n=1}^{\infty} C_n \left(L/2\right) \delta_{m,n}, \end{aligned}$$

hence

$$C_m = \frac{2}{L} \int_0^L \left[f(x) - u_E(x)\right] \sin\left(\frac{m\pi x}{L}\right) \mathrm{d}x.$$

But

$$u(x,t) = u_0(x,t) + u_E(x),$$

hence

$$\begin{aligned} u(x,t) &= u_E(x) + \sum_{n=1}^{\infty} C_n \sin\left(\frac{\pi n x}{L}\right) \mathrm{e}^{-n^2\pi^2 D t/L^2}, \\ C_n &= \frac{2}{L}\int_0^L [f(x) - u_E(x)] \sin\left(\frac{n\pi x}{L}\right) \mathrm{d}x. \end{aligned}$$

Chapter 7

Topics in Ordinary Differential Equations (*)

7.1 Outline

In this section we look at some advanced topics from the theory of Ordinary Differential equations. We then look at periodic motion in the general Dynamical Systems framework. This will pave the way for an analysis of nonlinear oscillations in the next chapter.

- Nonlinear first-order ODEs with exact solutions
- On the impossibility of periodic motion in certain ODE systems
- Conditions for periodic motion in two-equation ODE systems
- Glycolysis models

7.2 Nonlinear first-order ODEs with exact solutions

We start by studying the so-called **Bernoulli ODE**:

$$\frac{dy}{dx} + P(x)y = Q(x)y^n, \qquad n \neq 1. \tag{7.1}$$

We also take $n \neq 0$, since setting $n = 0$ reproduces the linear equation solvable by separation of variables. Thus, we have a term y^n in the equation, with $n \neq 0$ and $n \neq 1$, meaning the equation is **nonlinear**. Unlike almost all other nonlinear ODEs, there is a solution method that enables one to derive an exact solution.

As such, we divide both sides of the equation by y^n to obtain

$$\frac{1}{y^n}\frac{dy}{dx} + P(x)\frac{1}{y^{n-1}} = Q(x). \tag{7.2}$$

Introduce

$$w(x) = \frac{1}{y^{n-1}} = y^{1-n}.$$

We have

$$\frac{dw}{dx} = (1-n)y^{-n}\frac{dy}{dx} = (1-n)\frac{1}{y^n}\frac{dy}{dx},$$

hence

$$\frac{1}{y^n}\frac{dy}{dx} = \frac{1}{1-n}\frac{dw}{dx}.$$

Substitute this into Equation (7.2)

$$\frac{1}{1-n}\frac{dw}{dx} + P(x)w = Q(x), \tag{7.3}$$

or

$$\frac{dw}{dx} + P(x)(1-n)w = Q(x)(1-n).$$

But this is now a linear first-order ODE that can be solved using the integrating-factor technique. We have

$$\mu(x) = \exp\left[(1-n)\int P(x)\,\mathrm{d}x\right],$$

hence

$$w(x) = \frac{C}{\mu(x)} + \frac{1-n}{\mu(x)}\int \mu(x)Q(x)\,\mathrm{d}x,$$

and

$$y(x) = w(x)^{1/(1-n)}.$$

Example: Find the exact solution of

$$\frac{dy}{dx} - \frac{2y}{x} = -x^2y^2. \tag{7.4}$$

Solution: This is a Bernoulli ODE with $n = 2$. We divide both sides by y^2 to obtain

$$\frac{1}{y^2}\frac{dy}{dx} - \frac{2}{x}\frac{1}{y} = -x^2. \tag{7.5}$$

We let $w(x) = 1/y$. We have

$$\frac{dw}{dx} = -\frac{1}{y^2}\frac{dy}{dx},$$

hence $-(dw/dx) = (1/y^2)(dy/dx)$. Substitute into Equation (7.5) to obtain

$$-\frac{dw}{dx} - \frac{2}{x}w = -x^2,$$

or

$$\frac{dw}{dx} + \frac{2}{x}w = x^2.$$

This is a linear first-order ODE amenable to the integrating-factor technique. We compute

$$\mu = \exp\left[2\int \frac{\mathrm{d}x}{x}\right] = \exp\left(2\log x\right) = \exp\left(\log x^2\right) = x^2.$$

Hence,

$$\frac{d}{dx}\left(wx^2\right) = x^4,$$

hence

$$wx^2 = C + \tfrac{1}{5}x^5,$$

and

$$w = \frac{C}{x^2} + \tfrac{1}{5}x^3.$$

But $w = 1/y$, hence

$$y = \frac{1}{\frac{C}{x^2} + \frac{1}{5}x^3} = \frac{x^2}{C + \frac{1}{5}x^5}.$$

Note that the ODE admits a second independent solution, $y = 0$.

We next look at the **Riccati ODE**:

$$\frac{dy}{dx} = Q_0(x) + Q_1(x)y(x) + Q_2(x)y^2(x), \tag{7.6}$$

where we assume that $Q_0(x)$ and $Q_2(x)$ are not identically zero. If $Q_0(x) = 0$, then the equtaion reduces to a Bernoulli ODE; if $Q_2(x) = 0$ then the equation reduces to a standard linear ODE.

As such, we make a transformation $v = yQ_2$, with

$$\begin{aligned} v' &= y'Q_2 + yQ_2', \\ &= \left[Q_0(x) + Q_1(x)y(x) + Q_2(x)y^2(x)\right]Q_2 + yQ_2', \\ &= Q_0Q_2 + y\left(Q_1Q_2 + Q_2'\right) + Q_2^2y^2, \\ &= Q_0Q_2 + Q_1 + \left(Q_1 + Q_2'/Q_2\right)v + v^2, \\ &= v^2 + R(x)v + S(x), \end{aligned}$$

with $R(x) = Q_1 + (Q_2'/Q_2)$ and $S(x) = Q_0Q_2$.

We make the further transformation $v = -u'/u$, hence

$$\begin{aligned} v' &= -\frac{u''}{u} + \frac{u'^2}{u^2}, \\ &= v^2 + R(x)v + S(x), \\ &= \frac{u'^2}{u^2} - R(u'/u) + S(x), \end{aligned}$$

and finally,

$$u'' - R(x)u' + S(x)u = 0,$$

which is a second-order linear homogeneous ODE. Finally, with $-u'/u = v = yQ_2$, a solution to the original nonlinear ODE is recovered by taking

$$y = -u'/(uQ_2).$$

Example: Find a non-trivial solution of the ODE

$$\frac{dy}{dx} = -2 - y + y^2.$$

Solution: We first of all try the Riccati solution method, with $Q_0 = -2$, $Q_1 = -1$, and $Q_2 = 1$, hence $R(x) = -1$ and $S(x) = -2$. The transformed ODE to solve is

$$u'' + u' - 2 = 0,$$

with solutions $u = \mathrm{e}^{\lambda x}$, and $\lambda = 1$ or $\lambda = -2$. Hence, $u'/u = \lambda$, and

$$y = -u'/(uQ_2) = -\lambda/Q_2 = -\lambda = -1, 2.$$

Howver, these are constant solutions – precisely the fixed points of the ODE, or equivalently, the roots of $-2 - y + y^2 = 0$.

Howver, there is a second aspect to the Ricatti equation. If $y = y_1$ is a solution of the ODE (in this case, a constant solution), we attempt a transformation

$$z = \frac{1}{y - y_1}.$$

We have

$$\begin{aligned}\frac{dz}{dx} &= -\frac{1}{z^2}\frac{dy}{dx}, \\ &= -\frac{1}{z^2}\left(Q_0 + Q_1 y + Q_2 y^2\right).\end{aligned}$$

But $z = 1/(y - y_1)$, hence $y = y_1 + (1/z)$, hence

$$\begin{aligned}\frac{dz}{dx} &= -\frac{1}{z^2}\left[Q_0 + Q_1\left(y_1 + \frac{1}{z}\right) + Q_2\left(y_1 + \frac{1}{z}\right)^2\right], \\ &= -z^2\cancel{\left(Q_0 + Q_1 y_1 + Q_2 y_1^2\right)} - z\left(Q_1 + 2Q_2 y_1\right) - Q_2.\end{aligned}$$

Hence, z satisfies a standard first-order linear ODE with constant coefficients:

$$\frac{dz}{dt} + z\left(Q_1 + 2Q_2 y_1\right) = -Q_2(x),$$

which can be solved by the integrating-factor technique.

For us, $Q_1 = -1$, $Q_2 = 1$, and $y_1 = 2$ (say), hence

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

with solution

$$z = -\tfrac{1}{3} + C\mathrm{e}^{-3x}.$$

With the transformation $y = y_1 + (1/z)$ we obtain a non-constant solution

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

A further, more general solution can be built up by taking

$$y_1(x) := 2 + \frac{1}{-\frac{1}{3} + C\mathrm{e}^{-3x}},$$

and

$$\widetilde{z}(x) = \frac{1}{y - y_1(x)}.$$

Then, $\widetilde{z}(x)$ satisfies the same z-equation as before, but with y_1 no longer constant:

$$\frac{d\widetilde{z}}{dx} + \widetilde{z}\left[Q_1 + 2Q_2 y_1(x)\right] = -Q_2(x),$$

or

$$\frac{d\widetilde{z}}{dx} + \widetilde{z}\,[Q_1 + 2Q_2 y_1(x)] = -Q_2(x).$$

Filling in for $Q_1(x)$ etc.,

$$\frac{d\widetilde{z}}{dx} = 3\widetilde{z}\left(\frac{1 + C\mathrm{e}^{-3x}}{1 - C\mathrm{e}^{-3x}}\right) - 1, \qquad C \to C/3,$$

with a rather involved solution $\widetilde{z}(x; C, D)$ that can be obtained by the integrating-factor technique, and depends on a second constant of integration, D. Thus, the general solution is then

$$y = y_1(x; C) + \frac{1}{\widetilde{z}(x; C, D)}.$$

7.3 On the impossibility of periodic motion in certain ODE systems

We look at two scenarios where periodic motion is impossible in ODE systems. The first is in one-dimensional systems on the line:

Theorem 7.1 *Periodic motion is impossible in a strictly one-dimensional dynamical system on the line, $dx/dt = f(x)$, $-\infty < x < \infty$.*

This is essentially a topological fact. Fixed points dominate the dynamics of the equation $dx/dt = f(x)$. As such,

- If there are no fixed points, then $x(t)$ increases / decreases monotonically on the line.
- If there are fixed points, they are either stable or unstable. $x(t)$ will increase / decrease monotonically away from an unstable fixed point, where as $x(t)$ will approach a stable fixed point asymptotically.

These are all the possibilities for one-dimensional motion.

Remark 7.1 *By introducing explicit time dependence, via $dx/dt = f(x,t)$ one has already gone over to a two-dimensional system of equations,*

$$\frac{d}{dt}\begin{pmatrix} x \\ \tau \end{pmatrix} = \begin{pmatrix} f(x,\tau) \\ 1 \end{pmatrix},$$

and the possibility of periodic motion is regained.

For the remainder of this chapter, we look at such two-dimensional (and higher) systems of equations, which can generically be written in the form

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_1, \cdots, x_n)^T, \tag{7.7}$$

where $\boldsymbol{f}$ is a vector of functions,

$$\boldsymbol{f} = (f_1(x_1, \cdots, x_n), \cdots, f_n(x_1, \cdots, x_n)),$$

i.e. a vector field. A solution of Equation (7.7) with a particular initial condition $\boldsymbol{x}(t=0) = \boldsymbol{x}_0$ is called a **trajectory** or an **orbit**. A **periodic orbit** is one that repeats itself every T units of time:

$$\boldsymbol{x}(t+T) = \boldsymbol{x}(t),$$

for all t. For some of the applications, we will look at a general value of n; for the applications involving periodic orbits and the Poincaré–Bendixson Theorem, we specialize to $n = 2$, which is adequate for our purposes.

Theorem 7.2 *Periodic orbits are impossible under gradient dynamics $\boldsymbol{f} = \nabla V$, where $V(\boldsymbol{x})$ is a continuously differentiable scalar field, i.e. the following ODE system admits no periodic orbits:*

$$\frac{d\boldsymbol{x}}{dt} = \nabla V(\boldsymbol{x}, t), \qquad V \neq \textit{Const.}$$

Proof: Consider

$$\begin{aligned} \frac{dV}{dt} &= \frac{\partial V}{\partial x_i}\frac{dx_i}{dt}, \\ &= \frac{\partial V}{\partial x_i}\frac{\partial V}{\partial x_i}, \\ &= (\nabla V)^2 . \end{aligned}$$

Integrate from $t = t_1$ to $t = t_2$:

$$V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) = \int_{t_1}^{t_2} (\nabla V)^2 \, \mathrm{d}t \neq 0.$$

Suppose there is a periodic orbit, with $\boldsymbol{x}(t_2) = \boldsymbol{x}(t_1)$, and $t_2 > t_1$. Then,

$$V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) = 0.$$

But $V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) \neq 0$, so this is a contradiction. Hence, periodic orbits are impossible in gradient systems.

Definition 7.1 *Let $d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x})$ be a system of ODEs with a fixed point at $\boldsymbol{x}_*$. A continuously differentiable function $V(\boldsymbol{x})$ is a called a* **Lyapunov function** *for the dynamics if*

- $V(\boldsymbol{x}) > 0$ *for all* $\boldsymbol{x} \neq \boldsymbol{x}_*$ *and* $V(\boldsymbol{x}_*) = 0$.
- $dV/dt < 0$ *for all* $\boldsymbol{x} \neq \boldsymbol{x}_*$.

In other words, the flow is 'downhill' towards the minimum value of $V(\boldsymbol{x})$, which occurs at the fixed point.

Theorem 7.3 *If the system $d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x})$ has a Lyapunov function, then no periodic orbits exist.*

Proof: Again, look at dV/dt on a trajectory, and integrate from $t = t_1$ to $t = t_2$. Then,

$$V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) = \int_{t_1}^{t_2} \dot{V}(\boldsymbol{x}(t))\,\mathrm{d}t < 0.$$

Suppose that there is a periodic orbit between t_1 and t_2, with $\boldsymbol{x}(t_2)$ and $\boldsymbol{x}(t_1)$ (neither of which is equal to a fixed point). Assuming periodicity, we have

$$V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) = 0.$$

But $V(\boldsymbol{x}(t_2)) - V(\boldsymbol{x}(t_1)) < 0$, which is a contradiction. Hence, no periodic orbits exist in this case.

There is a third powerful method that enables one to rule out periodic orbits in certain circumstances (again for $n = 2$), called **Dulac's criterion**:

Theorem 7.4 *Let $d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x})$ be two coupled ODEs where the trajectories are confined to an open simply-conected region R in the plane. If there exists a continuously differentiable real-valued function $g(\boldsymbol{x})$ such that $\nabla \cdot [g\boldsymbol{f}(\boldsymbol{x})]$ has the same sign throughout R, then there are no closed periodic orbits lying entirely in R.*

Proof: Suppose for contradiction that a periodic orbit is possible inside the region R. Then, there will be a closed trajectory C lying entirely in the region R. Parametrize the trajectory as $\boldsymbol{x}_C(t)$, with $d\boldsymbol{x}_C/dt = \boldsymbol{f}(\boldsymbol{x}_C)$. Consider the integral

$$I = \oint_C (g\boldsymbol{f}) \cdot \widehat{\boldsymbol{n}} \mathrm{d}\ell,$$

where

- $\widehat{\boldsymbol{n}}$ is normal to the trajectory at all points along the trajectory;

- $\mathrm{d}\ell = |d\boldsymbol{x}/dt|dt$ is an element of length along the trajectory.

But

$$\begin{aligned} I &= \int_{t_1}^{t_2} g(\boldsymbol{x})\,[(d\boldsymbol{x}/dt)\widehat{\boldsymbol{n}}]\,|d\boldsymbol{x}/dt|dt, \\ &= 0, \end{aligned}$$

since $d\boldsymbol{x}/dt$ is tangent to the trajectory, so $(d\boldsymbol{x}/dt)\widehat{\boldsymbol{n}} = 0$.

But by Green's theorem in the plane,

$$I = \iint_A \nabla \cdot (g\boldsymbol{f})\ \mathrm{d}A,$$

where A is the area enclosed by the curve C. By the assumption that $\nabla \cdot (g\boldsymbol{f})$ has one sign throughout R, we have $I > 0$. But this is a contradiction. So we conclude that $I = 0$ is impossible, hence no closed trajectories (periodic orbits) can occur.

7.4 Conditions for periodic orbits in two-equation ODE systems

Having shown how to rule out periodic motion in a range of different ODE systems, we look at an important way of ruling in periodic motion, as such periodic motion is important in applications (physiology, mechanics, etc.). The central result of this section is the famous **Poincaré–Bendixson Theorem**.

Theorem 7.5 *Let $d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x})$ be two coupled ODEs in the plane, and let R be a closed, bounded subset of the plane. Suppose that*

- *R does not contain any fixed points;*
- *There exists a trajectory C that is confined in R – i.e. starts in R and remains in R for all times*

Then either C is a closed orbit, or it spirals towards a closed orbit as $t \to \infty$.

In either case, R contains a closed orbit – e.g. Figure 7.1. We don't prove the theorem in this module; rather, we use it in various settings to establish the existence of a periodic orbit.

When applying the Poincaré–Bendixson Theorem, it is easy the first condition (no fixed points); it is more difficult to satisfy the second. However, this can be done by constructing a so-called **trapping**

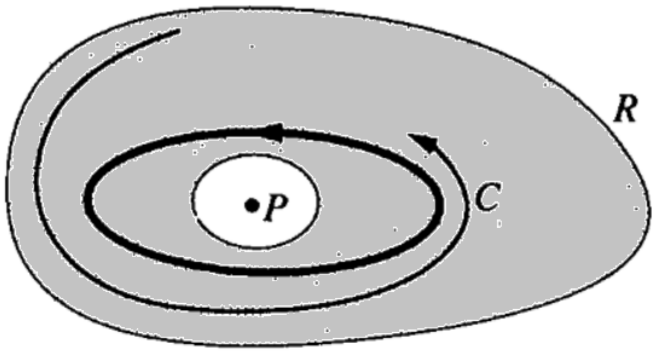


Figure 7.1: Schematic description of Poincaré–Bendixson Theorem. From Reference [Str01].

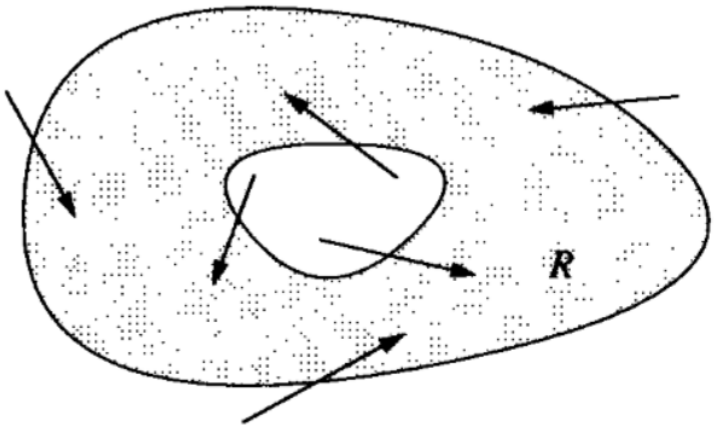


Figure 7.2: The idea of the trapping region. From Reference [Str01].

region – a closed connected region R such that the vector field points inwards everywhere on R. Then, all trajectories in R are confined (e.g. Figure 7.2).

Example: Consider the following set of ODEs in polar coordinates:

$$\begin{aligned} \frac{dr}{dt} &= r(1-r^2) + \mu r\cos\theta, \\ \frac{d\theta}{dt} &= 1. \end{aligned}$$

When $\mu = 0$ there is a periodic orbit at $r = 1$; moreover, this can be shown to be a stable orbit. Show that a closed orbit still exists for $\mu > 0$, so long as μ is sufficiently small.

Solution: We notice that the system has no fixed point, since $d\theta/dt = 1$ always. Therefore, if we can find two values $r_{\min}$ and $r_{\max}$ such that $dr/dt > 0$ on $r_{\min}$ and $dr/dt < 0$ on $r_{\max}$, then the annulus

$$R = \{r | r_{\min} < r < r_{\max}\}$$

is the trapping region, and trajectories spiral into the region from outside. Therefore, we require

$$dr/dt = r(1-r^2) + \mu r\cos\theta > 0 \text{ for } r = r_{\min},$$

hence

$$r_{\min} < \sqrt{1-\mu},$$

so $r_{\min} = 0.999\sqrt{1-\mu}$ would work.

Some steps have been skipped here! To see the missing steps, write $\mathrm{d}r/\mathrm{d}t = rf(r)$, where $f(r) = (1-r^2) + \mu\cos\theta$. We require $f(r) > 0$ on the inner circle, for all θ, in other words,

$$1 - r^2 + \mu\cos\theta > 0.$$

Re-arranging gives

$$r^2 < 1 + \mu\cos\theta, \qquad \text{for all } \theta,$$

in particular, this should be true in a worst-case scenario when the left-hand side is as small as possible, i.e. when $\cos\theta = -1$:

$$r^2 < 1 - \mu,$$

hence $r < \sqrt{1-\mu}$.

By a similar argument on the outer circle, we obtain

$$r_{\max} > \sqrt{1+\mu},$$

so $r_{\min} = 1.001\sqrt{1+\mu}$ would work. Hence, R is a trapping region, and a periodic orbit exists in R. This can be confirmed by numerical experiments, e.g. Figure 7.3. In this figure, the periodic orbit (closed curve C) splits the plane into two parts:

- All trajectories outside of C spiral towards C through decreasing values of r.
- All trajectories inside C spiral towards C through increasing values of r.

Hence, the periodic orbit is isolated – it is a **limit cycle**. Since the trajectories on either side spiral towards the limit cycle, it is stable – hence, a **stable limit cycle**

7.5 Glycolysis model

In this section we look at a mathematical model of **glycolysis**. This is the process by which living cells obtain energy by breaking down sugar. In intact yeast cells as well as in yeast or muscle extracts,

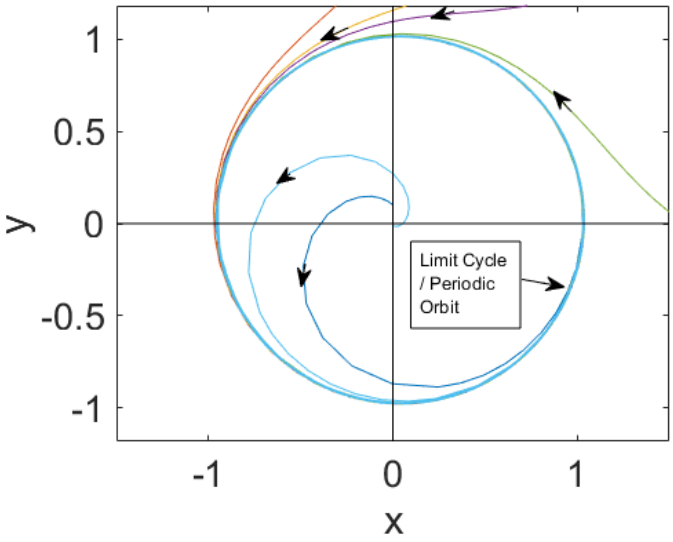


Figure 7.3: Numerical solution showing the limit cycle / periodic orbit with $\mu = 0.1$

glycolysis can proceed in an oscillatory fashion, with the concentrations of various intermediate products increasing and decreasing periodically over a timescale of several minutes.

A simple mathematical model of glycoloysis is provided in Reference [Str01], involving a pair of (dimensionless) equations:

$$\begin{aligned} \frac{dx}{dt} &= -x + ay + x^2 y, \\ \frac{dy}{dt} &= b - ay - x^2 y, \end{aligned}$$

where

- x represents the concentration of ADP (adenosine diphosphate),
- y represents the concentration of F6P (fructose-6-phosphate),
- a and b are positive rate constants.

We construct a trapping region for this set of equations, and thereby demonstrate the existence of a periodic orbit.

We first of all note that the trapping region should be in the positive quadrant, since x and y are concentrations. We next focus on the **nullclines** of the motion, i.e. where $dx/dt = 0$ and separately, where $dy/dt = 0$ (notice that the intersection of the nullclines gives the fixed point(s)).

- $dx/dt = 0$ gives the nullcline

$$y = \frac{x}{a + x^2}.$$

On this nullcline, $dx/dt = 0$, and

$$\frac{dy}{dt} = b - ay - x^2 y = b - y(a + x^2) = b - \left(\frac{x}{a + x^2}\right)(a + x^2) = b - x.$$

Thus, $dx/dt = 0$ and $dy/dt = b - x$, and the vector field is $\boldsymbol{f} = (0, b - x)$, with arrows pointing down for large x and up for small x.

- $dy/dt = 0$ gives the nullcline

$$y = \frac{b}{a + x^2}.$$

Thus, $dy/dt = 0$ and $dx/dt = -x + b$, and the vector field is $\boldsymbol{f} = (-x + b, 0)$, with arrows pointing to the left for large x and to the right for small x.

- $dx/dt = 0$ and $dy/dt = 0$ simultaneously corresponds to the intersection of the nullclines and hence, $x = b$ and then, $y = b/(a + b^2)$. This is the fixed point.

The nullclines and the fixed point are plotted in Figure 7.4. The direction fo the vector field $\boldsymbol{f}$ on the nullclines is indicated at the points A, B, C, and D. The direction of the vector field in the other parts of the plane can similarly be reasoned out.

We now attempt to construct a trapping region in the positive quadrant. On $y = 0$, we have $dx/dt = -x$ and $dy/dt = b$, so the vector field points updwards and to the left, i.e. into the positive quadrant. Similarly, on $x = 0$, we have $dx/dt = y$ and $dy/dt = b - ay$. Provided we take $y < b/a$, the vector field points downward and to the right, i.e. into the positive quadrant again. These insights help to construct the trapping region in Figure 7.5.

Unfortunately, the region R constructed in Figure 7.5 is not yet suitable for the application of the Poincaré–Bendixson theorem, as it contains the fixed point

$$\boldsymbol{x}_* = (x_*, y_*) = \left(b, \frac{b}{a + b^2}\right).$$

However, if we take the punctured region

$$R_* = R - \{(x_*, y_*)\}$$

we are nearly there, **provided the fixed point is unstable**. For, if the fixed point is unstable, then the vector field in a small circular region around the fixed point will have arrows pointing away from $\boldsymbol{x}_*$ and into R_*, and then R_* will really meet the criteria of the theorem. This idea is shown schematically in Figure 7.6. It remains to ensure that the fixed point is unstable. With

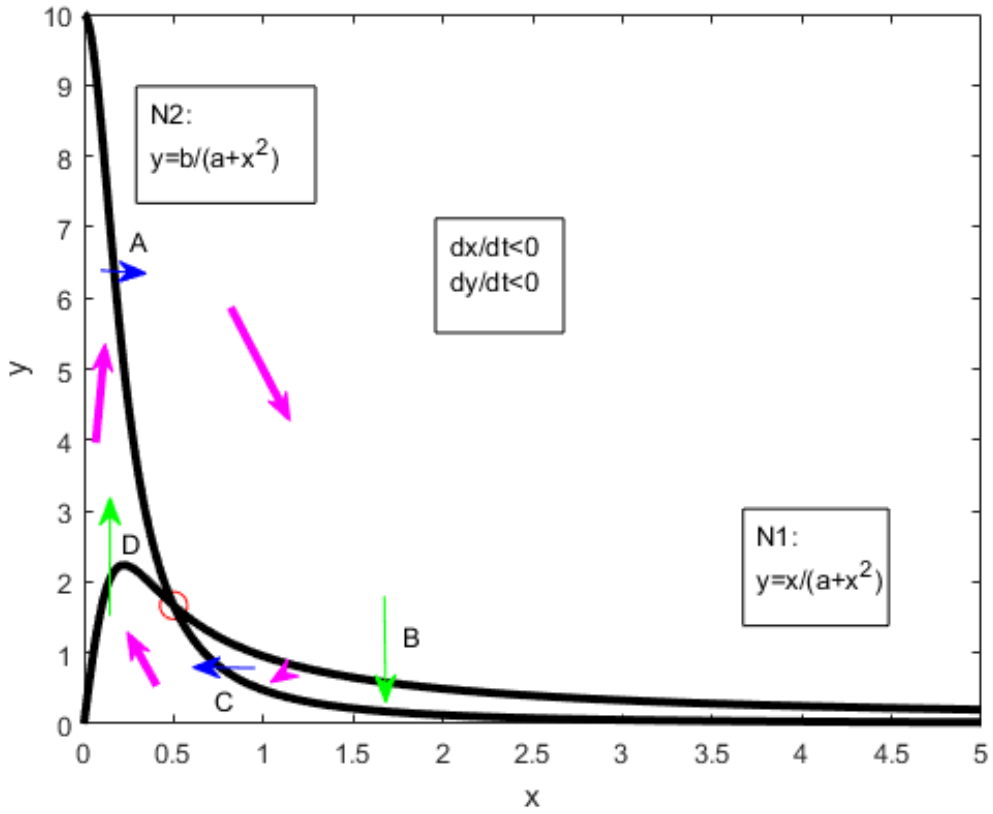


Figure 7.4: Schematic description of the nullclines and the direction of the vector field $\boldsymbol{f}$ for the glycolysis model, with $a = 0.05$ and $b = 0.5$.

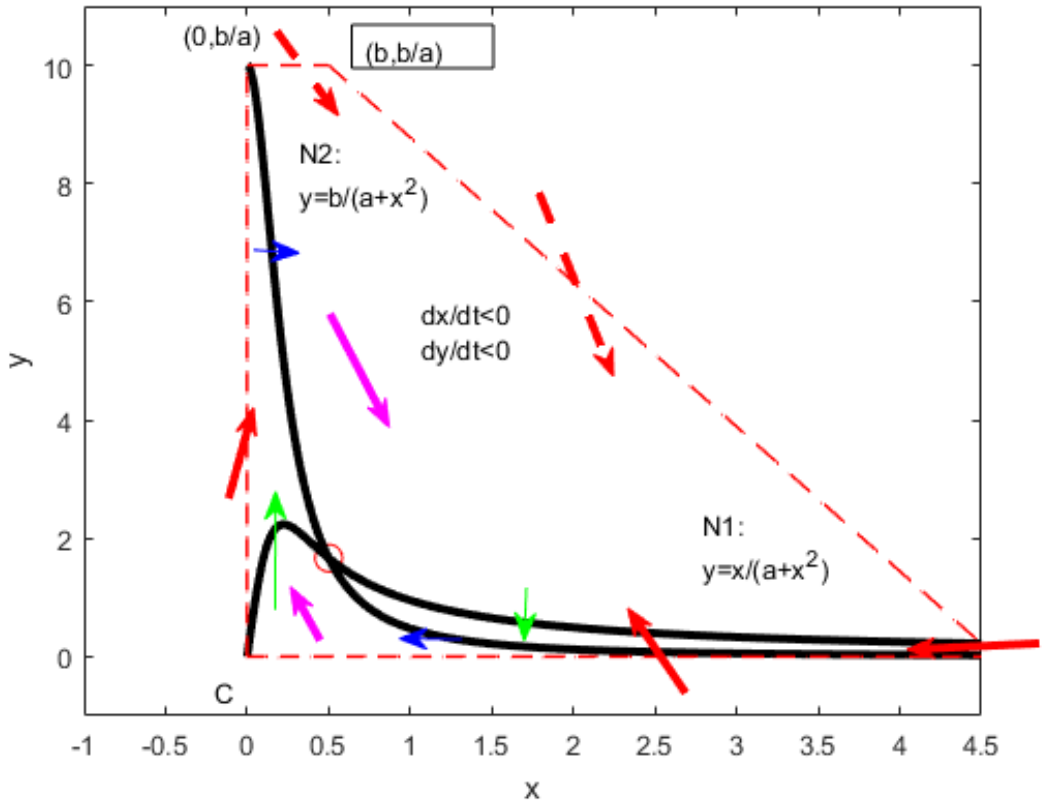


Figure 7.5: Trapping region R for the glycolysis model, with $a = 0.05$ and $b = 0.5$.

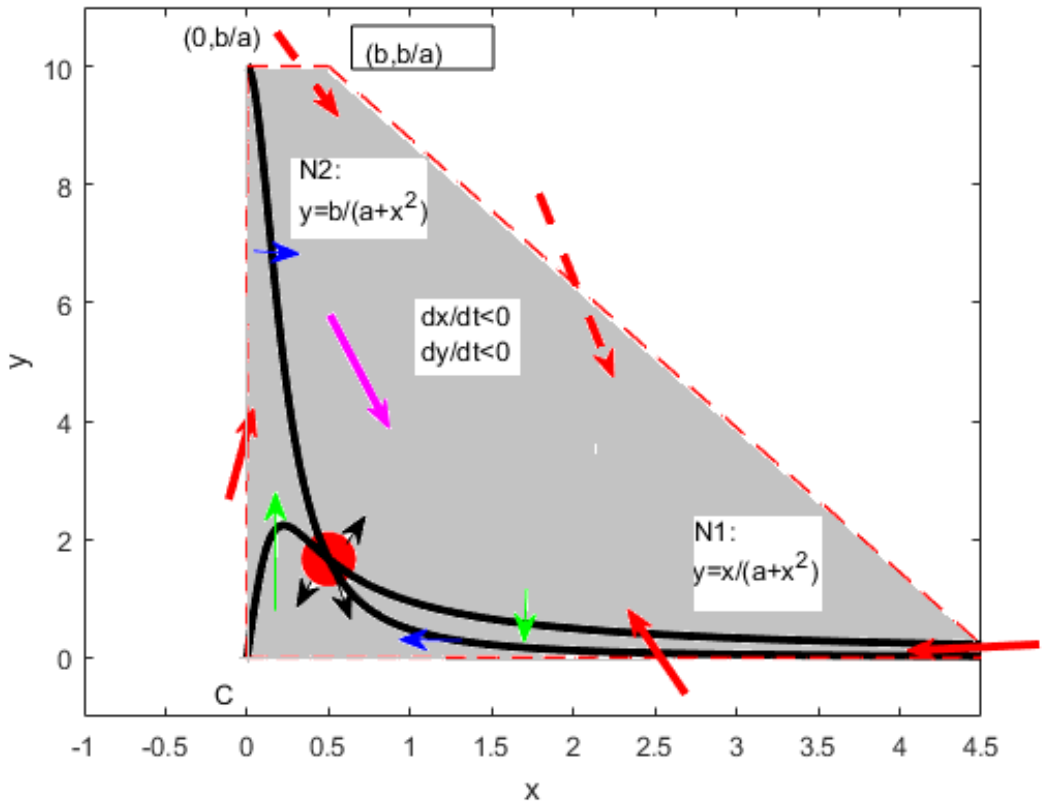


Figure 7.6: Proper trapping region R_* for the glycolysis model, with $a = 0.05$ and $b = 0.5$, showing the puncture around the fixed point.

$$\begin{aligned}\frac{dx}{dt} &= F(x,y) = -x + ay + x^2 y,\\ \frac{dy}{dt} &= G(x,y) = b - ay - x^2 y\end{aligned}$$

we have

$$\boldsymbol{A} = \begin{pmatrix} F_x & F_y \\ G_x & G_y \end{pmatrix} = \begin{pmatrix} -1+2xy & a+x^2 \\ -2xy & -a-x^2 \end{pmatrix}.$$

At the fixed point $\boldsymbol{x}_* = (b, b/(a+b^2))$ this is

$$\boldsymbol{A} = \begin{pmatrix} -1+\frac{2b^2}{a+b^2} & a+b^2 \\ -\frac{2b^2}{a+b^2} & -a-b^2 \end{pmatrix},$$

There is a pattern here – the Jacobian can be abbreviated in an obvious way as

$$\boldsymbol{A} = \begin{pmatrix} -1+X & Y \\ -X & -Y \end{pmatrix},$$

We take a short cut to the characeteristic equation using the generic formula

$$\lambda^2 - \mathrm{tr}(\boldsymbol{A})\lambda + \det(\boldsymbol{A}) = 0.$$

We have

$$\begin{aligned}\det(\boldsymbol{A}) &= Y = a + b^2 > 0,\\ \mathrm{tr}(\boldsymbol{A}) &= -1 + X - Y = -\left[\frac{b^4 + (2a-a)b^2 + (a+a^2)}{a+b^2}\right],\end{aligned}$$

and finally,

$$\lambda = \frac{\mathrm{tr}(\boldsymbol{A}) \pm \sqrt{[\mathrm{tr}(\boldsymbol{A})]^2 - 4\det(\boldsymbol{A})}}{2}.$$

Clearly, there is a threshold at $\mathrm{tr}(\boldsymbol{A}) = 0$, where the system is neutrally stable and exhibits small oscillations with

$$\lambda = \pm \mathrm{i}\sqrt{\det(\boldsymbol{A})} = \pm \mathrm{i}\sqrt{a+b^2}.$$

- One one side of the threshold, $\lambda_\mathrm{r} < 0$ for all eigenvalues, and the fixed point is stable.
- On the other side of the threshold, $\lambda_\mathrm{r} > 0$ for at least one eigenvalue and the fixed point is unstable – desired for the application of the Poincaré–Bendixson Theorem.

The threshold $\mathrm{tr}(\boldsymbol{A}) = 0$ therefore defines a curve in the parameter space:

$$\mathrm{tr}(\boldsymbol{A}) = 0 \implies b^2 = \tfrac{1}{2}\left(1 - 2a \pm \sqrt{1-8a}\right),$$

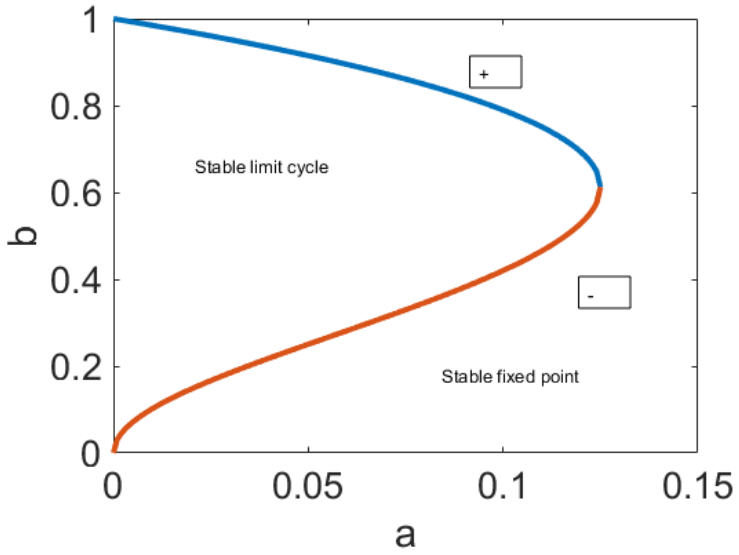


Figure 7.7: Parameter space for the glycolysis model showing the region where a stable limit cycle (periodic orbits) exist.

shown in Figure 7.7. Some plots of evolution towards the limit cycle for the paramter values $a = 0.05$ and $b = 0.5$ are shown in Figure 7.8.

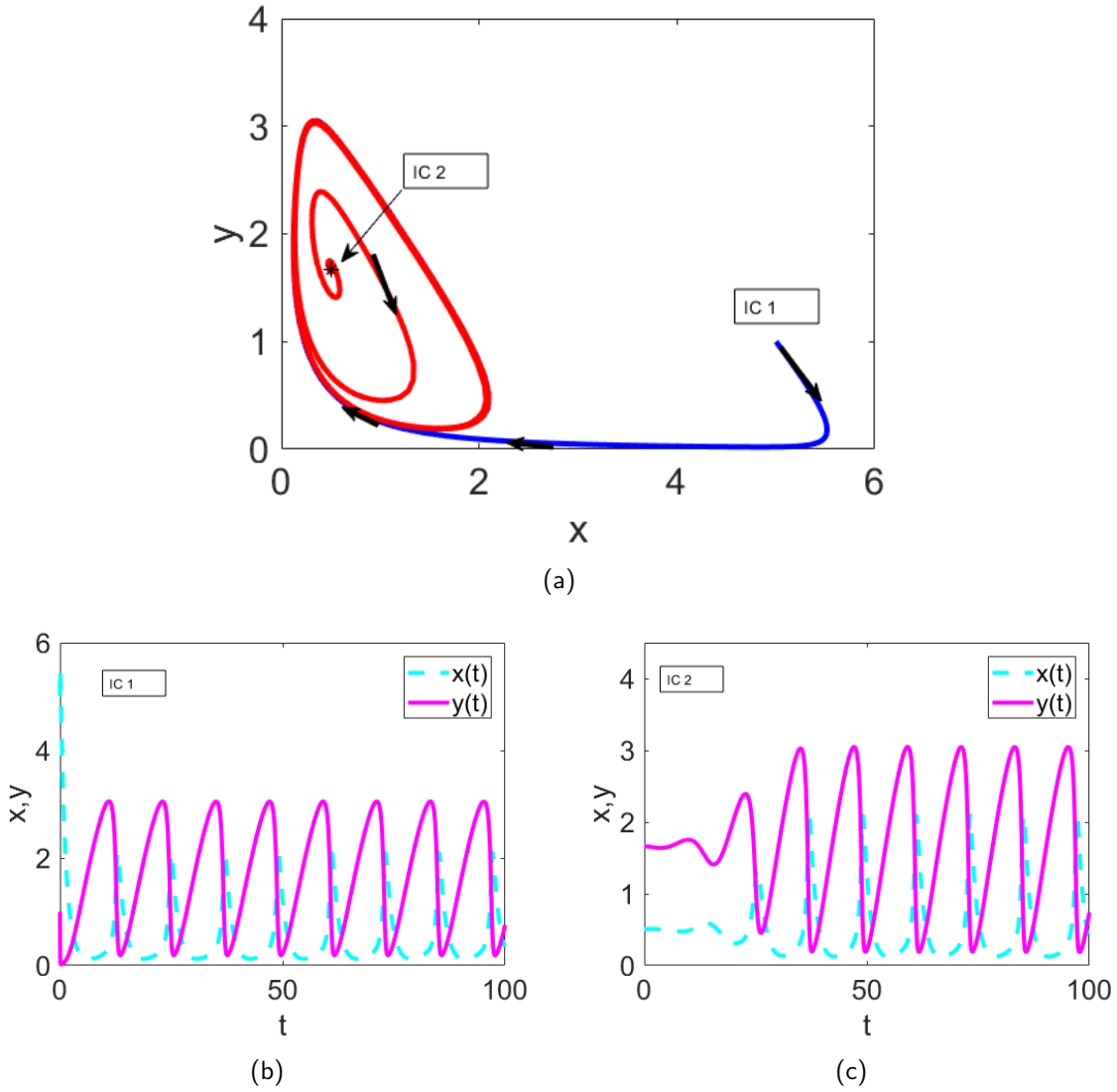


Figure 7.8:

Chapter 8

Theory of Nonlinear Oscillations (*)

8.1 Outline

In the previous chapter we developed a powerful method of determining when a two-dimnesional system admitted oscillations. However, this was a kind of existence theory, and we did not construct explicity analytical solutions for the oscillatory motion. We now do this in certain asymptotic limits, by identifying a small parameter ϵ in the problem and developing a perturbation theory as $\epsilon \to 0$.

8.2 Preliminaries

The generic problem we work on will be of the form

$$\frac{d^2x}{dt^2} + x + \epsilon h(x, dx/dt) = 0.$$

This can be brought to the standard form $d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x})$ by writing

$$\boldsymbol{x} = \begin{pmatrix} x \\ v \end{pmatrix}, \qquad v = dx/dt.$$

Then,

$$\begin{aligned} \frac{d\boldsymbol{x}}{dt} &= \begin{pmatrix} v \\ d^2x/dt^2 \end{pmatrix}, \\ &= \begin{pmatrix} v \\ -x - \epsilon h(x, v) \end{pmatrix}, \\ &= \boldsymbol{f}. \end{aligned}$$

We further introduce the notation $dx/dt = \dot{x}$, etc.

8.3 Regular perturbation theory and its failure

To fix ideas, we first of all look at a linear equation where the exact solution is known:

$$\ddot{x} + 2\epsilon\dot{x} + x = 0, \tag{8.1}$$

subject to the initial conditions

$$x(0) = 0, \qquad \dot{x}(0) = 1. \tag{8.2}$$

Using the trial solution $x = A\mathrm{e}^{\lambda x}$ and the characteristic equation $\lambda^2 + 2\epsilon\lambda + 1 = 0$, the following exact solution is obtained:

$$x(t;\epsilon) = \frac{\mathrm{e}^{-\epsilon t}}{\sqrt{1-\epsilon^2}} \sin\left(\sqrt{1-\epsilon^2}t\right), \tag{8.3}$$

where $\epsilon \ll 1$ is assumed. Notice that this solution satisfies the initial conditions (8.2).

We now attempt to solve Equation (8.1) via regular perturbation theory. We expand the solution $x(t)$ in a poer series in ϵ:

$$x(t) = x_0(t) + \epsilon x_1(t) + \epsilon^2 x_2(t) + \cdots .$$

We substitute this expansion into Equation (8.1) and equate coefficients of powers of ϵ. The result is

$$\begin{aligned} O(1) &: \ \ddot{x}_0 + x_0 = 0, \\ O(\epsilon) &: \ \ddot{x}_1 + 2\dot{x}_0 + x_1 = 0, \end{aligned}$$

where we focus only on the lowest orders in the perturbation theory for now.

The solution at $O(1)$ satisfies the initial condition $x_0(0) = 0$ and $\dot{x}_0(0) = 1$. The only possible solution is thus

$$x_0(t) = \sin(t).$$

At $O(\epsilon)$ the solution $x_1(t)$ has the initial condition $x_1(0) = \dot{x}_1(0) = 0$. However, the $O(\epsilon)$ equation is inhomogeneous in x_1:

$$\ddot{x}_1 + x_1 = -2\dot{x}_0 = -2\cos(t).$$

The solution here via the method of variation or parameters[1] (or just a clever trial solution) is

$$x_1(t) = -t\,\sin t.$$

[1] see e.g. `https://en.wikipedia.org/wiki/Variation_of_parameters`

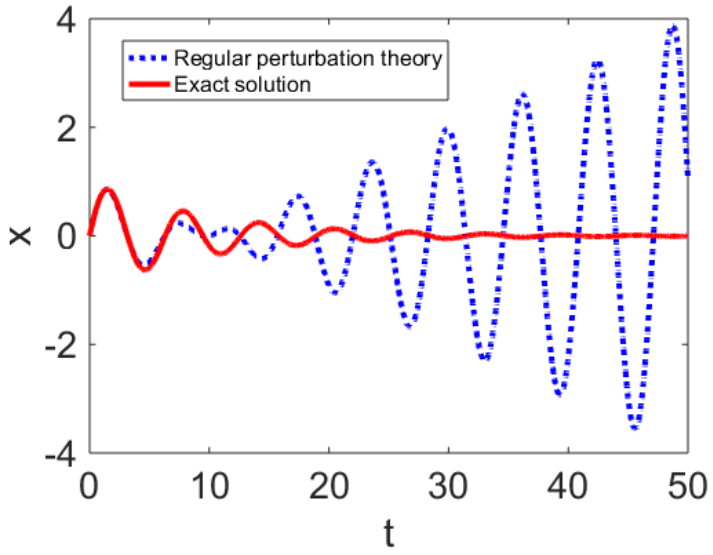


Figure 8.1: Comparison between exact solution and regular perturbation approximation for $\epsilon = 0.1$ – the two curves diverge from each over very rapidly over a timescale $t \ll 1/\epsilon = 10$.

Notice that this solution has the hallmark of resonant forcing, as $x_1(t)$ grows algebraically in time.

Putting these results together, the low-order approximation to the full solution is

$$x(t) = \sin t - \epsilon t \, \sin(t) + O(\epsilon^2).$$

There is clearly a problem with this solution, as it grows without bound at large t– **we have fixed** ϵ **but** t **can vary**, meaning that the amplitude of the proposed solution is not controlled. In contrast, the amplitude of the correct solution is bounded – indeed, it decays exponentially.

Otherwise put, the perturbation solution is valid only for $t \ll 1/\epsilon$ – as evidenced in Figure 8.1.

We can understand why the regular perturbation theory fails by looking at the timescales in the exact solution:

- A slow timescale $T = 1/\epsilon$ (i.e. $T \to \infty$) over which the amplitude of the solution decays.
- A fast timescale $T = 2\pi/\sqrt{1-\epsilon^2}$ (i.e. $T = O(1)$) corresponding to the oscillations of the sine function.

The regular perturbation theory misrepresents both these timescales:

- The perturbation theory takes the exponential factor $\mathrm{e}^{-\epsilon t}$ in the exact solution and approximates it as $\mathrm{e}^{-\epsilon t} \approx 1 - \epsilon t$. Thus, instead of a decaying amplitude, the perturbation theory wrongly makes the amplitude look as though it is growing as $-\epsilon t$.

To correct this, we should have to sum up all of the terms in the series $\mathrm{e}^{-\epsilon t} = 1 - \epsilon t + (1/2)\epsilon^2 t^2 + \cdots$. But this would correspond to solving the perturbation theory to all orders, when in fact we want a theory that gives an accurate solution with only a few terms included.

- The perturbation theory takes the frequency $\omega = \sqrt{1-\epsilon^2}$ in the argument of the sine wave and approximates it as $\omega = 1$. After a long time $t = 1/\epsilon^2$ this difference is significant, and there will be a finite phase shift between the true solution and the perturbed one.

The solution to these difficulties is to recognize that the system of ODEs contains two very separate timescales, and to treat motion on those timescales as completely independent.

As such, we identify the **fast** timescale $\tau = t$ – the rapid timescale on which the oscillations occur. We also identify the **slow** timescale T as the separate timescale on which the damping occurs. We now make the solution $x(t;\epsilon)$ be in effect a function of both these timescales:

$$x(t;\epsilon) = x_0(\tau, T) + \epsilon x_1(\tau, T) + O(\epsilon^2).$$

The time derivative dx/dt is thereby expanded using the chain rule:

$$\frac{dx}{dt} = \frac{\partial x}{\partial \tau}\frac{d\tau}{dt} + \frac{\partial x}{\partial T}\frac{dT}{dt}. \tag{8.4}$$

We have $\tau = t$, hence $d\tau/dt = 1$. We make T the slow timescale by taking

$$T = \epsilon t,$$

hence $dT/dt = \epsilon$, and so Equation (8.4) implies that the time derivative splits up into partial derivatives:

$$\frac{d}{dt} = \frac{\partial}{\partial \tau} + \epsilon\frac{\partial}{\partial T}. \tag{8.5a}$$

Similarly,

$$\frac{d^2}{dt^2} = \frac{\partial^2}{\partial \tau^2} + 2\epsilon\frac{\partial}{\partial \tau}\frac{\partial}{\partial T} + \epsilon^2\frac{\partial^2}{\partial T^2}. \tag{8.5b}$$

We now substitute Equation (8.5) and the perturbation expansion $x(t;\epsilon) = x_0(\tau,T)+\epsilon x_1(\tau,T)+\cdots$ into the basic ODE (8.1) and equate coefficients of powers of ϵ as before. We obtain:

$$\begin{aligned} O(1) &: \partial_{\tau\tau}x_0 + x_0 = 0, \\ O(\epsilon) &: \partial_{\tau\tau}x_1 + \partial_{\tau T}x_0 + 2\partial_\tau x_0 + x_1 = 0. \end{aligned}$$

At lowest order, the solution of $\partial_{\tau\tau}x_0 + x_0 = 0$ is

$$x_0 = A\sin\tau + B\cos\tau.$$

But $x_0 = x_0(\tau, T)$ – so the coefficients A and B are functions of the slow time T:

$$x_0(T,\tau) = A(T)\sin\tau + B(T)\cos\tau.$$

To determine $A(T)$ and $B(T)$ we go to the $O(1)$ equation, which can now be written as

$$\begin{aligned} \partial_{\tau\tau}x_1 + x_1 &= -2\left(\partial_{\tau T}x_0 + \partial_\tau x_0\right), \\ -2\left(A' + A\right)\cos\tau + 2\left(B' + B\right)\sin\tau. & \end{aligned}$$

The right-hand side apparently contains resonant forcing terms $(\cdots)\cos\tau$ and $(\cdots)\sin\tau$ – these are precisely the kind of forcing terms that ruined the earlier perturbation theory., However, the two-timing approach gives us an important freedom: we can choose A and B so that these forcing terms are zero. As such, we choose

$$\begin{aligned} A' + A &= 0, \\ B' + B &= 0. \end{aligned}$$

Hence,

$$A(T) = A(0)\mathrm{e}^{-T}, \qquad B(T) = B(0)\mathrm{e}^{-T},$$

where $A(0) = A_0$ and $B(0) = B_0$ now really are constants. Hence,

$$x_0(\tau, T) = A_0\mathrm{e}^{-T}\sin\tau + B_0\mathrm{e}^{-T}\cos\tau$$

is the lowest-order solution – up to the constants to be determined.

At lowest order, from the initial conditions, we clearly have $x_0(0,0) = 0$, hence $B_0 = 0$, and

$$x_0(\tau, T) = A_0\mathrm{e}^{-T}\sin\tau.$$

It remains to work out A_0. We have

$$\frac{dx_0}{dt} = \frac{\partial x_0}{\partial\tau} + O(\epsilon),$$

hence the initial condition $(dx_0/dt) = 1$ is equivalent to

$$\frac{\partial x_0}{\partial\tau} = 1, \qquad \tau = T = 0,$$

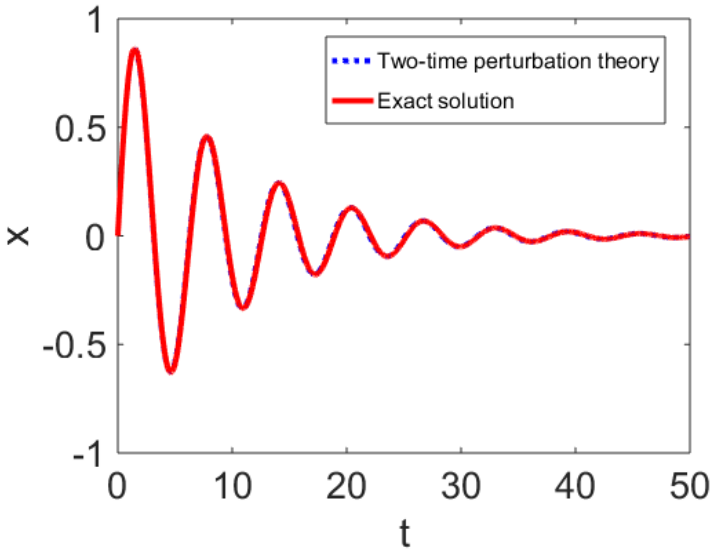


Figure 8.2: Agreement between exact solution and two-time perturbation approximation for $\epsilon = 0.1$.

hence $A_0 = 1$, and the x_0 solution is thus $x_0 = \mathrm{e}^{-T}\sin\tau$. But $x(t;\epsilon) = x_0(\tau, T) + O(\epsilon)$, hence

$$\begin{aligned} x(t;\epsilon) &= \mathrm{e}^{-T}\sin\tau + O(\epsilon), \\ &= \mathrm{e}^{-\epsilon t}\sin(t) + O(\epsilon). \end{aligned}$$

A comparison between the exact solution and the two-time perturbation approximation is shown in Figure 8.2 – the agreement is remarkably good.

8.4 Nonlinear oscillations

We now apply the two-time perturbation theory to the Van der Pol oscillator

$$\ddot{x} + x + \epsilon(x^2 - 1)\dot{x} = 0. \tag{8.6}$$

By introducing the two-time expansion as before, we obtain the following set of equations:

$$\begin{aligned} O(1) &: \partial_{\tau\tau}x_0 + x_0 = 0, \\ O(\epsilon) &: \partial_{\tau\tau}x_1 + x_1 = -2\partial_{\tau T}x_0 - (x_0^2 - 1)\partial_\tau x_0. \end{aligned}$$

The equation at $O(1)$ is that of simple harmonic motion (this will always be the case for the model perturbation equation $\ddot{x}+x+\epsilon h(x,\dot{x}) = 0$). The solution is therefore $x_0 = A(T)\cos\tau + B(T)\sin\tau$.

We choose here to write that in equivalent amplitude-phase form as

$$x_0 = r(T)\cos\left[\tau + \varphi(T)\right]$$

To fix $r(T)$ and $\varphi(T)$ we substitute this expression into the $O(\epsilon)$ equation. We obtain:

$$\partial_{\tau\tau}x_1 + x_1 = -2\left[r'\sin(\tau+\varphi) + r\varphi'\cos(\tau+\varphi)\right] - r\sin(\tau+\varphi)\left[r^2\cos^2(\tau+\varphi) - 1\right].$$

As before, we need to avoid the resonant terms on the right. Clearly, $\sin(\tau+\varphi)$ and $\cos(\tau+\varphi)$ are resonant terms. However, the cross term proportional to $\sin(\tau+\varphi)\cos^2(\tau+\varphi)$ is also a resonant term, since by trigonometry,

$$\sin(\tau+\varphi)\cos^2(\tau+\varphi) = \tfrac{1}{4}\left[\sin(\tau+\varphi) + \sin 3(\tau+\varphi)\right],$$

and the first term on the right-hand side is resonant. Gathering up the terms, we have

$$\partial_{\tau\tau}x_1 + x_1 = \sin(\tau+\varphi)\left[-2r' + r - \tfrac{1}{4}r^3\right] + \left[-2r\varphi'\right]\cos(\tau+\varphi) - \tfrac{1}{4}r^3\sin 3(\tau+\varphi).$$

As such, we require

$$-2r' + r - \tfrac{1}{4}r^3 = 0, \qquad r\varphi' = 0$$

for non-resonance.

The r-equation reduces to

$$r' = \tfrac{1}{8}r\left(4 - r^2\right).$$

For now, we just look at the fixed points of this equation. By a one-dimensional vector field analysis, the fixed point $r = 0$ is unstable and $r = 2$ is stable, so

$$r \to 2 \text{ as } T \to \infty.$$

Hence, $\varphi' = 0$, so $x_0 = r(T)\cos(\tau+\varphi_0) \to 2\cos(\tau+\varphi_0)$ as $T \to \infty$. Here, φ_0 is constant. Back to $\tau = t$:

$$x(t) = 2\cos(t+\varphi_0) \text{ as } t \to \infty.$$

Thus, the Van der Pol oscillator has a stable limit cycle in the limit of small ϵ.

Moreover, for specific initial conditions, we can study the spiral towards the limit cycle. For instance, consider

$$x(0) = 1, \qquad \dot{x}(0) = 0.$$

Then,

$$x_0 = 1, \qquad \frac{\partial x_0}{\partial \tau} = 0, \qquad T = \tau = 0.$$

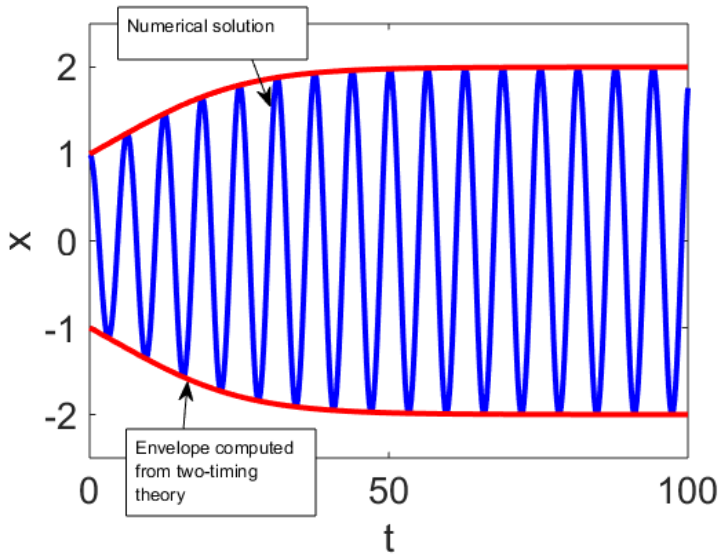


Figure 8.3:

This gives $\varphi = 0$ and $r(0) = 1$. We therefore solve

$$r' = \tfrac{1}{8} r \left(4 - r^2\right), \qquad r(0) = 1,$$

with solution (up to quadratures)

$$t = \int \frac{\mathrm{d}r}{r(4-r^2)} = \tfrac{1}{4}\log(r) - \tfrac{1}{8}\log(4-r^2) + C,$$

hence

$$\frac{4C\mathrm{e}^T}{1 + C\mathrm{e}^T} = r^2.$$

We **assume** the initial condition $x(0) = 1$, hence $r(0) = 1$. These assumptions give $C = 1/3$, hence

$$r(T) = \frac{2}{\sqrt{1 + 3\mathrm{e}^{-T}}},$$

Hence,

$$\begin{aligned} x(t;\epsilon) &= x_0(\tau, T) + O(\epsilon), \\ &= \frac{2}{\sqrt{1 + 3\mathrm{e}^{-\epsilon t}}} \cos(t) + O(\epsilon). \end{aligned}$$

This solution is plotted in Figure 8.3 and a comparison is made with a numerical solution, with $\epsilon = 0.1$. Excellent agreement is obtained. One can also see from the numerical solution what happens to the limit cycle for larger values of the control parameter ϵ (Figure 8.4)

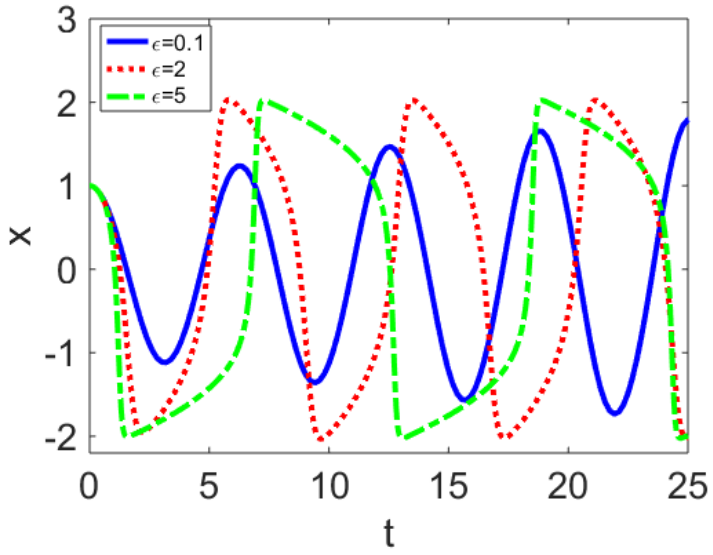


Figure 8.4:

8.5 Averaged Equations – General formulation

We return to the general weakly nonlinear equation

$$\ddot{x} + x + \epsilon h(x, \dot{x}) = 0,$$

with a view to formulating a general two-time perturbation theory.

We apply two-timing to get

$$\begin{aligned} O(1) \quad &: \quad \partial_{\tau\tau} x_0 + x_0 = 0, \\ O(\epsilon) \quad &: \quad \partial_{\tau\tau} x_1 + x_1 = -2\partial_{\tau T} x_0 - h(x_0, \partial_\tau x_0) \end{aligned}$$

We previously observed that the structure of the weakly nonlinear eqeuation is such that the $O(1)$ equation is always that of simple harmonic motion. As such, the solution for x_0 carries over from the Van der Pol oscillator study:

$$x_0(\tau, T) = R(T)\cos\left[\tau + \varphi(T)\right].$$

Correspondingly, the $O(\epsilon)$ equation is

$$\partial_{\tau\tau} x_1 + x_1 = 2\left[r'\sin(\tau+\varphi) + r\varphi'\cos(\tau+\varphi)\right] - h\left[r\cos(\tau+\varphi), -r\sin(\tau+\varphi)\right].$$

It now remains to conduct a general analysis to remove the resonant terms from the right-hand side.

We let $\theta = \tau + \varphi$ and re-write the right-hand side as

$$2\left[r' \sin\theta + r\varphi' \cos\theta\right] - h(\theta).$$

Since

$$\begin{aligned} h &= h\left[r\cos(\tau+\varphi), -r\sin(\tau+\varphi)\right], \\ &= h\left(r\cos\theta, -r\sin\theta\right), \\ &= h(\theta), \end{aligned}$$

and h is therefore manifestly a 2π-periodic function in θ. As such, we can expand $h(\theta)$ in a Fourier series:

$$h(\theta) = a_0 + \sum_{k=1}^{\infty} a_k \cos(k\theta) + \sum_{k=1}^{\infty} b_k \sin(k\theta),$$

with

$$\begin{aligned} a_0 &= \frac{1}{2\pi}\int_0^{2\pi} h(\theta)\,\mathrm{d}\theta, \\ a_k &= \frac{1}{\pi}\int_0^{2\pi} h(\theta)\cos(k\theta)\,\mathrm{d}\theta, \qquad k \geq 1 \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} h(\theta)\sin(k\theta)\,\mathrm{d}\theta, \qquad k \geq 1. \end{aligned}$$

Hence, the x_1-equation becomes

$$\partial_{\tau\tau} x_1 + x_1 = 2\left[r'\sin\theta + r\varphi'\cos\theta\right] - \sum_{k=0}^{\infty} a_k \cos(k\theta) - \sum_{k=1}^{\infty} b_k \sin(k\theta).$$

The resonant terms are associated with $k = 1$ only – the other terms in the series are non-resonant. As such, in order to kill off resonant motions, we require

$$2r' - b_1 = 0, \qquad r\varphi' = a_1/2.$$

In other words,

$$r' = \frac{1}{2\pi}\int_0^{2\pi} h(\theta)\sin(\theta)\,\mathrm{d}\theta := \langle h\sin\theta\rangle, \tag{8.7a}$$

$$r\varphi' = \frac{1}{2\pi}\int_0^{2\pi} h(\theta)\cos(\theta)\,\mathrm{d}\theta := \langle h\cos\theta\rangle, \tag{8.7b}$$

$$\tag{8.7c}$$

where the angle brackets denote averaging over one periodic cycle, from $\theta = 0$ to $\theta = 2\pi$.

Equations (**??**) are the general **slow-time equations**. They are written in general terms, i.e. in terms of a generic h-function alone. The averaged quantities in the equations can be computed quite independently of knowledge of the solution of the underlying ODE – the averaging just requires knowledge of calculus, so this is a powerful technique.

Example: Compute the amplitude and period of the weakly nonlinear motion of the **Duffing oscillator**

$$\ddot{x} + x + \epsilon x^3 = 0.$$

Solution: Identify $h(x, \dot{x}) = x^3$, so

$$h(\theta) = r^3 \cos^3 \theta,$$

hence

$$r' = \langle h(\theta) \sin\theta \rangle = r^3 \langle \cos^3\theta \sin\theta \rangle = 0,$$

hence $r = \text{const.} = a$.

Remark 8.1 *The amplitude of the oscillation is constant for the Duffing oscillator. This makes sense, as the Duffing oscillator is a conservative mechanical system: the energy*

$$E = \tfrac{1}{2}\dot{x}^2 + \tfrac{1}{4}\epsilon x^4.$$

Also,

$$r\varphi' = \langle h(\theta)\cos\theta \rangle = r^3 \langle \cos^4\theta \rangle = \tfrac{3}{8} r^3.$$

Hence, $\varphi' = 3a^2/8$, hence $\varphi = (3a^2/8)T + \varphi_0$, where φ_0 is a constant. To lowest order in the perturbation theory, the solution is thus

$$\begin{aligned} x(t) &= a\cos\left(\tau + (3a^2/8)T + \varphi_0\right), \\ &= a\cos\left[t + (3a^2/8)\epsilon t + \varphi_0\right], \\ &= a\cos\left[t\left(1 + \tfrac{3}{8}\epsilon a^2\right) + \varphi_0\right], \end{aligned}$$

so the frequency of the oscillation is clearly

$$\omega = 1 + \tfrac{3}{8}\epsilon a^2 + O(\epsilon^2).$$

Remark 8.2 *The frequency of oscillation depends on the amplitude. This is a generic feature of nonlinear oscillators. In contrast, for linear oscillators, the frequency is independent of amplitude.*

Chapter 9

Averaging methods for ODEs (*)

9.1 Outline

In the previous chapter we developed a powerful method for constructing approximate solutions of weakly nonlinear oscillator equations. The idea was based on the separation of timescales and averaging. We extend that idea in this chapter to more generic higher-dimensional systems of ODEs. This will pave the way to extend the same idea again to PDEs in later chapters.

9.2 General framework

We look at an $n+m$-dimensional system of ODEs,

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{y}), \tag{9.1a}$$

$$\frac{d\boldsymbol{y}}{dt} = \frac{1}{\epsilon}\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}), \tag{9.1b}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$; together, the vector $(\boldsymbol{x},\boldsymbol{y})$ lies in $\mathbb{R}^{n+m}$ where the system is defined. Correspondingly, $(\boldsymbol{f},\boldsymbol{g})$ is a vector field in $\mathbb{R}^{n+m}$. We shall work in the limit where $\epsilon \to 0$.

- The set of variables $\boldsymbol{x}$ are the **slow variables**.
- The set of variables $\boldsymbol{y}$ are the **fast variables**.

We shall carry out a separation of scales based on the slow and fast variables – similar to the ideas used previously in Section 9, the difference now being the more general setting.

The idea – at a very crude level – is to notice that in order to make Equation (9.1b) self-consistent as $\epsilon \to 0$, we require $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = 0$. This then gives a functional relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Assuming this implicit equation can be inverted, we obtain $\boldsymbol{y} = \Psi(\boldsymbol{x})$. This can then be substituted into Equation (9.1b) to give a much reduced system of n slow equations:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, \Psi(\boldsymbol{x})).$$

This basic idea will be elaborated on by way of a formal perturbation theory in the next section.

9.3 The expansion

As suggested previously, we propose the following solution for the fast variables:

$$\boldsymbol{y} = \Psi(\boldsymbol{x}),$$

i.e. $y_i = \Psi_i(\boldsymbol{x})$. Differentiate this solution to obtain

$$\begin{aligned} \frac{dy_i}{dt} &= \frac{\partial \Psi_i}{\partial x_j}\frac{dx_j}{dt}, \\ &= \frac{\partial \Psi_i}{\partial x_j} f_j(\boldsymbol{x}, \Psi(\boldsymbol{x})), \\ &= \frac{1}{\epsilon} g_i(\boldsymbol{x}, \Psi(\boldsymbol{x})). \end{aligned}$$

We therefore have the following equation for the unknown function $\Psi(\boldsymbol{x})$:

$$\frac{\partial \Psi_i}{\partial x_j} f_j(\boldsymbol{x}, \Psi(\boldsymbol{x})) = \frac{1}{\epsilon} g_i(\boldsymbol{x}, \Psi(\boldsymbol{x})).$$

We solve for Ψ by making an expansion in powers of ϵ:

$$\Psi(\boldsymbol{x}) = \Psi_0(\boldsymbol{x}) + \epsilon \Psi_1(\boldsymbol{x}) + O(\epsilon^2).$$

We obtain:

$$\begin{aligned} O\left(\frac{1}{\epsilon}\right) &: \quad \boldsymbol{g}(\boldsymbol{x}, \Psi_0(\boldsymbol{x})) = 0, \\ O(1) &: \quad \frac{\partial \Psi_{0i}}{\partial x_j} f_j(\boldsymbol{x}, \Psi_0(\boldsymbol{x})) = \left(\frac{\partial g_i}{\partial y_j}\right)_{(\boldsymbol{x}, \Psi_0(\boldsymbol{x}))} \Psi_{1j}. \end{aligned}$$

Calling

$$J_{ij} = \left(\frac{\partial g_i}{\partial y_j}\right)_{(\boldsymbol{x}, \Psi_0(\boldsymbol{x}))},$$

the condition for this method to work clearly is that $\det J \neq 0$, hence

$$\Psi_{1j} = \left(J^{-1}\right)_{ij} \frac{\partial \Psi_{0i}}{\partial x_k} f_k(\boldsymbol{x}, \Psi_0(\boldsymbol{x}))$$

However, for our purposes, the $O(1/\epsilon)$ terms will be sufficient. As such,

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{f}(\boldsymbol{x}, \Psi(\boldsymbol{x})), \\ &= \boldsymbol{f}(\boldsymbol{x}, \Psi_0(\boldsymbol{x})) + \epsilon \left(\frac{\partial f_i}{\partial y_j}\right)_{(\boldsymbol{x}, \Psi_0(\boldsymbol{x}))} \Psi_{1i}(\boldsymbol{x}) + O(\epsilon^2), \\ &= \boldsymbol{f}(\boldsymbol{x}, \Psi_0(\boldsymbol{x})) + O(\epsilon), \end{aligned}$$

and the remaining dynamics are

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, \Psi_0(\boldsymbol{x})).$$

Remark 9.1 *The set $\boldsymbol{y} = \Psi_0(\boldsymbol{x})$ is an approximate* **invariant manifold** *of the set of ODEs* (9.1)*: if*

$$(\boldsymbol{x}(0), \boldsymbol{y}(0)) = (\boldsymbol{x}_0, \Psi(\boldsymbol{x}_0))$$

at time zero, then under the evolution (9.1)*, the system (by construction of Ψ_0) remains (to lowest order in ϵ) on the general surface $\boldsymbol{y} = \Psi_0(\boldsymbol{x})$ for all later times.*

Remark 9.2 *The dynamics*

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{f}(\boldsymbol{X}, \Psi_0(\boldsymbol{X}))$$

are an appropriate approximation for the dynamics – but only for $\epsilon \ll 1$ and time t up to $O(1)$. Also, underlying the above derivation is an assumption that $\boldsymbol{y}(0)$ is initialized close to $\Psi(\boldsymbol{x}(0))$. When this fails, further aguments are required to deal with what is essentially a **boundary / initial layer**.

9.4 Linear fast dynamics

A common generic kind of system is the following:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}), \tag{9.2a}$$

$$\frac{d\boldsymbol{y}}{dt} = \frac{1}{\epsilon}\left[-\boldsymbol{y} + \boldsymbol{g}(\boldsymbol{x})\right], \tag{9.2b}$$

where the approximate invariant manifold is obviously

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}),$$

and the slow dynamics are approximated by solution sof the equation

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{g}(\boldsymbol{X})).$$

Construct the approximate slow dynamics of the system

$$\begin{aligned}
\frac{dx_1}{dt} &= -x_2 - x_3, && \text{(9.3a)}\\
\frac{dx_2}{dt} &= x_1 + \tfrac{1}{5}x_2, && \text{(9.3b)}\\
\frac{dx_3}{dt} &= \tfrac{1}{5} + y - 5x_3, && \text{(9.3c)}\\
\frac{dy}{dt} &= -\frac{y}{\epsilon} + \frac{x_1 x_3}{\epsilon} && \text{(9.3d)}
\end{aligned}$$

Solution: The approximate invariant manifold is obvioulsy $y = x_1 x_3$. Substitute this back into the first three equations in the system (9.3) to obtain

$$\begin{aligned}
\frac{dX_1}{dt} &= -X_2 - X_3, && \text{(9.4a)}\\
\frac{dX_2}{dt} &= X_1 + \tfrac{1}{5}X_2, && \text{(9.4b)}\\
\frac{dX_3}{dt} &= \tfrac{1}{5} + X_1 X_3 - 5X_3, && \text{(9.4c)}
\end{aligned}$$

This is a Rösler system of equations – although we have simplified the original set of equations from four down to three equations, the remaining Rösler system is still known to possess chaotic orbits. By plotting the attactor of both systems we can see that the approximate or averaged set of equations does a good job of approximating the qualitative features of the full set of equations (Figure 9.1).

9.5 Centre Manifold

Another kind of commonly-encountered system in applications is the following two-dimensional system:

$$\begin{aligned}
\frac{dx}{dt} &= \lambda x + \sum_{i=0}^{2} a_i x^i y^{2-i}, && \text{(9.5a)}\\
\frac{dy}{dt} &= x - y + \sum_{i=0}^{2} b_i x^i y^{2-i}, && \text{(9.5b)}
\end{aligned}$$

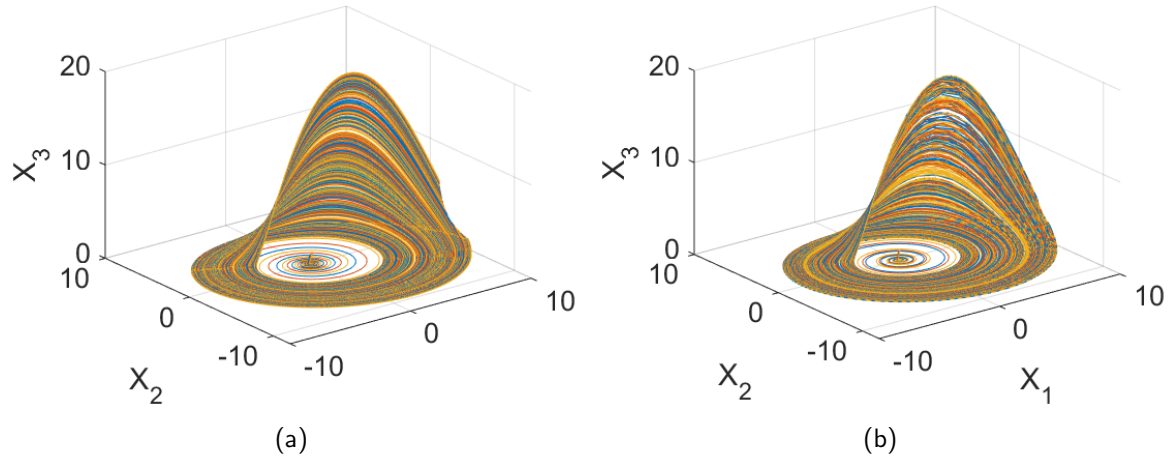


Figure 9.1: (a) Attractor of full set of four equations, projected into the (x_1, x_2, x_3) subspace; (b) Attractor of the set of three approximate equations. In both cases, $\epsilon = 0.01$.

where the nonlinearities are quadratic and are encoded in the homogeneous multinomials $\sum a_i x^i y^{2-i}$ and $\sum b_i x^i y^{2-i}$, and where λ is a real paramter.

The linearization of this system around the fixed point $(x, y) = (0, 0)$ gives

$$\frac{dx}{dt} = \lambda x, \tag{9.6a}$$

$$\frac{dy}{dt} = x - y, \tag{9.6b}$$

with Jacobian

$$\boldsymbol{A} = \begin{pmatrix} \lambda & 0 \\ 1 & -1 \end{pmatrix},$$

and hence, eigenvalues λ and -1. As λ passes through zero the linear stability peroperty of the orgin changes from stable to unstable. Therfore, studying the system (9.5) in the vicinity of $\lambda = 0$ is of interest. In particular, we can find a **centre manifold** at $\lambda = 0$, i.e. an invariant manifold tangent to the eigenspace correspoinding to the zero eigenvalue.

To obtain the centre manifold, we rescale the equations: we let

$$x \to \epsilon x, \qquad y \to \epsilon y, \qquad \lambda \to \epsilon, \qquad t \to t/\epsilon.$$

The rescaled equations read

$$\frac{dx}{dt} = \lambda x + \sum_{i=0}^{2} a_i x^i y^{2-i}, \tag{9.7a}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon}(x - y) + \sum_{i=0}^{2} b_i x^i y^{2-i}, \tag{9.7b}$$

and we can now apply the perturbation theory to obtain the lowest-order approximation to the invariant manifold: $y = x$. We substitute this into the first equation of the above pair to obtain the dynamics on the invariant manifold:

$$\frac{dX}{dt} = \lambda X + AX^2, \qquad A = \sum_{i=0}^{2} a_i.$$

Then, $\lambda = 0$ gives the centre manifold, while $\lambda < 0$ gives the stable manifold.

Chapter 10

Classification of Partial Differential Equations (*)

10.1 Outline

So far our knowledge of PDEs is based solely on the diffusion equation. This is a good start! In this chapter we place the diffusion equation in the context of general, linear partial differential equations. We move on to formulate other, more general PDEs, including quasilinear and fully nonlinear equations. For the case of second-orer quasi-linear PDEs, we make a convenient classification of PDEs according to whether they are elliptic, parabolic, or hyperbolic.

10.2 Linear operators

Definition 10.1 *Let V_1 and V_2 be real vector spaces. A linear operator T is a map*

$$\begin{aligned} T : V_1 &\rightarrow V_2, \\ \boldsymbol{x} &\rightarrow T\boldsymbol{x}, \end{aligned}$$

such that

$$\begin{aligned} T(\boldsymbol{x}+\boldsymbol{y}) &= T\boldsymbol{x} + T\boldsymbol{y}, \\ T(\lambda\boldsymbol{x}) &= \lambda T\boldsymbol{x}, \end{aligned}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in V_1$ and $\lambda \in \mathbb{R}$.

Definition 10.2 *Let V_1 and V_2 be real vector spaces and let T be a linear operator, $T : V_1 \rightarrow V_2$.*

The **kernel** *of the linear operator is the set*

$$\ker(T) = \{\boldsymbol{x} \in V_1 | T\boldsymbol{x} = 0\}.$$

The kernel of T is a vector subspace of V_1.

Examples:

- An $n \times n$ matrix is a linear operator on $\mathbb{R}^n$, and maps $\mathbb{R}^n$ to itself.
- The matrix

  $$A = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

  where a and b are nonzero real numbers is a linear operator that maps vectors $\boldsymbol{x} = (x, y)$ in $\mathbb{R}^2$ to vectors in $\mathbb{R}^2$. The kernel of A is the set of all vectors (x, y) such that

  $$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

  In other words,

  $$y = -(a/b)x,$$

  which is a one-dimensional subspace of $\mathbb{R}^2$.
- Let $C^r(\omega)$ be the space of all r-times continuously differentiable functions on the open interval $\Omega \subset \mathbb{R}$. Then the usual derivative operation is a linear operator:

  $$\begin{aligned} d/dx : C^r(\Omega) &\rightarrow C^{r-1}(\Omega), \\ f(x) &\rightarrow (df/dx), \end{aligned}$$

  since

  $$\begin{aligned} (d/dx)\left[f(x) + g(x)\right] &= (df/dx) + (dg/dx), \\ (d/dx)\left[\lambda f(x)\right] &= \lambda(df/dx), \end{aligned}$$

  for all $f(x), g(x) \in C^r(\Omega)$ and $\lambda \in \mathbb{R}$.

  The kernel of d/dx is the vector subspace of all constant functions.
- Let

  $$\Omega_t = \{(x, t) | x \in [0, L], t \in (0, \infty)\},$$
  $$\overline{\Omega_t} = \{(x, t) | x \in [0, L], t \in [0, \infty)\},$$

and let

$$\mathcal{C}^{2,1}(\Omega_t) = \left\{ u(x,t) \middle| u_{xx} \text{ is continuous in } \Omega_t \text{ and } u_t \text{ is continuous in } \Omega_t \right\},$$
$$\mathcal{C}^{2,1}(\overline{\Omega_t}) = \left\{ u(x,t) \middle| u_{xx} \text{ is continuous in } \overline{\Omega_t} \text{ and } u_t \text{ is continuous in } \overline{\Omega_t} \right\}.$$

Then the **diffusion operator**

$$\mathcal{L} = \frac{\partial}{\partial t} - D\frac{\partial^2}{\partial x^2}$$

is a linear operator that acts on the space $\mathcal{C}^{2,1}(\overline{\Omega_t})$.

10.3 The principle of superposition

As before, consider the set

$$\Omega_t = \{(x,t) | x \in [0,L], t \in (0,\infty)\},$$

and consider the diffusion operator

$$\mathcal{L} = \frac{\partial}{\partial t} - D\frac{\partial^2}{\partial x^2}.$$

on the space of functions $C^{2,1}(\Omega_t)$. The kernel of the operator is the set of solutions of $\mathcal{L}u(x,t) = 0$:

$$u(x,t) \in \ker(\mathcal{L}) \text{ iff } \mathcal{L}u(x,t) = 0,$$

This is a vector subspace of $C^{2,1}(\Omega_t)$. But a vector subspace is closed under addition and scalar multiplication:

$$u_1, u_2 \in \ker(\mathcal{L}) \implies \lambda_1 u_1 + \lambda_2 u_2 \in (\ker \mathcal{L}).$$

This result is not unique to the diffusion operator. Indeed, we have the following **principle of superposition**:

Theorem 10.1 *Let $\mathcal{M}$ be a linear differential operator on some space of suitably differentiable functions $C(\Omega)$, where Ω is the domain of definition of the operator. Then,*

$$u_1, u_2 \in \ker(\mathcal{M}) \implies \lambda_1 u_1 + \lambda_2 u_2 \in (\ker \mathcal{M}).$$

In other words, **if u_1 and u_2 satisfy the linear PDE $\mathcal{M}u = 0$, then any linear combination of these two solutions also satisfies the PDE.**

Worked example

Let $u_1(x,t)$ solve

$$\mathcal{M}u = 0, \qquad x \in (0,L),$$

where $\mathcal{M}$ is some linear operator in space and time, subject to the following **linear** boundary and initial conditions:

$$\begin{aligned} \text{BC}: &\quad u(0,t>0) = 0, \\ \text{BC}: &\quad u(L,t>0) = 0, \\ \text{IC}: &\quad u(x,t=0) = f(x), \qquad 0 \le x \le L, \end{aligned}$$

and let $u_2(x,t)$ solve

$$\mathcal{M}u = 0, \qquad x \in (0,L),$$

subject to the following **linear** boundary and initial conditions:

$$\begin{aligned} \text{BC}: &\quad u(0,t>0) = g(t), \\ \text{BC}: &\quad u(L,t>0) = 0, \\ \text{IC}: &\quad u(x,t=0) = 0, \qquad 0 \le x \le L, \end{aligned}$$

Then the linear combination

$$\lambda_1 u_1 + \lambda_2 u_2$$

solves

$$\mathcal{M}u = 0, \qquad x \in (0,L),$$

with boundary and initial conditions:

$$\begin{aligned} \text{BC}: &\quad u(0,t>0) = \lambda_2 g(t), \\ \text{BC}: &\quad u(L,t>0) = 0, \\ \text{IC}: &\quad u(x,t=0) = \lambda_1 f(x), \qquad 0 \le x \le L. \end{aligned}$$

10.4 Beyond the diffusion equation

We have mentioned that the superposition principle holds not only for the diffusion equation, but for any linear operator. Let us think about other operators we might encounter in applied mathematics:

- The wave equation for a disturbance $\phi(x,t)$:

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{\partial^2\phi}{\partial x^2},$$

- Laplace's equation in a two-dimensional domain, for a scalar field $\phi(x,y)$

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0,$$

- The linear advection equation for a concentration (scalar) field $\phi(x,t)$:

$$\frac{\partial\phi}{\partial t} + c(x,t)\frac{\partial\phi}{\partial x} = 0.$$

Definition 10.3 *Let $\phi(x,t)$ be a function of space and time.*

- *A second-order PDE in the function ϕ is a relationship of the form*

$$F\left[x,t,\phi,\partial_x\phi,\partial_t\phi,\partial_{xt}\phi,\partial_{xx}\phi,\partial_{tt}\phi\right] = 0.$$

  The PDE is **linear** *if the function F is linear in its last six variables. It is called* **quasi-linear** *if F is linear in its last three variables.*

- *A first-order PDE in the function ϕ is a relationship of the form*

$$F\left[x,t,\phi,\partial_x\phi,\partial_t\phi\right] = 0.$$

  The PDE is **linear** *if the function F is linear in its three variables. It is called* **quasi-linear** *if F is linear in its last two variables.*

1. The heat equation is a linear second-order PDE, with

$$F\left[\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7,\alpha_8\right] = \alpha_5 - D\alpha_7.$$

2. The wave equation is a linear second-order PDE, with

$$F\left[\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7,\alpha_8\right] = \frac{1}{c^2}\alpha_8 - \alpha_7$$

3. The linear advection equation is a linear first-order PDE, with

$$F\left[\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5\right] = \alpha_5 + c\left(\alpha_1,\alpha_2\right)\alpha_4.$$

4. Burgers equation,
$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = D\frac{\partial^2 u}{\partial x^2}$$
is a quasilinear PDE:
$$F\left[\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8\right] = \alpha_5 + \alpha_3\alpha_4 - D\alpha_7$$

Remark 10.1 *Other, higher-order variants exist. In general, a PDE is called linear if the function ϕ and its derivatives appear in a linear way in the PDE relation. It is called quasi-linear if the highest-order derivatives appear in a linear way.*

10.5 Classification of second-order PDEs

We focus for definiteness on a generic two-dimensional **quasi-linear** PDE of the form
$$a(x,y)u_{xx} + 2b(x,y)u_{xy} + c(x,y)u_{yy} + f(x,y,u,u_x,u_y). \tag{10.1}$$

This equation admits several special cases.

- If $a = 1, b = 0, c = -1$ and y represents time, then we get the wave equation:
$$u_{xx} - u_{yy} = f,$$
with the units chosen so that the speed of wave propagation is 1. Notice that the wave equation can be put in the alternative form
$$u_{xy} = f \tag{10.2}$$
by means of the change of variables $\xi = x - y$, $\eta = x + y$, followed by a relabelling of the independent variables.

  We no not study the wave equation at all in this module, as it is already covered in detail in Oscillations & Waves (ACM 20020).

- If $a = 1, b = 0, c = 0$, and $f = u_y - F(x,y)$, then we get the one-dimensional diffusion equation:
$$u_y = u_{xx} + F(x,y),$$
with the diffusivity normalized to 1.

  Of all the PDEs studied in this module, the diffusion equation is the one we cover in most detail. It is also referred to as the **heat equation**.

- Finally, if $a = 1, b = 0, c = 1$, and $f = F(x, y)$ we recover Poisson's equation:

$$u_{xx} + u_{yy} = F(x, y).$$

It is possible – via a suitable transformation – to reduce this general equation (10.1) down to one of the standard forms (wave equation, heat equation, Laplace/Poisson equation), which gives the required classification. This requires a lot of algebra, which we tackle now.

To achieve one of these canonical forms, we introduce a differentiable change of variable

$$\alpha = \phi(x, y), \qquad \beta = \psi(x, y), \tag{10.3}$$

and we impose that the Jacobian

$$J := \frac{\partial(\phi, \psi)}{\partial(x, y)} = \begin{vmatrix} \phi_x & \phi_y \\ \psi_x & \psi_y \end{vmatrix} \neq 0, \tag{10.4}$$

which guarantees (at least locally) that the inverse transformation

$$x = \Phi(\alpha, \beta), \qquad y = \Psi(\alpha, \beta)$$

exists. As such, we write the solution u of the PDE (10.1) in terms of the new variables:

$$\tilde{u}(\alpha, \beta) = u(\Phi(\alpha, \beta), \Psi(\alpha, \beta)).$$

In a standard abuse of notation beloved of all applied mathematicians, we now just write $u(\alpha, \beta)$ when we really mean $\tilde{u}(\alpha, \beta)$. We use the chain rule to get

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial u}{\partial \alpha}\frac{\partial \alpha}{\partial x} + \frac{\partial u}{\partial \beta}\frac{\partial \beta}{\partial x}, \\ &= u_\alpha \phi_x + u_\beta \psi_x, \end{aligned}$$

and similarly,

$$\frac{\partial u}{\partial y} = u_\alpha \phi_y + u_\beta \psi_y.$$

We have to do the same thing for the second-order derivatives. For instance,

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x}\left(u_\alpha \phi_x + u_\beta \psi_x\right), \\ &= \phi_x \frac{\partial u_\alpha}{\partial x} + \phi_{xx} u_\alpha + \psi_x \frac{\partial u_\beta}{\partial x} + \psi_{xx} u_\beta, \\ &= \phi_x \left(u_{\alpha\alpha}\phi_x + u_{\alpha\beta}\psi_x\right) + \phi_{xx} u_\alpha + \\ &\qquad \psi_x \left(u_{\alpha\beta}\phi_x + u_{\beta\beta}\psi_x\right) + \psi_{xx} u_\beta, \\ &= u_{\alpha\alpha}\phi_x^2 + 2u_{\alpha\beta}\phi_x\psi_x + u_{\beta\beta}\psi_x^2 + u_\alpha \phi_{xx} + u_\beta \psi_{xx}.\end{aligned}$$

Similarly (and no less painfully),

$$\begin{aligned}u_{yy} &= u_{\alpha\alpha}\phi_y^2 + 2u_{\alpha\beta}\phi_y\psi_y + u_{\beta\beta}\psi_y^2 + u_\alpha \phi_{yy} + u_\beta \psi_{yy}, \\ u_{xy} &= u_{\alpha\alpha}\phi_{xy} + u_{\alpha\beta}\left(\phi_x\psi_y + \phi_y\psi_x\right) + u_{\beta\beta}\psi_{xy} + u_\alpha \phi_{xy} + u_\beta \psi_{xy}.\end{aligned}$$

Now substitute these expressions into the PDE (10.1):

$$\begin{aligned}a(x,y)u_{xx} + 2b(x,y)u_{xy} + c(x,y)u_{yy} &= a\left[u_{\alpha\alpha}\phi_x^2 + 2u_{\alpha\beta}\phi_x\psi_x + u_{\beta\beta}\psi_x^2 + u_\alpha\phi_{xx} + u_\beta\psi_{xx}\right] \\ &\quad + 2b\left[u_{\alpha\alpha}\phi_{xy} + u_{\alpha\beta}\left(\phi_x\psi_y + \phi_y\psi_x\right) + u_{\beta\beta}\psi_{xy} + u_\alpha\phi_{xy} + u_\beta\psi_{xy}\right] \\ &\quad + c\left[u_{\alpha\alpha}\phi_y^2 + 2u_{\alpha\beta}\phi_y\psi_y + u_{\beta\beta}\psi_y^2 + u_\alpha\phi_{yy} + u_\beta\psi_{yy}\right], \\ &= u_{\alpha\alpha}\left(a\phi_x^2 + 2b\phi_{xy} + c\phi_y^2\right) + 2u_{\alpha\beta}\left[a\phi_x\psi_x + b\left(\phi_x\psi_y + \phi_y\psi_x\right) + c\phi_y\psi_y\right] \\ &\quad + u_{\beta\beta}\left(a\psi_x^2 + 2b\psi_{xy} + c\psi_y^2\right) + u_\alpha\left(a\phi_{xx} + 2b\phi_{xy} + c\phi_{yy}\right) + u_\beta\left(a\psi_{xx} + 2b\psi_{xy} + c\psi_{yy}\right) = f.\end{aligned}$$

Hence, the transformed PDE reads

$$A(\alpha,\beta)u_{\alpha\alpha} + 2B(\alpha,\beta)u_{\alpha\beta} + C(\alpha,\beta)u_{\beta\beta} = F_1(\alpha,\beta,u_\alpha,u_\beta), \tag{10.5}$$

where

$$\begin{aligned}A(\alpha,\beta) &= a\phi_x^2 + 2b\phi_{xy} + c\phi_y^2, \\ C(\alpha,\beta) &= a\psi_x^2 + 2b\psi_{xy} + c\psi_y^2, \\ B(\alpha,\beta) &= a\phi_x\psi_x + b\left(\phi_x\psi_y + \phi_y\psi_x\right) + c\phi_y\psi_y, \\ F_1(\alpha,\beta,u_\alpha,u_\beta) &= f - u_\alpha\left(a\phi_{xx} + 2b\phi_{xy} + c\phi_{yy}\right) \\ &\qquad - u_\beta\left(a\psi_{xx} + 2b\psi_{xy} + c\psi_{yy}\right).\end{aligned}$$

Until now, the coordinate transformation (10.3) has been left completely general. However, we have complete freedom in selecting what this is. We therefore select it so that Equation (10.5) reduces down to a simpler form, with the ultimate aim reducing the PDE down to one of the three standard

forms.

Now, if $a = c = 0$ in a region of interest, the original PDE reduces down to $2b(x, y)u_{xy} = f$. If we further assume that $b \neq 0$, then this reduces to a standard wave equation $u_{xy} = f/(2b)$ (*cf.* Equation (10.2)). If this is the case, we are done – the given PDE (10.1) is manifestly hyperbolic.

So we instead assume that one of a and c is nonzero. By symmetry of the x- and y-variables in Equation (10.1), we may assume that $a(x, y) \neq 0$ in a region of interest. **We now attempt to choose the coordinate transformation such that A and C both vanish in Equation** (10.5) – this will give the transformed PDE a very simple form.

Now, A and B both vanish if ϕ and ψ both satisfy the generic ODE

$$av_x^2 + 2bv_xv_y + cv_y^2 = 0, \tag{10.6}$$

for $v = \phi$ or $v = \psi$. This condition can be simplified by completing the square:

$$a\left(v_x - \frac{-b + \sqrt{b^2 - ac}}{a}v_y\right)\left(v_x - \frac{-b - \sqrt{b^2 - ac}}{a}v_y\right) = 0,$$

and because $a \neq 0$ in the region of interest, the solution of the PDE depends on making the terms in the $(\cdots)$ vanish. There are now three cases to consider, depending on the sign of the discriminant $b^2 - ac$.

We now look at the different possibilities for the sign of the discriminant.

10.5.1 Case 1. $b^2 - ac > 0$ in the region of interest.

In this case we call Equation (10.1) a **hyperbolic** equation. Then, Equation (10.6) reduces to

$$\phi_x - \frac{-b + \sqrt{b^2 - ac}}{a}\phi_y = 0, \qquad \psi_x - \frac{-b - \sqrt{b^2 - ac}}{a}\psi_y = 0. \tag{10.7}$$

The reason for picking the positive and negative branches of the square root in this manner is shown now – take the Jacobian (10.4)

$$J := \frac{\partial(\phi, \psi)}{\partial(x, y)} = \begin{vmatrix} \phi_x & \phi_y \\ \psi_x & \psi_y \end{vmatrix} = \phi_x\psi_y - \phi_y\psi_x.$$

Write

$$\phi_x = \frac{-b + \sqrt{b^2 - ac}}{a}\phi_y, \qquad \psi_x = \frac{-b - \sqrt{b^2 - ac}}{a}\psi_y, \tag{10.8}$$

and eliminate all x-derivatives for the Jacobian to get

$$J = \frac{2}{a}\sqrt{b^2 - ac}\,(\phi_y \psi_y)\,.$$

So, $J \neq 0$, the transformation from (x, y)-variables to (α, β)-variables is locally invertible provided $\phi_y, \psi_y \neq 0$ (If we took two square-roots with the same sign in front for both the ψ-equation and the ψ-equation, we would end up with $J = 0$). To solve Equation (10.7), focus (say) on the first equation:

$$\phi_x + s(x, y)\phi_y = 0.$$

This can be solved using the method of characteristics, which we look at in more detail later in the module, in the context of one-dimensional advection equations. For now, consider the **ordinary** differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = s(x, y).$$

This is an ordinary differential equation, and hence has a unique (local) solution, provided $s(x, y)$ is a 'reasonable' function. Denote the solution by

$$y = \tilde{y}(x),$$

with an initial condition $y = y_0$ at $x = x_0$, where (x_0, y_0) is taken to be a point in the region of interest. Then, let

$$\tilde{\phi} = \phi(x, \tilde{y}(x)).$$

Consider

$$\begin{aligned} \frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}x} &= \frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}x}, \\ &= \frac{\partial \phi}{\partial x} + s(x, y)\frac{\partial \phi}{\partial y}, \\ &= 0, \end{aligned}$$

hence, $\phi = \mathrm{const}$ $(= \phi_0 = \phi(x_0, y_0))$ along $y = \tilde{y}(x)$. This solves the PDE:

$$\phi(x, y) = \phi_0 \text{ along } y = \tilde{y}(x).$$

We justify finally the fact that $\phi_y \neq 0$. For, consider ϕ viewed momentarily as a function of the initial value ϕ_0. Then, along a trajectory,

$$1 = \frac{\partial \phi}{\partial \phi_0} = \frac{\partial \phi}{\partial x}\frac{\partial x}{\partial \phi_0} + \frac{\partial \phi}{\partial y}\frac{\partial \tilde{y(x)}}{\partial \phi_0}.$$

If $\phi_y = 0$, then by Equation (10.8) $\phi_x = 0$ also, giving $1 = 0$ and hence a contradiction in the above equation. Hence, $\phi_y \neq 0$. In the same way, we infer that $\psi_y \neq 0$, where $\psi(x, y)$ is the solution of the PDE

$$\psi_x - \left(\frac{-b - \sqrt{b^2 - ac}}{a}\right)\psi_y = 0,$$

which can be found by taking $\psi = \text{const.}$ along

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\left(\frac{-b - \sqrt{b^2 - ac}}{a}\right).$$

Finally, back-substitute into Equation (10.6):

$$\begin{aligned} B(\alpha, \beta) &= a\phi_x\psi_x + b\left(\phi_x\psi_y + \phi_y\psi_x\right) + c\phi_y\psi_y, \\ &= a\left(\frac{-b + \sqrt{\cdots}}{a}\right)\left(\frac{-b - \sqrt{\cdots}}{a}\right)\phi_y\psi_y + b\left(\frac{-b + \sqrt{\cdots}}{a} + \frac{-b - \sqrt{\cdots}}{a}\right)\phi_y\psi_y + c\phi_y\psi_y, \\ &= (ac/a^2)\phi_y\psi_y + c\phi_y\psi_y - (2b^2/a)\phi_y\psi_y, \\ &= 2\left(\frac{ac - b^2}{a}\right)\phi_y\psi_y \neq 0. \end{aligned}$$

Hence, the transformed equation (10.5) with $A = C = 0$ reads

$$u_{\alpha\beta} = F(\alpha, \beta, u, u_\alpha, u_\beta), \qquad F = F_1/2B.$$

Thus, the hyperbolic equation can be transformed into the wave equation.

10.5.2 Case 2. $b^2 - ac = 0$ in the region of interest.

If $a = 0$, then $b = 0$. Also, if $c = 0$ the PDE (10.1) makes no sense, so take $c \neq 0$. Then, the PDE is just $u_{yy} = f/c$, which is already a heat equation and thus in the correct (reduced) form. Therefore, take $a \neq 0$.

We endeavour to make $A = 0$ in the transformed PDE. This requires now that ϕ should satisfy

$$a\phi_x^2 + 2b\phi_x\phi_y + c\phi_y^2 = 0.$$

Completing the square (with $b^2 - ac = 0$) gives

$$\phi_x + (b/a)\phi_y = 0.$$

This can be solves as in Case 1 by using the method of characteristics. This fixes the coordinate transformation $\alpha = \phi(x, y)$ as required, and makes $A = 0$ identically, in the region of interest.

Happily, it also makes $B = 0$:

$$\begin{aligned} B &= a\phi_x\psi_x + b\left(\phi_x\psi_y + \phi_y\psi_x\right) + c\phi_y\psi_y, \\ &= \phi_x\left(a\phi_x + b\phi_y\right) + \psi_y\left(b\phi_x + c\phi_y\right), \\ &= a\phi_x\left(\phi_x + \frac{b}{a}\phi_y\right) + \psi_y\left(b\phi_x + c\phi_y\right), \\ &= \frac{a}{a}\psi_y\left(ab\phi_x + ac\phi_y\right), \\ &= \left(\psi_y/a\right)\left(ab\phi_x + b^2\phi_y\right), \qquad b^2 = ac, \\ &= \frac{ab}{a}\psi_y\left(\frac{b}{a}\psi_y + \phi_x\right), \\ &= 0. \end{aligned}$$

Thus, $B = 0$ identically. Therefore, $\beta = \psi(x, y)$ can be chosen to be any convenient, continuously twice-differentiable function which is linearly independent of ϕ to obtain the normal or canonical form

$$u_{\beta\beta} = F(\alpha, \beta, u_\alpha, u_\beta), \qquad F = F_1/C,$$

for a **parabolic** equation.

10.5.3 Case 3. $b^2 - ac < 0$ in the region of interest.

In this case, Equation (10.1) is called **elliptic**. It is still possible to go through the analysis already done in Case 1, leading to the first-order PDEs

$$\phi_x = \frac{-b + \sqrt{b^2 - ac}}{a}\phi_y, \qquad \psi_x = \frac{-b - \sqrt{b^2 - ac}}{a}\psi_y. \tag{10.9}$$

These are first-order linear PDEs with complex coefficients. Thus, the characteristics are complex – and the method of characteristics does not really work. (For instance, consider a 'solution' $\phi = \text{const.}$ along $dx/dt =$ complex function, where x and t are real variables. However, just because the method of characteristics fails does not mean that the PDEs (10.9) have no solution. Indeed, in many important cases, the coefficients $a(x, y)$, $b(x, y)$, and $c(x, y)$ are analytic functions of x and y (i.e. the possess convergent power-series expansions in x and y) an din this case, a solution for ϕ and ψ can be found. As before, one arrives at an equation of the form

$$u_{\alpha\beta} = G(\alpha, \beta, u, u_\alpha, u_\beta),$$

where $\alpha = \phi(x, y)$ and $\beta = \psi(x, y)$ are complex variables. To eliminate the complex variables, one takes

$$\alpha = \xi + \mathrm{i}\eta, \qquad \beta = \xi - \mathrm{i}\eta,$$

where ξ and η are real variables. The inverse transformation is got by adding/subtracting:

$$\xi = \tfrac{1}{2}(\alpha+\beta), \qquad \eta = \tfrac{1}{2}(\alpha-\beta),$$

and by the chain rule,

$$\begin{aligned}\frac{\partial\Psi}{\partial\alpha} &= \frac{\partial\Psi}{\partial\xi}\frac{\partial\xi}{\partial\alpha}+\frac{\partial\Psi}{\partial\eta}\frac{\partial\eta}{\partial\alpha},\\ &= \tfrac{1}{2}\left(\frac{\partial\Psi}{\partial\xi}+\frac{\partial\Psi}{\partial\eta}\right).\end{aligned}$$

Similarly,

$$\frac{\partial\Psi}{\partial\beta} = \tfrac{1}{2}\left(\frac{\partial\Psi}{\partial\xi}-\frac{\partial\Psi}{\partial\eta}\right)$$

Hence,

$$\begin{aligned}u_{\alpha\beta} &= \tfrac{1}{4}\left(\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\eta}\right)\left(\frac{\partial\Psi}{\partial\xi}-\frac{\partial\Psi}{\partial\eta}\right),\\ &= \tfrac{1}{4}\left(\frac{\partial^2 u}{\partial\xi^2}+\frac{\partial^2 u}{\partial\eta^2}\right).\end{aligned}$$

Hence, the normal or canonical form for elliptic equations is obtained:

$$u_{\xi\xi}+u_{\eta\eta} = F(\xi,\eta,u,u_\xi,u_\eta), \qquad F = 4G.$$

Worked example

Consider the PDE

$$yu_{xx}-xu_{yy}=0$$

in the open first quadrant $x>0$, and $y>0$. Here, $a=y$, $b=0$, and $c=-x$, so $b^2-ac=xy>0$ and the equation is hyperbolic.

We moreover construct the solution for the transformations $\alpha=\phi(x,y)$ and $\beta=\psi(x,y)$. We have

$$\phi_x-\sqrt{\frac{x}{y}}\phi_y=0.$$

We look for a solution $\phi=\text{const.}$ along

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\sqrt{\frac{x}{y}}.$$

Hence,

$$y^{3/2}-y_0^{3/2} = -x^{3/2}+x_0^{3/2}. \tag{10.10}$$

When solving the PDE, we need to look only for a solution (not 'the solution'), meaning that we can apply any initial/boundary conditions so as to get a sensible transformation. We choose to apply initial/boundary conditions

$$\phi(0,y) = y, \qquad x = 0, \qquad y > 0.$$

As such, the we put $x_0 = 0$ in the trajectory equation (10.10), to yield a solution

$$\phi = y_0, \qquad \text{on } y = \left(y_0^{3/2} - x^{3/2}\right)^{2/3}.$$

In the very final solution, y_0 is sort of an irrelevance, so we re-write

$$y_0^{3/2} = y^{3/2} + x^{3/2}, \text{ hence } y_0 = \left(x^{3/2} + y^{3/2}\right)^{2/3}.$$

But $\phi = y_0$, hence

$$\alpha = \phi(x,y) = \left(y^{3/2} + x^{3/2}\right)^{2/3}.$$

Similarly,

$$\beta = \psi(x,y) = \left(y^{3/2} - x^{3/2}\right)^{2/3}.$$

Notice that the transformation is locally valid for the triangular region $y > x$ – another transformation can be constructed for $y < x$.

In any case, the transformed PDE reads (Equation (10.6))

$$2Bu_{\alpha\beta} = -u_\alpha\left(a\phi_{xx} + 2b\phi_{xy} + c\phi_{yy}\right) - u_\beta\left(a\psi_{xx} + 2b\psi_{xy} + c\psi_{yy}\right),$$
$$B = a\phi_x\psi_x + b\left(\phi_x\psi_y + \phi_y\psi_x\right) + c\phi_y\psi_y. \quad (10.11)$$

I'm afraid there is a lot of algebra to get this down to a more acceptable form – Start with

$$\begin{aligned}
\phi_x &= \frac{x^{1/2}}{(x^{3/2}+y^{3/2})^{1/3}},\\
\phi_{xx} &= \tfrac{1}{2}x^{-1/2}\left(x^{3/2}+y^{3/2}\right)^{-1/3} - \tfrac{1}{2}x\left(x^{3/2}+y^{3/2}\right)^{-4/3},\\
\phi_y &= \frac{y^{1/2}}{(x^{3/2}+y^{3/2})^{1/3}},\\
\phi_{yy} &= \tfrac{1}{2}y^{-1/2}\left(x^{3/2}+y^{3/2}\right)^{-1/3} - \tfrac{1}{2}y\left(x^{3/2}+y^{3/2}\right)^{-4/3},\\
\psi_x &= \frac{-x^{1/2}}{(y^{3/2}-x^{3/2})^{1/3}},\\
\psi_{xx} &= -\tfrac{1}{2}x^{-1/2}\left(y^{3/2}-x^{3/2}\right)^{-1/3} + \tfrac{1}{2}x\left(y^{3/2}-x^{3/2}\right)^{-4/3},\\
\psi_y &= \frac{y^{1/2}}{(y^{3/2}-x^{3/2})^{1/3}},\\
\psi_{yy} &= \tfrac{1}{2}y^{-1/2}\left(y^{3/2}-x^{3/2}\right)^{-1/3} - \tfrac{1}{2}y\left(y^{3/2}-x^{3/2}\right)^{-4/3}.
\end{aligned}$$

Next,

$$\begin{aligned} B &= a\phi_x\psi_x + c\phi_y\psi_y, \qquad b = 0, \\ &= y\left[\frac{x^{1/2}}{(y^{3/2}+y^{3/2})^{1/3}}\right]\left[\frac{-x^{1/2}}{(y^{3/2}-x^{3/2})^{1/3}}\right] \\ &\qquad - x\left[\frac{y^{1/2}}{(y^{3/2}+y^{3/2})^{1/3}}\right]\left[\frac{y^{1/2}}{(y^{3/2}-x^{3/2})^{1/3}}\right], \\ &= -\frac{2xy}{(y^3-x^3)^{1/3}}. \end{aligned}$$

Also,

$$\begin{aligned} a\phi_{xx} + 2b\phi_{xy} + c\phi_{yy} &= y\phi_{xx} - x\phi_{yy}, \\ &= y\left[\tfrac{1}{2}x^{-1/2}(\cdots)^{-1/3} - \tfrac{1}{2}x(\cdots)^{-4/3}\right] - x\left[\tfrac{1}{2}y^{-1/2}(\cdots)^{-1/3} - \tfrac{1}{2}y(\cdots)^{-4/3}\right], \\ &= \tfrac{1}{2}\left(\frac{y}{x^{1/2}} - \frac{x}{y^{1/2}}\right)\frac{1}{(x^{3/2}+y^{3/2})^{1/3}}, \\ &= \tfrac{1}{2}\frac{1}{x^{1/2}y^{1/2}}\frac{y^{3/2}-x^{3/2}}{(x^{3/2}+y^{3/2})}, \\ &= \tfrac{1}{2}\frac{1}{x^{1/2}y^{1/2}}\frac{\beta^{3/2}}{\alpha^{1/2}}. \end{aligned}$$

Similarly (!),

$$a\psi_{xx} + 2b\psi_{xy} + c\psi_{yy} = -\tfrac{1}{2}\frac{1}{x^{1/2}y^{1/2}}\frac{\alpha^{3/2}}{\beta^{1/2}}$$

Put it all together, starting with Equation (10.11):

$$-\frac{4xy}{(y^3-x^3)^{1/3}}u_{\alpha\beta} = -\tfrac{1}{2}u_\alpha\frac{1}{x^{1/2}y^{1/2}}\frac{\beta^{3/2}}{\alpha^{1/2}} + \tfrac{1}{2}\frac{1}{x^{1/2}y^{1/2}}\frac{\alpha^{3/2}}{\beta^{1/2}}.$$

Hence,

$$8u_{\alpha\beta} = u_\alpha\frac{(y^3-x^3)^{1/3}}{x^{3/2}y^{3/2}}\frac{\beta^{3/2}}{\alpha^{1/2}} - u_\beta\frac{(y^3-x^3)^{1/3}}{x^{3/2}y^{3/2}}\frac{\alpha^{3/2}}{\beta^{1/2}}$$

Use

$$\begin{aligned} \alpha^3 - \beta^3 &= (x^{3/2}+y^{3/2})^2 - (y^{3/2}-x^{3/2})^2, \\ &= x^3 + y^3 + 2x^{3/2}y^{3/2} - y^3 - x^3 + 2x^{3/2}y^{3/2}, \\ &= 4x^{3/2}y^{3/2}, \end{aligned}$$

hence

$$x^{1/2}y^{1/2} = \left(\frac{\alpha^3-\beta^3}{4}\right)^{1/3}.$$

Also,

$$\begin{aligned}(\alpha\beta)^{1/2} &= (x^{3/2}+y^{3/2})^{1/3}(y^{3/2}-x^{3/2})^{1/3},\\ &= (y^3-x^3)^{1/3},\end{aligned}$$

hence

$$\alpha^{3/2}\beta^{3/2} = y^3 - x^3.$$

Finally then,

$$\begin{aligned}8u_{\alpha\beta} &= u_\alpha \frac{(y^3-x^3)^{1/3}}{x^{3/2}y^{3/2}}\frac{\beta^{3/2}}{\alpha^{1/2}} - u_\beta \frac{(y^3-x^3)^{1/3}}{x^{3/2}y^{3/2}}\frac{\alpha^{3/2}}{\beta^{1/2}},\\ &= u_\alpha \frac{4\alpha^{1/2}\beta^{1/2}}{\alpha^3-\beta^3}\frac{\beta^{3/2}}{\alpha^{1/2}} - u_\beta \frac{4\alpha^{1/2}\beta^{1/2}}{\alpha^3-\beta^3}\frac{\alpha^{3/2}}{\beta^{1/2}},\\ &= \frac{4u_\alpha\beta^2}{\alpha^3-\beta^3} - \frac{4u_\beta\alpha^2}{\alpha^3-\beta^3}.\end{aligned}$$

Hence,

$$u_{\alpha\beta} = \tfrac{1}{2}\frac{1}{\alpha^3-\beta^3}\left(\beta^2 u_\alpha - \alpha^2 u_\beta\right)$$

is the transformed PDE, which is manifestly a hyperbolic wave equation.

10.6 Classification – A more general framework

In the context of PDEs in more than two variables, the same hyperbolic/parabolic/elliptic classification can be achieved for second-order quasi-linear equations. In this context, it is helpful to denote the variables by $t, x_1, x_2, \cdots, x_n$ (t cna be thought of as a 'time-like' variable, although this is not necessary). As such, let $\phi(t, x_1, x_2, \cdots, x_n)$ be a smooth function of t and the coordinates $\boldsymbol{x} = (x_1, \cdots, x_n) \in \Omega$, where Ω is some open subset of $\mathbb{R}^n$. **Furthermore, let $A_{ij}(\boldsymbol{x})$, or possibly $A_{ij}(\boldsymbol{x}, t)$, be a real, symmetric matrix**.

1. Consider the PDE

$$\frac{\partial\phi}{\partial t} = \sum_{i,j=1}^{n} A_{ij}(\boldsymbol{x},t)\frac{\partial^2\phi}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\boldsymbol{x},t)\frac{\partial\phi}{\partial x_i} + c(\boldsymbol{x},t)\phi + d(\boldsymbol{x},t).$$

   The PDE is called **parabolic** if the eigenvalues of A_{ij} are all real and positive, for all t and for all $\boldsymbol{x} \in \Omega$.

2. Consider the PDE

$$\frac{\partial^2\phi}{\partial t^2} = \sum_{i,j=1}^{n} A_{ij}(\boldsymbol{x},t)\frac{\partial^2\phi}{\partial x_i x_j} + \sum_{i=1}^{n} b_i(\boldsymbol{x},t)\frac{\partial\phi}{\partial x_i} + c(\boldsymbol{x},t)\phi + d(\boldsymbol{x},t).$$

The PDE is called **hyperbolic** if the eigenvalues of A_{ij} are all real and positive, for all t and for all $\boldsymbol{x} \in \Omega$.

3. No t-dependence: Consider the PDE

$$0 = \sum_{i,j=1}^{n} A_{ij}(\boldsymbol{x})\frac{\partial^2\phi}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\boldsymbol{x})\frac{\partial\phi}{\partial x_i} + c(\boldsymbol{x})\phi + d(\boldsymbol{x}).$$

The PDE is called **elliptic** if the eigenvalues of A_{ij} are real and all have the same sign, for all $\boldsymbol{x} \in \Omega$

Notice that the type of the equation (hyperbolic/parabolic/elliptic) can change in different regions of the domain.

Worked example

To see why these classifications are important, consider the following example of a parabolic PDE – the **anisotropic diffusion equation**

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^{2} A_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j} \tag{10.12}$$

where $(x_1, x_2) = (x, y) \in \mathbb{R}^2$ and A_{ij} is a constant, **symmetric** matrix. We have the following theorem:

Theorem 10.2 *Equation* (10.12) *has a unique solution. More precisely, we have the following statement.*

Consider the PDE (10.12) *with domain Ω and boundary $\partial\Omega$, where $\partial\Omega$ is a piecewise smooth closed path. The PDE is endowed with one of the following sets of boundary conditions:*

- *Homogeneous Dirichlet: $u = 0$ on $\partial\Omega$,*
- *Homogeneous Neumann:*

$$\sum_{i,j=1}^{2} A_{ij}\widehat{n}_i\frac{\partial\phi}{\partial x_j} = 0, \qquad \text{on } \partial\Omega,$$

where $\widehat{\boldsymbol{n}}$ is the outward-pointing unit normal to the boundary.

Consider also two smooth solutions of the PDE with identical boundary and initial conditions. Then the two solutions remain identical for all time $t > 0$.

To prove this, consider two smooth solutions u_1 and u_2. Then form the difference $\phi = u_1 - u_2$. As an initial condition, the function ϕ is zero everywhere. By linearity, ϕ satisfies the anisotropic diffusion equation (10.12):

$$\frac{\partial \phi}{\partial t} = \sum_{i,j=1}^{2} A_{ij} \frac{\partial^2 \phi}{\partial x_i \partial x_j}.$$

Now multiply by ϕ and integrate over the domain Ω:

$$\begin{aligned}
\phi \frac{\partial \phi}{\partial t} &= \sum_{i,j=1}^{2} A_{ij} \phi \frac{\partial^2 \phi}{\partial x_i \partial x_j}, \\
\int_\Omega \phi \frac{\partial \phi}{\partial t} \mathrm{d}x\, \mathrm{d}y &= \int_\Omega \sum_{i,j=1}^{2} A_{ij} \phi \frac{\partial^2 \phi}{\partial x_i \partial x_j} \mathrm{d}x\, \mathrm{d}y, \\
\tfrac{1}{2} \frac{d}{dt} \int_\Omega \phi^2 \, \mathrm{d}x\, \mathrm{d}y &= \int_\Omega \sum_{i,j=1}^{2} A_{ij} \phi \frac{\partial^2 \phi}{\partial x_i \partial x_j} \mathrm{d}x\, \mathrm{d}y, \\
\tfrac{1}{2} \frac{d}{dt} \|\phi\|_2^2 &= \int_\Omega \sum_{i,j=1}^{2} A_{ij} \left[\frac{\partial}{\partial x_i} \phi \frac{\partial \phi}{\partial x_j} - \frac{\partial \phi}{\partial x_i} \frac{\partial \phi}{\partial x_j} \right] \mathrm{d}x\, \mathrm{d}y,
\end{aligned}$$

Use Gauss's theorem in the plane (Fig. 10.1):

$$\begin{aligned}
\int_\Omega \nabla \cdot \boldsymbol{v} \, \mathrm{d}x\mathrm{d}y &= \int_{\partial\Omega} \boldsymbol{u} \cdot \hat{\boldsymbol{n}} \, \mathrm{d}\ell, \\
\int_\Omega \nabla \psi \, \mathrm{d}x\mathrm{d}y &= \int_{\partial\Omega} \psi \hat{\boldsymbol{n}} \, \mathrm{d}\ell, \\
\int_\Omega \partial_i v_j \, \mathrm{d}x\mathrm{d}y &= \int_{\partial\Omega} v_j n_i \, \mathrm{d}\ell.
\end{aligned}$$

where $\hat{\boldsymbol{n}} = (n_1, n_2)$ is the unit vector **normal** to the boundary curve $\partial\Omega$ which encloses the area Ω, and where $\mathrm{d}\ell$ is the line element along the curve. Hence,

$$\tfrac{1}{2} \frac{d}{dt} \|\phi\|_2^2 = \sum_{i,j=1}^{2} A_{ij} \left(\int_{\partial\Omega} \phi \frac{\partial \phi}{\partial x_j} n_i \, \mathrm{d}\ell - \sum_{i,j=1}^{2} \int_\Omega \frac{\partial \phi}{\partial x_i} \frac{\partial \phi}{\partial x_j} \mathrm{d}x\, \mathrm{d}y \right),$$

One or other of the sets of boundary conditions will force the first term here to vanish. Thus,

$$\begin{aligned}
\tfrac{1}{2} \frac{d}{dt} \|\phi\|_2^2 &= -\sum_{i,j=1}^{2} \int_\Omega \frac{\partial \phi}{\partial x_i} \frac{\partial \phi}{\partial x_j} \mathrm{d}x\, \mathrm{d}y, \\
&:= -\int_\Omega (\nabla \phi | A | \nabla \phi) \mathrm{d}x\, \mathrm{d}y,
\end{aligned}$$

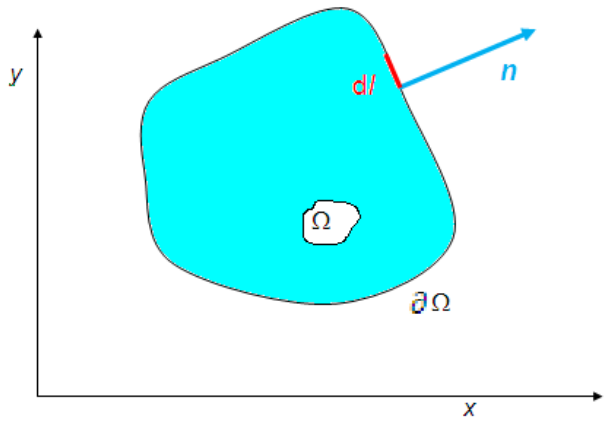


Figure 10.1: Gauss's theorem in the plane

where

$$(\boldsymbol{x}|A|\boldsymbol{x}) = (x,y)A(x,y)^T = \boldsymbol{x}\cdot(A\boldsymbol{x})$$

for all $\boldsymbol{x} := (x,y) \in \mathbb{R}^2$ is a **quadratic form**. However, we are told that the equation is parabolic, and hence A is a symmetric matrix with positive eigenvalues. That means that the quadratic form is **positive-definite**:

$$\begin{aligned}
\boldsymbol{x} &= a_1\boldsymbol{f}_1 + a_2\boldsymbol{f}_2, \qquad A\boldsymbol{f}_{(i)} = \lambda_{(i)}\boldsymbol{f}_{(i)}, \qquad \boldsymbol{f}_{(i)}\cdot\boldsymbol{f}_{(j)} = \delta_{ij},\\
A\boldsymbol{x} &= a_1\lambda_{(1)}\boldsymbol{f}_{(1)} + a_2\lambda_{(2)}\boldsymbol{f}_{(2)},\\
\boldsymbol{x}\cdot(A\boldsymbol{x}) &= (a_1\boldsymbol{f}_1 + a_2\boldsymbol{f}_2)\cdot\left(a_1\lambda_{(1)}\boldsymbol{f}_{(1)} + a_2\lambda_{(2)}\boldsymbol{f}_{(2)}\right),\\
&= \lambda_1 a_1^2 + \lambda_2 a_2^2 \geq 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
\tfrac{1}{2}\frac{d}{dt}\|\phi\|_2^2 &= -\int_\Omega (\nabla\phi|A|\nabla\phi)\mathrm{d}x\,\mathrm{d}y,\\
&\leq 0,
\end{aligned}$$

and the L^2-norm of ϕ is decreasing. In particular,

$$0 \leq \|\phi\|_2(T) \leq \|\phi\|_2(0) = 0, \qquad T \geq 0.$$

hence

$$\|\phi\|_2(T) = 0.$$

For a smooth function, the only way to satisfy this equation is for $\phi = 0$, hence

$$u_1 = u_2.$$

Thus, we conclude that **the parabolic property is essential for obtaining the uniqueness of diffusion-type equations**.

Chapter 11

Averaging methods for PDEs (*)

11.1 Outline

In this section we look at a way of computing approximate solutions to a diffusion equation with non-constant diffusivity $D(\boldsymbol{x})$ that varies rapidly in space. The general approach will be similar to the averaging approach used previously for ODEs. However, much of the theory we develop in this chapter relies on the linearity of the diffusion equation – whereas before the averaging techniques for ODEs worked quite generally.

11.2 General framework

We study the diffusion equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(\boldsymbol{x})\nabla u), \qquad (\boldsymbol{x}/\epsilon, t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{11.1a}$$

$$u = g(\boldsymbol{x}), \qquad (\boldsymbol{x}, t) \in \mathbb{R}^d \times \{0\}, \tag{11.1b}$$

where $D(\boldsymbol{x})$ is assumed to be a smooth periodic function with period ϵ in all spatial directions. Crucially, D is assumed to be **strictly positive**: $D(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$. We take $0 < \epsilon \ll 1$, meaning that the diffusion coefficient is rapidly varying whereas the initial condition is slowly varying.

11.3 Separation of scales

Because of the two lengthscales in the problem, we assume a separation of scales whereby u depends on a slowly-varying scale $\boldsymbol{x}$ and a rapidly-varying scale $\boldsymbol{y} = \boldsymbol{x}/\epsilon$:

$$u(\boldsymbol{x}; \epsilon) = u(\boldsymbol{x}, \boldsymbol{y}).$$

Assuming that these variables are indpendent, the derivative ∇ separates into to parts. For instance, in one dimension,

$$\begin{aligned} \frac{\partial}{\partial x} u(x, y) &= \frac{\partial u}{\partial x} + \frac{\partial y}{\partial x}\frac{\partial u}{\partial y}, \\ &= \frac{\partial u}{\partial x} + \frac{1}{\epsilon}\frac{\partial u}{\partial x}. \end{aligned}$$

Extending to higher dimensions, we immediately obtain (in an obvious notation)

$$\nabla = \nabla_x + \frac{1}{\epsilon}\nabla_y.$$

Hence, Equation (11.1) becomes

$$\begin{aligned} \frac{\partial u}{\partial t} &= \left(\nabla_x + \frac{1}{\epsilon}\nabla_y\right) \cdot \left[D(y)\left(\nabla_x + \frac{1}{\epsilon}\nabla_y\right) u(\boldsymbol{x}, \boldsymbol{y})\right], \\ &= D(\boldsymbol{y})\nabla_x^2 + \frac{1}{\epsilon} D\nabla_x \cdot \nabla_y u + \frac{1}{\epsilon}\nabla_y \cdot (D\nabla_x u) + \frac{1}{\epsilon^2}\nabla_y \cdot (D\nabla_y u)\,. \end{aligned} \tag{11.2}$$

We make a power-series expansion in the small parameter ϵ to obtain

$$u(\boldsymbol{x}, \boldsymbol{y}) = u_0(\boldsymbol{x}, \boldsymbol{y}) + \epsilon u_1(\boldsymbol{x}, \boldsymbol{y}) + \epsilon^2 u_2(\boldsymbol{x}, \boldsymbol{y}) + O(\epsilon^3).$$

We substitute this expansion into Equation (11.2) and equate coefficients of like powers:

$$\begin{aligned} O\left(\frac{1}{\epsilon^2}\right) &: \quad \nabla_y \cdot (D\nabla_y u_0) = 0, \\ O\left(\frac{1}{\epsilon}\right) &: \quad \nabla_y \cdot (D\nabla_x u_0) + D\nabla_x \cdot \nabla_y u_0 + \nabla_y \cdot (D\nabla_y u_1) = 0, \\ O(1) &: \quad \frac{\partial u_0}{\partial t} = D\nabla_x^2 u_0 + D\nabla_x \cdot \nabla_y u_1 + \nabla_y \cdot (D\nabla_x u_1) + \nabla_y \cdot (D\nabla_y u_2)\,. \end{aligned}$$

We solve each of these equations in turn. Because the diffusion coefficient $D(\boldsymbol{y})$ is periodic in each spatial direction with period 1, it is reasonable to impose **periodic boundary conditions** on all variations in the $\boldsymbol{y}$-direction:

$$u_0(y_i + 1) = u_0(y_i),$$

for each component y_i of the vector $\boldsymbol{y}$. For definiteness, we henceforth assume that the vector $\boldsymbol{y}$ is n-dimensional, hence $i \in \{1, \cdots n\}$.

As such, we start with the $O(\epsilon^{-2})$ equation:

$$\nabla_y \cdot (D \nabla_y u_0) = 0.$$

We multiply both sides by u_0 and integrate over the periodic domain $\boldsymbol{y} \in \mathbb{T}^n$:

$$\int_{\mathbb{T}^n} u_0 \frac{\partial}{\partial y_i} \left(D \frac{\partial u_0}{\partial y_i} \right) \mathrm{d}^n y = 0, \tag{11.3}$$

where the summation convention is assumed – i.e. the appearance of a repeated index implies summation over that index. We apply integration by parts and the periodic boundary conditions to Equation (11.3) to obtain

$$\int_{\mathbb{T}^n} D(\boldsymbol{y}) \left| \nabla_y u_0 \right|^2 \mathrm{d}^n y = 0.$$

Since $D(\boldsymbol{y}) > 0$, the only way for the above integral vanish is for $|\nabla_y u_0|^2 = 0$ identically, hence u_0 is independent of $\boldsymbol{y}$, and so the solution to the $O(1/\epsilon^2)$ problem is

$$u_0 = u_0(\boldsymbol{x}, t). \tag{11.4}$$

We now look at the $O(1/\epsilon)$ problem. The term $D \nabla_x \cdot \nabla_y u_0$ now drops out, because of Equation (11.4). So we are left with

$$\nabla_y \cdot (D \nabla_y u_1) = - \left(\nabla_y D \right) \cdot \left(\nabla_x u_0 \right). \tag{11.5}$$

We attempt a separation-of-variables solution

$$u_1(\boldsymbol{x}, \boldsymbol{y}, t) = b_i(\boldsymbol{y}) \frac{\partial u_0}{\partial x}.$$

Substitution into Equation (11.5) yields

$$\nabla_y \cdot (D \nabla_y b_i) \, \frac{\partial u_0}{\partial x} = - \frac{\partial D}{\partial y_i} \frac{\partial u_0}{\partial x}, \tag{11.6}$$

hence

$$\nabla_y \cdot (D \nabla_y b_i) = - \frac{\partial D}{\partial y_i}, \qquad i \in \{1, \cdots, n\}, \tag{11.7a}$$

with periodic boundary conditions

$$b_j(y_i + 1) = b_j(y_i), \qquad i, j \in \{1, \cdots, n\} \tag{11.7b}$$

Equation (11.7) is called the **cell problem**.

We finally move up to the $O(1)$ problem, which reads

$$\frac{\partial u_0}{\partial t} = D\nabla_x^2 u_0 + D\nabla_x \cdot \nabla_y u_1 + \nabla_y \cdot (D\nabla_x u_1) + \nabla_y \cdot (D\nabla_y u_2)\,. \tag{11.8}$$

The strategy is to integrate this equation over the $\boldsymbol{y}$-variable. We introduce the notation

$$\langle\phi\rangle(\boldsymbol{x}) = \int_{\mathbb{T}^n} \phi(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}^n y.$$

We also use the fact that the spatial dependence in the $\boldsymbol{y}$-variable is assumed to be periodic, such that

$$\int_{\mathbb{T}^n} \nabla_y \cdot (D\nabla_y u_2)\,\mathrm{d}^n y = 0.$$

This is a key step, as the dependence of the lower-order terms on the higher-order terms in the perturbation theory is thereby broken. As such, Equation (11.8) becomes

$$\begin{aligned}
\frac{\partial\langle u_0\rangle}{\partial t} &= \langle D\rangle\nabla_x^2 + \int_{\mathbb{T}^n} D(\boldsymbol{y})\nabla_x \cdot \nabla_y b_i(\boldsymbol{y})\frac{\partial u_0}{\partial x_i}\mathrm{d}^n y,\\
&= \langle D\rangle\nabla_x^2 + \left(\frac{\partial}{\partial x_j}\frac{\partial u_0}{\partial i}\right)\left(\int_{\mathbb{T}^n} D(\boldsymbol{y})\frac{\partial b_i}{\partial x_j}\mathrm{d}^n y\right),\\
&= \left[\langle D\rangle\delta_{ij} + \left(\int_{\mathbb{T}^n} D(\boldsymbol{y})\frac{\partial b_i}{\partial x_j}\mathrm{d}^n y\right)\right]\frac{\partial}{\partial x_j}\frac{\partial u_0}{\partial x_i},\\
&= (D_{\mathrm{eff}})_{ij}\,\frac{\partial}{\partial x_j}\frac{\partial u_0}{\partial x_i}.
\end{aligned}$$

Hence, the effects of the small-scale variations in the diffusivity average out and instead manifest themselves as an anisotropic **effective diffusion**

$$(D_{\mathrm{eff}})_{ij} = \left\langle D\left(\delta_{ij} + \frac{\partial b_i}{\partial y_j}\right)\right\rangle. \tag{11.9}$$

This averaging technique is called **homogenization theory**.

11.4 Regularity of the homogenization theory

The system that we started with (i.e. Equation (11.1)) is clearly parabolic, since it can be written as

$$\frac{\partial u}{\partial t} = A_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j} + \frac{\partial D}{\partial x_i}\frac{\partial u}{\partial x_i},$$

where $A_{ij} = D\delta_{ij}$ with n repeated eigenvalues $\lambda = D > 0$, thereby ensuring that the model has the parabolic property. By reference to Chapter 10, this is an extremly desirable property of PDEs of the

form $u_t = A_{ij}u_{x_ix_j} + \cdots$, as it guarantees uniqueness of (smooth solutions) and hence contributes to the well-posedness of the model. It would therefore be a great shame if the parabolic property were lost in averaging out the small scales. Therefore, in this section, we prove the following theorem:

Theorem 11.1 *The effective diffusion*

$$(D_{\text{eff}})_{ij} = \left\langle D\left(\delta_{ij} + \frac{\partial b_i}{\partial y_j}\right)\right\rangle.$$

has all real positive eigenvalues.

Proof: It suffices to show that $(D_{\text{eff}})_{ij}$ is a symmetric positive definite matrix, as a symmetric positive definite matrix has real positive eigenvalues.

Proof: By direct computation, we have

$$\begin{aligned}
(D_{\text{eff}})_{ij} &= \langle D\rangle\delta_{ij} + \int_{\mathbb{T}^n} D\frac{\partial b_i}{\partial y_j}\mathrm{d}^n y, \\
&\overset{\text{I.B.P.}}{=} \langle D\rangle\delta_{ij} - \int_{\mathbb{T}^n} \frac{\partial D}{\partial y_j} b_i \mathrm{d}^n y, \\
&\overset{\text{Cell problem}}{=} \langle D\rangle\delta_{ij} - \int_{\mathbb{T}^n} \nabla_y \cdot (D\nabla_y b_j)\, b_i \,\mathrm{d}^n y, \\
&= \langle D\rangle\delta_{ij} - \int_{\mathbb{T}^n} \frac{\partial}{\partial y_k}\left(D\frac{\partial b_j}{\partial y_k}\right) b_i \,\mathrm{d}^n y, \\
&\overset{\text{I.B.P.}}{=} \langle D\rangle\delta_{ij} + \int_{\mathbb{T}^n} D\frac{\partial b_i}{\partial y_k}\frac{\partial b_j}{\partial y_k}\,\mathrm{d}^n y.
\end{aligned}$$

Thus,

$$(D_{\text{eff}})_{ij} = \langle D\rangle\delta_{ij} + \int_{\mathbb{T}^n} D\frac{\partial b_i}{\partial y_k}\frac{\partial b_j}{\partial y_k}\,\mathrm{d}^n y,$$

which is manifestly symmetric.

We move on to the second part. We take a **constant** vector $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ in $\mathbb{R}^n$. We compute

$$\begin{aligned}
\xi_i\,(D_{\text{eff}})_{ij}\,\xi_j &= \langle \boldsymbol{\xi}, D_{\text{eff}}\boldsymbol{\xi}\rangle, \\
&= \langle D\rangle\|\boldsymbol{\xi}\|_2^2 + \int_{\mathbb{T}^n} D\left[\frac{\partial}{\partial y_k}(\xi_i b_i)\right]\left[\frac{\partial}{\partial y_k}(\xi_j b_j)\right]\mathrm{d}^n y, \\
&= \langle D\rangle\|\boldsymbol{\xi}\|_2^2 + \int_{\mathbb{T}^n} \left[\frac{\partial}{\partial y_k}(\boldsymbol{\xi}\cdot\boldsymbol{b})\right]^2 \mathrm{d}^n y;
\end{aligned}$$

since $D(\boldsymbol{y}) > 0$, we have

$$\langle \boldsymbol{\xi}, D_{\text{eff}}\boldsymbol{\xi}\rangle > 0$$

for all nonzero vectors $\boldsymbol{\xi} \in \mathbb{R}^n$, with $\langle \boldsymbol{\xi}, D_{\text{eff}}\boldsymbol{\xi}\rangle = 0$ if and only if $\boldsymbol{\xi} = 0$. Hence, the effective diffusion is positive definite, and so the homogenized diffusion equation retains the parabolic.

11.5 Worked example

Work on one spatial dimension, with a basic (unhomogenized) equation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(D(x/\epsilon)\frac{\partial u}{\partial x}\right).$$

We apply the homogenization theory, going right to the cell problem, which now takes the form

$$\frac{d}{dy}\left(D\frac{db}{dy}\right) = -\frac{dD}{dy},$$

with first integral

$$D(y)\frac{db}{dy} = -D(y) + c_0, \qquad c_0 = \text{Const.}, \tag{11.10}$$

hence

$$\frac{db}{dy} = -1 + \frac{c_0}{D(y)}. \tag{11.11}$$

We impose that $b(y)$ be a periodic function, hence $b(1) - b(0) = 0$. In other words, $\langle db/dy\rangle = 0$, hence $-1 + c_0\langle(1/D)\rangle = 0$. This fixes c_0:

$$c_0 = \langle D^{-1}\rangle^{-1}.$$

We now fill into the formula for the effective diffusion:

$$\begin{aligned} D_{\text{eff}} &= \langle D\rangle + \langle D(db/dy)\rangle, \\ &= \langle D\rangle + \langle D\left(-1 + \frac{c_0}{D}\right)\rangle, \\ &= c_0, \\ &= \langle D^{-1}\rangle^{-1}. \end{aligned}$$

Next, with $D > 0$, we write

$$1 = D^{1/2}D^{-1/2},$$

and apply the Cauchy–Schwartz inequality:

$$\begin{aligned} 1 &= \int_0^1 D^{1/2}D^{-1/2}\,\mathrm{d}y, \\ &= \langle D^{1/2}, D^{-1/2}\rangle, \\ &\leq \|D^{1/2}\|_2\|D^{-1/2}\|_2, \end{aligned}$$

hence

$$\|D^{1/2}\|_2\|D^{-1/2}\|_2 \geq 1,$$
$$\|D^{1/2}\|_2^2\|D^{-1/2}\|_2^2 \geq 1,$$
$$\langle D\rangle\langle D^{-1}\rangle \geq 1,$$

and so

$$\langle D^{-1}\rangle^{-1} \leq \langle D\rangle.$$

But $D_{\text{eff}} = \langle D^{-1}\rangle^{-1}$, hence

$$D_{\text{eff}} = \langle D^{-1}\rangle^{-1} \leq \langle D\rangle,$$

and so the effective diffusion is less than the mean diffusion.

Chapter 12

Maximum principles for second-order PDEs (*)

12.1 Outline

In this section we develop methods to characterize *a priori* the properties of the solutions of both the diffusion equation and Laplace's equation. These are important equations, as they are prototypical parabolic and elliptic equations respectively. We **assume** that smooth solutions of these equations exist, and deduce the solution propterties in the absence of knowledge of the solution's existence.

Finally, deducing *a priori* the properties of such solutions is not so silly, as this knowledge can then be turned around to find really existing solutions of these PDEs in many situations.

12.2 The basic idea

We look at the heat equation on a finite interval:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \qquad x \in (0, L), \qquad t \in (0, \infty), \tag{12.1a}$$

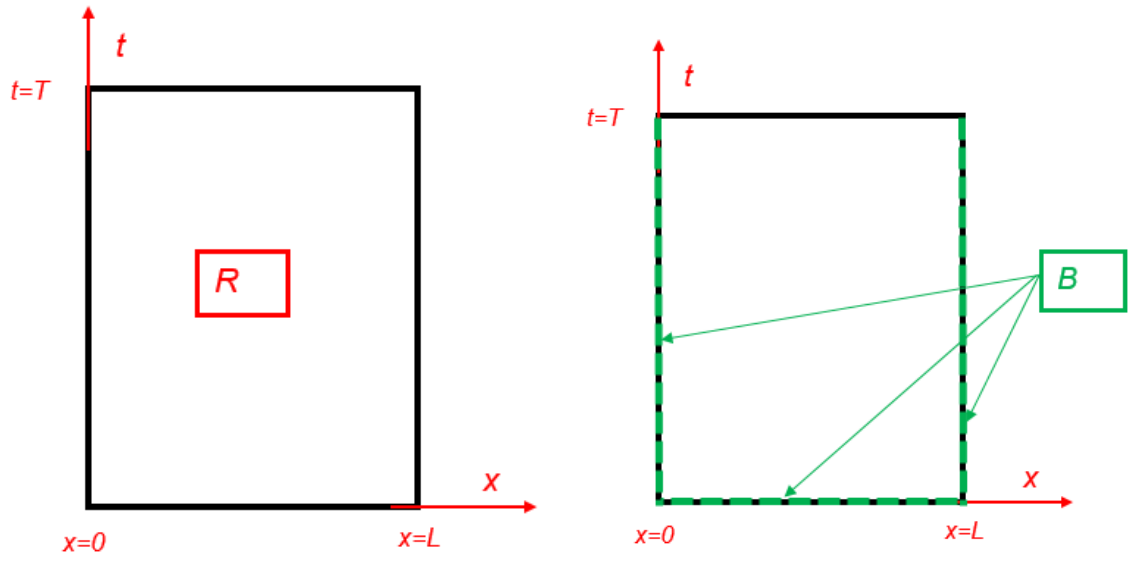
$$u(x, t = 0) = f(x), \qquad x \in [0, L], \tag{12.1b}$$

where $D > 0$ is the diffusion coefficient and $f(x)$ is a smooth function on the interval $(0, L)$ and continuous on $[0, L]$. Also, some boundary condition has to be given for $x = 0, L$, for $t \in (0, \infty)$.

We fix an arbitrary time $T > 0$ and we form the spacetime regions (Figure 12.1).

$$R = [0, L] \times [0, T], \qquad \text{(Rectangle in spacetime)},$$

$$B = ([0, L] \times \{0\}) \cup (\{0\} \times [0, T)) \cup (\{L\} \times [0, T)), \qquad \text{(Three sides of the rectangle } R\text{)}$$



(a) (b)

Figure 12.1: The different spacetime regions of intrest for the diffusion equation (12.2)

We prove the following maximum principle:

Theorem 12.1 *Let $u(x,t)$ denote a smooth solution of the diffusion equation* (12.2)*, such that*

- *u is continuous in the closed region R,*
- *u is a C^1 function of time and a C^2 function of space in the open region*

$$\mathit{Int}(R) = (0,L) \times (0,T).$$

Then the maximum of $u(x,t)$ over R is the same as the maximum of $u(x,t)$ over B:

$$\max_R u(x,t) = \max_R u(x,t).$$

In other words, the function $u(x,t)$ attains its maximum on the boundary.

Proof – we start with the basic idea. Suppose that the maximum is at (x_0,t_0). Then, we would like to conclude from basic calculus $\partial_t u = 0$ and $\partial_{xx} u < 0$, to produce

$$u_t - Du_{xx} > 0, \qquad \text{at } (x_0,t_0)$$

On the other hand, if (x_0,t_0) is in the interior of the domain we have

$$u_t - Du_{xx} = 0, \qquad \text{at } (x_0,t_0),$$

and these two equations contradict, meaning that we are forced to take (x_0, t_0) to be a boundary point. This is the basic idea of the theorem. However, we have to rule out degenerate critical points, e.g. $u_t = u_{xx} = 0$. As such, we proceed in a more formal manner, as follows.

We introduce an auxiliary function

$$v = u + \epsilon x^2,$$

hence

$$\begin{aligned} v_t - Dv_{xx} &= u_t - D\partial_{xx}\left[u + \epsilon x^2\right], \\ &= u_t - Du_{xx} - 2D\epsilon, \\ &= -2D, \qquad (x,t) \in \text{Int}(R). \end{aligned}$$

Hence

$$v_t - Dv_{xx} < 0, \qquad (x,t) \in \text{Int}(R),$$

Because v is by construction a continuous function in R, it attains a maximum in R. Let the maximum be at (x_1, t_1). Suppose that $(x_1, t_1) \in \text{Int}(R)$. Then,

$$\begin{aligned} v_t &= 0, \qquad \text{at } (x_1, t_1), \\ v_{xx} &\leq 0, \qquad \text{at } (x_1, t_1), \end{aligned}$$

where the non-strict nature of the inequality allows for the degeneracy of the maximum. Hence,

$$v_t - v_{xx} \geq 0 \qquad \text{at } (x_1, t_1).$$

Assuming $(x_1, t_1) \in \text{Int}(R)$, we also have

$$v_t - v_{xx} < 0 \qquad \text{at } (x_1, t_1).$$

These inequalities contradict, so

$$(x_1, t_1) \neq \text{Int}(R).$$

We also rule out $t_1 = T$ and $0 < x_1 < L$. Because the solution to the heat equation is defined for all $0 \leq t < \infty$, the upper limit T is arbitary. Given information $0 \leq t < T$, we can define the t-derivative of v at T.

$$v_t(x_1, T) = \lim_{h \uparrow 0} \frac{v(x_1, T) - v(x_1, T-h)}{h}.$$

Clearly, if $t_1 = T$, then $v_t(x_1, T) \geq 0$ and we again have

$$v_t - v_{xx} \geq 0 \qquad \text{at } (x_1, t_1 = T).$$

But the heat equation can be assumed to hold at $t = T$ and $x = x_1$ (with $0 < x_1 < L$), so we have

$$v_t - v_{xx} < 0 \qquad \text{at } (x_1, t_1).$$

This again produces a contradiction, we we are forced to conclude that (x_1, t_1) does not lie on the top boundary of the spacetime region. This leaves

$$(x_1, t_1) \in B.$$

Hence,

$$\begin{aligned} \max_R(v) &= \max_B(v), \\ &= \max_B(u + \epsilon x^2), \\ &\leq \max_B u + \max_B(\epsilon x^2), \\ &\leq \max_B u + \epsilon L^2. \end{aligned}$$

But also,

$$\max_B(u) \leq \max_R(u),$$

because $B \subset R$. So we have,

$$\begin{aligned} \max_R(u) &\leq \max_B(u) + \epsilon L^2, \qquad \text{for all } \epsilon > 0, \\ \max_R(u) &\geq \max_B(u). \end{aligned}$$

Since the first of these inequalities is true for all $\epsilon > 0$, we can take $\epsilon \to 0$ (through positive values) to obtain

$$\begin{aligned} \max_R(u) &\leq \max_B(u), \\ \max_R(u) &\geq \max_B(u), \end{aligned}$$

the only solution of which is

$$\max_R(u) = \max_B(u).$$

Remark 12.1 *The Maximum principle is independent of the details of the boundary conditions.*

By following essentially the same derivation as above, one can also obtain the following **minimum principle**:

Theorem 12.2 *Let $u(x, t)$ denote a smooth solution of the diffusion equation* (12.2)*, such that*

- *u is continuous in the closed region R,*
- *u is a C^1 function of time and a C^2 function of space in the open region Int(R).*

Then the **minimum** *of $u(x,t)$ over R is the same as the minimum of $u(x,t)$ over B:*

$$\max_R u(x,t) = \max_R u(x,t).$$

In other words, the function $u(x,t)$ attains its **minimum** on the boundary.

These results can be used to provide very elegant proofs to various properties of the heat equation (12.2), as described in the following section.

12.3 Consequences of the Maximum Principle

Equation (12.2) has a unique smooth solution, as specified in the following theorem:

Theorem 12.3 *Let $u_1(x,t)$ and u_2 be two smooth solutions of the diffusion equation* (12.2) *with common Dirichlet boundary conditions:*

$$u_{1,2}(0,t) = b_L(t), \qquad u_{1,2}(L,t) = b_R(t), \qquad 0 < t < \infty,$$

and a common initial condition $u_{1,2}(x,0) = f(x)$, with $x \in [0,L]$. Then $u_1(x,t) = u_(x,t)$ for all $(x,t) \in R$, for all $0 < T < \infty$.

Here, as usual, by 'smooth solution' we mean that such that

- $u_{1,2}$ are continuous in the closed region R,
- $u_{1,2}$ are C^1 functions of time and C^2 functions of space in the open region Int(R).

Proof – the idea is very standard – we let

$$u = u_1 - u_2.$$

Then, by linearity, u satisfies the diffusion equation with zero Dirichlet conditions at $x = 0$ and $x = L$, and zero initial condition. Hence,

$$u(x,t) = 0, \qquad (x,t) \in B.$$

By the maximum and minimum principles,

$$\begin{aligned} \min_R u(x,t) &= \min_B u(x,t) \overset{\text{B.C.}}{=} 0, \\ \max_R u(x,t) &= \max_B u(x,t) \overset{\text{B.C.}}{=} 0. \end{aligned}$$

Hence,

$$0 = \min_R u(x,t) \leq u(x,t) \leq \max_R u(x,t) = 0,$$

so

$$u(x,t) = 0, \qquad (x,t) \in R,$$

and so $u(x,t)$ is zero identically, hence $u_1(x,t) = u_2(x,t)$ and the smooth solution is unique.

Equation (12.2) with Dirichlet boundary conditions has the nice property that different smooth solutions that are 'close together' on the boundary are close together throughout the whole problem domain. This is an embodiment of the notion that a model should demonstrate **continuous dependence on parameters**. In precise terms, we have the following theorem:

Theorem 12.4 *Let $u_1(x,t)$ and u_2 be two smooth solutions of the diffusion equation* (12.2) *with Dirichlet boundary and initial data that are ' close' in a sense to be specified as follows. Then the solutions remain close throughout the entire domain of the solution. In other words, if*

$$\begin{aligned} u_{1,t} &= Du_{1,xx}, \quad x \in (0,L), \quad 0 < t < T & \qquad u_{2,t} &= Du_{2,xx}, \quad x \in (0,L), \quad 0 < t < T \\ u_1(0,t) &= b_L(t), \quad 0 < t < T & \qquad u_2(0,t) &= c_L(t), \quad 0 < t < T \\ u_1(L,t) &= b_R(t), \quad 0 < t < T & \qquad u_2(L,t) &= c_R(t), \quad 0 < t < T \\ u_1(x,0) &= f(x), \quad x \in [0,L] & \qquad u_2(x,0) &= g(x), \quad x \in [0,L] \end{aligned}$$

with

$$\begin{aligned} \left|b_(t) - c_L(t)\right| &\leq \epsilon, \quad t \in (0,T), \\ |b_R(t) - c_R(t)| &\leq \epsilon, \quad t \in (0,T), \\ |f(x) - g(x)| &\leq \epsilon, \quad x \in [0,L], \end{aligned}$$

for all $t \in (0,T)$ and $x \in [0,L]$, and for any $T \in (0,\infty)$, then the solutions remain similarly close for all time:

$$|u_1(x,t) - u_2(x,t)| \leq \epsilon, \qquad (x,t) \in R.$$

Proof: As in the proof of uniqueness, let $v(x,t) = u_1 - u_2$. Then, $v_t = Dv_{xx}$ on Int(R), and

$$\begin{aligned} |v(0,t)| &\leq \epsilon, \qquad t \in (0,T), \\ |v(L,t)| &\leq \epsilon, \qquad t \in (0,T), \\ |v(x,0)| &\leq \epsilon, \qquad t \in (0,T). \end{aligned}$$

In other words,

$$\max_B v \leq \epsilon, \qquad \min_B v \geq -\epsilon.$$

By the maximum / minimum principles,

$$\max_R v \leq \epsilon, \qquad \min_R v \geq -\epsilon,$$

hence

$$-\epsilon \leq v(x,t) \leq \epsilon, \qquad (x,t) \in R,$$

hence,

$$|v(x,t)| = |u_1(x,t) - u_2(x,t)| \leq \epsilon \qquad (x,t) \in R,$$

Summarizing,

- The diffusion equation (12.2) with Dirichlet boundary conditions has a smooth solution – *this can be constructed using Fourier series, as in Chapter 6*.
- The diffusion equation with Dirichlet boundary conditions has a **unique** smooth solution – *shown using the maximum / minimum principles as in Theorem 12.3, or using energy methods, as in Chapter 10*.
- The diffusion equation with Dirichlet boundary conditions has exhibits continuous dependence on the model parameters (boundary and initial data) – *shown again using the maximum / minimum principles and Theorem 12.4*.

In this way, the diffusion equation (12.2) with Dirichlet boundary conditions possesses all three conditions for a **well posed model**.

Remark 12.2 *A smooth solution of diffusion equation* (12.2) *with homogeneous Neumann boundary conditions*

$$\partial_x u(0,t) = \partial_x u(L,t), \qquad t \in (0,\infty)$$

is unique **only up to a constant**. *In other words, if* $u_1(x,t)$ *and* $u_2(x,t)$ *are two smooth solutions*

with homogeneous Neumann boundary conditions and the same initial conditions, then

$$u_2(x,t) - u_1(x,t) = \textit{Const.}$$

We look one last time at our maximum-principle result (Theorem 12.1):

$$\max_R u = \max_R u.$$

This is rather weak statement – it says that the solution $u(x,t)$ of the heat equation attains its maximum on the boundary – the statement does not rule out the possibility that $u(x,t)$ attains its maximum elsewhere. However, by going through the theorem carefully, we see from the auxiliary function $v(x,t) = u(x,t) + \epsilon x^2$ that the maximum (x_1, t_1) is necessarily on the boundary B. The only slightly anomalouse case is when u is a constant, which gives rise to the following **strong maximum principle**:

Theorem 12.5 *Let $u(x,t)$ be a smooth solution of Equation* (12.2)*. If $u(x,t)$ attains its maximum at $(x_1, t_1) \in$ Int(R) then $u(x,t)$ is a constant.*

Something has been glossed over here, although the abouve intuition suffices for our purposes. Indeed, proving the strong maximum principle rigorously relies on the notion of a **heat kernel**, which for reasons of time we do not go into in this module.

12.4 Extensions of the theorem

The maximum principle extends to non-constant diffusion models – very similar to the ones we looked at in Chapter 11:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(D(x)\frac{\partial u}{\partial x}\right) \qquad x \in (0,L), \qquad t \in (0,\infty), \tag{12.2a}$$

$$u(x, t=0) = f(x), \qquad x \in [0,L], \tag{12.2b}$$

where $D(x) > 0$ is a positive smooth function, and $f(x)$ is a smooth function on the interval $(0,L)$ and continuous on $[0,L]$. Also, some boundary condition has to be given for $x = 0, L$, for $t \in (0,\infty)$.

To prove the maximum theorem in this instance, one proceeds by analogy with Theorem 12.1; however, the auxiliary function should be

$$v(x,t) = u + \epsilon \int_0^x \frac{x}{D(x)} \mathrm{d}x,$$

with

$$\begin{aligned} v_t - \partial_x (Dv_x) &= u_t - \partial_x (Du_x) - \epsilon \frac{\partial}{\partial x}\left(D\frac{x}{D}\right), \\ &= -\epsilon, \qquad (x,t) \in \text{Int}(R), \end{aligned}$$

hence

$$v_t - \partial_x (Dv_x) < 0, \qquad (x,t) \in \text{Int}(R)$$

As before, let (x,t) attain its maximum at (x_1,t_1). Then

$$\begin{aligned} v_t - \partial_x(Dv_x) &= -\epsilon, \qquad \text{for all } (x,t) \in \text{Int}(R), \\ &= v_t - (D_x)(v_x) - Dv_{xx}, \\ &= -Dv_{xx}, \qquad \text{at } (x_1,t_1) \text{ assumed to be in Int}(R), \\ &\geq 0, \qquad \text{at } (x_1,t_1), \end{aligned}$$

which gives the necessary contradiction, such that

$$(x_1,t_1) \in B.$$

The remaining parts of the proof are indentical to Theorem 12.1.

We can also extend the maximum principle to higher dimensions. We look at the heat equation defined on a domain Ω, where Ω is a bounded, simply connected region of $\mathbb{R}^n$ with smooth curve C. The model equation to solve is

$$\frac{\partial u}{\partial t} = D\nabla^2 u, \qquad \boldsymbol{x} \in \Omega, \qquad t \in (0,\infty), \tag{12.3a}$$

$$u(\boldsymbol{x}, t=0) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \overline{\Omega}, \tag{12.3b}$$

where $D > 0$ is the diffusion coefficient and $f(x)$ is a smooth function in the interval Ω and continuous on the closure $\overline{\Omega}$. Also, some boundary conditions have to be imposed on $u(\boldsymbol{x},t)$ for $\boldsymbol{x} \in C$, for $t \in (0,\infty)$ – these would typically be Neumann or Dirichlet.

Then, both the weak and strong versions of the Maximum principle carry over – the only difference now is the understanding of the spacetime regions R and $B - R$ is a solid cylinder and B is its boundary with the top cut off (Figure 12.2).

We can clearly extend things further – to problems such as

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(\boldsymbol{x})\nabla u), \qquad \boldsymbol{x} \in \Omega, \qquad t \in (0,\infty), \tag{12.4a}$$

$$u(\boldsymbol{x}, t=0) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \overline{\Omega}, \tag{12.4b}$$

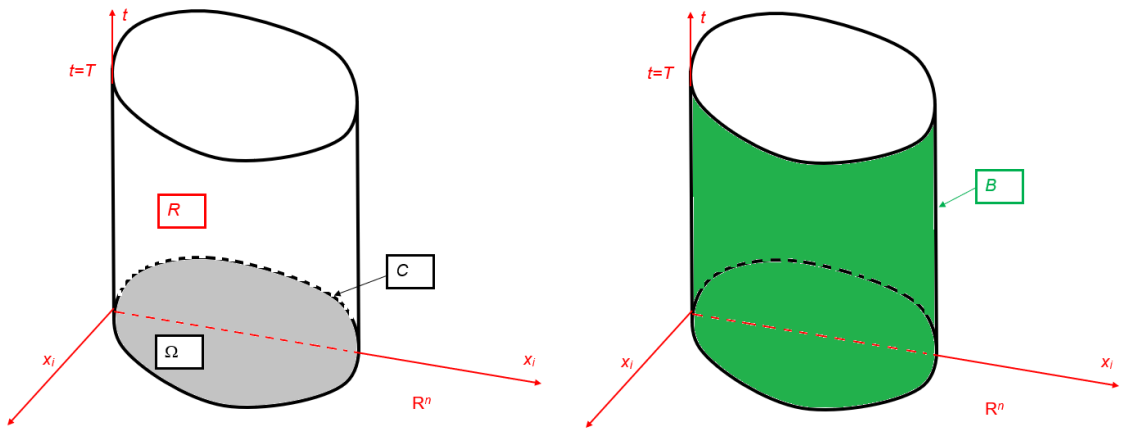


Figure 12.2: The different spacetime regions of intrest for the diffusion equation (12.3)

where $D(\boldsymbol{x}) > 0$ is the diffusion coefficient (smooth function). Even more generalizations are permitted:

- Diffusion tensor with $D_{ij}(\boldsymbol{x})$ as a smooth function and also, a positive-definite matrix at each point $\boldsymbol{x} \in \Omega$
- Certain extra terms in the model, e.g. $u_t = D\nabla^2 u + F(u)$ or $u_t = D\nabla^2 u + s(\boldsymbol{x})$
- Combinations of the above.

All of these models are of relevance in applications, some of which we will see again in the final chapter of the notes.

12.5 Relaxation to Laplace's Equation

We look again at the heat equation defined on a bounded domain $\Omega \subset \mathbb{R}^n$, with Dirichlet boundary conditions:

$$\frac{\partial u}{\partial t} = D\nabla^2 u, \qquad \boldsymbol{x} \in \Omega, \qquad t \in (0, \infty), \tag{12.5a}$$

$$u(\boldsymbol{x}, t = 0) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \overline{\Omega}, \tag{12.5b}$$

$$u(\boldsymbol{x}, t) = g(\boldsymbol{x}), \qquad \boldsymbol{x} \in C, \qquad t \in (0, \infty) \tag{12.5c}$$

where $D > 0$ is the diffusion coefficient and $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are smooth functions and as usual, Ω is a bounded simply connnected region of $\mathbb{R}^n$, with smooth boundary C. Separately, we look at

Laplace's equation on the same domain:

$$0 = D\nabla^2 u_0, \qquad \boldsymbol{x} \in \Omega, \qquad t \in (0,\infty), \tag{12.6a}$$

$$u_0(\boldsymbol{x},t) = g(\boldsymbol{x}), \qquad \boldsymbol{x} \in C, \qquad t \in (0,\infty) \tag{12.6b}$$

It turns out that the solution of the diffusion equation (12.5) relaxes to the solution of Laplace's equation (12.6), as $t \to \infty$:

Theorem 12.6 *Let $u(\boldsymbol{x},t)$ be a smooth solution of Equation* (12.5) *and let $u_0(\boldsymbol{x})$ be a smooth solution of Equation* (12.6)*. Then*

$$\lim_{t\to\infty} u(\boldsymbol{x},t) = u_0(\boldsymbol{x}).$$

To prove this result we first of all need to auxiliary results:

Theorem 12.7 (Poincaré's inequality) *Let Ω be an open, bounded and connected subset of $\mathbb{R}^n$. Denote by $H_1^(\Omega)$ the set of all functions $f : \Omega \to \mathbb{R}$ such that f and its gradient ∇f are both square integrable. Then, there exist positive constants C_1 and C_2 such that*

$$\int_\Omega [f(\boldsymbol{x})]^2 \mathrm{d}^n x \le C_1 \int_\Omega |\nabla f|^2 \mathrm{d}^n x + C_2 \left(\int_\Omega f(\boldsymbol{x})\, \mathrm{d}^n \right)^2,$$

for all functions $f \in H^1(\Omega)$.

The constants C_1 and C_2 are independent of the function f and depend only on the shape of the domain. For mean-zero functions, the inequality reduces to

$$\int_\Omega [f(\boldsymbol{x})]^2 \mathrm{d}^n x \le C_1 \int_\Omega |\nabla f|^2 \mathrm{d}^n x.$$

The general proof of Poincarés inequality eludes us here; we however prove it for the special case where $\Omega = (0,L)^n$, where $f(\boldsymbol{x})$ is a periodic mean-zero function, which we write here in compact exponential notation as

$$f(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} f_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}},$$

where the sum is over all vectors $\boldsymbol{k} = (k_1, \cdots, k_n)$, with $k_j = (2\pi/L)j$, with $j \in \mathbb{Z}$, but $\boldsymbol{k} \neq 0$ because the function has mean zero. By orthogonality of the exponentials,

$$\int_{[0,L]^n} \mathrm{e}^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}} \mathrm{d}^n x = \delta_{\boldsymbol{k},\boldsymbol{k}'} L^n,$$

we obtain

$$f_{\boldsymbol{k}} = \int_{[0,L]^n} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}x,$$

hence

$$f^*_{\boldsymbol{k}} = f_{-\boldsymbol{k}},$$

for real-valued functions $f(\boldsymbol{x})$. Similarly, we obtain Parseval's identity,

$$\int_{[0,L]^n} |f(\boldsymbol{x})|^2 \mathrm{d}^n x = L^n \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} |f_{\boldsymbol{k}}|^2.$$

We apply the same reasoning to ∇f to obtain

$$\begin{aligned} \nabla f &= \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} \mathrm{i}\boldsymbol{k} f_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \\ \int_{[0,L]^n} |\nabla f(\boldsymbol{x})|^2 \mathrm{d}^n x &= L^n \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} k^2 |f_{\boldsymbol{k}}|^2. \end{aligned}$$

Hence,

$$\begin{aligned} \int_{[0,L]^n} |\nabla f(\boldsymbol{x})|^2 \mathrm{d}^n x &= L^n \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} k^2 |f_{\boldsymbol{k}}|^2, \\ &\geq \min(k^2) L^n \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} |f_{\boldsymbol{k}}|^2, \\ &= (2\pi/L)^2 L^n \sum_{\substack{\boldsymbol{k} \\ \boldsymbol{k}\neq 0}} |f_{\boldsymbol{k}}|^2, \\ &= (2\pi/L)^2 \int_{[0,L]^n} |f(\boldsymbol{x})|^2 \mathrm{d}^n, \end{aligned}$$

hence

$$\int_{[0,L]^n} |\nabla f(\boldsymbol{x})|^2 \mathrm{d}^n x \geq (2\pi/L)^2 \int_{[0,L]^n} |f(\boldsymbol{x})|^2 \mathrm{d}^n,$$

and hence, $C_1 = (2\pi/L)^2$ and $C_2 = 0$, for periodic mean-zero functions on $\Omega = (0, L)^n$.

A second result that is required is **Gronwall's inquality**:

Theorem 12.8 *Let the following differential* **inequality** *be given:*

$$\frac{dy}{dx} \leq Q(x) y(x), \qquad x \in (a, b), \qquad y(a) = y_0, \qquad a < b,$$

where $Q(x)$ is a smooth function. Then $y(x)$ is bounded by the solution of the corresponding

differential equation $dy/dx = Q(x)y$, *i.e.*

$$y(x) \leq y(a) \exp\left(\int_a^x Q(x')\mathrm{d}x'\right).$$

A proof is not given here but it is readly available by an integrating-factor argument: as the integrating factor $\mu = \exp(\int_a^x Q(x')\mathrm{d}x')$ is positive, the inequality can be multiplied across by μ without changing the direction of the inequality. The proof proceeds from this observation.

In any case, we now look again at the solution $u(x,t)$ of the diffusion equation (12.5) and the solution $u_0(x)$ of Laplace's eqeuation (**??**); we form the difference

$$v(\boldsymbol{x},t) = u(\boldsymbol{x},t) - u_0(\boldsymbol{x}).$$

By linearity, $v(\boldsymbol{x},t)$ satisfies the diffusion equation

$$v_t = D\nabla^2 v, \qquad \boldsymbol{x} \in \Omega, \qquad t \in (0,\infty), \tag{12.7}$$

with zero Dirichlet boundary conditions and an initial condition $v(\boldsymbol{x},0) = f(\vec{)} - u_0(\boldsymbol{x})$. We multiply both sides of Equation (12.7) by v and integrate over Ω:

$$\begin{aligned}
\int_\Omega v v_t \,\mathrm{d}^n x &= \int v\nabla^2 v \,\mathrm{d}^n x, \\
\tfrac{1}{2}\frac{d}{dt}\int_\Omega v^2 \,\mathrm{d}^n x &= \int_\Omega \left[\nabla\left(v\nabla v\right) - |\nabla v|^2\right], \\
\tfrac{1}{2}\frac{d}{dt}\|v\|_2^2 &\overset{\text{Gauss=}}{=} \int_C v(\nabla v)\cdot\boldsymbol{S} - \int_\Omega |\nabla v|^2 \,\mathrm{d}^n, \\
&= -\int_\Omega |\nabla v|^2 \,\mathrm{d}^n,
\end{aligned}$$

where we have used Gauss's theorem with $\mathrm{d}\boldsymbol{S} = \widehat{\boldsymbol{n}}\mathrm{d}S$, i.e. $\widehat{\boldsymbol{n}}$ is the outward-pointing unit normal on the surface C and $\mathrm{d}S$ and element of surface area on the same. In the same place we have further used $v = 0$ on C.

We now apply Poincaré's inequality:

$$\begin{aligned}
\tfrac{1}{2}\frac{d}{dt}\|v\|_2^2 &= -\int_\Omega |\nabla v|^2 \,\mathrm{d}^n, \\
&\leq -C_2\|v\|_2^2,
\end{aligned}$$

and so by Gronwall's inequality,

$$\|v\|_2^2(t) \leq \|v\|_2^2(0)\mathrm{e}^{-2C_2 t},$$

hence

$$\lim_{t\to\infty} \|v\|_2^2(t) = 0,$$

and since $v(x,t)$ is a continuous function,

$$\lim_{t\to\infty} v(\boldsymbol{x},t) = 0,$$

i.e.

$$\lim_{t\to\infty} u(\boldsymbol{x},t) = u_0(\boldsymbol{x}).$$

Now that we have established that the diffusion equation relaxes to a solution of a corresponding Laplace equation for the case of Dirichlet boundary conditions (but also, Neumann boundary conditions), it makes sense to look briefly at Laplace's equation in its own right, which we do in the next section.

12.6 Laplace's Equation – Properties

We start with the following PDE:

$$\nabla^2 u = 0 \qquad \boldsymbol{x}\in\Omega, \qquad u = f(\boldsymbol{x}), \qquad \boldsymbol{x}\in\partial\Omega, \tag{12.8}$$

where $\Omega\subset\mathbb{R}^n$ is a bounded, simply connected domain with the smooth boundary $\partial\Omega$, and $f(\boldsymbol{x})$ is a smooth function.

We have the following definitions:

Definition 12.1 *Let Ω be an open, bounded, and simply connected subset of $\mathbb{R}^n$, and let $u(\boldsymbol{x})$ be harmonic on Ω. Let $\boldsymbol{x}_0\in\Omega$. Then, there exists a real number $r>0$ such that the open ball of radius r and centred at $\boldsymbol{x}_0$ is entirely contained in Ω. We have some notation:*

- *$B(\boldsymbol{x}_0,r)$ denotes the open ball of radius r and centred at $\boldsymbol{x}_0$ is entirely contained in Ω. The volume of the ball B is denoted by $|B|$.*
- *$S(\boldsymbol{x}_0,r)$ is the boundary sphere of $B(\boldsymbol{x}_0,r)$. The area of the boundary sphere is denoted by $|S|$.*
- *The symbol $\mathrm{d}\omega_n$ denotes the differential element of solid angle in $\mathbb{R}^n$, and the following identity holds:*

$$\mathrm{d}^n x = r^{n-1}\mathrm{d}r\,\mathrm{d}\omega_n, \qquad r=|\boldsymbol{x}|, \tag{12.9}$$

- *Integrate both sides of Equation* (12.9) *to obtain* $|B|$*, the volume of the ball in* $\mathbb{R}^n$*, of radius* R*:*

$$|B| = \int_0^R r^{n-1}\mathrm{d}r \int_{\omega_n} \mathrm{d}\omega_n,$$

where the subscript in $\int_{\omega_n}$ *denotes integration over all solid angles. Thus,*

$$|B| = \frac{1}{n}R^n|\omega_n|,$$

where $|\omega_n| = \int_{\omega_n} \mathrm{d}\omega_n$ *is the area of the unit sphere in* $\mathbb{R}^n$*.*

Example: In $\mathbb{R}^2$, we have $\mathrm{d}\omega_{n=2} = \mathrm{d}\varphi$, where φ is the polar angle in the usual polar coordinates. Thus, $|\omega_2| = \int_0^{2\pi} \mathrm{d}\varphi = 2\pi$. Also, $|B_{n=2}| = (1/2)R^2(2\pi) = \pi R^2$.

In $\mathbb{R}^3$, we have $\mathrm{d}\omega_{n=3} = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi$, where again, (θ, φ) denote the usual polar coordinates: θ is the polar angle, and φ is the azimuthal angle. Thus, $|\omega_3| = \int_0^{\pi} \sin\theta\,\mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi = 4\pi$. Also, $|B_{n=3}| = (1/3)R^3(4\pi) = (4/3)\pi R^3$.

Next, we define boundary-averages and area-averages of $u(\boldsymbol{x})$ as follows:

- Boundary average:

$$\mathrm{av}_{S(\boldsymbol{x}_0,r)}u := \frac{1}{|\omega_n|}\int_{\omega_n} u(\boldsymbol{x}_0 + r\hat{\boldsymbol{r}})\,\mathrm{d}\omega_n$$

where $\hat{\boldsymbol{r}}$ is the unit radial vector expressed as a function of the pertinent angular variables in $\mathbb{R}^n$.

- Volume average:

$$\mathrm{av}_{B(\boldsymbol{x}_0,r)}u := \frac{1}{|B|}\int_{B(\boldsymbol{x}_0,r)} u(\boldsymbol{x})\mathrm{d}^n x,$$

12.6.1 The Mean-Value Theorem for Harmonic Functions

We have the following theorem:

Theorem 12.9 (Mean-value theorem, harmonic functions) *Let* Ω *be an open, bounded, and simply-connected subset of* $\mathbb{R}^n$*, and let* $u(\boldsymbol{x})$ *be harmonic on* Ω*. Specifically, let* $u \in C^2(\Omega)\cap C^0(\overline{\Omega})$*. Let* $\boldsymbol{x}_0 \in \Omega$*. Then, for a ball* $B(\boldsymbol{x}_0, r)$ *contained entirely in* Ω*,*

$$u(\boldsymbol{x}_0) = \mathrm{av}_{S(\boldsymbol{x}_0,r)}u = \mathrm{av}_{B(\boldsymbol{x}_0,r)}u.$$

Proof: We start with the boundary average. We call

$$\phi(r, \boldsymbol{x}_0) := \mathrm{av}_{S(\boldsymbol{x}_0,r)}u = \frac{1}{|\omega_n|}\int_{\omega_n} u(\boldsymbol{x}_0 + r\hat{\boldsymbol{r}})\,\mathrm{d}\omega_n.$$

We note that $\phi(0, \boldsymbol{x}_0) = u(\boldsymbol{x}_0)$. If we could show that $\partial\phi/\partial r = 0$, then we would be done, since we would then have that

$$\phi(r, \boldsymbol{x}_0) = \phi(0, \boldsymbol{x}_0) = u(\boldsymbol{x}_0).$$

We compute:

$$\begin{aligned}
\frac{\partial \phi}{\partial r} &= \frac{1}{|\omega_n|} \int_{\omega_n} \left[\frac{\partial}{\partial r} u(\boldsymbol{x}_0 + r\hat{\boldsymbol{r}}) \right] \mathrm{d}\omega_n, \\
&= \frac{1}{|\omega_n|} \int_{\omega_n} \left[\hat{\boldsymbol{r}} \cdot \nabla u \right]_{\boldsymbol{x}_0 + r\hat{\boldsymbol{r}}} \mathrm{d}\omega_n, \\
&= \frac{1}{|\omega_n|} \int_{\omega_n} \left[(\nabla u)_{\boldsymbol{x}} \right] \cdot (\hat{\boldsymbol{r}}\, \mathrm{d}\omega_n), \qquad \boldsymbol{x} = \boldsymbol{x}_0 + r\hat{\boldsymbol{r}}, \\
&= \frac{1}{|\omega_n| r^{n-1}} \int_{\omega_n} (\nabla u) \cdot \mathrm{d}\boldsymbol{S}, \qquad \mathrm{d}\boldsymbol{S} = r^{n-1} \mathrm{d}\omega_n \hat{\boldsymbol{r}}, \\
&= \frac{1}{|\omega_n| r^{n-1}} \int_B \nabla^2 u \, \mathrm{d}^n x, \qquad \ldots \qquad \text{Gauss's theorem} \\
&= 0.
\end{aligned}$$

Hence, the first part is shown.

For the second part, we also do a direct calculation:

$$\begin{aligned}
\frac{1}{|B|} \int_B u(\boldsymbol{x}) \mathrm{d}^n x &= \frac{1}{|B|} \int_0^r r^{n-1} \mathrm{d}r \int_{\omega_n} u(\boldsymbol{x}_0 + r\hat{\boldsymbol{r}})\, \mathrm{d}\omega_n, \\
&= \frac{|\omega_n|}{|B|} \int_0^r r^{n-1}\, \mathrm{d}r \left[\mathrm{av}_{S(\boldsymbol{x}_0, r)} u \right], \\
&= \frac{n}{r^n} u(\boldsymbol{x}_0) \int_0^r r^{n-1}\, \mathrm{d}r, \\
&= u(\boldsymbol{x}_0).
\end{aligned}$$

Putting it all together, we have the following **mean-value theorem** for harmonic functions:

$$u(\boldsymbol{x}_0) = u(\boldsymbol{x}_0) = \mathrm{av}_{S(\boldsymbol{x}_0, r)} = \mathrm{av}_{B(\boldsymbol{x}_0, r)}.$$

12.6.2 The Maximum Principle for Harmonic Functions

The weak maximum principle also follows from this result:

Theorem 12.10 (Maximum principle, harmonic functions) *Let Ω be an open, bounded, and simply-connected subset of $\mathbb{R}^n$ and let $u(\boldsymbol{x})$ be harmonic on Ω, with $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$. Then*

$$\max_{\overline{\Omega}} u(\boldsymbol{x}) = \max_{\partial\Omega} u(\boldsymbol{x}).$$

At the same time, the strong maximum principle also follows:

Theorem 12.11 *Let Ω be an open, bounded, and simply-connected subset of $\mathbb{R}^n$ and let $u(\boldsymbol{x})$ be harmonic on Ω, with $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$. If $u(\boldsymbol{x})$ attains its maximum in Ω then $u(\boldsymbol{x})$ is a constant function.*

Proof: Let $u(\boldsymbol{x})$ attain its maximum over $\overline{\Omega}$ at $\boldsymbol{x}_0$. If $\boldsymbol{x}_0 \in \partial\Omega$, the theorem is proved. Thus, consider the case where $\boldsymbol{x}_0 \in \Omega$, with $M = u(\boldsymbol{x}_0)$. Then, by the topology of the set Ω, and by the Mean-Value Theorem, we can write

$$u(\boldsymbol{x}_0) = \frac{1}{|B|} \int_{B(\boldsymbol{x}_0, r)} u(\boldsymbol{x}) \mathrm{d}^n x,$$

where $r > 0$ is a positive number. Hence,

$$\max_{B(\boldsymbol{x}_0, r)} u(\boldsymbol{x}) = \mathrm{avg}_{B(\boldsymbol{x}_0, r)} u(\boldsymbol{x}), \tag{12.10}$$

and this result extends to the closed ball $\overline{B(\boldsymbol{x}_0, r)}$ because of the mean value theorem (boundary averages). Thus, the maximum of the function is actually the mean value of the function on $\overline{B(\boldsymbol{x}_0, r)}$, and hence

$$u(\boldsymbol{x}) = M, \qquad \boldsymbol{x} \in \overline{B(\boldsymbol{x}_0, r)}. \tag{12.11}$$

We now extend this result to cover the entire domain Ω. Thus, choose a point $\boldsymbol{x}_1 \in \partial B(\boldsymbol{x}_0, r)$, with $u(\boldsymbol{x}_1) = M$. Choose a ball $B(\boldsymbol{x}_1, r')$ contained entirely in Ω and conclude that

$$u(\boldsymbol{x}) = M, \qquad \boldsymbol{x} \in \overline{B(\boldsymbol{x}_1, r')}.$$

By covering the set Ω with a collection of overlapping balls in this manner, it follows that

$$u(\boldsymbol{x}) = M, \qquad \boldsymbol{x} \in \Omega. \tag{12.12}$$

By continuity (for $u \in C^0(\overline{\Omega})$), we have $u(\boldsymbol{x}) = M$ on $\overline{\Omega}$. Thus, in this second case, the maximum is attained everywhere, in particular, it is attained on the boundary. Therefore, in both cases, we have

$$\boldsymbol{x}_0 \in \partial\Omega,$$

and the result is proved.

Remark 12.3 *This result only holds for Ω a connected set.*

Chapter 13

Reaction-Diffusion Equations (*)

13.1 Outline

We look at solutions of the Fisher–KPP equation – a widely studied type of reaction-diffusion equation.

13.2 Problem setup

We look at the following reaction-diffusion equation:

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2} + rP\left(1 - \frac{P}{P_0}\right), \qquad -\infty < x < \infty, \qquad t > 0. \tag{13.1}$$

where r, D, and P_0 are positive constants. For simplicity, much of this chapter is concerned with the one-dimensional case. This model is called the Fisher–Kolmogorov–Petrovsky–Piskunov equation (or FKPP equation). The model is also an example of a **reaction-diffusion equation**, as there is both a reaction term (i.e. the autocatalytic term $rP[1-(P/P_0)]$), and a diffusion term. These are very common in applications, e.g. Mathematical Biology, Chemistry, Combustion Modelling.

The idea of this model is to take the basic population model of logistic growth,

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{P_0}\right), \qquad t > 0, \qquad P(0) \text{ given}, \tag{13.2}$$

and to add spatial inhomogeneities through the additional diffusion term. This has interesting consequences, as we see throughout the chapter.

13.3 Solutions in the absence of spatial inhomogeneity

For completeness, we first of all review the solutions of Equation (13.2). From a one-dimensional vector field analysis, we see that the fixed points are $P = 0$ and $P = P_0$. Also, $P = 0$ is unstable and $P = P_0$ is stable. We further require that $P(t) \geq 0$, since P is supposed to be a population.

Away from the fixed points, the solution is obtained by separation of variables and integration:

$$\frac{\mathrm{d}P}{P\left(1 - \frac{P}{P_0}\right)} = r\mathrm{d}t,$$

The left-hand side separates further via partial fractions:

$$\frac{\mathrm{d}P}{P} + \frac{\mathrm{d}P}{P - P_0} = r\mathrm{d}t,$$

which integrates to

$$P(t) = \frac{P_0 \mathrm{e}^{rt}}{\frac{P_0 - P(0)}{P(0)} + \mathrm{e}^{rt}}$$

13.4 Front solutions

We henceforth work with the variables $u = P/P_0$, $\tau = rt$, such that Equation (13.1) can be rewritten with fewer parameters:

$$\frac{\partial u}{\partial \tau} = \kappa \frac{\partial^2 u}{\partial x^2} + u\left(1 - u\right), \qquad -\infty < x < \infty, \qquad \tau > 0. \tag{13.3}$$

where $\kappa = D/r$. Equation (13.3) is a nonlinear PDE (specifically, it is quasilinar, as the nonlinearity appears as u^2, and highest-order derivative Du_{xx} appears in a linear fashion. Therefore, we can't for instance find analytical solutions via separation of variables. However, we can still make analytical progress in certain circumstances.

Speficially, we propose a so-called **front solution** (or travelling-wave solution)

$$u(x, \tau) = f(\eta), \qquad \eta = x - v\tau, \tag{13.4}$$

where v is a **positive** constant which we will identify as the front velocity.

Physically, the front solution will be comprehensible by way of analogy with combustion: the front is going to advance with velocity v ('burning' as it goes) from a state where $u = 1$ (already burnt) to $u = 0$ (not yet burnt). Thus, the front solution is like a flame propagation – the flame advances into regions not yet burnt. It is clearly no coincidence that FKPP-type equations play a role in combustion modelling.

With $u(x,\tau)$ given as in Equation (13.3) we have

$$\frac{\partial u}{\partial x} = f'(\eta), \qquad \frac{\partial u}{\partial \tau} = -vf'(\eta),$$

where the second identity follows because of the chain rule. Substituting into Equation (13.3), we have

$$\kappa f'' + vf' + f(1-f) = 0. \tag{13.5}$$

Thus, we have reduced the PDE (complicated) into an ODE (a bit less complicated).

It is instructive to look at Equation (13.5) as a system of two first-order ODEs, with $x = f$ and $y = f'$ (with apologies for the abuse of notation). As such,

$$\frac{d}{d\eta}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ -\frac{vy - x(1-x)}{\kappa} \end{pmatrix}, \tag{13.6}$$

with fixed points $(x,y) = (0,0)$ and $(0,1)$ – i.e. precisely the fixed points of the model with no spatial inhomogeneity, corresponding to the low-fuel and rich-fuel states respectively.

However, we now need to be very careful in our comparison between the homogeneous model and the new inhomgeoenous model with front propagation. We therefore analyse the vector field

$$\boldsymbol{f} = \begin{pmatrix} y \\ -\frac{vy - x(1-x)}{\kappa} \end{pmatrix}$$

associated with the front dynamics (13.5). The vector field is shown in Figure 13.1. We examine the stability of the fixed points, starting with the Jacobian of the vector field $\boldsymbol{f}$:

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -\frac{1}{\kappa}(1-2x) & -\frac{v}{\kappa} \end{pmatrix},$$

We look at the two fixed points separately.

- Case 1. $(x_*, y_*) = (0,0)$. The Jacobian is

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -\frac{1}{\kappa} & -\frac{v}{\kappa} \end{pmatrix}, \qquad (x_*, y_*) = (0,0).$$

  The characteristic equation is

$$\kappa\lambda^2 + \lambda v + 1 = 0,$$

  with eigenvalues

$$\lambda = \frac{-v \pm \sqrt{v^2 - 4\kappa}}{2}.$$

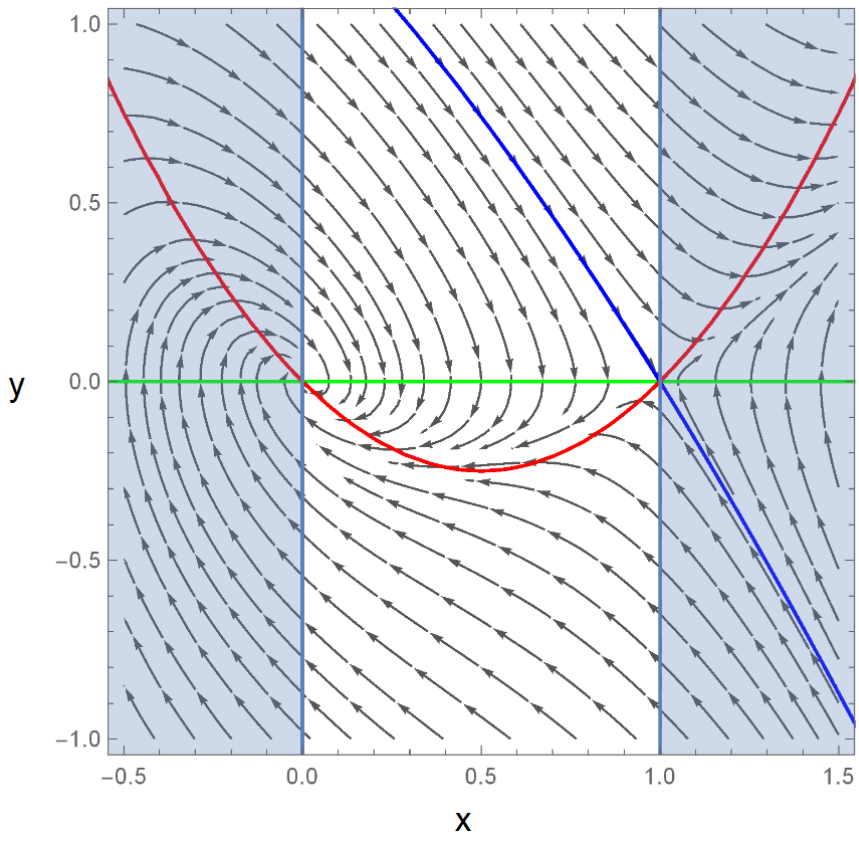


Figure 13.1: Vector field for Equation (13.6). Taken from Reference [url].

Hence,

- The fixed point $(0,0)$ is stable. If $v^2 > 4\kappa$ the fixed point is further classified as a node – both eigenvalues real and negative.
- If $v^2 < 4\kappa$ the fixed point is a stable spiral – both eigenvalues have negative real part and nonzero imaginary part.

- Case 2. $(x_*, y_*) = (1,0)$. The Jacobian is

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ \frac{1}{\kappa} & -\frac{v}{\kappa} \end{pmatrix}, \qquad (x_*, y_*) = (1,0).$$

The characteristic equation is

$$\kappa\lambda^2 + \lambda v - 1 = 0.$$

$$\lambda = \frac{-v \pm \sqrt{v^2 + 4\kappa}}{2}.$$

Hence, the fixed point $(0,0)$ is **unstable**. In fact, it is a **hyperbolic fixed point**, such that

– There is one stable eigenvalue,

$$\lambda = \frac{-v - \sqrt{v^2 + 4\kappa}}{2}.$$

Trajectories moving strictly along the corresponding eigendirection will converge towards the fixed point.

– There is one unstable eigenvalue,

$$\lambda = \frac{-v + \sqrt{v^2 + 4\kappa}}{2}.$$

Trajectories moving along the corresponding eigendirection will diverge the fixed pont.

A general trajectory will have a component in both the stable and unstable eigendirections, and will therefore diverge from the fixed point.

Our aim is no much clearer – we have to solve the $(d/dt)(x, y)^T = \boldsymbol{f}$ such that

- The system starts off at $(1, 0)$ at $\eta = -\infty$
- The system ends at $(1, 0)$ at $\eta = +\infty$

Such a trajectory – starting at one fixed point and finishing at another – is called a **homoclinic orbit**.

We now proceed to solve Equation (13.5) numerically, using the intuition gained from the proceeding fixed-point analysis. We therefore view Equation (13.5) as a boundary-value problem on a symmetric interval $(-L, L)$, with $L \to \infty$. On the left-hand boundary we impose a boundary condition that forces f to be 'close' to $(1, 0)$, and crucially, also forces the solution to approach $(1, 0)$ along a stable eigendirection (stable in backwards time, $\eta \to -\infty$):

$$f(-L) \sim 1 + \epsilon \mathrm{e}^{\lambda_L(-L)}, \qquad f'(-L) \sim \lambda_L \epsilon \mathrm{e}^{\lambda(-L)}, \qquad L \to \infty,$$

where ϵ is a (small) arbitrary constant, and λ_L is chosen to be the **positive** eigenvalue at the $(1, 0)$ fixed point:

$$\lambda_L = \frac{-v + \sqrt{v^2 + 4\kappa}}{2} > 0$$

The unknown constant can be omitted by imposing the following equivalent boundary condition:

$$\frac{f'(-L)}{f(-L) - 1} = \lambda_L \qquad L \to \infty,$$

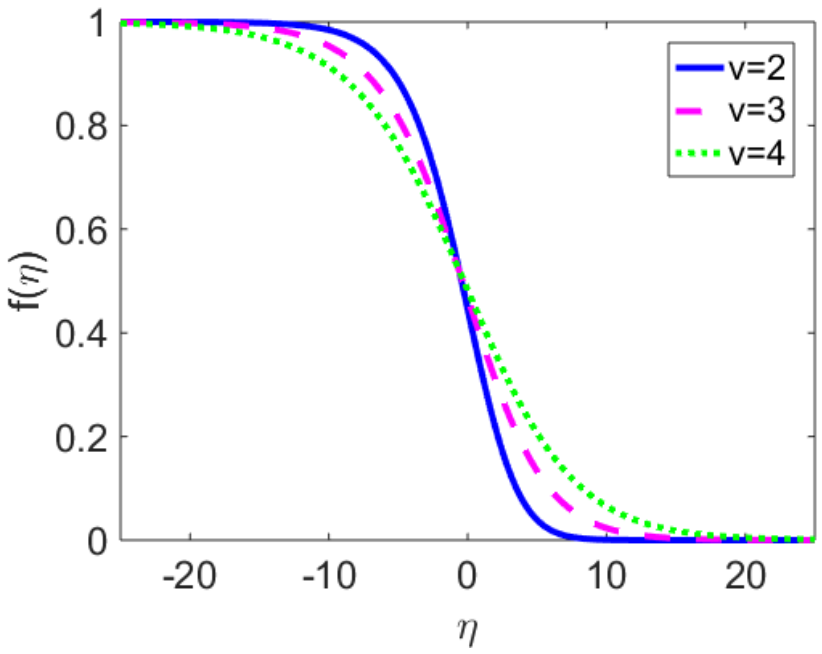


Figure 13.2:

Instead of viewing these conditions as forcing the trajectory on to 'the stable eigendirection in negative time', one can more straightforwardly (but equivalently) view these conditions as enforcing a bounded solution of $|f(\eta)| < \infty$ as $\eta \to -\infty$.

Similarly, on the right-hand boundary we impose a boundary condition that forces f to be 'close' to $(0,0)$, and crucially, also forces the solution to approach $(0,0)$ along a the most stable of the two eigendirection (stable in forwards time, $\eta \to +\infty$):

$$f(L) \sim \epsilon \mathrm{e}^{\lambda_R(L)}, \qquad f'(L) \sim \lambda_R \epsilon \mathrm{e}^{\lambda_R(L)}, \qquad L \to \infty,$$

where ϵ is a (small) arbitrary constant, and λ_R is chosen to be the eigenvalue at the $(0,0)$ fixed point (positive branch!):

$$\lambda_R = \frac{-v + \sqrt{v^2 - 4\kappa}}{2} > 0$$

The unknown constant can be omitted by imposing the following equivalent boundary condition:

$$f(L)\mathrm{e}^{-\lambda_R L} \sim 1 \qquad L \to \infty,$$

Again, these conditions enforce a bounded solution of $|f(\eta)| < \infty$ as $\eta \to +\infty$.

Some numerical results based on this procedure are shown in Figure 13.2 for $\kappa = 1$ and different values of the front velocity, with $v^2 \geq 4\kappa$ chosen deliberately. The matlab code used to solve the boundary-value problem (13.5) is also given in the included listings.

```
function [x,c,dc]=bvp_fisher(v,kappa)

L=10;
maxL=30;

x_init = -L:0.1:L;
solinit = bvpinit(x_init,@guess);

sol = bvp4c(@shockODE,@shockBC,solinit);

for L = L+1:0.05:maxL
    solinit = bvpinit(sol,[-L L]);
    % Extend the solution to Bnew.
    sol = bvp4c(@shockODE,@shockBC,solinit);
    x_coord = sol.x;
    mm = sol.y;
    plot(x_coord,mm(1,:),'-o')
    drawnow
end

x = sol.x;
c = sol.y(1,:);
dc=sol.y(2,:);

% ----------------------------------------------------

function y_init = guess(x_init)

%front initial guess

y_init =    [(1/2)*(1-tanh(x_init))
             -(1/2)*sech(x_init).*sech(x_init)
             ];

end

% ----------------------------------------------------
```

```
38
39 function dydx = shockODE(~,y)
40
41 dydx = [ y(2)
42          (-v*y(2)-y(1)*(1-y(1)))/kappa];
43
44 end
45
46 % ************************************************
47
48 function res = shockBC(ya,yb)
49
50 lambdaL= (-v+sqrt(v^2+4*kappa))/2;
51 lambdaR= (-v+sqrt(v^2-4*kappa))/2;
52
53 res=      [ (ya(2)/(ya(1)-1))-lambdaL
54             yb(1)*exp(-lambdaR*L)-1];
55
56 end
57
58 end
```

13.5 Stability of fronts

We further look at the issue of stability of the front solutions, i.e. what happens if small perturbations to the fronts are introduced. As such, we imagine a fully time-dependent solution of the FKPP equation, but written in the frame of reference of the moving front, hence in (τ, η) variables:

$$\frac{\partial u}{\partial \tau} - v\frac{\partial u}{\partial \eta} = \kappa\frac{\partial^2 u}{\partial \eta^2} + u\left(1-u\right), \tau > 0. \tag{13.7}$$

With $\partial u/\partial \tau = 0$ we recover the front solution. The idea now is that the front solution is an equilibrium solution of Equation (13.9) – and we can look at small disturbances away from the equilibrium. As such, let

$$u(\tau, \eta) = f(\eta) + \epsilon u_1(\tau, \eta),$$

where $f(\eta)$ is the equilibrium (front) solution, and u_1 is a perturbation. We substitute this trial solution into Equation (13.9) and linearize, assuming that ϵ is infinitesimally small. We obtain

$$\frac{\partial u_1}{\partial \tau} = \kappa \frac{\partial^2 u_1}{\partial \eta^2} + v \frac{\partial u_1}{\partial \eta} + [1 - 2f(\eta)]\, u_1. \tag{13.8}$$

This is of the form

$$\frac{\partial u_1}{\partial \tau} = \mathcal{L}(\eta) u_1, \qquad \mathcal{L}(\eta) = \kappa \partial_{\eta\eta} + v \partial_\eta + [1 - 2f_0(\eta)] \tag{13.9}$$

where $\mathcal{L}(\eta)$ is a linear operator. We propose a trial solution

$$u_1 = \mathrm{e}^{\sigma\tau} \widehat{u_1}(\eta), \tag{13.10}$$

hence

$$\sigma \widehat{u_1}(\eta) = \mathcal{L}(\eta) \widehat{u_1}(\eta), \tag{13.11}$$

which is yet another eigenvalue problem for the eigenvalue σ – although this time, the operator $\mathcal{L}(\eta)$ is a differential operator that depends continuously on the variable η, and not a matrix.

We make a further (inspired) trial solution

$$\widehat{u_1}(\eta) = \phi(\eta) \mathrm{e}^{-v\eta/2\kappa}, \tag{13.12}$$

such that Equation (13.11) becomes

$$\sigma\phi = \left\{ \kappa \partial_{\eta\eta} - \frac{v^2}{2D} + [1 - 2f(\eta)] \right\} \phi, \tag{13.13}$$

i.e. the derivative $v\partial/\partial\eta$ has been knocked out by the transformation (13.10).

We focus on compact perturbations, such that

$$\phi(\eta) = 0, \qquad |\eta| \geq R,$$

for some finite radius R. Hence, by multiplying Equation (13.13) by ϕ and integrating from $-R$ to R, we obtain

$$\sigma \|\phi\|_2^2 = -\kappa \|\partial_\eta \phi\|_2^2 - 2 \int_{-R}^{R} f(\eta) |\phi|^2 \mathrm{d}\eta + \left(1 - \frac{v^2}{4\kappa}\right) \|\phi\|_2^2,$$

where

$$\|\phi\|_2^2 = \int_{-R}^{R} |\phi|^2 \mathrm{d}\eta,$$

and

$$\int_{-R}^{R} f(\eta)|\phi|^2 \mathrm{d}\eta \geq 0,$$

since the front solution $f(\eta)$ is non-negative. Hence,

$$\begin{aligned} \sigma &= \left(1 - \frac{v^2}{2\kappa}\right) - \left(\frac{\kappa\|\partial_\eta\phi\|_2^2 + 2\int_{-R}^{R} f(\eta)|\phi|^2 \mathrm{d}\eta}{\|\phi\|_2^2}\right), \\ &= \left(1 - \frac{v^2}{4\kappa}\right) - A^2, \end{aligned}$$

where A is a real number. Therefore, in order to make the eigenvalue negative, a **sufficient condition** is

$$v^2 > 4\kappa.$$

Remark 13.1 *The condition $v^2 > 4\kappa$ is a sufficient condition for stability. In principle, the eigenfunction ϕ could assume a certain profile such that $\left(1 - \frac{v^2}{4\kappa}\right)$ is positive whereas the whole sum $[1 - (v^2/4\kappa)] - A^2$ is negative. Thus, it is not clear that $v^2 > 4\kappa$ is a* **necessary condition**. *However, using more advanced techniques [url], it can be shown that the condition $v^2 > 4\kappa$ is both necessary and sufficient for stability of the fronts.*

13.6 Numerical solutions

We look in more depth at numerical solutions of the Fisher-KPP equation, with a view to characterizing the dynamics in several dimensions. Specifically, we look at solutions to We look at the following reaction-diffusion equation:

$$\frac{\partial u}{\partial \tau} = \kappa\nabla^2 u + u(1-u), \qquad \boldsymbol{x} \in \Omega, \qquad t > 0, \tag{13.14}$$

where Ω is taken to be the doubly periodic domain $[0, L]^2$, hence $\boldsymbol{x} = (x, y) \in [0, L]^2$, with

$$u(x + L, y, t) = u(x, y, t), \qquad u(x, y + L, t) = u(x, y, t),$$

for all $t > 0$.

Because of the periodic boundary conditions, we can expand the solution $u(\boldsymbol{x}, t)$ in a Fourier series,

$$u(\boldsymbol{x}, t) = \sum_{\boldsymbol{k}} \widehat{u}_{\boldsymbol{k}}(t) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\mathrm{d}\boldsymbol{x}},$$

where

$$\boldsymbol{k} = (2\pi/L)\boldsymbol{n}, \qquad \boldsymbol{n} \in \mathbb{Z}^2,$$

and

$$\widehat{u}_{\boldsymbol{k}}(t) = \frac{1}{L^2}\iint \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} u(\boldsymbol{x},t)\mathrm{d}^2x.$$

We further multiply Equation (13.14) by $\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}$ and integrate over $[0,L]^2$. We obtain

$$\frac{\mathrm{d}\widehat{u}_{\boldsymbol{k}}}{=} - \kappa k^2 \widehat{u}_{\boldsymbol{k}} + \widehat{u}_{\boldsymbol{k}} - \iint \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} u^2(\boldsymbol{x},t)\mathrm{d}^2x, \tag{13.15}$$

or

$$\frac{\mathrm{d}\widehat{u}_{\boldsymbol{k}}}{=} - \kappa k^2 \widehat{u}_{\boldsymbol{k}} + \widehat{u}_{\boldsymbol{k}} - \widehat{Q}_{\boldsymbol{k}}, \tag{13.16}$$

where

$$Q = u^2, \qquad \widehat{Q}_{\boldsymbol{k}} = \iint \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} u^2(\boldsymbol{x},t)\mathrm{d}^2x.$$

Assuming perfect knowledge of all the Fourier coefficients of u and Q, Equation (13.16) can be discretized in time by defininining a solution at discrete time points

$$\widehat{u}^n_{\boldsymbol{k}} = \widehat{u}_{\boldsymbol{k}}(t = j\Delta t), \qquad j \in \{0,1,2,\cdots\}.$$

For instance, Equation (13.16) can be discretized using a backward-Euler scheme for maximum numerical stability:

$$\frac{\widehat{u}^{n+1}_{\boldsymbol{k}} - \widehat{u}^n_{\boldsymbol{k}}}{\Delta t} = -\kappa k^2 \widehat{u}^{n+1}_{\boldsymbol{k}} + \widehat{u}^n_{\boldsymbol{k}} - \widehat{Q}^n_{\boldsymbol{k}},$$

hence

$$\widehat{u}^{n+1}_{\boldsymbol{k}} \left(1 + \kappa k^2 \Delta t\right) = (1+\Delta t)\widehat{u}^n_{\boldsymbol{k}} - \widehat{Q}^n_{\boldsymbol{k}}\Delta t,$$

hence

$$\widehat{u}^{n+1}_{\boldsymbol{k}} = \frac{(1+\Delta t)\widehat{u}^n_{\boldsymbol{k}} - \widehat{Q}^n_{\boldsymbol{k}}\Delta t}{1 + \Delta t \kappa k^2}.$$

We now solve an approximation of Equation (13.15) numerically. We truncate the Fourier expansions such that

$$\boldsymbol{k} = (2\pi/L)(n,m), \qquad |n|, |m| < N/2$$

As such, we replace the Fourier transform of $u(\boldsymbol{x},t)$ with the discrete (fast) Fourier transform analogue, to produce the following algorithm:

1. Start with initial data $u(\boldsymbol{x}, t=0)$, and $Q = u^2(\boldsymbol{x}, t=0)$ perform a discrete Fourier transform to obtain $\widehat{u}^0_{\boldsymbol{k}}$ and $\widehat{Q}^0_{\boldsymbol{k}}$

2. Obtain $\widehat{u}^1_{\boldsymbol{k}}$ from

$$\widehat{u}^1_{\boldsymbol{k}} = \frac{(1+\Delta t)\widehat{u}^0_{\boldsymbol{k}} - \widehat{Q}^0_{\boldsymbol{k}}\Delta t}{1 + \Delta t \kappa k^2}.$$

3. Perform the inverse Fourier transform to obtain $u(\boldsymbol{x}, t=\Delta t)$ and hence, $Q = u^2(\boldsymbol{x}, t=\Delta t)$.

4. Repeat Steps 1-3.

This is an efficient algorithm, as the differentiation ∇^2 is carried out in Fourier space, where it manifests itself as multiplication ($\nabla^2 \to -k^2$). Equally, the convolution

$$\widehat{Q}_{\boldsymbol{k}} = \iint \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} u^2(\boldsymbol{x},t)\mathrm{d}^2x = \sum_{\boldsymbol{k}'} \widehat{u}_{\boldsymbol{k}'}\widehat{u}_{\boldsymbol{k}-\boldsymbol{k}'}$$

is carried out in real space, where it manifests iteself just as ordinary multiplication, i.e. $Q = u^2$. This algorithm therefore has the best of both worlds. The algorithm is called **pseudospectral** – it is not a fully spectral algorithm, as the numerical solution is not computed entirely in terms of the Fourieramplitudes $\widehat{u}_{\boldsymbol{k}}$. Instead, at each timestep, we transform back into treal space, where Q is computed highlighy efficiently. One then reverts to spectral (Fourier) space for the next timestep.

13.7 Validation of numerical solutions

It is helpful to be able to have some known analytical solution against which an implementation of the numerical method in Section 13.6 can be compared. Unfortunately, the Fisher-KPP equation is nonlinear, and analytical solutions are difficult to come by. Certainly, we could use the quasi-analytical front solutions as a test bed, although another (equally valid) proposal is suggested here for simulations in higher spatial dimensions. As such, we consider a modified Fisher–KPP equation with a source term: We look at the following reaction-diffusion equation:

$$\frac{\partial u}{\partial \tau} = \kappa\nabla^2 u + u(1-u) + s(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \qquad t > 0, \tag{13.17}$$

where Ω is again for definiteness taken to be the doubly periodic domain $[0, L]^2$, and $s(\boldsymbol{x})$ is a source term. We assume that the system evolves so that an equilibrium is reached after a large but finite time, corresponding to the following special case of Equation (13.17):

$$0 = \kappa\nabla^2 u_0 + u_0(1-u_0) + s(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \qquad t > 0, \tag{13.18}$$

We assume that $s(\boldsymbol{x}) = \epsilon s_0(\boldsymbol{x})$, where ϵ is a small positive parameter and $s_0(\boldsymbol{x})$ is a function with unit magnitude, in the sense that $\|s_0\|^2_{2,\Omega} = 1$. As such, assume that $u_0(\boldsymbol{x})$ has relaxed (up to small fluctuations) to the stable fixed point $u_0 = 1$, and we propose the following power-series solution for $u_0(\boldsymbol{x})$:

$$u_0(\boldsymbol{x}) = 1 + \epsilon f_1(\boldsymbol{x}) + \epsilon^2 f_2(\boldsymbol{x}) + \cdots .$$

ϵ	N	Res	$\|f_1\|_2^2$	N	Res	$\|f_1\|_2^2$
0.001	128	4.13×10^{-8}	4.13×10^{-8}	256	4.13×10^{-8}	4.13×10^{-8}
0.005	128	1.03×10^{-6}	4.13×10^{-8}	256	1.03×10^{-6}	4.13×10^{-8}
0.01	128	4.10×10^{-6}	4.13×10^{-8}	256	4.10×10^{-6}	4.13×10^{-8}
0.05	128	1.00×10^{-4}	4.13×10^{-8}	256	1.00×10^{-4}	4.13×10^{-8}
0.1	128	3.89×10^{-4}	4.13×10^{-8}	256	4.89×10^{-4}	4.13×10^{-8}

Table 13.1: Numerical benchmark case using the Fisher-KPP equation with Gaussian source $s(\boldsymbol{x}) = \epsilon \mathrm{e}^{-r^2\sigma^2}$, where $r^2 = [x - (L/2)]^2 + [y - (L/2)]^2$ and $\sigma^2 \approx 22.36$. Physical parameters: $\kappa = 0.01$ and $L = 1$. The system settles down to a steady state by $t \approx 10$ whereas the norms are computed long after, at $t = 100$. Simulation results for grids for $N = 128, 256$, where N^2 is the number of Fourier modes in the pseudospectral method. The chosen timestep is $dt = 10^{-3}$. Here, Res $= \|u(\boldsymbol{x}, t = 100) - 1\|_2^2$. Finally, the initial conditions are chosen such that $C(\boldsymbol{x}, t = 0) = \rho$, where ρ is a random number between 0 and 0.01 – a different (and independent) random number at each point $\boldsymbol{x} \in [0, L]^2$.

We substitute this trial solution into Equation (13.18) and equate coefficients of like powers of ϵ. At lowest order (i.e. $O(\epsilon)$) we obtain

$$\kappa \nabla^2 f_1 - f_1 = s_0(\boldsymbol{x}).$$

We assume a Fourier-series representation for $s_0(\boldsymbol{x})$, with

$$s_0(\boldsymbol{x}) = \sum \boldsymbol{k} \widehat{s}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}.$$

The corresponding solution for $f_1(\boldsymbol{x})$ is therefore

$$f_1(\boldsymbol{x}) = -\sum_{\boldsymbol{k}} \frac{\widehat{s}_{\boldsymbol{k}}}{\kappa k^2 + 1} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}$$

hence

$$\|f_1\|_2^2 = \sum_{\boldsymbol{k}} \frac{|\widehat{s}_{\boldsymbol{k}}|^2}{(\kappa k^2 + 1)^2}.$$

Hence,

$$\|u_0(\boldsymbol{x}) - 1\|_2^2 \epsilon^2 = \epsilon^2 \|f_1\|_2^2 + O(\epsilon^3),$$

and this is a quantity that can be measured in a numerical computation.

The results of this procedure are tabulated in Table 13.1 for a Gaussian source term and the corresponding plot is shown in Figure 13.4. There is excellent agreement between the theory and the numerics (both at low resolution $N = 128$ and high resolution $N = 256$, confirming not only the correctness of the numerical code but also, grid convergence). Finally, some snapshots of the evolution of the spatial structure of $u(\boldsymbol{x}, t)$ are shown in Figure **??**.

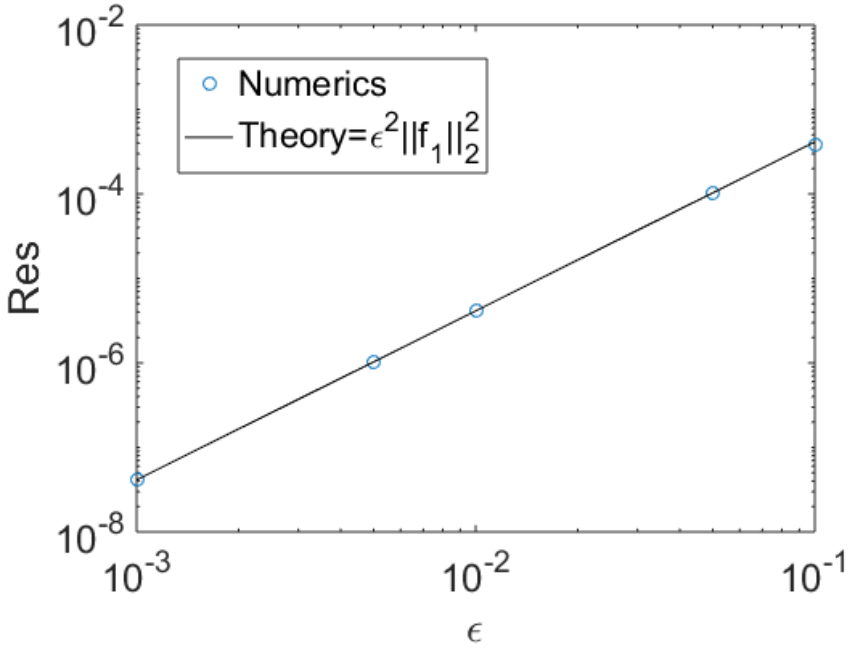


Figure 13.3: Graph corresponding to Table 13.1

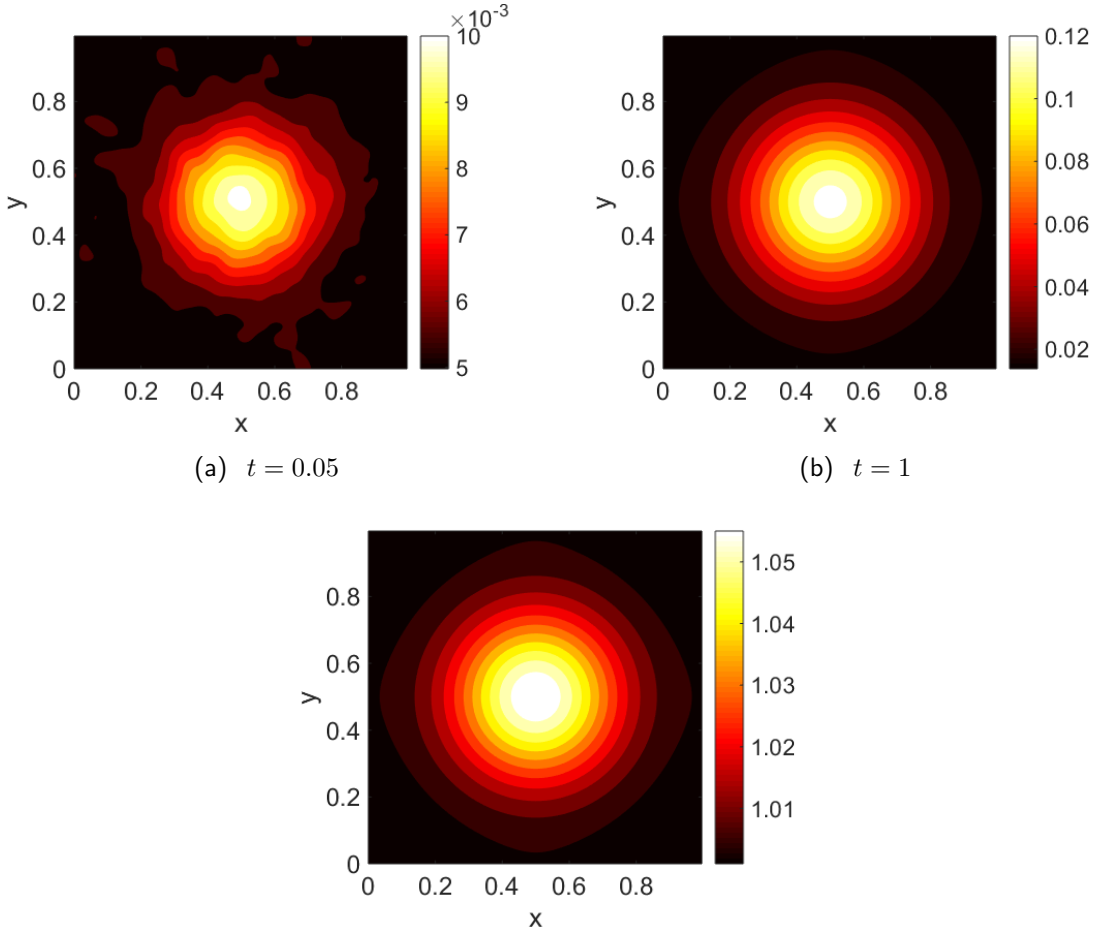


Figure 13.4: Snapshots of the spatial structure of $u(\boldsymbol{x}, t)$ at various times. Simulation data the same as in Table 13.1.

Bibliography

[GL96] Ronald B Guenther and John W Lee. *Partial differential equations of mathematical physics and integral equations*. Courier Corporation, 1996.

[PS08] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.

[Str01] S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Perseus Books Group, New York, 2001.

[url] Traveling wave solutions of reaction-diffusion equations in population dynamics. `https://www.math.leidenuniv.nl/scripties/jonkhout.pdf`. Accessed: 08/12/2017.