Chapter 5
Finite-Element Methods

The finite-difference approach with equidistant grids is easy to understand and straightforward to implement. Resulting uniform rectangular grids are comfortable, but in many applications not flexible enough. Steep gradients of the solution require a finer grid locally such that the difference quotients provide good approximations of the differentials. On the other hand, a flat gradient may be well modeled on a coarse grid. Arranging such a flexibility of the grid with finite-difference methods is possible but cumbersome.

An alternative type of methods for solving PDEs that does provide high flexibility is the class of finite-element methods (FEM). A "finite element" designates a mathematical topic such as an interval and thereupon defined a piece of function. There are alternative names such as *variational methods*, or *weighted residuals*, or *Ritz–Galerkin methods*. These names hint at underlying principles that serve to derive suitable equations. As these different names suggest, there are several different approaches leading to finite elements. The methods are closely related.

The flexibility of finite-element methods is not only favorable to approximate functions, but also to approximate domains of computation that are not rectangular. This is important for multifactor options. For the one-dimensional situation of standard options, the possible improvement of a finite-element method over the standard methods of the previous chapter is not significant. With the focus on standard options, Chap. 5 may be skipped on first reading. But options with several underlyings may lead to domains of computation that are more "fancy."

For example, a two-asset basket with portfolio value $\alpha_1 S_1 + \alpha_2 S_2$ in the case of a call option leads to a payoff of type $\Psi(S_1, S_2) = (\alpha_1 S_1 + \alpha_2 S_2 - K)^+$. If such an option is endowed with barriers, then it is reasonable to set up barriers such that the payoff takes a constant value. For the two-asset basket, this amounts to barrier lines $\alpha_1 S_1 + \alpha_2 S_2 =$constant. This naturally leads to trapezoidal shapes of domains. For a special case with two knock-out barriers the payoff and the domain are illustrated by Fig. 5.1. This example will be considered in Sect. 5.4, see the domain in Fig. 5.5. In more complicated examples, the domain may be elliptic ($\longrightarrow$ Exercise 5.1). In

R.U. Seydel, *Tools for Computational Finance*, Universitext,
DOI 10.1007/978-1-4471-7338-0_5

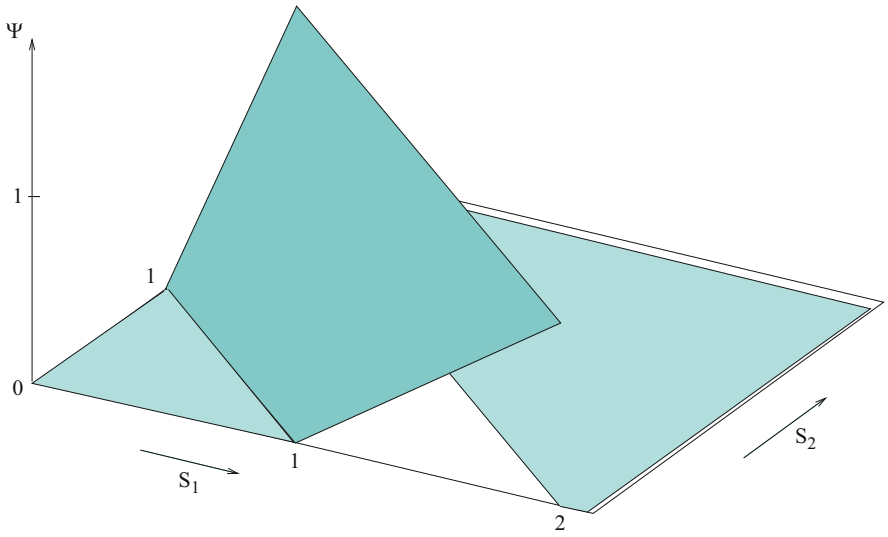

Fig. 5.1 Payoff $\Psi(S_1, S_2)$ of a call on a two-asset basket, with knock-out barrier (Example 5.6)

such situations of non-rectangular domains, finite elements are ideally applicable and highly recommendable.

Faced with the huge field of finite-element methods, in this chapter we confine ourselves to a step-by-step exposition towards the solution of two-asset options. We start with an overview on basic approaches and ideas (in Sect. 5.1). Then, in Sect. 5.2, we describe the approximation with the simplest finite elements, namely, piecewise straight-line segments, and apply this to a stationary model problem. These approaches will be applied to the time-dependent situation of pricing standard options, in Sect. 5.3. This sets the stage to the main application of FEM in financial engineering, options on two or more assets. Section 5.4 will present an application to an exotic option with two underlyings. Here we derive a weak form of the PDE, and discuss boundary conditions. Finally, in Sect. 5.5, we will introduce to error estimates. Methods more subtle than just the Taylor expansion of the discretization error are required to show that quadratic convergence is possible with unstructured grids and nonsmooth solutions. To keep the exposition of an error analysis short, we concentrate on the one-dimensional situation. But the ideas extend to multidimensional scenarios.

5.1 Weighted Residuals

Many of the principles on which finite-element methods are based, can be interpreted as weighted residuals. What does this mean? This heading points at ways in which a discretization can be set up, and how an approximation can be defined.

Fig. 5.2 Discretization of a continuum

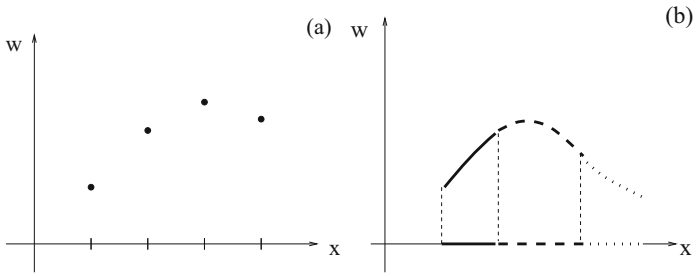

Fig. 5.3 Two kinds of approximations (one-dimensional situation)

There lies a duality in a discretization. This is illustrated by means of Fig. 5.2, which shows a partition of an x-axis. This discretization is either represented by

(a) discrete grid points x_i, or by
(b) a set of subintervals.

The two ways to see a discretization lead to different approaches of constructing an approximation w. Let us illustrate this with the one-dimensional situation of Fig. 5.3. An approximation w based on finite differences is built on the grid points and primarily consists of discrete points (Fig. 5.3a). In contrast, finite elements are founded on subdomains (intervals in Fig. 5.3b) with piecewise functions, which are defined by suitable criteria and constitute a global approximation w. In a narrower sense, a finite element is a pair consisting of one piece of subdomain and the corresponding function defined thereupon, mostly a polynomial. Figure 5.3 reflects the respective basic approaches; in a second step the isolated points of a finite-difference calculation can well be extended to continuous piecewise functions by means of interpolation (⟶ Appendix C.1).

A two-dimensional domain can be partitioned into triangles, for example, where w is again represented by piecewise polynomials. Figure 5.4 depicts the simplest such situation, namely, a triangle in an (x, y)-plane, and a piece of a linear function defined thereupon. Figure 5.5 below will provide an example how triangles easily fill a seemingly "irregular" domain.

As will be shown next, the approaches of finite-element methods use integrals. If done properly, integrals require less smoothness. This often matches applications better and adds to the flexibility of finite-element methods. The integrals can be derived in a natural way from minimum principles, or are constructed artificially. Finite elements based on polynomials make the calculation of the integrals easy.

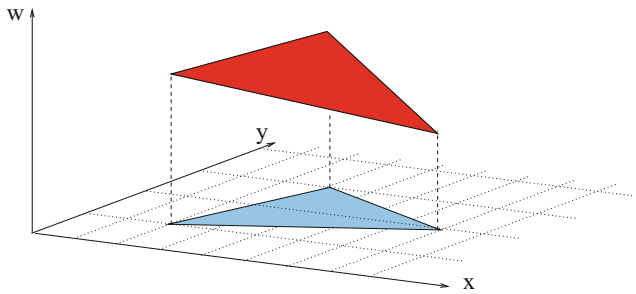

Fig. 5.4 A simple finite element in two dimensions, based on a triangle

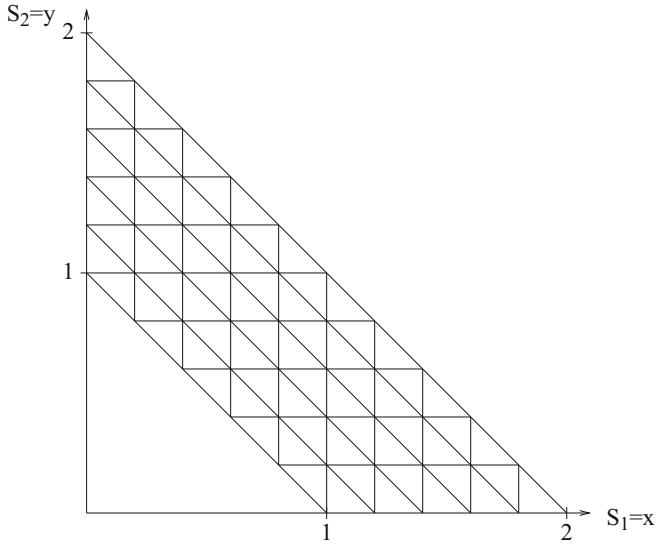

Fig. 5.5 A simple regular finite-element discretization of a domain $\mathcal{D}$ into triangles $\mathcal{D}_k$ (see Example 5.6)

5.1.1 The Principle of Weighted Residuals

To explain the principle of weighted residuals we discuss the formally simple case of the differential equation

$$Lu = f\,. \tag{5.1}$$

Here L symbolizes a linear differential operator. Important examples are

$$Lu := -u'' \text{ for } u(x), \text{ or} \tag{5.2}$$

$$Lu := -u_{xx} - u_{yy} \text{ for } u(x,y)\,. \tag{5.3}$$

The right-hand side f is a problem-dependent function. Solutions u of the differential equation (5.1) are studied on a domain $\mathcal{D} \subseteq \mathbb{R}^n$, with $n = 1$ in (5.2) and $n = 2$ in (5.3). The piecewise approach starts with a partition of the domain into a finite number m of subdomains $\mathcal{D}_k$,

$$\mathcal{D} = \bigcup_{k=1}^{m} \mathcal{D}_k \,. \tag{5.4}$$

All boundaries of $\mathcal{D}$ should be included, and approximations to u are calculated on the closure of $\mathcal{D}$. The partition is assumed disjoint up to the boundaries of $\mathcal{D}_k$, so $\mathcal{D}_j^\circ \cap \mathcal{D}_k^\circ = \emptyset$ for $j \neq k$. In the one-dimensional case ($n = 1$), for example, the $\mathcal{D}_k$ are subintervals of a whole interval $\mathcal{D}$. In the two-dimensional case, (5.4) may describe a partition into triangles, as illustrated in Fig. 5.5.

The ansatz for approximations w to a solution u is a basis representation with N basis functions φ_i,

$$w := \sum_{i=1}^{N} c_i \, \varphi_i \,. \tag{5.5}$$

The functions φ_i are also called *trial functions*. In the case of one independent variable x the $c_i \in \mathbb{R}$ are constant coefficients, and the φ_i are functions of x. Typically, N is chosen and $\varphi_1, \ldots, \varphi_N$ are prescribed. Depending on this choice, the free parameters $c_1, \ldots, c_N$ are to be determined such that $w \approx u$. The ansatz (5.5) was suggested by Ritz in 1908.

We have m subdomains and N basis functions. In the one-dimensional situation ($n = 1$), nodes and subintervals interlace, and m and N essentially can be identified. For $n = 1$ the two numbers m and N differ by at most one, depending on whether the solution is known or unknown at the end points of the interval $\mathcal{D}$. In the latter case it is convenient to have the summation index in (5.5) run as $i = 0, \ldots, m$. For dimensions $n > 1$ the number m of subdomains (e.g. triangles in case $n = 2$) in general is different from the number N of basis functions (nodes[1]). For example, in Fig. 5.5 we have 75 triangles and 51 nodes; 26 of the nodes are interior nodes and 25 are placed along the boundary. That is, $1 \leq k \leq 75$. The number N refers to the number of nodes for which a value of u is to be approximated.

One strategy to determine the coefficients c_i is based on the residual function

$$R(w) := Lw - f \,. \tag{5.6}$$

We look for a w such that the residual R becomes "small." Since the φ_i are considered prescribed, in view of (5.5) N conditions or equations must be established to define

[1]Basis functions can be constructed such that there is one for each node. Then N represents also the number of nodes.

and calculate the unknown $c_1, \ldots, c_N$. To this end we weight the residual R by introducing N weighting functions *(test functions)* $\psi_1, \ldots, \psi_N$ and require

$$\int_{\mathcal{D}} R(w)\, \psi_j \, \mathrm{d}\mathcal{D} = 0 \quad \text{for } j = 1, \ldots, N\,. \tag{5.7}$$

This amounts to the requirement that the residual be orthogonal to the set of weighting functions ψ_j. The "d$\mathcal{D}$" in (5.7) symbolizes the integration that matches $\mathcal{D} \subseteq \mathbb{R}^n$, as d$x$ for $n = 1$. For ease of notation, we frequently drop dx as well as the $\mathcal{D}$ at the n-dimensional integral. For the model problem (5.1) the system of Eqs. (5.7) consists of the N equations

$$\int_{\mathcal{D}} Lw\, \psi_j = \int_{\mathcal{D}} f\, \psi_j \quad (j = 1, \ldots, N) \tag{5.8}$$

for the N unknowns $c_1, \ldots, c_N$, which define w. Often the equations in (5.8) are written using a formulation with inner products,

$$(Lw, \psi_j) = (f, \psi_j)\,,$$

defined as the corresponding integrals in (5.8). For linear L the ansatz (5.5) implies

$$\int Lw\, \psi_j = \int \left(\sum_i c_i L\varphi_i \right) \psi_j = \sum_i c_i \underbrace{\int L\varphi_i \psi_j}_{=:a_{ij}}\,.$$

The integrals a_{ij} constitute a matrix A. The $r_j := \int f\psi_j$ set up the elements of a vector r and the coefficients c_j a vector $c = (c_1, \ldots, c_N)^{\mathrm{tr}}$. In vector notation the system of equations is rewritten as

$$Ac = r\,. \tag{5.9}$$

This outlines the general principle, but leaves open the questions how to handle boundary conditions and how to select basis functions φ_i and weighting functions ψ_j. The freedom to choose trial functions φ_i and test functions ψ_j allows to construct several different methods. For the time being suppose that these functions have sufficient potential to be differentiated or integrated. We will enter a discussion of relevant function spaces in Sect. 5.5.

5.1.2 Examples of Weighting Functions

We postpone the choice of basis functions φ_i and begin with listing important examples of how to select weighting functions ψ:

1.) **Galerkin's choice:**
Choose $\psi_j := \varphi_j$ for all j. Then $a_{ij} = \int L\varphi_i \varphi_j$.

2.) **Collocation:**
Choose $\psi_j := \delta(x - x_j)$. Here δ denotes Dirac's delta function, which in $\mathbb{R}^1$ satisfies $\int f\delta(x - x_j)\,\mathrm{d}x = f(x_j)$. As a consequence,

$$\int Lw\,\psi_j = Lw(x_j)\,, \quad \int f\psi_j = f(x_j)\,.$$

That is, a system of equations $Lw(x_j) = f(x_j)$ results, which amounts to evaluating the differential equation at selected points x_j.

3.) **Least squares:**
Choose

$$\psi_j := \frac{\partial R}{\partial c_j}\,.$$

This choice of test functions deserves its name *least-squares*, because to minimize $\int (R(c_1, \ldots, c_N))^2$ the necessary criterion is the vanishing of the gradient, so

$$\int_{\mathcal{D}} R\frac{\partial R}{\partial c_j} = 0 \quad \text{for all } j\,.$$

5.1.3 Examples of Basis Functions

The construction of suitable basis functions φ_i observes the underlying partition into subdomains $\mathcal{D}_k$. Our concern will be to meet two aims: resulting methods must be accurate, and their implementation should become efficient.

The efficiency can be focused on the sparsity of matrices. In particular, if the matrix A of the linear equations is sparse, then the system can be solved efficiently even when it is large. In order to achieve sparsity we require that $\varphi_i \equiv 0$ on most of the subdomains $\mathcal{D}_k$. Figure 5.6 illustrates an example for the one-dimensional

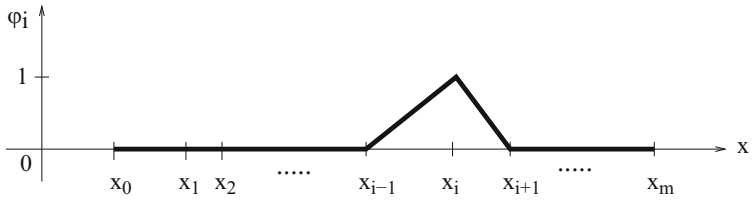

Fig. 5.6 "Hat function": simple choice of finite elements

case $n = 1$. This *hat function* of Fig. 5.6 is the simplest example related to finite elements. It is piecewise linear, and each function φ_i has a support consisting of only two subintervals, $\varphi_i(x) \neq 0$ for $x \in$ support. A consequence is

$$\int_{\mathcal{D}} \varphi_i \varphi_j = 0 \quad \text{for } |i-j| > 1\,, \tag{5.10}$$

as well as an analogous relation for $\int \varphi_i' \varphi_j'$. We will discuss hat functions in the following Sect. 5.2. Basis functions more advanced than the canonical hat functions are constructed using piecewise polynomials of higher degree. In this way, basis functions can be obtained with $\mathcal{C}^1$- or $\mathcal{C}^2$-smoothness (⟶ Exercise 5.2). Recall from interpolation (⟶ Appendix C.1) that polynomials of degree three can lead to $\mathcal{C}^2$-smooth splines.

5.1.4 Smoothness

We have left open how close an approximation w of (5.5)/(5.9) is to the solution u of (5.1). Clearly, $R(u) = 0$ and u satisfies (5.7). But w in general does not solve (5.1). The differential equation (5.1) is a stronger requirement than the integral relations (5.7).

The accuracy depends on the smoothness of the basis functions. Depending on the chosen method, different kinds of smoothness are relevant. Let us illustrate this matter on the model problem (5.2),

$$Lu = -u'', \quad \text{with } u, \varphi, \psi \in \{\, u \mid u(0) = u(1) = 0 \,\}\,.$$

Integration by parts formally implies

$$\int_0^1 \varphi'' \psi = -\int_0^1 \varphi' \psi' = \int_0^1 \varphi \psi''\,,$$

because the boundary conditions $u(0) = u(1) = 0$ let the nonintegral terms vanish. These three versions of the integral can be distinguished by the smoothness requirements on φ and ψ, and by the question whether the integrals exist. One will choose the integral version that corresponds to the underlying method, and to the smoothness of the solution. For example, for Galerkin's approach the elements a_{ij} of A consist of the integrals

$$-\int_0^1 \varphi_i' \varphi_j'\,.$$

We will return to the topics of accuracy, convergence, and function spaces in Sect. 5.5 (with Appendix C.3).

5.2 Ritz–Galerkin Method with One-Dimensional Hat Functions

As mentioned before, any required flexibility is provided by finite-element methods. This holds to a larger extent in higher-dimensional spaces. In this section, for simplicity, we stick to the one-dimensional situation, $x \in \mathbb{R}$. The dependence on the time variable t will be postponed to Sect. 5.3.

Assume a partition of the x-domain by a set of increasing mesh points $x_0, \ldots, x_m$. A nonuniform spacing is advisable in several instances in order to improve the accuracy. For example, close to the strike, a denser grid is appropriate to mollify the lack of smoothness of a payoff. In contrast, to model infinity, one rarefies the nodes for larger x and shifts the final node x_m to a large value. One strategy is to select a spacing such that locally (up to additional scaling and shifts) $\sinh(x_i) = \eta_i$, where η_i are chosen equidistantly. A dense spacing is also advisable for barrier options close to the barrier, where the gradient of option prices is high.

5.2.1 Hat Functions

The prototype of a finite-element method makes use of the hat functions, which we define formally (compare Figs. 5.6 and 5.7).

Definition 5.1 (Hat Functions) For $1 \le i \le m-1$ set $\varphi_i(x) := 0$ on all subintervals except two:

$$\varphi_i(x) := \frac{x - x_{i-1}}{x_i - x_{i-1}} \quad \text{for } x_{i-1} \le x < x_i\,,$$

$$\varphi_i(x) := \frac{x_{i+1} - x}{x_{i+1} - x_i} \quad \text{for } x_i \le x < x_{i+1}\,,$$

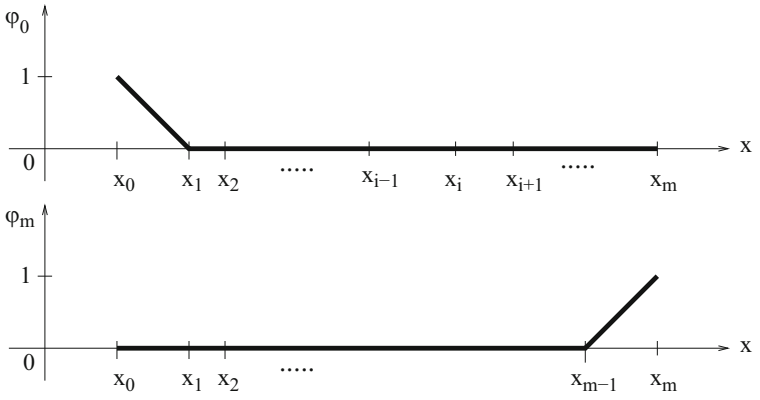

Fig. 5.7 Special "hat functions" φ_0 and φ_m

and boundary functions φ_0, φ_m nonzero on just one subinterval:

$$\varphi_0(x) := \frac{x_1 - x}{x_1 - x_0} \quad \text{for } x_0 \le x < x_1 \,,$$
$$\varphi_m(x) := \frac{x - x_{m-1}}{x_m - x_{m-1}} \quad \text{for } x_{m-1} \le x \le x_m \,.$$

For each node x_i there is one hat function. These $m + 1$ hat functions satisfy the following properties.

Properties 5.2 (Hat Functions) *The following properties (a)–(e) hold:*

(a) The $\varphi_0, \ldots, \varphi_m$ form a basis of the space of polygons

$$\{\, g \in \mathcal{C}^0[x_0, x_m] \;\mid\; g \text{ straight line on } \mathcal{D}_k := [x_k, x_{k+1}] \,, \text{ for all } k = 0, \ldots, m-1 \,\} \,.$$

That is to say, for each polygon v on the union of $\mathcal{D}_0, \ldots, \mathcal{D}_{m-1}$ there are unique coefficients $c_0, \ldots, c_m$ such that

$$v = \sum_{i=0}^{m} c_i \varphi_i \,.$$

(b) On any $\mathcal{D}_k$ only φ_k and $\varphi_{k+1} \neq 0$ are nonzero. Hence

$$\varphi_i \varphi_j = 0 \;\text{ for }\; |i - j| > 1 \,,$$

which explains (5.10).

(c) A simple approximation of the integral $\int_{x_0}^{x_m} f \varphi_j \, \mathrm{d}x$ can be calculated as follows: Substitute f by the interpolating polygon

$$f_{\mathrm{p}} := \sum_{i=0}^{m} f_i \varphi_i \quad \text{, where} \quad f_i := f(x_i) \,,$$

and obtain for each j the approximating integral

$$I_j := \int_{x_0}^{x_m} f_{\mathrm{p}} \varphi_j \, \mathrm{d}x = \int_{x_0}^{x_m} \sum_{i=0}^{m} f_i \varphi_i \varphi_j \, \mathrm{d}x = \sum_{i=0}^{m} f_i \underbrace{\int_{x_0}^{x_m} \varphi_i \varphi_j \, \mathrm{d}x}_{=: b_{ij}} \,.$$

The b_{ij} constitute a symmetric matrix B and the f_i a vector $\bar{f}$. If we arrange all integrals $I_j\,(0 \le j \le m)$ into a vector, then all integrals can be written in a compact way in vector notation as

$$B\bar{f}\,.$$

This will approximate the vector r in (5.9).

(d) The "large" $(m+1)^2$–matrix $B := (b_{ij})$ can be set up $\mathcal{D}_k$-elementwise by (2×2)-matrices (discussed below in Sect. 5.2.2). The (2×2)-matrices are those integrals that integrate only over a single subdomain $\mathcal{D}_k$. For each $\mathcal{D}_k$ in our one-dimensional setting exactly the four integrals $\int \varphi_i \varphi_j \mathrm{d}x$ for $i,j \in \{k, k+1\}$ are nonzero. They can be arranged into a (2×2)-matrix

$$\int_{x_k}^{x_{k+1}} \begin{pmatrix} \varphi_k^2 & \varphi_k\varphi_{k+1} \\ \varphi_{k+1}\varphi_k & \varphi_{k+1}^2 \end{pmatrix} \mathrm{d}x\,.$$

(The integral over a matrix is understood elementwise.) These are the integrals on $\mathcal{D}_k$, where the integrand is a product of the factors

$$\frac{x_{k+1}-x}{x_{k+1}-x_k} \quad \text{and} \quad \frac{x-x_k}{x_{k+1}-x_k}\,.$$

The four numbers

$$\frac{1}{(x_{k+1}-x_k)^2} \int_{x_k}^{x_{k+1}} \begin{pmatrix} (x_{k+1}-x)^2 & (x_{k+1}-x)(x-x_k) \\ (x-x_k)(x_{k+1}-x) & (x-x_k)^2 \end{pmatrix} \mathrm{d}x$$

result. With $h_k := x_{k+1} - x_k$ integration yields the *element-mass matrix* ($\longrightarrow$ Exercise 5.3)

$$\frac{1}{6} h_k \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\,.$$

(e) Analogously, integrating $\varphi_i' \varphi_j'$ yields

$$\begin{aligned} &\int_{x_k}^{x_{k+1}} \begin{pmatrix} \varphi_k'^2 & \varphi_k'\varphi_{k+1}' \\ \varphi_{k+1}'\varphi_k' & \varphi_{k+1}'^2 \end{pmatrix} \mathrm{d}x \\ &= \frac{1}{h_k^2} \int_{x_k}^{x_{k+1}} \begin{pmatrix} (-1)^2 & (-1)1 \\ 1(-1) & 1^2 \end{pmatrix} \mathrm{d}x = \frac{1}{h_k} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\,. \end{aligned}$$

These matrices are called *element-stiffness matrices*. They are used to set up the matrix A.

5.2.2 Assembling

The next step is to assemble the matrices A and B. It might be tempting to organize this task as follows: run a double loop on all basis indices i, j (N node indices) and check for each (i, j) on which $\mathcal{D}_k$ the integral

$$\int_{\mathcal{D}_k} \varphi_i \varphi_j$$

is nonzero. Such a procedure of performing a double loop has the complexity of $O(N^2 m)$. This is cumbersome as compared to the alternative of running a single loop on the subdomain index k and benefit from all relevant integrals on $\mathcal{D}_k$, which are precalculated above (Fig. 5.8).

To this end, split the integrals

$$\int_{x_0}^{x_m} = \sum_{k=0}^{m-1} \int_{\mathcal{D}_k}$$

to construct the $(m+1) \times (m+1)$-matrices $A = (a_{ij})$ and $B = (b_{ij})$ *additively* out of the small element matrices. For the case of the one-dimensional hat functions with

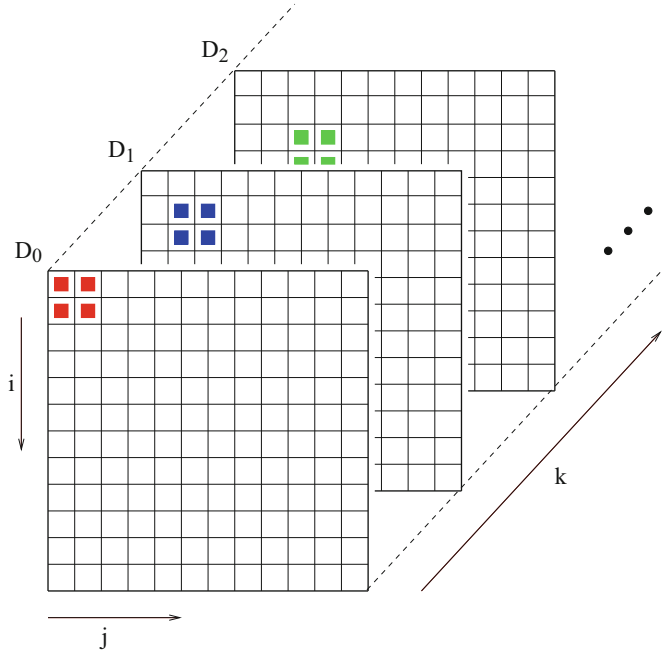

Fig. 5.8 Assembling in the one-dimensional setting

subintervals

$$\mathcal{D}_k = \{\, x \mid x_k \le x \le x_{k+1} \,\}$$

the element matrices are (2×2), see above. In this case only those integrals of $\varphi_i' \varphi_j'$ and $\varphi_i \varphi_j$ are nonzero, for which $i, j \in \mathcal{I}_k$, where

$$i, j \in \mathcal{I}_k := \{k, k+1\}\,. \tag{5.11}$$

$\mathcal{I}_k$ is the set of indices of those products of basis functions that are nonzero on $\mathcal{D}_k$. The *assembling algorithm* performs a loop over the subdomain index $k = 0, 1, \ldots, m-1$ and distributes the (2×2)-element matrices additively to the positions $i, j \in \mathcal{I}_k$. Before the assembling is started, the matrices A and B must be initialized with zeros. For $k = 0, \ldots, m-1$ one obtains for A the $(m+1)^2$-matrix

$$A = \begin{pmatrix} \frac{1}{h_0} & -\frac{1}{h_0} & & & \\ -\frac{1}{h_0} & \frac{1}{h_0} + \frac{1}{h_1} & -\frac{1}{h_1} & & \\ & -\frac{1}{h_1} & \frac{1}{h_1} + \frac{1}{h_2} & -\frac{1}{h_2} & \\ & & -\frac{1}{h_2} & \ddots & \ddots \\ & & & \ddots & \end{pmatrix}. \tag{5.12}$$

The matrix B is assembled in an analogous way. In the one-dimensional situation the matrices are tridiagonal. For an equidistant grid with $h = h_k$ the matrix A specializes to

$$A = \frac{1}{h} \begin{pmatrix} 1 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & 2 & -1 \\ 0 & & & & -1 & 1 \end{pmatrix} \tag{5.13}$$

and B to

$$B = \frac{h}{6} \begin{pmatrix} 2 & 1 & & & 0 \\ 1 & 4 & 1 & & \\ & 1 & 4 & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & 4 & 1 \\ 0 & & & & 1 & 2 \end{pmatrix}. \tag{5.14}$$

5.2.3 A Simple Application

In order to demonstrate the procedure, let us consider the simple time-independent ("stationary") model boundary-value problem

$$Lu := -u'' = f \quad \text{with} \quad u(x_0) = u(x_m) = 0\,. \tag{5.15}$$

Substituting $w := \sum_{i=0}^{m} c_i\varphi_i$ into the differential equation, in view of (5.8), leads to

$$\sum_{i=0}^{m} c_i \int_{x_0}^{x_m} L\varphi_i\,\varphi_j\,\mathrm{d}x = \int_{x_0}^{x_m} f\varphi_j\,\mathrm{d}x\,.$$

This is the result of the Ritz–Galerkin approach. Next we apply integration by parts on the left-hand side, and invoke Property 5.2(c) on the right-hand side. The resulting system of equations is

$$\sum_{i=0}^{m} c_i \underbrace{\int_{x_0}^{x_m} \varphi_i'\varphi_j'\,\mathrm{d}x}_{a_{ij}} = \sum_{i=0}^{m} f_i \underbrace{\int_{x_0}^{x_m} \varphi_i\varphi_j\,\mathrm{d}x}_{b_{ij}}, \quad j = 0, 1, \ldots, m\,. \tag{5.16}$$

This system is preliminary because the homogeneous boundary conditions $u(x_0) = u(x_m) = 0$ are not yet taken into account.

At this state, the preliminary system of Eqs. (5.16) can be written as

$$Ac = B\bar{f}\,. \tag{5.17}$$

It is easy to see that the matrix A from (5.13) is singular, because

$$A(1, 1, \ldots, 1)^{tr} = 0\,.$$

The singularity reflects the fact that the system (5.17) does not have a unique solution. This is consistent with the differential equation $-u'' = f(x)$: If $u(x)$ is solution, then also $u(x) + \alpha$ for arbitrary α. Unique solvability is attained by satisfying the boundary conditions; a solution u of $-u'' = f$ must be fixed by at least one essential boundary condition. For our example (5.15) we know in view of $u(x_0) = u(x_m) = 0$ the coefficients $c_0 = c_m = 0$. This information can be inserted into the system of equations in such a way that the matrix A changes to a nonsingular matrix without losing symmetry. To this end, cancel the first and the last of the $n + 1$ equations in (5.17), and make use of $c_0 = c_m = 0$. Now the inner part of size $(m - 1) \times (m - 1)$ of A remains. The matrix B is $(m - 1) \times (m + 1)$.

Finally, for the special case of an equidistant grid, the system of equations is

$$\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{m-2} \\ c_{m-1} \end{pmatrix} = \frac{h^2}{6} \begin{pmatrix} 1 & 4 & 1 & & & 0 \\ & 1 & 4 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 4 & 1 \\ 0 & & & & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} \bar{f}_0 \\ \bar{f}_1 \\ \vdots \\ \bar{f}_{m-1} \\ \bar{f}_m \end{pmatrix}. \tag{5.18}$$

In (5.18) we have used an equidistant grid for sake of a lucid exposition. Our main focus is the nonequidistant version, which is also implemented easily. In case nonhomogeneous boundary conditions are prescribed, appropriate values of c_0 or c_m are predefined. The importance of finite-element methods in structural engineering has lead to call the global matrix A the stiffness matrix, and B is called the mass matrix.

5.3 Application to Standard Options

Finite elements are especially advantageous in higher-dimensional spaces (several underlyings). But it also works for the one-dimensional case of standard options. This is the theme of this section. In contrast to the previous section, time must be included.

5.3.1 European Options

We know that the valuation of single-asset European options with vanilla payoff makes use of the Black–Scholes formula. But for the sake of exposition, and for non-vanilla payoff, let us briefly sketch a finite-element approach. Here we apply the FEM approach to the transformed version $y_\tau = y_{xx}$ of the Black–Scholes equation with constant parameters. In view of the general basis representation in (5.5) one may think of starting from $w = \sum w_i \varphi_i(x, \tau)$ with constant coefficients w_i. This would require two-dimensional basis functions. (We shall come back to such functions in Sect. 5.4.) To make use of one-dimensional hat functions, apply a separation ansatz in the form $\sum w_i(\tau)\varphi_i(x)$ with functions $w_i(\tau)$. As a consequence

of this simple approach, the same x-grid is applied for all τ, which results in a rectangular grid in the (x, τ)-plane. Dirichlet boundary conditions

$$y(x_{\min}, \tau) = \alpha(\tau),\ y(x_{\max}, \tau) = \beta(\tau)$$

mean that in view of the shape of φ_0, φ_m (Definition 5.1, Fig. 5.7) the values $w_0 = \alpha$ or $w_m = \beta$ would be known. It is practical to separate known terms and restrict the sum to the terms with unknown weights w_i. This can be managed by introducing a special function φ_b that compensates for Dirichlet boundary conditions on y. The function $\varphi_\mathrm{b}(x, \tau)$ is no basis function, and is constructed in advance. For example,

$$\varphi_\mathrm{b}(x, \tau) := (\beta(\tau) - \alpha(\tau)) \frac{x - x_{\min}}{x_{\max} - x_{\min}} + \alpha(\tau)$$

does the job for the above boundary conditions. So φ_b can be considered to be known, and the sum $\sum w_i \varphi_i$ does not reflect any nonzero Dirichlet boundary conditions on y. Then the final ansatz is

$$\sum_i w_i(\tau)\varphi_i(x) + \varphi_\mathrm{b}(x, \tau)\,, \tag{5.19}$$

and the index i counts those nodes x_i for which no boundary conditions of the above type are prescribed, $1 \le i \le m - 1$ in case two Dirichlet boundary conditions are given. The basis functions $\varphi_1, \ldots, \varphi_N$ are chosen to be the hat functions, which incorporate the discretization of the x-axis. Hence, $N = m-1$, and x_0 corresponds to $x_{\min}$, and x_m to $x_{\max}$. The functions $w_1, \ldots, w_{m-1}$ are unknown, and $w_0 = w_m = 0$.

Calculating derivatives of (5.19) and substituting into $y_\tau = y_{xx}$ leads to the Ritz–Galerkin approach

$$\int_{x_0}^{x_m} \left[\sum_{i=1}^{m-1} \dot{w}_i \varphi_i + \dot{\varphi}_\mathrm{b} \right] \varphi_j \,\mathrm{d}x = \int_{x_0}^{x_m} \left[\sum_{i=1}^{m-1} w_i \varphi_i'' + \varphi_\mathrm{b}'' \right] \varphi_j \,\mathrm{d}x$$

for $j = 1, \ldots, m - 1$. The overdot represents differentiation with respect to τ, and the prime with respect to x. Arranging the terms that involve derivatives of φ_b into vectors $a(\tau)$, $b(\tau)$,

$$a(\tau) := \begin{pmatrix} \int \varphi_\mathrm{b}''(x, \tau)\, \varphi_1(x)\,\mathrm{d}x \\ \vdots \\ \int \varphi_\mathrm{b}''(x, \tau)\, \varphi_{m-1}(x)\,\mathrm{d}x \end{pmatrix}, \quad b(\tau) := \begin{pmatrix} \int \dot{\varphi}_\mathrm{b}(x, \tau)\, \varphi_1(x)\,\mathrm{d}x \\ \vdots \\ \int \dot{\varphi}_\mathrm{b}(x, \tau)\, \varphi_{m-1}(x)\,\mathrm{d}x \end{pmatrix},$$

and using the matrices A, B as in (5.13)/(5.14), we arrive after integration by parts at

$$B\dot{w} + b = -Aw - a\,. \tag{5.20}$$

Note that for the specific φ_b from above $\varphi_b'' = 0$ and $a = 0$. For vanilla options, α and β can be drawn from (4.28), and b can be set up analytically; a and b can be considered as known. This completes the semidiscretization. Time τ is still continuous, and (5.20) defines the unknown vector function $w(\tau) := (w_1(\tau), \ldots, w_{m-1}(\tau))^{tr}$ as solution of a system of ordinary differential equations. This is a method of lines approach. The lines are defined by $x = x_i$ for $1 \le i \le m-1$, and the approximations along the lines are given by $w_i(\tau)$.

Initial conditions for $\tau = 0$ are derived from (5.19). Assume the initial condition from the payoff as $y(x, 0) = \gamma(x)$, then

$$\sum_{i=1}^{N} w_i(0)\varphi_i(x) + \varphi_b(x, 0) = \gamma(x) .$$

For vanilla payoff, γ is given by (4.5)/(4.6). Specifically for $x = x_j$ the sum reduces to $w_j(0) \cdot 1$, leading to

$$w_j(0) = \gamma(x_j) - \varphi_b(x_j, 0) .$$

To complete the discretization, time τ must be discretized. Standard software for ODEs can be applied to (5.20), in particular, codes for stiff systems. For discretizing with difference quotients consult Sect. 4.2.1. For example, apply the ODE trapezoidal rule as in (4.20) for the discretization of $\dot{w}$ in (5.20). We leave the derivation of the resulting Crank–Nicolson type discretization as an exercise to the reader. With the usual notation of the vector $w^{(\nu)}$ approximating $w(\tau_\nu)$, the result can be written

$$\begin{aligned} (B + \tfrac{\Delta\tau}{2}A)\, w^{(\nu+1)} &= (B - \tfrac{\Delta\tau}{2}A)\, w^{(\nu)} \\ &\quad - \tfrac{\Delta\tau}{2}\,(a^{(\nu)} + a^{(\nu+1)} + b^{(\nu)} + b^{(\nu+1)}) . \end{aligned} \tag{5.21}$$

The structure of (5.21) strongly resembles the finite-difference approach (4.24). This similarity suggests that the order is the same, because for the finite-element A's and B's we have (compare (5.13)/(5.14))

$$A = O\left(\frac{1}{\Delta x}\right) , \quad B = O(\Delta x) .$$

The separation of the variables x and τ in (5.19) allows to investigate the orders of the discretizations separately. In $\Delta\tau$, the order $O(\Delta\tau^2)$ of the Crank–Nicolson type approach (5.21) is clear from the ODE trapezoidal rule. It remains to derive the order of convergence with respect to the discretization in x. Because of the separation of variables it is sufficient to derive the convergence for a one-dimensional model problem. This will be done in Sect. 5.5.

5.3.2 Variational Form of the Obstacle Problem

To warm up for the discussion of the American option case, let us return to the simple obstacle problem of Sect. 4.5.5 with the obstacle function $g(x)$, or $g(x,\tau)$. This problem can be formulated as a variational inequality. The function u solving the obstacle problem can be characterized by comparing it to functions v out of a set $\mathcal{K}$ of *competing functions*

$$\mathcal{K} := \{\ v \in \mathcal{C}^0[-1,1] \mid\ v(-1) = v(1) = 0\,,$$
$$v(x) \geq g(x)\ \text{ for }\ -1 \leq x \leq 1,\ v \text{ piecewise } \in \mathcal{C}^1\}\,.$$

The requirements on u imply $u \in \mathcal{K}$. For $v \in \mathcal{K}$ we have $v - g \geq 0$ and in view of $-u'' \geq 0$ also $-u''(v-g) \geq 0$. Hence for all $v \in \mathcal{K}$ the inequality

$$\int_{-1}^{1} -u''(v-g)\,\mathrm{d}x \geq 0$$

must hold. By the LCP formulation (4.39) the integral

$$\int_{-1}^{1} -u''(u-g)\,\mathrm{d}x = 0$$

vanishes. Subtracting yields

$$\int_{-1}^{1} -u''(v-u)\,\mathrm{d}x \geq 0\ \text{ for any } v \in \mathcal{K}\,.$$

The obstacle function g does not occur explicitly in this formulation; the obstacle is implicitly defined in $\mathcal{K}$. Integration by parts leads to

$$\underbrace{[-u'(v-u)]_{-1}^{1}}_{=0} + \int_{-1}^{1} u'(v-u)'\,\mathrm{d}x \geq 0\,.$$

The integral-free term vanishes because of $u(-1) = v(-1)$, $u(1) = v(1)$. In summary, we have derived the statement:

$$\text{If } u \text{ solves the obstacle problem (4.39), then } \int_{-1}^{1} u'(v-u)'\,\mathrm{d}x \geq 0 \quad \text{for all } v \in \mathcal{K}\,. \tag{5.22}$$

Since v varies in the set $\mathcal{K}$ of competing functions, an inequality such as in (5.22) is called *variational inequality*. The characterization of u by (5.22) can be used to construct an approximation w: Instead of u, find a $w \in \mathcal{K}$ such that the

inequality (5.22) is satisfied for all $v \in \mathcal{K}$,

$$\int_{-1}^{1} w'(v-w)' \, \mathrm{d}x \geq 0 \quad \text{for all } v \in \mathcal{K} \, .$$

The characterization (5.22) is related to a minimum problem, because the integral vanishes for $v = u$.

5.3.3 *Variational Form of an American Option*

Analogously as the simple obstacle problem also the problem of calculating American options can be formulated as variational problem, compare Problem 4.7. The class of competing functions must be redefined as

$$\begin{aligned} \mathcal{K} := \{ \; & v \in \mathcal{C}^0[x_{\min}, x_{\max}] \;\mid\; \tfrac{\partial v}{\partial x} \text{ piecewise } \mathcal{C}^0 \, , \\ & v(x,\tau) \geq g(x,\tau) \;\text{ for all } x, \tau \; , \; v(x,0) = g(x,0) \, , \\ & v(x_{\max},\tau) = g(x_{\max},\tau), \; v(x_{\min},\tau) = g(x_{\min},\tau) \, \} \, . \end{aligned} \tag{5.23}$$

For the following, $v \in \mathcal{K}$ for the $\mathcal{K}$ from (5.23). Let y denote the exact solution of Problem 4.7. As solution of the partial differential inequality, y is $\mathcal{C}^2$-smooth on the continuation region, and $y \in \mathcal{K}$. From

$$v \geq g, \quad \frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} \geq 0$$

we deduce

$$\int_{x_{\min}}^{x_{\max}} \left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} \right) (v-g) \, \mathrm{d}x \geq 0 \, .$$

Invoking the complementarity

$$\int_{x_{\min}}^{x_{\max}} \left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} \right) (y-g) \, \mathrm{d}x = 0$$

and subtraction gives

$$\int_{x_{\min}}^{x_{\max}} \left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} \right) (v-y) \, \mathrm{d}x \geq 0 \, .$$

Integration by parts leads to the inequality

$$\int_{x_{\min}}^{x_{\max}} \left(\frac{\partial y}{\partial \tau}(v-y) + \frac{\partial y}{\partial x}\left(\frac{\partial v}{\partial x} - \frac{\partial y}{\partial x} \right) \right) \mathrm{d}x - \frac{\partial y}{\partial x}(v-y) \Bigg|_{x_{\min}}^{x_{\max}} \geq 0 \,.$$

The nonintegral term vanishes, because at the boundary for $x_{\min}$, $x_{\max}$, in view of $v = g$, $y = g$, the equality $v = y$ holds. The final result is

$$I(y;v) := \int_{x_{\min}}^{x_{\max}} \left(\frac{\partial y}{\partial \tau} \cdot (v-y) + \frac{\partial y}{\partial x}\left(\frac{\partial v}{\partial x} - \frac{\partial y}{\partial x} \right) \right) \mathrm{d}x \geq 0 \quad \text{for all } v \in \mathcal{K} \,. \tag{5.24}$$

The exact y is characterized by the fact that the inequality (5.24) holds for all comparison functions $v \in \mathcal{K}$. For the special choice $v = y$ the integral takes its minimal value,

$$\min_{v \in \mathcal{K}} I(y;v) = I(y;y) = 0 \,.$$

A more general question is, whether the inequality (5.24) holds for a $\widehat{y} \in \mathcal{K}$ that is not $\mathcal{C}^2$-smooth on the continuation region.[2] The aim is:

Problem 5.3 (Weak Version) *Construct a $\widehat{y} \in \mathcal{K}$ such that $I(\widehat{y}; v) \geq 0$ for all $v \in \mathcal{K}$.*

This formulation of our problem is called *weak version*, because it does *not* use $\widehat{y} \in \mathcal{C}^2$. Solutions $\widehat{y}$ of Problem 5.3, which are globally continuous but only piecewise $\in \mathcal{C}^1$, are called *weak solutions*. The original partial differential equation requires $y \in \mathcal{C}^2$ and hence more smoothness. Such $\mathcal{C}^2$-solutions are called *strong solutions* or *classical solutions* (⟶ Sect. 5.5).

5.3.4 Implementation of Finite Elements

A discretized version of the weak problem is obtained by replacing the space $\mathcal{K}$ by a finite-dimensional subspace $\widehat{\mathcal{K}}$, which is spanned by a finite number of basis functions. That is, we search for a $\widehat{y} \in \widehat{\mathcal{K}}$ such that

$$I(\widehat{y}; \widehat{v}) \geq 0 \quad \text{for all } \widehat{v} \in \widehat{\mathcal{K}} \,,$$

where $I(y;v)$ is defined in (5.24). This sets the arena for finite element methods.

[2] For the Black–Scholes $y(x,\tau)$ or $V(S,t)$ the weaker $y \in \mathcal{C}^{2,1}$ suffices. Recall that the American option is widely $\mathcal{C}^2$-smooth, except across the early-exercise curve.

As a first step to approximately solve the minimum problem, assume as in Sect. 5.3.1 separation approximations for $\widehat{y}$ and $\widehat{v}$ in the similar forms

$$\begin{aligned} \widehat{y} &= \sum_i w_i(\tau)\varphi_i(x)\,, \\ \widehat{v} &= \sum_i v_i(\tau)\varphi_i(x)\,. \end{aligned} \tag{5.25}$$

Summation is over a finite number of terms, which represents $\widehat{y},\ \widehat{v} \in \widehat{\mathcal{K}}$. The reduced smoothness of these expressions match the requirements of $\mathcal{K}$ from (5.23); time dependence is incorporated in the coefficient functions w_i and v_i. Since the basis functions φ_i represent the x_i-grid, we again perform a semidiscretization. Plugging the ansatz (5.25) into $I(\widehat{y};\widehat{v})$ from (5.24) gives

$$\begin{aligned} &\int \left\{ \left(\sum_i \frac{\mathrm{d}w_i}{\mathrm{d}\tau}\varphi_i \right) \left(\sum_j (v_j - w_j)\varphi_j \right) + \right. \\ &\qquad \left. \left(\sum_i w_i \varphi_i' \right) \left(\sum_j (v_j - w_j)\varphi_j' \right) \right\} \mathrm{d}x \\ &= \sum_i \sum_j \frac{\mathrm{d}w_i}{\mathrm{d}\tau}(v_j - w_j) \int \varphi_i \varphi_j \,\mathrm{d}x + \sum_i \sum_j w_i (v_j - w_j) \int \varphi_i' \varphi_j' \,\mathrm{d}x \geq 0\,. \end{aligned}$$

Translated into vector notation for the coefficient functions $w_i(\tau)$, $v_i(\tau)$, this is equivalent to

$$\left(\frac{\mathrm{d}w}{\mathrm{d}\tau} \right)^{\mathrm{tr}} B(v - w) + w^{\mathrm{tr}} A(v - w) \geq 0$$

or[3]

$$(v - w)^{\mathrm{tr}} \left(B \frac{\mathrm{d}w}{\mathrm{d}\tau} + Aw \right) \geq 0\,.$$

This is the (semi-)discretized weak version of $I(\widehat{y};\widehat{v}) \geq 0$. The matrices A and B are defined via the assembling described above; for equidistant steps the special versions in (5.13), (5.14) arise.

As a second step, the time τ is discretized as well. To this end let us define the vectors

$$w^{(v)} := w(\tau_v), \quad v^{(v)} := v(\tau_v)\,.$$

[3]Notation: Now v is the vector of the coefficient functions.

Upon substituting, and θ-averaging the Aw term as in Sect. 4.6.1, we arrive at the inequalities

$$\left(v^{(\nu+1)} - w^{(\nu+1)}\right)^{tr} \left(B\frac{1}{\Delta\tau}(w^{(\nu+1)} - w^{(\nu)}) + \theta A w^{(\nu+1)} + (1-\theta)Aw^{(\nu)} \right) \geq 0 \tag{5.26}$$

for all ν. For $\theta = 1/2$ this is a Crank–Nicolson-type method. Rearranging (5.26) leads to

$$\left(v^{(\nu+1)} - w^{(\nu+1)}\right)^{tr} \left((B + \Delta\tau\,\theta A)\, w^{(\nu+1)} + (\Delta\tau(1-\theta)A - B)\, w^{(\nu)} \right) \geq 0\,.$$

With the abbreviations

$$\begin{aligned} r &:= (B - \Delta\tau(1-\theta)A)\, w^{(\nu)}\,, \\ C &:= B + \Delta\tau\,\theta A\,, \end{aligned} \tag{5.27}$$

the inequality can be rewritten as

$$\left(v^{(\nu+1)} - w^{(\nu+1)}\right)^{tr} \left(Cw^{(\nu+1)} - r \right) \geq 0\,. \tag{5.28}$$

This is the fully discretized version of $I(\widehat{y}; v) \geq 0$.

5.3.4.1 Side Conditions

To match the requirements of $\mathcal{K}$, the inequalities $\widehat{y} \geq g$ and $\widehat{v} \geq g$ must hold. $\widehat{y}(x,\tau) \geq g(x,\tau)$ amounts to

$$\sum w_i(\tau)\varphi_i(x) \geq g(x,\tau)\,.$$

For hat functions φ_i (with $\varphi_i(x_i) = 1$ and $\varphi_i(x_j) = 0$ for $j \neq i$) and $x = x_j$ this implies $w_j(\tau) \geq g(x_j,\tau)$. With $\tau = \tau_\nu$ we have

$$w^{(\nu)} \geq g^{(\nu)}; \quad \text{analogously } v^{(\nu)} \geq g^{(\nu)}\,.$$

For each time level ν we must find a solution that satisfies both the inequality (5.26)–(5.28) and the side condition

$$w^{(\nu+1)} \geq g^{(\nu+1)} \text{ for all } v^{(\nu+1)} \geq g^{(\nu+1)}\,.$$

In summary, the algorithm is

Algorithm 5.4 (Finite Elements for American Standard Options)

Choose θ $(\theta = 1/2)$. *Calculate* $w^{(0)}$, *and* C *from (5.27).*

For $\nu = 1, \ldots, \nu_{\max}$:

Calculate $r = (B - \Delta\tau(1-\theta)A)w^{(\nu-1)}$ *and* $g = g^{(\nu)}$.

Construct a w *such that for all* $v \geq g$

$$(v-w)^{tr}(Cw - r) \geq 0, \quad w \geq g.$$

Set $w^{(\nu)} := w$.

This algorithm generates a discretized solution of the weak Problem 5.3: The vectors w define $\widehat{y} \in \widehat{\mathcal{K}}$ via (5.25); $\widehat{v}$ is not needed explicitly. Let us emphasize again the main step (FE), which is the kernel of this algorithm and the main labor: Construct w such that

$$\begin{aligned} &\textbf{(FE)} \quad \text{for all } v \geq g \\ &\qquad (v-w)^{tr}(Cw - r) \geq 0\,, \quad w \geq g\,. \end{aligned} \tag{5.29}$$

This task (FE) can be reformulated into a task we already solved in Sect. 4.6. To this end recall the finite-difference equation (4.44), replacing A by C, and b by r. There the following holds for w:

$$\begin{aligned} &\textbf{(FD)} \quad Cw - r \geq 0\,, \quad w \geq g\,, \\ &\qquad (Cw - r)^{tr}(w - g) = 0\,. \end{aligned} \tag{5.30}$$

Theorem 5.5 (Equivalence) *The solution of the problem (FE) is equivalent to the solution of problem (FD).*

Proof

a) (FD) $\Longrightarrow$ (FE):
Let w solve (FD), so $w \geq g$, and

$$(v-w)^{tr}(Cw - r) = (v-g)^{tr}\underbrace{(Cw - r)}_{\geq 0} - \underbrace{(w-g)^{tr}(Cw-r)}_{=0}$$

hence $(v-w)^{tr}(Cw - r) \geq 0$ for all $v \geq g$.

b) (FE) $\Longrightarrow$ (FD):
Let w solve (FE), so $w \geq g$, and

$$v^{tr}(Cw - r) \geq w^{tr}(Cw - r) \quad \text{for all } v \geq g\,.$$

Suppose the kth component of $Cw - r$ is negative, and make v_k arbitrarily large. Then the left-hand side becomes arbitrarily small, which is a contradiction. So $Cw - r \geq 0$. Now

$$w \geq g \implies (w-g)^{tr}(Cw-r) \geq 0\,.$$

Set in (FE) $v = g$, then $(w-g)^{tr}(Cw-r) \leq 0$. Therefore $(w-g)^{tr}(Cw-r) = 0$.

5.3.4.2 Implementation

As a consequence of this equivalence, the solution of the finite-element problem (FE) can be calculated with the methods we applied to solve problem (FD) in Sect. 4.6. Following the exposition in Sect. 4.6.2, the kernel of the finite-element Algorithm 5.4 can be written as follows

(FE′) Solve $Cw = r$ componentwise such that

the side condition $w \geq g$ is obeyed.

The vector v is not calculated. Boundary conditions on w are set up in the same way as discussed in Sect. 4.4 and summarized in Algorithm 4.14. Consequently, the finite-element algorithm parallels Algorithm 4.14 closely in the special case of an equidistant x-grid; there is no need to repeat this algorithm ($\longrightarrow$ Exercise 5.4). In the general nonequidistant case, the off-diagonal and the diagonal elements of the tridiagonal matrix C vary with i. Then the formulation of the SOR-loop gets more involved. The details of the implementation are technical and omitted. The Algorithm 4.15 is the same in the finite-element case.

The computational results match those of Chap. 4 and are not repeated. The costs of the presented simple version of a finite-element approach are slightly lower than that of the finite-difference approach, because we can take advantage of an optimal spacing of the mesh points x_i. For arguments discussing the closeness of $\widehat{y}$ to y, we refer to Sect. 5.5.

5.4 Two-Asset Options

In Sect. 3.5.5 we discussed an option based on two assets with prices S_1, S_2. There we applied Monte Carlo to simulate the GBM model, see Example 3.9. For the mathematical model we have chosen the Black–Scholes market. The corresponding PDE for the value function $V(S_1, S_2, t)$ is

$$\begin{aligned}\frac{\partial V}{\partial t} &+ \tfrac{1}{2}\sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} + (r-\delta_1)S_1 \frac{\partial V}{\partial S_1} - rV \\ &+ \tfrac{1}{2}\sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2} + (r-\delta_2)S_2 \frac{\partial V}{\partial S_2} + \rho\sigma_1\sigma_2 S_1 S_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} = 0\,,\end{aligned} \tag{5.31}$$

with dividend rates δ_1, δ_2. (For the general case see Sect. 6.2.) Notice that for $S_2 = 0$ the familiar one-dimensional Black–Scholes equation results. The model is completed by a payoff function $\Psi(S_1, S_2)$ and the terminal condition $V(S_1, S_2, T) = \Psi(S_1, S_2)$. The computational domain $\mathcal{D}$ is two-dimensional, $\mathcal{D} \subset \mathbb{R}^2$ (disregarding time t).

Example 5.6 (European Call on a Basket with Double Barrier) We consider a call on a two-asset basket with two knock-out barriers. The payoff of this exotic European-style option is

$$\Psi(S_1, S_2) = (S_1 + S_2 - K)^+ ,$$

up to the barriers (see Fig. 5.1). In the underlying basket the two assets are of equal weight. The two knock-out barriers are given by B_1 and B_2, down-and-out at B_1, and up-and-out at B_2. That is, the option ceases to exist when $S_1 + S_2 \le B_1$, or when $S_1 + S_2 \ge B_2$; in both cases $V = 0$. In this example, the computational domain $\mathcal{D}$ is easy to define: The value function is zero outside the barriers. Hence the domain is bounded by the two lines $S_1 + S_2 = B_1$ and $S_1 + S_2 = B_2$. This shape of $\mathcal{D}$ naturally suggests to tile the domain into a grid of triangular elements $\mathcal{D}_k$. One possible triangulation is shown in Fig. 5.5, where a structured regular subdivision is applied. For this example we choose the parameters

$$\begin{aligned} &K = 1\,,\ T = 1\,,\ \sigma_1 = \sigma_2 = 0.25\,,\ \rho = 0.7\,,\ r = 0.05\,, \\ &\delta_1 = \delta_2 = 0\,,\ \ B_1 = 1\,,\ B_2 = 2\,. \end{aligned}$$

The values V for $S_1 \to 0$ and $S_2 \to 0$ are known by the one-dimensional Black–Scholes equation; just set either $S_1 = 0$ or $S_2 = 0$ in (5.31). These values of single-asset double-barrier options for $B_1 \le S \le B_2$ can be evaluated by a closed-form formula, see [172]. We shall come back to this example below.

5.4.1 Analytical Preparations

It is convenient to solve the Black–Scholes equation in divergence form. To this end, use standard PDE variables $x := S_1$, $y := S_2$ for the independent variables, and $u(x, y, t)$ for the dependent variable, and derive the vector PDE for u

$$-\nabla \cdot (D(x, y)\nabla u) + b(x, y)^{tr} \nabla u + ru = u_t\,. \tag{5.32}$$

This makes use of the formal "nabla" vector $\nabla := (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^{tr}$, and

$$\begin{aligned} D(x, y) &:= \tfrac{1}{2} \begin{pmatrix} \sigma_1^2 x^2 & \rho\sigma_1\sigma_2 xy \\ \rho\sigma_1\sigma_2 xy & \sigma_2^2 y^2 \end{pmatrix}, \\ b(x, y) &:= -\begin{pmatrix} (r - \delta_1 - \sigma_1^2 - \rho\sigma_1\sigma_2/2)\, x \\ (r - \delta_2 - \sigma_2^2 - \rho\sigma_1\sigma_2/2)\, y \end{pmatrix}. \end{aligned} \tag{5.33}$$

∇u is the gradient of u, and the dot-product notation

$$\nabla \cdot U = \frac{\partial U_1}{\partial x} + \frac{\partial U_2}{\partial y}$$

for a vector function U denotes the divergence; the $\cdot$ corresponds to the scalar product, similar as ^{tr} for vectors. The reader is invited to check the equivalence with (5.31) ($\longrightarrow$ Exercise 5.5). The advantage of version (5.32) over (5.31) lies in a simple treatment of the second-order derivatives; they can be removed, and a weak version can be derived. This will become apparent below.

5.4.2 Weighted Residuals

The partial differential equation (5.32) can be represented by $R(u, x, y, t) = 0$, where

$$\begin{aligned} R(u, x, y, t) := & - \nabla \cdot (D(x, y) \nabla u(x, y, t)) + b(x, y)^{tr} \nabla u(x, y, t) \\ & + r u(x, y, t) - \frac{\partial u(x, y, t)}{\partial t} \end{aligned}$$

denotes the residual. As in Sect. 5.1, the residual is used to set up an integral equation. To this end, introduce weighting functions v, multiply the residual of the PDE with $v(x, y, t)$ and request

$$\int_{\mathcal{D}} R(u, x, y, t)\, v \, \mathrm{d}x \, \mathrm{d}y = 0 \,. \tag{5.34}$$

This integral over the computational domain $\mathcal{D} \subset \mathbb{R}^2$ is a double integral. It depends on t, and should vanish for all $0 \leq t \leq T$ and arbitrary v. We consider u to be a solution in case (5.34) holds for "all" v. This is a weak version of the PDE and requires less regularity of its "weak" solutions u. Aspects of accuracy are postponed to Sect. 5.5.

To exploit the potential of the integral version (5.34), we transform the second-order derivatives to first order, comparable to integration by parts. The leading integral over the second-order term is

$$\int_{\mathcal{D}} -\nabla \cdot (D \nabla u)\, v \, \mathrm{d}x \, \mathrm{d}y \,.$$

The reader may check for the vector $U := v D \nabla u$ the formula for the divergence $\nabla \cdot U$, namely,

$$\nabla \cdot (v D \nabla u) = (\nabla v)^{tr} D \nabla u + v \nabla \cdot D \nabla u \,,$$

and hence

$$-\int_{\mathcal{D}} v\,\nabla\cdot(D\nabla u)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathcal{D}} (\nabla v)^{tr} D\nabla u\,\mathrm{d}x\,\mathrm{d}y - \int_{\mathcal{D}} \nabla\cdot(vD\nabla u)\,\mathrm{d}x\,\mathrm{d}y\,.$$

Next we quote the divergence theorem, here for the two-dimensional situation:

$$\int_{\mathcal{D}} \nabla\cdot U\,\mathrm{d}x\,\mathrm{d}y = \int_{\partial\mathcal{D}} U^{tr} n\,\mathrm{d}s\,, \tag{5.35}$$

where $\partial\mathcal{D}$ denotes the boundary of $\mathcal{D}$, and n is the outward unit normal vector on $\partial\mathcal{D}$. (n is perpendicular to the curve $\partial\mathcal{D}$ and points away from $\mathcal{D}$.) The parameter s measures the arclength along the boundary $\partial\mathcal{D}$.[4] We apply the divergence theorem to the specific vector $U := vD\nabla u$, and arrive at the result for the second-order term

$$-\int_{\mathcal{D}} v\,\nabla\cdot(D\nabla u)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathcal{D}} (\nabla v)^{tr} D\nabla u\,\mathrm{d}x\,\mathrm{d}y - \int_{\partial\mathcal{D}} (vD\nabla u)^{tr} n\,\mathrm{d}s\,.$$

In (5.32)/(5.33) the matrix D is symmetric, $D = D^{tr}$. For symmetric D the integrand in the boundary integral is $v(\nabla u)^{tr} Dn$. After the above transformations of the leading integral, we rewrite (5.34) into

$$\int_{\mathcal{D}} \left[(\nabla v)^{tr} D\nabla u + vb^{tr}\nabla u + ruv - \frac{\partial u}{\partial t} v \right] \mathrm{d}x\,\mathrm{d}y - \int_{\partial\mathcal{D}} v(\nabla u)^{tr} Dn\,\mathrm{d}s = 0\,. \tag{5.36}$$

Recall that both u and v as well as ∇u and ∇v depend on x, y, t, and the integrals on t. This is the weak version of the PDE (5.32).

Next discretize the time $0 \le t \le T$ as in Chap. 4, say, with equidistant steps Δt. For the simplest implicit approach, the derivative with respect to time t is resolved by the first-order difference quotient,

$$\frac{\partial u(x,y,t)}{\partial t} \approx \frac{u(x,y,t+\Delta t) - u(x,y,t)}{\Delta t}\,.$$

For backward running time t,

$$u_{\mathrm{pre}} := u(x,y,t+\Delta t)$$

is known at time t from the calculation of the previous time level. The analogue of the fully implicit time-stepping method is then to solve (5.36) at time level t for $\frac{\partial u}{\partial t}$

[4] Recall from calculus the definition $\int_C f(x,y)\mathrm{d}s = \int_a^b f(g(\xi),h(\xi))\frac{\mathrm{d}s}{\mathrm{d}\xi}\mathrm{d}\xi$ where $(g(\xi),h(\xi))$ for $a \le \xi \le b$ is a parameterization of a planar curve C; ξ is the curve parameter. The value of this *line integral* is independent of the orientation of the curve C and independent of the particular parameterization.

replaced by

$$\frac{1}{\Delta t}(u_{\text{pre}} - u)\,,$$

starting at $t = T - \Delta t$ with the payoff, $u_{\text{pre}} = \Psi$. With this approximation, the function u in (5.36) approximates the value function V at time level t. Alternatively, a second-order time-discretization can be applied, similar as in Sect. 4.3. For the required regularity of the functions u and v, consult Sect. 5.5.

5.4.3 Boundary

Boundary conditions enter via the boundary integral around the boundary $\partial\mathcal{D}$. In practice, the computational domain $\mathcal{D}$ is defined by specifying $\partial\mathcal{D}$. To this end, express the curve $\partial\mathcal{D}$ as the union of a finite number of non-overlapping piecewise smooth boundary curves $\partial\mathcal{D}_1, \partial\mathcal{D}_2, \ldots$. Each of these curves must be parameterized as in

$$\partial\mathcal{D}_1 := \{\,(g_1(\xi), h_1(\xi)) \mid a_1 \le \xi \le b_1\,\}\,.$$

In this way, an orientation is given by starting the curve at the parameter value $\xi = a_1$ and ending at $\xi = b_1$. By specifying parameter intervals as $a_1 \le \xi \le b_1$ and parametric functions as g_1, h_1, the entire boundary is defined. The convention is that the orientation is done such that the domain $\mathcal{D}$ is *on the left-hand side*, as we run through the parameterizations for increasing parameter values ξ.

Now the curve $\partial\mathcal{D}$ is defined and we address the boundary integral along that curve. It is split into a sum of integrals according to the piecewise smooth curves $\partial\mathcal{D}_1, \partial\mathcal{D}_2, \ldots$. For example, the boundary of the domain in Fig. 5.5 consists of four such parts (⟶ Exercise 5.6).

The product-type integrand $f(x, y) := v(\nabla u)^{tr} Dn$ suggests to place emphasis on two specific kinds of boundary condition, namely,

- v is prescribed (Dirichlet boundary conditions),
- $(\nabla u)^{tr} Dn$ is prescribed (Neumann boundary conditions).

The boundary differential operator $(\nabla u)^{tr} Dn = n^{tr} D \nabla u$ can be considered as a generalized directional derivative since $\frac{\partial u}{\partial n} = n^{tr} \nabla u$. Mixed boundary conditions are possible as well. If we cast the components of the vector $n^{tr} D$ into a vector (α_1, α_2), then all type of boundary conditions can be written in the form

$$\alpha_1(x, y)\frac{\partial u}{\partial x} + \alpha_2(x, y)\frac{\partial u}{\partial y} = \alpha_0(x, y)\, u + \beta(x, y)$$

with proper functions α_0 and β. Then

$$v\,(\alpha_0(x,y)\,u+\beta(x,y))$$

is substituted into the boundary integral, which is approximated numerically using the edges of the triangulation of $\mathcal{D}$.

Fortunately, boundary conditions are frequently of simple form. In particular one encounters the two types

- $u=0$ (or $v=0$), which is of Dirichlet type with $\alpha_1=\alpha_2=\beta=0$ and $\alpha_0\neq 0$.
- $(\nabla u)^{tr}Dn=0$, which is of Neumann type with $\alpha_0=\beta=0$ and nonzero vector (α_1,α_2).

The boundary $\partial\mathcal{D}$ may consist, for example, of two parts $\partial\mathcal{D}_{\mathrm{D}}$ and $\partial\mathcal{D}_{\mathrm{N}}$ with $\partial\mathcal{D}=\partial\mathcal{D}_{\mathrm{D}}\cup\partial\mathcal{D}_{\mathrm{N}}$, $\partial\mathcal{D}_{\mathrm{D}}\cap\partial\mathcal{D}_{\mathrm{N}}=\emptyset$, and Dirichlet conditions on $\partial\mathcal{D}_{\mathrm{D}}$ and Neumann conditions on $\partial\mathcal{D}_{\mathrm{N}}$. Clearly, boundary integrals vanish for the special cases $v=0$ or $(\nabla u)^{tr}Dn=0$. Neumann conditions are advantageous in that they need not be specified for weak formulations. This entails an advantage of FEM over discretizing the PDEs by finite differences. In the latter case, *all* boundary conditions must be implemented. For FEM it suffices to implement Dirichlet conditions. Defining the right boundary conditions can be demanding. Aside to be financially meaningful, another aim is the problem to be well-posed—that is, it defines a unique solution. To some extent, defining proper boundary conditions is an art.

Example 5.7 (European Binary Put as in Example 3.9) In Chap. 3 the Example 3.9 of a binary put was simulated with Monte Carlo, and no boundary or boundary conditions were needed. Here we prepare the example to be solved by FEM. Again, $x:=S_1$, $y:=S_2$. As in Chap. 4, the domain $0<x<\infty$, $0<y<\infty$ must be truncated to finite size. A simple choice of a computational domain is a rectangle

$$\mathcal{D}=\{\,(x,y)\mid\ 0\le x\le x_{\max},\ 0\le y\le y_{\max}\,\}$$

with $x_{\max},y_{\max}$ large enough such that zero boundary conditions $u=0$ can be chosen as approximation for $x=x_{\max}$ or $y=y_{\max}$. The rectangle is bounded by four straight lines, which can be parameterized, for example, by

$$\begin{aligned}
\partial\mathcal{D}_1&:=\{\,x=\xi,\ y=0\ \mid\quad 0\le\xi\le x_{\max}\,\},\\
\partial\mathcal{D}_2&:=\{\,x=x_{\max},\ y=\xi\ \mid\quad 0\le\xi\le y_{\max}\,\},\\
\partial\mathcal{D}_3&:=\{\,x=x_{\max}-\xi,\ y=y_{\max}\ \mid\quad 0\le\xi\le x_{\max}\,\},\\
\partial\mathcal{D}_4&:=\{\,x=0,\ y=y_{\max}-\xi\ \mid\quad 0\le\xi\le y_{\max}\,\}.
\end{aligned}$$

Now $\partial\mathcal{D}=\partial\mathcal{D}_1\cup\partial\mathcal{D}_2\cup\partial\mathcal{D}_3\cup\partial\mathcal{D}_4$, and the parameterized curve has the domain on the left.

Dirichlet conditions are imposed for $\partial\mathcal{D}_2$ and $\partial\mathcal{D}_3$, where we have chosen to approximate boundary values by requesting $u = 0$. For $y = 0$ the boundary conditions can be chosen as the values of the one-dimensional European binary put. An analytic formula for the one-dimensional case of a European binary put is

$$V_{\text{binP}}^{\text{Eur}}(S,t) := c\,\mathrm{e}^{-r(T-t)}\,F\left(-\frac{\log(S/K)+(r-\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right),$$

for a face value c, with standard normal distribution F [172]. For $y = 0$ we set $S = x$. The same formula can be applied for the boundary with $x = 0$; then $S = y$. In this way, on $\partial\mathcal{D}_1$ and $\partial\mathcal{D}_4$ the boundary conditions are of Dirichlet type with $u = V_{\text{binP}}^{\text{Eur}}$. With this choice of boundary conditions, $\partial\mathcal{D}_{\text{D}} = \partial\mathcal{D}$ and $\partial\mathcal{D}_{\text{N}} = \emptyset$. But there is a simpler choice: As [300] points out, this Dirichlet condition is implicitly defined by the PDE, because the one-dimensional PDE is embedded in (5.31) for $S_1 = 0$ or $S_2 = 0$. So no boundary condition needs to be specified along $\partial\mathcal{D}_1$ and $\partial\mathcal{D}_4$. This amounts to zero Neumann conditions. Both the Dirichlet version and the Neumann version work. The latter has the advantage of avoiding the effort of evaluating $V_{\text{binP}}^{\text{Eur}}$.

The implementation of the weak form in (5.36) is straightforward when, for example, the package `FreeFem++` is applied. Thereby a figure similar as Fig. 3.7 is produced easily.

5.4.4 Involved Matrices

The accuracy of FEM depends on how the grid is chosen. Algorithms for mesh generation and mesh adaption are needed, but these are demanding topics. It is cumbersome to implement a two-dimensional FEM yourself. For first results, one may work with a fixed structured grid. But in general it is advisable and comfortable to apply a FEM package to solve (5.36). Here we merely focus on how the two-dimensional analogue of the hat functions enters.

For the Ritz–Galerkin approach we apply the basis representation

$$w(x,y,t) = \sum_i w_i(t)\,\varphi_i(x,y) \tag{5.37}$$

as approximation for u, and set $v = \varphi_j$. This ansatz separates time τ and "space" (x,y). The functions φ_i are defined on $\mathcal{D}$.

For basis functions, we choose the two-dimensional hat functions, which perfectly match triangular elements. The situation is shown schematically in Fig. 5.9. There the central node l is node of several adjacent triangles, which constitute the support (shaded) on which φ_l is built by planar pieces. This approach defines a tent-like hat function φ_l, which is zero "outside." By linear combination of such basis functions, piecewise planar surfaces above the computational domain are constructed. Locally, for one triangle, this may look like the element in Fig. 5.4.

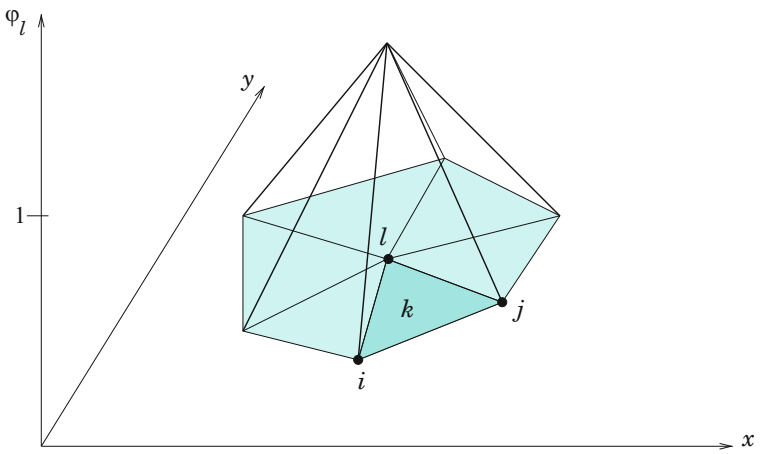

Fig. 5.9 Two-dimensional hat function $\varphi_l(x, y)$ (zero outside the *shaded area*)

Notice that $\nabla w = \sum w_i \nabla \varphi_i$. The weak form of (5.36) leads to

$$\int_{\mathcal{D}} (\nabla \varphi_j)^{tr} D \sum w_i \nabla \varphi_i + \varphi_j \left[b^{tr} (\sum w_i \nabla \varphi_i) + r \sum w_i \varphi_i - \sum \frac{\partial w_i}{\partial t} \varphi_i \right] \mathrm{d}x\,\mathrm{d}y - \int_{\partial \mathcal{D}} \varphi_j (\sum w_i \nabla \varphi_i)^{tr} D n \,\mathrm{d}s = 0\,,$$

for all j. This is a system of ODEs

$$\sum_i w_i \int_{\mathcal{D}} \left[(\nabla \varphi_j)^{tr} D \nabla \varphi_i + \varphi_j b^{tr} \nabla \varphi_i + \varphi_j r \varphi_i \right] \mathrm{d}x\,\mathrm{d}y - \sum_i \frac{\partial w_i}{\partial t} \int_{\mathcal{D}} \varphi_i \varphi_j \,\mathrm{d}x\,\mathrm{d}y - \sum_i w_i \int_{\partial \mathcal{D}} \varphi_j (\nabla \varphi_i)^{tr} D n \,\mathrm{d}s = 0\,. \tag{5.38}$$

As an exercise, the reader should rewrite this ODE system in matrix-vector notation. In summary, FEM needs the integrals over the domain $\mathcal{D}$

$$\int (\nabla \varphi_j)^{tr} D \nabla \varphi_i \quad (\text{“diffusion terms”})\,,$$

$$\int \varphi_j b^{tr} \nabla \varphi_i \quad (\text{“convection terms”})\,,$$

$$\int \gamma \varphi_j \varphi_i \quad (\text{“reaction terms”})\,,$$

where γ is chosen appropriately, and in addition boundary integrals along $\partial \mathcal{D}$.

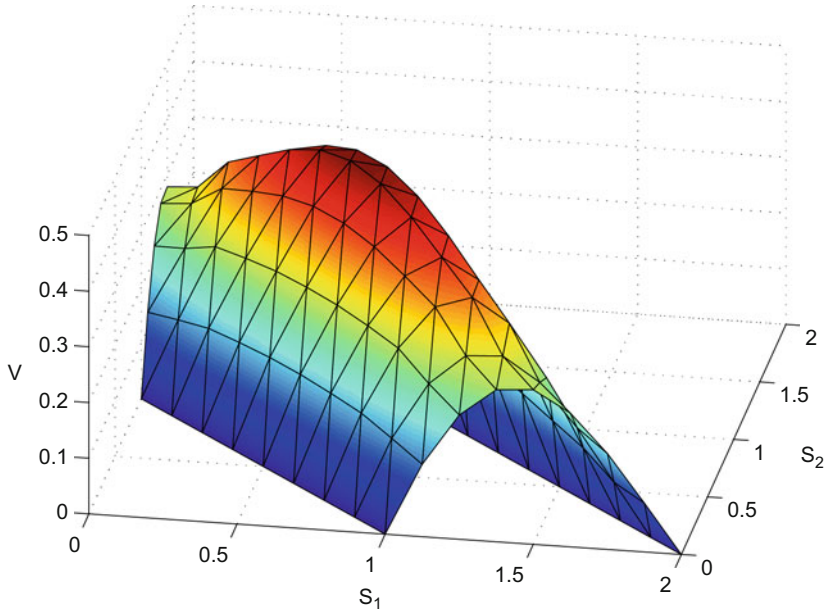

Fig. 5.10 Rough approximation of the value function $V(S_1, S_2, 0)$ of a basket double-barrier call option, Example 5.6. With kind permission of Anna Kvetnaia

For each number k of a triangle, there are three vertices of the triangle, with node numbers i, j, l in Fig. 5.9. Hence the table $\mathcal{I}$ of index sets that assigns nodes to triangles includes the entry

$$\mathcal{I}_k := \{i, j, l\}\,.$$

Only for the three node numbers $i, j, l \in \mathcal{I}_k$ the local integrals on $\mathcal{D}_k$ are nonzero. They can be arranged into 3×3 element matrices. For the derivation of the integrals, it makes sense to use a local numbering $1_k, 2_k, 3_k$ for the nodes of $\mathcal{D}_k$. For each global matrix, the assembling loop over k distributes up to 27 local integrals calculated on $\mathcal{D}_k$, nine integrals of each of the above three types.[5]

Back to Example 5.6, we solve (5.36) with FEM. Figure 5.10 shows a FEM solution with 192 triangles. Figure 5.11 illustrates a mesh structure for higher resolution obtained with `FreeFem++`. In the two-dimensional case, because of higher costs, we typically confine ourselves to an accuracy lower than in the one-dimensional situation. Based on our results we state

$$V(1.25,\, 0.25,\, 0) \approx 0.2949\,.$$

[5]Basic ingredients for the calculation of the local integrals on an arbitrary triangle $\mathcal{D}_k$ are the relations in Exercise 5.7. See also Exercises 5.8 and 5.9.

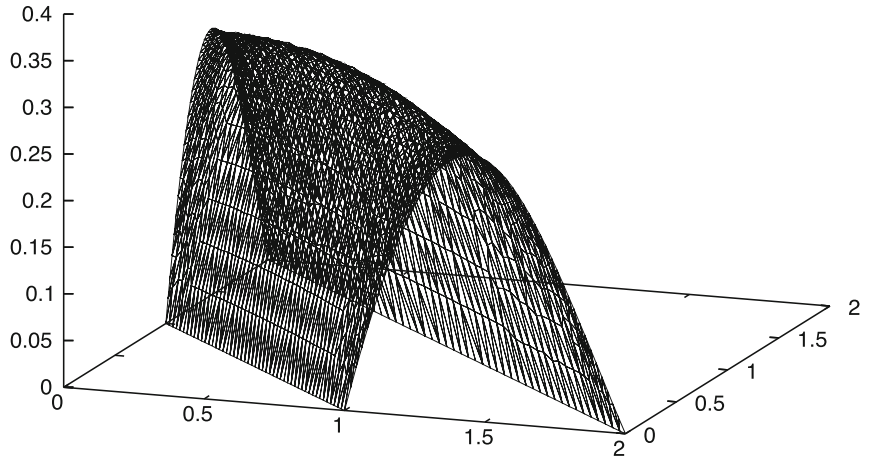

Fig. 5.11 Finer approximation of the value function $V(S_1, S_2, 0)$ of a basket double-barrier call option, Example 5.6

Example 5.8 (Heston's PDE) In Example 1.16 Heston's model was introduced, where v denotes a stochastic volatility. The corresponding PDE from [178] is

$$\begin{aligned} &\frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2} \sigma_v^2 v \frac{\partial^2 V}{\partial v^2} + \rho \sigma_v v S \frac{\partial^2 V}{\partial S \partial v} \\ &\quad + rS \frac{\partial V}{\partial S} + [\kappa(\theta - v) - \lambda v] \frac{\partial V}{\partial v} - rV = 0\,, \end{aligned} \tag{5.39}$$

with parameters as in (1.59), and λ standing for the market price of volatility risk. Here we are interested in solutions $V(S, v, t)$ on part of a two-dimensional (S, v)-plane. The PDE (5.39) can be cast into version (5.32). As an exercise, the reader is encouraged to derive D and b, and with the payoff of a call and an own choice of parameters, to think about suitable boundary conditions, and to do experiments with (5.39). Note that for a call a reasonable requirement for maximum values of the volatility v is $V = S$. When in addition the interest rate r is replaced by a stochastic variable, the PDE is based on a three-dimensional domain [163].

5.5 Error Estimates

The similarity of the finite-element equation (5.21) with the finite-difference equation (4.24) suggests that the errors may be of the same order. In fact, numerical experiments confirm that the finite-element approach with the linear basis functions from Definition 5.1 produces errors decaying quadratically with the mesh size. Applying the finite-element Algorithm 5.4 and entering the calculated data into a

diagram as Fig. 4.14, confirms the quadratic order experimentally. The proof of this order of the error is more difficult for finite-element methods because weak solutions assume less smoothness. For standard options, the separation of variables in (5.19) also separates the discussion of the order, and an analysis of the one-dimensional situation suffices. This section explains some basic ideas of how to derive error estimates. We begin with reconsidering some of the related topics that have been introduced in previous sections.

5.5.1 Strong and Weak Solutions

Our exposition will be based on the model problem (5.15). That is, the simple second-order differential equation

$$-u'' = f(x) \quad \text{for } \alpha < x < \beta \tag{5.40}$$

with given f, and homogeneous Dirichlet-boundary conditions

$$u(\alpha) = u(\beta) = 0 \tag{5.41}$$

will serve as illustration. The differential equation is of the form $Lu = f$, compare (5.2). The domain $\mathcal{D} \subseteq \mathbb{R}^n$ on which functions u are defined specializes for $n = 1$ to the open and bounded interval $\mathcal{D} = \{x \in \mathbb{R}^1 \mid \alpha < x < \beta\}$. For continuous f, solutions of the differential equation (5.40) satisfy $u \in \mathcal{C}^2(\mathcal{D})$. In order to have operative boundary conditions, solutions u must be continuous on $\mathcal{D}$ including its boundary, which is denoted $\partial\mathcal{D}$. Therefore we require $u \in \mathcal{C}^0(\overline{\mathcal{D}})$ where $\overline{\mathcal{D}} := \mathcal{D} \cup \partial\mathcal{D}$. In summary, classical solutions of second-order differential equations require

$$u \in \mathcal{C}^2(\mathcal{D}) \cap \mathcal{C}^0(\overline{\mathcal{D}})\,. \tag{5.42}$$

The function space $\mathcal{C}^2(\mathcal{D}) \cap \mathcal{C}^0(\overline{\mathcal{D}})$ must be reduced further to comply with the boundary conditions.

For weak solutions the function space is larger ($\longrightarrow$ Appendix C.3). For functions u and v we define the inner product

$$(u, v) := \int_{\mathcal{D}} uv \,\mathrm{d}x\,. \tag{5.43}$$

Strong solutions u of $Lu = f$ satisfy also

$$(Lu, v) = (f, v) \quad \text{for all } v\,. \tag{5.44}$$

Specifically for the model problem (5.40)/(5.41) integration by parts leads to

$$(Lu, v) = -\int_\alpha^\beta u'' v \, \mathrm{d}x = -u'v\Big|_\alpha^\beta + \int_\alpha^\beta u'v' \, \mathrm{d}x \,.$$

The nonintegral term on the right-hand side of the equation vanishes in case also v satisfies the homogeneous boundary conditions (5.41). The remaining integral is a **bilinear form**, which we abbreviate

$$b(u, v) := \int_\alpha^\beta u'v' \, \mathrm{d}x \,. \tag{5.45}$$

Bilinear forms as $b(u, v)$ from (5.45) are linear in each of the two arguments u and v. For example, $b(u_1 + u_2, v) = b(u_1, v) + b(u_2, v)$ holds. The bilinear form (5.45) is symmetric, $b(u, v) = b(v, u)$. For several classes of more general differential equations analogous bilinear forms are obtained. Formally, (5.44) can be rewritten as

$$b(u, v) = (f, v) \,, \tag{5.46}$$

where we assume that v satisfies the homogeneous boundary conditions (5.41).

The Eq. (5.46) has been derived out of the differential equation, for the solutions of which we have assumed smoothness in the sense of (5.42). Many "solutions" of practical importance do not satisfy (5.42) and, accordingly, are not smooth. In several applications, u or derivatives of u have discontinuities. For instance consider the obstacle problem of Sect. 4.5.5: The second derivative u'' of the solution fails to be continuous at α and β. Therefore $u \notin \mathcal{C}^2(-1, 1)$ no matter how smooth the data function is, compare Fig. 4.10. As mentioned earlier, integral relations require less smoothness.

In the derivation of (5.46) the integral version has resulted as a consequence of the primary differential equation. This is contrary to wide areas of applied mathematics, where an integral relation is based on first principles, and the differential equation is derived in a second step. For example, in the calculus of variations a minimization problem may be described by an integral performance measure, and the differential equation is a necessary criterion [350]. This situation suggests considering the integral relation as an equation of its own right rather than as offspring of a differential equation. This leads to the question, *what is the maximal function space* such that (5.46) with (5.43), (5.45) is meaningful? That means to ask, for which functions u and v do the integrals exist? For a more detailed background we refer to Appendix C.3. For the introductory exposition of this section it may suffice to sketch the maximal function space briefly. The suitable function space is denoted $\mathcal{H}^1$, the version equipped with the boundary conditions is denoted $\mathcal{H}_0^1$. This *Sobolev space* consists of those functions that are continuous on $\mathcal{D}$ and that are *piecewise differentiable* and satisfy the boundary conditions (5.41). This function space corresponds to the class of functions $\mathcal{K}$ in (5.23). By means of the Sobolev space $\mathcal{H}_0^1$ a weak solution of $Lu = f$ is defined, where L is a second-order differential operator and b the corresponding bilinear form.

Definition 5.9 (Weak Solution) $u \in \mathcal{H}_0^1$ is called weak solution [of $Lu = f$], if $b(u, v) = (f, v)$ holds for all $v \in \mathcal{H}_0^1$.
This definition implicitly expresses the task: find a $u \in \mathcal{H}_0^1$ *such that* $b(u, v) = (f, v)$ *for all* $v \in \mathcal{H}_0^1$. This problem is called *variational problem*. The model problem (5.40)/(5.41) serves as example for $Lu = f$; the corresponding bilinear form $b(u, v)$ is defined in (5.45) and (f, v) in (5.43). For the integrals (5.43) to exist, we in addition require f to be square integrable ($f \in \mathcal{L}^2$, compare Appendix C.3). Then (f, v) exists because of the Schwarzian inequality (C.16). In a similar way, weak solutions are introduced for more general problems; the formulation of Definition 5.9 applies.

5.5.2 *Approximation on Finite-Dimensional Subspaces*

For a practical computation of a weak solution the infinite-dimensional space $\mathcal{H}_0^1$ is replaced by a finite-dimensional subspace. Such finite-dimensional subspaces are spanned by basis functions φ_i. Simple examples are the hat functions of Sect. 5.2. Reminding of the important role splines play as basis functions, the finite-dimensional subspaces are denoted $\mathcal{S}$, and are called *finite-element spaces*. As stated in Property 5.2(a), the hat functions $\varphi_0, \ldots, \varphi_m$ span the space of polygons. Recall that each such polygon v can be represented as linear combination

$$v = \sum_{i=0}^{m} c_i \varphi_i \,.$$

The coefficients c_i are uniquely determined by the values of v at the nodes, $c_i = v(x_i)$. We call hat functions "linear elements" because they consist of piecewise straight lines. Apart from linear elements, for example, also quadratic or cubic elements are used, which are piecewise polynomials of second or third degree [79, 335, 382]. The attainable accuracy is different for basis functions consisting of higher-degree polynomials.

Since by definition the functions of the Sobolev space $\mathcal{H}_0^1$ fulfill the homogeneous boundary conditions, each subspace does so as well. Again the subscript $_0$ indicates the realization of the homogeneous boundary conditions (5.41).[6] A finite-dimensional subspace of $\mathcal{H}_0^1$ is defined by

$$\mathcal{S}_0 := \left\{ v = \sum_{i=0}^{m} c_i \varphi_i \;\mid\; \varphi_i \in \mathcal{H}_0^1 \right\} \,. \tag{5.47}$$

[6] In this subsection the meaning of the index $_0$ is twofold: It is the index of the "first" hat function, and serves as symbol of the homogeneous boundary conditions (5.41).

Properties of $\mathcal{S}_0$ are determined by the basis functions φ_i. As mentioned earlier, basis functions with small supports give rise to sparse matrices. The partition (5.4) of $\mathcal{D}$ is implicitly included in the definition $\mathcal{S}_0$ because this information is contained in the definition of the φ_i. For our purposes the hat functions suffice. The larger m is, the better $\mathcal{S}_0$ approximates the space $\mathcal{H}_0^1$, since a finer discretization (smaller $\mathcal{D}_k$) allows to approximate the functions from $\mathcal{H}_0^1$ better by polygons. We denote the largest diameter of the $\mathcal{D}_k$ by h, and ask for convergence. That is, we study the behavior of the error for $h \to 0$ (basically $m \to \infty$).

In analogy to the variational problem expressed in connection with Definition 5.9, a *discrete* weak solution w is defined by replacing the space $\mathcal{H}_0^1$ by a finite-dimensional subspace $\mathcal{S}_0$:

Problem 5.10 (Discrete Weak Solution) *Find a $w \in \mathcal{S}_0$ such that $b(w, v) = (f, v)$ for all $v \in \mathcal{S}_0$.*

The quality of the approximation relies on the discretization fineness h of $\mathcal{S}_0$, which is occasionally emphasized by writing w_h.

5.5.3 *Quadratic Convergence*

Having defined a weak solution u and a discrete approximation w, we turn to the error $u - w$. To measure the distance between functions in $\mathcal{H}_0^1$ we use the norm $\| \ \|_1$ (⟶ Appendix C.3). That is, our first aim is to construct a bound on $\|u - w\|_1$. Let us suppose that the bilinear form is continuous and $\mathcal{H}^1$-*elliptic*:

Assumptions 5.11 (Continuous $\mathcal{H}^1$-Elliptic Bilinear Form)

(a) There is a $\gamma_1 > 0$ such that $|b(u, v)| \le \gamma_1 \|u\|_1 \|v\|_1$ for all $u, v \in \mathcal{H}^1$.
(b) There is a $\gamma_2 > 0$ such that $b(v, v) \ge \gamma_2 \|v\|_1^2$ for all $v \in \mathcal{H}^1$.

The assumption (a) is the continuity, and the property in (b) is called $\mathcal{H}^1$-ellipticity. Under the Assumptions 5.11, the problem to find a weak solution following Definition 5.9, possesses exactly one solution $u \in \mathcal{H}_0^1$; the same holds true for Problem 5.10. This is guaranteed by the Theorem of Lax–Milgram [53, 79]. In view of $\mathcal{S}_0 \subseteq \mathcal{H}_0^1$,

$$b(u, v) = (f, v) \quad \text{for all } v \in \mathcal{S}_0 .$$

Subtracting $b(w, v) = (f, v)$ and invoking the bilinearity implies

$$b(w - u, v) = 0 \quad \text{for all } v \in \mathcal{S}_0 . \tag{5.48}$$

The property of (5.48) is called *error-projection property*. The Assumptions 5.11 and the error projection are the basic ingredients to obtain a bound on the error $\|u - w\|_1$:

Lemma 5.12 (Céa) *Suppose the Assumptions 5.11 are satisfied. Then*

$$\|u - w\|_1 \leq \frac{\gamma_1}{\gamma_2} \inf_{v \in \mathcal{S}_0} \|u - v\|_1 \,. \tag{5.49}$$

Proof $v \in \mathcal{S}_0$ implies $\tilde{v} := w - v \in \mathcal{S}_0$. Applying (5.48) for $\tilde{v}$ yields

$$b(w - u, w - v) = 0 \quad \text{for all } v \in \mathcal{S}_0 \,.$$

Therefore

$$\begin{aligned} b(w - u, w - u) &= b(w - u, w - u) - b(w - u, w - v) \\ &= b(w - u, v - u) \,. \end{aligned}$$

Applying the assumptions shows

$$\begin{aligned} \gamma_2 \|w - u\|_1^2 \leq |b(w - u, w - u)| &= |b(w - u, v - u)| \\ &\leq \gamma_1 \|w - u\|_1 \|v - u\|_1 \,, \end{aligned}$$

from which

$$\|w - u\|_1 \leq \frac{\gamma_1}{\gamma_2} \|v - u\|_1$$

follows. Since this holds for all $v \in \mathcal{S}_0$, the assertion of the lemma is proven.

Let us check whether the Assumptions 5.11 are fulfilled by the model problem (5.40)/(5.41). For (a) this follows from the Schwarzian inequality (C.16) with the norms

$$\|u\|_1 = \left(\int_\alpha^\beta (u^2 + u'^2) \, \mathrm{d}x \right)^{1/2} , \quad \|u\|_0 = \left(\int_\alpha^\beta u^2 \, \mathrm{d}x \right)^{1/2} ,$$

because

$$\left(\int_\alpha^\beta u' v' \, \mathrm{d}x \right)^2 \leq \left(\int_\alpha^\beta u'^2 \, \mathrm{d}x \right) \left(\int_\alpha^\beta v'^2 \, \mathrm{d}x \right) \leq \|u\|_1^2 \, \|v\|_1^2 \,.$$

The Assumption 5.11(b) can be derived from the inequality of the Poincaré-type

$$\int_\alpha^\beta v^2 \, \mathrm{d}x \leq (\beta - \alpha)^2 \int_\alpha^\beta v'^2 \, \mathrm{d}x \,,$$

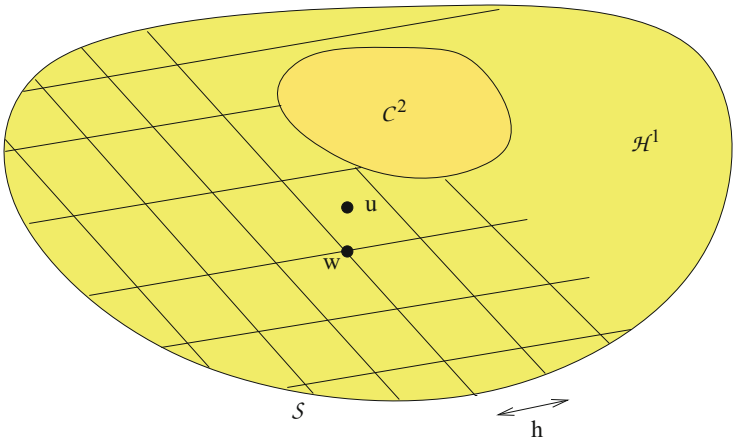

Fig. 5.12 Approximation spaces

which in turn is proven with the Schwarzian inequality (⟶ Exercise 5.10). Adding $\int v v'^2 \, dx$ on both sides leads to

$$\|v\|_1^2 \leq [(\beta - \alpha)^2 + 1] \, b(v, v) \, ,$$

from which the constant γ_2 of Assumption 5.11(b) results. Hence Céa's lemma applies to the model problem.

The next question is, how small the infimum in (5.49) may be. This is equivalent to the question, how close the subspace $\mathcal{S}_0$ can approximate the space $\mathcal{H}_0^1$ (⟶ Fig. 5.12). We will show that for hat functions and $\mathcal{S}_0$ from (5.47) the infimum is of the order $O(h)$. Again h denotes the maximum mesh size, and the notation w_h reminds us that the discrete solution depends on the grid with a spacing symbolized by h. To apply Céa's lemma, we need an upper bound for the infimum of $\|u - v\|_1$. Such a bound is found easily by a specific choice of v, which is taken as an arbitrary interpolating polygon u_I. Then by (5.49)

$$\|u - w_h\|_1 \leq \frac{\gamma_1}{\gamma_2} \inf_{v \in \mathcal{S}_0} \|u - v\|_1 \leq \frac{\gamma_1}{\gamma_2} \|u - u_\mathrm{I}\|_1 \, . \tag{5.50}$$

It remains to bound the error of interpolating polygons. This bound is provided by the following lemma, which is formulated for C^2-smooth functions u:

Lemma 5.13 (Error of an Interpolating Polygon) *For $u \in C^2$ let u_I be an arbitrary interpolating polygon and h the maximal distance between two consecutive nodes. Then*

(a) $\max\limits_x |u(x) - u_\mathrm{I}(x)| \leq \frac{h^2}{8} \max |u''(x)| \, ,$

(b) $\max\limits_x |u'(x) - u_\mathrm{I}'(x)| \leq h \max |u''(x)| \, .$

We leave the proof to the reader (⟶ Exercise 5.11). Lemma 5.13 asserts

$$\|u - u_1\|_1 = O(h)\,,$$

which together with (5.50) implies the claimed error statement

$$\|u - w_h\|_1 = O(h)\,. \tag{5.51}$$

Recall that this assertion is based on a continuous and $\mathcal{H}^1$-elliptic bilinear form and on hat functions φ_i. The $O(h)$-order in (5.51) is dominated by the unfavorable $O(h)$-order of the first-order derivative in Lemma 5.13(b). This low order is at variance with the actually observed $O(h^2)$-order attained by the approximation w_h itself (not its derivative). In fact, the square order holds. The final result is

$$\|u - w_h\|_0 \le Ch^2\|u\|_2 \tag{5.52}$$

for a constant C. This result is proven with the following lemma, which is based on a tricky idea due to Nitsche.

Lemma 5.14 (Nitsche) *Assume b is a symmetric bilinear form satisfying Assumptions 5.11, and u and w are defined as above. Then*

$$\|u - w\|_1 \le Kh^1\|f\|_0 \;\; \textit{implies} \;\; \|u - w\|_0 \le Ch^2\|f\|_0\,.$$

Proof Consider the auxiliary problem $Lz = \tilde{f} := u - w$, with weak version

$$b(z, \tilde{v}) = (\tilde{f}, \tilde{v})_0 \quad \text{for all } \tilde{v} \in \mathcal{H}_0^1\,,$$

which defines z. Choose specifically $\tilde{v} = u - w = \tilde{f}$. Then

$$b(z, u - w) = (u - w, u - w)_0 = \|u - w\|_0^2\,.$$

Invoking the error-projection property (5.48) we note

$$0 = b(u - w, v) = b(v, u - w) \quad \text{for all } v \in \mathcal{S}_0\,.$$

Subtracting this, yields

$$b(z - v, u - w) = \|u - w\|_0^2 \quad \text{for all } \; v \in \mathcal{S}_0\,.$$

We apply the continuity of b,

$$\|u - w\|_0^2 \le \gamma_1\|z - v\|_1\,\|u - w\|_1 \quad \text{for all } v \in \mathcal{S}_0\,,$$

and choose specifically v as the finite-element approximation of z. Then

$$\|u - w\|_0^2 \le \gamma_1 K_1 h^1 \|\tilde{f}\|_0 \cdot K_2 h^1 \|f\|_0 = Ch^2 \|u - w\|_0 \, \|f\|_0 \,,$$

from which the assertion follows.

This error of the order h^2 can be observed for the examples of Sect. 5.4, but not easily. The error is somewhat hidden among the other errors, namely, localization error, interpolation error, and the error of the time discretization.

The derivations of this section have been focused on the model problem (5.40)/(5.41) with a second-order differential equation and one independent variable $x\,(n = 1)$, and have been based on linear elements. Most of the assertions can be generalized to higher-order differential equations, to higher-dimensional domains $(n > 1)$, and to nonlinear elements. For example, in case the elements in $\mathcal{S}$ are polynomials of degree k, and the differential equation is of order $2l$, $\mathcal{S} \subseteq \mathcal{H}^l$, and the corresponding bilinear form on $\mathcal{H}^l$ satisfies the Assumptions 5.11 with norm $\| \; \|_l$, then the inequality

$$\|u - w_h\|_l \le Ch^{k+1-l} \|u\|_{k+1}$$

holds. This general statement includes for $k = 1$, $l = 1$ the special case of Eq. (5.52) discussed above. For the analysis of the general case, we refer to [79, 162]. This includes boundary conditions more general than the homogeneous Dirichlet conditions of (5.41).

5.6 Notes and Comments

On Sect. 5.1

As an alternative to piecewise defined finite elements one may use polynomials φ_j that are defined globally on $\mathcal{D}$, and that are pairwise orthogonal. Then the orthogonality is the reason for the vanishing of many integrals. Such type of methods are called spectral methods. Since the φ_i are globally smooth on $\mathcal{D}$, spectral methods can produce high accuracies. In other context, spectral methods were applied in [142]. For historical remarks on Ritz–Galerkin type methods, see [145].

Specifically designed basis functions can be generated by some low-dimensional approximation, comparable to PCA in finite dimensions ($\longrightarrow$ Exercise 2.16). Functions are suitable that represent preferred patterns of the solution. Then the number N of modes φ_i can be small. Such methods are described under the heading *principal orthogonal decomposition* (POD), or Karhunen–Loève expansion.

On Sect. 5.2

In the early stages of their development, finite-element methods have been applied intensively in structural engineering. In this field, stiffness matrix and mass matrix have a physical meaning leading to these names [382].

On Sect. 5.3

The approximation $\sum w_i(\tau)\varphi_i(x)$ for $\hat{y}$ is a one-dimensional finite-element approach. The geometry of the grid and the accuracy resemble the finite-difference approach. A two-dimensional approach as in

$$\sum w_i \varphi_i(x, \tau)$$

with two-dimensional hat functions and constant w_i is more involved and more flexible. Sections 5.3.2–5.3.4 widely follow [376].

On Sect. 5.4

For the calculation of the local integrals on an arbitrary triangle $\mathcal{D}_k$ consult the special FEM literature, such as [335]. In general an irregular triangulation better exploits the potential adaptivity of FEM. In particular, close to the barriers a fine mesh is required for high accuracy [304]. Since the gradient of u varies with time, a dynamic mesh refinement might be advisable, provided accuracy or stability do not deteriorate. For American options, boundary conditions $V = \Psi$ along the boundary are recommendable. For an illustration of assembling, see Topic 12 of the *Topics fCF*.

On Sect. 5.5

The assumption $u \in \mathcal{C}^2$ in Lemma 5.13 can be weakened to $u'' \in \mathcal{L}^2$ [351]. For domains $\mathcal{D} \in \mathbb{R}^2$ the claim of Lemma 5.13 holds analogously; then the second-order derivative u'' is replaced by the Hessian matrix of the second-order derivatives of u. This can be applied to mesh adaption, where one attempts to place nodes such that the Hessian is equilibrated across the mesh. The finite-dimensional function space $\mathcal{S}_0$ in (5.47) is assumed to be *sub*space of $\mathcal{H}_0^1$. Elements with this property are called *conforming elements*. A more accurate notation for $\mathcal{S}_0$ of (5.47) is $\mathcal{S}_0^1$. In the general case, conforming elements are characterized by $\mathcal{S}^l \subseteq \mathcal{H}^l$. In the representation of

v in Eq. (5.47) we avoid discussing the technical issue of how to organize different types of boundary conditions.

There are also smooth basis functions φ, for example, cubic Hermite polynomials. For sufficiently smooth solutions, such basis functions produce higher accuracy than hat functions do. For the accuracy of finite-element methods consult, for example, [2, 19, 53, 79, 162, 351].

On Other Methods

Finite-element methods are frequently used for approximating exotic options, in particular in multidimensional situations. For different types of options special methods have been developed. For applications, computational results and accuracies see also [2, 361, 362]. Front-fixing has been applied with finite elements in [188]. The accuracy aspect is also treated in [144]. Ritz–Galerkin methods are used with wavelet functions in [185, 263]; the latter paper is specifically devoted to stochastic volatility. A penalty approach with FEM is discussed in [230], where rectangular subdomains are furnished with basis functions as product of one-dimensional hat functions of the type $\varphi(x, y) = \varphi_i(x)\varphi_j(y)$.

5.7 Exercises

5.1 (Elliptical Probability Curves)
Suppose the situation of two asset prices $S_1(t)$ and $S_2(t)$ for $t > 0$ governed by GBM (3.35), with initial price point $(S_1(0), S_2(0))$. Barriers of a barrier option can be aligned such that the probability of $(S_1(t), S_2(t))$ reaching the barrier has the same constant value. Define $Y_1 := \log S_1$, $Y_2 := \log S_2$.

(a) Show that the curve of constant probability in the (Y_1, Y_2)-plane has an elliptical shape.
(b) Let the covariance matrix be

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Calculate its eigenvalues and eigenvectors.
(c) Sketch representative ellipses in a (Y_1, Y_2)-plane. How do they depend on ρ?

5.2 (Cubic B-Spline)
Suppose an equidistant partition of an interval be given with mesh size $h = x_{k+1} - x_k$. Cubic B-splines have a support of four subintervals. In each subinterval the spline is a piece of polynomial of degree three. Apart from special boundary splines, the

cubic B-splines φ_i are determined by the requirements

$$\begin{aligned} \varphi_i(x_i) &= 1 \\ \varphi_i(x) &\equiv 0 \quad \text{for } x < x_{i-2} \\ \varphi_i(x) &\equiv 0 \quad \text{for } x > x_{i+2} \\ \varphi &\in \mathcal{C}^2(-\infty, \infty)\,. \end{aligned}$$

To construct these φ_i proceed as follows:

(a) Construct a spline $S(x)$ that satisfies the above requirements for the special nodes

$$\tilde{x}_k := -2 + k \quad \text{for } k = 0, 1, \ldots, 4\,.$$

(b) Find a transformation $T_i(x)$, such that $\varphi_i = S(T_i(x))$ satisfies the requirements for the original nodes.
(c) For which i, j does $\varphi_i \varphi_j = 0$ hold?

5.3 (Finite-Element Matrices)
For the hat functions φ from Sect. 5.2 calculate for arbitrary subinterval $\mathcal{D}_k$ all nonzero integrals of the form

$$\int \varphi_i \varphi_j \,\mathrm{d}x, \quad \int \varphi_i' \varphi_j \,\mathrm{d}x, \quad \int \varphi_i' \varphi_j' \,\mathrm{d}x$$

and represent them as local 2×2 matrices.

5.4 (Calculating Options with Finite Elements)
Design an algorithm for the pricing of standard options by means of finite elements. To this end proceed as outlined in Sect. 5.3. Start with a simple version using an equidistant discretization step Δx. If this is working properly change the algorithm to a version with nonequidistant x-grid. Distribute the nodes x_i closer around $x = 0$. Always place a node at the strike.

5.5 (Black-Scholes Equation in Divergence-Free Form)

(a) Prove the equivalence of (5.31) and (5.32), where D and b are given by (5.33). Specialize this to the one-dimensional case of the Black–Scholes equation.
(b) Show

$$b^{tr} \nabla u + ru = \nabla \cdot (bu) + \gamma u$$

and determine γ for the two-dimensional case, and for the Black–Scholes equation.

(c) With the transformation

$$x := \log(\frac{S_1}{K_1}),\ y := \log(\frac{S_2}{K_2})$$

and writing $u(x, y, t)$ for V leads to the PDE

$$u_t + \frac{1}{2}\sigma_1^2 u_{xx} + (r - \delta_1 - \frac{1}{2}\sigma_1^2)u_x - ru$$
$$+ \frac{1}{2}\sigma_2^2 u_{yy} + (r - \delta_2 - \frac{1}{2}\sigma_2^2)u_y + \rho\sigma_1\sigma_2 u_{xy} = 0\,.$$

What are the matrix D and the vector b such that we arrive at (5.32)?

5.6 (Outward Normals)
The boundary $\partial\mathcal{D}$ of the trapezoidal domain $\mathcal{D}$ in Fig. 5.5 consists of four straight lines. What are the four unit outward vectors n orthogonal to $\partial\mathcal{D}$? Give a parameter representation of the boundary.

5.7 (Gradient on a Triangle)
Consider hat functions φ on a triangular element $\mathcal{D}_k$ with vertex nodes numbers $\mathcal{I}_k = \{i, j, l\}$, and the local plane on $\mathcal{D}_k$ represented by

$$w(x, y) = w_i\varphi_i(x, y) + w_j\varphi_j(x, y) + w_l\varphi_l(x, y)\,.$$

(a) In the three-dimensional (x, y, w)-space let the plane $w(x, y) = c_1 + c_2\,x + c_3\,y$ interpolate the three points $(x_i,\ y_i,\ w_i)$, $i = 1, 2, 3$ (local node numbering). That is,

$$\begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix},$$

shortly $Ac = w$. Establish a formula for the gradient $\nabla w = (c_2, c_3)^{tr}$, showing that there is a (2×3)-matrix G_k such that

$$\nabla w = G_k w\,.$$

Hint: Use Cramer's rule; $|F_k|$ is the area of the triangle, where

$$F_k := \frac{1}{2}\det(A)\,.$$

(b) Show

$$(\nabla\varphi_i \,|\, \nabla\varphi_j \,|\, \nabla\varphi_l) = G_k \,.$$

(c) Show

$$\int_{\mathcal{D}_k} \nabla\varphi_i^{tr} \nabla\varphi_j \,\mathrm{d}x\,\mathrm{d}y = \nabla\varphi_i^{tr} \nabla\varphi_j \,|F_k| \,,$$

and all nine integrals of the element stiffness matrix are obtained by

$$|F_k| G_k^{tr} G_k \,.$$

5.8 (Assembling)
Consider the domain $\mathcal{D} := \{(x, y) \,|\, x \geq 0,\ y \geq 0,\ \ 1 \leq x + y \leq 2\}$ tiled by 12 triangles $\mathcal{D}_k$, where triangles and vertices are numbered as in Fig. 5.13.

(a) Set up the index set $\mathcal{I}$ with entries $\mathcal{I}_k = \{i_k, j_k, l_k\}$, which assigns node numbers to the kth triangle, for $1 \leq k \leq 12$.
(b) Formulate the assembling algorithm that builds up the global stiffness matrix out of the element stiffness matrices

$$\begin{pmatrix} s_{11}^{(k)} & s_{12}^{(k)} & s_{13}^{(k)} \\ s_{21}^{(k)} & s_{22}^{(k)} & s_{23}^{(k)} \\ s_{31}^{(k)} & s_{32}^{(k)} & s_{33}^{(k)} \end{pmatrix}$$

for a general index set $\mathcal{I}$ and $1 \leq k \leq m$.

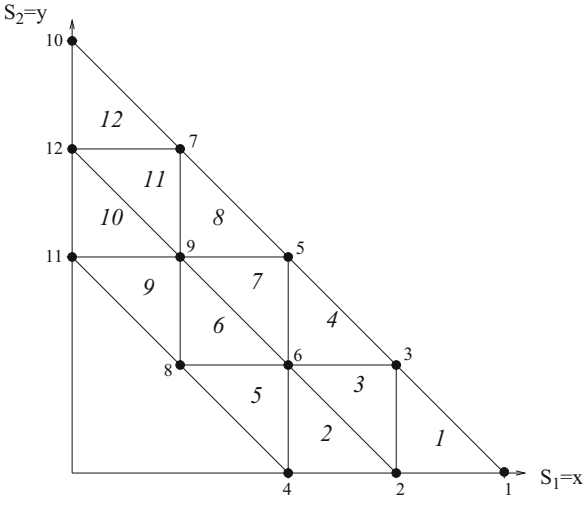

Fig. 5.13 Specific triangulation and numbering, see Exercise 5.8

(c) The example of Fig. 5.13 leads to a banded stiffness matrix. What is the bandwidth?

5.9 (Variable Volatility (Project))
For variable volatility $\sigma(S,t)$ and constant K, T, r, δ, PDEs of the type

$$\frac{\partial y}{\partial \tau} - \frac{1}{2}\hat{\sigma}^2(x,\tau)\left(\frac{\partial^2 y}{\partial x^2} - \frac{1}{4}y\right) = 0$$

are to be solved, with $\tau = T - t$ and transformations $S \leftrightarrow x$, $V \leftrightarrow y$ from the Black–Scholes model given by (A.25), (A.26); consult Appendix A.6.

(a) For an American put, apply these transformations to derive from $V(S,t) \geq (K-S)^+$ an inequality $y(x,\tau) \geq g(x,\tau)$.
(b) Carry out the finite-element formulation for the linear complementarity problem analogously as in Sect. 5.3.4.
(c) Integrals will include local integrals

$$\int \sigma^2(x,\tau)\varphi_i\varphi_j \,\mathrm{d}x\,, \quad \int \sigma^2(x,\tau)\varphi_i'\varphi_j \,\mathrm{d}x\,.$$

Apply Simpson's quadrature rule

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]$$

to approximate the above local integrals.
(d) Set up a finite-element code, and test it with the artificial function [128]

$$\sigma(S) := 0.3 - \frac{0.2}{\log(S/K)^2 + 1}\,.$$

5.10 Assume a function $v(\zeta)$ with $\alpha \leq \zeta \leq \beta$ and $v(\alpha) = 0$.

(a) Show

$$(v(\zeta))^2 \leq (\zeta - \alpha)\int_\alpha^\zeta (v'(x))^2 \,\mathrm{d}x\,.$$

Hint: Recall $v(\zeta) = \int_\alpha^\zeta v'(x)\,\mathrm{d}x$, and apply the Schwarzian inequality (C.16).
(b) Use (a) to show

$$\int_\alpha^\beta (v(\zeta))^2 \,\mathrm{d}\zeta \leq \frac{1}{2}(\beta-\alpha)^2 \int_\alpha^\beta (v'(x))^2 \,\mathrm{d}x\,.$$

5.11 Prove Lemma 5.13, and for $u \in \mathcal{C}^2$ the assertion $\|u - w_h\|_1 = O(h)$.