PDEs and numerical methods

Lecture notes

ENSIMAG 2A MMIS and Master 1 MSIAM

Eric Blayo, Fall 2021

Warning: the objective of this document is only to summarize some key notions and gather some lecture notes. It is thus far from being exhaustive nor self-sufficient.

Contents

Chapter 1

Some basic notions about PDEs

1.1 ODEs and PDEs

Definition 1.1. A **differential equation** is a relationship involving a function u and (some of) its derivatives. It is called an **ordinary differential equation (ODE)** if u depends on one single variable, or a **partial differential equation (PDE)** if u depends on several variables.

Examples

- $-u''(x) + x^2\,u'(x) - x\,u(x) = \sin x$ *is an ODE.*
- $\dfrac{\partial^2 u}{\partial x^2}(x,y,z) + \dfrac{\partial u}{\partial y}(x,y,z)\dfrac{\partial u}{\partial z}(x,y,z) = 0$ *is a PDE.*

1.2 Usual partial differential operators

Definition 1.2. Let $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. The **directional (or Gâteaux) derivative** of u at point $\mathbf{x} \in \Omega$ in direction $\mathbf{d} \in \mathbb{R}^n$ is

$$\frac{\partial u}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \to 0} \frac{u(\mathbf{x} + \alpha \mathbf{d}) - u(\mathbf{x})}{\alpha}$$

Examples

- *A partial derivative is a directional derivative in a direction belonging to the canonical basis.*
- *Let* $u(x,y) = x^2 - 2xy + y$ *and* $\mathbf{d} = (1,2)$. *The directional derivative of* u *in direction* $\mathbf{d}$ *is*

$$\frac{\partial u}{\partial \mathbf{d}}(\mathbf{x}) = -2x - 2y + 2$$

Definition 1.3. Let $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. The **gradient** of u at point $\mathbf{x}$ is

$$\text{grad } u(\mathbf{x}) = \nabla u(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial u}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \dfrac{\partial u}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Theorem 1.1. An important relation: $\dfrac{\partial u}{\partial \mathbf{d}}(\mathbf{x}) = \nabla u(\mathbf{x}).\mathbf{d}$

Definition 1.4. Let $\mathbf{u} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ denoted by $\mathbf{u}(\mathbf{x}) = \begin{pmatrix} u_1(x_1, \ldots, x_n) \\ \vdots \\ u_n(x_1, \ldots, x_n) \end{pmatrix}$.

The **divergence** of $\mathbf{u}$ is: $\text{div}\,\mathbf{u} = \displaystyle\sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i}$. It can also be denoted formally by $\nabla.\mathbf{u}$

Definition 1.5. Let $\mathbf{u} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^3$. The **curl** of $\mathbf{u}$ is defined by:

$$\text{curl}\,\mathbf{u} = \begin{pmatrix} \dfrac{\partial u_3}{\partial x_2} - \dfrac{\partial u_2}{\partial x_3} \\ \dfrac{\partial u_1}{\partial x_3} - \dfrac{\partial u_3}{\partial x_1} \\ \dfrac{\partial u_2}{\partial x_1} - \dfrac{\partial u_1}{\partial x_2} \end{pmatrix}$$

It can also be denoted formally by $\nabla \wedge \mathbf{u}$.

Definition 1.6. Let $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. The **Laplacian** of u is defined by $\Delta u = \displaystyle\sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}$

It can also be defined for $\mathbf{u} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ by $\Delta\mathbf{u} = \begin{pmatrix} \Delta u_1 \\ \vdots \\ \Delta u_n \end{pmatrix}$

Δu is sometimes denoted by $\nabla^2 u$.

1.3 Green formulas

Let Ω an open bounded subset of $\mathbb{R}^n$, with a piecewise smooth boundary $\partial\Omega$. The external normal vector to $\partial\Omega$ is denoted by $\mathbf{n}$. So called **Green formulas**[1] are actually particular cases

[1] e.g. Gauss (or Ostrogradsky, or divergence) theorem, Stokes theorem, Green-Riemann theorem...

of integration by parts.

The basic Green formula reads:

$$\int_\Omega \frac{\partial u}{\partial x_k}\, v\, d\mathbf{x} = -\int_\Omega u\, \frac{\partial v}{\partial x_k}\, d\mathbf{x} + \int_{\partial\Omega} u\, v\, (\mathbf{e}_k.\mathbf{n})\, d\mathbf{s}$$

where $\mathbf{e}_k$ is the unit vector in direction x_k, and where u and v are continuously differentiable functions on $\overline{\Omega}$.

All other formulas derive from it, like for instance:

$$\int_\Omega \Delta u\, v\, d\mathbf{x} = -\int_\Omega \nabla u\,.\nabla v\, d\mathbf{x} + \int_{\partial\Omega} \frac{\partial u}{\partial \mathbf{n}}\, v\, d\mathbf{s}$$

$$\int_\Omega u\, \mathrm{div}\, \mathbf{E}\, d\mathbf{x} = -\int_\Omega \nabla u.\mathbf{E}\, d\mathbf{x} + \int_{\partial\Omega} u\, (\mathbf{E}.\mathbf{n})\, d\mathbf{s}$$

1.4 Some definitions related to PDEs

Definition 1.7. Like for ODEs, the **order** of a PDE is the highest degree of derivation that appears in the PDE.

Definition 1.8. Like for ODEs, a PDE is **linear** iff the relation is linear w.r.t. u and its partial derivatives. The PDE is said to be **non linear** otherwise.

Definition 1.9. Like for ODEs, a PDE is **quasi linear** if each nonlinear term is actually a n^{th} derivative multiplied by a coefficient which depends only on $\mathbf{x}$, u and its derivatives up to order $n-1$.

Examples

- *The inviscid Burgers equation* $\dfrac{\partial u}{\partial t} + u\,\dfrac{\partial u}{\partial x} = f$ *is a non linear first-order PDE. It is actually a quasi linear PDE.*
- *The transport-diffusion equation* $\dfrac{\partial u}{\partial t} + \mathbf{c}\,.\nabla u - \nu\,\Delta u = f$ *is a linear second-order PDE.*

Definition 1.10. A PDE that involves the time variable is a **time-dependent** (or evolution) equation. Otherwise it is a **steady-state** (or time-independent, or stationary) equation.

Definition 1.11. PDEs are generally complemented with **boundary conditions (BCs)**, prescribed on the limits of the domain, and with an **initial condition** (generally the value of the solution at the initial time) if the PDE is time-dependent.

Some usual boundary conditions are:

Dirichlet $u = g$ on $\partial\Omega$

Neumann $\dfrac{\partial u}{\partial \mathbf{n}} = g$ on $\partial\Omega$

Robin (or Fourier) $\dfrac{\partial u}{\partial \mathbf{n}} + r\,u = g$ on $\partial\Omega$

Mixed Dirichlet-Neumann $\begin{cases} u = g \text{ on } \Gamma_0 \\ \dfrac{\partial u}{\partial \mathbf{n}} = h \text{ on } \Gamma_1 \end{cases}$ where $\Gamma_0 \cup \Gamma_1 = \partial\Omega$ and $\Gamma_0 \cap \Gamma_1 = \emptyset$

Definition 1.12. A steady-state PDE with boundary conditions is also called a **boundary value problem**.
A time-dependent PDE with initial conditions is also called an **initial value problem**, or a **Cauchy problem**.

Definition 1.13. A problem (PDE + initial and/or boundary conditions) is **well-posed** (in the sense of Hadamard) iff it has a unique solution, that continuously depends on the "parameters" of the problem (shape of the domain, coefficients in the equation, initial and/or boundary conditions...). Otherwise the problem is said to be **ill-posed**.

This continuous dependence is a crucial property for numerical simulation, since numerical solutions result from perturbations of the original problem. It is thus a necessary condition for numerical solutions to hopefully be correct approximations of the true solution.

1.5 Classification of PDEs

Many PDEs can roughly be classified into three main categories, which can generally be loosely described as follows:

- **elliptic**: time-independent, describing smooth equilibrium states
- **parabolic**: time-dependent and diffusive
- **hyperbolic**: time-dependent and wave-like, with finite speed of propagation

This classification can be made mathematically precise in particular for second-order linear PDEs. Let consider such a PDE on $\Omega \subset \mathbb{R}^n$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial u}{\partial x_i} + c(\mathbf{x})u = f \qquad (E)$$

The quadratic form corresponding to its second-order part is

$$Q_{\mathbf{x}}(X_1, \ldots, X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\mathbf{x}) X_i X_j$$

(E) is said to be:

elliptic at point $\mathbf{x}$ iff $Q_{\mathbf{x}}$ is definite (positive or negative)

parabolic at point $\mathbf{x}$ iff $Q_{\mathbf{x}}$ is positive or negative, but not definite

hyperbolic at point $\mathbf{x}$ iff $Q_{\mathbf{x}}$ is neither definite, nor positive or negative

Examples

- *The wave equation* $\dfrac{\partial^2 u}{\partial t^2} - c^2 \, \Delta u \; = 0$ *is a hyperbolic equation.*
- *The Laplace equation* $\Delta u = 0$ *is an elliptic equation.*
- *The diffusion equation* $\dfrac{\partial u}{\partial t} - \nu \Delta u = 0$ *is a parabolic equation.*
- *The equation* $x \dfrac{\partial^2 u}{\partial x^2} + y \dfrac{\partial^2 u}{\partial y^2} \; = 0$ *is elliptic for* $xy > 0$*, parabolic for* $xy = 0$*, and hyperbolic for* $xy < 0$*.*
- *The Schrödinger equation* $\dfrac{\partial u}{\partial t} - i\nu \Delta u = 0$ *does not fall in the preceding classes, due to its non real coefficient.*

We will study typical equations of each of these three categories in the following chapters.

1.6 Some tools for the analytical study of PDEs

For simple PDEs (e.g. linear, and/or with constant coefficients, and/or with a null right-hand side), it may be possible to get the analytical expression of (some of) their solutions. Some basic tools for such calculations are the following.

- **Method of characteristics** The analytical expression of the solutions of general linear first-order PDEs:

  $$\sum_{k=1}^{n} a_k(x_1, \ldots, x_n) \, \frac{\partial u}{\partial x_k}(x_1, \ldots, x_n) + r(x_1, \ldots, x_n) \, u(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$$

  can be computed by the so-called *method of characteristics* described in §5.2.

- **Fourier transform** A huge difficulty with PDEs is to deal with the different variables, while it is easy to solve linear first-order ODEs, or linear second-order ODEs with constant coefficients (see Appendix A for a reminder). Therefore a way to solve a linear PDE in $\mathbb{R}^n$ may be to take its Fourier transform with respect to all variables except one. One then obtains a linear ODE in the Fourier space with respect to the remaining variable, that can easily be solved. An inverse Fourier transform, if simple enough, then leads to the solution of the original PDE. See Appendix B for some reminders on Fourier series and Fourier transforms.

  Examples

  - *Let consider the PDE* $\dfrac{\partial u}{\partial x}(x,y) + a\dfrac{\partial^2 u}{\partial y^2}(x,y) + b\,u(x,y) = 0$ *on* $\mathbb{R}^2$. *Its Fourier transform with respect to* y *is the linear ODE* $\dfrac{d\hat{u}}{dx}(x,\xi) + (b - 4a\pi^2\xi^2)\,\hat{u}(x,\xi) = 0$. *Its solutions are* $\hat{u}(x,\xi) = \hat{u}(0,\xi)\,e^{(4a\pi^2\xi^2 - b)x}$. *Hence by inverse Fourier transform:*

$$u(x,y) = u(0,y) * FT^{-1}\left(e^{(4a\pi^2\xi^2-b)x}\right) = u(0,y) * \frac{e^{-bx}}{2\sqrt{a\pi\,x}}\,e^{\frac{y^2}{4ax}}$$

$$\text{i.e.} \qquad u(x,y) = \frac{e^{-bx}}{2\sqrt{a\pi\,x}}\int_{\mathbb{R}} u(0,z)\,e^{\frac{(y-z)^2}{4ax}}\,dz$$

    We thus have the expression of the solution as a function of u *at* $x = 0$.

  - *Let consider now the Laplace equation* $\Delta u(x_1,\dots,x_n) = 0$. *A Fourier transform with respect to* $x_2,\dots,x_n$ *leads to* $\dfrac{d^2\hat{u}}{dx_1^2}(x_1,\xi_2,\dots,\xi_n) - 4\pi^2\left(\sum_{k=2}^{n}\xi_k^2\right)\hat{u}(x_1,\xi_2,\dots,\xi_n) = 0$. *The solutions of this second-order ODE are the functions*

$$\hat{u}(x_1,\xi_2,\dots,\xi_n) = \alpha\,e^{2\pi\|\boldsymbol{\xi}\|x_1} + \beta\,e^{-2\pi\|\boldsymbol{\xi}\|x_1} \qquad \alpha,\beta\in\mathbb{R} \qquad \text{where } \|\boldsymbol{\xi}\| = \sqrt{\sum_{k=2}^{n}\xi_k^2}$$

    Hence the formal expression of the solutions:

$$u(x_1,x_2,\dots,x_n) = \alpha\,TF^{-1}\left(e^{2\pi\|\boldsymbol{\xi}\|x_1}\right) + \beta\,TF^{-1}\left(e^{-2\pi\|\boldsymbol{\xi}\|x_1}\right) \qquad \alpha,\beta\in\mathbb{R}$$

    However this inverse Fourier transform is quite complex and does not allow for a simple expression of the solutions, which shows that this approach can fail.

- **Separation of variables** Solutions of a PDE can also be searched for under the particular form of a product of several functions, each one depending on one single variable:

$$u(x_1,\dots,x_n) = u_1(x_1)\,u_2(x_2)\,\dots u_n(x_n)$$

  This approach can transform the PDE into a set of n ODEs.

Example *Let consider again the Laplace equation* $\Delta u(x_1, \ldots, x_n) = 0$. *Introducing the expression above leads to*

$$u_1''(x_1)u_2(x_2)\ldots u_n(x_n) + u_1(x_1)u_2''(x_2)\ldots u_n(x_n) + \cdots + u_1(x_1)u_2(x_2)\ldots u_n''(x_n) = 0$$

Assuming that those functions never vanish, one gets $\dfrac{u_1''(x_1)}{u_1(x_1)} + \cdots + \dfrac{u_n''(x_n)}{u_n(x_n)} = 0$ *for all values of* $x_1, x_2, \ldots, x_n$. *A simple reasoning then leads to the fact that there exists constants* $\lambda_1, \ldots, \lambda_n$ *with* $\displaystyle\sum_{k=1}^{n} \lambda_k = 0$ *such that* $\forall k = 1, \ldots, n$, $\forall x_k$, $\dfrac{u_k''(x_k)}{u_k(x_k)} = \lambda_k$. *Hence the expression of* u_k *as a linear combination of simple sinus, cosinus or exponential functions involving* λ_k. *The admissible values for the* λ_k*s are linked to the boundary conditions associated to the PDE.*

This method is used for instance in §3.2.2 and §7.2.1.
For second-order PDEs with non-constant coefficients, the determination of the functions u_k is closely linked to the Sturm-Liouville theory.

1.7 Fourier analysis of continuous or discretized PDEs

As will be seen in these notes, Fourier (or spectral) analysis is a powerful tool for studying PDEs and their approximations, and it will be frequently used in the following chapters. Its relevance can be justified from several points of view.

- **Plane-wave solutions** The basic idea is that a linear homogeneous (i.e. with a null right-hand side) PDE with constant coefficients admits plane-wave solutions of the form $u(\mathbf{x}, t) = e^{i(\mathbf{p}.\mathbf{x}+\chi t)}$, $\mathbf{p} \in \mathbb{R}^n, \chi \in \mathbb{C}$ (or $u(\mathbf{x}) = e^{i\mathbf{p}.\mathbf{x}}$ if it is a steady-state equation). In other words, one can observe that if an initial data $u_0(\mathbf{x}) = e^{i\mathbf{p}.\mathbf{x}}$ is supplied to such a time-dependent PDE, then it has a solution $u(\mathbf{x}, t) = e^{i\chi t}\, u_0(\mathbf{x})$ (where χ depends on $\mathbf{p}$), i.e. the initial condition multiplied by an oscillatory factor.

  The relationship between χ and $\mathbf{p}$ is called the **dispersion relation**. Comparing this relation with the ones corresponding to approximate equations obtained by numerical methods (such as the finite difference method that is described in these notes) is a way to assess the quality of these approximations.

  Examples *Making* $u(\mathbf{x}, t) = e^{i(\mathbf{p}.\mathbf{x}+\chi t)}$ *in the equations introduced in Section 1.5 leads to the following dispersion relations:*

  - *wave equation:* $\chi^2 = c^2 \|\mathbf{p}\|^2$
  - *diffusion equation:* $\chi = i\nu \|\mathbf{p}\|^2$
  - *Schrödinger equation:* $\chi = -\nu \|\mathbf{p}\|^2$

- **Fourier decomposition** Another point of view consists in considering that any regular enough function can be seen as the superposition of single complex exponential functions (via its inverse Fourier transform or its Fourier decomposition - see Appendix B). Therefore the effect of a linear operator on this function is the superposition of its effect on single complex exponential functions. A way to investigate some properties of linear PDEs and of their approximation schemes is thus to compute their effect on single complex exponential functions.

Fourier analysis will also lead to the notions of dissipation and dispersion errors, and numerical stability, that will be introduced later.

1.8 Dimensional and non-dimensional PDEs

When representing a physical phenomenon, a PDE is considered as a dimensional equation, which means that the function as well as the variables represent physical quantities (e.g. a concentration, a temperature or a pressure for the function, time and/or space coordinates for the variables). For instance, when dealing with the diffusion of heat in a material (see chapter 7), the PDE describes the evolution of the temperature (in $^\circ K$) with respect to space (in m) and time (in s). This means that the unit for $\partial u/\partial t$ is $K.s^{-1}$, and $K.m^{-2}$ for Δu. Those two quantities are linked in the diffusion equation through a diffusion coefficient ν (in $m^2.s^{-1}$), such that $\dfrac{\partial u}{\partial t} - \nu\,\Delta u = 0$ makes sense in terms of units.

On the contrary, from a purely mathematical point of view, one may write PDEs which would not make sense from a physical point of view, like $\dfrac{\partial u}{\partial t} - \Delta u = 0$. An implicit assumption is then that the function and all variables are non dimensional (i.e. they have no units).

Note also that any dimensional PDE can be transformed into a non dimensional one, by introducing physical scales (i.e. orders of magnitude) and performing changes of variables. Let denote for instance L, T and U scales respectively for length, time and temperature. Let introduce the new non dimensional variables $\mathbf{x}' = \mathbf{x}/L$, $t' = t/T$ and the new non dimensional function $v(\mathbf{x}', t') = u(\mathbf{x}, t)/U$. The dimensional equation

$$\frac{\partial u}{\partial t}(\mathbf{x}, t) - \nu\,\Delta u(\mathbf{x}, t) = 0$$

then reads

$$\frac{U}{T}\,\frac{\partial v}{\partial t'}(\mathbf{x}', t') - \nu\,\frac{U}{L^2}\,\Delta v(\mathbf{x}', t') = 0$$

i.e. the non dimensional PDE

$$\frac{\partial v}{\partial t'}(\mathbf{x}', t') - \nu'\,\Delta v(\mathbf{x}', t') = 0$$

with $\nu' = \dfrac{\nu T}{L^2}$ a non dimensional parameter.

Chapter 2

Introduction to finite differences

The **finite difference method** provides a numerical approximation of the solutions of ODEs and PDEs. It consists in

- defining a **mesh**, also called a **grid**, approximating the physical domain Ω where the equation is defined. For the finite difference method, contrary to the finite element method, this grid is almost always structured. This means that the **grid points** (or **nodes**) are regularly spaced (i.e. the **space step** is constant), or can be transformed into such a form by a simple function (see Figure 2.1). If one deals with a time-dependent PDE, then a mesh of the time interval is also defined (the time interval is divided into **time steps**, see Figure 2.2)
- looking for an approximation u_i^n of the exact solution at each node i and at each time step n (or more simply for an approximation u_i of the exact solution at each node i if the problem does not depend on time). This is achieved by replacing the exact equation at each node i and at each time step n by an approximate equation involving only the u_j^k, $j = \ldots, i-1, i, i+1, \ldots$, $k = \ldots, n-1, n, n+1, \ldots$ The basic tool for building this approximation is the Taylor formula.

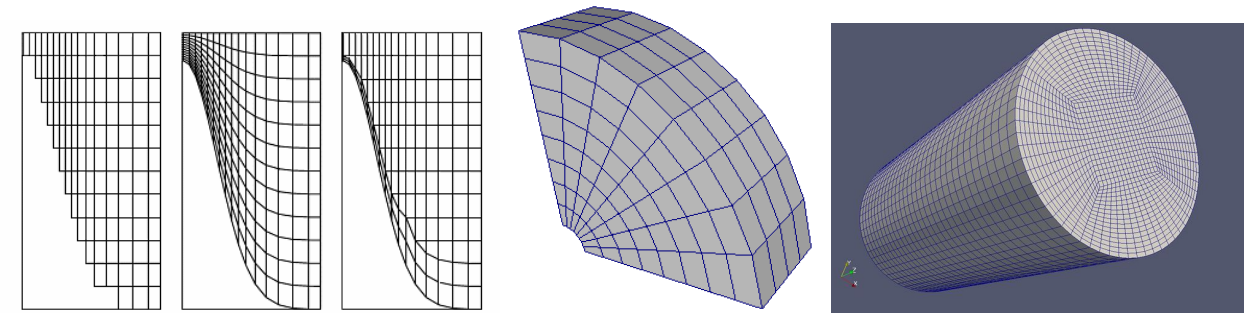

Figure 2.1: some examples of finite difference grids

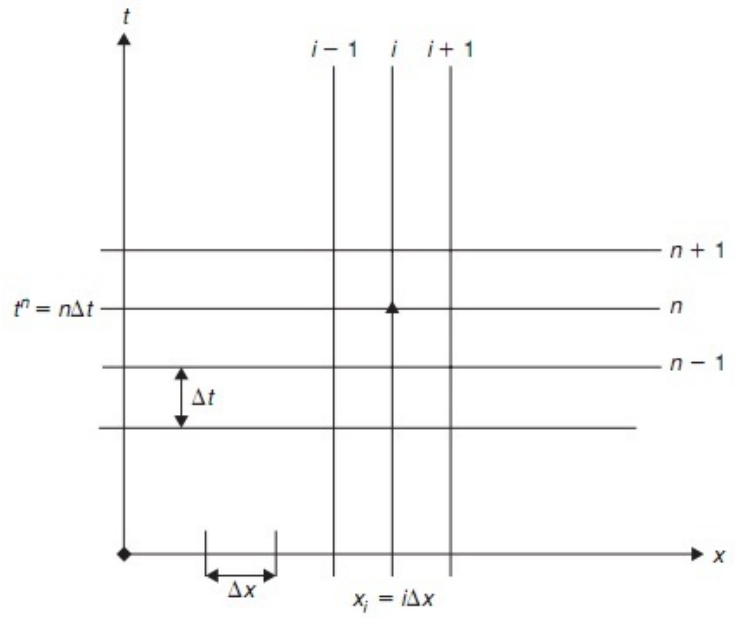

Figure 2.2: Regular space-time discretization

2.1 1-D Taylor formulas

Taylor formulas provide local polynomial approximations of regular functions. Let us recall for instance the Taylor-Lagrange formula for a function of one variable.

Theorem 2.1. Let u with $\mathcal{C}^n$ regularity on $[a, b]$, and with a $(n+1)^{\text{th}}$ derivative on (a, b). Then there exists $\zeta \in (a, b)$ such that

$$u(b) = u(a) + (b-a)u'(a) + \cdots + \frac{(b-a)^n}{n!}\, u^{(n)}(a) + \frac{(b-a)^{n+1}}{(n+1)!}\, u^{(n+1)}(\zeta) \tag{2.1}$$

This can be rewritten in the following way:

Theorem 2.2. Let u with $\mathcal{C}^n$ regularity in a neighborhood of x, and with a $(n+1)^{\text{th}}$ derivative in this neighborhood. Then, for h sufficiently small:

$$u(x+h) = u(x) + h\,u'(x) + \cdots + \frac{h^n}{n!}\, u^{(n)}(x) + \mathcal{O}\left(h^{n+1}\right)$$

This leads to the well-known **Taylor polynomials** approximating common functions. The exact meaning of the notation $\mathcal{O}$ is given in Appendix D.2.

This 1-D formula is sufficient in most cases to derive finite difference schemes, even for multi-dimensional problems. However, for some specific schemes, multidimensional Taylor expansion might be necessary. An example in the 2-D case is given in §2.6.

2.2 Finite difference schemes

In this section, we will explain how finite difference schemes are built, and introduce usual schemes for the approximation of first- and second-order derivatives. Then we will introduce a way to analyze finite difference schemes.
We will only deal with 1-D functions, and will consider that functions are regular enough, so that their high order derivatives exist and Taylor expansions make sense.

2.2.1 Approximation schemes

Let $x_0, x_1, \ldots, x_q$ distinct points and $p \in \mathbb{N}$. If we find real coefficients $\alpha_j \, (j = 0, \ldots, q)$ such that

$$u^{(p)}(x_0) \simeq \sum_{j=0}^{q} \alpha_j \, u(x_j)$$

then this linear combination is called an **approximation scheme**, or a **finite difference scheme**, for $u^{(p)}(x_0)$.

This scheme is said to be **consistent** iff $u^{(p)}(x_0) - \sum_{j=0}^{q} \alpha_j u(x_j) \longrightarrow 0$ as $h \longrightarrow 0$, h being a common order of magnitude for all the $|x_j - x_0|$.

The scheme is said to be **k^{th}-order accurate** iff $u^{(p)}(x_0) = \sum_{j=0}^{q} \alpha_j \, u(x_j) + \mathcal{O}(h^k)$. k is the **order of accuracy** of the scheme.

The grid points involved in a finite difference scheme form its so-called associated **stencil**.

2.2.2 A general method for deriving a finite difference scheme

Let $x_0, x_1, \ldots, x_q$ distinct points, and $h_j = x_j - x_0 \;\; (j = 1, \ldots, q)$. We intend to build a consistent approximation scheme for $u^{(p)}(x_0)$, for $p \leq q$.

The Taylor-Lagrange formula applied to u at point $x_j \;\; (j = 1, \ldots, q)$ at order q reads

$$u(x_j) = u(x_0) + h_j u'(x_0) + \cdots + \frac{h_j^q}{q!} \, u^{(q)}(x_0) + \mathcal{O}(h_j^{q+1})$$

Let build an arbitrary linear combination of these expansions:

$$\sum_{j=1}^{q} \alpha_j u(x_j) = \left(\sum_{j=1}^{q} \alpha_j \right) u(x_0) + \left(\sum_{j=1}^{q} \alpha_j h_j \right) u'(x_0) + \cdots + \frac{1}{q!} \left(\sum_{j=1}^{q} \alpha_j h_j^q \right) u^{(q)}(x_0) + \left(\sum_{j=1}^{q} \alpha_j \mathcal{O}(h_j^{q+1}) \right) \tag{2.2}$$

This combination is an approximation scheme for $u^{(p)}(x_0)$ as soon as the coefficients of $u^{(k)}(x_0)$ vanish for $k = 1, \dots, q,\ k \neq p$:

$$\left\{\begin{array}{rcl} \displaystyle\sum_{j=1}^{q} \alpha_j h_j & = & 0 \\ & \vdots & \\ \displaystyle\sum_{j=1}^{q} \alpha_j h_j^{p-1} & = & 0 \\ \displaystyle\sum_{j=1}^{q} \alpha_j h_j^{p} & = & p! \\ \displaystyle\sum_{j=1}^{q} \alpha_j h_j^{p+1} & = & 0 \\ & \vdots & \\ \displaystyle\sum_{j=1}^{q} \alpha_j h_j^{q} & = & 0 \end{array}\right. \quad \text{i.e.} \quad \begin{pmatrix} h_1 & h_2 & \cdots & h_q \\ \vdots & & & \vdots \\ h_1^{p-1} & h_2^{p-1} & \cdots & h_q^{p-1} \\ h_1^{p} & h_2^{p} & \cdots & h_q^{p} \\ h_1^{p+1} & h_2^{p+1} & \cdots & h_q^{p+1} \\ \vdots & & & \vdots \\ h_1^{q} & h_2^{q} & \cdots & h_q^{q} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{p-1} \\ \alpha_p \\ \alpha_{p+1} \\ \vdots \\ \alpha_q \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ p! \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.3}$$

This is a $q \times q$ linear system of Vandermonde type. Thus it has a unique solution $\alpha_1, \dots, \alpha_q$ iff the h_js are q distinct values, which is obviously the case since the x_js are distinct points. Moreover the only non homogeneous equation in this system implies that $\alpha_j = \mathcal{O}(h^{-p})$, $j = 1, \dots, q$, where h is a common order of magnitude for the h_js (for instance h can be taken as the minimum, mean, or maximum value of the h_js).
The linear system (2.3) being satisfied, (2.2) becomes

$$\sum_{j=1}^{q} \alpha_j u(x_j) = \left(\sum_{j=1}^{q} \alpha_j\right) u(x_0) + u^{(p)}(x_0) + \mathcal{O}(h^{q+1-p})$$

i.e.

$$u^{(p)}(x_0) = \sum_{j=1}^{q} \alpha_j u(x_j) - \left(\sum_{j=1}^{q} \alpha_j\right) u(x_0) + \mathcal{O}(h^{q+1-p}) \tag{2.4}$$

Theorem 2.3. Using q additional grid points $x_1, \dots, x_q$, one can build an approximation of $u^{(p)}(x_0)$ at order $q + 1 - p$. Moreover (2.2) proves that if p is even and if the scheme is symmetric, this order of the approximation becomes $q + 2 - p$.

Theorem 2.4. A direct consequence of the preceding result is that a k^{th}-order scheme is exact for any polynomial function u which degree is lower than or equal to k.

This is simply due to the fact that, following (2.1), the error term $\mathcal{O}(h^{q+1-p})$ in (2.4) is a combination of $(q+1)^{\text{th}}$-order derivatives of u. If the scheme is k^{th}-order accurate, then $q+1-p = k$, i.e. $q + 1 = k + p > k$. Given the fact that, for a polynomial of degree k, the derivatives of order greater than k are zero, the error term is thus also zero.

2.2.3 Usual schemes for the first- and second-order derivatives

Using the Taylor-Lagrange formula at points $x+h$ and $x-h$ with a positive increment h leads to the following usual schemes for the first-order derivative:

- $u'(x) = \dfrac{u(x+h)-u(x)}{h} + \mathcal{O}(h)$: first-order **right-sided** (or **downstream**) scheme
- $u'(x) = \dfrac{u(x)-u(x-h)}{h} + \mathcal{O}(h)$: first-order **left-sided** (or **upwind**) scheme
- $u'(x) = \dfrac{u(x+h)-u(x-h)}{2h} + \mathcal{O}(h^2)$: second-order **centered** scheme

The two first-order schemes are said to be **one-sided**, since they involve grid points only on one side of the current grid point of interest x, while the second-order scheme is said to be **two-sided**.

Similarly, the most usual scheme for the second derivative is the second-order centered scheme:

$$u''(x) = \frac{u(x-h)-2u(x)+u(x+h)}{h^2} + \mathcal{O}(h^2) \tag{2.5}$$

We can see that the orders of accuracy of these schemes follow the rule described by Theorem 2.3. The order is indeed equal to $q+1-p$ for the schemes approximating u' ($p=1$, $q=1$ for the one-sided schemes and $q=2$ for the two-sided scheme), and $q+2-p$ for the second-order centered scheme for u'' ($p=q=2$).

2.2.4 Fourier analysis of finite difference schemes

As mentioned in §1.7, Fourier analysis is a powerful tool to study the properties and the quality of approximation schemes. It consists in comparing the effect of a numerical scheme to the effect of the exact continuous operator in the frequency space.
Since any regular function can be written as an integral or a series of complex exponential functions (see Appendix B), we only need to consider the effect on a generic complex exponential function $u_\omega(x) = e^{i\omega x}$, $\omega \in \mathbb{R}$.
Let then define the **transfer function** T of an operator S as $S(u_\omega) = T(\omega)\, u_\omega$. Such a transfer function exists for all derivation or integration operators, since u_ω are eigenfunctions of those operators. And it also exists for all linear finite difference schemes, since $u_\omega(x+h) = e^{i\omega(x+h)} = e^{i\omega h}e^{i\omega x} = e^{i\omega h}\, u_\omega(x)$. For normalization purpose, it is defined taking $h=1$, and $\omega \in [0,\pi]$ (to fulfill the Nyquist-Shannon criterion[1]).

Example *Let consider the derivation operator:*

$$S : u \longrightarrow u'$$

[1] The Nyquist-Shannon criterion states that a sufficient condition for a sample to capture all the information from a continuous signal is that the sample rate is larger than, or equal to, twice the maximum frequency contained in the continuous signal. On a regular grid of step h, the sampling rate is equal to $1/h$, i.e. 1 if we take $h=1$ for the sake of normalization. The continuous signal $e^{i\omega x}$ is made of one single frequency $\omega/(2\pi)$. Then the Nyquist-Shannon criterion reads: $1 \geq 2\omega/(2\pi)$, i.e. $\omega \leq \pi$.

We have obviously $S(u_\omega)(x) = (e^{i\omega x})' = i\omega\, e^{i\omega x}$, *i.e.* $S(u_\omega) = i\omega\, u_\omega$. *The transfer function corresponding to the first-order derivation is thus* $T(\omega) = i\omega$. *Moreover, since* $i\omega\, e^{i\omega x} = \omega\, e^{i(\omega x + \pi/2)}$, *the derivation changes the amplitude of* u_ω *(by a factor of* ω*) and its phase (by adding a* $\pi/2$ *delay).*
Let now consider the two following finite difference schemes approximating the first derivative:

$$S_h^{(1)} : u \longrightarrow \frac{u(.+h) - u(.)}{h} \qquad \textit{and} \qquad S_h^{(2)} : u \longrightarrow \frac{u(.+h) - u(.-h)}{2h}$$

$S_1^{(1)}(u_\omega) = e^{i\omega(x+1)} - e^{i\omega x}$ *and* $S_1^{(2)}(u_\omega) = \frac{1}{2}\left(e^{i\omega(x+1)} - e^{i\omega(x-1)}\right)$, *which implies that their transfer functions are respectively* $T^{(1)}(\omega) = e^{i\omega} - 1$ *and* $T^{(2)}(\omega) = i\sin\omega$. *They are compared with the transfer function* $T(\omega) = i\omega$ *of the exact continuous derivation in Figure 2.3.*

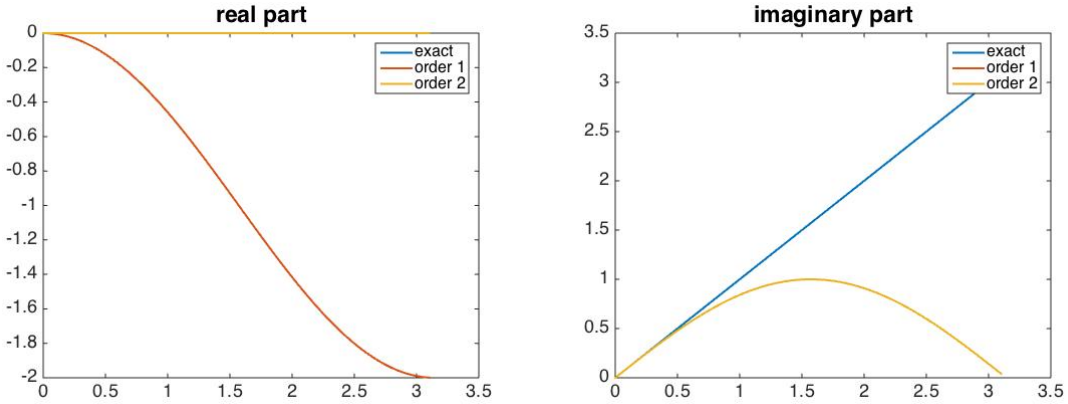

Figure 2.3: Real (left panel) and imaginary (right panel) parts of the transfer function for the exact derivation operator (blue curves), and the $S_h^{(1)}$ (red curves) and $S_h^{(2)}$ (yellow curves) finite difference schemes

An important aspect contributing to the quality of a finite difference scheme is its ability to modify as few as possible the exact transfer function, nor its phase neither its amplitude. In the present example, some simple algebra leads to

$$T^{(1)}(\omega) = \mathrm{sinc}\left(\frac{\omega}{2}\right) e^{i\omega/2}\, T(\omega) \qquad \textit{and} \qquad T^{(2)}(\omega) = \mathrm{sinc}\,(\omega)\; T(\omega)$$

where $\mathrm{sinc}(\omega) = \dfrac{\sin\omega}{\omega}$ *is the cardinal sine function (see Figure B.1 in Appendix B). Under this form, it is clear that* $S^{(1)}$ *modifies both the amplitude and the phase w.r.t.* S, *while* $S^{(2)}$ *modifies the amplitude but not the phase.*

Definition 2.1. A scheme S_h that modifies the amplitude of Fourier components w.r.t. to the exact operator S is said to be **diffusive** or **dissipative**. The modification of this amplitude is called the **diffusion error** or **dissipation error** of the scheme.

Definition 2.2. A scheme S_h that modifies the phase of Fourier components w.r.t. to the exact operator S is said to be **dispersive**. The modification of this phase is called the **dispersion error** of the scheme.

Both errors appear clearly by looking at the ratio of their transfer functions: there is a dissipation error[2] as soon as the amplitude of this ratio is not equal to 1, and a dispersion error as soon as its phase is non zero (i.e. the ratio is not a real number).
As can be seen in the above example, these errors are not constant, but generally depend on the wavenumber ω. For most schemes, the errors are small for small wavenumbers, and increase for larger ones.

Some generic calculations facilitating the computation of transfer functions are given in Appendix D.1.

2.3 The finite difference method: a simple example

Let now illustrate the principle of the finite difference method on the very simple example of the ODE: $-u''(x) = f(x)$ for $x \in (a, b)$, with boundary conditions $u(a) = 0$ and $u(b) = 0$.

The finite difference method consists in:

- building a mesh of the domain $[a, b]$. Let take here for instance the regular mesh defined by $x_i = a + ih$ $(i = 0, \dots, N + 1)$ with $h = (b - a)/(N + 1)$.
- considering the ODE on grid points only. The original ODE on (a, b) is replaced by

$$\begin{cases} -u''(x_i) = f(x_i) \quad i = 1, \dots, N \\ u(x_0) = u(x_{N+1}) = 0 \end{cases}$$

- replacing the differential operator by a finite difference scheme. Here, we can use the second-order scheme (2.5) seen previously, and the ODE at point x_i reads

$$\begin{cases} -\dfrac{1}{h^2}\left(u(x_{i-1}) - 2u(x_i) + u(x_{i+1})\right) + \varepsilon_i = f(x_i) \quad , i = 1, \dots, N \quad \text{with } \varepsilon_i = \mathcal{O}(h^2) \\ u(x_0) = u(x_{N+1}) = 0 \end{cases} \tag{2.6}$$

- neglecting the error terms ε_i, and then actually solving the remaining system. In the present case, it is thus the simple linear system:

$$\begin{cases} -u_{i-1} + 2u_i - u_{i+1} = h^2 f(x_i) \quad , i = 1, \dots, N \\ u_0 = u_{N+1} = 0 \end{cases} \tag{2.7}$$

  where u_i is the approximation of $u(x_i)$.

[2] Note that one speaks about "dissipation error" even if the amplitude of T_h/T is greater than 1.

The preceding problem can of course be written in matrix form. Let

$$A_h = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix}, \quad F = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix}, \quad E = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

$$U = \begin{pmatrix} u(x_1) \\ \vdots \\ u(x_N) \end{pmatrix}, \quad \text{and} \quad U_h = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}$$

Then (2.6) reads: $A_h\, U + E = F$, while (2.7) reads:

$$A_h\, U_h = F \tag{2.8}$$

Before addressing the issues of the existence and uniqueness of U_h and of its convergence towards U, let present some numerical results, for the particular case where the exact solution is $u(x) = e^{-x/8}\,\sin x$ for $x \in [0, 6\pi]$ (the right-hand side is then $f(x) = \frac{-1}{64} e^{-x/8}\,(63 \sin x + 16 \cos x)$). Figure 2.4 compares this exact solution u with the numerical approximation u_h for several values of h. It clearly illustrates the convergence of the finite difference solution towards the true solution as h decreases. The rate of this convergence can be quantified by computing the norm of the error $\| U_h - U\|$ for the different values of h. This quantity is displayed in Figure 2.5 in log-log scale. As can be seen, the error decreases almost linearly, with a slope which is very close to 2. This means that the error behaves like $C\, h^2$, which is coherent with the second-order discretization of the numerical scheme.

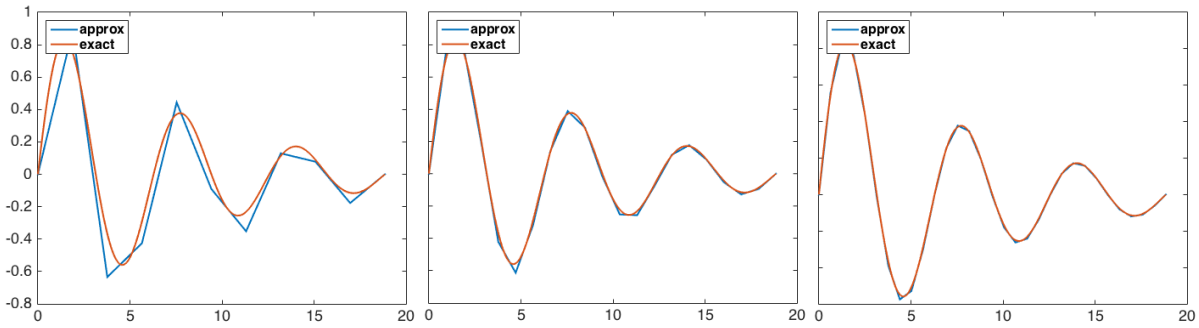

Figure 2.4: $u(x)$ (red curve) and $u_h(x)$ (blue curve) for $N = 10$, $N = 20$ and $N = 30$

2.4 Properties of the scheme and of the numerical solution

2.4.1 Consistence, stability and convergence

As seen before, when applied to an ODE or a PDE, the finite difference method leads (at least for linear equations) to a linear system $A_h U_h = F$, while the exact equations are $A_h U + E_h = F$. At this stage, two questions arise naturally:

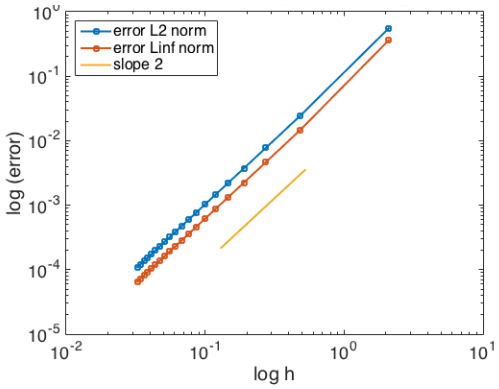

Figure 2.5: $\| U_h - U\|_2$ (blue curve) and $\| U_h - U\|_\infty$ (red curve) compared to the theoretical h^2 slope (yellow line)

- Is the problem well posed, i.e. does the system have a unique solution U_h ?
- If the problem is well posed, does its numerical solution U_h converge to the exact continuous solution U as h tends to 0 ?

Definition 2.3. The scheme is said to be **consistent** iff $E_h \longrightarrow 0$ as $h \longrightarrow 0$.

Definition 2.4. The error of the numerical solution is $\|U_h - U\| = \|A_h^{-1} E_h\| \leq \|A_h^{-1}\| \, \|E_h\|$. The fact that $\|A_h^{-1}\|$ is bounded independently of h is called **stability**.

Theorem 2.5. Stability and consistency lead obviously to **convergence**: $U_h \longrightarrow U$ as $h \longrightarrow 0$. This is even an equivalence for simple linear problems.

Example *In the preceding example, A_h is invertible, since it is a symmetric positive definite matrix. Therefore (2.8) is well posed.*
Given its expression, E_h obviously tends to 0 as h tends to 0: the numerical scheme is consistent. To get the convergence of the approximation method, we have thus to show that $\|A_h^{-1}\|$ is bounded independently of h. Note that this is not obvious at all, since A_h is a $N \times N$ matrix, with N tending to infinity as h tends to 0. In the present case, it can be shown for instance that $\|A_h^{-1}\| \leq \max\left(1, \dfrac{2(b-a)^2}{\pi^2}\right)$ for h sufficiently small.

2.4.2 Equivalent equation

Associated to convergence properties is the notion of **equivalent equation**. As a matter of fact, using Taylor expansion, the numerical scheme can be reformulated as a series (w.r.t. h), the first term of which (i.e. corresponding to h^0) is the original equation (iff the discretization is consistent). This expression is the so-called equivalent equation associated to the numerical scheme. However the most interesting term in the error is the one corresponding to the lowest power of h, also called **dominant error term**, which may give an indication on the way the

numerical scheme modifies the true solution. That is why, by extension, the term **equivalent equation** is also frequently employed to indicate the original equation with only the dominant error term in addition.

Example *Coming back to the preceding example, we have:*

$$\frac{u(x+h)-2u(x)+u(x-h)}{h^2} = u''(x) + \frac{h^2}{12}u^{(4)}(x) + \frac{h^4}{360}u^{(6)}(x) + \dots$$

Therefore the finite difference method actually solves the equivalent equation

$$-u''(x) - \frac{h^2}{12}u^{(4)}(x) - \frac{h^4}{360}u^{(6)}(x) - \dots = f(x)$$

The dominant error term is: $-\frac{h^2}{12}u^{(4)}(x)$*, and the equation with this additional term only*

$$-u''(x) - \frac{h^2}{12}u^{(4)}(x) = f(x)$$

is also called the equivalent equation.

Some generic calculations facilitating the computation of equivalent PDEs are given in Appendix D.3 and a general result allowing for the interpretation of its dominant error term is given in Appendix D.4.

2.4.3 Other properties

Additional issues may also be of interest. For instance, it might be important that U_h also satisfies some specific mathematical or physical properties satisfied by U (e.g. conservation laws, symmetry, positivity...). Such properties are called **mimetic properties** of the numerical scheme.

Example *Let consider the homogeneous ODE* $-u''(x) = 0$ *on* (a,b)*, with Dirichlet boundary conditions* $u(a) = \alpha$ *and* $u(b) = \beta$*. The exact solution is obviously* $u(x) = \alpha + \frac{x-a}{b-a}(\beta - \alpha)$*. This function satisfies a so-called "maximum principle" (we will come back on this notion in the following chapters) in the sense that the extrema of the function are reached only on the boundary of the domain (for* $x = a$ *and* $x = b$*).*

Following the discretization used previously, the corresponding finite difference solution satisfies:

$$\begin{cases} -u_{i-1} + 2u_i - u_{i+1} = 0 \quad , i = 1, \dots, N \\ u_0 = \alpha \ , \ u_{N+1} = \beta \end{cases}$$

Remarking that $u_i = \frac{1}{2}(u_{i-1} + u_{i+1})$*, a simple proof by contradiction shows that the approximate solution also reaches its extrema for* $x = a$ *and* $x = b$*. This is a mimetic property of the numerical scheme.*

2.5 Considering boundary conditions

Until this point, we did not detail the management of boundary conditions in the finite difference method. This aspect was also hidden in the example of §2.3, due to the fact that we had homogeneous Dirichlet conditions $u(a) = u(b) = 0$.

The way boundary conditions must be accounted for depends on each particular case. One must check that the finite difference schemes are still valid in the vicinity of the boundary and, if it is not the case, locally use other schemes. Moreover, boundary conditions must be integrated in the system of discretized equations.

2.5.1 Validity of finite difference schemes near boundaries

Numerical schemes often cannot be used in the vicinity of the boundary. As an example, with the same notations as in section 2.3, if one approximates $u''(x)$ by

$$u''(x_i) = \frac{-u_{i-2} + 16\,u_{i-1} - 30\,u_i + 16\,u_{i+1} - u_{i+2}}{12\,h^2} + \mathcal{O}(h^4)$$

this scheme cannot be used for grid points x_1 and x_N, since x_{-1} and x_{N+2} do not exist. Other schemes must be considered, for instance the usual centered scheme:

$$u''(x_1) \simeq \frac{u_0 - 2u_1 + u_2}{h^2} \qquad \text{and} \qquad u''(x_N) \simeq \frac{u_{N-1} - 2u_N + u_{N+1}}{h^2}$$

However, as seen before, this scheme is only second-order accurate. Discretization errors will then be larger at these two points than elsewhere, which may corrupt the overall quality of the numerical solution. To avoid this, another possibility consists in using one-sided fourth-order schemes, the price to pay being a larger stencil.

2.5.2 Dirichlet conditions

Non homogeneous Dirichlet boundary conditions can be easily integrated within numerical schemes. For instance, if one replaces the homogeneous Dirichlet conditions $u(a) = u(b) = 0$ by non homogeneous ones $u(a) = \alpha$ and $u(b) = \beta$ in the example of §2.3, system (2.8) becomes

$$\frac{1}{h^2}\begin{pmatrix} 1 & 0 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & 0 & 1 \end{pmatrix}\begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \\ u_{N+1} \end{pmatrix} = \begin{pmatrix} \alpha/h^2 \\ f(x_1) \\ \vdots \\ f(x_N) \\ \beta/h^2 \end{pmatrix}$$

or, if directly eliminating u_0 and u_{N+1}:

$$\frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} f(x_1) + \dfrac{\alpha}{h^2} \\ \vdots \\ \\ \vdots \\ f(x_N) + \dfrac{\beta}{h^2} \end{pmatrix}$$

2.5.3 Neumann conditions

In case of a Neumann boundary condition, the derivative must of course also be approximated by a finite difference scheme. For instance, if one replaces the Dirichlet condition $u(a) = \alpha$ by the Neumann condition $u'(a) = \alpha$ in the previous example, one can use the first-order approximation:

$$\frac{u_1 - u_0}{h} = \alpha$$

which leads to the system

$$\frac{1}{h^2}\begin{pmatrix} -h & h & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & 0 & 1 \end{pmatrix}\begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \\ u_{N+1} \end{pmatrix} = \begin{pmatrix} \alpha \\ f(x_1) \\ \vdots \\ f(x_N) \\ \beta \end{pmatrix}$$

One could also use the second-order scheme

$$\frac{-3u_0 + 4u_1 - u_2}{2h} = \alpha$$

leading in that case to

$$\frac{1}{h^2}\begin{pmatrix} -3h & 4h & -h & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & 0 & 1 \end{pmatrix}\begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \\ u_{N+1} \end{pmatrix} = \begin{pmatrix} 2\alpha \\ f(x_1) \\ \vdots \\ f(x_N) \\ \beta \end{pmatrix}$$

2.6 The n-D case

Most PDEs involve more than one space variable. However, even in n-D with $n > 1$, the discretization of differential operators very often requires 1-D finite difference schemes only, since it can be done in each direction independently.

Example *Let consider the 2-D Laplacian operator* $\Delta u(x,y) = \dfrac{\partial^2 u}{\partial x^2}(x,y) + \dfrac{\partial^2 u}{\partial y^2}(x,y)$. *Using the usual second-order centered approximation* (2.5) *for the second derivative, one gets immediately:*

$$\frac{u(x+h,y) - 2u(x,y) + u(x-h,y)}{h^2} + \frac{u(x,y+k) - 2u(x,y) + u(x,y-k)}{k^2} = \Delta u(x,y) + \mathcal{O}(h^2+k^2)$$

or, taking $h = k$*:*

$$\frac{u(x+h,y) + u(x-h,y) + u(x,y+h) + u(x,y-h) - 4u(x,y)}{h^2} = \Delta u(x,y) + \mathcal{O}(h^2) \quad (2.9)$$

However, if one considers a more complex stencil involving other grid points, it may be necessary to use a multi-dimensional Taylor formula to build the finite difference scheme. For instance, in 2-D:

Theorem 2.6. The **Taylor formula in two real variables** reads:

$$\begin{aligned} u(x+h,y+k) = u(x,y) &+ h\,\frac{\partial u}{\partial x}(x,y) + k\,\frac{\partial u}{\partial y}(x,y) \\ &+ \frac{h^2}{2}\,\frac{\partial^2 u}{\partial x^2}(x,y) + hk\,\frac{\partial^2 u}{\partial x \partial y}(x,y) + \frac{k^2}{2}\,\frac{\partial^2 u}{\partial y^2}(x,y) \\ &\vdots \\ &+ \sum_{p=0}^{n} \frac{h^p k^{n-p}}{p!\,(n-p)!}\,\frac{\partial^n u}{\partial x^p \partial y^{n-p}}(x,y) \\ &+ \mathcal{O}(h^{n+1}+k^{n+1}) \end{aligned}$$

Example *Coming back to the 2-D Laplacian operator, the preceding Taylor formula can be used to prove that*

$$\frac{u(x+h,y+h)+u(x-h,y+h)+u(x+h,y-h)+u(x-h,y-h)-4u(x,y)}{2h^2} = \Delta u(x,y) + \mathcal{O}(h^2)$$

More precisely:

$$\begin{aligned} &\frac{u(x+h,y+h)+u(x-h,y+h)+u(x+h,y-h)+u(x-h,y-h)-4u(x,y)}{2h^2} = \\ &\qquad \Delta u(x,y) + \frac{h^2}{12}\left(\frac{\partial^4 u}{\partial x^4}(x,y) + 6\frac{\partial^4 u}{\partial x^2\,\partial y^2}(x,y) + \frac{\partial^4 u}{\partial y^4}(x,y)\right) + o(h^2) \end{aligned}$$

while the more usual scheme (2.9) *satisfies*

$$\begin{aligned} &\frac{u(x+h,y)+u(x-h,y)+u(x,y+h)+u(x,y-h)-4u(x,y)}{h^2} = \\ &\qquad \Delta u(x,y) + \frac{h^2}{12}\left(\frac{\partial^4 u}{\partial x^4}(x,y) + \frac{\partial^4 u}{\partial y^4}(x,y)\right) + o(h^2) \end{aligned}$$

Note however that a multidimensional Taylor formula is a direct consequence of the 1-D Taylor formula. For example, the preceding Taylor formula in two variables can easily be proved by performing a 1-D Taylor expansion in the x-direction w.r.t. $(x, y+k)$ and then several 1-D

Taylor expansions in the y-direction w.r.t. (x, y):

$$
\begin{aligned}
u(x+h, y+k) &= u(x, y+k) + h\,\frac{\partial u}{\partial x}(x, y+k) + \frac{h^2}{2}\,\frac{\partial^2 u}{\partial x^2}(x, y+k) + \dots \\
&= u(x, y) + k\,\frac{\partial u}{\partial y}(x, y) + \frac{k^2}{2}\,\frac{\partial^2 u}{\partial y^2}(x, y) + \dots \\
&\quad + h\left(\frac{\partial u}{\partial x}(x, y) + k\,\frac{\partial^2 u}{\partial x \partial y}(x, y) + \dots\right) \\
&\quad + \frac{h^2}{2}\left(\frac{\partial^2 u}{\partial x^2}(x, y) + \dots\right) + \dots \\
&= u(x, y) + h\,\frac{\partial u}{\partial x}(x, y) + k\,\frac{\partial u}{\partial y}(x, y) + \frac{h^2}{2}\,\frac{\partial^2 u}{\partial x^2}(x, y) + hk\,\frac{\partial^2 u}{\partial x \partial y}(x, y) + \frac{k^2}{2}\,\frac{\partial^2 u}{\partial y^2}(x, y) + \dots
\end{aligned}
$$

Chapter 3

Laplace and Poisson problems

3.1 Some vocabulary

The steady-state solution of a physical phenomenon governed by diffusion satisfies

$$-\text{div}\ (\nu(\mathbf{x})\,\nabla u(\mathbf{x})) = f$$

where u is the state variable (temperature, chemical concentration...), ν is the diffusion coefficient and f the forcing term (source/sink).

If k is actually a constant, the PDE becomes $-\nu\,\Delta u = f$, called a **Poisson equation**.

Moreover, if f is equal to zero, the PDE becomes $\Delta u = 0$, called a **Laplace equation**. The solutions of Laplace equation are called **harmonic functions**.

3.2 Some general remarks on harmonic functions

3.2.1 Harmonic functions in $\mathbb{R}^2$

Examples of harmonic functions in $\mathbb{R}^2$ are:

- $u(x,y) = a(x^2 - y^2) + bxy + cx + dy + e$
- $\begin{cases} u^1_\lambda(x,y) = (a\cos\lambda x + b\sin\lambda x)(ce^{\lambda y} + de^{-\lambda y}) \\ u^2_\lambda(x,y) = (ae^{\lambda x} + be^{-\lambda x})(c\cos\lambda y + d\sin\lambda y) \end{cases} \quad \forall\lambda\in\mathbb{R}, \forall a,b,c,d\in\mathbb{R}$
- In polar coordinates:

$$\begin{cases} u_0(r,\theta) = c_0\ln r + d_0 \\ u_n(r,\theta) = (a_n\cos n\theta + b_n\sin n\theta)\left(c_n r^n + \dfrac{d_n}{r^n}\right) \end{cases} \quad \forall n\in\mathbb{N}^* \quad \text{for } r\neq 0,\ \forall a_n,b_n,c_n,d_n\in\mathbb{R}$$

Moreover, Δ being a linear operator, any linear combination of harmonic functions is also an harmonic function.
Therefore there are "many" harmonic functions since Span $\{u^1_\lambda, u^2_\lambda, \lambda\in\mathbb{R}\}$, which is a space of uncountable infinite dimension, is included in the set of harmonic functions in $\mathbb{R}^2$.

3.2.2 Harmonic functions in bounded domains in $\mathbb{R}^2$

PDEs are often defined on bounded domains, rather than on $\mathbb{R}^n$. Analytical solutions can be found in some cases, in particular for domains with simple geometries, like rectangles or disks in $\mathbb{R}^2$. This is the case for the Laplace equation, where preceding elementary harmonic functions can be combined to get solutions on particular domains. For instance:

- The solution to the Laplace equation in $\Omega = (0, L_x) \times (0, L_y)$ with Dirichlet boundary conditions may be obtained by a separation of variables technique and a superposition principle. For instance, for $u(0, y) = h(y), u(L_x, y) = u(x, 0) = u(x, L_y) = 0$, the solution reads

$$u(x, y) = \sum_{k \geq 1} \alpha_k \left(e^{\lambda_k x} - e^{\lambda_k (2L_x - x)} \right) \sin(\lambda_k y)$$

  where $\lambda_k = \dfrac{k\pi}{L_y}$ and $\alpha_k = \dfrac{2}{L_y \left(1 - e^{2\lambda_k L_x}\right)} \displaystyle\int_0^{L_y} h(y) \sin(\lambda_k y)\, dy.$

- The solution to the Laplace equation on the open disk Ω of center $(0, 0)$ and radius R with Dirichlet boundary conditions $u = g(\theta)$ is

$$u(r, \theta) = K(r, \theta) * g(\theta) = \frac{1}{2\pi} \int_0^{2\pi} K(r, \theta - \alpha) g(\alpha)\, d\alpha \quad \text{where } K(r, \theta) = \frac{R^2 - r^2}{R^2 + r^2 - 2rR\cos\theta}$$

  Note that this result can actually be extended to any dimension n:

$$u(\mathbf{x}) = \frac{R^2 - \|\mathbf{x}\|^2}{R\ |\partial B(O, 1)|} \int_{\partial B(O,R)} \frac{g(\mathbf{y})}{\|\mathbf{x} - \mathbf{y}\|^n}\, dS(\mathbf{y})$$

  where $B(O, R)$ is the ball of center $(0, \ldots, 0)$ and radius R in $\mathbb{R}^n$.

3.2.3 Some properties of harmonic functions

The harmonic functions share a number of properties. For an open set $\Omega \subset \mathbb{R}^n$:

- **Global influence of boundary values**: u changes everywhere in Ω as soon as the Dirichlet boundary data changes somewhere on $\partial\Omega$.
- **Regularity**: If the Dirichlet boundary data $g \in \mathcal{C}^0(\partial\Omega)$ then $u \in \mathcal{C}^\infty(\Omega)$.
- **Mean value property**: Let $B(\mathbf{x}, r)$ denote the ball of center $\mathbf{x}$ and radius r. For each closed ball $B(\mathbf{x}, r) \subset \Omega$:

$$u(\mathbf{x}) = \frac{1}{|B(\mathbf{x}, r)|} \int_{B(\mathbf{x},r)} u(\mathbf{y})\, d\mathbf{y} = \frac{1}{|\partial B(\mathbf{x}, r)|} \int_{\partial B(\mathbf{x},r)} u(\boldsymbol{\sigma})\, d\boldsymbol{\sigma}$$

  This means that the value of a harmonic function at point $\mathbf{x}$ is the average of its values over every ball and every sphere which center is $\mathbf{x}$ and which is contained in the domain. One can even prove that, if u is a $\mathcal{C}^2$ function that satisfies the mean value property, then u is a harmonic function.

- **Maximum principle**: If $u \in \mathcal{C}^2(\Omega)$ and $u \in \mathcal{C}^0(\overline{\Omega})$, then u has no extreme values in Ω.

3.3 Poisson equation in $\mathbb{R}^2$ and $\mathbb{R}^3$

In $\mathbb{R}^2$, the Laplacian operator in polar coordinates reads $\Delta u = \dfrac{\partial^2 u}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial u}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2 u}{\partial \theta^2}$.
The corresponding radial harmonic functions, defined on $\mathbb{R}^2 \setminus \{(0,0)\}$, are

$$u(r,\theta) = u(r) = a \ln r + b \qquad \forall a, b \in \mathbb{R}$$

In $\mathbb{R}^3$, the Laplacian operator in spherical coordinates reads

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2 \sin\phi}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r^2 \sin\phi}\frac{\partial}{\partial \phi}\left(\sin\phi \frac{\partial u}{\partial \phi}\right)$$

The corresponding radial harmonic functions, defined on $\mathbb{R}^3 \setminus \{(0,0,0)\}$, are

$$u(r,\theta,\phi) = u(r) = \frac{a}{r} + b \qquad \forall a, b \in \mathbb{R}$$

The function

$$K(r) = \begin{cases} \dfrac{1}{2\pi} \ln r & \text{in } \mathbb{R}^2 \setminus \{(0,0)\} \\ \dfrac{-1}{4\pi r} & \text{in } \mathbb{R}^3 \setminus \{(0,0,0)\} \end{cases}$$

which corresponds to particular cases of the preceding radial harmonic functions, is called **Poisson kernel** in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Theorem 3.1. Let consider the Poisson problem $\Delta u(\mathbf{x}) = f(\mathbf{x})$ in $\mathbb{R}^n$ ($n = 2$ or 3), with $\|u\| \to 0$ as $\|\mathbf{x}\| \to \infty$. Then $u = K * f$ is a solution to this problem.
Moreover, if $f \in \mathcal{C}^2$ and is zero far away from 0, then $u \in \mathcal{C}^2$.

The proof of this theorem implies the use of the so-called distribution theory. This result can also be extended to bounded domains.

3.4 Generalization to any linear operator on $\mathbb{R}^n$

The proof of theorem 3.1 can actually easily be extended to any linear operator in $\mathbb{R}^n$. This leads to the following result:

Theorem 3.2. Let consider the PDE $Lu(\mathbf{x}) = f(\mathbf{x})$ in $\mathbb{R}^n$, where L is a linear partial differential operator. If $K(\mathbf{x})$ is a distribution that satisfies $LK = \delta$ where δ is the Dirac distribution, then $u = K * f$ is a solution to the PDE.

K is the kernel associated to L on $\mathbb{R}^n$.

3.5 Companion equations and operators

Several other operators are close to, or derived from, the Laplacian operator. Some well-known ones are:

- the **Helmholtz operator** $\Delta + \lambda^2 \,\mathrm{Id}$. $\Delta u + \lambda^2\ u = 0$ is the Helmholtz equation. It appears for instance in acoustics, seismology, electromagnetic radiation..., and actually for every problem linked to the diffusion equation or to the wave equation. As a matter of fact, as will be discussed later (see §6.2.4, §7.2.1 and Appendix C), solving the Helmholtz equation corresponds to looking for the eigenvalues and eigenfunctions of the Laplacian operator, which are fundamental components of the solutions of these equations.
- the **biharmonic operator** Δ^2: $\Delta^2 u = \Delta(\Delta u)$. It appears for instance in continuum mechanics. A famous example is also the *Chladni figures*, where an experimental device (sand on vibrating plates) highlights the zero isolines of its eigenfunctions.
- and more generally the iterated Laplacian operators Δ^p ($p \in \mathbb{N}$). They appear in particular in the parameterization of dissipation processes in fluid mechanics.

3.6 Finite difference schemes

As seen in Chapter 2, the usual discretization scheme for the second-order derivative is given by (2.5):

$$u''(x) = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} + \mathcal{O}(h^2)$$

Hence the usual second-order finite difference scheme for the Laplacian:

$$\Delta u_{i_1,i_2,\dots,i_N} = \frac{u_{i_1-1,i_2,\dots,i_N} - 2u_{i_1,i_2,\dots,i_N} + u_{i_1+1,i_2,\dots,i_N}}{h_1^2} + \cdots + \frac{u_{i_1,i_2,\dots,i_N-1} - 2u_{i_1,i_2,\dots,i_N} + u_{i_1,i_2,\dots,i_N+1}}{h_N^2} + \mathcal{O}(h^2)$$

with the convention $h^2 = \sum_{i=1}^{N} h_i^2$.

In the particular case $N = 2$, it reads:

$$\Delta u_{i,j} = \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h_x^2} + \frac{u_{i,j-1} - 2u_{ij} + u_{i,j+1}}{h_y^2} + \mathcal{O}(h^2)\ ,$$

which reduces to the well-known five-point scheme if $h_x = h_y$:

$$\Delta u_{i,j} = \frac{u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} - 4u_{ij}}{h^2} + \mathcal{O}(h^2)$$

The discretization of the Laplace equation with these second-order schemes leads to a numerical solution that obviously satisfies the maximum principle, since $u_{i_1,i_2,\dots,i_N}$ appears as a weighted average, with positive weights, of neighboring points.

Some properties of this scheme, and of an alternative 9-point scheme, are given in the exercise sheet.

Chapter 4

Dealing with the time variable

This chapter focuses time-dependent PDEs, and in particular on aspects related to their finite difference discretization.

Compared to other variables, time is quite specific. Indeed, time is running forward, and the very large majority of time-dependent problems are seeking for the evolution in time of a system, given its initial state. Time has also its own vocabulary: as seen in Definition 1.12, solving a time-dependent PDE with initial conditions is called an *initial value problem*, or a *Cauchy problem*, while solving more generally a steady-state PDE with boundary conditions is called a *boundary value problem*.

4.1 Some basic behaviors of solutions of first-order in time PDEs

Let consider the generic equation $\dfrac{\partial u}{\partial t} = F(u)$ where F is a partial differential operator involving only derivatives with respect to space coordinates. Simple cases are $F(u) = -r(\mathbf{x},t)\,u$ (reaction equation), or $F(u) = -\mathbf{c}(\mathbf{x},t).\nabla u$ (transport equation — see Chapter 5), or $F(u) = \text{div}(k(\mathbf{x},t)\nabla u)$ (diffusion equation — see Chapter 7). Since many PDEs involve such terms, it is quite interesting to know a priori, at least in a qualitative way, what is the individual effect of each of these terms on the solution. In more complex cases mixing these different aspects, the behavior of the solution will also be a mix of these elementary behaviors.

As will be seen in the following chapters, the expressions of the elementary solutions on $\mathbb{R}^n$ in the case of constant coefficients are the following:

Reaction	$\dfrac{\partial u}{\partial t}(\mathbf{x},t) + r\,u(\mathbf{x},t) = 0$	$u(\mathbf{x},t) = u(\mathbf{x},0)\,e^{-rt}$
Transport	$\dfrac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}.\nabla u(\mathbf{x},t) = 0$	$u(\mathbf{x},t) = u(\mathbf{x}-\mathbf{c}t,0)$
Diffusion	$\dfrac{\partial u}{\partial t}(\mathbf{x},t) - \nu\,\Delta u(\mathbf{x},t) = 0$	$u(\mathbf{x},t) = (u(.,0) * K(.,t))\,(\mathbf{x})$ with $K(\mathbf{x},t) = \left(2\sqrt{\pi\nu t}\right)^{-n} e^{-\frac{\|\mathbf{x}\|^2}{4\nu t}}$

An illustration of their behavior in the 1-D case is given in Figure 4.1.

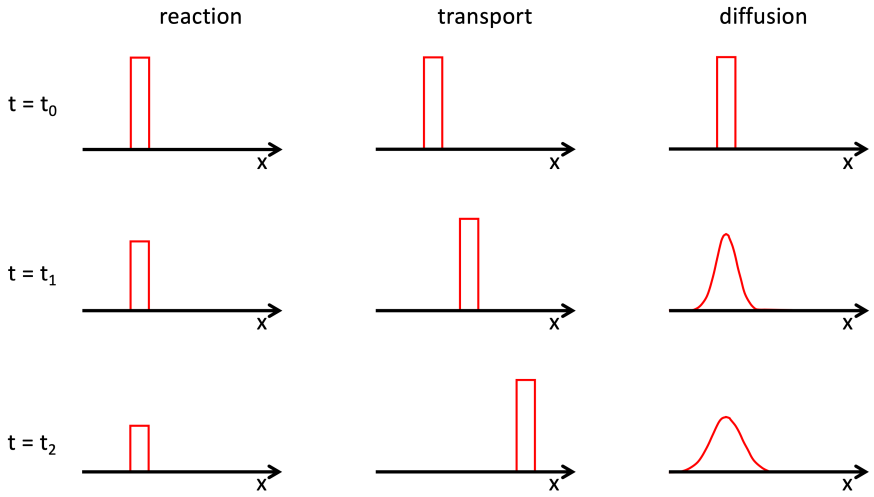

Figure 4.1: Schematic view of the behavior of the solutions of reaction, transport, and diffusion equations, in the 1-D case with constant coefficients.

4.2 Discretization of $\frac{\partial u}{\partial t}$

The most usual schemes for the time derivative are:

- **Euler forward**: $\dfrac{\partial u}{\partial t}(x,t) = \dfrac{u(x,t+\delta t) - u(x,t)}{\delta t} + \mathcal{O}(\delta t)$
- **Euler backward**: $\dfrac{\partial u}{\partial t}(x,t) = \dfrac{u(x,t) - u(x,t-\delta t)}{\delta t} + \mathcal{O}(\delta t)$
- **Leap-frog**: $\dfrac{\partial u}{\partial t}(x,t) = \dfrac{u(x,t+\delta t) - u(x,t-\delta t)}{2\delta t} + \mathcal{O}(\delta t^2)$

Euler schemes are obviously first-order accurate, while the leap-frog scheme is second-order accurate.

These schemes are widely used, mostly because of their simplicity. However many other time integration schemes are available, some of them being given in §4.5.

4.3 Time discretization of $F(u)$

When discretizing in time the equation $\dfrac{\partial u}{\partial t} = F(u)$, several choices can be made for the evaluation of $F(u)$.

- If it is evaluated at time t, the scheme is said to be **explicit**, since $u(x,t+\delta t)$ can be explicitly expressed from the values of u at previous time steps.

Example *Euler forward + explicit scheme*

$$\frac{u(x,t+\delta t)-u(x,t)}{\delta t} \simeq F(u(x,t)) \quad \longrightarrow \quad u(x,t+\delta t) \simeq u(x,t) + \delta t\, F(u(x,t))$$

- On the opposite, if $F(u)$ is evaluated at time $t+\delta t$, the scheme is said to be **implicit**, since it is necessary to solve an equation to deduce $u(x,t+\delta t)$ from the values of u at previous time steps.

  Example *Euler forward + implicit scheme*

  $$\frac{u(x,t+\delta t)-u(x,t)}{\delta t} \simeq F(u(x,t+\delta t)) \quad \longrightarrow \quad u(x,t+\delta t) - \delta t\, F(u(x,t+\delta t)) \simeq u(x,t)$$

- A combination of both approaches is also possible:

  $$\alpha\, F(u(x,t+\delta t)) + (1-\alpha)\, F(u(x,t)), \qquad \text{with } 0 \leq \alpha \leq 1$$

  $\alpha = 0$ corresponds to the explicit scheme and $\alpha = 1$ to the implicit one. For $0 < \alpha < 1$, the scheme is said to be **semi-implicit**.

- The particular case $\alpha = 0.5$ is called the **Crank-Nicolson** scheme. One of its main interests comes from the fact that an Euler forward scheme associated with a Crank-Nicolson discretization of $F(u)$ corresponds actually to a second-order approximation of $\dfrac{\partial u}{\partial t} = F(u)$ at point $(x, t+\delta t/2)$ (and not only to an obvious first-order approximation of this equation at point (x,t)).

An explicit scheme is simpler and generally requires less calculations at each time step than an implicit scheme. But an implicit scheme is often more stable (see §4.4) and allows larger time steps than an explicit scheme.

4.4 Stability

4.4.1 Numerical stability

The finite difference method leads to an approximation of the solution of a differential equation, the main error source being the truncation error, related to the approximation of derivatives by numerical schemes[1]. This error can be reduced by decreasing the values of the time and space steps, and/or by using higher-order schemes. However, nothing ensures that a small initial error will not grow up into a large error after a number of time steps.

Definition 4.1. A finite difference scheme for a time dependent differential equation is said to be **stable** iff errors remain bounded during the computations. This implies that the numerical solution remains bounded (whenever the exact solution is bounded, of course).

[1] The other error sources are the error associated to the resolution of linear systems, if any, and (to a much lesser extent) the rounding error, due to the representation of real numbers in a computer.

This notion of (in)stability can be illustrated for instance by considering the ODE $\frac{\partial u}{\partial t} = -\lambda u$ ($\lambda > 0$), with the initial condition $u(0) = 1$. Its exact solution is $u(t) = \exp(-\lambda t)$.

An explicit forward Euler discretization reads: $\frac{u_{n+1} - u_n}{\delta t} = -\lambda\, u_n$ where u_n is an approximation of $u(n\delta t)$. This leads to $u_n = (1 - \lambda\delta t)\, u_{n-1} = \cdots = (1 - \lambda\delta t)^n u_0$. Choosing a time step δt such that $|1 - \lambda\delta t| > 1$, i.e. $\delta t > 2/\lambda$, will thus make u_n tend to infinity. This phenomenon is called a **numerical blow-up**. On the contrary, choosing $\delta t \leq 2/\lambda$ ensures that the u_n's remain bounded. Hence the **stability condition**: $\delta t \leq \frac{2}{\lambda}$.

Thus **the stability of numerical schemes must be investigated when discretizing a time dependent differential equation.**

Warning Numerical stability does not imply convergence towards the exact solution, but only ensures that no numerical blow-up will occur (to be convinced, choose for instance $\delta t = 2/\lambda$ in the preceding example).

4.4.2 Investigating the stability: the Fourier method

As mentioned in §1.7, the **Fourier method**, also called **Von Neumann method**, is the most common approach for investigating the stability of a numerical scheme discretizing a **linear** PDE. It consists in considering the expansion into Fourier series (see Appendix B) of the initial condition $u(x, 0)$:

$$u(x,0) = \sum_{k=-\infty}^{+\infty} c_k\, e^{\frac{2i\pi\, kx}{L}} \tag{4.1}$$

The numerical scheme is stable iff it is stable for any individual component. Let $u_n(x)$ the numerical approximation of the exact solution u at time $n\delta t$. For linear schemes, it can easily be proved by recurrence that, if $u_0(x) = e^{ipx}$ ($p \in \mathbb{R}$), then $u_n(x) = \xi^n e^{ipx}$, where ξ is a complex number that depends on p and on the numerical scheme (in particular on the time and space steps). The Fourier stability criterion then consists in imposing that $|\xi| \leq 1$ for any p, which implies constraints on the time and space steps.

Example *Let consider the 1D transport equation* $\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = 0$ *for* $x \in [0, L], t > 0$, *with* $c > 0$, *and with the initial condition* $u(x, 0) = u_0(x)$ *and some boundary condition at* $x = 0$. *Let a regular mesh of* $[0, L]$, *with* δx *the mesh step. Let also* δt *the time step, and* u_j^n *the approximation of the exact solution at time* $n\,\delta t$ *at grid point* j. *Let consider the discretization scheme:*

$$\frac{u_j^{n+1} - u_j^n}{\delta t} + c\,\frac{u_j^n - u_{j-1}^n}{\delta x} = 0$$

Replacing u_j^n *by* $\xi^n e^{ipj\delta x}$ *leads to* $\frac{\xi - 1}{\delta t} + c\,\frac{1 - e^{-ip\delta x}}{\delta x} = 0$ *i.e.* $\xi = 1 - c\,\frac{\delta t}{\delta x}\left(1 - e^{-ip\delta x}\right)$. *The modulus of* ξ *is thus:*

$$|\xi|^2 = 1 + 2\,(1 - \cos(p\delta x))\, C(C-1) \qquad \text{with } C = c\,\frac{\delta t}{\delta x}$$

The stability condition $|\xi| \leq 1, \forall p \in \mathbb{R}$, *is thus equivalent to* $C \leq 1$, *i.e.* $\dfrac{\delta x}{\delta t} \geq c$. *This condition is called the* **Courant-Friedrichs-Lewy (or CFL) condition**. *The dimensionless quantity* $c\dfrac{\delta t}{\delta x}$ *is called the* **Courant (or CFL) number**.

Remarks

- The coefficient ξ is of course closely linked to the transfer functions (see §2.2.4) of the different individual time and space schemes involved in the finite difference approximation of the PDE.
- Note that stability must be investigated considering the homogeneous equation (i.e. without any right-hand side). As a matter of fact, a source or sink term would continuously inject or consume energy, which could make the solution blow up or remain bounded independently of the intrinsic properties of the numerical scheme.
- For this approach and the expansion (4.1) to be rigorous, the PDE and the numerical scheme must be linear, with constant coefficients. However this Fourier/Von Neumann method is actually used for much more general cases (by linearizing the equation and/or freezing the coefficients), and it generally provides a good guess at the possible constraints on the time and space steps.
- In the same way, this approach is usually applied to the PDE on all space with no boundaries ($x \in \mathbb{R}$). It can also be used to study the stability of problems with periodic boundary conditions. But studying the stability of problems with more general boundary conditions can be quite difficult: Von Neumann analysis actually addresses the issue of stability of the PDE discretization alone.
- This method can be generalized to 2D and 3D cases.

Some generic calculations facilitating the computation of stability criteria are given in Appendix D.1.

4.4.3 Other methods for investigating stability issues

Other methods are available to study the stability properties of finite difference schemes.

- The **eigenvalue method** consists in writing the matrix form of the scheme:

$$U^{n+1} = MU^n \qquad \text{with } U^n = \begin{pmatrix} u_1^n \\ \vdots \\ u_J^n \end{pmatrix}$$

  We have then obviously $U^n = M^n U^0$. A stability condition is thus that the spectral radius of M (i.e. the maximum of the moduli of its eigenvalues) be less or equal to 1. One has thus to study the eigenvalues of M.

- Stability can be also be directly assessed by the **energy method**. Let define the energy of the discrete solution as $E^n = \sum_{j=1}^{J} (u_j^n)^2$. The goal is then to prove that this quantity remains bounded independently of n, as n tends to infinity.
  Note that, thanks to the Parseval's theorem (see Appendix B), this corresponds to the Fourier method, but in the real space.

- These three methods (Fourier, eigenvalue, energy) aim at proving the L^2 stability of the scheme, i.e. the fact that the L^2 norm of the numerical solution remains bounded independently of the time. Other stability criteria may also be used, like for instance the L^∞ stability (i.e. working on the L^∞ norm of the numerical solution).

4.5 Some time discretization schemes

Numerous time integration schemes are available for ODEs and PDEs. Without looking for exhaustiveness, let us mention a few.

Let the differential equation $\frac{\partial u}{\partial t} = F(u,t)$ with the initial condition $u(0) = u_0$. Let δt the time step, $t_n = n\,\delta t$ and u_n the approximation of $u(t_n)$.

4.5.1 One step methods

One step methods are integration schemes determining the value at the next time step using only the value at the current time step. The simplest one is the already mentioned Euler method. Let cite also the **midpoint method** (second order):

$$\begin{aligned} k_1 &= \delta t\, F(u_n, t_n) \\ k_2 &= \delta t\, F(u_n + \frac{k_1}{2}, t_n + \frac{\delta t}{2}) \\ u_{n+1} &= u_n + k_2 \end{aligned}$$

and more generally the **Runge-Kutta methods**, the most famous one being probably the fourth-order scheme:

$$\begin{aligned} k_1 &= \delta t\, F(u_n, t_n) \\ k_2 &= \delta t\, F(u_n + \frac{k_1}{2}, t_n + \frac{\delta t}{2}) \\ k_3 &= \delta t\, F(u_n + \frac{k_2}{2}, t_n + \frac{\delta t}{2}) \\ k_4 &= \delta t\, F(u_n + k_3, t_n + \delta t) \\ u_{n+1} &= u_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} \end{aligned}$$

4.5.2 Multi-step methods

Contrary to one-step methods, the value at the next time step is obtained using values at several preceding time steps. Let cite **Adams-Bashforth explicit methods** (k^{th}-order):

$$k=1 \qquad u_{n+1} = u_n + \delta t\, F(u_n, t_n) \qquad \text{(i.e. explicit Euler scheme)}$$

$$k=2 \qquad u_{n+2} = u_{n+1} + \frac{\delta t}{2}\left[3F(u_{n+1}, t_{n+1}) - F(u_n, t_n)\right]$$

$$k=3 \qquad u_{n+3} = u_{n+2} + \frac{\delta t}{12}\left[23F(u_{n+2}, t_{n+2}) - 16F(u_{n+1}, t_{n+1}) + 5F(u_n, t_n)\right]$$

$$k=4 \qquad u_{n+4} = u_{n+3} + \frac{\delta t}{24}\left[55F(u_{n+3}, t_{n+3}) - 59F(u_{n+2}, t_{n+2}) + 37F(u_{n+1}, t_{n+1}) - 9F(u_n, t_n)\right]$$

and the **Adams-Moulton implicit methods** (k^{th}-order):

$$k=1 \qquad u_{n+1} = u_n + \delta t\, F(u_{n+1}, t_{n+1}) \qquad \text{(i.e. implicit Euler scheme)}$$

$$k=2 \qquad u_{n+1} = u_n + \frac{\delta t}{2}\left[F(u_{n+1}, t_{n+1}) + F(u_n, t_n)\right]$$

$$k=3 \qquad u_{n+2} = u_{n+1} + \frac{\delta t}{12}\left[5F(u_{n+2}, t_{n+2}) + 8F(u_{n+1}, t_{n+1}) - F(u_n, t_n)\right]$$

$$k=4 \qquad u_{n+3} = u_{n+2} + \frac{\delta t}{24}\left[9F(u_{n+3}, t_{n+3}) + 19F(u_{n+2}, t_{n+2}) - 5F(u_{n+1}, t_{n+1}) + F(u_n, t_n)\right]$$

4.5.3 Predictor-corrector schemes

The idea underlying this approach is to get performances close to implicit schemes without their major drawback, i.e. without solving a linear system. It is decomposed in two stages:

- **Prediction**: use an explicit scheme to get a first approximation $\widetilde{u}_{n+1}$.
- **Correction**: use an implicit scheme, but replacing u_{n+1} by its approximation $\widetilde{u}_{n+1}$ in the implicit part of the equation.

Example *Prediction with an explicit Euler scheme and correction with an implicit Euler scheme reads:*

Prediction: $\widetilde{u}_{n+1} = u_n + \delta t\, F(u_n, t_n)$.

Correction: $u_{n+1} = u_n + \delta t\, F(\widetilde{u}_{n+1}, t_{n+1})$.

Chapter 5

The transport equation and first-order linear PDEs

This chapter focuses on the **transport equation** $\frac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}(\mathbf{x},t).\nabla u(\mathbf{x},t) = 0$, and more generally on linear first-order PDEs: $\frac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}(\mathbf{x},t).\nabla u(\mathbf{x},t) + r(\mathbf{x},t)\, u(\mathbf{x},t) = f(\mathbf{x},t)$.

5.1 Some generalities

5.1.1 Physical interpretation

Let consider a scalar quantity $u(\mathbf{x},t)$ transported by a given velocity field $\mathbf{c}(\mathbf{x},t)$ in a domain $\Omega \subset \mathbb{R}^n$ (e.g. the concentration of some chemical species in a fluid, an oil spill in the ocean...). In case of a pure transport, i.e. without any interaction of the quantity with the surrounding medium, the behavior of u is governed by the equation

$$\frac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}(\mathbf{x},t).\nabla u(\mathbf{x},t) = 0$$

$$\text{i.e.} \quad \frac{\partial u}{\partial t}(\mathbf{x},t) + \sum_{i=1}^{n} c_i(\mathbf{x},t)\, \frac{\partial u}{\partial x_i}(\mathbf{x},t) = 0$$

This can easily be proved by stating that, for an infinitesimal interval of time δt, $u(\mathbf{x}+\mathbf{c}(\mathbf{x},t)\,\delta t, t+\delta t) = u(\mathbf{x},t)$, then developing the first term using a Taylor expansion, and making δt tend to zero.

5.1.2 Boundary conditions

The basic principle regarding boundary conditions is that, for the transport equation to be well-posed, information is required at the boundary where and when the transport field is incoming, but not where and when it is outgoing. Denoting $\mathbf{n}(\mathbf{x})$ the outward normal vector for $\mathbf{x} \in \partial\Omega$, this incoming portion of the boundary is $\partial\Omega^-(t) = \{\mathbf{x} \in \partial\Omega \,/\, \mathbf{c}(\mathbf{x},t).\,\mathbf{n}(\mathbf{x}) < 0\}$.

5.2 Analytical resolution: the method of characteristics

5.2.1 Eulerian vs Lagrangian representations

In the context of continuum mechanics (fluid mechanics, solid mechanics), the system state can be described using either an Eulerian or a Lagrangian point of view. The **Eulerian** point of view consists in quantifying the properties of the medium (mass, velocity, temperature...) at every time t for any given location $\mathbf{x}$. On the other hand, in the **Lagrangian** point of view, the observer follows a specific material particle all along its motion.

We will use in the following the notation $X(s;\mathbf{x},t)$ to indicate the location at time s of the material particle that is located at point $\mathbf{x}$ at time t. $\mathbf{x}$ is the Eulerian coordinate, while X is the Lagrangian one.

5.2.2 Method of characteristics: general principle

Let consider the general linear first-order PDE in $\mathbb{R}^n$:

$$\begin{cases} \dfrac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}(\mathbf{x},t).\nabla u(\mathbf{x},t) + r(\mathbf{x},t)\, u(\mathbf{x},t) = f(\mathbf{x},t) \qquad \mathbf{x}\in\mathbb{R}^n, t>0 \\ u(\mathbf{x},0) = u_0(\mathbf{x}) \end{cases}$$

with: $\mathbf{c}(\mathbf{x},t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ a continuous function, Lipschitz continuous w.r.t. t
$r(\mathbf{x},t)$ and $f(\mathbf{x},t)$: $\mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ continuous functions
$u_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ a continuous function.

The **method of characteristics** to solve this PDE consists in the following steps:

- Let $(\mathbf{x},t)$ fixed.
- Solve the differential system (n equations, n unknowns):

$$\begin{cases} X'(s) = \mathbf{c}(X(s),s) \\ X(t) = \mathbf{x} \end{cases}$$

  Its solution is denoted $X(s)$, or $X(s;\mathbf{x},t)$. It is called the **characteristic curve** or **characteristic function** related to $(\mathbf{x},t)$. It corresponds to the trajectory that the material particle would follow in the pure transport case (i.e. $r = f = 0$).

  We have then $\dfrac{d}{ds}u(X(s),s) = f(X(s),s) - r(X(s),s)\, u(X(s),s)$, which means that $Z(s) = u(X(s),s)$ satisfies $Z'(s) = f(X(s),s) - r(X(s),s)\, Z(s)$.

- Solve the ODE:

$$\begin{cases} Z'(s) + r(X(s),s)\, Z(s) = f(X(s),s) \quad s>0 \\ Z(0) = u(X(0),0) = u_0(X(0;\mathbf{x},t)) \end{cases}$$

- Then: $u(\mathbf{x},t) = u(X(t),t) = Z(t;\mathbf{x},t)$

Example *Let consider the PDE*

$$\begin{cases} \dfrac{\partial u}{\partial t} + x\dfrac{\partial u}{\partial x} + y\dfrac{\partial u}{\partial y} - 3u = 0 \qquad (x,y) \in \mathbb{R}^2, t > 0 \\ u(x,y,0) = x + y \end{cases}$$

With previous notations, $\mathbf{c}(\mathbf{x},t) = (c_1(x,y,t), c_2(x,y,t)) = (x,y)$, $r(\mathbf{x},t) = -3$ *and* $f(\mathbf{x},t) = 0$. *For fixed* (x,y,t), *the first step consists in solving*

$$\begin{cases} X_1'(s) = c_1(X(s),s) = X_1(s) \\ X_1(t) = x \end{cases} \quad \text{and} \quad \begin{cases} X_2'(s) = c_2(X(s),s) = X_2(s) \\ X_2(t) = y \end{cases}$$

This leads to $X_1(s) = x\,e^{s-t}$ *and* $X_2(s) = y\,e^{s-t}$.

The second step consists then in solving

$$\begin{cases} Z'(s) - 3Z(s) = 0 \qquad s > 0 \\ Z(0) = u(X(0),0) = u_0(X(0;x,y,t)) = u_0(x\,e^{-t}, y\,e^{-t}) = (x+y)\,e^{-t} \end{cases}$$

which leads to $Z(s) = (x+y)\,e^{3s-t}$. *Hence the solution* $u(x,y,t) = Z(t;x,y,t) = (x+y)\,e^{2t}$.

5.2.3 Case of a pure transport equation

The particular case where $r(\mathbf{x},t) = f(\mathbf{x},t) = 0$ corresponds to a pure transport equation:

$$\begin{cases} \dfrac{\partial u}{\partial t}(\mathbf{x},t) + \mathbf{c}(\mathbf{x},t).\nabla u(\mathbf{x},t) = 0 \\ u(\mathbf{x},0) = u_0(\mathbf{x}) \end{cases}$$

The previous methodology leads to the solution $u(\mathbf{x},t) = u_0(X(0;\mathbf{x},t))$, which becomes $u_0(\mathbf{x} - \mathbf{c}t)$ in the particular case of a constant $\mathbf{c}$.

5.2.4 Case of a bounded domain

If the PDE is defined on a bounded domain $\Omega \subset \mathbb{R}^n$, the computation of $X(0;\mathbf{x},t)$ may become impossible, since the trajectory $s \longrightarrow X(s;\mathbf{x},t)$ may be out of Ω for $s < \tau_{in}$, with $\tau_{in} > 0$.
Let thus $\tau_{in}(\mathbf{x},t) = \inf\{s \in [0,t] / X(s;\mathbf{x},t) \in \Omega\}$ (τ_{in} exists since the set is not empty and has 0 as a lower bound).

- If $\tau_{in}(\mathbf{x},t) > 0$, then $X(\tau_{in}(\mathbf{x},t);\mathbf{x},t)$ is located on $\partial\Omega$ and one has thus to modify the initial condition in the ODE for Z, which becomes

$$\begin{cases} Z'(s) + r(X(s),s)\,Z(s) = f(X(s),s) \quad s > \tau_{in}(\mathbf{x},t) \\ Z(\tau_{in}(\mathbf{x},t)) = u(X(\tau_{in}(\mathbf{x},t)), \tau_{in}(\mathbf{x},t)) = \text{a value given by the boundary condition} \end{cases}$$

- If $\tau_{in}(\mathbf{x},t)=0$, then $X(\tau_{in}(\mathbf{x},t);\mathbf{x},t)$ is not necessarily located on $\partial\Omega$, but $X(\varepsilon;\mathbf{x},t)\in\Omega,\ \forall\varepsilon>0$. Some additional regularity conditions are thus required to go back in time up to $s=0$ and use $u_0(X(0);\mathbf{x},t)$ again as the initial condition.

5.3 Numerical schemes for the 1D transport equation

Let consider the 1D transport equation

$$\begin{cases} \dfrac{\partial u}{\partial t}+c\dfrac{\partial u}{\partial x}=0 & x\in\mathbb{R}, t>0 \\ u(x,0)=u_0(x) & x\in\mathbb{R} \end{cases} \qquad \text{with } c>0$$

Its exact solution is $u(x,t)=u_0(x-c\,t)$. The characteristic curves are straight lines $x-c\,t=constant$.

5.3.1 Euler one-sided explicit schemes

Let consider the two following first-order finite difference schemes (with usual notations):

- **Downstream scheme**: $\dfrac{u_j^{n+1}-u_j^n}{\delta t}+c\,\dfrac{u_{j+1}^n-u_j^n}{\delta x}=0$
- **Upwind scheme**: $\dfrac{u_j^{n+1}-u_j^n}{\delta t}+c\,\dfrac{u_j^n-u_{j-1}^n}{\delta x}=0$

A Fourier stability analysis (see §4.4.2) proves that the downstream scheme is unconditionally unstable, while the upwind scheme is stable for $C\leq 1$, where $C=c\,\dfrac{\delta t}{\delta x}$ is the Courant number.

Interpretation in terms of domain of dependence The true value for $u(j\delta x,(n+1)\delta t)$ is equal to $u(j\delta x-c\,\delta t,n\,\delta t)$ (see Figure 5.1).

- The downstream scheme can be rewritten as

  $$u_j^{n+1}=u_j^n-C\,(u_{j+1}^n-u_j^n)=(1+C)\,u_j^n-C\,u_{j+1}^n$$

  However the true location $j\delta x-c\,\delta t$ does not lie in the interval $[j\,\delta x;(j+1)\,\delta x]$ whatever the value of δt (see Figure 5.1), which gives some rationale to the fact that this scheme is unstable.

- The upwind scheme can be rewritten as

  $$u_j^{n+1}=u_j^n-C\,(u_j^n-u_{j-1}^n)=(1-C)\,u_j^n+C\,u_{j-1}^n$$

  This means that, if $C\leq 1$, u_j^{n+1} is a convex combination of u_{j-1}^n and u_j^n, while the true location $j\delta x-c\,\delta t$ is indeed in the interval $[(j-1)\,\delta x;j\,\delta x]$ (see Figure 5.1).

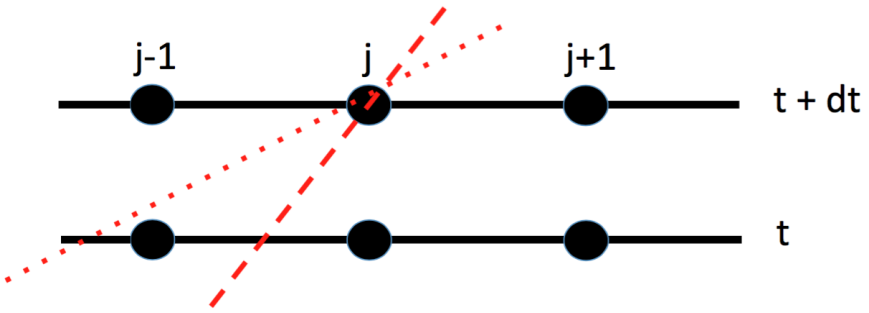

Figure 5.1: Schematic view of the transport from time t to time $t + \delta t$ w.r.t. the discretization grid. The dotted and dashed lines are characteristic lines, respectively for $C > 1$ and $C < 1$.

Interpretation in terms of equivalent PDE A Taylor expansion of the Euler scheme (see also Appendix D.3) leads to

$$\frac{u(j\,\delta x,(n+1)\delta t)-u(j\,\delta x,n\,\delta t)}{\delta t}=\frac{\partial u}{\partial t}(j\,\delta x,n\,\delta t)+\frac{\delta t}{2}\frac{\partial^2 u}{\partial t^2}(j\,\delta x,n\,\delta t)+\dots$$

Moreover:

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

which leads to:

$$\frac{u(j\,\delta x,(n+1)\delta t)-u(j\,\delta x,n\,\delta t)}{\delta t}=\frac{\partial u}{\partial t}(j\,\delta x,n\,\delta t)+c^2\,\frac{\delta t}{2}\frac{\partial^2 u}{\partial x^2}(j\,\delta x,n\,\delta t)+\dots$$

The leading order term of the error, $c^2\,\frac{\delta t}{2}\frac{\partial^2 u}{\partial x^2}$, is thus an antidiffusive term, which continuously injects energy in the numerical solution (see Chapter 7 and Appendix D.4), and may thus lead to numerical instability. In order for the numerical scheme to be stable, the discretization of the spatial derivative must therefore compensate for this antidiffusive term.

Taylor expansions applied to the the downstream and upwind schemes (see also Appendix D.3) lead to:

$$\begin{cases}\dfrac{u((j+1)\,\delta x,n\,\delta t)-u(j\,\delta x,n\,\delta t)}{\delta x}=\dfrac{\partial u}{\partial x}(j\,\delta x,n\,\delta t)+\dfrac{\delta x}{2}\dfrac{\partial^2 u}{\partial x^2}(j\,\delta x,n\,\delta t)+\dots\\[2ex]\dfrac{u(j\,\delta x,n\,\delta t)-u((j-1)\,\delta x,n\,\delta t)}{\delta x}=\dfrac{\partial u}{\partial x}(j\,\delta x,n\,\delta t)-\dfrac{\delta x}{2}\dfrac{\partial^2 u}{\partial x^2}(j\,\delta x,n\,\delta t)+\dots\end{cases}$$

which implies the following equivalent PDEs:

$$\begin{cases}\text{downstream scheme:} & \dfrac{\partial u}{\partial t}+c\,\dfrac{\partial u}{\partial x}+\underbrace{\left(c^2\dfrac{\delta t}{2}+c\dfrac{\delta x}{2}\right)}_{>0}\dfrac{\partial^2 u}{\partial x^2}+\dots=0\\[2ex]\text{upwind scheme:} & \dfrac{\partial u}{\partial t}+c\,\dfrac{\partial u}{\partial x}+\underbrace{\left(c^2\dfrac{\delta t}{2}-c\dfrac{\delta x}{2}\right)}_{\leq 0 \text{ for } C\leq 1}\dfrac{\partial^2 u}{\partial x^2}+\dots=0\end{cases}$$

The leading order term of the error is thus of antidiffusive nature for the downstream scheme, which is coherent with the fact that this scheme is unconditionnaly unstable.

The leading order term of the error is of diffusive nature for the upwind scheme iff $C \leq 1$, which is also coherent with the stability condition found previously. The upwind scheme is then said to be a **diffusive scheme**, which means that the main effect of its error is to smooth and damp out the solution (see also Definition 2.1 and Appendix D.4: the exact plane wave solution $\exp(ip(x-ct))$ becomes $\exp(ip(x-ct))\,\exp(-\nu p^2 t)$ with $\nu = c(\delta x - c\delta t)/2 = c\delta x(1-C)/2$, and $\nu \geq 0$ since $C \leq 1$).

5.3.2 Lax-Wendroff scheme

The preceding upwind scheme is only first-order accurate. A way to increase the accuracy is to use a two-sided scheme for the space derivative. However

$$\frac{u((j+1)\,\delta x, n\,\delta t) - u((j-1)\,\delta x, n\,\delta t)}{2\,\delta x} = \frac{\partial u}{\partial x}(j\,\delta x, n\,\delta t) + \frac{\delta x^2}{6}\frac{\partial^3 u}{\partial x^3}(j\,\delta x, n\,\delta t) + \dots$$

Such a discretization cannot therefore compensate for the leading error term $c^2\,\frac{\delta t}{2}\,\frac{\partial^2 u}{\partial x^2}$ in the Euler scheme. The idea of the so-called Lax-Wendroff scheme is thus to compensate almost exactly for this term by introducing an artificial additional term. It reads:

$$\frac{u_j^{n+1} - u_j^n}{\delta t} + c\,\frac{u_{j+1}^n - u_{j-1}^n}{2\delta x} - c^2\,\frac{\delta t}{2}\,\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\delta x^2} = 0$$

Its equivalent PDE (see also Appendix D.3) then reads

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} + \frac{\delta t^2}{6}\,\frac{\partial^3 u}{\partial t^3} + c\,\frac{\delta x^2}{6}\,\frac{\partial^3 u}{\partial x^3} + \dots = 0$$

which proves that the scheme is second-order accurate in time and space.

A Fourier stability analysis (see §4.4.2) proves that this scheme is stable iff $C \leq 1$.

Since $\frac{\partial^3 u}{\partial t^3} = -c^3\frac{\partial^3 u}{\partial x^3}$, the leading error term in the equivalent PDE can be rewritten as $\frac{c}{6}\left(\delta x^2 - c^2\,\delta t^2\right)\frac{\partial^3 u}{\partial x^3}$. The Lax-Wendroff scheme is thus said to be a **dispersive scheme**, since the main effect of its error is to modify the transport velocity (see also Definition 2.2 and Appendix D.4: the exact plane wave solution $\exp(ip(x-ct))$ becomes $\exp\left(ip\left[x-(c-p^2\mu)t\right]\right)$ with $\mu = c(\delta x^2 - c^2\delta t^2)/6$). Since $C \leq 1$ for stability reasons, $\mu \geq 0$ and the numerical transport is slower than the exact continuous one.

5.3.3 Other schemes

Many other schemes are of course available for this transport equation, several of them being detailed in the corresponding exercise sheet.

Chapter 6

The wave equation

This chapter focuses on the wave equation $\dfrac{\partial^2 u}{\partial t^2}(\mathbf{x},t) - c^2 \Delta u(\mathbf{x},t) = 0 \quad (c > 0)$, which is a prototype for linear hyperbolic second-order PDEs.

6.1 Some properties

6.1.1 Conservation of energy

Following some physical reasoning, the kinetic and potential energies of the solution are defined respectively by

$$KE(t) = \frac{1}{2}\int_\Omega \left(\frac{\partial u}{\partial t}(\mathbf{x},t)\right)^2 d\mathbf{x} \qquad \text{and} \qquad PE(t) = \frac{c^2}{2}\int_\Omega \|\nabla u(\mathbf{x},t)\|^2 \, d\mathbf{x}$$

If the domain $\Omega = \mathbb{R}^n$, or if it is bounded with either homogeneous Neumann boundary conditions $\dfrac{\partial u}{\partial \mathbf{n}} = 0$ or steady-state Dirichlet conditions $u(\mathbf{x},t) = g(\mathbf{x})\ \forall t$, then the total energy $E(t) = KE(t) + PE(t)$ is conserved.

6.1.2 Initial conditions

Since the wave equation involves a second-order time derivative, two initial conditions are required for the problem to be well-posed. Typically the values of the solution and of its time derivative at initial time will be provided:

$$u(\mathbf{x},0) = u_0(\mathbf{x}) \qquad \text{and} \qquad \frac{\partial u}{\partial t}(\mathbf{x},0) = v_0(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega$$

with $u_0 \in \mathcal{C}^2(\Omega)$ and $v_0 \in \mathcal{C}^1(\Omega)$.

6.2 Analytical solutions

The wave equation can be solved analytically, at least in 1-D, by several techniques.
The first technique presented in this section relies on a change of variables, and is rather specific

to the wave equation. The two other techniques are widely used for computing analytical solutions of time-dependent linear PDEs: separation of variables, in the case where the domain is bounded (already seen for the Laplace equation), and Fourier transform w.r.t. space variables, in the case where the domain is $\mathbb{R}^n$.

6.2.1 1-D solution: change of variables

The 1-D wave equation reads $\dfrac{\partial^2 u}{\partial t^2}(x,t) - c^2 \dfrac{\partial^2 u}{\partial x^2}(x,t) = 0.$

Let $p(x,t) = \dfrac{1}{c}\dfrac{\partial u}{\partial t}$ and $q(x,t) = \dfrac{\partial u}{\partial x}$. The wave equation can then be rewritten as

$$\frac{\partial p}{\partial t} - c\frac{\partial q}{\partial x} = 0 \tag{6.1}$$

while the link between p and q is:

$$\frac{\partial q}{\partial t} - c\frac{\partial p}{\partial x} = 0 \tag{6.2}$$

Adding and substracting (6.1) and (6.2) leads to the two following transport equations:

$$\begin{cases} \dfrac{\partial (p-q)}{\partial t} + c\dfrac{\partial (p-q)}{\partial x} = 0 \\ \dfrac{\partial (p+q)}{\partial t} - c\dfrac{\partial (p+q)}{\partial x} = 0 \end{cases}$$

Hence, following §5.2.3, $(p+q)(x,t) = A(x+ct)$ and $(p-q)(x,t) = B(x-ct)$. This leads to

$$\begin{cases} p(x,t) = \dfrac{1}{c}\dfrac{\partial u}{\partial t} = \dfrac{1}{2}[A(x+ct) + B(x-ct)] \\ q(x,t) = \dfrac{\partial u}{\partial x} = \dfrac{1}{2}[A(x+ct) - B(x-ct)] \end{cases}$$

which yields

$$u(x,t) = F(x+ct) + G(x-ct) \qquad \text{with } F, G \in \mathcal{C}^2 \tag{6.3}$$

This means that the solution is made of the superposition of two travelling waves, one propagating to the right at velocity c and the other one to the left at velocity $-c$ (cf Figure 6.1). Similarly to linear first-order PDEs, lines $x - ct = constant$ and $x + ct = constant$, where $F(x+ct)$ and $G(x-ct)$ are constant, are called *characteristic lines*.

Note that the same result can be obtained by the change of variables $X = x + ct, Y = x - ct$, which transforms the wave equation into $\dfrac{\partial^2 U}{\partial X\, \partial Y}(X,Y) = 0$, where $U(X,Y) = u(x,t)$.

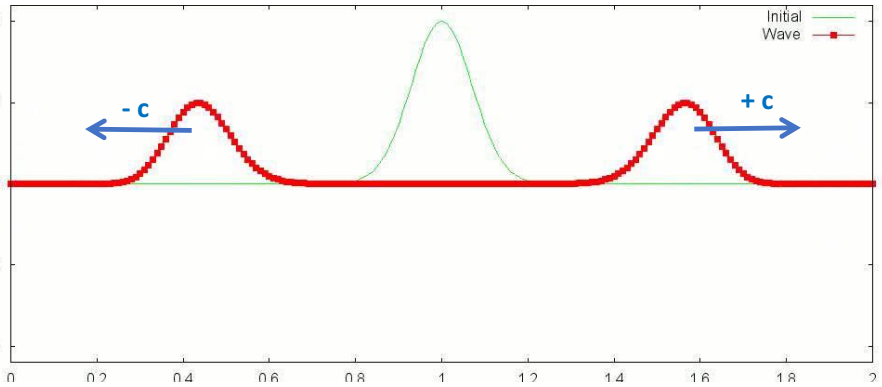

Figure 6.1: A solution of the 1-D wave equation. Green curve: initial condition; red curve: solution a while later.

6.2.2 1-D d'Alembert solution

Let complement the 1-D wave equation in $\mathbb{R}$ with initial conditions:

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2}(x,t) - c^2 \dfrac{\partial^2 u}{\partial x^2}(x,t) = 0 \qquad x \in \mathbb{R}, t > 0 \\ u(x,0) = u_0(x) \qquad u_0 \in \mathcal{C}^2(\mathbb{R}) \\ \dfrac{\partial u}{\partial t}(x,0) = v_0(x) \qquad v_0 \in \mathcal{C}^1(\mathbb{R}) \end{cases}$$

Thus, since $u(x,t) = F(x+ct) + G(x-ct)$ (from (6.3)), the initial conditions yield $F(x) + G(x) = u_0(x)$ and $c(F'(x) - G'(x)) = v_0(x)$. Integrating the last equation and combining it with the first one provides the expressions for F and G. Hence the so-called d'Alembert solution:

$$u(x,t) = \frac{1}{2}\left[u_0(x+ct) + u_0(x-ct)\right] + \frac{1}{2c}\int_{x-ct}^{x+ct} v_0(s)\, ds \tag{6.4}$$

This expression clearly shows that the domain of dependence of $u(x,t)$ is the whole interval $[x-ct; x+ct]$ (i.e. $u(x,t)$ depends on the initial conditions on $[x-ct; x+ct]$).

The same kind of approach can be generalized to higher dimensions, but leading to much more complex analytical expressions for $u(\mathbf{x},t)$.

6.2.3 Analytical solutions through Fourier transform

Let now go back to the n-D case in $\mathbb{R}^n$ with initial conditions

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2}(\mathbf{x},t) - c^2\, \Delta u(\mathbf{x},t) = 0 \qquad \mathbf{x} \text{ in } \mathbb{R}^n, t > 0 \\ u(\mathbf{x},0) = u_0(\mathbf{x})\ , \ \dfrac{\partial u}{\partial t}(\mathbf{x},0) = v_0(\mathbf{x}) \quad \mathbf{x} \text{ in } \mathbb{R}^n \end{cases} \tag{6.5}$$

with $\mathbf{x} \to u(\mathbf{x},t) \in \mathcal{L}^2(\mathbb{R}^n)\ \forall t$. A Fourier transform (see §B.2) of this problem in all spatial variables directly leads to the ODE

$$\begin{cases} \dfrac{\partial^2 \widehat{u}}{\partial t^2}(\boldsymbol{\xi},t) + 4\pi^2 c^2 \|\boldsymbol{\xi}\|^2 \widehat{u}(\boldsymbol{\xi},t) = 0 & \boldsymbol{\xi} \text{ in } \mathbb{R}^n, t>0 \\ \widehat{u}(\boldsymbol{\xi},0) = \widehat{u}_0(\boldsymbol{\xi})\ , \ \dfrac{\partial \widehat{u}}{\partial t}(\boldsymbol{\xi},0) = \widehat{v}_0(\boldsymbol{\xi}) & \boldsymbol{\xi} \text{ in } \mathbb{R}^n \end{cases}$$

which solution is $\widehat{u}(\boldsymbol{\xi},t) = A(\boldsymbol{\xi})\, e^{2i\pi c\|\boldsymbol{\xi}\|t} + B(\boldsymbol{\xi})\, e^{-2i\pi c\|\boldsymbol{\xi}\|t}$. $A(\boldsymbol{\xi})$ and $B(\boldsymbol{\xi})$ are obtained thanks to the initial conditions:

$$\begin{cases} A(\boldsymbol{\xi}) = \dfrac{\widehat{u}_0(\boldsymbol{\xi})}{2} + \dfrac{\widehat{v}_0(\boldsymbol{\xi})}{4i\pi c\,\|\boldsymbol{\xi}\|} \\ B(\boldsymbol{\xi}) = \dfrac{\widehat{u}_0(\boldsymbol{\xi})}{2} - \dfrac{\widehat{v}_0(\boldsymbol{\xi})}{4i\pi c\,\|\boldsymbol{\xi}\|} \end{cases}$$

Hence the solution by inverse Fourier transform.

In the 1-D case, one retrieves of course the d'Alembert solution (6.4).

6.2.4 Case of a bounded domain: separation of variables

We consider now the wave equation in a bounded domain $\Omega \subset \mathbb{R}^n$. As seen previously, if the domain is of infinite size, the solution $u(\mathbf{x},t)$ depends on the initial conditions in the ball $\mathcal{B}(\mathbf{x},ct)$. However, if the domain is bounded, this ball no longer entirely belongs to Ω for t sufficiently large. Moreover, additional boundary conditions are required.

A way to get the analytical solution in this case, at least for domains with simple geometry, is to use a separation of variables technique. Let the wave equation with null Dirichlet boundary conditions. The initial conditions will be provided later. This problem reads:

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2}(\mathbf{x},t) - c^2\,\Delta u(\mathbf{x},t) = 0 & \mathbf{x} \text{ in } \Omega, t>0 \\ u(\mathbf{x},t) = 0 & \mathbf{x} \text{ on } \partial\Omega, t>0 \end{cases} \tag{6.6}$$

Looking for a solution under the form $u(\mathbf{x},t) = X(\mathbf{x})\,T(t)$ leads to

$$\begin{cases} \dfrac{1}{c^2}\dfrac{T''(t)}{T(t)} = \dfrac{\Delta X(\mathbf{x})}{X(\mathbf{x})} = \lambda & \forall \mathbf{x} \text{ in } \Omega, \forall t>0 \\ X(\mathbf{x}) = 0 & \mathbf{x} \text{ on } \partial\Omega \end{cases}$$

The possible values for λ are thus the eigenvalues of the Laplacian operator on Ω with null Dirichlet boundary conditions. As explained in Appendix C, there is a countable set of such eigenvalues, which are all negative, and denoted $-\omega_k^2$, $k \in \mathbb{N}$. The associated eigenfunctions X_k form an orthonormal basis of $\mathcal{L}^2(\Omega)$.

For a given k, the corresponding function $T_k(t)$ satisfies $T_k''(t) + c^2\omega_k^2\,T_k(t) = 0$, i.e. $T_k(t) = \alpha_k \cos(c\,\omega_k t) + \beta_k \sin(c\,\omega_k t)$. We have thus built a family of functions $X_k(\mathbf{x})T_k(t)$ which are

elementary solutions of (6.6). This PDE being linear, any linear combination of these elementary solutions is also a solution. Therefore a general solution of (6.6) reads:

$$u(\mathbf{x},t) = \sum_k X_k(\mathbf{x}) \; [\; \alpha_k \cos(c\,\omega_k t) + \beta_k \sin(c\,\omega_k t)]$$

Adding initial conditions will then determine the α_ks and β_ks.

Let apply this technique in the 1-D case, for $\Omega = (0, L)$. The equation reads:

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2}(x,t) - c^2 \dfrac{\partial^2 u}{\partial x^2}(x,t) = 0 \quad x \in (0,L), t > 0 \\ u(0,t) = u(L,t) = 0 \quad t > 0 \\ u(x,0) = u_0(x), \quad \dfrac{\partial u}{\partial t}(x,0) = v_0(x) \qquad x \in (0,L) \end{cases}$$

The separation of variables yields $X_k''(x) + \omega_k^2 X_k(x) = 0$ with $X_k(0) = X_k(L) = 0$. Hence $X_k(x) = \sin \dfrac{k\pi x}{L}$, and $u(\mathbf{x},t) = \displaystyle\sum_{k=1}^{\infty} \sin \frac{k\pi x}{L} \left[\alpha_k \cos \frac{k\pi ct}{L} + \beta_k \sin \frac{k\pi ct}{L} \right]$.

The initial conditions imply $u_0(x) = \displaystyle\sum_{k=1}^{\infty} \alpha_k \sin \frac{k\pi x}{L}$ and $v_0(x) = \displaystyle\sum_{k=1}^{\infty} \beta_k \frac{k\pi c}{L} \sin \frac{k\pi x}{L}$. These expressions can be identified with expansions into Fourier series of $2L$-periodic odd functions coinciding with u_0 and v_0 on $(0, L)$. Hence

$$\alpha_k = \frac{2}{L} \int_0^L u_o(x) \sin \frac{k\pi x}{L} \quad \text{and} \quad \beta_k = \frac{2}{k\pi c} \int_0^L v_o(x) \sin \frac{k\pi x}{L} \qquad k = 1, 2, \ldots$$

6.3 Discretization schemes

6.3.1 Second-order standard explicit scheme

A simple standard explicit scheme for the 1-D wave equation is the following:

$$\frac{u_j^{n+1} - 2u_j^n + u_j^{n-1}}{\delta t^2} - c^2 \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\delta x^2} = 0 \tag{6.7}$$

It is obviously second-order accurate both in time and space. Its stability condition can be easily computed by the Fourier method, and is of CFL type: $|c|\dfrac{\delta t}{\delta x} \leq 1$.

The equivalent PDE of (6.7) (see Appendix D.3) is

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} + \frac{c^2\,\delta x^2}{12} \left(c^2 \frac{\delta t^2}{\delta x^2} - 1 \right) \frac{\partial^4 u}{\partial x^4} + \ldots = 0$$

The stability condition corresponds to a negative coefficient for the dominant error term, which effect is mainly to slow down the wave propagation for low frequencies, but also to modify their amplitude for high frequencies (see equations (D.4)-(D.5)).

6.3.2 Lax-Wendroff scheme

One can also use the structure of (6.1) and (6.2), and apply a Lax-Wendroff discretization to both (cf §5.3.2). This reads

$$\begin{cases} \dfrac{p_j^{n+1}-p_j^n}{\delta t} - c\,\dfrac{q_{j+1}^n - q_{j-1}^n}{2\delta x} - \dfrac{\delta t}{2}\,c^2\,\dfrac{p_{j+1}^n - 2p_j^n + p_{j-1}^n}{\delta x^2} = 0 \\ \dfrac{q_j^{n+1}-q_j^n}{\delta t} - c\,\dfrac{p_{j+1}^n - p_{j-1}^n}{2\delta x} - \dfrac{\delta t}{2}\,c^2\,\dfrac{q_{j+1}^n - 2q_j^n + q_{j-1}^n}{\delta x^2} = 0 \end{cases} \tag{6.8}$$

These schemes are second-order accurate both in time and space. Then u_j^n can be obtained, with the same accuracy, for instance by defining

$$p_j^n = \frac{1}{c}\,\frac{u_j^{n+1} - u_j^{n-1}}{2\delta t}$$

Hence $u_j^{n+1} = u_j^{n-1} + 2c\,\delta t\,p_j^n$.

The stability of (6.8) can be investigated by a Fourier analysis. Let $p_j^n = P_n e^{ipj\,\delta x}$ and $q_j^n = Q_n e^{ipj\,\delta x}$. Then (6.8) reads

$$\begin{pmatrix} p_j^{n+1} \\ q_j^{n+1} \end{pmatrix} = \begin{pmatrix} 1 - C^2(1-\cos(p\delta x)) & iC\sin(p\delta x) \\ iC\sin(p\delta x) & 1 - C^2(1-\cos(p\delta x)) \end{pmatrix} \begin{pmatrix} p_j^n \\ q_j^n \end{pmatrix}$$

which implies that

$$\begin{pmatrix} p_j^n \\ q_j^n \end{pmatrix} = A^n \begin{pmatrix} p_j^0 \\ q_j^0 \end{pmatrix} \qquad \text{with } A = \begin{pmatrix} 1 - C^2(1-\cos(p\delta x)) & iC\sin(p\delta x) \\ iC\sin(p\delta x) & 1 - C^2(1-\cos(p\delta x)) \end{pmatrix}$$

Thus the scheme is stable iff the spectral radius of A is less or equal to 1, which leads after a few calculations to the same stability condition as the previous scheme: $|c|\dfrac{\delta t}{\delta x} \leq 1$.

Chapter 7

The diffusion equation

This chapter focuses on the diffusion equation $\dfrac{\partial u}{\partial t}(\mathbf{x},t) - \text{div}\,(\nu(\mathbf{x},t)\,\nabla u(\mathbf{x},t)) = f(\mathbf{x},t)$, which is a prototype for linear parabolic second-order PDEs.

7.1 Physical interpretation

Let a scalar quantity (heat, mass of some given chemical species...). Its conservation in a domain ω reads

$$\frac{d}{dt}\int_\omega u(\mathbf{x},t)\,d\mathbf{x} = \underbrace{\int_\omega f(\mathbf{x},t)\,d\mathbf{x}}_{\text{internal sources/sinks}} + \underbrace{\int_{\partial\omega} \phi(\sigma,t)\,d\sigma}_{\text{input/output fluxes}}$$

where u is its corresponding volumetric variable (volumetric enthalpy, in $J.m^{-3}$, for heat, volumetric mass, in $Kg.m^{-3}$...).
A common physical assumption (Fourier's law in the context of heat conservation, Fick's law in the context of the conservation of a chemical species) states that $\phi = \nu\,\nabla u.\mathbf{n}$ where $\mathbf{n}$ is the outward normal vector to $\partial\omega$. ν is called **diffusivity** or **diffusion coefficient**, and has positive values. Using this assumption, performing an integration by parts, and making the size of ω tend to zero leads to

$$\frac{\partial u}{\partial t}(\mathbf{x},t) - \text{div}\,(\nu(\mathbf{x},t)\,\nabla u(\mathbf{x},t)) = f(\mathbf{x},t)$$

In the particular case where ν is a constant, this equation reduces to

$$\frac{\partial u}{\partial t}(\mathbf{x},t) - \nu\,\Delta u(\mathbf{x},t) = f(\mathbf{x},t)$$

7.2 Analytical solutions in n-D

The diffusion equation can be solved analytically, at least for simple domain geometries. This allows for highlighting some properties of the solution, which are actually more general. As already seen in the case of the wave equation, two techniques are widely used for computing analytical solutions of time-dependent linear PDEs. The first one is the separation of variables, in the case

where the domaine is bounded. The second one is the Fourier transform w.r.t. spatial variables, in the case where the domain is $\mathbb{R}^n$.

7.2.1 Diffusion in a bounded domain $\Omega \subset \mathbb{R}^n$

Let $\Omega \subset \mathbb{R}^n$ a bounded domain, and consider the diffusion equation with constant diffusivity, no right-hand side, and null Dirichlet boundary conditions. The initial conditions will be added later. This problem reads:

$$\begin{cases} \dfrac{\partial u}{\partial t}(\mathbf{x},t) - \nu\,\Delta u(\mathbf{x},t) = 0 & \mathbf{x} \text{ in } \Omega, t > 0 \\ u(\mathbf{x},t) = 0 & \mathbf{x} \text{ on } \partial\Omega, t > 0 \end{cases} \tag{7.1}$$

Looking for a solution under the form $u(\mathbf{x},t) = X(\mathbf{x})\,T(t)$ leads to

$$\begin{cases} \dfrac{1}{\nu}\dfrac{T'(t)}{T(t)} = \dfrac{\Delta X(\mathbf{x})}{X(\mathbf{x})} = \lambda & \forall \mathbf{x} \text{ in } \Omega, \forall t > 0 \\ X(\mathbf{x}) = 0 & \mathbf{x} \text{ on } \partial\Omega \end{cases}$$

The possible values for λ are thus the eigenvalues of the Laplacian operator on Ω with null Dirichlet boundary conditions. As explained in Appendix C, there is a countable set of such eigenvalues, which are all negative, and denoted $-\omega_k^2$, $k \in \mathbb{N}$. The associated eigenfunctions X_k form an orthonormal basis of $\mathcal{L}^2(\Omega)$.

For a given k, the corresponding function $T_k(t)$ satisfies $T_k'(t) + \nu\,\omega_k^2\,T(t) = 0$, i.e. $T_k(t) = \alpha_k e^{-\nu\omega_k^2 t}$. We have thus built a family of functions $X_k(\mathbf{x})T_k(t)$ which are elementary solutions of (7.1). This PDE being linear, any linear combination of these elementary solutions is also a solution. Therefore a general solution of (7.1) reads

$$u(\mathbf{x},t) = \sum_k \alpha_k\,X_k(\mathbf{x})\,e^{-\nu\omega_k^2 t}$$

Adding an initial condition will then determine the α_ks. See §7.3.1 for the application of this technique in the 1-D case.

7.2.2 Diffusion in $\mathbb{R}^n$

Let look now for solutions of

$$\begin{cases} \dfrac{\partial u}{\partial t}(\mathbf{x},t) - \nu\,\Delta u(\mathbf{x},t) = 0 & \mathbf{x} \text{ in } \mathbb{R}^n, t > 0 \\ u(\mathbf{x},0) = u_0(\mathbf{x}) & \mathbf{x} \text{ in } \mathbb{R}^n \end{cases} \tag{7.2}$$

with $\mathbf{x} \to u(\mathbf{x},t) \in \mathcal{L}^2(\mathbb{R}^n)\ \forall t$. A Fourier transform (see §B.2) of this problem in all spatial variables directly leads to

$$\begin{cases} \dfrac{\partial \widehat{u}}{\partial t}(\boldsymbol{\xi},t) + 4\pi^2\nu\,\|\boldsymbol{\xi}\|^2\,\widehat{u}(\boldsymbol{\xi},t) = 0 & \boldsymbol{\xi} \text{ in } \mathbb{R}^n, t > 0 \\ \widehat{u}(\boldsymbol{\xi},0) = \widehat{u}_0(\boldsymbol{\xi}) & \boldsymbol{\xi} \text{ in } \mathbb{R}^n \end{cases}$$

which solution is $\widehat{u}(\boldsymbol{\xi},t) = \widehat{u}_0(\boldsymbol{\xi})\, e^{-4\pi^2\|\boldsymbol{\xi}\|^2\nu t}$. Hence the solution by inverse Fourier transform:

$$u(\mathbf{x},t) = (u_0(.) * K(.,t))(\mathbf{x}) \qquad \text{with } K(\mathbf{x},t) = FT^{-1}\left(e^{-4\pi^2\|\boldsymbol{\xi}\|^2\nu t}\right) = \left(2\sqrt{\pi\nu t}\right)^{-n} e^{-\frac{\|\mathbf{x}\|^2}{4\nu t}}$$

See §7.3.2 for the application of this technique in the 1-D case.

7.3 Analytical solutions in 1-D

7.3.1 1-D diffusion in a bounded interval

We consider here the case where the domain of interest is an interval $(0, L)$. As explained in §7.2.1, the problem

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) - \nu \dfrac{\partial^2 u}{\partial x^2}(x,t) = 0 \qquad x \in (0,L), t > 0 \\ u(0,t) = u(L,t) = 0 \quad t > 0 \\ u(x,0) = u_0(x) \quad x \in (0,L) \end{cases}$$

can be solved by a separation of variables technique, i.e. looking for $u(x,t) = X(x)T(t)$. This leads, after some algebra and a Fourier series expansion (see §B.1) of u_0, to

$$u(x,t) = \sum_{k\geq 1} \alpha_k\, e^{-\dfrac{k^2\pi^2\nu t}{L^2}} \sin\frac{k\pi x}{L} \qquad \text{with } \alpha_k = \frac{2}{L}\int_0^L u_0(x) \sin\frac{k\pi x}{L}\, dx$$

7.3.2 1-D diffusion in $\mathbb{R}$

The problem is now

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) - \nu \dfrac{\partial^2 u}{\partial x^2}(x,t) = 0 \qquad x \in \mathbb{R}, t > 0 \\ u(x,0) = u_0(x) \qquad x \in \mathbb{R} \end{cases}$$

If we assume that the Fourier transform of $u(x,t)$ w.r.t. x exists for all t, then this PDE is transformed, as in §7.2.2, into the ODE

$$\begin{cases} \dfrac{\partial \widehat{u}}{\partial t}(\xi,t) + 4\pi^2\xi^2\nu\, \widehat{u}(\xi,t) = 0 \qquad t > 0 \\ \widehat{u}(\xi,0) = \widehat{u}_0(\xi) \end{cases}$$

which solution is $\widehat{u}(\xi,t) = \widehat{u}_0(\xi)\, e^{-4\pi^2\xi^2\nu t}$. An inverse Fourier transform thus yields $u(x,t) = (u_0(.) * K(.,t))(x)$ where $K(x,t) = \dfrac{1}{2\sqrt{\pi\nu\, t}}\, e^{-x^2/4\nu t}$ is the **kernel of the 1-D heat equation**. Hence

$$u(x,t) = \frac{1}{2\sqrt{\pi\nu\, t}} \int_{\mathbb{R}} u_0(s)\, e^{-\dfrac{(x-s)^2}{4\nu\, t}}\, ds$$

Note however that this solution is not unique without further conditions on u, like a boundedness assumption for instance. Other solutions may indeed exist, for which a Fourier transform is not allowed (which implies that the preceding calculations do not hold in such cases).

7.3.3 Some properties of the 1-D solution in $\mathbb{R}$

Let work in this paragraph with this preceding 1-D solution in $\mathbb{R}$. It actually illustrates some general properties of solutions of diffusion equations, even if their exact formulations may vary, depending of course of the nature and of the regularity of the diffusivity $\nu(\mathbf{x},t)$, of the initial condition $u_0(\mathbf{x},t)$ and of the physical domain $\Omega \subset \mathbb{R}^n$. Adding a forcing term $f(\mathbf{x},t)$ may also obviously change some of these properties.

Damping of the solution If $u_0 \in L^1(\mathbb{R})$, then $|u(x,t)| \leq \dfrac{1}{2\sqrt{\pi\nu\, t}}\, \|u_0\|_{L^1}$. Thus it is clear that $u(x,t) \to 0$ as $t \to \infty$.

Maximum principle If $u_0 \in L^\infty(\mathbb{R})$, i.e. $\|u_0\|_\infty = \sup\limits_{x\in\mathbb{R}} |u_0(x)\ | < +\infty$, then $|u(x,t)| \leq \ \|u_0\|_\infty$. The extreme values of $u(x,t)$ thus necessarily occur at the initial time, since diffusion progressively damps out the solution.

Positivity If $u_0(x) \geq 0$ (resp. ≤ 0) $\ \forall x \in \mathbb{R}$, then $u(x,t) \geq 0$ (resp. ≤ 0) $\ \forall x \in \mathbb{R}, \forall t > 0$.

Regularity If $u_0 \in \mathcal{C}^0(\mathbb{R})$, then $u(.,t) \in \mathcal{C}^\infty(\mathbb{R})\ \forall t > 0$ (regularizing character of the diffusion equation).

Dependence on initial conditions and propagation speed As can be seen from the analytical expression of $u(x,t)$, any change (even very local) in u_0 results in a change in $u(x,t)$, $\forall x,t$. In other words, the information is immediately transmitted everywhere: its propagation speed is infinite.
Note also that, since an infinite speed does not exist in the real world, this property highlights a limitation in the mathematical diffusion model.

7.3.4 Adding a source term

Let now the problem

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) - \nu \dfrac{\partial^2 u}{\partial x^2}(x,t) = f(x,t) & x \in \mathbb{R}, t > 0 \\ u(x,0) = u_0(x) & x \in \mathbb{R} \end{cases}$$

Using again the Fourier transform, we have to solve the ODE

$$\begin{cases} \dfrac{\partial \widehat{u}}{\partial t}(\xi,t) + 4\pi^2\xi^2\nu\, \widehat{u}(\xi,t) = \widehat{f}(\xi,t) & t > 0 \\ \widehat{u}(\xi,0) = \widehat{u}_0(\xi) \end{cases}$$

This is a linear first-order non homogeneous ODE. Its solution is thus the sum of the general solution of the corresponding homogeneous ODE with a particular solution of the full ODE (see §A.1). The solution of the corresponding homogeneous equation is $\widehat{u}_h(\xi,t) = C(\xi)\, e^{-4\pi^2\xi^2\nu\, t}$ (see

§7.3.2). A particular solution of the ODE with its right-hand side can be obtained by the variation of constants method, looking for a solution that reads: $\widehat{u}_p(\xi,t) = C(\xi,t)\, e^{-4\pi^2\xi^2\nu\, t}$. One gets $C(\xi,t) = \displaystyle\int_0^t \widehat{f}(\xi,\tau)\, e^{4\pi^2\xi^2\nu\,\tau}\, d\tau$. Hence $\widehat{u}_p(\xi,t) = \displaystyle\int_0^t \widehat{f}(\xi,\tau)\, e^{-4\pi^2\xi^2\nu\,(t-\tau)}\, d\tau$. Since $\widehat{u}(\xi,t) = \widehat{u}_h(\xi,t) + \widehat{u}_p(\xi,t)$, and given the initial condition, one finally gets $\widehat{u}(\xi,t) = \widehat{u}_0(\xi)\, e^{-4\pi^2\xi^2\nu\, t} + \widehat{u}_p(\xi,t)$. This leads, by inverse Fourier transform, to:

$$\begin{aligned} u(x,t) &= (u_0(.) * K(.,t))(x) + \int_0^t (f(.,\tau) * K(.,t-\tau))(x)\, d\tau \\ &= (u_0(.) * K(.,t))(x) + \int_0^t \int_{\mathbb{R}} f(s,\tau)\, K(x-s,t-\tau)\, ds\, d\tau \end{aligned}$$

7.3.5 Energy of the solution

Let consider here the more general formulation in $\Omega \subset \mathbb{R}^n$

$$\frac{\partial u}{\partial t}(\mathbf{x},t) - \text{div}\ (\nu(\mathbf{x},t)\, \nabla u(\mathbf{x},t)) = f(\mathbf{x},t)$$

where $\nu(\mathbf{x},t) \geq 0$. Let $E(t) = \dfrac{1}{2}\displaystyle\int_\Omega u^2(\mathbf{x},t)\, d\mathbf{x}$ the energy of the solution. Assuming that u is regular enough, then

$$\frac{dE(t)}{dt} = \frac{1}{2}\int_\Omega \frac{\partial u^2}{\partial t}(\mathbf{x},t)\, d\mathbf{x} = \frac{1}{2}\int_\Omega u(\mathbf{x},t)\, \text{div}\ (\nu(\mathbf{x},t)\, \nabla u(\mathbf{x},t))\ d\mathbf{x} + \frac{1}{2}\int_\Omega u(\mathbf{x},t)\, f(\mathbf{x},t)\, d\mathbf{x}$$

$$= -\frac{1}{2}\int_\Omega \nu(\mathbf{x},t)\, \|\nabla u(\mathbf{x},t)\|^2\, d\mathbf{x} + \frac{1}{2}\int_{\partial\Omega} \nu(\sigma,t)\, u(\sigma,t)\, \frac{\partial u}{\partial \mathbf{n}}(\sigma,t)\, d\sigma + \frac{1}{2}\int_\Omega u(\mathbf{x},t)\, f(\mathbf{x},t)\, d\mathbf{x}$$

Therefore it is clear that, if there is no energy source in the domain ($f = 0$) nor at the boundary ($u = 0$ or $\partial u/\partial \mathbf{n} = 0$ on $\partial\Omega$), then the energy of the solution continuously decreases as time increases.

7.4 Numerical schemes for the diffusion equation

7.4.1 Usual Euler explicit scheme

The simplest discretization of the 1D diffusion equation reads

$$\frac{u_j^{n+1} - u_j^n}{\delta t} - \nu\, \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\delta x^2} = 0$$

It can be rewritten as

$$u_j^{n+1} = \lambda\, u_{j-1}^n + (1-2\lambda)\, u_j^n + \lambda\, u_{j+1}^n \qquad \text{with } \lambda = \nu\, \frac{\delta t}{\delta x^2}$$

- **Stability** A Fourier stability analysis shows that this scheme is stable for $\lambda \leq \frac{1}{2}$. Stability can also be assessed by studying the spectral radius of the matrix of the system (see §4.4.3), which reads for example :

$$\begin{bmatrix} 1-2\lambda & \lambda & 0 & \cdots & 0 \\ \lambda & 1-2\lambda & \lambda & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \lambda \\ 0 & \cdots & 0 & \lambda & 1-2\lambda \end{bmatrix}$$

  in the case of null boundary conditions.

- **Accuracy and equivalent PDE** Its equivalent PDE is

$$\frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} + \left(\nu^2 \frac{\delta t}{2} - \nu \frac{\delta x^2}{12} \right) \frac{\partial^4 u}{\partial x^4} + \ldots = 0$$

  This scheme is first-order accurate in time and second-order accurate in space.

  Since $\eta = \nu^2 \frac{\delta t}{2} - \nu \frac{\delta x^2}{12} = \frac{\nu\, \delta x^2}{2} \left(\lambda - \frac{1}{6} \right) > 0$ iff $\lambda > 1/6$, then, using the analysis presented in Appendix D.4, the main effect of the numerical error is to amplify the solution if $0 < \lambda < 1/6$ and to damp the solution if $1/6 < \lambda < 1/2$.

- **Positivity** If $\lambda \leq 1/2$, then u_j^{n+1} is a convex weighted average of u_{j-1}^n, u_j^n and u_{j+1}^n. Thus, if $u_0(x) \geq 0\ \forall x$, then $u_j^n \geq 0\ \forall j, n$. The positivity property of the exact solution is also satisfied by the numerical solution.

- **Maximum principle** If $\lambda \leq 1/2$, then $|u_j^{n+1}| \leq \lambda\, |u_{j-1}^n| + (1-2\lambda)\, |u_j^n| + \lambda\, |u_{j+1}^n|$.
  Thus $|u_j^{n+1}| \leq (\lambda + (1-2\lambda) + \lambda) \|u^n\|_\infty = \|u^n\|_\infty$.
  Hence $\|u^{n+1}\|_\infty \leq \|u^n\|_\infty \leq \ldots \leq \|u^0\|_\infty$.
  The maximum principle satisfied by the exact solution is also satisfied by the numerical solution.

- **Dependence w.r.t. initial conditions** In the exact solution, any local change in $u_0(x)$ instantaneously affects the whole solution. This property can obviously not be satisfied with such an explicit scheme. Here, u_j^n only depends on $u_l^0, l = j-n, \ldots, j+n$.

7.4.2 Other schemes

Many other schemes are of course available for this diffusion equation, either in 1D or in 2D. Several of them are detailed in the corresponding exercise sheet.

Appendix A

Reminder on linear ODEs

A.1 First-order linear ODEs

Let consider the general linear first-order ordinary differential equation (ODE):

$$a(x)\,u'(x) + b(x)\,u(x) = c(x)$$

on an interval $I \subset \mathbb{R}$ where the functions $a(x), b(x)$ and $c(x)$ are continuous, and where $a(x)$ does not cancel. Dividing by $a(x)$, the equation becomes

$$u'(x) + \alpha(x)\,u(x) = \beta(x) \qquad x \in I \qquad (E)$$

Its so-called associated homogeneous equation is $u_0'(x) + \alpha(x)\,u_0(x) = 0 \qquad (E_0)$.

Theorem A.1. (**Principle of superposition**)
Let $u_p(x)$ a particular solution of (E). The solutions of (E) are the functions $u(x) = u_p(x) + u_0(x)$, where u_0 represents the solutions of (E_0).
In other words, the set of solutions of (E) on I is $\mathcal{S} = u_p + \mathcal{S}_0$, where $\mathcal{S}_0$ denotes the set of solutions of (E_0).

Theorem A.2. (**Solutions of the homogeneous equation**)
The solutions of (E_0) on I are the functions $u_0(x) = K\,e^{-A(x)},\ K \in \mathbb{R}$, where $A(x)$ is a primitive of $\alpha(x)$.

Proof Let $A(x)$ a primitive of $\alpha(x)$.
Multiplying (E_0) by $e^{A(x)}$, one gets $u_0'(x)\,e^{A(x)} + \alpha(x)\,e^{A(x)}\,u_0(x) = 0$, i.e. $\left(u_0(x)\,e^{A(x)}\right)' = 0$.
This is equivalent to $u_0(x)\,e^{A(x)} = K,\ K \in \mathbb{R}$, i.e. $u_0(x) = K\,e^{-A(x)},\ K \in \mathbb{R}$. □

Regarding u_p, a particular solution of (E), it can be obtained either by analogy or by the method of variation of constants:

- **Analogy** If the right-hand side $\beta(x)$ is a polynomial, or a linear combination of exponentials, or a linear combination of sine and cosine functions, and if $\alpha(x)$ is constant or of similar nature as $\beta(x)$, then it may exist a particular solution $u_p(x)$ in a form similar to that of $\beta(x)$. This method does not work systematically, but in general it is worth trying, since it is very simple.

- **Variation of constants** Contrary to the approach by analogy, this method always works, but is a little bit more demanding in terms of calculations. It consists in looking for u_p under the form $u_p(x) = K(x)\, e^{-A(x)}$. One then gets $K'(x) = \beta(x)\, e^{A(x)}$. Hence $K(x)$ by integration, and then $u_p(x)$.

A.2 Second-order linear ODEs with constant coefficients

We consider here the equation

$$au''(x) + bu'(x) + cu(x) = f(x) \qquad (E)$$

with $a, b, c \in \mathbb{R}$ and $a \neq 0$.

The principle of superposition still holds. So the set of solutions of (E) is $\mathcal{S} = u_p + \mathcal{S}_0$, where $u_p(x)$ is a particular solution of (E) and $\mathcal{S}_0$ denotes the set of solutions of the associated homogeneous equation (E_0).

The so-called *characteristic polynomial* associated to (E_0) is $aX^2 + bX + c$. Let denote $\Delta = b^2 - 4ac$ its discriminant. Then:

- if $\Delta > 0$, $u_0(x) = Ae^{r_1 x} + Be^{r_2 x} \quad A, B \in \mathbb{R}$, where $r_1 = \frac{-b-\sqrt{\Delta}}{2a}$ and $r_2 = \frac{-b+\sqrt{\Delta}}{2a}$ are the two real roots of the characteristic polynomial.
- if $\Delta = 0$, $u_0(x) = (Ax + B)e^{rx} \quad A, B \in \mathbb{R}$, where $r = \frac{-b}{2a}$ is the unique root of the characteristic polynomial.
- if $\Delta < 0$, $u_0(x) = (A\cos \alpha x + B \sin \alpha x)e^{\beta x} \quad A, B \in \mathbb{R}$, where $\alpha = \frac{\sqrt{-\Delta}}{2a}$ and $\beta = \frac{-b}{2a}$.

Similarly to the case of first-order equations, a particular solution u_p can be obtained either by analogy with the right-hand side (if simple), or by the method of variation of constants (i.e. replacing the constants A and B, or only one of them, by a function of x).

Appendix B

Reminder on Fourier series and Fourier transforms

B.1 Fourier series expansion

Let f a integrable and periodic function, with period L. One can then define:

$$F(x) = a_0 + \sum_{k\geq 1} \left(a_k \cos\frac{2\pi kx}{L} + b_k \sin\frac{2\pi kx}{L} \right)$$

$$\text{with } a_0 = \frac{1}{L}\int_0^L f(x)\,dx\ , \quad a_k = \frac{2}{L}\int_0^L f(x)\cos\frac{2\pi kx}{L}\,dx\ , \quad b_k = \frac{2}{L}\int_0^L f(x)\sin\frac{2\pi kx}{L}\,dx$$

F is the so-called **Fourier series expansion** of f.

This expansion also reads

$$F(x) = \sum_{k=-\infty}^{+\infty} c_k\, e^{\frac{2i\pi kx}{L}} \qquad \text{with } c_k = \frac{1}{L}\int_0^L f(x)\, e^{-\frac{2i\pi kx}{L}}\,dx$$

- If f is an even function, $b_k = 0 \quad \forall k \geq 1 \quad$ (i.e. $c_k = c_{-k} \quad \forall k$)
- If f is an odd function, $a_k = 0 \quad \forall k \geq 0 \quad$ (i.e. $c_k = -c_{-k} \quad \forall k$)

Theorem B.1. (**Pointwise convergence**)
If f is $\mathcal{C}^1(0, L)$, then $F = f$ (note that some similar results exist which require less regularity for f)

Theorem B.2. (**Parseval's equality**, or conservation of energy)
If $f \in \mathcal{L}^2(0, L)$, then

$$\frac{1}{L}\int_0^L |f(x)|^2\,dx = a_0^2 + \frac{1}{2}\sum_{k=1}^{+\infty}(a_k^2 + b_k^2) = \sum_{k=-\infty}^{+\infty} |c_k|^2$$

B.2 Fourier transform

Let f integrable on $\mathbb{R}$. The **Fourier transform** of f is

$$FT[f](\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}} f(x)\, e^{-2i\pi\xi x}\, dx$$

and the **inverse Fourier transform** of $\widehat{f}$ is

$$FT^{-1}[\widehat{f}](x) = \int_{\mathbb{R}} \widehat{f}(\xi)\, e^{2i\pi\xi x}\, d\xi$$

Some properties of the Fourier transfom

- If $f \in \mathcal{C}^1(\mathbb{R})$ and if $\widehat{f}$ is $L^1(\mathbb{R})$, then $FT^{-1}[\widehat{f}] = f$ (reciprocity of the Fourier transform)
- $\widehat{f}\,\widehat{g} = \widehat{f * g}$ Reminder: convolution product $(a * b)(x) = \int_{\mathbb{R}} a(y)\, b(x-y)\, dy$
- $\widehat{f} * \widehat{g} = \widehat{fg}$
- $\widehat{f}'(\xi) = 2i\pi\xi\, \widehat{f}(\xi)$
- If $g(x) = f(x - x_0)$, then $\widehat{g}(\xi) = e^{-2i\pi x_0 \xi}\, \widehat{f}(\xi)$
- The Fourier transform of the **Gaussian function** $\exp(-\pi\alpha x^2)$ is the Gaussian function $\dfrac{1}{\sqrt{\alpha}} \exp\left(-\dfrac{\pi}{\alpha}\xi^2\right)$
- The Fourier transform of the **gate function** $\Pi(x) = 1$ for $x \in (-1/2; 1/2)$ and 0 elsewhere is $\mathrm{sinc}(\pi\xi)$ where sinc is the **cardinal sine function** defined by $\mathrm{sinc}\, a = (\sin a)/a$.

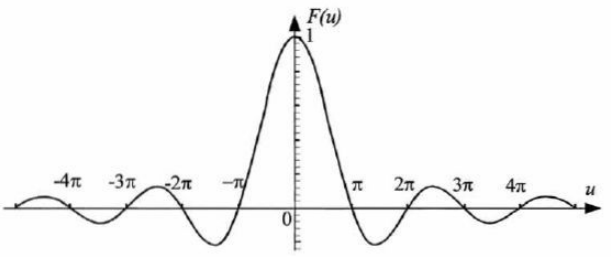

Figure B.1: Plot of the cardinal sine function $\mathrm{sinc}\, u = \dfrac{\sin u}{u}$

Theorem B.3. (**Parseval's equality**, or conservation of energy)
If $f \in \mathcal{L}^2(\mathbb{R})$, then

$$\int_{\mathbb{R}} |f(x)|^2\, dx = \int_{\mathbb{R}} \left|\widehat{f}(\xi)\right|^2\, d\xi$$

These definitions and properties of the Fourier transform in $\mathbb{R}$ can be directly generalized to $\mathbb{R}^n$.

Appendix C

The Laplacian operator and its spectrum

C.1 General results

Let $\Omega \subset \mathbb{R}^n$ a bounded domain, and consider the following eigenvalue problem:

$$\begin{cases} \Delta X(\mathbf{x}) = \sum_{i=1}^{n} \dfrac{\partial^2 X}{\partial x_i^2}(x_1, \ldots, x_n) = \lambda\, X(x_1, \ldots, x_n) & \mathbf{x} \in \Omega \\ X(\mathbf{x}) = 0 \quad \text{on } \partial\Omega \end{cases}$$

Theorem C.1. There is a countable set of eigenvalues, all of them being negative: $\lambda_k = -\omega_k^2$. The corresponding eigenfunctions $X_k(\mathbf{x})$ form an orthonormal basis of $\mathcal{L}^2(\Omega)$.

The proof of this theorem can be found for instance in Evans (1998). It is quite long and technical, and is out of the scope of these notes, but note that two aspects at least are obvious:

- All eigenvalues are negative:

  $$(\Delta X - \lambda X = 0) \Longrightarrow \int_\Omega X \Delta X = -\int_\Omega \|\nabla X\|^2 = \lambda \int_\Omega X^2$$

  Hence

  $$\lambda = -\frac{\int_\Omega \|\nabla X\|^2}{\int_\Omega X^2} \leq 0$$

- Eigenfunctions associated to different eigenvalues are orthogonal:

  Let X_k and X_l two eigenfunctions associated to two different eigenvalues $-\omega_k^2$ and $-\omega_l^2$.

  $$\begin{cases} \Delta X_k + \omega_k^2 X_k = 0 & \Longrightarrow \int_\Omega \Delta X_k\, X_l + \omega_k^2 \int_\Omega X_k\, X_l = -\int_\Omega \nabla X_k\, \nabla X_l + \omega_k^2 \int_\Omega X_k\, X_l = 0 \\ \Delta X_l + \omega_l^2 X_l = 0 & \Longrightarrow \int_\Omega \Delta X_l\, X_k + \omega_l^2 \int_\Omega X_l\, X_k = -\int_\Omega \nabla X_l\, \nabla X_k + \omega_l^2 \int_\Omega X_l\, X_k = 0 \end{cases}$$

  Making the difference between those two equations yields $(\omega_k^2 - \omega_l^2) \int_\Omega X_l\, X_k = 0$, hence $\int_\Omega X_l\, X_k = 0$. Note that this also implies $\int_\Omega \nabla X_l\, \nabla X_k = 0$. X_k and X_l are orthogonal both in $L^2(\Omega)$ and in $H^1(\Omega)$.

C.2 The 1-D case

In the 1-D case, let consider $\Omega = (0, L)$. The eigenvalue problem reads

$$\begin{cases} X''(x) = \lambda X(x) & x \in (0, L) \\ X(0) = X(L) = 0 \end{cases}$$

As previously, λ is negative and can be written $\lambda = -\omega^2$. Hence $X''(x) + \omega^2 X(x) = 0$, which yields $X(x) = \alpha \sin \omega x + \beta \cos \omega x$. $X(0) = 0$ implies $\beta = 0$, while $X(L) = 0$ implies $\alpha \sin \omega L = 0$. Non zero solutions are then obtained for

$$\omega_k = \frac{k\pi}{L} \quad \text{and} \quad X_k(x) = \sin \frac{k\pi x}{L} \quad , k \in \mathbb{N}$$

Appendix D

Some generic calculations related to finite difference schemes

D.1 Fourier analysis: computation of transfer functions and stability studies

Computing both transfer functions (§2.2.4) and stability criteria (§4.4.2) requires some repetitive calculations involving complex exponentials. Some generic formula are given below, in order to facilitate these computations.

Transfer functions With the same notations as in §2.2.4:

Scheme	Transfer function	
$\frac{u(x-h)+u(x+h)}{2}$	$\cos\omega$	
$\frac{u(x+h)-u(x)}{h}$	$e^{i\omega}-1$	$= e^{i\omega/2}\, 2i\, \sin(\omega/2)$
$\frac{u(x)-u(x-h)}{h}$	$1-e^{-i\omega}$	$= e^{-i\omega/2}\, 2i\, \sin(\omega/2)$
$\frac{u(x+h)-u(x-h)}{2h}$	$i\,\sin\omega$	
$\frac{u(x+h)-2u(x)+u(x-h)}{h^2}$	$2\,(\cos\omega-1)$	

Note that these results can be easily adapted to slightly different schemes. For instance, if the scheme $S_h = \frac{\cdots}{h^p}$ has a transfer function $T(\omega)$, then the transfer function of $S_{\lambda h}$ is $\frac{1}{\lambda^p}\,T(\lambda\omega)$. This is useful typically for $\lambda = 2$ or $\lambda = 1/2$.

APPENDIX D. SOME GENERIC CALCULATIONS RELATED TO FINITE DIFFERENCE SCHEMES

Amplification factors With usual notations, replacing u_j^n by $\xi^n e^{ipj\delta x}$ in a numerical scheme, one will obtain this term $\xi^n e^{ipj\delta x}$ multiplied by an amplification factor.

Scheme	Amplification factor	
$\dfrac{u_j^{n-1}+u_j^{n+1}}{2}$	$\dfrac{1/\xi+\xi}{2}$	$=\dfrac{\xi^2+1}{2\,\xi}$
$\dfrac{u_j^{n+1}-u_j^{n}}{\delta t}$	$\dfrac{\xi-1}{\delta t}$	
$\dfrac{u_j^{n+1}-u_j^{n-1}}{2\,\delta t}$	$\dfrac{\xi-1/\xi}{2\,\delta t}$	$=\dfrac{\xi^2-1}{2\,\xi\,\delta t}$
$\dfrac{u_j^{n+1}-2u_j^{n}+u_j^{n-1}}{\delta t^2}$	$\dfrac{\xi-2+1/\xi}{\delta t^2}$	$=\dfrac{(\xi-1)^2}{\xi\,\delta t^2}$
$\dfrac{u_{j-1}^{n}+u_{j+1}^{n}}{2}$	$\cos(p\,\delta x)$	
$\dfrac{u_{j+1}^{n}-u_j^{n}}{\delta x}$	$\dfrac{e^{ip\,\delta x}-1}{\delta x}$	$=e^{ip\,\delta x/2}\,\dfrac{2i\,\sin(p\delta x/2)}{\delta x}$
$\dfrac{u_j^{n}-u_{j-1}^{n}}{\delta x}$	$\dfrac{1-e^{-ip\,\delta x}}{\delta x}$	$=e^{-ip\,\delta x/2}\,\dfrac{2i\,\sin(p\delta x/2)}{\delta x}$
$\dfrac{u_{j+1}^{n}-u_{j-1}^{n}}{2\,\delta x}$	$\dfrac{i\,\sin(p\,\delta x)}{\delta x}$	
$\dfrac{u_{j+1}^{n}-2u_j^{n}+u_{j-1}^{n}}{\delta x^2}$	$2\,\dfrac{\cos(p\,\delta x)-1}{\delta x^2}$	

Note that these results can be easily adapted to slightly different schemes, for instance by multiplying the amplification factor by ξ if the scheme is at time $n+1$ instead of n, or by replacing δx by $2\delta x$ if the scheme involves $j-2$ and $j+2$ instead of $j-1$ and $j+1$.

Examples *Below are two applications of the previous tables.*

- *Let consider the Dufort-Frankel scheme for the 1-D diffusion equation:*

$$\frac{u_j^{n+1}-u_j^{n-1}}{2\delta t}-\nu\,\frac{u_{j+1}^{n}-2\,\dfrac{u_j^{n-1}+u_j^{n+1}}{2}+u_{j-1}^{n}}{\delta x^2}=0$$

Using the tables, replacing u_j^n by $\xi^n e^{ipj\delta x}$ immediatly leads to:

$$\frac{\xi^2-1}{2\,\xi\,\delta t}-\frac{\nu}{\delta x^2}\left(2\cos(p\,\delta x)-2\,\frac{\xi^2+1}{2\,\xi}\right)=0$$

i.e.

$$(2\lambda+1)\xi^2 - 4\lambda\cos(p\,\delta x)\,\xi + (2\lambda-1) = 0 \qquad \textit{with } \lambda = \frac{\nu\,\delta t}{\delta x^2}$$

One has then to study the modulus of ξ to prove the stability of the scheme.

- *Let consider the following scheme for the 1-D transport equation:*

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\delta t} + c\left(\frac{4}{3}\,\frac{u_{j+1}^{n+1} - u_{j-1}^{n+1}}{2\,\delta x} - \frac{1}{3}\,\frac{u_{j+2}^{n} - u_{j-2}^{n}}{4\,\delta x}\right) = 0$$

Using the tables, replacing u_j^n by $\xi^n e^{ipj\delta x}$ directly leads to:

$$\frac{\xi^2 - 1}{2\,\xi\,\delta t} + c\,\xi\left(\frac{4}{3}\,\frac{i\,\sin(p\,\delta x)}{\delta x} - \frac{1}{3}\,\frac{i\,\sin(2p\,\delta x)}{2\delta x}\right) = 0$$

Hence
$$\xi^2 = \left[1 + i\,\frac{2c\,\delta t}{3\,\delta x}\left(4\sin(p\,\delta x) - \frac{1}{2}\sin(2p\,\delta x)\right)\right]^{-1}$$

which implies that $|\xi| \leq 1$.

D.2 Small o and big O

Definition D.1. $f(x) = o(x^p)$ (pronounce *small o*) in the vicinity of 0 iff $f(x) = x^p\,\varepsilon(x)$ with $\varepsilon(x) \to 0$ as $x \to 0$. In other words, $f(x)$ is negligible w.r.t. x^p in the vicinity of 0.

Definition D.2. $f(x) = \mathcal{O}(x^p)$ (pronounce *big o*) in the vicinity of 0 iff there exists two positive constants α and β such that $\alpha|x|^p \leq |f(x)| \leq \beta|x|^p$ in a neighborhood of 0. In other words, $f(x)$ is of the same order as x^p in the vicinity of 0.

D.3 Computation of equivalent PDEs

Computing equivalent PDEs requires linear combinations of Taylor expansions. Some formulas corresponding to frequently used schemes are given below, in order to facilitate these computations.

APPENDIX D. SOME GENERIC CALCULATIONS RELATED TO FINITE DIFFERENCE SCHEMES

$$\frac{u(x,t+\delta t)+u(x,t-\delta t)}{2} = u(x,t) + \frac{\delta t^2}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\delta t^4}{24}\frac{\partial^4 u}{\partial t^4}(x,t) + \mathcal{O}(\delta t^6)$$

$$\frac{u(x,t+\delta t)-u(x,t)}{\delta t} = \frac{\partial u}{\partial t}(x,t) + \frac{\delta t}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\delta t^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) + \mathcal{O}(\delta t^3)$$

$$\frac{u(x,t)-u(x,t-\delta t)}{\delta t} = \frac{\partial u}{\partial t}(x,t) - \frac{\delta t}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\delta t^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) + \mathcal{O}(\delta t^3)$$

$$\frac{u(x,t+\delta t)-u(x,t-\delta t)}{2\,\delta t} = \frac{\partial u}{\partial t}(x,t) + \frac{\delta t^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) + \frac{\delta t^4}{120}\frac{\partial^5 u}{\partial t^5}(x,t) + \mathcal{O}(\delta t^6)$$

$$\frac{u(x,t+\delta t)-2u(x,t)+u(x,t-\delta t)}{\delta t^2} = \frac{\partial^2 u}{\partial t^2}(x,t) + \frac{\delta t^2}{12}\frac{\partial^4 u}{\partial t^4}(x,t) + \frac{\delta t^4}{360}\frac{\partial^6 u}{\partial t^6}(x,t) + \mathcal{O}(\delta t^6)$$

$$\frac{u(x+\delta x,t)+u(x-\delta x,t)}{2} = u(x,t) + \frac{\delta x^2}{2}\frac{\partial^2 u}{\partial x^2}(x,t) + \frac{\delta x^4}{24}\frac{\partial^4 u}{\partial x^4}(x,t) + \mathcal{O}(\delta x^6)$$

$$\frac{u(x+\delta x,t)-u(x,t)}{\delta x} = \frac{\partial u}{\partial x}(x,t) + \frac{\delta x}{2}\frac{\partial^2 u}{\partial x^2}(x,t) + \frac{\delta x^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) + \mathcal{O}(\delta x^3)$$

$$\frac{u(x,t)-u(x-\delta x,t)}{\delta x} = \frac{\partial u}{\partial x}(x,t) - \frac{\delta x}{2}\frac{\partial^2 u}{\partial x^2}(x,t) + \frac{\delta x^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) + \mathcal{O}(\delta x^3)$$

$$\frac{u(x+\delta x,t)-u(x-\delta x,t)}{2\,\delta x} = \frac{\partial u}{\partial x}(x,t) + \frac{\delta x^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) + \frac{\delta x^4}{120}\frac{\partial^5 u}{\partial x^5}(x,t) + \mathcal{O}(\delta x^6)$$

$$\frac{u(x+\delta x,t)-2u(x,t)+u(x-\delta x,t)}{\delta x^2} = \frac{\partial^2 u}{\partial x^2}(x,t) + \frac{\delta x^2}{12}\frac{\partial^4 u}{\partial x^4}(x,t) + \frac{\delta x^4}{360}\frac{\partial^6 u}{\partial x^6}(x,t) + \mathcal{O}(\delta x^6)$$

Example *Let consider the following explicit scheme for the 1-D diffusion equation:*

$$\frac{u_j^{n+1}-u_j^n}{\delta t} - \nu\,\frac{u_{j+1}^n - 2\,u_j^n + +u_{j-1}^n}{\delta x^2} = 0$$

Using the previous table, its equivalent PDE follows:

$$\frac{\partial u}{\partial t}(x,t) + \frac{\delta t}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \mathcal{O}(\delta t^2) - \nu\left(\frac{\partial^2 u}{\partial x^2}(x,t) + \frac{\delta x^2}{12}\frac{\partial^4 u}{\partial x^4}(x,t) + \mathcal{O}(\delta x^4)\right) = 0$$

Since $\dfrac{\partial^2 u}{\partial t^2} = \dfrac{\partial}{\partial t}\left(\dfrac{\partial u}{\partial t}\right) = \dfrac{\partial}{\partial t}\left(\nu\dfrac{\partial^2 u}{\partial x^2}\right) = \nu\,\dfrac{\partial^2}{\partial x^2}\left(\dfrac{\partial u}{\partial t}\right) = \nu^2\,\dfrac{\partial^4 u}{\partial x^4}$, *it becomes:*

$$\frac{\partial u}{\partial t}(x,t) - \nu\frac{\partial^2 u}{\partial x^2}(x,t) + \left(\nu^2\frac{\delta t}{2} - \nu\frac{\delta x^2}{12}\right)\frac{\partial^4 u}{\partial x^4}(x,t) + \mathcal{O}(\delta t^2) + \mathcal{O}(\delta x^4) = 0 \qquad \text{(D.1)}$$

Once an equivalent PDE has been computed, its dominant error term can be interpreted thanks to the following section.

D.4 Interpretation of the effect of the dominant error term

Let consider the generic 1-D PDE

$$\begin{cases} \dfrac{\partial u}{\partial t} + c_1 \dfrac{\partial u}{\partial x} + c_2 \dfrac{\partial^2 u}{\partial x^2} + c_3 \dfrac{\partial^3 u}{\partial x^3} + c_4 \dfrac{\partial^4 u}{\partial x^4} = 0 \qquad x \in \mathbb{R}, t > 0 \\ u_0(x) = e^{ipx} \qquad x \in \mathbb{R} \end{cases} \tag{D.2}$$

Looking for a plane wave solution $u(x,t) = e^{i(px+\chi t)}$ leads almost directly to

$$u(x,t) = \exp\left(\underbrace{i\, p \left[x - (c_1 - p^2 c_3)t\right]}_{\text{phase}} \right) \underbrace{\exp\left(\left[c_2 - c_4 p^2\right] p^2 t\right)}_{\text{amplitude}} \tag{D.3}$$

This general expression can then be used to interpret the effect of the dominant error term of equivalent PDEs of finite difference schemes. As indicated in §2.2.4, a scheme will be said

- **dissipative** if it modifies the amplitude of the wave
- **dispersive** if it modifies the phase (i.e. the velocity) of the wave

Example *Coming back to the preceding example, the equivalent PDE is given by (D.1). It thus corresponds to $c_1 = c_3 = 0$, $c_2 = -\nu$ and $c_4 = \nu^2 \dfrac{\delta t}{2} - \nu \dfrac{\delta x^2}{12}$ in (D.2). Looking now to (D.3), it appears that the dominant error term creates the artificial multiplicative factor $\exp(-c_4 p^4 t)$ with respect to the exact solution. It will thus result in an artificial damping or amplification of the solution, depending on the sign of c_4, i.e. depending on whether $\nu \delta t / \delta x^2$ is greater or larger than $1/6$.*

If we consider similarly the second-order in time equation (relevant to study the effect of dominant error terms in finite difference approximations of the wave equation)

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2 \dfrac{\partial^2 u}{\partial x^2} + \eta \dfrac{\partial^4 u}{\partial x^4} = 0 \qquad x \in \mathbb{R}, t > 0 \\ u_0(x) = e^{ipx} \qquad x \in \mathbb{R} \end{cases} \tag{D.4}$$

its plane wave solutions are

$$u(x,t) = A \exp\left(ip\left[x - (c^2 + \eta p^2)^{1/2} t\right]\right) + B \exp\left(ip\left[x + (c^2 + \eta p^2)^{1/2} t\right]\right) \tag{D.5}$$

References

[1] Allaire G., 2006: *Analyse numérique et optimisation*. Editions de l'Ecole Polytechnique.

[2] Demailly J.-P., 1996: *Analyse numérique et équations différentielles*. Presses Universitaires de Grenoble.

[3] Euvrard D., 1994: *Résolution numérique des équations aux dérivées partielles*. Masson.

[4] Evans L., 1998: *Partial Differential Equations*. Graduate Studies in Mathematics, Vol 19, American Math. Society.

[5] Lascaux P. et R. Théodor, 1986: *Analyse numérique matricielle appliquée à l'art de l'ingénieur*. Masson.

[6] Leveque R. J., 2007: *Finite difference methods for ordinary and partial differential equations*. SIAM.

[7] Mohammadi B. and J.H. Saiac, 2003: *Pratique de l'analyse numérique*. Dunod,

[8] Sainsaulieu L., 1996: *Calcul scientifique*. Masson.

[9] Trefethen L.N., 1996: *Finite difference and spectral methods for ordinary and partial differential equations*. Unpublished text, available on line.