Chapter 2
Elements of Numerical Methods for PDEs

In this chapter, we present some elements of numerical methods for partial differential equations (PDEs). The PDEs are classified into elliptic, parabolic and hyperbolic equations, and we indicate the corresponding type of problems that they model. PDEs arising in option pricing problems in finance are mostly parabolic. Occasionally, however, elliptic PDEs arise in connection with so-called "infinite horizon problems", and hyperbolic PDEs may appear in certain pure jump models with dominating drift.

Therefore, we consider in particular the heat equation and show how to solve it numerically using finite differences or finite elements. Finite difference methods (FDM) consist of finding an approximate solution on a grid by replacing the derivatives in the differential equation by difference quotients. Finite element methods (FEM) are based instead on variational formulations of the differential equations and determine approximate solutions that are usually piecewise polynomials on some partition of the (log) price domain. We start with recapitulating some function spaces as well as the classification of PDEs.

2.1 Function Spaces

The variational formulation and the analysis of the finite element method require tools from functional analysis, in particular Hilbert spaces (see Appendix A). Let G be a non-empty open subset of $\mathbb{R}^d$. If a function $u : G \to \mathbb{R}$ is sufficiently smooth, we denote the partial derivatives of u by

$$D^{\mathbf{n}}u(x) := \frac{\partial^{|\mathbf{n}|}u(x)}{\partial x_1^{n_1}\cdots\partial x_d^{n_d}} = \partial_{x_1}^{n_1}\cdots\partial_{x_d}^{n_d}u(x), \quad x = (x_1,\ldots,x_d) \in G, \tag{2.1}$$

where $\mathbf{n} = (n_1,\ldots,n_d) \in \mathbb{N}_0^d$ is a multi-index. The order of the partial derivative is given by $|\mathbf{n}| = \sum_{i=1}^d n_i$. For any integer $n \in \mathbb{N}_0$, we define

$$C^n(G) = \{u : D^{\mathbf{n}}u \text{ exists and is continuous on } G \text{ for } |\mathbf{n}| \le n\},$$

N. Hilber et al., *Computational Methods for Quantitative Finance*, Springer Finance,
DOI 10.1007/978-3-642-35401-4_2, © Springer-Verlag Berlin Heidelberg 2013

and set $C^\infty(G) = \bigcap_{n\geq 0} C^n(G)$. The support of u is denoted by $\operatorname{supp} u$, and we define $C_0^n(G)$, $C_0^\infty(G)$ consisting of all functions $u \in C^n(G)$, $C^\infty(G)$ with compact support $\operatorname{supp} u \Subset G$.

We denote by $L^p(G)$, $1 \leq p \leq \infty$ the usual space which consists of all Lebesgue measurable functions $u : G \to \mathbb{R}$ with finite L^p-norm,

$$\|u\|_{L^p(G)} := \begin{cases} \left(\int_G |u(x)|^p \,\mathrm{d}x\right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \operatorname{ess\,sup}_G |u(x)| & \text{if } p = \infty, \end{cases}$$

where ess sup means the *essential supremum* disregarding values on nullsets. The case $p = 2$ is of particular interest. The space $L^2(G)$ is a Hilbert space with respect to the inner product $(u, v) = \int_G u(x)v(x)\,\mathrm{d}x$.

Let $\mathcal{H}$ be a Hilbert space with the inner product $(\cdot,\cdot)_\mathcal{H}$ and norm $\|u\|_\mathcal{H} := (u, u)_\mathcal{H}^{1/2}$. We denote by $\mathcal{H}^*$ the dual space of $\mathcal{H}$ which consists of all bounded linear functionals $u^* : \mathcal{H} \to \mathbb{R}$ on $\mathcal{H}$. $\mathcal{H}^*$ can be identified with $\mathcal{H}$ by the Riesz representation theorem.

Theorem 2.1.1 (Riesz representation theorem) *For each $u^* \in \mathcal{H}^*$ there exists a unique element $u \in \mathcal{H}$ such that*

$$\langle u^*, v\rangle_{\mathcal{H}^*,\mathcal{H}} = (u, v)_\mathcal{H} \quad \forall v \in \mathcal{H}.$$

The mapping $u^* \mapsto u$ is a linear isomorphism of $\mathcal{H}^*$ onto $\mathcal{H}$.

The theory of parabolic partial differential equations requires the introduction of Hilbert space-valued L^p-functions. As above, let $\mathcal{H}$ be a Hilbert space with the norm $\|\cdot\|_\mathcal{H}$. Denote by J the interval $J := (0, T)$ with $T > 0$, and let $1 \leq p \leq \infty$. The space $L^p(J; \mathcal{H})$ is defined by

$$L^p(J; \mathcal{H}) := \{u : \overline{J} \to \mathcal{H} \text{ measurable} : \|u\|_{L^p(J;\mathcal{H})} < \infty\},$$

with the norm

$$\|u\|_{L^p(J;\mathcal{H})} := \begin{cases} \left(\int_J \|u(t)\|_\mathcal{H}^p \,\mathrm{d}t\right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \operatorname{ess\,sup}_J \|u(t)\|_\mathcal{H} & \text{if } p = \infty. \end{cases}$$

Furthermore, for $n \in \mathbb{N}_0$ let $C^n(J; \mathcal{H})$ be the space of $\mathcal{H}$-valued functions that are of the class C^n with respect to t.

2.2 Partial Differential Equations

For $k \in \mathbb{N}$ we let

$$D^k u(x) := \{D^\mathbf{n} u(x) : |\mathbf{n}| = k\}$$

be the set of all partial derivatives of order k. If $k = 1$, we regard the elements of $D^1u(x) =: Du(x)$ as being arranged in a row vector

$$Du = (\partial_{x_1}u, \ldots, \partial_{x_d}u).$$

If $k = 2$, we regard the elements of $D^2u(x)$ as being arranged in a matrix

$$D^2u = \begin{pmatrix} \partial_{x_1}\partial_{x_1}u & \ldots & \partial_{x_1}\partial_{x_d}u \\ \vdots & \ddots & \vdots \\ \partial_{x_d}\partial_{x_1}u & \ldots & \partial_{x_d}\partial_{x_d}u \end{pmatrix}.$$

Hence, the Laplacian Δu of u can be written as

$$\Delta u := \sum_{i=1}^{d} \partial_{x_i}\partial_{x_i}u = \mathrm{tr}[D^2u], \tag{2.2}$$

where $\mathrm{tr} : \mathbb{R}^{d\times d} \to \mathbb{R}$, $\mathbf{B} \mapsto \mathrm{tr}[\mathbf{B}] = \sum_{i=1}^{d} \mathbf{B}_{ii}$ is the trace of a $d \times d$-matrix $\mathbf{B}$. In the following, we write $\partial_{x_i x_j}$ instead of $\partial_{x_i}\partial_{x_j}$ to simplify the notation.

A *partial differential equation* is an equation involving an unknown function of two or more variables and certain of its derivatives. Let $G \subset \mathbb{R}^d$ be open, $x = (x_1, \ldots, x_d) \in G$, and $k \in \mathbb{N}$.

Definition 2.2.1 An expression of the form

$$F(D^k u(x), D^{k-1}u(x), \ldots, Du(x), u(x), x) = 0, \quad x \in G,$$

is called a kth order partial differential equation, where the function

$$F : \mathbb{R}^{d^k} \times \mathbb{R}^{d^{k-1}} \times \cdots \times \mathbb{R}^d \times \mathbb{R} \times G \to \mathbb{R}$$

is given and the function $u : G \to \mathbb{R}$ is the unknown.

Let $a_{ij}(x)$, $b_i(x)$, $c(x)$ and $f(x)$ be given functions. For a *linear first order PDE* in $d + 1$ variables, F has the form

$$F(Du, u, x) = \sum_{i=0}^{d} b_i(x)\partial_{x_i}u + c(x)u - f(x).$$

Setting $x_0 = t$, $x_1 = x$, $b(x) = (b_1(x), b_2(x))^\top = (1, b)^\top$, $b \in \mathbb{R}_+$, and $c = 0$, for example, we obtain the (hyperbolic) transport equation with constant speed b of propagation $\partial_t u + b\partial_x = f(t, x)$.

For a *linear second order PDE* in $d + 1$ variables, F takes the form

$$F(D^2u, Du, u, x) = -\sum_{i,j=0}^{d} a_{ij}(x)\partial_{x_i x_j}u + \sum_{i=0}^{d} b_i(x)\partial_{x_i}u + c(x)u - f(x).$$

Let $b(x) = (b_0(x), \dots, b_d(x))$ and assume that the matrix $A(x) = \{a_{ij}(x)\}_{i,j=0}^{d}$ is symmetric with real eigenvalues $\lambda_0(x) \le \lambda_1(x) \le \cdots \le \lambda_d(x)$. We can use the eigenvalues to distinguish three types of PDEs: elliptic, parabolic and hyperbolic.

Definition 2.2.2 Let $\mathcal{I} = \{0, \dots, d\}$. At $x \in \mathbb{R}^{d+1}$, a PDE is called

(i) Elliptic $\Leftrightarrow \lambda_i(x) \neq 0, \forall i \wedge \operatorname{sign}(\lambda_0(x)) = \cdots = \operatorname{sign}(\lambda_d(x))$,
(ii) Parabolic $\Leftrightarrow \exists! j \in \mathcal{I} : \lambda_j(x) = 0 \wedge \operatorname{rank}(A(x), b(x)) = d + 1$,
(iii) Hyperbolic $\Leftrightarrow (\lambda_i(x) \neq 0,\ \forall i) \wedge \exists! j \in \mathcal{I} : \operatorname{sign} \lambda_j(x) \neq \operatorname{sign} \lambda_k(x),\ k \in \mathcal{I} \setminus \{j\}$.

The PDE is called elliptic, parabolic, hyperbolic on G, if it is elliptic, parabolic, hyperbolic at all $x \in G$.

We give a typical example for each type:

(i) The Poisson equation $\Delta u = f(x)$ is elliptic.
(ii) The heat equation $\partial_t u - \Delta u = f(t, x)$ is parabolic (set $x_0 = t$).
(iii) The wave equation $\partial_{tt} u - \Delta u = f(t, x)$ is hyperbolic (set $x_0 = t$).
(iv) The Black–Scholes equation for the value of a European option price $v(t, s)$

$$\partial_t v - \frac{1}{2}\sigma^2 s^2 \partial_{ss} v - rs\partial_s v + rv = 0, \tag{2.3}$$

with volatility $\sigma > 0$ and interest rate $r \ge 0$ is parabolic at $(t, s) \in (0, T) \times (0, \infty)$ and degenerates to an ordinary differential equation as $s \to 0$.

Partial differential equations arising in finance, like the Black–Scholes equation (2.3), are mostly of parabolic type, i.e. they are of the form

$$\partial_t u - \sum_{i,j=1}^{d} a_{ij}(x)\partial_{x_i x_j} u + \sum_{i=1}^{d} b_i(x)\partial_{x_i} u + c(x)u = f(x).$$

Therefore, we introduce the basic concepts for solving parabolic equations in the next section. For illustration purpose, we consider the heat equation. Indeed, setting $s = e^x$, $t = 2\sigma^{-2}\tau$ and

$$v(t, s) = e^{\alpha x + \beta\tau} u(\tau, x), \quad \alpha = 1/2 - r\sigma^{-2}, \quad \beta = -(1/2 + r\sigma^{-2})^2,$$

the Black–Scholes equation (2.3) for $v(t, s)$ can be transformed to the heat equation $\partial_\tau u - \partial_{xx} u = 0$ for $u(\tau, x)$.

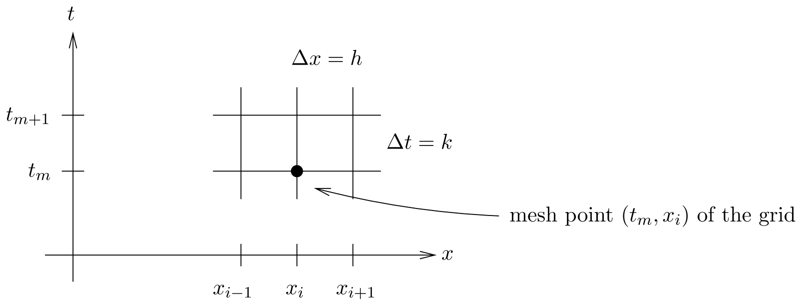


Fig. 2.1 Time-space grid

2.3 Numerical Methods for the Heat Equation

Let the space domain $G = (a, b) \subset \mathbb{R}$ be an open interval and let the time domain $J := (0, T)$ for $T > 0$. Consider the initial–boundary value problem:

$$\begin{aligned} &\text{Find } u : J \times G \to \mathbb{R} \text{ such that} \\ &\partial_t u - \partial_{xx} u = f(t, x), \qquad && \text{in } J \times G, \\ &u(t, x) = 0, && \text{on } J \times \partial G, \\ &u(0, x) = u_0, && \text{in } G, \end{aligned} \tag{2.4}$$

where $u(0, x) = u_0$ is the *initial condition* and $u(t, x) = 0$ on the boundary is called *the homogeneous Dirichlet boundary condition*. We explain two numerical methods to find approximations to the solution $u(t, x)$ of the problem (2.4). We start with the finite difference method.

2.3.1 *Finite Difference Method*

In the finite difference discretization, the domain $J \times G$ is replaced by discrete grid points (t_m, x_i) and the partial derivatives in (2.4) are approximated by difference quotients at the grid points. Let the *space grid points* be given by

$$x_i = a + ih,\ i = 0, 1, \ldots, N + 1,\ h := (b - a)/(N + 1) = \Delta x, \tag{2.5}$$

which are equidistant with mesh width h, and the *time levels* by

$$t_m = mk,\ m = 0, 1, \ldots, M,\ k := T/M = \Delta t. \tag{2.6}$$

The time-space grid is illustrated in Fig. 2.1.

Assume that $f \in C^2(G)$. Then, using Taylor's formula, we have

$$f'(x) = \frac{f(x + h) - f(x)}{h} - \frac{h}{2} f''(\xi), \quad \xi \in (x, x + h).$$

Setting $f_i := f(x_i)$, we obtain as $h \to 0$,

$$f'(x_i) = \frac{f_{i+1} - f_i}{h} + \mathcal{O}(h) =: (\delta_x^+ f)_i + \mathcal{O}(h),$$

where $(\delta_x^+ f)_i$ is called the *one-sided difference quotient* of f with respect to x at x_i. The difference quotient is said to be *accurate of first order* since the remainder term is $\mathcal{O}(h)$ as $h \to 0$. Analogous expressions hold for the time derivative ∂_t.

Higher order finite differences allow obtaining approximations of order $\mathcal{O}(h^p)$ with $p \geq 2$ rather than just $\mathcal{O}(h)$. If the function to be approximated has sufficient *regularity*, we have

$$\begin{aligned} f'(x_i) &= \frac{f_{i+1} - f_{i-1}}{2h} + \mathcal{O}(h^2) =: (\delta_x f)_i + \mathcal{O}(h^2), && \text{for } f \in C^3(G), \\ f''(x_i) &= \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + \mathcal{O}(h^2) =: (\delta_{xx}^2 f)_i + \mathcal{O}(h^2), && \text{for } f \in C^4(G). \end{aligned}$$

With the difference quotients we turn next to the finite difference discretization of the heat equation (2.4). Let u_i^m denote the approximate value of the solution u at grid point (t_m, x_i), i.e. $u_i^m \approx u(t_m, x_i)$. For a parameter $\theta \in [0, 1]$, we approximate the partial differential operator $\partial_t u - \partial_{xx} u$ at the grid point (t_m, x_i) by the finite difference operator

$$\begin{aligned} \mathcal{E}_i^m &:= \frac{u_i^{m+1} - u_i^m}{k} - \left[(1-\theta)(\delta_{xx}^2 u)_i^m + \theta(\delta_{xx}^2 u)_i^{m+1}\right] \\ &= \frac{u_i^{m+1} - u_i^m}{k} - \left[(1-\theta)\frac{u_{i+1}^m - 2u_i^m + u_{i-1}^m}{h^2} + \theta\frac{u_{i+1}^{m+1} - 2u_i^{m+1} + u_{i-1}^{m+1}}{h^2}\right], \end{aligned} \tag{2.7}$$

and replace the partial differential equation (2.4) by the *finite difference equations*

$$\mathcal{E}_i^m = \theta f_i^{m+1} + (1-\theta) f_i^m, \quad i = 1, \ldots, N, \quad m = 0, \ldots, M-1, \tag{2.8}$$

with initial conditions $u_i^0 = u_0(x_i)$, $i = 1, \ldots, N$, and boundary conditions $u_k^m = 0$, $k \in \{0, N+1\}$, $m = 0, \ldots, M$. We observe that for $\theta = 0$, u_i^{m+1}, $i = 1, \ldots, N$, are given in $\mathcal{E}_i^m = f_i^m$ *explicitly* in terms of u_i^m, i.e. the scheme (2.8) is explicit. For $\theta = 1$, a linear system of equations must be solved for u_i^{m+1} at each time step, i.e. the scheme is *implicit*.

We write (2.8) in matrix form. To this end, we introduce the column vectors

$$\underline{u}^m = (u_1^m, \ldots, u_N^m)^\top, \quad \underline{f}^m = (f_1^m, \ldots, f_N^m)^\top, \quad \underline{u}_0 = (u_0(x_1), \ldots, u_0(x_N))^\top,$$

and the tridiagonal matrices

$$\mathbf{G} = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{N\times N}_{\mathrm{sym}}, \quad \mathbf{I} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \in \mathbb{R}^{N\times N}_{\mathrm{sym}}. \tag{2.9}$$

Then, after multiplication by k, the finite difference scheme (2.8) becomes:

Find $\underline{u}^{m+1} \in \mathbb{R}^N$ such that for $m = 0, \ldots, M-1$,

$$\big(\mathbf{I} + \theta k \mathbf{G}\big)\underline{u}^{m+1} = \big(\mathbf{I} - (1-\theta)k\mathbf{G}\big)\underline{u}^m + k(\theta \underline{f}^{m+1} + (1-\theta)\underline{f}^m), \tag{2.10}$$

$$\underline{u}^0 = \underline{u}_0.$$

We now show that the vectors $\underline{u}^m$ converge towards the exact solution as $k \to 0$ and $h \to 0$.

2.3.2 Convergence of the Finite Difference Method

Naturally, by the transition from the PDE (2.4) to the finite difference equations (2.8), which is called the *discretization* of the PDE, an error is introduced, the so-called *discretization error*, which we analyze next. We begin with the definition of a related *consistency error*.

Definition 2.3.1 The consistency error E_i^m at (t_m, x_i) is the difference scheme (2.7) with u_i^m in $\mathcal{E}_i^m$ replaced by $u(t_m, x_i)$.

Using Taylor expansions of the exact solution at the grid point (t_m, x_i), we can readily estimate the consistency errors E_i^m in terms of powers of the mesh width h and the time step size k.

Proposition 2.3.2 *If the exact solution $u(t, x)$ of (2.4) is sufficiently smooth, then, as $h \to 0$, $k \to 0$, the following estimates hold for $m = 1, \ldots, M-1$ and $i = 1, \ldots, N$:*

$$|E_i^m| \le C(u)(h^2 + k), \quad 0 \le \theta \le 1, \tag{2.11}$$

$$|E_i^m| \le C(u)(h^2 + k^2), \quad \theta = \frac{1}{2}, \tag{2.12}$$

where the constant $C(u) > 0$ depends on the exact solution u and its derivatives.

For the convergence of the FDM, we are interested in estimating the error between the finite difference solution u_i^m and the exact solution $u(t, x)$ at the grid point (t_m, x_i). We collect the discretization errors in the grid points at time t_m in the *error vector* $\underline{\varepsilon}^m$, i.e.

$$\varepsilon_i^m := u(t_m, x_i) - u_i^m, \quad 0 \le i \le N+1, \quad 0 \le m \le M. \tag{2.13}$$

The error vectors $\{\underline{\varepsilon}^m\}_{m=0}^M$ satisfy the difference equation

$$\big(k^{-1}\mathbf{I} + \theta \mathbf{G}\big)\underline{\varepsilon}^{m+1} + \big(-k^{-1}\mathbf{I} + (1-\theta)\mathbf{G}\big)\underline{\varepsilon}^m = \underline{E}^m \tag{2.14}$$

or, in explicit form,

$$\underline{\varepsilon}^{m+1} = \mathbf{A}_\theta \underline{\varepsilon}^m + \underline{\eta}^m, \tag{2.15}$$

where $\underline{\eta}^m := (k^{-1}\mathbf{I} + \theta\mathbf{G})^{-1}\underline{E}^m$, and where $\mathbf{A}_\theta := (k^{-1}\mathbf{I} + \theta\mathbf{G})^{-1}(-k^{-1}\mathbf{I} + (1-\theta)\mathbf{G})$, is called an *amplification matrix.*

The recursion (2.15) shows that the discretization error ε_i^m is related to the consistency error E_i^m. Estimates on ε_i^m can be obtained by taking norms in the recursion (2.15). Using induction on m, we have

Proposition 2.3.3 *For all $M \in \mathbb{N}$, $1 \le m \le M$, one has*

$$\|\underline{\varepsilon}^m\|_{\ell_2} \le \|\mathbf{A}_\theta\|_2^m, \|\underline{\varepsilon}^0\|_{\ell_2} + k\sum_{n=0}^{m-1} \|\mathbf{A}_\theta\|_2^{m-1-n} \|\underline{E}^n\|_{\ell_2}, \tag{2.16}$$

where $\|\underline{\varepsilon}^m\|_{\ell_2}^2 = \sum_{i=0}^{N+1} |\varepsilon_i^m|^2$.

We see from (2.16) that the discretization errors ε_i^m can be controlled in terms of the consistency errors E_i^m *provided* the norm $\|\mathbf{A}_\theta\|_2$ is bounded by 1. The condition that the norm of the amplification matrix $\mathbf{A}_\theta$ is bounded by 1 is a *stability condition* for the FDM. We obtain immediately

Theorem 2.3.4 *If the stability condition*

$$\|\mathbf{A}_\theta\|_2 \le 1 \tag{2.17}$$

holds, then, as $M \to \infty$, $N \to \infty$, the FDM (2.10) *converges and, if $\underline{\varepsilon}^0 = \underline{0}$,*

$$\sup_m \|\underline{\varepsilon}^m\|_{\ell_2} \le T \sup_m \|\underline{E}^m\|_{\ell_2}. \tag{2.18}$$

We want to discuss the validity of the stability condition (2.17) for $\mathbf{G}$ as in (2.9). Therefore, we need the following lemma to obtain the eigenvalues for a tridiagonal matrix. It follows immediately by elementary calculations.

Lemma 2.3.5 *Let* $\mathbf{X} \in \mathbb{R}^{(N-1)\times(N-1)}$ *be a tridiagonal matrix given by*

$$\mathbf{X} = \begin{pmatrix} \alpha & \beta & & \\ \gamma & \alpha & \ddots & \\ & \ddots & \ddots & \beta \\ & & \gamma & \alpha \end{pmatrix}$$

Then, $\mathbf{X}\underline{v}^{(\ell)} = \mu_\ell \underline{v}^{(\ell)}$, $\ell = 1, \dots, N-1$, *with the eigensystem*

$$\mu_\ell = \alpha + 2\beta\sqrt{\beta^{-1}\gamma}\cos\bigl(N^{-1}\ell\pi\bigr), \quad \underline{v}^{(\ell)} = \bigl((\beta^{-1}\gamma)^{j/2}\sin(N^{-1}j\ell\pi)\bigr)_{j=1}^{N-1}.$$

For $\mathbf{G}$ resulting from the finite differences discretization, we obtain

Corollary 2.3.6 *The eigensystem of the matrix* $\mathbf{G}$ *in* (2.9) *is given by*

$$\mu_\ell = \frac{4}{h^2}\sin^2\Bigl(\frac{\ell\pi}{2(N+1)}\Bigr), \quad \underline{v}^{(\ell)} = \Bigl(\sin\Bigl(\frac{j\ell\pi}{N+1}\Bigr)\Bigr)_{j=1}^{N}, \quad \ell = 1, \dots, N. \tag{2.19}$$

Using Corollary 2.3.6, we find that

$$\lambda_\ell(\mathbf{A}_\theta) = \frac{4(1-\theta)h^{-2}k\sin^2(\ell\pi/(2(N+1))) - 1}{1 + 4\theta h^{-2}k\sin^2(\ell\pi/(2(N+1)))}, \quad \ell = 1, \dots, N.$$

Since $\|\mathbf{A}_\theta\|_2 = \max_\ell |\lambda_\ell|$, we have $\|\mathbf{A}_\theta\|_2 \le 1$, if $2(1-2\theta)h^{-2}k \le 1$. For $0 \le \theta < \frac{1}{2}$, we obtain the so-called *CFL-condition* (after the seminal paper of Courant, Friedrichs and Lewy [43]),

$$\frac{k}{h^2} \le \frac{1}{2(1-2\theta)}. \tag{2.20}$$

Therefore, we obtain directly from Theorem 2.3.4:

Lemma 2.3.7 (Stability of the θ-scheme)

(i) *If* $\frac{1}{2} \le \theta \le 1$, *the scheme* (2.10) *is stable for all* k *and* h.
(ii) *If* $0 \le \theta < \frac{1}{2}$, *the scheme* (2.10) *is stable if and only if the CFL-condition* (2.20) *holds.*

We can now combine the results obtained so far using Lemma 2.3.7, Proposition 2.3.3 and Proposition 2.3.2 to obtain a convergence result for the θ-scheme. We measure the error using the quantity $\sup_m h^{\frac{1}{2}}\|\underline{e}^m\|_{\ell_2}$ which is a discrete version of the $L^\infty(J; L^2(G))$-norm.

Theorem 2.3.8 *If $u \in C^4(\overline{J} \times \overline{G})$, we have*

(i) *For $\frac{1}{2} < \theta \leq 1$ or for $0 \leq \theta < \frac{1}{2}$ and (2.20),*

$$\sup_m h^{\frac{1}{2}} \|\underline{\varepsilon}^m\|_{\ell_2} \leq C(u)(h^2 + k);$$

(ii) *For $\theta = \frac{1}{2}$,*

$$\sup_m h^{\frac{1}{2}} \|\underline{\varepsilon}^m\|_{\ell_2} \leq C(u)(h^2 + k^2),$$

where the constant $C(u) > 0$ depends on the exact solution u and its derivatives.

We next explain the finite element method which is based on variational formulations of the differential equations.

2.3.3 Finite Element Method

For the discretization with finite elements, we use the *method of lines* where we first only discretize in space to obtain a system of coupled ordinary differential equations (ODEs). In a second step, a time discretization scheme is applied to solve the ODEs. We do not require the PDE (2.4) to hold pointwise in space but only in the variational sense. Therefore, we fix $t \in J$ and let $v \in C_0^\infty(G)$ be a smooth test function satisfying $v(a) = v(b) = 0$. We multiply the PDE with v, integrate with respect to the space variable x and use integration by parts to obtain

$$\int_a^b \partial_t u v \,\mathrm{d}x - \int_a^b \partial_{xx} u v \,\mathrm{d}x = \int_a^b f v \,\mathrm{d}x,$$

$$\Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t} \int_a^b u v \,\mathrm{d}x - \underbrace{\Big[\partial_x u(x,t) v(x)\Big]_{x=a}^{x=b}}_{=0} + \int_a^b \partial_x u \partial_x v \,\mathrm{d}x = \int_a^b f v \,\mathrm{d}x.$$

Therefore, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_a^b u v \,\mathrm{d}x + \int_a^b \partial_x u \partial_x v \,\mathrm{d}x = \int_a^b f v \,\mathrm{d}x, \quad \forall v \in C_0^\infty(G). \tag{2.21}$$

Since $C_0^\infty(G)$ is not a closed subspace of $L^2(G)$, we will consider test functions in the *Sobolev space* $H_0^1(G)$ which is the closure of $C_0^\infty(G)$ in the H^1-norm,

$$\|u\|_{H^1(G)}^2 = \|u\|_{L^2(G)}^2 + \|u'\|_{L^2(G)}^2.$$

The Sobolev space $H_0^1(G)$ consists of all continuous functions which are piecewise differentiable and vanish at the boundary. Since (2.21) holds for all $v \in C_0^\infty(G)$, (2.21) also holds for all $v \in H_0^1(G)$ because $C_0^\infty(G)$ is dense in $H_0^1(G)$. The *weak* or *variational formulation* of (2.4) reads:

$$\begin{aligned}
&\text{Find } u \in C(J, H_0^1(G)) \cap C^1(J, L^2(G)) \text{ such that for } t \in J \\
&\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b uv\,\mathrm{d}x + \int_a^b \partial_x u \partial_x v\,\mathrm{d}x = \int_a^b fv\,\mathrm{d}x, \quad \forall v \in H_0^1(G), \\
&u(0,\cdot) = u_0,
\end{aligned} \tag{2.22}$$

where we assume that the initial condition $u_0 \in L^2(G)$. The finite element method is based on the *Galerkin discretization* of (2.22). The idea is to project (2.22) to a finite dimensional subspace $V_N \subset H_0^1(G)$ and to replace (2.22) by:

$$\begin{aligned}
&\text{Find } u_N \in C^1(J, V_N), \text{ such that for } t \in J \\
&\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u_N v_N\,\mathrm{d}x + \int_a^b \partial_x u_N \partial_x v_N\,\mathrm{d}x = \int_a^b f v_N\,\mathrm{d}x, \quad \forall v_N \in V_N, \\
&u_N(0) = u_{N,0},
\end{aligned} \tag{2.23}$$

where $u_{N,0}$ is an approximation of u_0 in V_N. For example, $u_{N,0} = \mathcal{P}_N u_0$, the L^2-projection of u_0 on V_N, satisfying $\int_G \mathcal{P}_N u_0 v_N\,\mathrm{d}x = \int_G u_0 v_N\,\mathrm{d}x$ for all $v_N \in V_N$.

We show that (2.23) is equivalent to a linear system of ordinary differential equations (in time). Let b_j, $j = 1, \ldots, N$ be a basis of V_N. Since $u_N(t,\cdot) \in V_N$, we have

$$u_N(t,x) = \sum_{j=1}^N u_{N,j}(t) b_j(x),$$

where $u_{N,j}(t)$ denote the time dependent coefficients of u_N with respect to the basis of V_N. Inserting this series representations into (2.23) yields for $v_N(x) = b_i(x)$, $i = 1, \ldots, N$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u_N(t,x) v_N(x)\,\mathrm{d}x + \int_a^b \partial_x u_N(t,x) \partial_x v_N(x)\,\mathrm{d}x = \int_a^b f(t,x) v_N(x)\,\mathrm{d}x, \\
&\Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\int_a^b \sum_{j=1}^N u_{N,j}(t) b_j(x) b_i(x)\,\mathrm{d}x + \int_a^b \sum_{j=1}^N u_{N,j}(t) b_j'(x) b_i'(x)\,\mathrm{d}x \\
&= \int_a^b f(t,x) b_i(x)\,\mathrm{d}x, \\
&\Rightarrow \quad \sum_{j=1}^N \dot{u}_{N,j} \int_a^b b_j(x) b_i(x)\,\mathrm{d}x + \sum_{j=1}^N u_{N,j} \int_a^b b_j'(x) b_i'(x)\,\mathrm{d}x \\
&= \int_a^b f(t,x) b_i(x)\,\mathrm{d}x.
\end{aligned}$$

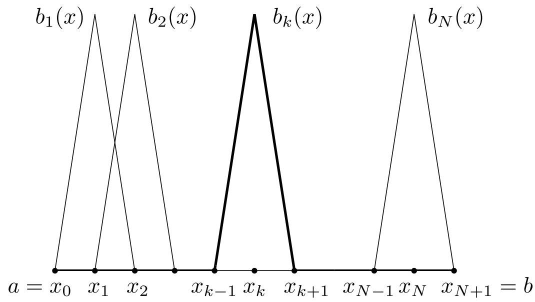


Fig. 2.2 Basis functions $b_i(x)$, $i = 1, \ldots, N$

With matrices $\mathbf{M}, \mathbf{A} \in \mathbb{R}^{N \times N}$ and vector $\underline{f}(t) \in \mathbb{R}^N$ given by

$$\mathbf{M}_{ij} := \int_a^b b_j(x) b_i(x)\,\mathrm{d}x, \quad \mathbf{A}_{ij} := \int_a^b b_j'(x) b_i'(x)\,\mathrm{d}x,$$
$$f_i(t) := \int_a^b f(t,x) b_i(x)\,\mathrm{d}x,$$

we obtain the weak semi-discretization (2.23) in *matrix form*:

$$\begin{aligned} &\text{Find } \underline{u}_N \in C^1(J; \mathbb{R}^N), \text{ such that for } t \in J \\ &\mathbf{M}\dot{\underline{u}}_N(t) + \mathbf{A}\underline{u}_N(t) = \underline{f}(t), \\ &\underline{u}_N(0) = \underline{u}_0, \end{aligned} \tag{2.24}$$

where $\underline{u}_0$ denotes the coefficient vector of $u_{N,0} = \sum_{i=1}^N u_{0,i} b_i$.

For the basis functions b_i of $V_N = \mathrm{span}\{b_i(x) : i = 1, \ldots, N\}$, we take the so-called *hat functions*

$$b_i : [a,b] \to \mathbb{R}_{\geq 0}, \quad b_i(x) = \max\{0, 1 - h^{-1}|x - x_i|\}, \quad i = 1, \ldots, N,$$

as illustrated in Fig. 2.2.

For equidistant mesh points x_i, $i = 1, \ldots, N$ with mesh width h as in (2.5),

$$x_i = a + ih, \ i = 0, 1, \ldots, N+1, \ h := (b-a)/(N+1) = \Delta x,$$

we obtain the matrices, $\mathbf{M}, \mathbf{A} \in \mathbb{R}^{N \times N}_{\mathrm{sym}}$, given by

$$\mathbf{M} = \frac{h}{6} \begin{pmatrix} 4 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 4 \end{pmatrix}, \quad \mathbf{A} = \frac{1}{h} \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}. \tag{2.25}$$

Table 2.1 Difference between finite differences and finite elements

	FDM	FEM
$\underline{u}^m$	vector of $u_i^m \approx u(t_m, x_i)$	coeff. vector of $u_N(t_m, x)$
$\mathbf{B}$	$\mathbf{I} + k\theta\mathbf{G}$	$\mathbf{M} + k\theta\mathbf{A}$
$\mathbf{C}$	$\mathbf{I} - k(1-\theta)\mathbf{G}$ $\mathbf{G} = h^{-2}\text{tridiag}(-1, 2, -1)$	$\mathbf{M} - k(1-\theta)\mathbf{A}$ $\mathbf{A} = h^{-1}\text{tridiag}(-1, 2, -1)$
f_i^m	$f(t_m, x_i)$	$\int_a^b f(t_m, x) b_i(x)\,\mathrm{d}x$

It remains to discretize the ODE (2.24). Proceeding exactly as in the FDM, we choose time levels t_m, $m = 0, \ldots, M$ as in (2.6)

$$t_m = mk,\ m = 0, 1, \ldots, M,\ k := T/M = \Delta t,$$

and denote $\underline{u}_N^m := u_N(t_m)$ and $\underline{f}^m := \underline{f}(t_m)$. Then, the fully discrete scheme reads:

$$\begin{aligned} &\text{Find } \underline{u}_N^{m+1} \in \mathbb{R}^N \text{ such that for } m = 0, \ldots, M-1, \\ &(\mathbf{M} + k\theta\mathbf{A})\underline{u}_N^{m+1} = (\mathbf{M} - k(1-\theta)\mathbf{A})\underline{u}_N^m + k\big(\theta\underline{f}^{m+1} + (1-\theta)\underline{f}^m\big), \\ &\underline{u}_N^0 = \underline{u}_0. \end{aligned} \tag{2.26}$$

Thus, in both the finite difference and the finite element method we have to solve M systems of N linear equations of the form

$$\mathbf{B}\underline{u}^{m+1} = \mathbf{C}\underline{u}^m + k\underline{F}^m, \quad m = 0, \ldots, M-1,$$

where $\underline{F}^m = \theta\underline{f}^{m+1} + (1-\theta)\underline{f}^m$. The difference between FDM and FEM is shown in Table 2.1.

From Table 2.1 we see that both discretization schemes for the heat equation lead to similar linear systems of equations to be solved in each timestep. For all the partial differential equations which we will encounter in these note, we will use both finite differences and finite elements for the discretization.

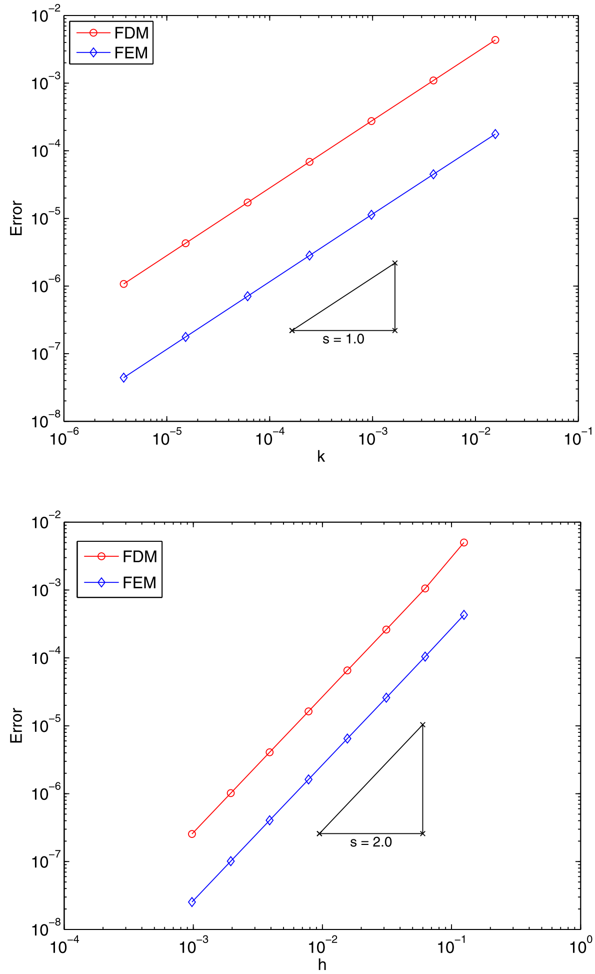
Example 2.3.9 Let $G = (0, 1)$, $T = 1$, and $u(t, x) = e^{-t} x \sin(\pi x)$. We measure the discrete $L^\infty(0, T; L^2(G))$-error defined by $\sup_m h^{\frac{1}{2}} \|\underline{\varepsilon}^m\|_{\ell_2}$ where

$$\|\underline{\varepsilon}^m\|_{\ell_2}^2 := \sum_{i=1}^N |u(t_m, x_i) - u_i^m|^2.$$

For $\theta = 1$ (backward Euler), we let $h = \mathcal{O}(\sqrt{k})$ and obtain first order convergence with respect to the time step k both for FDM and FEM, i.e.

$$\sup_m h^{\frac{1}{2}} \|\underline{\varepsilon}^m\|_{\ell_2} = \mathcal{O}(k). \tag{2.27}$$

Fig. 2.3 $L^\infty(J; L^2(G))$ convergence rates for $\theta = 1$ (*top*) and $\theta = \frac{1}{2}$ (*bottom*)



For $\theta = \frac{1}{2}$ (Crank–Nicolson), we let $k = \mathcal{O}(h)$ and obtain, in terms of the mesh width h, second order convergence for both FDM and FEM, i.e.

$$\sup_m h^{\frac{1}{2}} \|\underline{\varepsilon}^m\|_{\ell_2} = \mathcal{O}(h^2). \tag{2.28}$$

Both convergence rates are shown in Fig. 2.3.

The convergence rates (2.27)–(2.28) have been shown for the finite difference method in Theorem 2.3.8. In the next chapter, we show that these also hold for the finite element method.

2.4 Further Reading

A nice introduction to the mathematical theory of partial differential equations is given in Evans [65]. The mathematical theory of the finite difference and finite element methods for elliptic problems is introduced in the text of Braess [24], and for parabolic and hyperbolic equations in Larsson and Thomée [112]. Finite difference methods for time dependent problems are studied in more details in Gustafsson et al. [76]. For an elementary introduction to the finite element methods with particular attention to stabilized finite element methods for partial differential equations of the type which arise in finance, we refer to Johnson [99].