Lecture notes for
Fondements des méthodes numériques (4M106)

M1 - Master mention Mathématiques et Applications

Albert Cohen
Sorbonne Université

September 19, 2022

Contents

Preface

These notes are essentially the same as those written by Hervé Le Dret followed by Pascal Frey, Edwige Godlewski, Laurent Boudin, and Sidi-Mahmoud Kaber, who were in charge of this course before me. I would like to thank them for their help in the preparation of these materials.

Recommended books covering the prerequisites of this course:

- G. Allaire et S. M. Kaber : Algèbre linéaire numérique - Cours et exercices. Ellipses 2002
- G. Allaire, S. M. Kaber : Introduction à Scilab, exercices pratiques corrigés d'algèbre linéaire, Ellipses, 2002
- J.-P. Chancelier, F. Delebecque, C. Gomez, M. Goursat, R. Nikoukhah, S. Steer : Introduction à Scilab, Springer, 2001
- P. G. Ciarlet : Introduction à l'analyse numérique matricielle et à l'optimisation. Dunod, 2006
- M. Crouzeix et A.-L. Mignot : Analyse Numérique des Equations Différentielles, Masson, 1989
- S. Delabrière et M. Postel : Méthodes d'approximation. Equations différentielles, applications Scilab, Ellipses, 2004
- J.-P. Demailly : Analyse numérique et équations différentielles, EDP Sciences, 2006
- P. Lascaux, R. Théodor : Analyse numérique matricielle appliquée à l'art de l'ingénieur, Dunod 2000
- M. Schatzman : Analyse numérique - Une approche mathématique, Dunod 2001.

For the rest of the course, you can consult:

- G. Allaire : Analyse numérique et optimisation, Editions de l'Ecole Polytechnique, Palaiseau 20
- B. Mohammadi, J.-H. Saïac, Pratique de la simulation numérique, Dunod, 2003
- Laurent Di Menza : Analyse numérique des équations aux dérivées partielles, Cassini 2009
- F. Filbet : Analyse numérique, algorithme et étude mathématique, Dunod, 2009
- F. Hubert et J. Hubbard, Calcul scientifique 2, de la théorie à la pratique, Vuibert, 2006

- H. Le Dret, B. Lucquin, Partial differential equations : modeling, analysis and numerical approximation, Birkhaüser, 2016
- B. Lucquin : Equations aux dérivées partielles et leurs approximations, Ellipses, 2004
- B. Mohammadi, J.-H. Saïac, Pratique de la simulation numérique, Dunod, 2003
- P.-A. Raviart, J.-M. Thomas : Introduction l'analyse numérique des équations aux dérivées partielles, Dunod, 2004
- L. Sainsaulieu : Calcul scientifique, Masson enseignement des mathématiques, Dunod, 2000

Chapter 1

Introduction

"Over the course of a few decades, numerical simulations have become a privileged tool of investigation in science and technology. Their goal is to reproduce by computation the behavior of a system described by a model, which very often consists of partial differential equations. These equations correspond to the mathematical translation of scientific laws. The rise of numerical simulations thus reinforces the need for the mathematical study (analysis) of these equations and of their numerical resolution."

Translated from an extract of the inaugural lecture of Pierre-Louis Lions at Collège de France (Fayard, 2003)

Original: *"En quelques décennies, les simulations numériques sont devenues un outil privilégié d'investigation dans les sciences et les technologies. Elles ont pour but de reproduire par le calcul le comportement d'un système décrit par un modèle, très souvent constitué d'équations aux dérivées partielles. Ces équations correspondent à la traduction mathématique de lois scientifiques. L'essor des simulations numériques renforce donc la nécessité de l'étude mathématique (analyse) de ces équations et de leur résolution numérique."*

1.1 Overview of the course

We will present numerical methods in a very complete fashion based on some simple generic examples. We will not cover all numerical methods (such as those for solving linear systems, optimization, probabilistic and statistical methods, ...), but rather a certain number of those that allow the computation of approximate solutions of partial differential equations (PDEs), which in itself already represents an important domain and will offer us many exciting examples. Let us begin by outlining the overall structure of the course.

We introduce the principle of a numerical method (we take as an example the finite difference method, although at a later point we will also see the finite element and finite volume methods); we describe it precisely based on a generic associated example (example: the Laplacian for the Poisson equation), with prior analysis of the example, to ensure the existence and uniqueness of the solution. For example, for finite differences, the *principle* of the method is to compute approximate values of the solution at grid points, by approximating the derivatives

by finite differences (we will, of course, describe what this means in detail later); we then write the (or a) *scheme* which allows the computation of an approximate value of the solution (for example, we define a "discrete Laplacian"), we perform the *numerical analysis*, by studying its properties (consistency, stability, convergence, error estimation); we carry out the *implementation* by writing an algorithm and a computer program which performs the numerical solution, i.e. the effective calculation of the approximate solution; the program is written in such a way that to change the application scenario, it is sufficient to change some input data - this in effect leads to a *code*. The code then allows us to verify the theoretically demonstrated properties on particular examples (test cases) whose exact solution is known (the order of convergence): in this way, the *validation* of the method is ensured. When there are several possible methods or schemes, we compare them on test cases. Broadly speaking, these steps are the "basics" of the study of numerical methods.

The codes applied in practice (industry, applied research) use numerical methods that may be more elaborate but have been developed based on simplified examples and have followed the same validation process. However, whilst the implementation on real cases follows this same approach, it is more complicated and requires additional complementary procedures (construction of the mathematical model, choice of the model variables, adjustment of the parameters, discretization of the domain and choice of the numerical model, treatment of inherent multiple scales, comparison with measurements, interpretation of results, ...). Moreover, if the problem size of a real case (the number of parameters and variables) leads to a significant increase in the number of calculations to be carried out, it becomes essential to develop efficient methods which can carry out the calculations efficiently (domain decomposition, parallel computation, mesh adaptation, model order reduction, reduced bases,...), a topic which we will not discuss.

Complex systems are most often modelled by parts (geometrical elements or components), or elements corresponding to a type of functioning, behaviour, or physical phenomena that can be differentiated,... These parts are then *coupled*; each "elementary building block" requires a validation, just as the coupling must be validated on test cases. The equations are no longer generic equations as simple as those which we are going to study, however, the very process of modelling makes them fit into broad identifiable categories since the equations translate "physical" laws (the laws of the particular field of application) of processes which are understood to a larger or lesser extend (some of them still need to be studied in depth). A very current example is that of modelling and numerical simulation applied to various problems related to the construction of the experimental reactor ITER.

The work of an applied mathematician is required in the modelling process (in relation to the discipline of application or the profession), in the translation of the observed phenomenon into equations, in the analysis of the equations and of the behavior of the solutions (with respect to time, their dependence on parameters, data,...), in the coupling, in the choice of discretization methods, in the numerical analysis of these methods and also in the implementation. It is, therefore, useful to illustrate the whole chain of this activity based on well-understood examples, before extending this approach to more complex cases, treating "industrial" cases, or considering the definition of new numerical methods. Nevertheless, the field of research and applications of numerical methods is continuously developing: modern research actively studies new contexts, new multiphysics couplings, new methods whose behaviour is robust with respect to a change of scale (e.g. limiting cases when a parameter in the equation becomes very small), the effective treatment of high-dimensional problems, the control of the propagation of

uncertainty in the data and the introduction of uncertainty in the models...

We will therefore present the main numerical methods used in the context of approximation of PDEs and provide a detailed study in the 1-dimensional case. We will assume knowledge of the classical methods of interpolation by a polynomial, methods for approximating a derivative at a point (finite differences), methods for approximating an integral (quadrature formulas), and numerical methods for the approximate solution to an ODE (ordinary differential equation), such as the Euler method. Algorithmic and implementation aspects will be treated in detail in the travaux pratiques.

1.2 Modelling with PDEs

This course aims to introduce a certain number of methods for the numerical approximation of solutions to partial differential equations (PDE). It will consist of a relatively elementary exposure to these concepts and the examples treated will be, for the most part, limited to one spatial dimension, which of course somewhat lessens the significance of the term *partial derivative*. The "real" PDE problems, which are posed in two, three (or higher) spatial dimensions, and may also involve a variation with respect to time, require mathematical concepts and techniques which are largely beyond the limits of this course. But why are we interested in the approximation of PDE solutions?

It happens that many problems in physics, mechanics and engineering sciences, can be modelled in the form of PDEs. More precisely, many physical phenomena involve quantities which are, at least approximately, solutions (of systems) of PDEs. Let us provide a few examples. If we want to model the evolution of a pollutant in a given environment (atmosphere, ocean...), the quantity we are interested in is the concentration of the pollutant as a function of position $x \in \mathbb{R}^3$ and time $t \in \mathbb{R}$. Let us denote this function by

$$(x,t) \mapsto u(x,t).$$

In order to study its evolution in time, let us consider a domain $D \subset \mathbb{R}^3$ with a regular boundary (for example a ball). The amount of pollutant contained in D at time t is given by the integral

$$\int_D u(x,t)dx.$$

In the absence of sources for pollutant in the interior of D the evolution of this quantity depends on amount of pollutant passing into and out of D. This is modelled by a *flux*, i.e. a vector-valued function

$$(x,t) \mapsto q(x,t) = (q_1(x,t), q_2(x,t), q_3(x,t)),$$

which satisfies

$$\frac{d}{dt}\Big(\int_D u(x,t)dx\Big) = -\int_{\partial D} q(x,t)\cdot n(x)ds.$$

In this *conservation law*, the right hand side is a surface integral over the boundary ∂D of D and the integrand is the "outgoing flux", i.e. the scalar product of the flux with the exterior unit normal of the domain which is denoted by $n(x)$. It is easy to see that if the flux is, for example, tangent to the boundary of the domain then none of it leaves or enters the domain i.e. the amount of pollutant inside D remains constant.

Assuming that u and q are sufficiently regular, we can apply the divergence theorem to obtain

$$\frac{d}{dt}\Big(\int_D u(x,t)dx\Big) = -\int_D \mathrm{div}(q(x,t))dx, \tag{1.1}$$

where $\mathrm{div}(q) = \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3}$. We may also take the time derivative inside the integral which yields

$$\int_D \Big(\frac{\partial}{\partial t}u(x,t) + \mathrm{div}(q(x,t))\Big)dx = 0.$$

Finally, we note that if for a continuous function v we have

$$\int_D v(y)dy = 0,$$

for any domain D, then v is identically zero (exercise), which implies that u and q are related by the following partial differential equation

$$\frac{\partial u}{\partial t} + \mathrm{div}(q) = 0.$$

We will sometimes use the short-hand notation $\partial_t u$ for $\frac{\partial u}{\partial t}$.

In order to obtain an equation involving only u we can introduce a *constitutive equation* which describes the dependence of q on u. This depends on the particular physical phenomenon studied. If the fluid is moving, denoting its velocity at a point x by $b(x) = (b_1(x), b_2(x), b_3(x))$, the flux is given by the product of the product of this velocity field and the amount of pollutant, i.e.

$$q = bu.$$

We thus arrive at an equation called the *transport* or *convection equation*:

$$\partial_t u + \mathrm{div}(bu) = 0,$$

where we recall that $\mathrm{div}(bu) = \mathrm{div}(b)u + b \cdot \nabla u$. Thus, if b is such that $\mathrm{div}(b) = 0$ which corresponds to an incompressible fluid (for example in the case of constant velocity field) the equation takes the form

$$\partial_t u + b \cdot \nabla u = 0.$$

An alternative source of flux would be the diffusion of pollutant which may appear even in the absence of fluid movement. Fick's law tells us that the flux is negatively proportional to the gradient of u, i.e.

$$q = -\sigma \nabla u,$$

where $\sigma > 0$ intuitively speaking describes that the pollutant moves from regions of high concentration to regions of lower concentration. The constant σ describes the diffusivity of the fluid. More generally, σ may vary as a function of x when this diffusivity is not uniform. We thus obtain the *diffusion equation* which is given by

$$\partial_t u - \mathrm{div}(\sigma \nabla u) = 0,$$

and which can, in the case of constant σ also be written as

$$\partial_t u - \sigma \Delta u = 0, \tag{1.2}$$

where $\Delta u = \mathrm{div}(\nabla u) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2}$ is the Laplacian.

We may, of course, just as well consider both phenomena at the same time, which yields the advection-diffusion equation

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u = 0.$$

It is important to note that in all of those equations the only unknown is u: the quantities b and σ are physical parameters which are assumed to be known.

In the case when pollutant is injected into/generated in certain regions of the fluid, we have to modify the conservation law by adding to the right hand side a term of the form $\int_D f(x,t)dx$ where the function f represents the density of generated pollutant at time t and point x. Following the same reasoning as above this yields an equation of the form

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u = f.$$

We call f a *source term*. Another effect changing the amount of pollutant can be a reaction phenomenon which can be modelled by adding to the right hand side a term of the form $-\int_D c(x,t)u(x,t)dx$ where the function c represents the rate of reaction at time t and point x. We thus arrive at a transport-reaction-diffusion equation with a source term,

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u + cu = f.$$

Of course, we could equally consider models including only some of these terms, diffusion with source term, transport-reaction, etc.

It is a fortunate feature of mathematics that sometimes the same type of equations can be used to model physical phenomena of different nature. Indeed the diffusion equation (1.2) also models the evolution of temperature when heat diffuses in a medium of thermal conductivity σ, which explains why it is often called the *heat equation*.

It is worth noting the analogies and differences with ordinary differential equations (ODEs) that describe the evolution of a scalar quantity $u(t) \in \mathbb{R}$ according to

$$u'(t) = F(t, u(t)), \quad t > 0, \quad u(0) = u_0 \in \mathbb{R}$$

where F is a given function. The aforementioned PDEs describe the time-evolution of a function $u(\cdot,t)$ which represents the distribution of pollutant (or temperature for the heat equation) at time t. Similarly to the ODE case, it is natural to introduce an initial condition

$$u(\cdot,0) = u_0,$$

i.e. $u(x,0) = u_0(x)$ for every x, where u_0 is a given function, and to consider the PDE only for $t > 0$. This is called a *Cauchy problem*.

In practice the problem is often posed on a bounded domain $\Omega \subset \mathbb{R}^d$ and it is then necessary to specify the *boundary conditions* which one imposes on u over the boundary $\partial\Omega$ of the domain. The most commonly encountered boundary conditions are the following:

1. *Dirichlet boundary conditions:* we impose that the function value is given, i.e.

$$u_{|\partial\Omega} = g,$$

where g is a given function defined on $\partial\Omega$. In the case of the heat equation this means that we specify the value of the temperature on the boundary of the domain.

2. *Neumann boundary conditions:* Denoting by $n(x)$ the exterior normal vector to the boundary at a given point $x \in \partial\Omega$, we define on $\partial\Omega$ the function $\frac{\partial u}{\partial n} = \nabla u \cdot n$ which is called the *normal derivative*, and we impose its value

$$\frac{\partial u}{\partial n} = h,$$

where h is a given function defined on $\partial\Omega$. In the case of the heat equation this means that we specify the heat flux through the boundary of the domain.

When g or h are identically zero (which corresponds to isothermal or adiabatic conditions for heat), the boundary conditions are called homogeneous.

We are also interested in equations which describe a *stationary regime*, i.e. when the solution u no longer depends on time t. We thus obtain for the function $x \mapsto u(x)$ equations of the general form

$$b \cdot \nabla u - \sigma\Delta u + cu = f,$$

with boundary conditions on the boundary of the domain of consideration. The particular case of *Laplace's equation*

$$-\Delta u = f, \tag{1.3}$$

is fundamental, both as an object of mathematical study and due to the multitude of physical phenomena that can be described using this equation:

1. Diffusion: equilibrium concentration of a pollutant u in presence of a source f.
2. Heat: equilibrium temperature in presence of a source f.
3. Mechanics: displacement of an elastic membrane under the action of a force f.
4. Electromagnetism: Electric potential generated by a charge distribution f.
5. Gravitation: gravitational potential generated by a mass distribution f.

In order to be more precise, of course, one should specify both the boundary conditions as well as the physical measurement units used.

Some problems involve the second derivative (acceleration) in time. This is the case for the *wave equation*

$$\frac{\partial^2 u}{\partial t^2} - c^2\Delta u = 0,$$

which describes, for example, the evolution of the displacement of a plastic membrane (a drum), subjected to boundary conditions at the edge, where the constant c represents the speed of

propagation. Similarly to the ODE case, involving the second derivative in time requires us to specify two initial conditions: the initial value $u(\cdot,0) = u_0$ and the initial velocity $\partial_t u(\cdot,0) = u_1$.

The order of a PDE is the largest of the orders of derivatives present in the equation: For example, the wave equation is of order two in space and time, whilst the heat equation is second order in space and first order in time, and the transport equation is of order one both in space and in time.

The equations described above are linear: u depends *linearly* on the data f. They are written in the form $Lu = f$ where the operator L is a linear partial differential operator. This translates to important properties, such as the *principle of superposition*: the solution u for a given right hand side $f = f_1 + f_2$ is the sum of the solutions u_1 and u_2 for right hand sides f_1 and f_2 respectively (we must also sum the initial and boundary conditions). We also see that the uniqueness of the solution is equivalent to the zero function being the only solution to $Lu = 0$. Of course, many PDEs encountered in physics are nonlinear which often makes their study more difficult. Nevertheless, the study of linear PDEs is a first fundamental step, both from a theoretical and a practical point of view because these equations can sometimes be used as approximate models of nonlinear equations.

The following vocabulary is often used to characterize the nature of PDEs of order lower than or equal to 2: elliptic (example: Laplace's equation, Poisson's equation), parabolic (example: heat equation), hyperbolic (example: wave equation). For two independent variables (which we denote by x, y), the operator L has the following general form

$$Lu \equiv a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + d\frac{\partial u}{\partial y} + e\frac{\partial u}{\partial x} + a_0$$

Assuming the coefficients are constant, one considers the *symbol* $P(x,y) = ax^2 + bxy + cy^2 + dx + ey + a_0$ and the sign of $b^2 - 4ac$. The equation is called elliptic if $b^2 - 4ac < 0$, parabolic if $b^2 - 4ac = 0$, and hyperbolic if $b^2 - 4ac > 0$.

A problem is said to be *well-posed* if for any given data (i.e. the source f and the initial and/or boundary conditions), it has a unique solution which depends continuously on the data (implicitly requiring that we have specified on which space of functions the operator L is defined). The three preceding categories correspond to well-posed problems, provided that the type of initial and/or boundary conditions is correctly defined.

1.3 Some explicit solutions

Throughout this course we will focus on problems with a single spatial dimension, i.e. $x \in \mathbb{R}$ which simplifies the problems we consider significantly and which allows us, in certain cases, to have access to the explicit form of solutions.

Let us begin by considering the transport equation which takes the form

$$\frac{\partial u}{\partial t} + b\frac{\partial u}{\partial x} = 0, \tag{1.4}$$

which can alternatively be writing in the following notation

$$\partial_t u + b\partial_x u = 0,$$

and which is assumed here to be posed on all of $\mathbb{R}$ with b a constant independent of x and t, and with initial conditions

$$u(x,0) = u_0(x), \quad x \in \mathbb{R}.$$

It is easy to check that for every differentiable function $x \mapsto v(x)$, the function

$$u(x,t) := v(x-bt),$$

is a solution to the equation, because we have

$$\partial_t v(x,t) = -bv'(x-bt) = -b\partial_x v(x,t).$$

The initial condition then tells us that we have to take $v = u_0$. Consequently, the solution to the overall problem is given by

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

Visually, we see in figure 1.1 that the graph of the function $x \mapsto u(x,t)$ moves at the speed $|b|$ in the direction of the sign of b which clearly corresponds to the transport of a pollutant by a flow of constant velocity.

A more complete argument, left as an exercise to the reader, would allow us to show that indeed this is the unique solution. Hint: the following linear change of variable and unknowns

$$X = x - bt, \quad T = t, \quad U(X,T) = u(x,t)$$

transforms the equation to $\frac{\partial U}{\partial T} = 0$.

An important property is that

$$\max_{x\in\mathbb{R}} u(x,t) = \max_{x\in\mathbb{R}} u_0(x),$$

i.e. the maximum in x of the solution is conserved. This property remains true even when the speed of propagation b is not constant: this is the *maximum principle* for the transport equation. We also note that, contrary to the case of the heat equation which we will see later, the regularity of the initial solution is also preserved: if u_0 is k-times differentiable, but not $(k+1)$-times differentiable, then the same applies to the functions $t \mapsto u(x,t)$ et $x \mapsto u(x,t)$.

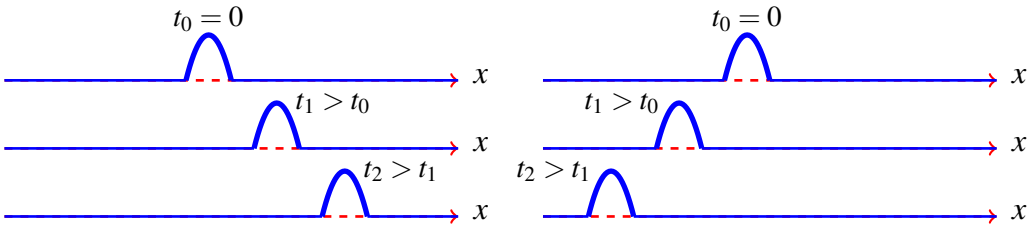

Figure 1.1: Solution of the Cauchy problem for the transport equation (1.4). Example of the evolution of an initial condition with compact support, in the case $b > 0$ (left) and $b < 0$ (right).

Since the solution to the equation (1.4) can be explicitly determined, one may rightly ask, why we should construct numerical methods for the solution to the equation. It turns out that

these methods can be generalized (more or less directly) to nonlinear transport equations of the form

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x} F(u) = 0, \quad x \in \mathbb{R}, \quad t > 0 \\ u(x,0) = u_0(x), \quad x \in \mathbb{R}, \end{cases}$$

where F is a given function, called the *flux*. In Burgers equation, cf. example 2, $F(u) = u^2/2$. In the linear case, the flux is $F(u) = bu$, and we recover the equation (1.4) if a is a constant (independent of x and t). Linear equations thus serve as a test case to validate numerical methods which are used in the more general case of nonlinear equations or even nonlinear PDE systems. Indeed, a first idea for the study of nonlinear PDE (systems) is to linearise them around a steady state, even though this approach is of course not sufficient for understanding nonlinear phenomena.

Let us now introduce a fundamental tool for the study of transport equations, in particular when the velocity b is non-constant.

Definition 1.3.1 *Consider the linear PDE*

$$\frac{\partial u}{\partial t} + b(x,t)\frac{\partial u}{\partial x} = 0, \tag{E'}$$

and denote, for a spatial point x_0, by C_{x_0} the parametrised curve $t \mapsto (X(t),t)$ where X is a solution to the following ODE

$$\begin{cases} X'(t) = b(X(t),t) \\ X(0) = x_0. \end{cases}$$

*We call C_{x_0} a characteristic curve of the PDE (*E'*) leaving from x_0 at $t = 0$.*

We will always assume that this ODE problem has a unique solution (guaranteed by the Cauchy–Lipschitz theorem).

Proposition 1.3.1 *Let $u \in \mathscr{C}^1(\mathbb{R} \times [0,\infty))$ be a solution to (*E'*). The restriction of u to a characteristic curve is constant.*

Proof. Let u be a function of class $\mathscr{C}^1$ and let $\varphi(t) = u(X(t),t)$ be its restriction to a characteristic curve. The change of $\varphi(t)$ (and thus of u along the characteristic) is

$$\frac{d\varphi}{dt}(t) = \Big\{\frac{\partial u}{\partial x}\frac{dX}{dt} + \frac{\partial u}{\partial t}\Big\}(X(t),t) = \Big\{\frac{\partial u}{\partial t} + b\frac{\partial u}{\partial x}\Big\}(X(t),t)$$

thus if u is a solution to (E'), we have $\varphi' = 0$ et $\varphi(t) = \varphi(t_0) = u(x_0,0) = u_0(x_0)$. □

Let us now return to our model equation (1.4) whose characteristics are straight lines; the characteristic passing through x_0 at time $t = 0$ has the following equation in the (x,t)-plane

$$x = x_0 + bt.$$

The point x_0 is called the "starting point/root" of the characteristic curve, cf. figure 1.2.

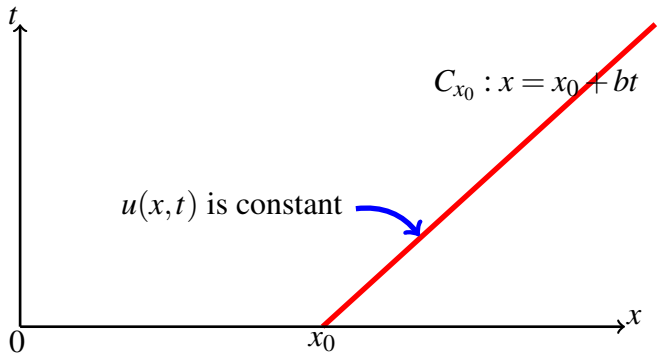

Figure 1.2: A characteristic curve of the equation (1.4).

The method of characteristics sometimes allows us to find the explicit solution to certain first order PDEs - more precisely of quasilinear (i.e. the PDE is linear in any derivatives of the unknown and possibly nonlinear in the unknown function itself) first order (i.e. involving only derivatives of order one) PDEs. Let us consider an example.

Example 1 *We consider the following linear PDE with variable coefficients*

$$\frac{\partial u}{\partial t} + x\frac{\partial u}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0. \tag{1.5}$$

It is easy to see that the characteristic curves of this PDE are the exponentials $X(t) = x_0 e^t$ (cf. figure 1.3). We thus deduce that the solution to the corresponding Cauchy problem is given by $u(x,t) = u_0(xe^{-t})$. By simply tracing the characteristics, and without computing the explicit form of the solution, we see that, for all x, $lim_{t\to+\infty} u(x,t) = u_0(0)$. Why is this?

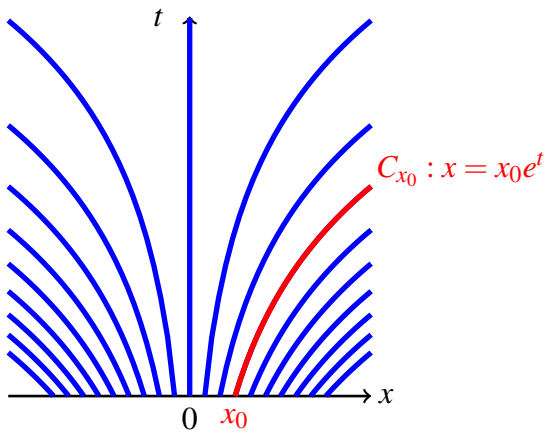

Figure 1.3: Caracteristics of the equation (1.5).

The method of characteristics can also be effectively applied to nonlinear transport equations, *i.e.* equations for which the speed of propagation depends on the solution u. In this case the equation takes the form

$$\frac{\partial u}{\partial t} + b(x,t,u)\frac{\partial u}{\partial x} = 0, \tag{E'}$$

and the differential equation of the characteristic C_{x_0} starting at the point x_0 is given by

$$\begin{cases} X'(t) = b(X(t),t,u(X(t),t)) \\ X(0) = x_0. \end{cases}$$

A similar calculation as in the linear case shows that $u(X(t),t)$ is constant along the characteristics and thus equal to $u_0(x_0)$. This shows that the equation of the characteristic can be written in the simpler form

$$\begin{cases} X'(t) = b(X(t),t,u_0(x_0)) \\ X(0) = x_0. \end{cases}$$

We can apply this method to the Burgers equation which is an exemplary model of a nonlinear transport equation.

Example 2 (Burgers equation)

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0, \\ u(x,0) = u_0(x), \quad x \in \mathbb{R}. \end{cases} \tag{1.6}$$

The characteristic curves of this equation with starting point x_0 at time $t = 0$, are defined by

$$\begin{cases} X'(t) = u(X(t),t) = u(X(0),0) = u_0(x_0) \\ X(0) = x_0. \end{cases}$$

Since u is constant along the characteristics, we have $X'(t) = u(X(0),0) = u_0(x_0)$. The characteristics are therefore straight lines given by the following equations

$$x = x_0 + u_0(x_0)t.$$

In order to determine $u(x,t)$, we note that

$$u(x,t) = u_0(x_0) = u_0(x - u(x,t)t).$$

It is therefore sufficient to solve the following equation (where the unknown function is $z = u(x,t)$)

$$z = u_0(x - zt).$$

This equation may have a unique solution (in the favourable case), have no solutions, or have multiple solutions (see figure 1.4). This situation typically arises at a critical time t^* beyond which the solution ceases to be regular and develops discontinuities called shock waves. In this last case, the characteristics method must be combined with other techniques to select the correct (physical) solution out of all possible ones. We will not deal with this difficult problem here as it is beyond the scope of this course.

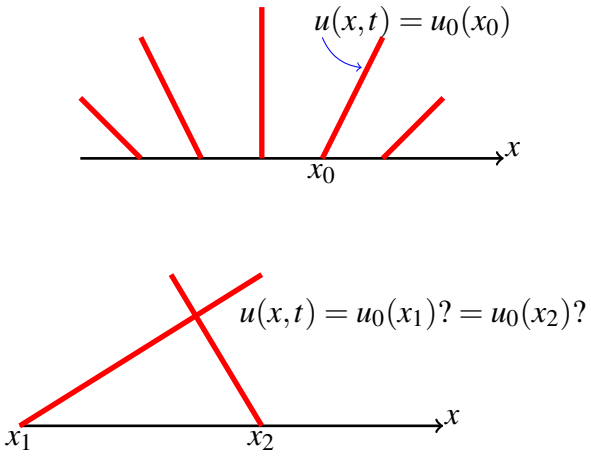

Figure 1.4: Characteristics of the Burgers equation (1.6). Upper figure: the case of increasing initial data. Through every point in the plane (x,t), passes a unique characteristic. Lower figure: the case of decreasing initial data. The characteristics starting at x_1 and x_2 cross, and the value of $u(x,t)$ is thus not uniquely determined.

This is already the end of our short nonlinear excursion. Let us return to the linear model equation (1.4). We are now interested in the problem posed on a bounded interval in space, for example the interval $[0,1]$. Suppose that $b > 0$. The characteristics of the PDE are shown in figure 1.5. For $x \in [0,1]$, we can compute $u(x,t)$ for every $t \leq x/b$ by "following/tracing" the characteristics

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

For $t > x/b$, we can no longer calculate $u(x,t)$ using the method of characteristics because the starting point of the characteristic passing through (x,t) is located outside of $[0,1]$.

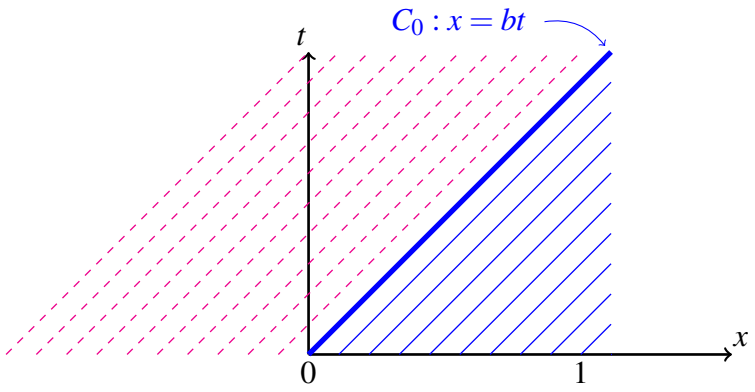

Figure 1.5: Characteristics of the transport equation (1.4) in the case $b > 0$.

With periodic boundary conditions,

$$u(x+1,t) = u(x,t),$$

it can be considered that the problem is posed on the full space. This lifts the aforementioned restriction on the computation of $u(x,t)$: we can uniquely determine the solution to the problem for all x and all times t.

Let us return to non-periodic boundary conditions, assuming again $b > 0$. Can we impose a boundary condition on the right boundary of the domain at $x = 1$? The answer is "no" because this would conflict with the fact that, for $t \leq 1/b$, the starting point x_0 of the characteristic passing through the point $(1,t)$ would be located in the interval $[0,1]$ and that would require $u(1,t) = u_0(x_0)$.

Can we impose a boundary condition on the left boundary of the domain, at $x = 0$? The answer is "yes" and we shall see that this allows us to determine $u(x,t)$ uniquely. Let us thus impose the boundary condition

$$u(0,t) = u_G(t), \quad t > 0, \tag{1.7}$$

where u_G is a given function. For $t > x/b$, we have $u(x,t) = u_G(\tau)$ where τ is determined by the intersection of the characteristic passing through (x,t) and the axis $\{x = 0\}$. We thus have

$$u(x,t) = u_G(t - x/b).$$

Conclusion: $u(x,t)$ is completely determined (cf. figure 1.6):

$$u(x,t) = \begin{cases} u_0(x-bt) & 0 \leq t \leq x/a, \\ u_G(t-x/b) & t > x/b. \end{cases}$$

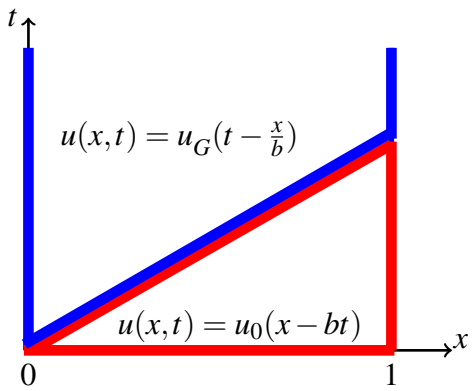

Figure 1.6: Solution of the transport equation (1.4) with the method of characteristics in the case $b > 0$. The function u_G corresponds to the boundary condition at $x = 0$. The line $y = bx$ divides the "cylinder" $[0,1] \times [0,+\infty[$ into two parts: in one of them, u is determined by the initial condition, and in the other, u is determined by the boundary condition.

Remark: for $b < 0$, we impose a boundary condition on the right boundary of the domain, at $x = 1$:

$$u(x,t) = \begin{cases} u_0(x-bt), & 0 \leq t \leq (x-1)/b \\ u_D(t+(1-x)/b), & t > (x-1)/b \end{cases}$$

Let us now consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0,$$

One can check that for every pair of twice-differentiable functions (v,w), the function

$$u(x,t) := v(x-ct) + w(x+ct),$$

is a solution on all of $\mathbb{R}$ (it is sufficient to check that each of the terms is a solution, since the equation is linear and with zero right hand side). Similarly to the transport equation the initial conditions can be used to determine v and w.

Finally, we consider the heat equation

$$\frac{\partial u}{\partial t} - \sigma \frac{\partial^2 u}{\partial x^2} = 0,$$

which we assume here is posed on all of $\mathbb{R}$ with $\sigma > 0$ a constant independent of x and t, and with initial condition

$$u(x,0) = u_0(x), \quad x \in \mathbb{R}.$$

Suppose that u_0 is continuous and uniformly bounded. The solution is then given for every $t > 0$ as the following convolution

$$u(x,t) = \int_{-\infty}^{+\infty} g_t(x-y) u_0(y) dy = (u_0 * g_t)(x)$$

where g_t is the Gaussian

$$g_t(x) := \frac{1}{\sqrt{4\pi\sigma t}} \exp\Big(-\frac{x^2}{4\sigma t}\Big).$$

In order to verify that the equation is satisfied by u as defined above, it suffices to compute the derivatives $\partial_t u$ and $\frac{\partial^2 u}{\partial x^2}$ by differentiation under the integral sign: we then find

$$\frac{\partial^2 u}{\partial x^2}(x,t) = \int_{-\infty}^{+\infty} \Big(\frac{(x-y)^2}{4\sigma^2 t^2} g_t(x-y) - \frac{1}{2\sigma t} g_t(x-y) \Big) u_0(y) dy,$$

and

$$\partial_t u(x,t) = \int_{-\infty}^{+\infty} \Big(\frac{(x-y)^2}{4\sigma t^2} g_t(x-y) - \frac{1}{2t} g_t(x-y) \Big) u_0(y) dy,$$

which shows that the equation is indeed satisfied.

In order to show that the initial condition is satisfied, let us consider the limit of the convolution as $t \to 0$. First of all, we note that g_t is an *approximation to the identity* as t tends to 0:

(i) g_t is non-negative and $\int_{\mathbb{R}} g_t = 1$.

(ii) for every $\delta > 0$ we have $\lim_{t\to 0} \int_{|x|>\delta} g_t(x) dx = 0$

In other words, the integral of $g_t$becomes increasingly concentrated around the origin as $t \to 0$. Using the assumption that u_0 is continuous and uniformly bounded over $\mathbb{R}$, this allows us to establish (exercise) that

$$\lim_{t\to 0} u_0 * g_t(x) = u_0(x), \quad x \in \mathbb{R},$$

which shows that the initial condition is indeed satisfied. Since g_t is non-negative and its integral is equal to 1, we see that for every $x \in \mathbb{R}$

$$u(x,t) \leq \max_{y\in\mathbb{R}} u_0(y) \int_{\mathbb{R}} g_t(x-y)dy = \max_{y\in\mathbb{R}} u_0(y),$$

in other words

$$\max_{x\in\mathbb{R}} u(x,t) \leq \max_{x\in\mathbb{R}} u_0(x),$$

i.e. we have arrived at the maximum principle for the heat equation. In a similar manner we obtain that $\min_{x\in\mathbb{R}} u(x,t) \geq \min_{x\in\mathbb{R}} u_0(x)$.

We note that for every $t > 0$ the function $v := u(\cdot,t) = u_0 * g_t$ is differentiable since we can differentiate the Gaussian inside the convolution integral, which gives $v' = u_0 * g_t'$. By iteration, we see that $u(\cdot,t)$ is infinitely differentiable, even if u_0 is not continuous: this is the effect of instantaneous regularisation of the heat equation.

As an exercise, the reader may seek a convolution formula which allows us to write the solution to the transport-diffusion equation

$$\frac{\partial u}{\partial t} + b\frac{\partial u}{\partial x} - \sigma\frac{\partial^2 u}{\partial x^2} = 0,$$

by trying to combine the formula for the heat equation with the translation due to transport.

For most PDEs in higher dimensions, indeed even in one dimension, explicit solutions are unfortunately not available. Consequently, we have no choice but to compute *numerical approximations* of their solutions. In this process, we distinguish the following stages:

1. Theoretical analysis: understanding if the problem is well-posed (existence and uniqueness of the solution, continuous dependence of this solution on initial conditions).

2. Development and numerical analysis: the approximation methods under consideration involve a discretization parameter, as in practice only a *finite number* of numerical values can be calculated. This parameter typically takes the form of a spatial step size δx or h (and a time step size δt in the case of an evolution equation). The goal is then to quantify the convergence of the approximation towards the exact solution as $h \to 0$. In other words, we want to show that we have indeed determined a an accurate approximation of the solution, and we want to show this in a precise sense.

3. Effective implementation of the methods on a computer: This raises some algorithmic questions, questions about the complexity and of the programming, none of which we will cover in detail in this course.

Finally, let us note that it remains important to test the numerical methods on simple test cases for which exact solutions are known: One may thus observe the exact error committed and compare this with the predicted error from convergence analysis.

1.4 Second order boundary value problems in dimension one

In the following, we will focus on problems in one spatial dimension. The analogue of Laplace's equation (1.3) with homogeneous Dirichlet boundary conditions in dimension one is as follows: we take $\Omega =]0,1[$ (the interval $]0,1[$ is usually chosen as a reference interval, the generalisation to an interval of general length is immediate), we take $f\colon]0,1[\to\mathbb{R}$ and aim to find $u\colon]0,1[\to\mathbb{R}$ such that

$$-u''(x) = f(x) \text{ on }]0,1[\quad \text{and} \quad u(0) = u(1) = 0. \tag{1.8}$$

Note that the boundary consists of two points, $\partial\Omega = \{0,1\}$. Let us consider the segment Ω as a subset of $\mathbb{R}^2$ which represents an elastic wire that is suspended between two points and where f models the vertical forces acting on the wire divided by the tension of the wire, then $u(x)$ represents the vertical displacement at the point x of the wire from its equilibrium state (cf. figure 1.7).

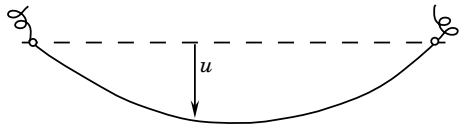

Figure 1.7: The elastic wire

It is crucial to observe that this *boundary value problem* is very different from the *Cauchy problem* for the ordinary differential equation (ODE):

$$-u''(x) = f(x) \text{ on }]0,1[\quad \text{and} \quad u(0) = \alpha, u'(0) = \beta. \tag{1.9}$$

In this initial value problem, we are given some *initial conditions*, denoted here by α, β, for u and u' at the *same point* $x = 0$. In this case, we should rather think of the evolution of a system from an initial state, the variable x being interpreted as time, $u(x)$ as the position at time x, $u'(x)$ as the velocity and $u''(x)$ as the acceleration. Consequently, $u(0)$ and $u'(0)$ are the initial position and velocity respectively.

In the boundary value problem, on the other hand, we are given the value of u at *two ends* of the interval and the variable x should be thought of as a spatial variable, as shown in the above example of an elastic wire. Due to this fundamental difference, methods for the analysis of the Cauchy problem for ODEs (ordinary differential equations) are not directly applicable to boundary value problems and methods for the numerical solution to evolution ODEs such as the forward and backward Euler method, Runge–Kutta methods, multi step methods, etc., are not suitable for boundary value problems and cannot be applied to the these problems. Indeed, for the latter we will have to sometimes implement entirely different ideas and methods.

To begin with, we will study the boundary value problem for an elliptic equation in dimension one in a classical way, i.e. in the usual function spaces $\mathcal{C}^k$, and using traditional tools, which we already know for the treatment of ODEs. This is possible in dimension one but no longer in higher dimensions. Afterwards, we will perform a similar study with methods that can also be generalised to higher dimensions. In addition to the existence and uniqueness, it is

useful to know the properties of the solution before trying to approximate it, in order to build numerical methods that can preserve a discrete analogue of these properties.

The numerical approximation of these boundary value problems involves the solution to linear systems and we shall therefore recall a few main results of linear algebra and matrix analysis.

The example (1.8) of the Poisson equation in one spatial dimension, an ODE (ordinary differential equation), leads us to recall the important Cauchy–Lipschitz theorem for more general ODEs. We recall its formulation in the case where u is a vector in $\mathbb{R}^n$, and F is Lipschitz continuous, although one can pose the problem in a more general context (Banach spaces) and with weaker hypothesis (F locally Lipschitz).

Theorem 1.4.1 *Let $I = [t_1, t_2] \subset \mathbb{R}$ be a compact interval with non-empty interior and let $F \in \mathcal{C}^0(I \times \mathbb{R}^n, \mathbb{R}^n)$ be Lipschitz continuous in its second argument:*

$$\forall t \in I, \forall u, v \in \mathbb{R}^n, \|F(t,v) - F(t,u)\| \leq L\|v-u\|.$$

Then $\forall t_0 \in I, \forall u_0 \in \mathbb{R}^n$, there is a unique function $u \in \mathcal{C}^1(I, \mathbb{R}^n)$ which satisfies

$$\begin{cases} u'(t) = F(t, u(t))) \quad \text{on }]t_0, t_1[, \\ \\ u(t_0) = u_0. \end{cases} \tag{1.10}$$

The proof uses the contraction mapping theorem.

Recall also that an ODE of order p

$$u^{(p)}(t) = F(t, u, u', \dots u^{(p-1)}), \tag{1.11}$$

with $u \in \mathcal{C}^p(I, \mathbb{R}^n)$ can be transformed to a first order ODE by introducing the vector of unknowns $U = (u, u', \dots u^{(p-1)})^T$ in the space $\mathbb{R}^{np}$. For the Cauchy problem, we must specify an initial condition on U, i.e. the p values $U(t_0) = (u, u', \dots u^{(p-1)})^T(t_0)$ of u and of its derivatives of order at most $p-1$ at t_0.

In the remainder of this section we are interested in the existence and uniqueness of solutions to boundary value problems. We will consider a slightly more general model problem than the problem of the elastic wire (1.8). The data of the problem are two continuous functions f and c on the interval $\bar{\Omega} = [0,1]$ and two reals α and β. We will try to solve the following limit problem: find $u \in \mathcal{C}^0([0,1]) \cap \mathcal{C}^2(]0,1[)$ such that

$$\text{(P)} \quad \begin{cases} -u''(x) + c(x)u(x) = f(x) \quad \text{on } \Omega =]0,1[, \\ \\ u(0) = \alpha, \quad u(1) = \beta. \end{cases}$$

Remark 1.4.1 *A quick comment on notation: u belongs to $\mathcal{C}^0([0,1]) \cap \mathcal{C}^2(]0,1[)$ means that u is a continuous function on the closed interval $[0,1]$, whose restriction to $]0,1[$ is twice continuously differentiable. The first of those conditions is essential in order to make sense of imposing non-homogeneous Dirichlet conditions such as those in the second line of problem (P). The second condition allows us to speak about the second derivative of u in the open interval, which allows us to make sense of the first line of (P). However, the equation $u'' = cu - f$ immediately implies that u'' can be extended continuously to 0 and 1 (similarly for u'): the solution to (P) is therefore automatically of class $\mathcal{C}^2([0,1])$.*

In this example, the function corresponding to the notation of (1.11) is $F(x,u,u') = c(x)u - f(x)$, thus F does not depend on the final argument and the Cauchy problem assumes that we are given the values of $u(0)$ and $u'(0)$. In the following, to study the existence of solutions of the Poisson equation in one spatial dimension, for which we will use the theorem 1.4.1 applied to a Cauchy problem (1.9). However, let us note here that we will later develop a different approach for the study of this equation since it is a *boundary value problem* and not an *initial value problem* (or Cauchy problem).

BEWARE! Unlike the Cauchy problem it is perfectly possible for this problem to have multiple solution or no solution at all. For example, if $c(x) = -\pi^2$, we see that the homogeneous equation

$$-\pi^2 u + u'' = 0,$$

together with homogeneous Dirichlet boundary conditions $u(0) = u(1) = 0$ admits a non-trivial solution $v(x) = \sin(\pi x)$. This shows that if u is a solution to (P), then so is $u + \lambda v$ for every $\lambda \in \mathbb{R}$ and thus we do not have uniqueness of solutions. Furthermore, with this choice of c and with $f(x) = 1$ and $\alpha = \beta = 0$, there is no solution to the problem (P). In order to show this, we suppose there is a solution, $u \in \mathcal{C}^0([0,1]) \cap \mathcal{C}^2(]0,1[)$, such that $-u'' - \pi^2 u = 1$ on $]0,1[$ and $u(0) = u(1) = 0$. Let us note immediately, since $u'' = -1 - \pi^2 u$, that the function u is in fact of class $\mathcal{C}^2([0,1])$ (*i.e.*, on the closed interval, and not just on the open interval), we will revisit this point in further detail later on. We can now multiply the differential equation by $\sin \pi x$, which yields

$$-u''(x)\sin \pi x - \pi^2 u(x)\sin \pi x = \sin \pi x \quad \text{on }]0,1[,$$

and then integrate the resulting equation from 0 to 1. This yields

$$-\int_0^1 u''(x)\sin \pi x\, dx - \pi^2 \int_0^1 u(x)\sin \pi x\, dx = \int_0^1 \sin \pi x\, dx = \frac{2}{\pi}.$$

On the other hand, by integration-by-parts on the first integral, which we are allowed to do because $u \in \mathcal{C}^2([0,1])$, we find that

$$-\int_0^1 u''(x)\sin \pi x\, dx = -[u'(x)\sin \pi x]_0^1 + \pi \int_0^1 u'(x)\cos \pi x\, dx.$$

Since $\sin 0 = \sin \pi = 0$, *i.e.*, the function $\sin \pi x$ satisfies the boundary conditions, the boundary term vanishes. We integrate by parts a second time to find

$$-\int_0^1 u''(x)\sin \pi x\, dx = [u(x)\cos \pi x]_0^1 + \pi^2 \int_0^1 u(x)\sin \pi x\, dx = \pi^2 \int_0^1 u(x)\sin \pi x\, dx$$

because the function u satisfies the boundary conditions. We thus obtain the equality $0 = 2/\pi$, which is a contradiction. Hence there are no solutions u to this problem.

Conclusion: in general, we do not have existence of solutions to boundary value problems for arbitrary data. Note the striking difference compared to the Cauchy problem for the ODE, which has a unique solution for any choice of continuous functions c and f.

Nevertheless, we will use the known results on the Cauchy problem to give an existence and uniqueness result for the boundary value problem in the case when c takes only non-negative values (it is possible to do even better than this).

Theorem 1.4.2 *For every $f, c \in \mathcal{C}^0([0,1])$ with $c \geq 0$ and for every $\alpha, \beta \in \mathbb{R}$, there is a unique solution to the problem (P). This solution is of class $\mathcal{C}^2([0,1])$.*

Proof. To begin with, note that if u solves (P) then necessarily, $u \in \mathcal{C}^2([0,1])$, *i.e.* the first and second derivatives of u continuously extend to the closed interval $[0,1]$. Indeed, since $u''(x) = c(x)u(x) - f(x)$ for $x \in]0,1[$ and since we assumed that c, f and u are continuous on the closed interval $[0,1]$, u'' extends by continuity to 0 and to 1, thus is bounded on $[0,1]$. Therefore, by integrating u'' from e.g. $x = 1/2$, the same applies to u'.

Let us begin by proving the *uniqueness*. We suppose thus that we have two solutions of the problem (P), u_1 and u_2. Taking $v = u_1 - u_2$, we easily see that v is a solution to the following boundary value problem:

$$\begin{cases} -v''(x) + c(x)v(x) = 0 \quad \text{on }]0,1[, \\ v(0) = v(1) = 0. \end{cases}$$

We multiply the equation by v and integrate the result over $]0,1[$ (a procedure which we will use many times in this course). It follows that

$$-\int_0^1 v''(x)v(x)\,dx + \int_0^1 c(x)v(x)^2\,dx = 0.$$

Integrating by parts the first of the two integrals we obtain that

$$-[v'v]_0^1 + \int_0^1 \left(v'(x)^2\,dx + c(x)v(x)^2\right)dx = 0,$$

because v vanishes at $x = 0$ and at $x = 1$, we have $[v'v]_0^1 = 0$. The function $v'(x)^2 + c(x)v(x)^2$ is continuous on $[0,1]$, and non-negative because c is non-negative *by assumption* and integrates to zero according to the above calculation. Thus v must be identically zero. This is a sum of two non-negative terms, thus both of those must be equal to zero. In particular $v'(x) = 0$, thus v is constant, which implies that $v(x) = v(0) = 0$ for every $x \in [0,1]$. In other words, $u_1 = u_2$ and we have uniqueness.

Remark 1.4.2 *From the above calculation, it can also be deduced that $c(x)v(x)^2 = 0$ on $[0,1]$. This does not directly imply that $v = 0$ because the hypotheses do not prevent c from cancelling over a subset of $[0,1]$. It is thus necessary to pass to v' and the boundary conditions in order to reach the desired conclusion. We can also see that if c takes strictly negative values, the same reasoning no longer works because the integrand need no longer be non-negative and thus may integrate to zero without being identically zero.*

For the *existence*, we use the *shooting method*. The Cauchy–Lipschitz theorem 1.4.1 guarantees that for every $\alpha \in \mathbb{R}$ and every $\lambda \in \mathbb{R}$, there is a unique $u_\lambda \in \mathcal{C}^2([0,1])$ which solves the Cauchy problem

$$\begin{cases} -u_\lambda''(x) + c(x)u_\lambda(x) = f(x) \quad \text{on }]0,1[, \\ u_\lambda(0) = \alpha, \quad u_\lambda'(0) = \lambda. \end{cases}$$

Indeed, we can write this ODE in the form of a system of linear first-order ODEs on $\mathbb{R}^2$

$$\begin{cases} u'_\lambda(x) = -w_\lambda(x) \\ w'_\lambda(x) = -c(x)u_\lambda(x) + f(x) \end{cases}$$

or, alternatively,

$$\begin{pmatrix} u_\lambda \\ w_\lambda \end{pmatrix}'(x) = A(x) \begin{pmatrix} u_\lambda \\ w_\lambda \end{pmatrix}(x) + \begin{pmatrix} 0 \\ f(x) \end{pmatrix}$$

with the matrix $A(x) = \begin{pmatrix} 0 & -1 \\ -c(x) & 0 \end{pmatrix}$ and the initial condition $(u_\lambda, w_\lambda)(0) = (\alpha, -\lambda)$. Let us denote $U = (u,w)^T$, $B(x) = (0, f(x))^T$ and $G(x,U) = A(x)U + B(x)$. The function G is Lipschitz with respect to the variable U

$$\|G(x,U) - G(x,V)\| \leq \|A(x)\| \, \|U - V\|$$

for every subordinate matrix norm, and, for example, if we take the ℓ^∞-norm (maximum norm), $\|A(x)\| \leq max(1, |c(x)|)$ is uniformly bounded in x because c is assumed to be continuous on the compact set $[0,1]$.

Clearly, we will have succeeded in constructing a solution to the problem (P) if we find a number λ such that $u_\lambda(1) = \beta$. Since this must be possible for any given $\beta \in \mathbb{R}$, it remains to show that the map

$$\begin{cases} g \colon \mathbb{R} \to \mathbb{R}, \\ \lambda \longmapsto u_\lambda(1), \end{cases}$$

is *surjective*. In fact, we will show that the map g is *affine*, i.e. there are two constants a and b such that $g(\lambda) = a\lambda + b$, with $a \neq 0$ for surjectivity. If this is verified, we have $g(0) = b$ i.e. $u_0(1) = b$, where the notation u_0 means that u_0 is the solution to $-u''_0 + cu_0 = f$ with $u_0(0) = \alpha$ and $u'_0(0) = 0$.

Thus take $b = u_0(1)$, and $v_\lambda(x) = u_\lambda(x) - u_0(x)$. Then we have

$$\begin{cases} -v''_\lambda(x) + c(x)v_\lambda(x) = 0 \quad \text{on }]0,1[, \\ v_\lambda(0) = 0, \quad v'_\lambda(0) = \lambda, \end{cases}$$

So if we introduce the function

$$\begin{cases} l \colon \mathbb{R} \to \mathbb{R}, \\ \lambda \longmapsto v_\lambda(1), \end{cases}$$

we clearly have

$$g(\lambda) = l(\lambda) + b$$

and it suffices to show that l is *linear* and *surjective*.

Let therefore be λ_1 and λ_2 two real numbers. The function $v_{\lambda_1} + v_{\lambda_2}$ clearly satisfies

$$\begin{cases} -(v_{\lambda_1} + v_{\lambda_2})''(x) + c(x)(v_{\lambda_1} + v_{\lambda_2})(x) = 0 \quad \text{on }]0,1[, \\ (v_{\lambda_1} + v_{\lambda_2})(0) = 0, \quad (v_{\lambda_1} + v_{\lambda_2})'(0) = \lambda_1 + \lambda_2, \end{cases}$$

i.e. the Cauchy problem corresponds to zero initial conditions for the function value and to initial condition $\lambda_1 + \lambda_2$ for its derivative. At this point the Cauchy–Lipschitz theorem ensures the *uniqueness* of the solution to this Cauchy problem, which we have previously denoted by

$v_{\lambda_1+\lambda_2}$. We thus have $v_{\lambda_1+\lambda_2}(x) = v_{\lambda_1}(x) + v_{\lambda_2}(x)$ for every $x \in [0,1]$. In similar fashion we can show that $v_{\mu\lambda_1} = \mu v_{\lambda_1}$. In particular, taking $x = 1$, we obtain

$$l(\lambda_1 + \lambda_2) = l(\lambda_1) + l(\lambda_2) \quad \text{and} \quad l(\mu\lambda_1) = \mu l(\lambda_1),$$

i.e., l is a linear map from $\mathbb{R}$ to $\mathbb{R}$.

A linear function from a finite dimensional vector space to itself is surjective if and only if it is injective. Therefore it suffices to show that l is injective, i.e. since it is a linear map, show that its kernel $\ker l = \{\lambda \in \mathbb{R}, l(\lambda) = 0\} = \{\lambda \in \mathbb{R}, v_\lambda(1) = 0\}$ is the zero space (i.e. $\dim \ker l = 0$). Now, if $\lambda \in \ker l$, it follows that

$$\begin{cases} -v_\lambda''(x) + c(x)v_\lambda(x) = 0 & \text{on }]0,1[, \\ v_\lambda(0) = v_\lambda(1) = 0. \end{cases}$$

Since $c \geq 0$, the same computation as for uniqueness, shows thus that $v_\lambda(x) = 0$, which implies of course that $v_\lambda'(x) = 0$. Now, by the definition of v_λ, we have $\lambda = v_\lambda'(0)$, and consequently we showed that $\lambda = 0$, *i.e.*, $\ker l = \{0\}$, l is injective, thus surjective. □

Remark 1.4.3 *i) The assumption $c \geq 0$ plays no role in the existence of a solution u_λ to the* Cauchy problem *by the Cauchy–Lipschitz theorem. However, it plays a crucial role in the proof of uniqueness (which here implies the existence) for the* boundary value problem. *Moreover, the example $c(x) = -\pi^2$ shows that we cannot do without imposing restrictive assumptions on c in order to be able to hope for an existence result (however, the assumption $c \geq 0$ is not optimal).*

ii) The origin of the terminology "shooting method" is clear (see figure 1.8): we start from

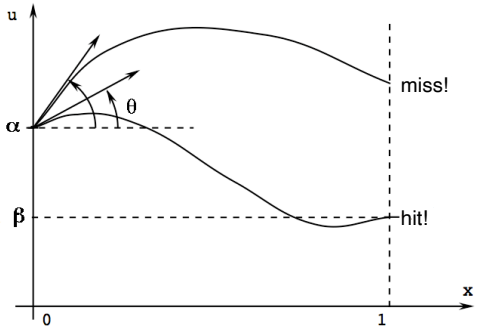

Figure 1.8: All you have to do is aim well

the value $u(0) = \alpha$ and seek to adjust the "firing angle" (denoted by θ in figure 1.8 and whose tangent is λ) in order to hit the target value β at $x = 1$. The trajectory of our fictitious projectile between the two points is governed by the differential equation of the problem (P). Unfortunately, this method is strictly mono-dimensional and there is no analogue in higher dimensions. Existence theory in this case is much more sophisticated and mostly beyond the scope of this course; we will provide just one example (in a simplified case) in the next chapter for a boundary value problem in two spatial dimensions.

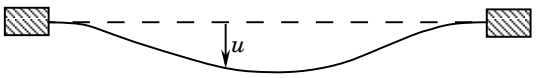

Figure 1.9: Beam fixed in bending encasing

iii) The method applies equally to more complicated problems, for example

$$\begin{cases} u''''(x) - (a(x)u'(x))' + c(x)u(x) = f(x) \quad on]0,1[, \\ u(0) = u(1) = u'(0) = u'(1) = 0, \end{cases}$$

which models the equilibrium state of an elastic beam which is fixed in bending encasing at its ends and on which we exert a force f (sketched in figure 1.9). Here, we take $a \geq 0$ and $c \geq 0$.

iv) We will show the existence and uniqueness of a solution to the boundary value problem, but, except for very special data c and f, we do not have an "analytic" expression which would allow us to numerically evaluate $u(x)$, hence the need arises to define approximate methods of solution.

We will now show that if we assume that the data f and c have more regularity, then so does the solution u of the boundary value problem (the regularity of the data propagates in a certain sense to the solution to the problem). Let us assume again that $c \geq 0$, a condition which will allow us to use theorem 1.4.2.

Theorem 1.4.3 *If f and c are of class $\mathcal{C}^m([0,1])$ with $m \in \mathbb{N}$, then u belongs to $\mathcal{C}^{m+2}([0,1])$.*

Proof. By induction on m.

For $m = 0$, the result is covered by theorem 1.4.2.

Assume thus that the result is true for $m-1$, $m \geq 1$. If f and c are of class $\mathcal{C}^m([0,1])$, then, *a fortiori*, $f, c \in \mathcal{C}^{m-1}([0,1])$. By our induction hypothesis, we thus have $u \in \mathcal{C}^{(m-1)+2}([0,1]) = \mathcal{C}^{m+1}([0,1]) \subset \mathcal{C}^m([0,1])$. Consequently, $u'' = cu - f \in \mathcal{C}^m([0,1])$ (because it is an algebra), i.e. $u \in \mathcal{C}^{m+2}([0,1])$. □

Remark 1.4.4 *This result is true and easy to prove in dimension one. It is false as such in higher dimensions in the spaces $\mathcal{C}^m(\bar{\Omega})$ (for example, there are functions u defined on an open subset Ω of $\mathbb{R}^2$ such that $\Delta u \in \mathcal{C}^0(\bar{\Omega})$ but for which $u \notin \mathcal{C}^2(\bar{\Omega})$). On the other hand, there are regularity results analogous to that for dimension 1, but these are more difficult to prove and are formulated on more complicated function spaces such as Hölder spaces $\mathcal{C}^{m,\alpha}(\bar{\Omega})$, $0 < \alpha < 1$, or Sobolev spaces $H^m(\Omega)$.*

Let us now provide a few results which can be grouped under the generic term of "maximum principle".

Theorem 1.4.4 *Suppose that $f \geq 0$ on $[0,1]$, $\alpha \geq 0$ and $\beta \geq 0$. Then the solution u of* (P) *is such that $u(x) \geq 0$ for every $x \in [0,1]$.*

Proof. Proof by contradiction. Suppose there is a point $x_0 \in [0,1]$ such that $u(x_0) < 0$. Since $u(0) = \alpha \geq 0$ and $u(1) = \beta \geq 0$, we have necessarily $x_0 \in]0,1[$ and since u is continuous, there is an interval $[a,b] \subset [0,1]$ containing x_0 such that $u(x) \leq 0$ on $[a,b]$ and $u(a) = u(b) = 0$ (this holds by the intermediate value theorem). On this interval, we thus have $u''(x) = c(x)u(x) - f(x) \leq 0$, since $c \geq 0$ by assumption. In other words, u is *concave* on this interval, i.e. $u(ta + (1-t)b) \geq tu(a) + (1-t)u(b) = 0$ for every $t \in [0,1]$. In particular, for $t = (b-x_0)/(b-a)$, which belongs to $[0,1]$ because $x_0 \in [a,b]$, it follows $u(x_0) \geq 0$, which contradicts the hypothesis $u(x_0) < 0$.□

Remark 1.4.5 *This result describes a monotone dependence of the solution u on the data f, α, β of the problem. This has a very nice physical interpretation: if we pull the elastic wire upwards (or downwards), the wire moves upwards (downwards, respectively).*

We can deduce from the previous theorem an *a priori* bound for the solution to the problem (P). Note, in the following result we impose no assumptions on the sign of α, β and f.

Theorem 1.4.5 *Suppose there is $\eta > 0$ such that $c(x) \geq \eta$ on $[0,1]$. Then the solution u of* (P) *is such that*

$$\max_{x\in[0,1]} |u(x)| \leq \max\left\{ |\alpha|, |\beta|, \eta^{-1} \max_{x\in[0,1]} |f(x)| \right\}.$$

Proof. In the following we let $\Gamma = \max\left\{ |\alpha|, |\beta|, \eta^{-1} \max_{x\in[0,1]} |f(x)| \right\}$.

i) Take $v^-(x) = u(x) - \Gamma$. Then $v^-(0) = u(0) - \Gamma = \alpha - \Gamma \leq \alpha - |\alpha| \leq 0$. Similarly, $v^-(1) = \beta - \Gamma \leq 0$. Finally,

$$-v^{-\prime\prime}(x) + c(x)v^-(x) = f(x) - \Gamma c(x) \leq f(x) - \Gamma\eta \leq f(x) - \max_{x\in[0,1]} |f(x)| \leq 0.$$

Theorem 1.4.4 thus implies that $v^-(x) \leq 0$ on $[0,1]$, i.e. $u(x) \leq \Gamma$ on $[0,1]$.

ii) Take $v^+(x) = u(x) + \Gamma$. We can see analogously that $v^+(0) \geq 0$, $v^+(1) \geq 0$ and

$$-v^{+\prime\prime}(x) + c(x)v^+(x) \geq 0.$$

Consequently, $v^+(x) \geq 0$ on $[0,1]$, i.e. $u(x) \geq -\Gamma$ on $[0,1]$. □

In the remainder of these notes we will sometimes use the notation

$$\|\varphi\|_{L^\infty} = \sup_{x\in\Omega} |\varphi(x)|,$$

for a function φ which is defined on a set Ω. The previous result can then be summarised as follows

$$\|u\|_{L^\infty} \leq \max\left\{ |\alpha|, |\beta|, \eta^{-1} \|f\|_{L^\infty} \right\}.$$

1.5 A few reminders about matrices

We recall a few results about matrices which we will use in what follows. To begin with we will, for practical reasons, often write a (column) vector V in $\mathbb{R}^n$

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

in the transpose form of a row vector

$$V = (v_1, v_2, \ldots, v_n)^T,$$

or, alternatively, $V^T = (v_1, v_2, \ldots, v_n)$.

For every natural number $n \in \mathbb{N}^*$, we denote by $M_n(\mathbb{R})$ the space of square matrices of size $n \times n$ with real coefficients. If we equip $\mathbb{R}^n$ with the canonical basis, i.e. with the n-tuple of vectors e_i, $1 \leq i \leq n$, defined by $e_1 = (1,0,\ldots,0)^T$, $e_2 = (0,1,\ldots,0)^T$, ..., $e_n = (0,0,\ldots,1)^T$, the space $M_n(\mathbb{R})$ can be identified with the space of endomorphisms on $\mathbb{R}^n$: if $A \in M_n(\mathbb{R})$ is the matrix with coefficients $a_{i,j}$, $A = (a_{i,j})$, $1 \leq i, j \leq n$, the column vectors $Ae_k = (a_{1,k}, a_{2,k}, \ldots, a_{n,k})^T$, $1 \leq k \leq n$ completely determine the corresponding endomorphism. We will implicitly use this identification throughout and denote by $\ker A$, $\operatorname{Im} A$, AV the nullspace/kernel, image and action of the endomorphism associated with the matrix A on a vector V in $\mathbb{R}^n$. The latter of those expressions is also the product of A, square matrix of size $n \times n$, with the column matrix (with n rows) whose entries are the coordinates of V in the canonical basis. This matrix product gives the column vector of coordinates of AV in the canonical basis. This kind of abuse of notation is perfectly fine for the type of applications we have in mind, as long as we have fixed a basis and, more fundamentally, as long as we know what we are doing.

The space $M_n(\mathbb{R})$ is itself an $\mathbb{R}$-vector space (of dimension n^2) under addition of matrices and usual coordinate-wise scalar multiplication. We can therefore, perfectly well, define norms on this space, i.e. maps $\|\cdot\| : M_n(\mathbb{R}) \to \mathbb{R}_+$ which have the following three properties:

$$\begin{cases} \forall A, B \in M_n(\mathbb{R}), \|A+B\| \leq \|A\| + \|B\| \text{ (triangle inequality)}, \\ \forall A \in M_n(\mathbb{R}), \forall \lambda \in \mathbb{R},\ \|\lambda A\| = |\lambda| \|A\|, \\ \|A\| = 0 \Longleftrightarrow A = 0. \end{cases}$$

In addition to the vector space structure, the space $M_n(\mathbb{R})$ has a structure which makes it a non-commutative ring. In general, a norm on $M_n(\mathbb{R})$ has no reason to be compatible with this multiplicative structure. However, those norms that do take into account the multiplicative structure are of particular interest.

Definition 1.5.1 *A norm on $M_n(\mathbb{R})$ is called a* matrix norm *if it satisfies*

$$\forall A, B \in M_n(\mathbb{R}), \qquad \|AB\| \leq \|A\| \|B\|.$$

Remark 1.5.1 *Note that there does not exist a norm which satisfies the stronger relation $\|AB\| = \|A\| \|B\|$ for all matrices (A,B). Indeed, there are pairs (A,B) such that $A \neq 0$, $B \neq 0$ but $AB = 0$ (the ring $M_n(\mathbb{R})$ admits zero divisors). This is the case if* $\operatorname{Im} B \subset \ker A$.

Following the definition of a matrix norm it remains to show that there are indeed norms which satisfy this requirement! Indeed, we shall see that there are a lot of matrix norms.

Proposition 1.5.1 (and definition) *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. The map*

$$\begin{cases} M_n(\mathbb{R}) & \longrightarrow \mathbb{R}_+ \\ A & \longmapsto \|A\| := \sup_{V\in\mathbb{R}^n\setminus\{0\}} \frac{\|AV\|}{\|V\|} \end{cases}$$

defines a matrix norm on $M_n(\mathbb{R})$ called the subordinate norm *induced by the vector norm $\|\cdot\|$.*

Proof. Let A be a given matrix. Take, for $V \in \mathbb{R}^n \setminus \{0\}$, $W = V/\|V\|$, we see that

$$\|A\| = \sup_{V\in\mathbb{R}^n\setminus\{0\}} \frac{\|AV\|}{\|V\|} = \sup_{V\in\mathbb{R}^n\setminus\{0\}} \Big\|A\Big(\frac{V}{\|V\|}\Big)\Big\| = \sup_{W\in\mathbb{R}^n,\|W\|=1} \|AW\|.$$

Because the unit ball is compact in $\mathbb{R}^n$ (we are working on a finite dimensional vector space) and because the chosen norm is a continuous function, we deduce immediately that $\|A\| < +\infty$, thus the function $A \mapsto \|A\|$ indeed maps $M_n(\mathbb{R})$ to $\mathbb{R}_+$. Let us first show that this is a norm. Let us consider a pair of matrices (A,B) and a scalar λ. For every $V \in \mathbb{R}^n \setminus \{0\}$,

$$\frac{\|(A+B)V\|}{\|V\|} = \frac{\|AV+BV\|}{\|V\|} \leq \frac{\|AV\|+\|BV\|}{\|V\|} = \frac{\|AV\|}{\|V\|} + \frac{\|BV\|}{\|V\|} \leq \|A\|+\|B\|,$$

since $\|\cdot\|$ satisfies the triangle inequality. Taking the sup on the left hand side, we obtain

$$\|A+B\| \leq \|A\|+\|B\|.$$

Similarly,

$$\|\lambda A\| = \sup_{V\in\mathbb{R}^n\setminus\{0\}} \frac{\|\lambda AV\|}{\|V\|} = \sup_{V\in\mathbb{R}^n\setminus\{0\}} |\lambda| \frac{\|AV\|}{\|V\|} = |\lambda|\,\|A\|.$$

Finally, if $\|A\| = 0$, we have $AV = 0$ for every $V \in \mathbb{R}^n$, whence $A = 0$. Thus we are working indeed with a norm on $M_n(\mathbb{R})$.

To show that it is a matrix norm, we firstly remark that for every $V \in \mathbb{R}^n$, by definition,

$$\|AV\| \leq \|A\|\,\|V\|. \tag{1.12}$$

We thus take two matrices A and B and note that for any $V \in \mathbb{R}^n \setminus \{0\}$,

$$\|ABV\| = \|A(BV)\| \leq \|A\|\,\|BV\| \leq \|A\|\,\|B\|\,\|V\|,$$

whence, by dividing by $\|V\|$,

$$\frac{\|ABV\|}{\|V\|} \leq \|A\|\,\|B\|,$$

and the result follows by taking the sup on the left hand side over all V in $\mathbb{R}^n \setminus \{0\}$. □

Proposition 1.5.2 *We have the following equivalent characterisation of the subordinate norm:*

$$\|A\| = \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu\|V\|, V \in \mathbb{R}^n\}.$$

Proof. By remark (1.12) from the preceding proof, it is clear that

$$\|A\| \geq \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu\|V\|, V \in \mathbb{R}^n\}.$$

Let thus $\mu \in \mathbb{R}_+$ such that $\|AV\| \leq \mu\|V\|$ for every $V \in \mathbb{R}^n$. We then have

$$\|A\| = \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\|AV\|}{\|V\|} \leq \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\mu\|V\|}{\|V\|} = \mu,$$

whence

$$\|A\| \leq \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu\|V\|, V \in \mathbb{R}^n\},$$

and the result follows. □

Remark 1.5.2 *i) Every subordinate matrix norm satisfies* $\|\operatorname{Id}\| = 1$.

ii) There are non-subordinate matrix norms, for example

$$\|A\| = \left(\operatorname{tr}(A^T A)\right)^{1/2} = \left(\sum_{ij} |a_{ij}|^2\right)^{1/2},$$

(exercise: proof).

iii) Sometimes the notion of the subordinate matrix norm is introduced by taking the sup over $\mathbb{C}^n \setminus \{0\}$ *instead of* $\mathbb{R}^n \setminus \{0\}$ *(thus considering the complexified space). It can happen, even for real matrices, that these two expressions do no coincide, the real sup naturally always being smaller than the complex sup. This is not the case for the examples that we will consider below.*

Proposition 1.5.3 (and definition) *For every* $p \in [1, +\infty]$*, the function* $\|V\|_p : \mathbb{R}^n \to \mathbb{R}_+$ *defined by*

$$\begin{cases} \|V\|_p = \left(\sum_{i=1}^n |v_i|^p\right)^{1/p} & \text{for } p < +\infty, \\ \|V\|_\infty = \max_{i=1,\ldots,n} |v_i| & \text{for } p = +\infty, \end{cases}$$

is a norm on $\mathbb{R}^n$ *called the "*l^p*-norm".*

We shall not prove this statement. For $p = 2$, we recover the usual Euclidean norm. We denote by $\|\cdot\|_p$ the subordinate matrix norm induced by the l^p-norm. We can more or less simply evaluate the subordinate norms for three values of p. Denote by $\rho(A)$ the *spectral radius* of the matrix A, i.e. the largest modulus of any eigenvalue of A. This is the radius of the smallest ball in $\mathbb{C}$ centred at 0 which contains all of the eigenvalues of A.

Theorem 1.5.1 *We have*

$$\begin{aligned} \|A\|_1 &= \max_j \left(\sum_i |a_{ij}|\right), \\ \|A\|_2 &= \sqrt{\rho(A^T A)}, \\ \|A\|_\infty &= \max_i \left(\sum_j |a_{ij}|\right) \end{aligned}$$

For a symmetric matrix we have $\|A\|_2 = \rho(A) = \max_{V \neq 0} \frac{\langle AV, V\rangle}{\|V\|^2}$.

Proof. Let us begin with the case $p = +\infty$. Let V be a non-zero vector. Since $(AV)_i = \sum_{j=1}^n a_{ij}v_j$, it follows that

$$\|AV\|_\infty = \max_i \Big| \sum_{j=1}^n a_{ij}v_j \Big| \leq \max_i \Big(\sum_{j=1}^n |a_{ij}||v_j| \Big)$$

by the triangle inequality. Now for any index j, $|v_j| \leq \|V\|_\infty$ by the definition of the l^∞-norm. We thus deduce that

$$\|AV\|_\infty \leq \Big(\max_i \sum_{j=1}^n |a_{ij}| \Big) \|V\|_\infty,$$

whence, due to proposition 1.5.2,

$$\|A\|_\infty \leq \max_i \sum_{j=1}^n |a_{ij}|.$$

To conclude, it suffices consequently to find a non-zero vector $W \in \mathbb{R}^n$, which is such that $\|AW\|_\infty = \big(\max_i \sum_{j=1}^n |a_{ij}|\big) \|W\|_\infty$. For this, we denote by i_0 a row-index such that

$$\max_i \sum_{j=1}^n |a_{ij}| = \sum_{j=1}^n |a_{i_0 j}|$$

(clearly there is at least one such index). We then introduce the vector W with components

$$w_j = \frac{a_{i_0 j}}{|a_{i_0 j}|} \text{ if } a_{i_0 j} \neq 0, \quad w_j = 1 \text{ if } a_{i_0 j} = 0.$$

In any possible case $|w_j| = 1$ and thus $\|W\|_\infty = 1$. In addition

$$(AW)_{i_0} = \sum_{j=1}^n a_{i_0 j} w_j = \sum_{j=1}^n |a_{i_0 j}|,$$

by construction of the w_j. We thus deduce that

$$\|AW\|_\infty \geq \sum_{j=1}^n |a_{i_0 j}| = \max_i \sum_{j=1}^n |a_{ij}|,$$

and so W is indeed of the desired form.

The case $p = 1$ can be treated analogously and is left as an exercise to the reader. Let us finally consider the case $p = 2$. We have

$$\|AV\|_2^2 = (AV)^T AV = V^T A^T AV, \text{ and } \|V\|_2^2 = V^T V.$$

The matrix $A^T A$ is symmetric and non-negative (since $\|AV\|_2^2 \geq 0$). Consequently, is is orthogonally diagonalisable and its eigenvalues λ_i are non-negative real. We sort them in increasing order, which implies that $\rho(A^T A) = \lambda_n$. Let us denote by f_i and orthogonal basis of eigenvectors of $A^T A$. Expanding V in this basis $V = \sum_{i=1}^n v_i f_i$, we see that $A^T AV = \sum_{i=1}^n \lambda_i v_i f_i$, whence

$$\|V\|_2^2 = \sum_{i=1}^n v_i^2 \quad \text{and} \quad \|AV\|_2^2 = \sum_{i=1}^n \lambda_i v_i^2.$$

On the one hand

$$\frac{\|AV\|_2^2}{\|V\|_2^2} = \frac{\sum_{i=1}^n \lambda_i v_i^2}{\sum_{i=1}^n v_i^2} \leq \frac{\lambda_n \sum_{i=1}^n v_i^2}{\sum_{i=1}^n v_i^2} = \lambda_n$$

whence $\|A\|_2 \leq \sqrt{\lambda_n}$, and on the other hand

$$\|Af_n\|_2^2 = \lambda_n,$$

whence $\|A\|_2 \geq \sqrt{\lambda_n}$, which completes the proof. □

Remark 1.5.3 *Beware of the potentially misleading notation. It is important not to confuse the subordinate norms $\|A\|_p$ and the "l^p-norms" on the vector space of matrices $M_n(\mathbb{R})$ (of dimension n^2, thus isomorphic to $\mathbb{R}^{n^2}$). Indeed, we see that $\|A\|_1 \neq \sum_{ij} |a_{ij}|$, $\|A\|_\infty \neq \max_{ij} |a_{ij}|$ and $\|A\|_2 \neq \left(\sum_{ij} |a_{ij}|^2\right)^{1/2}$...*