5 Finite Element Methods for Elliptic Equations

Over the last decades the *finite element method*, which was introduced by engineers in the 1960s, has become the perhaps most important numerical method for partial differential equations, particularly for equations of elliptic and parabolic types. This method is based on the variational form of the boundary value problem and approximates the exact solution by a piecewise polynomial function. It is more easily adapted to the geometry of the underlying domain than the finite difference method, and for symmetric positive definite elliptic problems it reduces to a finite linear system with a symmetric positive definite matrix.

We first introduce this method in Sect. 5.1 for the case of a two-point boundary value problem and show a number of error estimates. In Sect. 5.2 we then formulate the method for a two-dimensional model problem. Here the piecewise polynomial approximations are defined on triangulations of the spatial domain, and in the following Sect. 5.3 we study such approximation in more detail. In Sect. 5.4 we show basic error estimates for the finite element method for the model problem, using piecewise linear approximating functions. All error bounds derived up to this point contain a norm of the unknown exact solution and are therefore often referred to as *a priori* error estimates. In Sect. 5.5 we show a so-called *a posteriori* error estimate in which the error bound is expressed in terms of the data of the problem and the computed solution. In Sect. 5.6 we analyze the effect of numerical integration, which is often used when the finite element equations are assembled in a computer program. In Sect. 5.7 we briefly describe a so-called *mixed finite element method*.

5.1 A Two-Point Boundary Value Problem

We consider the special case $b = 0$ of the two-point boundary value problem treated in Sect. 2.3,

$$(5.1) \qquad \mathcal{A}u := -(au')' + cu = f \quad \text{in } \Omega := (0,1), \quad \text{with } u(0) = u(1) = 0,$$

where $a = a(x),\ c = c(x)$ are smooth functions with $a(x) \geq a_0 > 0,\ c(x) \geq 0$ in $\bar{\Omega}$, and $f \in L_2 = L_2(\Omega)$. We recall that the variational formulation of this problem is to find $u \in H_0^1$ such that

$$a(u,\varphi) = (f,\varphi), \quad \forall \varphi \in H_0^1, \tag{5.2}$$

where

$$a(v,w) = \int_\Omega (av'w' + cvw)\,\mathrm{d}x \quad \text{and} \quad (f,v) = \int_\Omega fv\,\mathrm{d}x,$$

and that this problem has a unique solution $u \in H^2$.

For the purpose of finding an approximate solution of (5.2) we introduce a partition of Ω,

$$0 = x_0 < x_1 < \cdots < x_M = 1,$$

and set

$$h_j = x_j - x_{j-1},\ K_j = [x_{j-1}, x_j], \quad \text{for } j = 1,\dots,M, \quad \text{and} \ \ h = \max_j h_j.$$

The discrete solution will be sought in the finite-dimensional space of functions

$$S_h = \big\{ v \in \mathcal{C} = \mathcal{C}(\bar{\Omega}) : v \text{ linear on each } K_j,\ v(0) = v(1) = 0 \big\}.$$

(By a linear function we understand a function of the form $f(x) = \alpha x + \beta$; strictly speaking such a function is called an affine function when $\beta \neq 0$.) It is easy to see that $S_h \subset H_0^1$. The set $\{\Phi_i\}_{i=1}^{M-1} \subset S_h$ of *hat functions* defined by

$$\Phi_i(x_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}$$

see Fig. 5.1, is a basis for S_h, and any $v \in S_h$ may be written as

$$v(x) = \sum_{i=1}^{M-1} v_i \Phi_i(x), \quad \text{with } v_i = v(x_i).$$

We now pose the finite-dimensional problem to find $u_h \in S_h$ such that

$$a(u_h, \chi) = (f, \chi), \quad \forall \chi \in S_h. \tag{5.3}$$

In terms of the basis $\{\Phi_i\}_{i=1}^{M-1}$ we write $u_h(x) = \sum_{j=1}^{M-1} U_j \Phi_j(x)$ and insert this into (5.3) to find that this equation is equivalent to

$$\sum_{j=1}^{M-1} U_j a(\Phi_j, \Phi_i) = (f, \Phi_i), \quad \text{for } i = 1,\dots,M-1. \tag{5.4}$$

This linear system of equations may be expressed in matrix form as

$$AU = b, \tag{5.5}$$

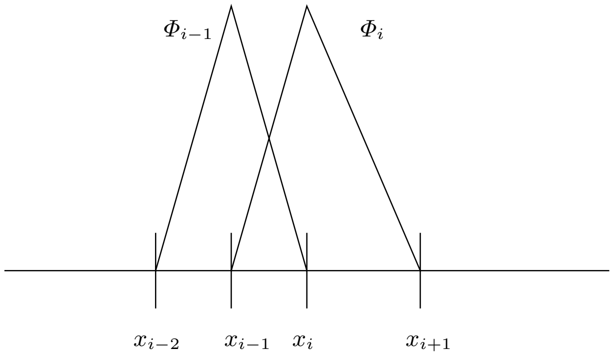

Fig. 5.1. Hat functions.

where $U = (U_i)$, $A = (a_{ij})$ is the *stiffness matrix* with elements $a_{ij} = a(\Phi_j, \Phi_i)$, and $b = (b_i)$ the *load vector* with elements $b_i = (f, \Phi_i)$. The matrix A is symmetric and positive definite, because for $V = (V_i)$ and $v(x) = \sum_{i=1}^{M-1} V_i \Phi_i(x)$ we have

$$V^{\mathrm{T}} A\, V = \sum_{i,j=1}^{M-1} V_i a_{ij} V_j = a\Big(\sum_{j=1}^{M-1} V_j \Phi_j, \sum_{i=1}^{M-1} V_i \Phi_i \Big) = a(v,v) \geq a_0 \|v'\|^2,$$

and hence $V^{\mathrm{T}} AV = 0$ implies $v' = 0$, so that v is constant $= 0$ because $v(0) = 0$, and thus $V = 0$. It follows that (5.5), and therefore also (5.3), has a unique solution, which is the *finite element solution* of (5.1). The matrix A is tridiagonal since $a_{ij} = 0$ when x_i and x_j are not neighbors, i.e., when $|i - j| \geq 2$, and the system (5.5) is therefore easy to solve, see App. B.1.

We note that when $\mathcal{A}u = -u''$ and the meshsize is constant, i.e., when $h_j = h = 1/M$ for $j = 1, \dots, M$, then, with the notation of Sect. 4.1, the equation (5.4) may be written

$$-\partial\bar{\partial} U_j = h^{-1}(f, \Phi_j), \quad j = 1, \dots, M-1 \tag{5.6}$$

(cf. Problem 5.2). The finite element method thus coincides with the finite difference equation (4.3), except that an average of f over $(x_j - h, x_j + h)$ is now used instead of the point-values $f_j = f(x_j)$.

The idea of replacing the space H_0^1 in (5.2) by a finite-dimensional subspace and to determine the coefficients of the corresponding approximate solution as in (5.4) is referred to as Galerkin's method. The finite element method is thus Galerkin's method, applied with a special choice of the finite-dimensional subspace, namely, in this case, the space of continuous, piecewise linear functions. The intervals K_j, together with the restriction of these functions to K_j, are then thought of as the finite elements.

Before we analyze the error in the finite element solution u_h, we discuss some approximation properties of the space S_h. We define the piecewise linear interpolant $I_h v \in S_h$ of a function $v \in \mathcal{C} = \mathcal{C}(\bar{\Omega})$ with $v(0) = v(1) = 0$ by

$$I_h v(x_j) = v(x_j), \quad j = 1, \dots, M-1.$$

Recall that $H_0^1 \subset \mathcal{C}$ in one dimension by Sobolev's inequality, Theorem A.5, so that $I_h v$ is defined for $v \in H_0^1$. It may be shown, which we leave as an exercise, see Problem 5.1, that, with $\|v\|_{K_j} = \|v\|_{L_2(K_j)}$ and $|v|_{2,K_j} = |v|_{H^2(K_j)}$,

$$\|I_h v - v\|_{K_j} \le C h_j^2 |v|_{2,K_j} \tag{5.7}$$

and

$$\|(I_h v - v)'\|_{K_j} \le C h_j |v|_{2,K_j}. \tag{5.8}$$

It follows that

$$\begin{aligned} \|I_h v - v\| &= \Big(\sum_{j=1}^{M} \|I_h v - v\|_{K_j}^2\Big)^{1/2} \le \Big(\sum_{j=1}^{M} C^2 h_j^4 |v|_{2,K_j}^2\Big)^{1/2} \\ &\le C h^2 \|v\|_2, \quad \forall v \in H^2, \end{aligned} \tag{5.9}$$

and similarly

$$\|(I_h v - v)'\| \le C h \|v\|_2, \quad \text{for } v \in H^2. \tag{5.10}$$

We now turn to the task of estimating the error in the finite element approximation u_h defined by (5.3). Since $a(\cdot,\cdot)$ is symmetric positive definite, it is an inner product on H_0^1, and the corresponding norm is the energy norm

$$\|v\|_a = a(v,v)^{1/2} = \Big(\int_0^1 \big(a(v')^2 + cv^2\big)\,\mathrm{d}x\Big)^{1/2}. \tag{5.11}$$

Theorem 5.1. *Let u_h and u be the solutions of* (5.3) *and* (5.2). *Then*

$$\|u_h - u\|_a = \min_{\chi \in S_h} \|\chi - u\|_a, \tag{5.12}$$

and

$$\|u_h' - u'\| \le C h \|u\|_2. \tag{5.13}$$

Proof. Since $S_h \subset H_0^1$ we may take $\varphi = \chi \in S_h$ in (5.2) and subtract it from (5.3) to obtain

$$a(u_h - u, \chi) = 0, \quad \forall \chi \in S_h. \tag{5.14}$$

This equation means that the finite element solution u_h may be described as the orthogonal projection of the exact solution u onto S_h with respect to the inner product $a(\cdot,\cdot)$. This also immediately implies that u_h is the best approximation of u in S_h with respect to the energy norm, and hence that (5.12) holds. This can be seen directly as follows: Using (5.14) we have, for any $\chi \in S_h$,

$$\|u_h - u\|_a^2 = a(u_h - u, u_h - u) = a(u_h - u, \chi - u) \le \|u_h - u\|_a \, \|\chi - u\|_a,$$

which shows (5.12) after cancellation of a factor $\|u_h - u\|_a$. By our assumptions we have, with C independent of h,

$$\sqrt{a_0}\|v'\| \le \|v\|_a \le C\|v'\|, \quad \text{for } v \in H_0^1,$$

where the first inequality is obvious by (5.11) and the second follows from (2.17). Hence, (5.12) implies

$$\|(u_h - u)'\| \le C\|u_h - u\|_a \le C \min_{\chi \in S_h} \|(\chi - u)'\|. \tag{5.15}$$

Taking $\chi = I_h u$ and using the interpolation error bound in (5.10), we obtain (5.13), and the proof is complete. □

Our next result concerns the L_2-norm of the error.

Theorem 5.2. *Let u_h and u be the solutions of* (5.3) *and* (5.2). *Then*

$$\|u_h - u\| \le Ch^2\|u\|_2. \tag{5.16}$$

Proof. We use a duality argument based on the auxiliary problem

$$\mathcal{A}\phi = e \quad \text{in } \Omega, \quad \text{with } \phi(0) = \phi(1) = 0, \quad \text{where } e = u_h - u. \tag{5.17}$$

Its weak formulation is to find $\phi \in H_0^1$ such that

$$a(w, \phi) = (w, e), \quad \forall w \in H_0^1. \tag{5.18}$$

We put the test function w on the left side, because (5.18) plays the role of the adjoint (or dual) problem to (5.2). Of course, this makes no difference here since $a(\cdot,\cdot)$ is symmetric, but is important in the case of a nonsymmetric differential operator $\mathcal{A}$, see Problem 5.7. By the regularity estimate (2.22) we have

$$\|\phi\|_2 \le C\|\mathcal{A}\phi\| = C\|e\|. \tag{5.19}$$

Taking $w = e$ in (5.18) and using (5.14) and (5.10), we therefore obtain

$$\begin{aligned}\|e\|^2 = a(e, \phi) = a(e, \phi - I_h\phi) &\le C\|e'\| \, \|(\phi - I_h\phi)'\| \\ &\le Ch\|e'\| \, \|\phi\|_2 \le Ch\|e'\| \, \|e\|.\end{aligned}$$

Cancelling one factor $\|e\|$ we see that we have gained one factor h over the error estimate for e',

$$\|e\| \le Ch\|e'\|, \tag{5.20}$$

and the proof may now be completed by using (5.13). □