Introduction

This course treats two essential subjects, among many others, in applied mathematics: numerical analysis and optimization. Before even presenting these two disciplines, let us immediately say that through their teaching the objective of this course is to introduce the reader to the world of mathematical modelling and numerical simulation which have gained considerable importance in these last decades in all areas of science and industrial applications (or engineering science). Mathematical modelling is the art (or the science, depending on the point of view) of representing (or transforming) a physical reality into abstract models which are accessible to analysis and to calculation. Numerical simulation is, of course, the process which allows us to calculate the solutions of these models on a computer, and thus to simulate physical reality.

But, first of all, what is applied mathematics? To say that it is mathematics turned towards applications would be a tautology and a false characterization. In effect, throughout time, mathematicians have been inspired by the practical problems that they have tried to solve, however, the emergence of applied mathematics as an independent discipline is relatively recent. In fact, everything changed with the appearance of the first computers shortly after the Second World War. More than for any other discipline the computer has been a revolution for mathematics: in effect it has opened up a new field, that of modelling and simulation. The computer has made mathematics an experimental science (we make ‘numerical experiments’ as others make physical experiments), and the design, as well as the analysis of methods of calculation on a computer, has become a new branch of mathematics: this is numerical simulation. This progress also made it possible for mathematics to attack much more complex and concrete problems, resulting from immediate industrial or scientific motivations, to which we can bring both qualitative and quantitative responses: this is mathematical modelling.

We can thus characterize applied mathematics as the mathematics of modelling and numerical simulation. From this point of view, applied mathematics lies at the intersection of many scientific disciplines: mathematics, computing, physical sciences, chemistry, mechanics, biology, economics, and engineering sciences (under this last term we usually group the different fields of industrial applications such as aeronautics, power generation, finance, etc.). The American mathematician Joseph Keller said as a joke that ‘pure mathematics is a branch of applied mathematics’. He wanted to highlight the multidisciplinary character of applied mathematics (but it is not excluded that he also wanted to pay back some ‘pure’ mathematicians who affect to despise applied mathematics).

Paraphrasing the title of a famous film, my colleague Pierre-Louis Lions claims that applied mathematics is characterized by three things: Sex, Lies, and Videotapes. Videocassettes are of course the symbols of digital simulation (and of the films that they produce), lies
correspond to models (not always faithful to reality), and the sex is obviously mathematical
analysis (inextinguishable engine of human passions and source of much pleasure).

After this (long) detour we can now return to the title of this course. Numerical analysis
is thus the discipline which conceives and analyses the methods or algorithms of numerical
calculation. In addition optimization is the theory of methods which allow us to improve
the operation, output, or the response of a system by maximizing or minimizing associated
functions. It is thus an essential tool for modelling.

The objectives of this course are to familiarize the reader with the principal models
(which are often partial differential equations), their methods of numerical solution and
their optimization. Of course, the ambition of this course is to give a foundation which
will allow future engineers either in a design department or in research and development
to create new models and new numerical algorithms for more complicated problems
not discussed here. However, even those not destined for such a career are interested in
understanding the fundamentals of numerical simulation. Indeed, many industrial or political
decisions will be taken from now on having faith in calculations or numerical simulations.
It is thus essential that the decision-makers are capable of judging the quality and of the
reliability of the calculations which are presented to them. This course will allow them
to understand the first criteria which guarantee the validity and the relevance of numerical
simulations.

The plan of this course is the following. After a first chapter of introduction to the prin-
cipal ‘classical’ models and to their numerical solution, Chapter 2 is dedicated to the study
of the numerical method of finite differences. These first two chapters allow us to go very
quickly to some essential numerical questions which motivate the theoretical developments
that follow. Chapters 3, 4, and 5 are dedicated to the theoretical solution by the variational
approach of stationary (independent of time) models. They also give the foundation of a
very important numerical method, called the finite element method, which is presented
in detail in Chapter 6. The finite element method is the basis of many pieces of industrial
or academic software. Chapters 7 and 8 discuss the solution of nonstationary problems
(or of evolution in time), from both the theoretical and numerical points of view. If the
first eight chapters are dedicated to numerical analysis, the last three treat optimization.
Chapter 9 presents a series of concrete examples of optimization problems and gives a theory
of existence of solutions to these problems. Chapter 10 derives the (necessary or sufficient)
conditions for optimality of the solutions. These conditions are important as much from the
theoretical as the numerical point of view. They allow us to characterize the optima, and
they are the foundation of the numerical algorithms that we describe. Finally, chapter 11
is an introduction to operational research. After having studied linear programming, we
give an outline of combinatorial optimization methods (that is to say optimization in dis-
crete variables) which are essential for the optimal planning of resources and tasks in all large
companies. Each chapter starts with an introduction which gives the plan and the principal
ideas.

The length of this course should not worry the reader: the course contains numerous
supplementary developments which allow the curious reader ‘to go a little further’ and to
make the link with other works or other disciplines. It is, therefore, more a work of reference
than the exact transcription of a lecture course.

To finish this introduction we give some practical information. As far as possible, this
course is intended to be ‘self-contained’ to avoid frequent references to other works. This
is particularly sensible for many results from analysis which here are only useful, but not
essential, technical tools. Statement without proof would amount to using them as ‘black
boxes’ which gives them the flavour of an artificial ‘recipe’. As far as possible, we have
therefore included their proof, but more as information and to ‘demystify’ than for the
theoretical interest of the mathematical arguments. In order to distinguish them we use, for all
these difficult passages or those of complementary interest, smaller characters like these. The reader
should therefore consider these passages in small characters as ‘outside of the programme’.
The statements of results or of definitions are in italic characters like these. The exercises
are in sans serif characters like these. The end of a proof is indicated by the character □, while the end of a remark or of an example is indicated by the character ◦. An index is available
at the end of the work.

The answers to exercises will be published in French. Most of the computer programs
which implement the numerical methods studied, and which have allowed us to produce the
figures in this work, are available on the website

http://www.cmap.polytechnique.fr/~allaire/course_X_annee2.html

where the reader can download them freely. The finite difference schemes, as well as the finite
element method in one dimension, have been programmed in the language Scilab developed
by INRIA and ENPC, available free on the website

http://www.scilab.org

while the results of the finite element method in two dimensions have been obtained with the
help of the program FreeFem++ developed by F. Hecht and O. Pironneau and also available
free on the website

http://www.freefem.org

In addition, most of the two-dimensional figures and all of the three-dimensional figures have
been drawn with the help of graphical program xd3d developed by François Jouve at the
École Polytechnique and also available free on the website

http://www.cmap.polytechnique.fr/~jouve/xd3d

Let us indicate another web address for the curious reader who wants to know more about
the history of mathematics or the life of some mathematicians cited in this course

http://www-history.mcs.st-and.ac.uk/history

The reader who would like to keep up to date with the progress and advances of applied
mathematics would benefit to consult the site of the Société de Mathématiques Appliquées
et Industrielles

http://smaie.math.fr

or that of its American colleague, the Society for Industrial and Applied Mathematics

http://www.siam.org

The level of this course is introductory and it does not need any other prerequisites
other than the level of knowledge gained in the first few years of university. We recognize
that it is difficult to show much originality in this subject which is already classical in
the literature. In particular, our course owes much to its predecessors and particularly
to the course by B. Larrouturou, P.-L. Lions, and P.-A. Raviart from which it sometimes
borrows heavily. The author thanks all those who have reread parts of the manuscript,
particularly Frédéric Bonnans, Bruno Després and Bertrand Maury. A special mention is
due to Stéphane Gaubert, who has co-written chapter 11, and also to Olivier Pantz, who has
reread the entire manuscript with great care and who has checked the exercises and written
the corrections. The author thanks in advance all those who will point out the inevitable errors or imperfections of this edition, for example, by email to the address gregoire.allaire@polytechnique.fr

G. Allaire
Paris, July 7, 2005
1 Introduction to mathematical modelling and numerical simulation

1.1 General introduction

This chapter is an introduction to two distinct, but closely linked, aspects of applied mathematics: mathematical modelling and numerical simulation. A mathematical model is a representation or an abstract interpretation of physical reality that is amenable to analysis and calculation. Numerical simulation allows us to calculate the solutions of these models on a computer, and therefore to simulate physical reality. In this book, the models we shall study will be partial differential equations (or PDEs), that is, differential equations in several variables (time and space, for example).

For the moment we shall put aside a third fundamental aspect of applied mathematics, that is, the mathematical analysis of models to which we shall return in a little more depth in later chapters. We need, in some way, to both motivate and justify this necessary intrusion of mathematical analysis. We shall see that the numerical calculation of the solutions of these physical models sometimes has some unpleasant surprises which can only be explained by a sound understanding of their mathematical properties. Once again we recall the fundamental multidisciplinary character of applied mathematics, and therefore of numerical simulation, which combines mathematics, computer science, and engineering.

Although most of the problems and applications which motivate applied mathematics are fundamentally nonlinear (see, for example, [12], [27]), we confine ourselves
in this work to linear problems for simplicity. Likewise, we only consider deterministic
problems, that is, with no random or stochastic components. Finally, in order for this
chapter to be introductory and easily accessible, we shall often be a little imprecise
in our mathematical arguments. The more rigorous reader can be reassured that we
shall return to the concepts introduced in this way more carefully in the following
chapter.

The plan of this chapter is the following. Section 1.2 is devoted to an elementary
example of modelling which leads to the heat flow equation. Section 1.3 is a quick
review of the principal PDEs that we meet in the usual models in mechanics, physics,
or engineering sciences. Section 1.4 is an informal introduction to numerical analysis
and the finite difference method. Finally, in the Section 1.5 we give the definition
of a well-posed problem as well as a (brief) classification of PDEs.

1.2 An example of modelling

Modelling represents a considerable part of the work of an applied mathematician
and requires a thorough knowledge, not only of applied mathematics, but also of the
scientific discipline to which it is applied. In fact, in many cases the mathematical
model may not yet be established, or we must select the pertinent one from among
several possibilities, or we must simplify known models which are too complex. How-
ever, in an introductory presentation of the discipline it is not possible to do justice
to this step of the modelling process: we must begin by learning the basic properties
of applied mathematics! This is why we limit ourselves to describing the derivation
of a well-known classical physical model, and we refer the reader who wishes to know
more to more specialised works.

The model which we shall describe is known as the heat flow equation, or the
diffusion equation.

Let us consider a domain $\Omega$ in $N$ space dimensions (denoted by $\mathbb{R}^N$, with in general
$N = 1, 2, \text{ or } 3$) which we assume is occupied by a homogeneous, isotropic material
which conducts heat. We denote the space variable by $x$, that is a point of $\Omega$, and the
time variable by $t$. The heat sources in $\Omega$ (possibly nonuniform in time and space) are
represented by a given function $f(x, t)$, while the temperature is an unknown function
$\theta(x, t)$. The quantity of the heat is proportional to the temperature $\theta$ and is therefore
$c\theta$ where $c$ is a physical constant (which depends on the material) called the specific
heat. To calculate the temperature $\theta$, we write down the law of conservation of
energy or of heat. In an elementary volume $V$ contained in $\Omega$, the variation in time
of the amount of heat is the balance of that produced by the sources and that which
leaves or returns through the element boundaries. In other words,

$$
\frac{d}{dt} \left( \int_V c\theta \, dx \right) = \int_V f \, dx - \int_{\partial V} q \cdot n \, ds,
$$

(1.1)

where $\partial V$ is the boundary of $V$ (with surface element $ds$), $n$ is the outward unit
normal from $V$, and $q$ is the heat flux vector. If we apply Gauss’s theorem we obtain

$$ \int_{\partial V} q \cdot n \, ds = \int_V \text{div} q \, dx. $$

Gathering together the different terms in (1.1) and using the fact that the elementary volume $V$ is independent of time, we deduce the energy conservation equation

$$ \frac{c}{\varepsilon} \frac{\partial \theta}{\partial t} + \text{div} q = f $$

(1.2)

which holds at every point $x \in \Omega$ and for all time $t$. We recall that the divergence operator is defined by

$$ \text{div} q = \sum_{i=1}^{N} \frac{\partial q_i}{\partial x_i} \text{ with } q = (q_1, \ldots, q_N)^T. $$

We must now link the heat flow to the temperature, by what is called a constitutive law. In this case, we use Fourier’s law which says that the heat flux is proportional to the temperature gradient

$$ q = -k \nabla \theta T $$

(1.3)

where $k$ is a positive constant (which depends on the material) called the thermal conductivity. Remember that the gradient operator is defined by

$$ \nabla \theta = \left( \frac{\partial \theta}{\partial x_1}, \ldots, \frac{\partial \theta}{\partial x_N} \right)^T. $$

By combining the conservation law (1.2) and the constitutive law (1.3), we obtain an equation for the temperature $\theta$

$$ \frac{c}{\varepsilon} \frac{\partial \theta}{\partial t} - k \Delta \theta = f, $$

where $\Delta = \text{div} \nabla$ is the Laplacian operator given by

$$ \Delta \theta = \sum_{i=1}^{N} \frac{\partial^2 \theta}{\partial x_i^2}. $$

This equation is valid in the entire domain $\Omega$ and we must add another relation, called a boundary condition, which describes what happens at the boundary $\partial \Omega$ of the domain, and another relation which describes the initial state of the temperature. By convention, we choose the instant $t = 0$ to be the initial time, and we impose an initial condition

$$ \theta(t = 0, x) = \theta_0(x), $$

(1.4)

where $\theta_0$ is the function giving the initial distribution of the temperature in the domain $\Omega$. The type of boundary condition depends on the physical context. If the domain is
surrounded by a region of constant temperature, then, by rescaling the temperature, the temperature satisfies the Dirichlet boundary condition

\[ \theta(t, x) = 0 \quad \text{for all } x \in \partial \Omega \text{ and } t > 0. \quad (1.5) \]

If the domain is assumed to be adiabatic or thermally isolated from the exterior, then the heat flux across the boundary is zero and the temperature satisfies the Neumann boundary condition

\[ \frac{\partial \theta}{\partial n}(t, x) \equiv n(x) \cdot \nabla \theta(t, x) = 0 \quad \text{for all } x \in \partial \Omega \text{ and } t > 0, \quad (1.6) \]

where \( n \) is the unit outward normal to \( \Omega \) (see Figure 1.1). An intermediate situation can happen: the heat flux across the boundary is proportional to the jump in temperature from the exterior to the interior, and the temperature satisfies the Fourier (or Robin) boundary condition

\[ \frac{\partial \theta}{\partial n}(t, x) + \alpha \theta(t, x) = 0 \quad \text{for all } x \in \partial \Omega, \text{ and } t > 0 \quad (1.7) \]

where \( \alpha \) is a positive constant. As we must choose a boundary condition (as one of the steps in the modelling), we shall take the Dirichlet boundary condition (1.5).

Finally, gathering together the equation, the initial value, and the boundary condition satisfied by the temperature, we obtain the heat equation

\[
\begin{cases}
\frac{\partial \theta}{\partial t} - k \Delta \theta = f & \text{for } (x, t) \in \Omega \times \mathbb{R}_t^+ \\
\theta(t, x) = 0 & \text{for } (x, t) \in \partial \Omega \times \mathbb{R}_t^+ \\
\theta(t = 0, x) = \theta_0(x) & \text{for } x \in \Omega
\end{cases}
\quad (1.8)
\]

Problem (1.8) therefore comprises a PDE equipped with boundary conditions and an initial value. Because of the boundary conditions, we say that (1.8) is a **boundary value problem**, but we also say that it is a **Cauchy problem** because of the initial value.
Remark 1.2.1 In this model of heat propagation we must make the physical units precise: the temperature $\theta$ is expressed in degrees Kelvin ($K$), the specific heat $c$ in Joules per kilogram per degree Kelvin ($J/(kg \times K)$), the thermal conductivity (per unit of mass) $k$ in Joules metre squared per kilogramme per degree Kelvin per second ($Jm^2/(kg \times K \times s)$). From a mathematical point of view, we shall frequently neglect these units, and also assume that the constants $c$ and $k$ are equal to 1 (this is equivalent to making the physical quantities nondimensional).

Remark 1.2.2 We have mentioned three types of boundary condition, Dirichlet, Neumann, and Fourier (but there are others) which hold on the entire boundary $\partial \Omega$. Of course, we can easily imagine situations where the boundary conditions are mixed: Dirichlet on $\partial \Omega_D$, Neumann on $\partial \Omega_N$, and Fourier on $\partial \Omega_F$, with $\partial \Omega_D, \partial \Omega_N, \partial \Omega_F$ being a partition of the boundary $\partial \Omega$.

Remark 1.2.3 The heat flow equation (1.8) is linear in the sense that its solution $\theta$ depends linearly on the data $(f, \theta_0)$. In physics, this property is often described in terms of a superposition principle: a linear combination of data $(f, \theta_0)$ leads to a solution $\theta$ which is the same linear combination of solutions corresponding to each term of the decomposition of data. From a physical point of view, linearity is only one hypothesis among many. Indeed, for problems with a strong variation in temperature, Fourier’s law is false, and it should be corrected by assuming that the thermal conductivity $k$ depends on the temperature $\theta$ and its gradient $\nabla \theta$ (which makes the problem nonlinear). Even worse, for very rapid phenomena (explosions, for example) it is necessary to abandon the assumption of the proportionality of the heat flux $q$ to the temperature gradient $\nabla \theta$. Indeed, this hypothesis (which initially appears ‘natural’) leads to the following paradox: the heat propagates with infinite velocity in the domain $\Omega$. We shall see later (see Remark 1.2.9) how to reach this paradox. Let us remember for the moment that modelling is making hypotheses and describing their domain of validity.

Remark 1.2.4 Problem (1.8) is not just a model of heat propagation. In fact it has a universal character, and we find it in many unrelated phenomena (we simply change the names of the variables). For example, (1.8) is also known as the diffusion equation, and models the diffusion or migration of a density or concentration across the domain $\Omega$ (imagine a pollutant diffusing in the atmosphere, or a chemical species migrating in a substrate). In this case, $\theta$ is the concentration or the density in question, $q$ is the mass flux, $k$ is the diffusivity, and $c$ is the volume density of the species. Likewise, the conservation law (1.2) is a mass balance, while the constitutive law (1.3) is called Fick’s law.

Remark 1.2.5 Problem (1.8) also occurs in finance where it is called the Black–Scholes model. A variant of (1.8) allows us to find the value of an option to buy (or call option) a stock, which is initially worth $x$, for price $k$ at some time in the future $T$. 

AN EXAMPLE OF MODELLING

5
This value is the solution $u$ of

$$
\begin{cases}
\frac{\partial u}{\partial t} - ru + \frac{1}{2}rx^2 \frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 u}{\partial x^2} = 0 & \text{for } (x,t) \in \mathbb{R} \times (0,T) \\
u(t = T, x) = \max(x - k, 0) & \text{for } x \in \mathbb{R}
\end{cases}
$$

(1.9)

More precisely, $u(0, x)$ is the value at time $t = 0$ of the call option with exercise price $k$ at the exercise time $T > 0$, and with value $x$ at $t = 0$. The volatility is denoted by $\sigma$ and the interest rate by $r$. We remark that (1.9) is a final value and not an initial value problem, but that the sign of the second space derivative is opposite to that in (1.8). Consequently, after reversing the time, (1.9) is a parabolic equation.

Numerous variants of the heat equation (1.8) exist, some of which we shall now explore. Up until now we have assumed that heat propagates in a fixed medium, or at least a still medium. Let us now assume that it propagates in a moving medium, for example, a fluid moving with velocity $V(x, t)$ (a vector valued function in $\mathbb{R}^N$). Then, we must now change the constitutive law since the heat flux is the sum of a diffusive flux (as before) and a convective flux (proportional to the velocity $V$), and proceeding similarly to the arguments above leads us to the convection–diffusion problem

$$
\begin{cases}
c \frac{\partial \theta}{\partial t} + cV \cdot \nabla \theta - k \Delta \theta = f & \text{in } \Omega \times \mathbb{R}_+^+
\theta = 0 & \text{on } \partial \Omega \times \mathbb{R}_+^+
\theta(t = 0, x) = \theta_0(x) & \text{in } \Omega
\end{cases}
$$

(1.10)

The difference between (1.8) and (1.10) is the appearance of a convection term. We measure the balance between this new convection term and the diffusion term by a dimensionless number called the Péclet number, defined by

$$
Pe = \frac{cVL}{k},
$$

(1.11)

where $L$ is a characteristic length of the problem (for example, the diameter of the domain $\Omega$). If the Péclet number is very small then the diffusive effects dominate the convective effects, and model (1.8) is sufficient to describe the phenomenon. If the Péclet number is neither small nor large (we say that it is the order of unity), then model (1.10) is more realistic than (1.8). On the other hand, if the Péclet number is very large, we can simplify (1.10) by removing the diffusion term. We then obtain the equation known as the advection equation

$$
\begin{cases}
c \frac{\partial \theta}{\partial t} + cV \cdot \nabla \theta = f & \text{in } \Omega \times \mathbb{R}_+^+
\theta(t, x) = 0 & \text{for } (x, t) \in \partial \Omega \times \mathbb{R}_+^+ \text{ if } V(x) \cdot n(x) < 0
\theta(t = 0, x) = \theta_0(x) & \text{in } \Omega
\end{cases}
$$

(1.12)

We note the difference in the boundary condition of (1.12) with respect to that of (1.10): we no longer impose that the temperature $\theta$ is zero everywhere on the boundary $\partial \Omega$ but only on those parts of the boundary where the velocity $V$ is re-entrant.
We have therefore described three models of heat propagation by convection and diffusion, (1.8), (1.10), (1.12), which have different regimes of validity depending on different values of the Péclet number. Of course, the analytical or numerical solution of these three problems is very different. This is reflected in the current state of mathematical modelling: there are several competing models and we must choose the 'best'.

In order to understand better the fundamental differences which exist between these models, we temporarily restrict ourselves to the case where \( \Omega = \mathbb{R} \) the whole real line (which rids us of the question of the boundary conditions), where the source term \( f \) is zero, and where the velocity \( V \) is constant. We can then explicitly calculate solutions of these models. For example, (1.10) becomes

\[
\begin{aligned}
\frac{\partial \theta}{\partial t} + V \frac{\partial \theta}{\partial x} - \nu \frac{\partial^2 \theta}{\partial x^2} &= 0 \quad \text{for} \ (x, t) \in \mathbb{R} \times \mathbb{R}_+^* \\
\theta(t = 0, x) &= \theta_0(x) \quad \text{for} \ x \in \mathbb{R}
\end{aligned}
\]

with \( \nu = k/c \), which has solution

\[
\theta(t, x) = \frac{1}{\sqrt{4\pi \nu t}} \int_{-\infty}^{+\infty} \theta_0(y) \exp \left( -\frac{(x - Vt - y)^2}{4\nu t} \right) dy.
\]

A solution of (1.8) is easily obtained by setting \( V = 0 \) in the expression (1.14).

**Exercise 1.2.1** We assume that the initial condition \( \theta_0 \) is continuous and uniformly bounded in \( \mathbb{R} \). Verify that (1.14) is a solution of (1.13).

With the same simplifying hypotheses, the advection equation becomes

\[
\begin{aligned}
\frac{\partial \theta}{\partial t} + V \frac{\partial \theta}{\partial x} &= 0 \quad \text{for} \ (x, t) \in \mathbb{R} \times \mathbb{R}_+^* \\
\theta(t = 0, x) &= \theta_0(x) \quad \text{for} \ x \in \mathbb{R}
\end{aligned}
\]

We verify that

\[
\theta(t, x) = \theta_0(x - Vt)
\]

is a solution of the equation (1.15).

**Exercise 1.2.2** We assume that the initial data \( \theta_0 \) is differentiable and uniformly bounded over \( \mathbb{R} \). Verify that (1.16) is a solution of (1.15). Show that (1.16) is the limit of (1.14) as the parameter \( \nu \) tends to zero.

**Remark 1.2.6** If we solve the heat flow equation (1.8) on a bounded interval (and not in the whole space), we can also calculate an explicit solution by using Fourier analysis (see [4], [38]). This solution would be a little less ‘explicit’ than (1.14) as it is defined as the sum of an infinite series. We remark that it was precisely to solve the heat flow equation that Fourier invented the analysis which takes his name.
Remark 1.2.7 The role of time is fundamentally different in equations (1.8) and (1.12). Indeed, assuming that the source term is zero, \( f = 0 \), if we change the sign of time \( t \) and that of the velocity, the advection equation (1.12) is unchanged (when we change the time we change the current). Conversely, a change in the sign of time in the heat flow equation (1.8) cannot be compensated by any variation in the sign of the data. This is obvious in the explicit form of the solution: (1.16) is invariant by changing the sign of \( t \) and \( V \), whereas (1.14) (with \( V = 0 \)) decreases in time, indicating the ‘arrow’ of time. We say that the advection equation is reversible in time, while the heat flow equation is irreversible in time. This mathematical observation is confirmed by physical intuition: some phenomena are reversible in time, others are not (like the diffusion of a drop of milk in a cup of tea).

Remark 1.2.8 Another fundamental difference between equations (1.8) and (1.12) lies with the property of invariance with respect to change of scale. Let us assume that the source term is zero, \( f = 0 \). It is easy to see that if \( \theta(x, t) \) is a solution of the heat flow equation (1.8), then, for all \( \lambda > 0 \), \( \theta(x/\lambda, t/\lambda^2) \) is also a solution of the same equation (for a different initial value). Likewise, assuming that the velocity \( V \) is constant, if \( \theta(x, t) \) is a solution of the advection equation (1.12), then \( \theta(x/\lambda, t/\lambda) \) is also a solution. We see that the scaling of time is not the same in both cases. We also remark that, in both cases, the equations are invariant under translation in space and in time.

Remark 1.2.9 A surprising property (from the physical point of view) of the heat flow equation (1.8) is that the solution in \((x, t)\) depends on all the initial values in \( \mathbb{R} \) (see formula (1.14)). In particular, in the case of (1.13), if the initial data is positive with compact support, then for all time \( t > 0 \) (no matter how small) the solution is strictly positive over all \( \mathbb{R} \): in other words, the heat propagates ‘instantaneously’ to infinity. We say that the heat propagates with an infinite velocity (which is clearly a limitation of the model). On the other hand, in the advection equation (1.15) the initial data is convected with velocity \( V \) (see formula (1.16)): therefore there is a finite velocity of propagation.

Remark 1.2.10 Thanks to the explicit formulas (1.14) and (1.16), we easily verify that the solutions of the convection–diffusion equation (1.13) and of the advection equation (1.15) satisfy the property

\[
\min_{x \in \mathbb{R}} \theta_0(x) \leq \theta(x, t) \leq \max_{x \in \mathbb{R}} \theta_0(x) \quad \text{for all} \quad (x, t) \in \mathbb{R} \times \mathbb{R}^+,
\]

which is called the maximum principle. This property (which is equally important from the point of view of both mathematics and physics) extends to more general forms of the convection–diffusion equation (1.10) and of the advection equation (1.12). We shall study it more carefully later.
1.3 Some classical models

In this section we shall quickly describe some classical models. Our goal is to present the principal classes of PDEs which we shall study later, and to show that these equations play a very important role in diverse scientific areas. From now on, we shall nondimensionalize all the variables, which will allow us to set the constants in the models equal to 1.

1.3.1 The heat flow equation

As we have seen, the heat flow equation appears as a model in many problems in science and engineering. It is written

\[
\begin{aligned}
\frac{\partial u}{\partial t} - \Delta u &= f \quad \text{in } \Omega \times \mathbb{R}^+ \\
u &= 0 \quad \text{on } \partial \Omega \times \mathbb{R}^+ \\
u(t = 0) &= u_0 \quad \text{in } \Omega.
\end{aligned}
\] (1.17)

This equation is first order in time and second order in space (the order is that of the highest partial derivatives). We shall say that this equation is parabolic (see Section 1.5.2). We have already seen some properties of this equation: irreversibility in time, propagation with infinite velocity, and the maximum principle.

Exercise 1.3.1 We shall find a property of exponential decrease in time (see formula (1.14)) of the solution of the heat flow equation (1.17) in a bounded domain. In one space dimension, we set \( \Omega = (0, 1) \) and we assume that \( f = 0 \). Let \( u(t, x) \) be a regular solution of (1.17). Multiplying the equation by \( u \) and integrating with respect to \( x \), establish the equality

\[
\frac{1}{2} \frac{d}{dt} \left( \int_0^1 u^2(t, x) \, dx \right) = - \int_0^1 \left| \frac{\partial u}{\partial x}(t, x) \right|^2 \, dx.
\]

Show that every continuously differentiable function \( v(x) \) on \([0, 1]\), such that \( v(0) = 0 \), satisfies the Poincaré inequality

\[
\int_0^1 v^2(x) \, dx \leq \int_0^1 \left| \frac{dv}{dx}(x) \right|^2 \, dx.
\]

From this, deduce the exponential decrease in time of \( \int_0^1 u^2(t, x) \, dx \).

1.3.2 The wave equation

The wave equation models propagation of waves or vibration. For example, in two space dimensions it is a model to study the vibration of a stretched elastic membrane (like the skin of a drum). In one space dimension, it is also called the vibrating cord equation. At rest, the membrane occupies a plane domain \( \Omega \). Under the action of
a force normal to the plane with intensity $f$, it deforms and its normal displacement is denoted by $u$ (see Figure 1.2). We assume that it is fixed at the boundary, which gives a Dirichlet boundary condition. The wave equation with solution $u$ is given by

$$
\begin{align*}
\frac{\partial^2 u}{\partial t^2} - \Delta u &= f \quad \text{in } \Omega \times \mathbb{R}^+ \\
\frac{\partial u}{\partial t}(t = 0) &= u_1 \quad \text{in } \Omega \\
\end{align*}
$$

(1.18)

We note that this equation is second order in time and that we therefore need two initial conditions for $u$. We say that this equation is hyperbolic (see Section 1.5.2).

**Exercise 1.3.2** We work in $N = 1$ space dimensions. We assume that the initial data $u_0$ and $u_1$ are regular functions, and that $f = 0$ with $\Omega = \mathbb{R}$. We note that $U_1$ is a primitive of $u_1$. Verify that

$$
u(t, x) = \frac{1}{2} (u_0(x + t) + u_0(x - t)) + \frac{1}{2} (U_1(x + t) - U_1(x - t)), \quad (1.19)$$

is the unique solution of (1.18) in the class of regular functions.

The wave equation shares, with the advection equation (1.12), the important property of **propagation with finite velocity**. Indeed, exercise 1.3.3 shows that the solution at a point $(x, t)$ does not depend on all the initial data but only on the values in a restricted interval called the **domain of dependence** (or light cone; see Figure 1.3). We recall that this property is not shared by the heat flow equation since it is clear, from formula (1.14), that the solution in $(x, t)$ depends on all the values of the initial data.

Another property of the wave equation is its invariance under the change of direction of time. If we change $t$ to $-t$, the form of the equation does not change. We can therefore ‘integrate’ the wave equation in the positive or negative time directions in the same way. We say that the wave equation is **reversible in time**.
Exercise 1.3.3 Verify that the solution (1.19) at the point \((x, t)\) only depends on the values of the initial data \(u_0\) and \(u_1\) in the segment \([x - t, x + t]\). Verify also that \(u(-t, x)\) is a solution of (1.18) in \(\Omega \times \mathbb{R}_-^*\) if we change the sign of the initial velocity \(u_1(x)\).

![Figure 1.3. Domain or cone of dependence of the wave equation.](image)

Exercise 1.3.4 We propose showing a principle of conservation of energy for the wave equation (1.18) without using the explicit formula (1.19). In one space dimension, we set \(\Omega = (0, 1)\) and we assume \(f = 0\). Let \(u(t, x)\) be a regular solution of (1.18). Multiplying the equation by \(\partial u / \partial t\) and integrating with respect to \(x\), establish the energy equality

\[
\frac{d}{dt} \left( \int_0^1 \left| \frac{\partial u}{\partial t}(t, x) \right|^2 dx + \int_0^1 \left| \frac{\partial u}{\partial x}(t, x) \right|^2 dx \right) = 0.
\]

Compare this with what happens for the heat equation.

### 1.3.3 The Laplacian

For certain choices of source term \(f\), the solution of the heat flow equation (1.17) reaches a steady (or stationary) state, that is, \(u(t, x)\) tends to a limit \(u_\infty(x)\) as time \(t\) tends to infinity. Often, it is interesting to calculate this steady state directly. In this case, for a source term \(f(x)\) which is independent of time, we solve an equation which is second order in space

\[
\begin{cases}
-\Delta u = f & \text{in } \Omega, \\
 u = 0 & \text{on } \partial \Omega,
\end{cases}
\]

which we call the Laplacian or Laplace’s equation. We say that this equation is elliptic (see Section 1.5.2). We remark that the Laplacian is also the stationary version of the wave equation (1.19). The Laplacian also occurs in numerous fields of science and engineering. For example, (1.20) models the vertical displacement of an elastic membrane subjected to a normal force \(f\) and fixed around its boundary.
1.3.4 Schrödinger’s equation

Schrödinger’s equation describes the evolution of the wave function $u$ of a particle subject to a potential $V$. Recall that $u(t, x)$ is a function of $\mathbb{R}^+ \times \mathbb{R}^N$ with values in $\mathbb{C}$ and that the square of its modulus $|u|^2$ is interpreted as the probability that the particle is found at the point $(t, x)$. The potential $V(x)$ is a real-valued function. The wave function is a solution of

$$
\begin{aligned}
&i \frac{\partial u}{\partial t} + \Delta u - V u = 0 \quad \text{in } \mathbb{R}^N \times \mathbb{R}^+ \\
u(t = 0) = u_0 \quad \text{in } \mathbb{R}^N
\end{aligned}
$$

(1.21)

There are no boundary conditions in (1.21) since the equation holds over the whole of space (which has no boundary). Nevertheless, we shall see that a ‘reasonable’ choice of function space in which to look for the solution implies de facto a condition of decay to infinity of $u$ which can be interpreted as a boundary condition at infinity.

**Exercise 1.3.5** We propose to show principles of energy conservation for Schrödinger’s equation (1.21). Let $u(t, x)$ be a regular solution of (1.21) in one space dimension which decreases to zero (as does $\partial u / \partial x$) as $|x| \to +\infty$. Show that for every differentiable function $v(t)$ we have

$$
\mathcal{R} \left( \frac{\partial v}{\partial t} \overline{v} \right) = \frac{1}{2} \frac{\partial |v|^2}{\partial t},
$$

where $\mathcal{R}$ denotes the real part and $\overline{v}$ the complex conjugate of $v$. Multiplying the equation by $\overline{u}$ and integrating with respect to $x$, establish the energy equality

$$
\int_\mathbb{R} |u(t, x)|^2 \, dx = \int_\mathbb{R} |u_0(x)|^2 \, dx.
$$

Multiplying the equation by $\partial \overline{u} / \partial t$, show that

$$
\int_\mathbb{R} \left( \left| \frac{\partial u}{\partial x}(t, x) \right|^2 + V(x) |u(t, x)|^2 \right) \, dx = \int_\mathbb{R} \left( \left| \frac{\partial u_0}{\partial x}(x) \right|^2 + V(x) |u_0(x)|^2 \right) \, dx.
$$

1.3.5 The Lamé system

The Lamé system is a particular case of the linearized stationary elasticity equations which model deformations of a solid under the assumption of small deformations and of small displacements (see Section 5.3.1 for further details on the modelling). To obtain the Lamé system, we assume that the solid is homogeneous and isotropic and that it is fixed at the boundary. The principal difference from the preceding models is that here we have a system of equations, that is, several coupled equations. The solid at rest occupies the domain $\Omega$ of the space $\mathbb{R}^N$. Under the action of a force $f$
it deforms, and each point \( x \) moves to \( x + u(x) \). The force \( f(x) \) is a vector-valued function of \( \Omega \) in \( \mathbb{R}^N \), as is the displacement \( u(x) \). This is a solution of

\[
\begin{cases}
-\mu \Delta u - (\mu + \lambda) \nabla (\text{div} u) = f & \text{in } \Omega \\
u = 0 & \text{on } \partial \Omega
\end{cases}
\]  

(1.22)

where \( \lambda \) and \( \mu \) are two constants, the Lamé constants, which are characteristics of the homogeneous, isotropic material which comprises the solid. For mechanical reasons these constants satisfy \( \mu > 0 \) and \( 2\mu + N\lambda > 0 \). The Dirichlet boundary condition for \( u \) reflects the fact that the solid is assumed fixed and immovable at its boundary, \( \partial \Omega \).

The system (1.22) has been written in vector notation. If we denote by \( f_i \) and \( u_i \), for \( 1 \leq i \leq N \), the components of \( f \) and \( u \) in the canonical basis of \( \mathbb{R}^N \), (1.22) is equivalent to

\[
\begin{cases}
-\mu \Delta u_i - (\mu + \lambda) \frac{\partial (\text{div} u)}{\partial x_i} = f_i & \text{in } \Omega \\
u_i = 0 & \text{on } \partial \Omega
\end{cases}
\]

for \( 1 \leq i \leq N \). We remark that, if \( (\mu + \lambda) \neq 0 \), then the equations for each component \( u_i \) are coupled by the divergence term. Obviously, in \( N = 1 \) dimension, the Lamé system has only one equation and reduces to the Laplacian.

### 1.3.6 The Stokes system

The Stokes system models the flow of a viscous incompressible fluid with small velocity. We assume that the fluid occupies a domain \( \Omega \) and that it adheres to the boundary, that is, its velocity is zero at the boundary (which leads to a Dirichlet boundary condition). Under the action of a force \( f(x) \) (a function of \( \Omega \) in \( \mathbb{R}^N \)), the velocity \( u(x) \) (a vector) and the pressure \( p(x) \) (a scalar) are solutions of

\[
\begin{cases}
\nabla p - \mu \Delta u = f & \text{in } \Omega \\
\text{div} u = 0 & \text{in } \Omega \\
u = 0 & \text{on } \partial \Omega
\end{cases}
\]  

(1.23)

where \( \mu > 0 \) is the fluid viscosity. We note that there are a further \( N \) equations \( \nabla p - \mu \Delta u = f \) (corresponding to the **conservation of momentum**), and one other equation \( \text{div} u = 0 \) called the **incompressibility condition** (which corresponds to **conservation of mass**). If the space dimension is \( N = 1 \), the Stokes system is uninteresting as we easily see that the velocity is zero and the pressure is a primitive of the force. On the other hand, in dimensions \( N \geq 2 \), the Stokes system makes good sense: in particular, there exist nontrivial incompressible velocity fields (take, for example, a curl).

### 1.3.7 The plate equations

We consider small elastic deformations of a thin plane plate (which is negligible in its other dimensions). If we denote by \( \Omega \) the average surface of the plate, and \( f(x) \)
(a function of $\Omega$ in $\mathbb{R}$) the resultant normal of the forces, then the normal component of the displacement $u(x)$ (a scalar) is the solution of the thin plate equation

$$\begin{cases}
\Delta (\Delta u) = f & \text{in } \Omega \\
u = 0 & \text{on } \partial \Omega \\
\frac{\partial u}{\partial n} = 0 & \text{on } \partial \Omega
\end{cases} \quad (1.24)$$

where we denote by $\frac{\partial u}{\partial n} = \nabla u \cdot n$ with $n$ the outward unit normal vector to $\partial \Omega$. We remark that this is a partial differential equation which is fourth order in space (also called the bi-Laplacian). This is why it is necessary to have two boundary conditions. These boundary conditions represent the clamping of the plate (there is neither displacement nor rotation of the edge of the plate).

We remark that it is possible to justify the plate equation (1.24) asymptotically from the Lamé system (1.22) by letting the thickness of the plate to tend to zero. This is an example of mathematical modelling.

### 1.4 Numerical calculation by finite differences

#### 1.4.1 Principles of the method

Apart from some very particular cases, it is impossible to calculate explicitly the solutions of the different models presented above. It is therefore necessary to have recourse to numerical calculation on a computer to estimate these solutions both qualitatively and quantitatively. The principle of all methods for the numerical solution of PDEs is to obtain discrete numerical values (that is, a finite number) which ‘approximate’ (in a suitable sense, to be made precise) the exact solution. In this process we must be aware of two fundamental points: first, we do not calculate exact solutions but approximate ones; second, we discretize the problem by representing functions by a finite number of values, that is, we move from the ‘continuous’ to the ‘discrete’.

There are numerous methods for the numerical approximation of PDEs. We present one of the oldest and simplest, called the finite difference method (later we shall see another method, called the finite element method). For simplicity, we limit ourselves to one space dimension (see Section 2.2.6 for higher dimensions). For the moment, we shall only consider the practical principles of this method, that is, the construction of what we call the numerical schemes. We reserve the theoretical justification of these schemes for Chapter 2, that is, the study of their convergence (in what way the approximate discrete solutions are close to the exact continuous solutions).

To discretise the spatio-temporal continuum, we introduce a space step $\Delta x > 0$ and a time step $\Delta t > 0$ which will be the smallest scales represented by the numerical method. We define a mesh or discrete coordinates in space and time (see Figure 1.4)

$$(t_n, x_j) = (n\Delta t, j\Delta x) \quad \text{for } n \geq 0, \; j \in \mathbb{Z}.$$
Figure 1.4. Finite difference mesh.

We denote by $u^n_j$ the value of the discrete solution at $(t_n, x_j)$, and $u(t, x)$ the (unknown) exact solution. The principle of the finite difference method is to replace the derivatives by finite differences by using Taylor series in which we neglect the remainders. For example, we approximate the second space derivative (the Laplacian in one dimension) by

$$
-\frac{\partial^2 u}{\partial x^2}(t_n, x_j) \approx \frac{-u^n_{j-1} + 2u^n_j - u^n_{j+1}}{(\Delta x)^2}
$$

(1.25)

where we recall the Taylor formula

$$
-u(t, x - \Delta x) + 2u(t, x) - u(t, x + \Delta x) = -(\Delta x)^2 \frac{\partial^2 u}{\partial x^2}(t, x)
$$

$$
-\frac{(\Delta x)^4}{12} \frac{\partial^4 u}{\partial x^4}(t, x) + O\left((\Delta x)^6\right)
$$

(1.26)

If $\Delta x$ is ‘small’, formula (1.25) is a ‘good’ approximation (it is natural, but not unique). The formula (1.25) is called centred since it is symmetric in $j$.

To discretize the convection–diffusion equation

$$
\frac{\partial u}{\partial t} + V \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = 0
$$

(1.27)

we must also discretize the convection term. A centred formula gives

$$
V \frac{\partial u}{\partial x}(t_n, x_j) \approx V \frac{u^n_{j+1} - u^n_{j-1}}{2\Delta x}
$$

It only remains to do the same thing for the time derivative. Again we have a choice between finite difference schemes: centred or one sided. Let us look at three ‘natural’ formulas.
1. As a first choice, the centred finite difference

\[ \frac{\partial u}{\partial t}(t_n, x_j) \approx \frac{u_{j}^{n+1} - u_{j}^{n-1}}{2\Delta t} \]

leads to a scheme which is completely symmetric with respect to \( n \) and \( j \) (called the centred scheme or Richardson’s scheme)

\[ \frac{u_{j}^{n+1} - u_{j}^{n-1}}{2\Delta t} + V \frac{u_{j+1}^{n} - u_{j-1}^{n}}{2\Delta x} + \nu \frac{-u_{j-1}^{n} + 2u_{j}^{n} - u_{j+1}^{n}}{(\Delta x)^2} = 0. \] (1.28)

Even though it is ‘natural’ this scheme cannot calculate approximate solutions of the convection–diffusion equation (1.27) (see the numerical example of Figure 1.5)! We shall justify the inability of this scheme to approximate the exact solution in Lemma 2.2.23. For the moment, we shall simply say that the difficulty comes from the centred character of the finite difference which approximates the time derivative.

2. A second choice is the one-sided upwind scheme (we go back in time) which gives the backward Euler scheme

\[ \frac{\partial u}{\partial t}(t_n, x_j) \approx \frac{u_{j}^{n} - u_{j}^{n-1}}{\Delta t} \]

which leads to

\[ \frac{u_{j}^{n} - u_{j}^{n-1}}{\Delta t} + V \frac{u_{j+1}^{n} - u_{j-1}^{n}}{2\Delta x} + \nu \frac{-u_{j-1}^{n} + 2u_{j}^{n} - u_{j+1}^{n}}{(\Delta x)^2} = 0. \] (1.29)

3. The third choice is the opposite of the preceding: the downwind one-sided finite difference (we go forward in time; we also talk of the forward Euler scheme)

\[ \frac{\partial u}{\partial t}(t_n, x_j) \approx \frac{u_{j}^{n+1} - u_{j}^{n}}{\Delta t} \]

which leads to

\[ \frac{u_{j}^{n+1} - u_{j}^{n}}{\Delta t} + V \frac{u_{j+1}^{n} - u_{j-1}^{n}}{2\Delta x} + \nu \frac{-u_{j-1}^{n} + 2u_{j}^{n} - u_{j+1}^{n}}{(\Delta x)^2} = 0. \] (1.30)

The principal difference between these last two schemes is that (1.29) is called implicit since we must solve a system of linear equations to calculate the values \((u_{j}^{n})_{j\in\mathbb{Z}}\) as functions of the preceding values \((u_{j}^{n-1})_{j\in\mathbb{Z}}\), while (1.30) is called explicit since it immediately gives the values \((u_{j}^{n+1})_{j\in\mathbb{Z}}\) as a function of \((u_{j}^{n})_{j\in\mathbb{Z}}\). The shift of 1 in the index \( n \) between the schemes (1.29) and (1.30) is only evident when we rewrite (1.30) in the form

\[ \frac{u_{j}^{n} - u_{j}^{n-1}}{\Delta t} + V \frac{u_{j+1}^{n-1} - u_{j-1}^{n-1}}{2\Delta x} + \nu \frac{-u_{j-1}^{n-1} + 2u_{j}^{n-1} - u_{j+1}^{n-1}}{(\Delta x)^2} = 0. \]
In the three schemes which we have defined, there must be initial data to start the iterations in \( n \): the initial values \( (u_0^j)_{j \in \mathbb{Z}} \) are defined by, for example, \( u_0^j = u_0(j\Delta x) \) where \( u_0 \) is the initial data of the convection–diffusion equation (1.27). We remark that the ‘bad’ centred scheme (1.28) has an additional difficulty in starting: for \( n = 1 \) we also have to know the values \( (u_1^j)_{j \in \mathbb{Z}} \) which, therefore, must be calculated in another way (for example, by applying one of the two other schemes).

### 1.4.2 Numerical results for the heat flow equation

We start by making some simple numerical tests in the case where \( V = 0 \) and \( \nu = 1 \), that is, we solve the heat flow equation numerically. As initial condition, we choose the function

\[
 u_0(x) = \max(1 - x^2, 0).
\]

To be able to compare the numerical solutions with the exact (1.14), we want to work on the infinite domain \( \Omega = \mathbb{R} \), that is, calculate, for each \( n \geq 0 \), an infinite number of values \( (u_n^j)_{j \in \mathbb{Z}} \), but the computer will not allow this as the memory is finite! To a first approximation, we therefore replace \( \mathbb{R} \) by the ‘large’ domain \( \Omega = (-10, +10) \) equipped with Dirichlet boundary conditions. The validity of this approximation is confirmed by the numerical calculations below. We fix the space step at \( \Delta x = 0.05 \): there are therefore 401 values \( (u_n^j)_{-200 \leq j \leq +200} \) to calculate. We should remember that the values \( u_n^j \) calculated by the computer are subject to rounding errors and are therefore not the exact values of the difference scheme: nevertheless, in the calculations presented here, these rounding errors are completely negligible and are in no way
responsible for the phenomena which we shall observe. On all the figures we show the exact solution, calculated by the explicit formula (1.14), and the approximate numerical solution under consideration.

Let us first look at the outcome of the centred scheme (1.28): since as we have said, this scheme is not able to calculate approximate solutions of the heat flow equation. Whatever the choice of the time step $\Delta t$, this scheme is \textbf{unstable}, that is the numerical solution oscillates unboundedly if we decrease the step sizes $\Delta x$ and $\Delta t$. This highly characteristic phenomenon (which appears rapidly) is illustrated by Figure 1.5. We emphasize that \textbf{whatever the choice} of steps $\Delta t$ and $\Delta x$, we see these oscillations (which are nonphysical). We say that the scheme is unconditionally unstable. A rigorous justification will be given in the following chapter (see lemma 2.2.23).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{heat_equation Implicit scheme, CFL=2., 200}
\caption{Implicit scheme with $\nu \Delta t = 2(\Delta x)^2$.}
\end{figure}

Contrary to the preceding scheme, the implicit scheme (1.29) calculates ‘good’ approximate solutions of the heat flow equation \textbf{whatever} the time step $\Delta t$ (see Figure 1.6). In particular, we never see numerical oscillation for any choice of steps $\Delta t$ and $\Delta x$. We say that the implicit scheme is unconditionally stable.

Let us now consider the explicit scheme (1.30): numerical experiments show that we obtain numerical oscillations depending on the time step $\Delta t$ (see Figure 1.7). The stability limit is easy to find experimentally: if the choice of steps $\Delta t$ and $\Delta x$ satisfy the condition
\begin{equation}
2\nu \Delta t \leq (\Delta x)^2
\end{equation}
the scheme is stable, while if (1.31) is not satisfied, then the scheme is unstable. We say that the explicit scheme is conditionally stable. The stability condition (1.31)
is one of the simplest but most profound observations in numerical analysis. It was discovered in 1928 (before the appearance of the first computers!) by Courant, Friedrichs, and Lewy. It takes the name CFL condition or the Courant, Friedrichs, Lewy condition.

\[ \nu \Delta t = 0.4(\Delta x)^2 \] (top) and \[ \nu \Delta t = 0.51(\Delta x)^2 \] (bottom).

Figure 1.7. Explicit scheme with \( \nu \Delta t = 0.4(\Delta x)^2 \) (top) and \( \nu \Delta t = 0.51(\Delta x)^2 \) (bottom).

We shall briefly justify this stability condition (a more thorough analysis will be carried out in the next chapter). Rewriting the explicit scheme in the form

\[ u_{j}^{n+1} = \frac{\nu \Delta t}{(\Delta x)^2} u_{j-1}^{n} + \left(1 - 2 \frac{\nu \Delta t}{(\Delta x)^2}\right) u_{j}^{n} + \frac{\nu \Delta t}{(\Delta x)^2} u_{j+1}^{n}. \] (1.32)
If the CFL condition is satisfied, then (1.32) shows that $u_{j}^{n+1}$ is a convex combination of the values at the preceding time $u_{j-1}^{n}, u_{j}^{n}, u_{j+1}^{n}$ (all of the coefficients on the right-hand side of (1.32) are positive and their sum is 1). In particular, if the initial data $u_0$ is bounded by two constants $m$ and $M$ such that

$$m \leq u_{j}^{0} \leq M \text{ for all } j \in \mathbb{Z},$$

then a recurrence easily shows that the same inequalities remain true for all time

$$m \leq u_{j}^{n} \leq M \text{ for all } j \in \mathbb{Z} \text{ and for all } n \geq 0. \quad (1.33)$$

Property (1.33) prevents the scheme from oscillating unboundedly: it is therefore stable subject to the CFL condition. Property (1.33) is called a discrete maximum principle: it is the discrete equivalent of the continuous maximum principle for exact solutions which we have seen in remark 1.2.10.

Suppose, on the other hand, the CFL condition is not satisfied, that is,

$$2\nu \Delta t > (\Delta x)^2,$$

then, for certain initial data the scheme is unstable (it may be stable for certain ‘exceptional’ initial data: for example, if $u_0 \equiv 0!$). Let us take the initial data defined by

$$u_{j}^{0} = (-1)^{j}$$

which is uniformly bounded. A simple calculation shows that

$$u_{j}^{n} = (-1)^{j} \left(1 - 4 \frac{\nu \Delta t}{(\Delta x)^2}\right)^{n}$$

which tends, in modulus, to infinity as $n$ tends to infinity since $1 - 4\nu \Delta t/(\Delta x)^2 < -1$. The explicit scheme is therefore unstable if the CFL condition is not satisfied.

**Exercise 1.4.1** The aim of this exercise is to show that the implicit scheme (1.29), with $V = 0$, also satisfies the discrete maximum principle. We impose Dirichlet boundary conditions, that is, formula (1.29) is valid for $1 \leq j \leq J$ and we fix $u_{j}^{0} = u_{j+1}^{0} = 0$ for all $n \in \mathbb{N}$. Take two constants $m \leq 0 \leq M$ such that $m \leq u_{j}^{0} \leq M$ for $1 \leq j \leq J$. Verify that we can uniquely calculate the $u_{j}^{n+1}$ as a function of $u_{j}^{n}$. Show that for all time $n \geq 0$ we again have the inequalities $m \leq u_{j}^{n} \leq M$ for $1 \leq j \leq J$ (without any condition on $\Delta t$ and $\Delta x$).

If we have illuminated the question of the stability of the explicit scheme a little, we have not said anything about its convergence, that is, its capacity to approximate the exact solution. We shall answer this question rigorously in the following chapter. We remark that stability is a necessary condition for convergence, but it is not sufficient. We shall be content for the moment with experimentally verifying the convergence of the scheme, that is, when the space and time steps become smaller and smaller,
the corresponding numerical solutions converge and their limit is the exact solution (we can check this as the exact solution is available). In Figure 1.8 we numerically verify that if we reduce the space step \( \Delta x \) (which has values 0.5, 0.1, and 0.05) and the time step \( \Delta t \) by keeping the ratio \( \nu \Delta t/(\Delta x)^2 \) (the CFL number) constant, then the numerical solution becomes closer and closer to the exact solution. (The comparison is carried out at the same final time \( t = 1 \), therefore the number of time steps grows as the time step \( \Delta t \) decreases.) This process of ‘numerical verification of convergence’ is very simple and we should never hesitate to use it if nothing better is available (that is, if the theoretical convergence analysis is impossible or too difficult).

1.4.3 Numerical results for the advection equation

We shall carry out a second series of numerical experiments on the convection–diffusion equation (1.27) with a nonzero velocity \( V = 1 \). We take the same data as before and we choose the explicit scheme with \( \nu \Delta t = 0.4(\Delta x)^2 \). We look at the influence of the diffusion constant \( \nu \) (or the inverse of the Péclet number) on the stability of the scheme. Figure 1.9 shows that the scheme is stable when \( \nu = 1 \), unstable for \( \nu = 0.01 \), and that for the intermediate value \( \nu = 0.1 \), the scheme seems stable but the approximate solution is slightly different from the exact solution. Clearly, the smaller the inverse of the Péclet number \( \nu \) is, the more the convective term dominates the diffusive term. Consequently, the CFL condition (1.31), obtained when the velocity \( V \) is zero, is less and less valid as \( \nu \) decreases.
Figure 1.9. Explicit scheme for the convection–diffusion equation with $\nu \Delta t = 0.4(\Delta x)^2$ and $V = 1$. At the top, $\nu = 1$, in the middle $\nu = 0.1$, and at the bottom $\nu = 0.01$. 
To understand this phenomenon, we examine the advection equation which is obtained in the limit $\nu = 0$. We remark first that the CFL condition (1.31) is automatically satisfied when $\nu = 0$ (whatever $\Delta t$ and $\Delta x$), which seems to contradict the experimental result at the bottom of Figure 1.9.

![TRANSPORT EQUATION, FINAL TIME=1](image)

Figure 1.10. Explicit centred scheme for the advection equation with $\Delta t = 0.9\Delta x$, $V = 1$, $\nu = 0$.

For the advection equation (that is, (1.27) with $\nu = 0$), the explicit scheme (1.30) may be rewritten

$$u_{j}^{n+1} = \frac{V \Delta t}{2 \Delta x} u_{j}^{n-1} + u_{j}^{n} - \frac{V \Delta t}{2 \Delta x} u_{j+1}^{n}. \tag{1.34}$$

This scheme leads to the oscillations in Figure 1.10 under the same experimental conditions as the bottom of Figure 1.9. We see that $u_{j}^{n+1}$ is never (no matter what $\Delta t$) a convex combination of $u_{j-1}^{n}$, $u_{j}^{n}$, and $u_{j+1}^{n}$. Therefore, there cannot be a discrete maximum principle for this scheme, which is an additional indication of its instability (a rigorous proof will be given in lemma 2.3.1). This instability occurs because, in the explicit scheme (1.34), we have chosen to use a centred approximation to the convective term. We can, however, make this term one-sided as we have done for the time derivative. Two choices are possible: weighting to the right or left. The sign of the velocity $V$ is crucial: here we assume that $V > 0$ (a symmetric argument is possible if $V < 0$). For $V > 0$, the weighting to the right is called downwinding: we obtain

$$V \frac{\partial u}{\partial x}(t_{n}, x_{j}) \approx V \frac{u_{j+1}^{n} - u_{j}^{n}}{\Delta x}$$

we try to find ‘information’ by following the current. This leads to a ‘disastrous’
downwind scheme
\[ \frac{u_{j}^{n+1} - u_{j}^{n}}{\Delta t} + V \frac{u_{j+1}^{n} - u_{j}^{n}}{\Delta x} = 0 \] (1.35)

which is as unstable as the centred scheme. On the other hand, the **upwinding** (which is to the left if \( V > 0 \)), looks for ‘information’ by going against the current

\[ V \frac{\partial u}{\partial x}(t_{n}, x_{j}) \approx V \frac{u_{j}^{n} - u_{j-1}^{n}}{\Delta x} \]

leading to an explicit upwind scheme

\[ \frac{u_{j}^{n+1} - u_{j}^{n}}{\Delta t} + V \frac{u_{j}^{n} - u_{j-1}^{n}}{\Delta x} = 0 \] (1.36)

which gives the results of Figure 1.11. We verify easily that the scheme (1.36) is stable

Figure 1.11. Explicit upwind scheme for the advection equation with \( \Delta t = 0.9 \) \( \Delta x \), \( V = 1 \).

under a new CFL condition (different from the preceding CFL condition (1.31))

\[ |V| \Delta t \leq \Delta x. \] (1.37)

Indeed, we can rewrite (1.36) in the form

\[ u_{j}^{n+1} = \frac{V \Delta t}{\Delta x} u_{j-1}^{n} + \left( 1 - \frac{V \Delta t}{\Delta x} \right) u_{j}^{n}, \]
which shows that, if condition (1.37) is satisfied, $u_{n+1}^j$ is a convex combination of $u_{n-1}^j$ and $u_{n}^j$. Consequently, the one-sided upwind scheme (1.36) satisfies a discrete maximum principle, which implies conditional stability. The idea of upwinded methods is another major idea in numerical analysis. It is particularly important in all fluid mechanics problems where it was first discovered, but it appears in many other models.

The conclusion of this study on the advection equation is that for the convection–diffusion model with a small diffusion constant $\nu$, we must upwind the convective term and obey the CFL condition (1.37) rather than (1.31). With this price we can improve the results of Figure 1.9.

**Exercise 1.4.2** Show that if (1.37) is not satisfied, the upwind scheme (1.36) for the advection equation is unstable for the initial data $u_0^j = (-1)^j$.

**Exercise 1.4.3** Write an explicit scheme centred in space for the wave equation (1.18) in one space dimension and without source term. Specify how to start the iterations in time. Verify the existence of a discrete cone of dependence analogous to the continuous one shown in figure 1.3. Deduce that, if this scheme converges, the time and space steps must satisfy the (CFL-like) condition $\Delta t \leq \Delta x$.

The conclusions of this section are numerous and will feed the reflections of the subsequent chapter. First of all, all ‘reasonable’ numerical schemes do not work, far from it. We meet stability problems (without even considering convergence) which require us to analyse these schemes: this is the raison d’être of numerical analysis which reconciles practical objectives and theoretical studies. Finally, the ‘good’ numerical schemes must have a certain number of properties (for example, the discrete maximum principle, or upwinding) which are the expression (at the discrete level) of the physical properties or the mathematics of the PDE. We cannot therefore skimp on a good understanding of the physical modelling and of the mathematical properties of the models if we want to have good numerical simulations.

## 1.5 Remarks on mathematical models

We finish this chapter with a number of definitions which allow the reader to understand the terms in classical works on numerical analysis.

### 1.5.1 The idea of a well-posed problem

**Definition 1.5.1** We use the term **boundary value problem** to refer to a PDE equipped with boundary conditions on the entire boundary of the domain in which it is posed.
For example, the Laplacian (1.20) is a boundary value problem. Conversely, the ordinary differential equation

\begin{align*}
\frac{dy}{dt} &= f(t, y) \quad \text{for } 0 < t < T \\
y(t = 0) &= y_0
\end{align*}

is not a boundary value problem as it is posed on an interval \((0, T)\), with \(0 < T \leq +\infty\), it only has ‘boundary’ conditions at \(t = 0\) (and not at \(t = T\)).

**Definition 1.5.2** We say **Cauchy problem** to mean a PDE where, for at least one variable (usually time \(t\)), the ‘boundary’ conditions are initial conditions (that is, only hold at the boundary \(t = 0\), and not at \(t = T\)).

For example, the ordinary differential equation (1.38) is a Cauchy problem, but the Laplacian (1.20) is not (no matter which component of the space variable \(x\) we make to play the role of time).

Numerous models are, at the same time, boundary value problems and Cauchy problems. Thus, the heat flow equation (1.8) is a Cauchy problem with respect to the time variable \(t\) and a boundary value problem with respect to the space variable \(x\). All the models we shall study in this course belong to one of these two categories of problem.

The fact that a mathematical model is a Cauchy problem or a boundary value problem does not automatically imply that it is a ‘good’ model. The expression **good model** is not used here in the sense of the physical relevance of the model and of its results, but in the sense of its mathematical coherence. As we shall see, this mathematical coherence is a necessary condition before we can consider numerical simulations and physical interpretations. The mathematician Jacques Hadamard gave a definition of what is a ‘good’ model, while speaking about **well-posed problems** (an ill-posed problem is the opposite of a well-posed problem). We denote by \(f\) the data (the right-hand side, the initial conditions, the domain, etc.), \(u\) the solution sought, and \(A\) ‘the operator’ which acts on \(u\). We are using abstract notation, \(A\) denotes simultaneously the PDE and the type of initial or boundary conditions. The problem is therefore to find \(u\), the solution of

\[
A(u) = f. \tag{1.39}
\]

**Definition 1.5.3** We say that problem (1.39) is **well-posed** if for all data \(f\) it has a unique solution \(u\), and if this solution \(u\) depends continuously on the data \(f\).

Let us examine Hadamard’s definition in detail: it contains, in fact, three conditions for the problem to be well-posed. First, a solution must at least exist: this is the least we can ask of a model supposed to represent reality! Second, the solution must be unique: this is more delicate since, while it is clear that, if we want to predict tomorrow’s weather, it is better to have ‘sun’ or ‘rain’ (with an exclusive
REMARKS ON MATHEMATICAL MODELS

'or') but not both with equal chance, there are other problems which 'reasonably' have several or an infinity of solutions. For example, problems involving finding the best route often have several solutions: to travel from the South to the North Pole then any meridian will do, likewise, to travel by plane from Paris to New York, your travel agency sometimes makes you go via Brussels or London, rather than directly, because it can be more economic. Hadamard excludes this type of problem from his definition since the multiplicity of solutions means that the model is indeterminate: to make the final choice between all of those that are best, we use another criterion (which has been ‘forgotten’ until now), for example, the most practical or most comfortable journey. This is a situation of current interest in applied mathematics: when a model has many solutions, we must add a selection criterion to obtain the ‘good’ solution (see, for a typical example, problems in gas dynamics [23]).

Third, and this is the least obvious condition 

a priori,

the solution must depend continuously on the data. At first sight, this seems a mathematical fantasy, but it is crucial from the perspective of numerical approximation. Indeed, numerically calculating an approximate solution of (1.39) amounts to perturbing the data (when continuous becomes discrete) and solving (1.39) for the perturbed data. If small perturbations of the data lead to large perturbations of the solution, there is no chance that the numerical approximation will be close to reality (or at least to the exact solution). Consequently, this continuous dependence of the solution on the data is an absolutely necessary condition for accurate numerical simulations. We note that this condition is also very important from the physical point of view since measuring apparatus will not give us absolute precision: if we are unable to distinguish between two close sets of data which can lead to very different phenomena, the model represented by (1.39) has no predictive value, and therefore is of almost no practical interest.

We finish by acknowledging that, at this level of generality, the definition (1.5.3) is a little fuzzy, and that to give it a precise mathematical sense we should say in which function spaces we put the data or look for the solution, and which norms or topologies we use for the continuity. It is not uncommon that changing the space (which can appear anodyne) implies very different properties of existence or uniqueness!

Exercise 1.5.1 The point of this exercise is to show that the Cauchy problem for the Laplacian is ill-posed. Take a two-dimensional domain \( \Omega = (0, 1) \times (0, 2\pi) \). We consider the following Cauchy problem in \( x \) and boundary value problem in \( y \)

\[
\begin{align*}
&-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0 & \text{in } \Omega \\
u(x, 0) = u(x, 2\pi) = 0 & \text{for } 0 < x < 1 \\
u(0, y) = 0, \ & \frac{\partial u}{\partial x}(0, y) = -e^{-\sqrt{n}} \sin(ny) & \text{for } 0 < y < 2\pi
\end{align*}
\]

Verify that \( u(x, y) = (e^{-\sqrt{n}}/n) \sin(ny) \text{sh}(nx) \) is a solution. Show that the initial condition and all its derivatives at \( x = 0 \) converge uniformly to 0, while, for all \( x > 0 \), the solution \( u(x, y) \) and all its derivatives are unbounded as \( n \) tends to infinity.
1.5.2 Classification of PDEs

Definition 1.5.4 The order of a partial differential equation is the order of the highest derivative in the equation.

For example, the Laplacian (1.20) is a second order equation, while the plate equation (1.24) is a fourth order equation. We often distinguish between the order with respect to the time variable \( t \) and with respect to the space variable \( x \). Therefore, we say that heat flow equation (1.8) is first order in time and second order in space; likewise, the wave equation (1.18) is second order in space-time.

In order to understand the vocabulary often used with PDEs, that is, elliptic, parabolic, or hyperbolic, we shall briefly classify linear, second order PDEs acting on real functions of two real variables \( u(x, y) \) (we shall not carry out a systematic classification for all PDEs). Such an equation is written

\[
a \frac{\partial^2u}{\partial x^2} + b \frac{\partial^2u}{\partial x \partial y} + c \frac{\partial^2u}{\partial y^2} + d \frac{\partial u}{\partial x} + e \frac{\partial u}{\partial y} + f u = g.
\]  

(1.40)

For simplicity we assume that the coefficients \( a, b, c, d, e, f \) are constant.

Definition 1.5.5 We say that the equation (1.40) is elliptic if \( b^2 - 4ac < 0 \), parabolic if \( b^2 - 4ac = 0 \), and hyperbolic if \( b^2 - 4ac > 0 \).

The origin of this vocabulary is in the classification of conic sections, from which Definition 1.5.5 is copied. Indeed, it is well-known that the second degree equation

\[
ax^2 + bxy + cy^2 + dx + ey + f = 0
\]

defines a plane curve which is (except in some degenerate cases) an ellipse if \( b^2 - 4ac < 0 \), a parabola if \( b^2 - 4ac = 0 \), and a hyperbola if \( b^2 - 4ac > 0 \).

If we apply Definition 1.5.5 to the various second order models we have stated in this chapter (replacing the variables \( x, y \) by the variables \( t, x \) in one space dimension), we conclude that the heat flow equation is parabolic (as is the convection–diffusion equation), that the Laplacian is elliptic, and that the wave equation is hyperbolic. A suitable generalisation of this definition allows us to check that the advection equation is hyperbolic, and that the Stokes, elasticity, and plate equations are elliptic. In general, stationary problems (independent of time) are modelled by elliptic PDEs, while evolution problems are modelled by parabolic or hyperbolic PDEs.

We shall see later that boundary value problems are well posed for elliptic PDEs, while problems which are Cauchy in time and boundary value problems in space are well-posed for parabolic or hyperbolic PDEs. There are therefore important differences in behaviour between these two types of equation.

Remark 1.5.6 The elliptic, hyperbolic or parabolic character of the equations (1.40) is not modified by a change of variable. Let \( (x, y) \to (X, Y) \) be such a change of variable which is nonsingular, that is, its Jacobian \( J = X_x Y_y - X_y Y_x \) is not zero (denoting by \( Z_z \) the derivative of \( Z \) with respect to \( z \)). A simple but tedious calculation shows that (1.40) becomes

\[
A \frac{\partial^2u}{\partial X^2} + B \frac{\partial^2u}{\partial X \partial Y} + C \frac{\partial^2u}{\partial Y^2} + D \frac{\partial u}{\partial X} + E \frac{\partial u}{\partial Y} + F u = G,
\]
with \( A = aX_x^2 + bX_xX_y + cX_y^2 \), \( B = 2aX_xY_x + b(X_xY_y + X_yY_x) + 2cX_yY_y \), \( C = aY_x^2 + bY_xY_y + cY_y^2 \), and we verify that \( B^2 - 4AC = J^2(b^2 - 4ac) \). In particular, a suitable change of variables allows us to simplify the PDE (1.40) and return it to its ‘canonical’ form. Thus, any elliptic equation can be reduced to the Laplacian \( \frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} \), any parabolic equation to the heat flow equation \( \frac{\partial}{\partial X} - \frac{\partial^2}{\partial Y^2} \), and any hyperbolic equation to the wave equation \( \frac{\partial^2}{\partial X^2} - \frac{\partial^2}{\partial Y^2} \).

Remark 1.5.7 It is well-known that the general conic equation has a number of degenerate cases when it no longer describes a cone but a set of lines. The same situation can hold with the PDE (1.40). For example, the equation \( \frac{\partial^2 u}{\partial X^2} = 1 \) with \( a = 1 \) and \( b = c = d = e = f = 0 \) is not parabolic in two dimensions (even though \( b^2 - 4ac = 0 \)) but elliptic in one dimension (the variable \( y \) plays no role here). It is therefore necessary to be careful before deciding on the type of these ‘degenerate’ equations.