2 Finite difference method

2.1 Introduction

In this chapter we analyse numerical schemes of finite differences. We define the **stability** and **consistency** of a scheme and show that, for linear, constant coefficient, partial differential equations, stability plus consistency of a scheme implies its **convergence**.

The plan of the chapter is the following. Section 2.2 treats the case of the heat equation introduced in Chapter 1. Section 2.3 generalizes the preceding results to the case of the wave equation or the advection equation. One of the aims of this chapter is the construction and analysis of finite differences schemes for much more general models. The reader should not be afraid of extending the concepts presented here to his preferred model and to construct original numerical schemes.

We finish this introduction by saying that the finite difference method is one of the oldest methods of numerical approximation which is still used in applications, such as wave propagation (seismic or electromagnetic) or compressible fluid mechanics. For other applications, such as solid mechanics or incompressible fluids, we often prefer the finite element method. Nevertheless, many concepts in finite differences are found in other numerical methods. Thus, the numerical schemes of Chapter 8 will combine finite elements for the space discretization and finite differences for the time discretization. The generality and simplicity of the finite difference method motivates our detailed study at the beginning of this work.

2.2 Finite differences for the heat equation

2.2.1 Various examples of schemes

We restrict ourselves to one space dimension and we refer to Section 2.2.6 for the case of several space dimensions. We consider the heat equation in the bounded domain $(0,1)$

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nu \dfrac{\partial^2 u}{\partial x^2} = 0 & \text{for } (x,t) \in (0,1) \times \mathbb{R}_*^+ \\ u(0,x) = u_0(x) & \text{for } x \in (0,1). \end{cases} \tag{2.1}$$

To discretize the domain $(0,1) \times \mathbb{R}^+$, we introduce a space step $\Delta x = 1/(N+1) > 0$ (with N a positive integer) and a time step $\Delta t > 0$, and we define the nodes of a regular mesh

$$(t_n, x_j) = (n\Delta t, j\Delta x) \quad \text{for } n \geq 0, j \in \{0, 1, \ldots, N+1\}.$$

We denote by u_j^n the value of a discrete approximate solution at the point (t_n, x_j), and $u(t,x)$ the exact solution of (2.1). The initial data is discretized by

$$u_j^0 = u_0(x_j) \quad \text{for } j \in \{0, 1, \ldots, N+1\}.$$

The boundary conditions of (2.1) can be of several types, but their choice is not involved in the definition of the schemes. Here, we use Dirichlet boundary conditions

$$u(t,0) = u(t,1) = 0 \quad \text{for all } t \in \mathbb{R}_*^+$$

which imply

$$u_0^n = u_{N+1}^n = 0 \quad \text{for all } n > 0.$$

Consequently, at each time step we have to calculate the values $(u_j^n)_{1 \leq j \leq N}$ which form a vector of $\mathbb{R}^N$. We now give several possible schemes for the heat equation (2.1). All of them are defined by N equations (at each point x_j, $1 \leq j \leq N$) which allow us to calculate the N values u_j^n. In Chapter 1 we have already talked of the **explicit scheme**

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \nu \frac{-u_{j-1}^n + 2u_j^n - u_{j+1}^n}{(\Delta x)^2} = 0 \tag{2.2}$$

for $n \geq 0$ and $j \in \{1, \ldots, N\}$, and also of the **implicit scheme**

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \nu \frac{-u_{j-1}^{n+1} + 2u_j^{n+1} - u_{j+1}^{n+1}}{(\Delta x)^2} = 0. \tag{2.3}$$

It is easy to verify that the implicit scheme (2.3) is well defined, that is, we can calculate the values u_j^{n+1} as a function of the u_j^n: in effect, we must invert the square

tridiagonal matrix of dimension N

$$\begin{pmatrix} 1+2c & -c & & & 0 \\ -c & 1+2c & -c & & \\ & \ddots & \ddots & \ddots & \\ & & -c & 1+2c & -c \\ 0 & & & -c & 1+2c \end{pmatrix} \quad \text{with } c = \frac{\nu\Delta t}{(\Delta x)^2}, \tag{2.4}$$

which is easily verified to be positive definite, therefore invertible. By making a convex combination of (2.2) and (2.3), for $0 \leq \theta \leq 1$, we obtain the **θ-scheme**

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \theta\nu \frac{-u_{j-1}^{n+1} + 2u_j^{n+1} - u_{j+1}^{n+1}}{(\Delta x)^2} + (1-\theta)\nu \frac{-u_{j-1}^{n} + 2u_j^{n} - u_{j+1}^{n}}{(\Delta x)^2} = 0. \tag{2.5}$$

We recover the explicit scheme (2.2) if $\theta = 0$, and the implicit scheme (2.3) if $\theta = 1$. The θ-scheme (2.5) is implicit when $\theta \neq 0$. For the value $\theta = 1/2$, we obtain the **Crank–Nicolson scheme**. Another implicit scheme, called the six point scheme, is given by

$$\begin{aligned} &\frac{u_{j+1}^{n+1} - u_{j+1}^n}{12\Delta t} + \frac{5(u_j^{n+1} - u_j^n)}{6\Delta t} + \frac{u_{j-1}^{n+1} - u_{j-1}^n}{12\Delta t} \\ &+\nu \frac{-u_{j-1}^{n+1} + 2u_j^{n+1} - u_{j+1}^{n+1}}{2(\Delta x)^2} + \nu \frac{-u_{j-1}^{n} + 2u_j^{n} - u_{j+1}^{n}}{2(\Delta x)^2} = 0. \end{aligned} \tag{2.6}$$

Exercise 2.2.1 Show that the scheme (2.6) is nothing more than the θ-scheme with $\theta = 1/2 - (\Delta x)^2/12\nu\Delta t$.

All the schemes above are called **two level** since they only involve two time indices. We can construct multilevel schemes: the most popular have three levels. In addition to the (unstable) Richardson scheme seen in Chapter 1, we cite the **DuFort–Frankel scheme**

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} + \nu \frac{-u_{j-1}^{n} + u_j^{n+1} + u_j^{n-1} - u_{j+1}^{n}}{(\Delta x)^2} = 0, \tag{2.7}$$

the **Gear scheme**

$$\frac{3u_j^{n+1} - 4u_j^n + u_j^{n-1}}{2\Delta t} + \nu \frac{-u_{j-1}^{n+1} + 2u_j^{n+1} - u_{j+1}^{n+1}}{(\Delta x)^2} = 0. \tag{2.8}$$

We have too many schemes! And the list above is not exhaustive! One of the aims of numerical analysis is to compare and to choose the best schemes following criteria of accuracy, cost, or robustness.

Remark 2.2.1 If there is a right-hand side $f(t,x)$ in the heat equation (2.1), then the schemes are modified by replacing zero in the right-hand side by a consistent approximation of $f(t,x)$ at the point (t_n, x_j). For example, if we choose the approximation $f(t_n, x_j)$, the explicit scheme (2.2) becomes

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \nu \frac{-u_{j-1}^n + 2u_j^n - u_{j+1}^n}{(\Delta x)^2} = f(t_n, x_j).$$

•

Remark 2.2.2 The schemes above are written compactly, that is, they involve a finite number of values u_j^n. The set of the couples (n', j') which appear in the discrete equation at the point (n, j) is called the **stencil** of the scheme. In general, the larger the stencil, the more costly and difficult it is to program the scheme (partly because of the 'boundary effects', that is, the case where some of the couples (n', j') leave the domain of calculation). •

Remark 2.2.3 We can replace the Dirichlet boundary conditions in (2.1) by Neumann boundary conditions, or by periodic (or other) boundary conditions. We start by describing two different ways of discretizing Neumann conditions

$$\frac{\partial u}{\partial x}(t,0) = 0 \quad \text{and} \quad \frac{\partial u}{\partial x}(t,1) = 0.$$

First, we can write

$$\frac{u_1^n - u_0^n}{\Delta x} = 0 \quad \text{and} \quad \frac{u_{N+1}^n - u_N^n}{\Delta x} = 0$$

which allow us to eliminate the values u_0^n and u_{N+1}^n and only to calculate the N values $(u_j^n)_{1 \le j \le N}$. This discretization of the Neumann condition is only first order. If the scheme is second order, this causes a loss of accuracy close to the boundary. This is why we propose another discretization (of second order)

$$\frac{u_1^n - u_{-1}^n}{2\Delta x} = 0 \quad \text{and} \quad \frac{u_{N+2}^n - u_N^n}{2\Delta x} = 0$$

which is more accurate, but needs us to add 2 'fictitious points' x_{-1} and x_{N+2}. We eliminate the values u_{-1}^n and u_{N+2}^n, corresponding to these fictitious points, and there now remains $N+2$ values to calculate, that is, $(u_j^n)_{0 \le j \le N+1}$.

On the other hand, periodic boundary conditions are written

$$u(t, x+1) = u(t,x) \quad \text{for all } x \in [0,1], \ t \ge 0.$$

These are discretized by the equations $u_0^n = u_{N+1}^n$ for all $n \ge 0$, and more generally $u_j^n = u_{N+1+j}^n$. •

2.2.2 Consistency and accuracy

Of course, the formulas of the schemes above are not chosen by accident: they follow from an approximation of the equation by Taylor expansion as we have explained in Chapter 1. To formalize this approximation of the partial differential equation by finite differences, we introduce the ideas of **consistency** and of **accuracy**. Although for the moment we only consider the heat equation (2.1), we shall give a definition of the consistency which is valid for every partial differential equation which we write $F(u) = 0$. We remark that $F(u)$ is notation for a function of u and its partial derivatives at every point (t, x). Generally a finite difference scheme is defined, for all possible indices n, j, by the formula

$$F_{\Delta t,\Delta x}\left(\{u_{j+k}^{n+m}\}_{m^-\leq m\leq m^+,\ k^-\leq k\leq k^+}\right) = 0 \tag{2.9}$$

where the integers m^-, m^+, k^-, k^+ define the width of the stencil of the scheme (see remark 2.2.2).

Definition 2.2.4 *The finite difference scheme (2.9) is called consistent with the partial differential equation $F(u) = 0$, if, for every sufficiently regular solution $u(t, x)$ of this equation,* **the truncation error** *of the scheme, defined by*

$$F_{\Delta t,\Delta x}\left(\{u(t + m\Delta t, x + k\Delta x)\}_{m^-\leq m\leq m^+,\ k^-\leq k\leq k^+}\right), \tag{2.10}$$

tends to zero, uniformly with respect to (t, x), as Δt and Δx tend to zero independently.

Further, we say that the scheme has accuracy of order p in space and order q in time if the truncation error (2.10) tends to zero as $\mathcal{O}\Big((\Delta x)^p + (\Delta t)^q\Big)$ when Δt and Δx tend to zero.

Remark 2.2.5 We must take care with the formula (2.9) since there is a small ambiguity in the definition of the scheme. Indeed, we can always multiply any formula by a sufficiently high power of Δt and Δx so that the truncation error tends to zero. This will make any scheme consistent! To avoid this problem, we always assume that the formula $F_{\Delta t,\Delta x}(\{u_{j+k}^{n+m}\}) = 0$ has been written so that, for a regular function $u(t, x)$ which is not a solution of the equation $F(u) = 0$, the limit of the truncation error is not zero. •

Concretely we calculate the truncation error of a scheme by replacing u_{j+k}^{n+m} in formula (2.9) by $u(t + m\Delta t, x + k\Delta x)$. As an application of the definition 2.2.4, we shall show the following lemma.

Lemma 2.2.6 *The explicit scheme (2.2) is consistent, accurate with order 1 in time and 2 in space. Further, if we choose to keep the ratio $\nu\Delta t/(\Delta x)^2 = 1/6$ constant, then this scheme is accurate with order 2 in time and 4 in space.*

Remark 2.2.7 In the second phrase of the statement of lemma 2.2.6 we slightly modified the definition of the consistency by specifying the ratio of Δt and Δx as they tend to zero. This allows us to take advantage of potential cancellation between terms in the truncation error. In practice, we see such improvements in the accuracy if we adopt a good relationship between the terms Δt and Δx. •

Proof. Let $v(t,x)$ be a function of class $\mathcal{C}^6$. By Taylor expansion around the point (t,x), we calculate the truncation error of the scheme (2.2)

$$\frac{v(t+\Delta t,x)-v(t,x)}{\Delta t}+\nu\frac{-v(t,x-\Delta x)+2v(t,x)-v(t,x+\Delta x)}{(\Delta x)^2}$$

$$=\Big(v_t-\nu v_{xx}\Big)+\frac{\Delta t}{2}v_{tt}-\frac{\nu(\Delta x)^2}{12}v_{xxxx}+\mathcal{O}\Big((\Delta t)^2+(\Delta x)^4\Big),$$

where v_t, v_x denote the partial derivatives of v. If v is a solution of the heat equation (2.1), we thus easily obtain the consistency as well as accuracy of order 1 in time and 2 in space. If further we assume that $\nu\Delta t/(\Delta x)^2 = 1/6$, then the terms in Δt and $(\Delta x)^2$ cancel since $v_{tt} = \nu v_{txx} = \nu^2 v_{xxxx}$. □

Scheme	Truncation error	Stability
Explicit (2.2)	$\mathcal{O}\Big(\Delta t+(\Delta x)^2\Big)$	Stable in L^2 and L^∞ for CFL condition $2\nu\Delta t \le (\Delta x)^2$
Implicit (2.3)	$\mathcal{O}\Big(\Delta t+(\Delta x)^2\Big)$	Stable in L^2 and L^∞
Crank–Nicolson (2.5) (with $\theta = 1/2$)	$\mathcal{O}\Big((\Delta t)^2+(\Delta x)^2\Big)$	Stable in L^2
θ-scheme (2.5) (with $\theta \neq 1/2$)	$\mathcal{O}\Big(\Delta t+(\Delta x)^2\Big)$	Stable in L^2 for CFL condition $2(1-2\theta)\nu\Delta t \le (\Delta x)^2$
Six point scheme (2.6)	$\mathcal{O}\Big((\Delta t)^2+(\Delta x)^4\Big)$	Stable in L^2
DuFort–Frankel (2.7)	$\mathcal{O}\Big((\Delta t/\Delta x)^2+(\Delta x)^2\Big)$	Stable in L^2 for CFL condition $\Delta t/(\Delta x)^2$ bounded
Gear (2.8)	$\mathcal{O}\Big((\Delta t)^2+(\Delta x)^2\Big)$	Stable in L^2

Table 2.1. Truncation errors and stability of various schemes for the heat equation

Exercise 2.2.2 For each of the schemes of Section 2.2.1, verify that the truncation error is of the type stated in Table 2.1. (We remark that all these schemes are consistent except for DuFort–Frankel.)

2.2.3 Stability and Fourier analysis

In Chapter 1 we introduced the stability of finite differences schemes without giving a precise definition. We have explained that, numerically, instability is shown by unbounded oscillations of the numerical solution. It is therefore time to give a mathematical definition of stability. For this we need to define a norm for the numerical solution $u^n = (u_j^n)_{1\leq j\leq N}$. We take the classical norms on $\mathbb{R}^N$ which we scale by the space step Δx:

$$\|u^n\|_p = \left(\sum_{j=1}^{N} \Delta x |u_j^n|^p\right)^{1/p} \quad \text{for } 1 \leq p \leq +\infty, \tag{2.11}$$

where the limiting case $p = +\infty$ should be understood in the sense $\|u^n\|_\infty = \max_{1\leq j\leq N} |u_j^n|$. We remark that the norm defined therefore depends on Δx through the weighting but also on the integer N since $\Delta x = 1/(N+1)$. Thanks to the weighting by Δx, the norm $\|u^n\|_p$ is identical to the norm $L^p(0,1)$ for piecewise constant functions over the subintervals $[x_j, x_{j+1}[$ of $[0,1]$. Often, we shall call this the 'L^p norm'. In practice, we most often use the norms corresponding to the values $p = 2, +\infty$.

Definition 2.2.8 *A finite difference scheme is called* **stable** *for the norm $\|\ \|$, defined by (2.11), if there exists a constant $K > 0$ independent of Δt and Δx (as these values tend to zero) such that*

$$\|u^n\| \leq K\|u^0\| \quad \text{for all } n \geq 0, \tag{2.12}$$

for arbitrary initial data u^0.

If (2.12) only hold for steps Δt and Δx defined by certain inequalities, we say that the scheme is **conditionally stable**.

Remark 2.2.9 Since all norms are equivalent in $\mathbb{R}^N$, the hasty reader might believe that stability with respect to one norm implies stability with respect to all norms. Unfortunately, this is not true and there exist schemes which are stable with respect to one norm but not with respect to another (see later the example of the Lax–Wendroff scheme in exercises 2.3.2 and 2.3.3). In effect, the crucial point in the definition 2.2.8 is that the bound is uniform with respect to Δx while the norms (2.11) depend on Δx. •

Definition 2.2.10 *A finite difference scheme is called* **linear** *if its formula $F_{\Delta t,\Delta x}(\{u_{j+k}^{n+m}\}) = 0$ is linear with respect to its arguments u_{j+k}^{n+m}.*

The stability of a two level linear scheme is very easy to interpret. Indeed, by linearity every two level linear scheme can be written in the condensed form

$$u^{n+1} = Au^n, \tag{2.13}$$

where A is a linear operator (a matrix, called the iteration matrix) from $\mathbb{R}^N$ into $\mathbb{R}^N$. For example, for the explicit scheme (2.2) the matrix A becomes

$$\begin{pmatrix} 1-2c & c & & & 0 \\ c & 1-2c & c & & \\ & \ddots & \ddots & \ddots & \\ & & c & 1-2c & c \\ 0 & & & c & 1-2c \end{pmatrix} \quad \text{with } c = \frac{\nu \Delta t}{(\Delta x)^2}, \tag{2.14}$$

while for the implicit scheme (2.3) the matrix A is the inverse of the matrix (2.4). With the help of this iteration matrix, we have $u^n = A^n u^0$ (take care, the notation A^n denotes the nth power of A), and consequently the stability of the scheme is equivalent to

$$\|A^n u^0\| \leq K \|u^0\| \quad \forall\, n \geq 0,\ \forall\, u^0 \in \mathbb{R}^N.$$

Introducing the subordinate matrix norm (see definition 13.1.1)

$$\|M\| = \sup_{u \in \mathbb{R}^N, u \neq 0} \frac{\|Mu\|}{\|u\|},$$

the stability of the scheme is equivalent to

$$\|A^n\| \leq K \quad \forall\, n \geq 0, \tag{2.15}$$

which is the same as saying the sequence of the powers of A is bounded.

Stability in the L^∞ norm

The stability in the L^∞ norm is closely linked with the discrete maximum principle which we have seen in Chapter 1. Let us recall the definition of this principle.

Definition 2.2.11 *A finite difference scheme satisfies the* **discrete maximum principle** *if for all $n \geq 0$ and all $1 \leq j \leq N$ we have*

$$\min\left(0, \min_{0 \leq j \leq N+1} u_j^0\right) \leq u_j^n \leq \max\left(0, \max_{0 \leq j \leq N+1} u_j^0\right)$$

for arbitrary initial data u^0.

Remark 2.2.12 In definition 2.2.11 the inequalities take account not only of the minimum and maximum of u^0 but also of zero which is the value imposed on the boundary by the Dirichlet boundary conditions. This is necessary if the initial data u^0 does not satisfy the Dirichlet boundary conditions (which is not required), and superfluous in the complementary case. •

As we have seen in Chapter 1 (see (1.33) and exercise 1.4.1), the discrete maximum principle allows us to prove the following lemma.

Lemma 2.2.13 *The explicit scheme (2.2) is stable in the L^∞ norm if and only if the CFL condition $2\nu\Delta t \leq (\Delta x)^2$ is satisfied. The implicit scheme (2.3) is stable in the L^∞ norm no matter what the time step Δt and space step Δx (we say that it is unconditionally stable).*

Exercise 2.2.3 Show that the Crank–Nicolson scheme (2.5) (with $\theta = 1/2$) is stable in the L^∞ norm if $\nu\Delta t \leq (\Delta x)^2$, and that the DuFort–Frankel scheme (2.7) is stable in the L^∞ norm if $2\nu\Delta t \leq (\Delta x)^2$.

Stability in the L^2 norm

Many schemes do not satisfy the discrete maximum principle but are nevertheless 'good' schemes. For this, we must verify the stability in a norm other than the L^∞ norm. The L^2 norm lends itself very well to the study of stability thanks to the very powerful tool of Fourier analysis which we now present. To do this, we assume from now on that the boundary conditions for the heat equation are **periodic boundary conditions**, which are written $u(t, x+1) = u(t, x)$ for all $x \in [0, 1]$ and all $t \geq 0$. For numerical schemes, these lead to the equations $u_0^n = u_{N+1}^n$ for all $n \geq 0$, and more generally $u_j^n = u_{N+1+j}^n$. We therefore have to calculate $N+1$ values u_j^n.

With each vector $u^n = (u_j^n)_{0 \leq j \leq N}$ we associate a function $u^n(x)$, piecewise constant, periodic with period 1, defined on $[0, 1]$ by

$$u^n(x) = u_j^n \quad \text{if } x_{j-1/2} < x < x_{j+1/2}$$

with $x_{j+1/2} = (j + 1/2)\Delta x$ for $0 \leq j \leq N$, $x_{-1/2} = 0$, and $x_{N+1+1/2} = 1$. The function $u^n(x)$ belongs to $L^2(0, 1)$. Now, from Fourier analysis, every function of $L^2(0, 1)$ can be decomposed into a Fourier sum (see [4], [35], [38]). More precisely we have

$$u^n(x) = \sum_{k \in \mathbb{Z}} \hat{u}^n(k) \exp(2i\pi kx), \tag{2.16}$$

with $\hat{u}^n(k) = \int_0^1 u^n(x) \exp(-2i\pi kx)\, dx$ and the Plancherel formula

$$\int_0^1 |u^n(x)|^2\, dx = \sum_{k \in \mathbb{Z}} |\hat{u}^n(k)|^2. \tag{2.17}$$

We remark that even if u^n is a real function, the coefficients $\hat{u}^n(k)$ of the Fourier series are complex. An important property for the Fourier transform of periodic functions is the following: if we denote by $v^n(x) = u^n(x + \Delta x)$, then $\hat{v}^n(k) = \hat{u}^n(k) \exp(2i\pi k\Delta x)$.

Let us now explain the method using the example of the explicit scheme (2.2). Under our notation, we can rewrite this scheme, for $0 \leq x \leq 1$,

$$\frac{u^{n+1}(x) - u^n(x)}{\Delta t} + \nu \frac{-u^n(x - \Delta x) + 2u^n(x) - u^n(x + \Delta x)}{(\Delta x)^2} = 0.$$

By application of the Fourier transform, this becomes

$$\hat{u}^{n+1}(k) = \left(1 - \frac{\nu\Delta t}{(\Delta x)^2}\left(-\exp(-2i\pi k\Delta x) + 2 - \exp(2i\pi k\Delta x)\right)\right)\hat{u}^n(k).$$

In other words,

$$\hat{u}^{n+1}(k) = A(k)\hat{u}^n(k) = A(k)^{n+1}\hat{u}^0(k) \quad \text{with } A(k) = 1 - \frac{4\nu\Delta t}{(\Delta x)^2}(\sin(\pi k\Delta x))^2.$$

For $k \in \mathbb{Z}$, the Fourier coefficient $\hat{u}^n(k)$ is bounded as n tends to infinity if and only if the amplification factor satisfies $|A(k)| \leq 1$, that is,

$$2\nu\Delta t(\sin(\pi k\Delta x))^2 \leq (\Delta x)^2. \tag{2.18}$$

If the CFL condition (1.31), that is, $2\nu\Delta t \leq (\Delta x)^2$, is satisfied, then inequality (2.18) is true for every Fourier mode $k \in \mathbb{Z}$, and by the Plancherel formula we deduce

$$\|u^n\|_2^2 = \int_0^1 |u^n(x)|^2 dx = \sum_{k\in\mathbb{Z}} |\hat{u}^n(k)|^2 \leq \sum_{k\in\mathbb{Z}} |\hat{u}^0(k)|^2 = \int_0^1 |u^0(x)|^2 dx = \|u^0\|_2^2,$$

which is nothing other than the L^2 stability of the explicit scheme. If the CFL condition is not satisfied, the scheme is unstable. In effect, it is enough to choose Δx (possibly sufficiently small) and k_0 (sufficiently large) and initial data with only one nonzero Fourier component $\hat{u}^0(k_0) \neq 0$ with $\pi k_0\Delta x \approx \pi/2$ (modulo π) in such a way that $|A(k_0)| > 1$. We have therefore proved the following lemma.

Lemma 2.2.14 *The explicit scheme (2.2) is stable in the L^2 norm if and only if the CFL condition $2\nu\Delta t \leq (\Delta x)^2$ is satisfied.*

In the same way we shall prove the stability of the implicit scheme.

Lemma 2.2.15 *The implicit scheme (2.3) is stable in the L^2 norm.*

Remark 2.2.16 For explicit (2.2) and implicit (2.3) schemes the L^2 stability condition is the same as that of the L^∞ stability. This is not always the case for other schemes. •

Proof. Similar reasoning to that used for the explicit scheme leads, for $0 \leq x \leq 1$, to

$$\frac{u^{n+1}(x) - u^n(x)}{\Delta t} + \nu\frac{-u^{n+1}(x-\Delta x) + 2u^{n+1}(x) - u^{n+1}(x+\Delta x)}{(\Delta x)^2} = 0,$$

and by application of the Fourier transform

$$\hat{u}^{n+1}(k)\left(1 + \frac{\nu\Delta t}{(\Delta x)^2}\left(-\exp(-2i\pi k\Delta x) + 2 - \exp(2i\pi k\Delta x)\right)\right) = \hat{u}^n(k).$$

In other words,

$$\hat{u}^{n+1}(k) = A(k)\hat{u}^n(k) = A(k)^{n+1}\hat{u}^0(k) \text{ with } A(k) = \left(1 + \frac{4\nu\Delta t}{(\Delta x)^2}(\sin(\pi k \Delta x))^2\right)^{-1}.$$

As $|A(k)| \leq 1$ for all Fourier modes k, the Plancherel formula gives us the L^2 stability of the scheme. □

Remark 2.2.17 The Fourier analysis relies on the choice of periodic boundary conditions. We can also carry it out if the partial differential equation holds over all $\mathbb{R}$ instead of $[0, 1]$ (we have then to deal with a Fourier integral instead of a Fourier series). Nevertheless, it is not very realistic to talk about a numerical scheme over all $\mathbb{R}$ since this implies an infinite number of values u_j^n at each time step n when a computer can only treat a finite number of values.

The L^2 stability can also be proved in the case of Dirichlet boundary conditions. We must then adapt the ideas of Fourier analysis. For example, what replaces the Fourier transform in this case is the decomposition over a basis of eigenvectors of the iteration matrix (2.13) which allows us to move from the vector u^n to the vector u^{n+1}. •

Remark 2.2.18 (Essential from a practical point of view) Let us give a 'recipe' for Fourier analysis to prove the L^2 stability of a scheme. We put Fourier modes into the scheme

$$u_j^n = A(k)^n \exp(2i\pi k x_j) \quad \text{with} \quad x_j = j\Delta x,$$

and we deduce the value of the amplification factor $A(k)$. Recall that, for the moment, we restrict ourselves to the scalar case, that is, $A(k)$ is a complex number in $\mathbb{C}$. The inequality

$$|A(k)| \leq 1 \quad \text{for all modes } k \in \mathbb{Z} \tag{2.19}$$

is called the **Von Neumann stability condition**. If the Von Neumann stability condition is satisfied (with possibly restrictions on Δt and Δx), then the scheme is stable for the L^2 norm, if not it is unstable.

In general, a stable (and consistent) scheme is convergent (see Section 2.2.4). In practice, an unstable scheme is totally 'useless'. In effect, even if we start from initial data specially designed so that none of the unstable Fourier modes are excited, the inevitable rounding errors will create nonzero components (although very small) of the solution in the unstable modes. The exponential increase of the unstable modes implies that after only a few time steps these 'small' modes become 'enormous' and completely pollute the rest of the numerical solution. •

Exercise 2.2.4 Show that the θ-scheme (2.5) is unconditionally stable in the L^2 norm if $1/2 \leq \theta \leq 1$, and stable under the CFL condition $2(1-2\theta)\nu\Delta t \leq (\Delta x)^2$ if $0 \leq \theta < 1/2$.

Exercise 2.2.5 Show that the 6-point scheme (2.6) is unconditionally stable in the L^2 norm.

Remark 2.2.19 Some authors use another definition of the stability, which is less restrictive than definition 2.2.8 but more complex. In this definition the scheme is called stable for the norm $\| \ \|$ if for all time $T > 0$ there exists a constant $K(T) > 0$ independent of Δt and Δx such that

$$\|u^n\| \leq K(T)\|u^0\| \quad \text{for all } 0 \leq n \leq T/\Delta t,$$

whatever the initial data u^0. This new definition allows the solution to grow with time as is the case, for example, for the solution of the equation

$$\frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} = cu \quad \text{for } (t, x) \in \mathbb{R}^+ \times \mathbb{R},$$

which, by changing the unknown $v(t, x) = e^{-ct}u(t, x)$, reduces to the heat equation (then the solution u grows exponentially in time). With such a definition of the stability, the Von Neumann stability condition becomes the inequality

$$|A(k)| \leq 1 + C\Delta t \quad \text{for all modes } k \in \mathbb{Z}.$$

For simplicity, we shall take the definition 2.2.8 of the stability. •

2.2.4 Convergence of the schemes

We now have all the tools to prove convergence of the finite differences schemes. The principal result of this section is the Lax theorem which shows that, for a linear scheme, **consistency and stability implies convergence**. The importance of this result far exceeds the finite difference method. For every numerical method (finite differences, finite elements, etc.) convergence is shown by combining two arguments: stability and consistency (their precise definitions change from one method to the other). From a practical point of view, the Lax theorem is very reassuring: if we use a consistent scheme (we can construct this generally) and we do not observe numerical oscillations (that is, it is stable), then the numerical solution is close to the exact solution (the scheme converges).

Theorem 2.2.20 (Lax) *Let $u(t, x)$ be the sufficiently regular solution of the heat equation (2.1) (with the appropriate boundary conditions). Let u_j^n be the discrete numerical solution obtained by a finite difference scheme with the initial data $u_j^0 = u_0(x_j)$. We assume that the scheme is linear, two level, consistent, and stable for a norm $\| \ \|$. Then the scheme is convergent in the sense where*

$$\forall\, T > 0, \quad \lim_{\Delta t, \Delta x \to 0} \left(\sup_{t_n \leq T} \|e^n\| \right) = 0, \tag{2.20}$$

with e^n is the 'error' vector defined by its components $e_j^n = u_j^n - u(t_n, x_j)$.

Further, if the scheme has accuracy of order p in space and order q in time, then for all time $T > 0$ there exists a constant $C_T > 0$ such that

$$\sup_{t_n \leq T} \|e^n\| \leq C_T \Big((\Delta x)^p + (\Delta t)^q \Big). \tag{2.21}$$

Remark 2.2.21 We have not proved the existence and uniqueness of the solution of the heat equation (2.1) (with Dirichlet or periodic boundary conditions). For the moment, we therefore make the hypothesis of the existence and uniqueness of such a solution (as well as its regularity), but we see in Chapter 8 that this result is generally true. •

Proof. For simplicity, we assume that the boundary conditions are Dirichlet. The same proof is also true for periodic boundary conditions or for Neumann boundary conditions (assuming they are discretized with the same order of accuracy as the scheme). A two level linear scheme can be written in the condensed form (2.13), that is,

$$u^{n+1} = Au^n,$$

where A is the iteration matrix (square of size N). Let u be the solution (assumed sufficiently regular) of the heat equation (2.1). We denote by $\tilde{u}^n = (\tilde{u}_j^n)_{1\leq j\leq N}$ with $\tilde{u}_j^n = u(t_n, x_j)$. As the scheme is consistent, there exists a vector ϵ^n such that

$$\tilde{u}^{n+1} = A\tilde{u}^n + \Delta t\epsilon^n \quad \text{with} \lim_{\Delta t,\Delta x\to 0} \|\epsilon^n\| = 0, \tag{2.22}$$

and the convergence of ϵ^n is uniform for all time $0 \leq t_n \leq T$. If the scheme is accurate with order p in space and order q in time, then $\|\epsilon^n\| \leq C((\Delta x)^p + (\Delta t)^q)$. By setting $e_j^n = u_j^n - u(t_n, x_j)$ we obtain by subtraction of (2.22) from (2.13)

$$e^{n+1} = Ae^n - \Delta t\epsilon^n$$

from which, by induction

$$e^n = A^n e^0 - \Delta t \sum_{k=1}^{n} A^{n-k}\epsilon^{k-1}. \tag{2.23}$$

Now, the stability of the scheme means that $\|u^n\| = \|A^n u^0\| \leq K\|u^0\|$ for all initial data, that is, $\|A^n\| \leq K$ where the constant K does not depend on n. On the other hand, $e^0 = 0$, therefore (2.23) gives

$$\|e^n\| \leq \Delta t \sum_{k=1}^{n} \|A^{n-k}\|\|\epsilon^{k-1}\| \leq \Delta t n K C\Big((\Delta x)^p + (\Delta t)^q\Big),$$

which gives the inequality (2.21) with the constant $C_T = TKC$. The proof of (2.20) is similar. □

Remark 2.2.22 The Lax theorem 2.2.20 is in fact valid for all linear partial differential equations. It has a converse in the sense that if a two level linear consistent scheme is convergent then it must be stable. We remark that the rate of convergence in (2.21) is exactly the accuracy of the scheme. Finally, it is good to note that the estimate (2.21) is only valid on a bounded time interval $[0, T]$ but it is independent of the number of points of discretization N. •