Chapter 2

One-Dimensional Finite Element Methods

2.1 Introduction

The piecewise-linear Galerkin finite element method of Chapter 1 can be extended in several directions. The most important of these is multi-dimensional problems; however, we'll postpone this until the next chapter. Here, we'll address and answer some other questions that may be inferred from our brief encounter with the method.

1. Is the Galerkin method the best way to construct a variational principal for a partial differential system?

2. How do we construct variational principals for more complex problems? Specifically, how do we treat boundary conditions other than Dirichlet?

3. The finite element method appeared to converge as $O(h)$ in strain energy and $O(h^2)$ in L^2 for the example of Section 1.3. Is this true more generally?

4. Can the finite element solution be improved by using higher-degree piecewise-polynomial approximations? What are the costs and benefits of doing this?

We'll tackle the Galerkin formulations in the next two sections, examine higher-degree piecewise polynomials in Sections 2.4 and 2.5, and conclude with a discussion of approximation errors in Section 2.6.

2.2 Galerkin's Method and Extremal Principles

> "For since the fabric of the universe is most perfect and the work of a most wise creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear."

- Leonhard Euler

Although the construction of variational principles from differential equations is an important aspect of the finite element method it will not be our main objective. We'll explore some properties of variational principles with a goal of developing a more thorough understanding of Galerkin's method and of answering the questions raised in Section 2.1. In particular, we'll focus on boundary conditions, approximating spaces, and extremal properties of Galerkin's method. Once again, we'll use the model two-point Dirichlet problem

$$\mathcal{L}[u] := -[p(x)u']' + q(x)u = f(x), \qquad 0 < x < 1, \tag{2.2.1a}$$

$$u(0) = u(1) = 0, \tag{2.2.1b}$$

with $p(x) > 0$, $q(x) \geq 0$, and $f(x)$ being smooth functions on $0 \leq x \leq 1$.

As described in Chapter 1, the Galerkin form of (2.2.1) is obtained by multiplying (2.2.1a) by a test function $v \in H_0^1$, integrating the result on $[0, 1]$, and integrating the second-order term by parts to obtain

$$A(v, u) = (v, f), \qquad \forall v \in H_0^1, \tag{2.2.2a}$$

where

$$(v, f) = \int_0^1 vf dx, \tag{2.2.2b}$$

and

$$A(v, u) = (v', pu') + (v, qu) = \int_0^1 (v'pu' + vqu)dx, \tag{2.2.2c}$$

and functions v belonging to the Sobolev space H^1 have bounded values of

$$\int_0^1 [(v')^2 + v^2]dx.$$

For (2.2.1), a function v is in H_0^1 if it also satisfies the trivial boundary conditions $v(0) = v(1) = 0$. As we shall discover in Section 2.3, the definition of H_0^1 will depend on the type of boundary conditions being applied to the differential equation.

There is a connection between self-adjoint differential problems such as (2.2.1) and the minimum problem: find $w \in H_0^1$ that minimizes

$$I[w] = A(w, w) - 2(w, f) = \int_0^1 [p(w')^2 + qw^2 - 2wf]dx. \tag{2.2.3}$$

Maximum and minimum variational principles occur throughout mathematics and physics and a discipline called the Calculus of Variations arose in order to study them. The initial goal of this field was to extend the elementary theory of the calculus of the maxima and minima of functions to problems of finding the extrema of functionals such as $I[w]$. (A *functional* is an operator that maps functions onto real numbers.)

The construction of the Galerkin form (2.2.2) of a problem from the differential form (2.2.1) is straight forward; however, the construction of the extremal problem (2.2.3) is not. We do not pursue this matter here. Instead, we refer readers to a text on the calculus of variations such as Courant and Hilbert [4]. Accepting (2.2.3), we establish that the solution u of Galerkin's method (2.2.2) is optimal in the sense of minimizing (2.2.3).

Theorem 2.2.1. *The function $u \in H_0^1$ that minimizes (2.2.3) is the one that satisfies (2.2.2a) and conversely.*

Proof. Suppose first that $u(x)$ is the solution of (2.2.2a). We choose a real parameter ϵ and any function $v(x) \in H_0^1$ and define the *comparison function*

$$w(x) = u(x) + \epsilon v(x). \tag{2.2.4}$$

For each function $v(x)$ we have a one parameter family of comparison functions $w(x) \in H_0^1$ with the solution $u(x)$ of (2.2.2a) obtained when $\epsilon = 0$. By a suitable choice of ϵ and $v(x)$ we can use (2.2.4) to represent any function in H_0^1. A comparison function $w(x)$ and its *variation* $\epsilon v(x)$ are shown in Figure 2.2.1.

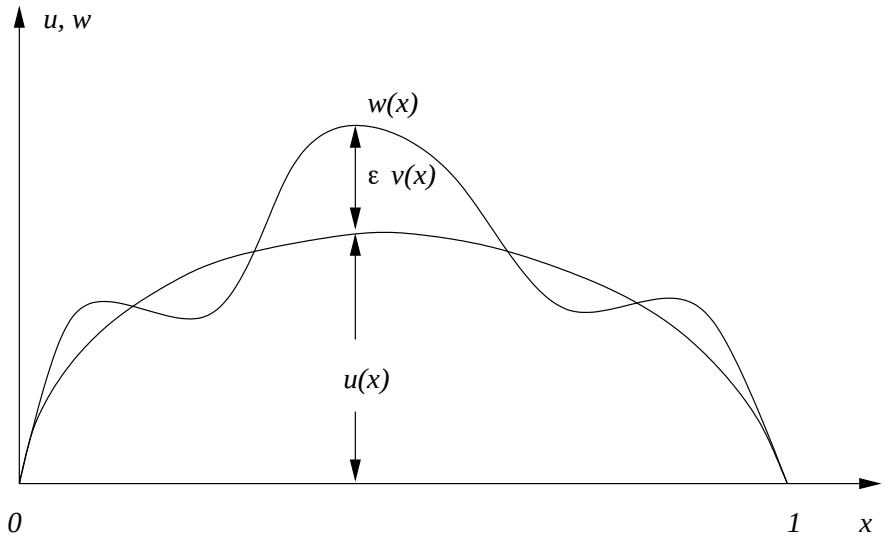

Figure 2.2.1: A comparison function $w(x)$ and its variation $\epsilon v(x)$ from $u(x)$.

Substituting (2.2.4) into (2.2.3)

$$I[w] = I[u + \epsilon v] = A(u + \epsilon v, u + \epsilon v) - 2(u + \epsilon v, f).$$

Expanding the strain energy and L^2 inner products using (2.2.2b,c)

$$I[w] = A(u,u) - 2(u,f) + 2\epsilon[A(v,u) - (v,f)] + \epsilon^2 A(v,v).$$

By hypothesis, u satisfies (2.2.2a), so the $O(\epsilon)$ term vanishes. Using (2.2.3), we have

$$I[w] = I[u] + \epsilon^2 A(v,v).$$

With $p > 0$ and $q \geq 0$, we have $A(v,v) \geq 0$; thus, u minimizes (2.2.3).

In order to prove the converse, assume that $u(x)$ minimizes (2.2.3) and use (2.2.4) to obtain

$$I[u] \leq I[u + \epsilon v].$$

For a particular choice of $v(x)$, let us regard $I[u+\epsilon v]$ as a function $\Phi(\epsilon)$, *i.e.*,

$$I[u+\epsilon v] := \Phi(\epsilon) = A(u+\epsilon v, u+\epsilon v) - 2(u+\epsilon v, f).$$

A necessary condition for a minimum to occur at $\epsilon = 0$ is $\Phi'(0) = 0$; thus, differentiating

$$\Phi'(\epsilon) = 2\epsilon A(v,v) + 2A(v,u) - 2(v,f)$$

and setting $\epsilon = 0$

$$\Phi'(0) = 2[A(v,u) - (v,f)] = 0.$$

Thus, u is a solution of (2.2.2a). □

The following corollary verifies that the minimizing function u is also unique.

Corollary 2.2.1. *The solution u of (2.2.2a) (or (2.2.3)) is unique.*

Proof. Suppose there are two functions $u_1, u_2 \in H_0^1$ satisfying (2.2.2a), *i.e.*,

$$A(v,u_1) = (v,f), \qquad A(v,u_2) = (v,f), \qquad \forall v \in H_0^1.$$

Subtracting

$$A(v, u_1 - u_2) = 0, \qquad \forall v \in H_0^1.$$

Since this relation is valid for *all* $v \in H_0^1$, choose $v = u_1 - u_2$ to obtain

$$A(u_1 - u_2, u_1 - u_2) = 0.$$

If $q(x) > 0$, $x \in (0,1)$, then $A(u_1 - u_2, u_1 - u_2)$ is positive unless $u_1 = u_2$. Thus, it suffices to consider cases when either (i) $q(x) \equiv 0$, $x \in [0,1]$, or (ii) $q(x)$ vanishes at isolated points or subintervals of $(0,1)$. For simplicity, let us consider the former case. The analysis of the latter case is similar.

When $q(x) \equiv 0$, $x \in [0,1]$, $A(u_1 - u_2, u_1 - u_2)$ can vanish when $u_1' - u_2' = 0$. Thus, $u_1 - u_2$ is a constant. However, both u_1 and u_2 satisfy the trivial boundary conditions (2.2.1b); thus, the constant is zero and $u_1 = u_2$. □

Corollary 2.2.2. *If u, w are smooth enough to permit integrating $A(u, v)$ by parts then the minimizer of (2.2.3), the solution of the Galerkin problem (2.2.2a), and the solution of the two-point boundary value problem (2.2.1) are all equivalent.*

Proof. Integrate the differentiated term in (2.2.3) by parts to obtain

$$I[w] = \int_0^1 [-w(pw')' + qw^2 - 2fw]dx + wpw'|_0^1.$$

The last term vanishes since $w \in H_0^1$; thus, using (2.2.1a) and (2.2.2b) we have

$$I[w] = (w, \mathcal{L}[w]) - 2(w, f). \tag{2.2.5}$$

Now, follow the steps used in Theorem 2.2.1 to show

$$A(v, u) - (v, f) = (v, \mathcal{L}[u] - f) = 0, \qquad \forall v \in H_0^1,$$

and, hence, establish the result. □

The minimization problems (2.2.3) and (2.2.5) are equivalent when w has sufficient smoothness. However, minimizers of (2.2.3) may lack the smoothness to satisfy (2.2.5). When this occurs, the solutions with less smoothness are often the ones of physical interest.

Problems

1. Consider the "stationary value" problem: find functions $w(x)$ that give stationary values (maxima, minima, or saddle points) of

$$I[w] = \int_0^1 F(x, w, w')dx \tag{2.2.6a}$$

when w satisfies the "essential" (Dirichlet) boundary conditions

$$w(0) = \alpha, \qquad w(1) = \beta. \tag{2.2.6b}$$

Let $w \in H_E^1$, where the subscript E denotes that w satisfies (2.2.6b), and consider comparison functions of the form (2.2.4) where $u \in H_E^1$ is the function that makes $I[w]$ stationary and $v \in H_0^1$ is arbitrary. (Functions in H_0^1 satisfy trivial versions of (2.2.6b), *i.e.*, $v(0) = v(1) = 0$.)

Using (2.2.1) as an example, we would have

$$F(x, w, w') = p(x)(w')^2 + q(x)w^2 - 2wf(x), \qquad \alpha = \beta = 0.$$

Smooth stationary values of (2.2.6) would be minima in this case and correspond to solutions of the differential equation (2.2.1a) and boundary conditions (2.2.1b).

Differential equations arising from minimum principles like (2.2.3) or from stationary value principles like (2.2.6) are called *Euler-Lagrange equations.*

Beginning with (2.2.6), follow the steps used in proving Theorem 2.2.1 to determine the Galerkin equations satisfied by u. Also determine the Euler-Lagrange equations for smooth stationary values of (2.2.6).

2.3 Essential and Natural Boundary Conditions

The analyses of Section 2.2 readily extend to problems having nontrivial Dirichlet boundary conditions of the form

$$u(0) = \alpha, \qquad u(1) = \beta. \tag{2.3.1a}$$

In this case, functions u satisfying (2.2.2a) or w satisfying (2.2.3) must be members of H^1 and satisfy (2.3.1a). We'll indicate this by writing $u, w \in H^1_E$, with the subscript E denoting that u and w satisfy the *essential* Dirichlet boundary conditions (2.3.1a). Since u and w satisfy (2.3.1a), we may use (2.2.4) or the interpretation of ϵv as a variation shown in Figure 2.2.1, to conclude that v should still vanish at $x = 0$ and 1 and, hence, belong to H^1_0.

When u is not prescribed at $x = 0$ and/or 1, the function v need not vanish there. Let us illustrate this when (2.2.1a) is subject to conditions

$$u(0) = \alpha, \qquad p(1)u'(1) = \beta. \tag{2.3.1b}$$

Thus, an essential or Dirichlet condition is specified at $x = 0$ and a *Neumann* condition is specified at $x = 1$. Let us construct a Galerkin form of the problem by again multiplying (2.2.1a) by a test function v, integrating on $[0, 1]$, and integrating the second derivative terms by parts to obtain

$$\int_0^1 v[-(pu')' + qu - f]dx = A(v, u) - (v, f) - vpu'|_0^1 = 0. \tag{2.3.2}$$

With an essential boundary condition at $x = 0$, we specify $u(0) = \alpha$ and $v(0) = 0$; however, $u(1)$ and $v(1)$ remain unspecified. We still classify $u \in H^1_E$ and $v \in H^1_0$ since they satisfy, respectively, the essential and trivial essential boundary conditions specified with the problem.

With $v(0) = 0$ and $p(1)u'(1) = \beta$, we use (2.3.2) to establish the Galerkin problem for (2.2.1a, 2.3.1b) as: determine $u \in H^1_E$ satisfying

$$A(v, u) = (v, f) + v(1)\beta, \qquad \forall v \in H^1_0. \tag{2.3.3}$$

Let us reiterate that the subscript E on H^1 restricts functions to satisfy Dirichlet (essential) boundary conditions, but not any Neumann conditions. The subscript 0 restricts functions to satisfy trivial versions of any Dirichlet conditions but, once again, Neumann conditions are not imposed.

As with problem (2.2.1), there is a minimization problem corresponding to (2.2.3): determine $w \in H^1_E$ that minimizes

$$I[w] = A(w, w) - 2(w, f) - 2w(1)\beta. \tag{2.3.4}$$

Furthermore, in analogy with Theorem 2.2.1, we have an equivalence between the Galerkin (2.3.3) and minimization (2.3.4) problems.

Theorem 2.3.1. *The function $u \in H^1_E$ that minimizes (2.3.4) is the one that satisfies (2.3.3) and conversely.*

Proof. The proof is so similar to that of Theorem 2.2.1 that we'll only prove that the function u that minimizes (2.3.4) also satisfies (2.3.3). (The remainder of the proof is stated as Problem 1 as the end of this section.)

Again, create the comparison function

$$w(x) = u(x) + \epsilon v(x); \tag{2.3.5}$$

however, as shown in Figure 2.3.1, $v(1)$ need not vanish. By hypothesis we have

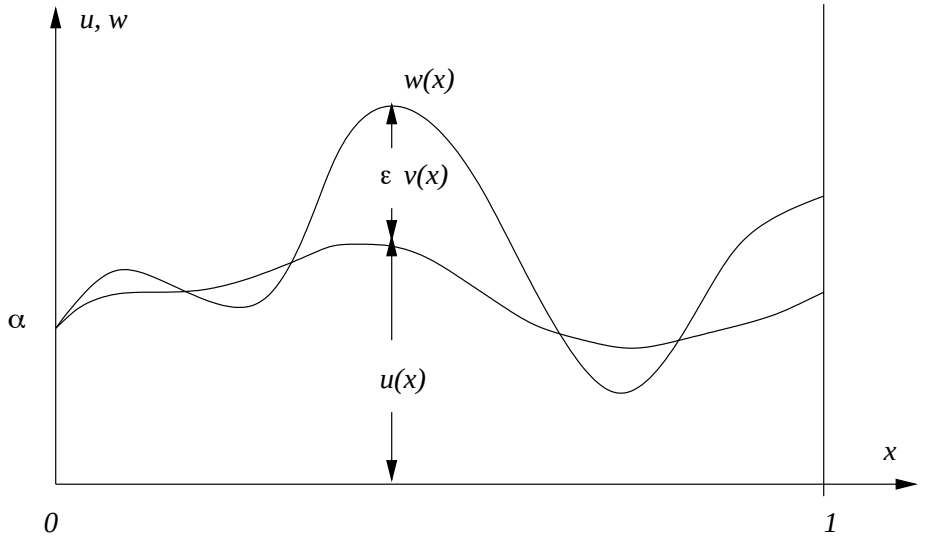

Figure 2.3.1: Comparison function $w(x)$ and variation $\epsilon v(x)$ when Dirichlet data is prescribed at $x = 0$ and Neumann data is prescribed at $x = 1$.

$$I[u] \le I[u + \epsilon v] = \Phi(\epsilon) = A(u + \epsilon v, u + \epsilon v) - 2(u + \epsilon v, f) - 2[u(1) + \epsilon v(1)]\beta.$$

Differentiating with respect to ϵ yields the necessary condition for a minimum as

$$\Phi'(0) = 2[A(v,u) - (v,f) - v(1)\beta] = 0;$$

thus, u satisfies (2.3.3). □

As expected, Theorem 2.3.1 can be extended when the minimizing function u is smooth.

Corollary 2.3.1. *Smooth functions $u \in H_E^1$ satisfying (2.3.3) or minimizing (2.3.4) also satisfy (2.2.1a, 2.3.1b).*

Proof. Using (2.2.2c), integrate the differentiated term in (2.3.3) by parts to obtain

$$\int_0^1 v[-(pu')' + qu - f]dx + v(1)[p(1)u'(1) - \beta] = 0, \qquad \forall v \in H_0^1. \tag{2.3.6}$$

Since (2.3.6) must be satisfied for all possible test functions, it must vanish for those functions satisfying $v(1) = 0$. Thus, we conclude that (2.2.1a) is satisfied. Similarly, by considering test functions v that are nonzero in just a small neighborhood of $x = 1$, we conclude that the boundary condition (2.3.1b) must be satisfied. Since (2.3.6) must be satisfied for *all* test functions v, the solution u must satisfy (2.2.1a) in the interior of the domain and (2.3.1b) at $x = 1$. □

Neumann boundary conditions, or other boundary conditions prescribing derivatives (*cf.* Problem 2 at the end of this section), are called *natural boundary conditions* because they follow directly from the variational principle and are not explicitly imposed. Essential boundary conditions constrain the space of functions that may be used as trial or comparison functions. Natural boundary conditions impose no constraints on the function spaces but, rather, alter the variational principle.

Problems

1. Prove the remainder of Theorem 2.3.1, *i.e.*, show that functions that satisfy (2.3.3) also minimize (2.3.4).

2. Show that the Galerkin form (2.2.1a) with the Robin boundary conditions

   $$p(0)u'(0) + \gamma_0 u(0) = \alpha_0, \qquad p(1)u'(1) + \gamma_1 u(1) = \alpha_1$$

   is: determine $u \in H^1$ satisfying

   $$A(v,u) = (v,f) + v(1)(\alpha_1 - \gamma_1 u(1)) - v(0)(\alpha_0 - \gamma_0 u(0)), \qquad \forall v \in H^1.$$

   Also show that the function $w \in H^1$ that minimizes

   $$I[w] = A(w,w) - 2(w,f) - 2\alpha_1 w(1) + \gamma_1 w(1)^2 + 2\alpha_0 w(0) - \gamma_0 w(0)^2$$

   is u, the solution of the Galerkin problem.

3. Construct the Galerkin form of (2.2.1) when

$$p(x) = \begin{cases} 1, & \text{if } 0 \le x < 1/2 \\ 2, & \text{if } 1/2 \le x \le 1 \end{cases} .$$

Such a situation can arise in a steady heat-conduction problem when the medium is made of two different materials that are joined at $x = 1/2$. What conditions must u satisfy at $x = 1/2$?

2.4 Piecewise Lagrange Polynomials

The finite element method is not limited to piecewise-linear polynomial approximations and its extention to higher-degree polynomials is straight forward. There is, however, a question of the best basis. Many possibilities are available from design and approximation theory. Of these, splines and Hermite approximations [5] are generally not used because they offer more smoothness and/or a larger support than needed or desired. Lagrange interpolation [2] and a hierarchical approximation in the spirit of Newton's divided-difference polynomials will be our choices. The piecewise-linear "hat" function

$$\phi_j(x) = \begin{cases} \frac{x - x_{j-1}}{x_j - x_{j-1}}, & \text{if } x_{j-1} \le x < x_j \\ \frac{x_{j+1} - x}{x_{j+1} - x_j}, & \text{if } x_j \le x < x_{j+1} \\ 0, & \text{otherwise} \end{cases} \tag{2.4.1a}$$

on the mesh

$$x_0 < x_1 < \ldots < x_N \tag{2.4.1b}$$

is a member of both classes. It has two desirable properties: (i) $\phi_j(x)$ is unity at node j and vanishes at all other nodes and (ii) ϕ_j is only nonzero on those elements containing node j. The first property simplifies the determination of solutions at nodes while the second simplifies the solution of the algebraic system that results from the finite element discretization. The Lagrangian basis maintains these properties with increasing polynomial degree. Hierarchical approximations, on the other hand, maintain only the second property. They are constructed by adding high-degree corrections to lower-degree members of the series.

We will examine Lagrange bases in this section, beginning with the quadratic polynomial basis. These are constructed by adding an extra node $x_{j-1/2}$ at the midpoint of each element $[x_{j-1}, x_j]$, $j = 1, 2, \ldots, N$ (Figure 2.4.1). As with the piecewise-linear basis (2.4.1a), one basis function is associated with each node. Those associated with vertices are

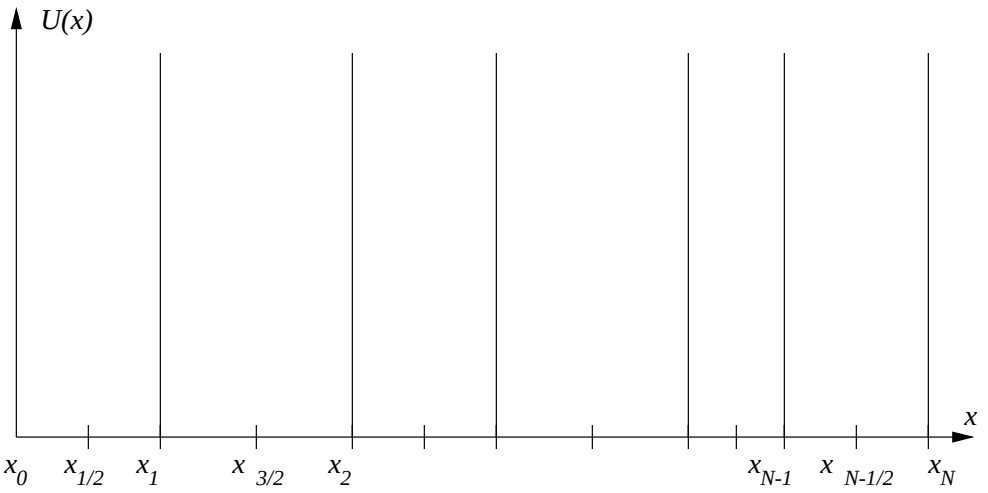

Figure 2.4.1: Finite element mesh for piecewise-quadratic Lagrange polynomial approximations.

$$\phi_j(x) = \begin{cases} 1 + 3(\frac{x-x_j}{h_j}) + 2(\frac{x-x_j}{h_j})^2, & \text{if } x_{j-1} \le x < x_j \\ 1 - 3(\frac{x-x_j}{h_{j+1}}) + 2(\frac{x-x_j}{h_{j+1}})^2, & \text{if } x_j \le x < x_{j+1} \\ 0, & \text{otherwise} \end{cases}, \qquad j = 0, 1, \ldots, N, \tag{2.4.2a}$$

and those associated with element midpoints are

$$\phi_{j-1/2}(x) = \begin{cases} 1 - 4(\frac{x-x_{j-1/2}}{h_j})^2, & \text{if } x_{j-1} \le x < x_j \\ 0, & \text{otherwise} \end{cases}, \qquad j = 1, 2, \ldots, N. \tag{2.4.2b}$$

Here

$$h_j = x_j - x_{j-1}, \qquad j = 1, 2, \ldots, N. \tag{2.4.2c}$$

These functions are shown in Figure 2.4.2. Their construction (to be described) invovles satsifying

$$\phi_j(x_k) = \begin{cases} 1, & \text{if } j = k \\ 0, & \text{otherwise} \end{cases}, \qquad j, k = 0, 1/2, 1, \ldots, N-1, N-1/2, N. \tag{2.4.3}$$

Basis functions associated with a vertex are nonzero on at most two elements and those associated with an element midpoint are nonzero on only one element. Thus, as noted, the Lagrange basis function ϕ_j is nonzero only on elements containing node j. The functions (2.4.2a,b) are quadratic polynomials on each element. Their construction and trivial extension to other finite elements guarantees that they are continuous over the entire mesh and, like (2.4.1), are members of H^1.

The finite element trial function $U(x)$ is a linear combination of (2.4.2a,b) over the vertices and element midpoints of the mesh that may be written as

$$U(x) = \sum_{j=0}^{N} c_j \phi_j(x) + \sum_{j=1}^{N} c_{j-1/2} \phi_{j-1/2}(x) = \sum_{j=0}^{2N} c_{j/2} \phi_{j/2}(x). \tag{2.4.4}$$

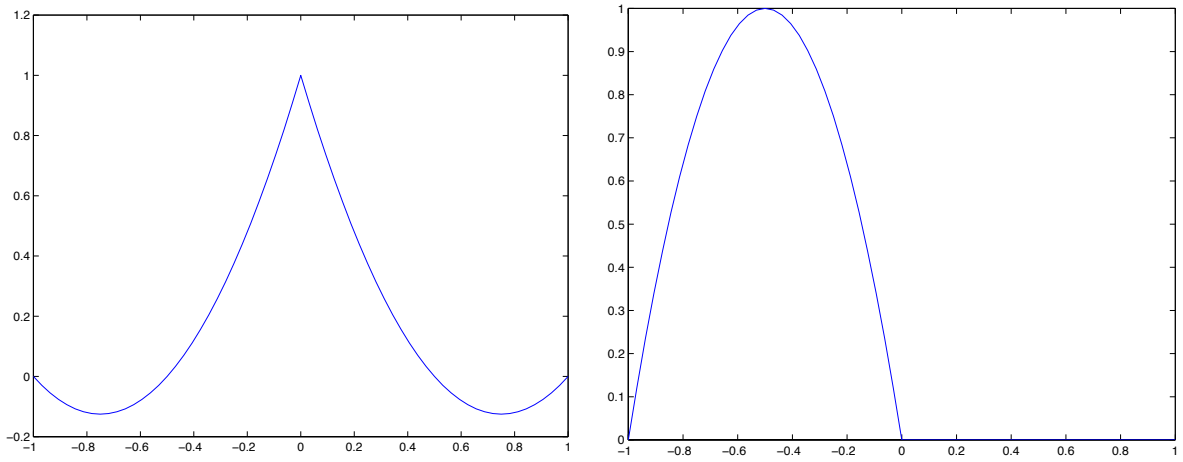

Figure 2.4.2: Piecewise-quadratic Lagrange basis functions for a vertex at $x = 0$ (left) and an element midpoint at $x = -0.5$ (right). When comparing with (2.4.2), set $x_{j-1} = -1$, $x_{j-1/2} = -0.5$, $x_j = 0$, $x_{j+1/2} = 0.5$, and $x_{j+1} = 1$.

Using (2.4.3), we see that $U(x_k) = c_k$, $k = 0, 1/2, 1, \ldots, N - 1/2, N$.

Cubic, quartic, *etc.* Lagrangian polynomials are generated by adding nodes to element interiors. However, prior to constructing them, let's introduce some terminology and simplify the node numbering to better suit our task. Finite element bases are constructed implicitly in an element-by-element manner in terms of shape functions. A *shape function* is the restriction of a basis function to an element. Thus, for the piecewise-quadratic Lagrange polynomial, there are three nontrivial shape functions on the element $\Omega_j := [x_{j-1}, x_j]$:

- the right portion of $\phi_{j-1}(x)$

$$N_{j-1,j}(x) = 1 - 3(\frac{x - x_{j-1}}{h_j}) + 2(\frac{x - x_{j-1}}{h_j})^2, \tag{2.4.5a}$$

- $\phi_{j-1/2}(x)$

$$N_{j-1/2,j}(x) = 1 - 4(\frac{x - x_{j-1/2}}{h_j})^2, \tag{2.4.5b}$$

- and the left portion of $\phi_j(x)$

$$N_{j,j}(x) = 1 + 3(\frac{x - x_j}{h_j}) + 2(\frac{x - x_j}{h_j})^2, \qquad x \in \Omega_j, \tag{2.4.5c}$$

(Figure 2.4.3). In these equations, $N_{k,j}$ is the shape function associated with node k, $k = j - 1, j - 1/2, j$, of element j (the subinterval Ω_j). We may use (2.4.4) and (2.4.5) to write the restriction of $U(x)$ to Ω_j as

$$U(x) = c_{j-1}N_{j-1,j} + c_{j-1/2}N_{j-1/2,j} + c_jN_{j,j}, \qquad x \in \Omega_j.$$

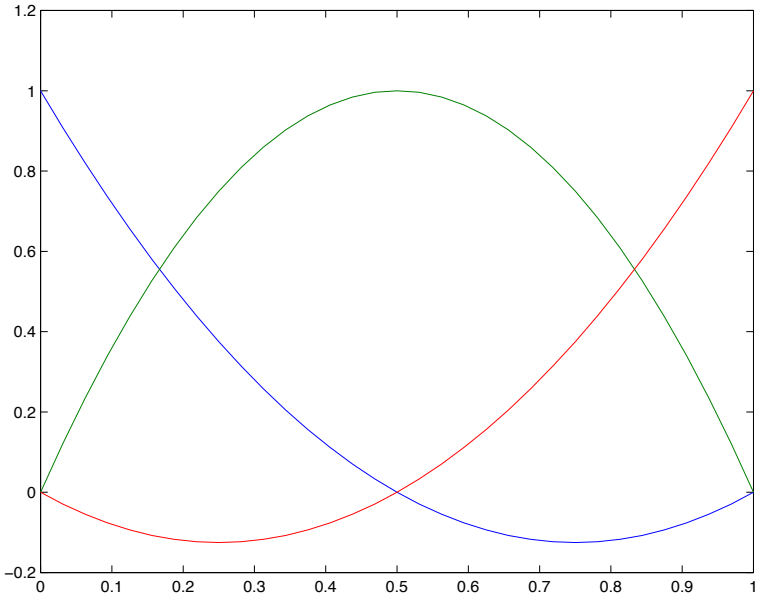

Figure 2.4.3: The three quadratic Lagrangian shape functions on the element $[x_{j-1}, x_j]$. When comparing with (2.4.5), set $x_{j-1} = 0$, $x_{j-1/2} = 0.5$, and $x_j = 1$.

More generally, we will associate the shape function $N_{k,e}(x)$ with *mesh entity* k of element e. At present, the only mesh entities that we know of are vertices and (nodes on) elements; however, edges and faces will be introduced in two and three dimensions. The key construction concept is that the shape function $N_{k,e}(x)$ is

1. nonzero only on element e and

2. nonzero only if mesh entity k belongs to element e.

A one-dimensional Lagrange polynomial shape function of degree p is constructed on an element e using two vertex nodes and $p-1$ nodes interior to the element. The generation of shape functions is straight forward, but it is customary and convenient to do this on a "canonical element." Thus, we map an arbitrary element $\Omega_e = [x_{j-1}, x_j]$ onto $-1 \le \xi \le 1$ by the linear transformation

$$x(\xi) = \frac{1-\xi}{2} x_{j-1} + \frac{1+\xi}{2} x_j, \qquad \xi \in [-1, 1]. \tag{2.4.6}$$

Nodes on the canonical element are numbered according to some simple scheme, *i.e.*, 0 to p with $\xi_0 = -1$, $\xi_p = 1$, and $0 < \xi_1 < \xi_2 < \ldots < \xi_{p-1} < 1$ (Figure 2.4.4). These are mapped to the actual physical nodes $x_{j-1}, x_{j-1+1/p}, \ldots, x_j$ on Ω_e using (2.4.6). Thus,

$$x_{j-1+i/p} = \frac{1-\xi_i}{2} x_{j-1} + \frac{1+\xi_i}{2} x_j, \qquad i = 0, 1, \ldots, p.$$

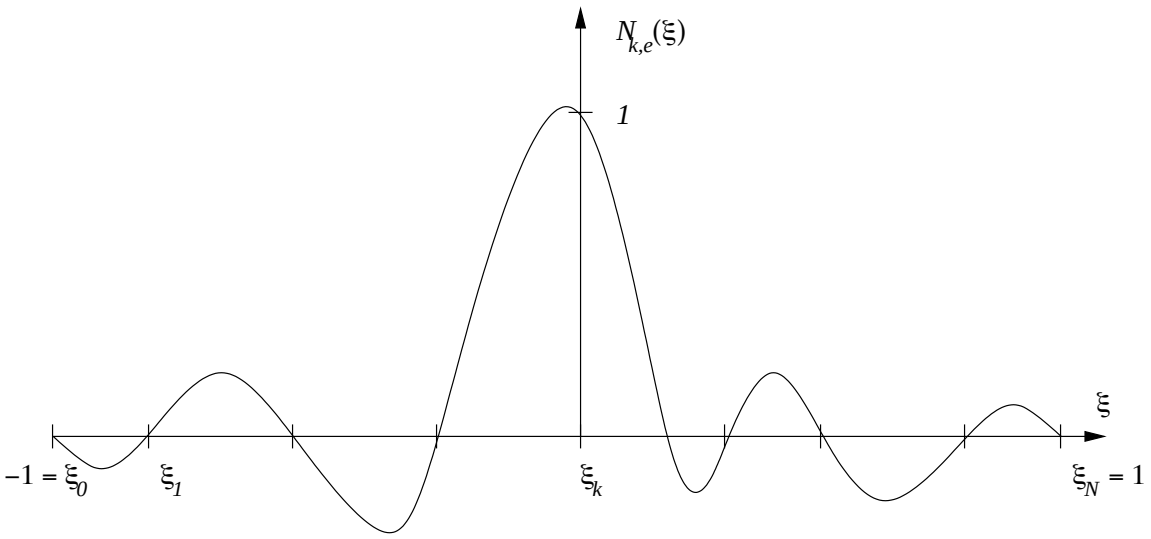

Figure 2.4.4: An element e used to construct a p *th*-degree Lagrangian shape function and the shape function $N_{k,e}(x)$ associated with node k.

The Lagrangian shape function $N_{k,e}(\xi)$ of degree p has a unit value at node k of element e and vanishes at all other nodes; thus,

$$N_{k,e}(\xi_l) = \delta_{kl} = \begin{cases} 1, & \text{if } k = l \\ 0, & \text{otherwise} \end{cases}, \qquad l = 0, 1, \ldots, p. \tag{2.4.7a}$$

It is extended trivially when $\xi \notin [-1, 1]$. The conditions expressed by (2.4.7a) imply that

$$N_{k,e}(\xi) = \prod_{l=0,\ l\neq k}^{p} \frac{\xi - \xi_l}{\xi_k - \xi_l} = \frac{(\xi - \xi_0)(\xi - \xi_1)\ldots(\xi - \xi_{k-1})(\xi - \xi_{k+1})\ldots(\xi - \xi_p)}{(\xi_k - \xi_0)(\xi_k - \xi_1)\ldots(\xi_k - \xi_{k-1})(\xi_k - \xi_{k+1})\ldots(\xi_k - \xi_p)}. \tag{2.4.7b}$$

We easily check that $N_{k,e}$ (i) is a polynomial of degree p in ξ and (ii) it satisfies conditions (2.4.7a). It is shown in Figure 2.4.4. Written in terms of shape function, the restriction of U to the canonical element is

$$U(\xi) = \sum_{k=0}^{p} c_k N_{k,e}(\xi). \tag{2.4.8}$$

Example 2.4.1. Let us construct the quadratic Lagrange shape functions on the canonical element by setting $p = 2$ in (2.4.7b) to obtain

$$N_{0,e}(\xi) = \frac{(\xi - \xi_1)(\xi - \xi_2)}{(\xi_0 - \xi_1)(\xi_0 - \xi_2)}, \qquad N_{1,e}(\xi) = \frac{(\xi - \xi_0)(\xi - \xi_2)}{(\xi_1 - \xi_0)(\xi_1 - \xi_2)},$$

$$N_{2,e}(\xi) = \frac{(\xi - \xi_0)(\xi - \xi_1)}{(\xi_2 - \xi_0)(\xi_2 - \xi_1)}.$$

Setting $\xi_0 = -1$, $\xi_1 = 0$, and $\xi_2 = 1$ yields

$$N_{0,e}(\xi) = \frac{\xi(\xi-1)}{2}, \qquad N_{1,e}(\xi) = (1-\xi^2), \qquad N_{2,e}(\xi) = \frac{(\xi+1)\xi}{2}. \tag{2.4.9}$$

These may easily be shown to be identical to (2.4.2) by using the transformation (2.4.6) (see Problem 1 at the end of this section).

Example 2.4.2. Setting $p = 1$ in (2.4.7b), we obtain the linear shape functions on the canonical element as

$$N_{0,e} = \frac{1-\xi}{2}, \qquad N_{1,e} = \frac{1+\xi}{2}. \tag{2.4.10}$$

The two nodes needed for these shape functions are at the vertices $\xi_0 = -1$ and $\xi_1 = 1$. Using the transformation (2.4.6), these yield the two pieces of the hat function (2.4.1a). We also note that these shape functions were used in the linear coordinate transformation (2.4.6). This will arise again in Chapter 5.

Problems

1. Show the the quadratic Lagrange shape functions (2.4.9) on the canonical $[-1, 1]$ element transform to those on the physical element (2.4.2) upon use of (2.4.6)

2. Construct the shape functions for a cubic Lagrange polynomial from the general formula (2.4.7) by using two vertex nodes and two interior nodes equally spaced on the canonical $[-1, 1]$ element. Sketch the shape functions. Write the basis functions for a vertex and an interior node.

2.5 Hierarchical Bases

With a hierarchical polynomial representation the basis of degree $p+1$ is obtained as a correction to that of degree p. Thus, the entire basis need not be reconstructed when increasing the polynomial degree. With finite element methods, they produce algebraic systems that are less susceptible to round-off error accumulation at high order than those produced by a Lagrange basis.

With the linear hierarchical basis being the usual hat functions (2.4.1), let us begin with the piecewise-quadratic hierarchical polynomial. The restriction of this function to element $\Omega_e = [x_{j-1}, x_j]$ has the form

$$U^2(x) = U^1(x) + c_{j-1/2} N^2_{j-1/2,e}(x), \qquad x \in \Omega_e, \tag{2.5.1a}$$

where $U^1(x)$ is the piecewise-linear finite element approximation on Ω_e

$$U^1(x) = c_{j-1} N^1_{j-1,e}(x) + c_j N^1_{j,e}(x). \tag{2.5.1b}$$

Superscripts have been added to U and $N_{j,e}$ to identify their polynomial degree. Thus,

$$N^1_{j-1,e}(x) = \begin{cases} \frac{x_j - x}{h_j}, & \text{if } x \in \Omega_e \\ 0, & \text{otherwise} \end{cases} , \tag{2.5.1c}$$

$$N^1_{j,e}(x) = \begin{cases} \frac{x - x_{j-1}}{h_j}, & \text{if } x \in \Omega_e \\ 0, & \text{otherwise} \end{cases} \tag{2.5.1d}$$

are the usual hat function (2.4.1) associated with a piecewise-linear approximation $U^1(x)$. The quadratic correction $N^2_{j-1/2,e}(x)$ is required to (i) be a quadratic polynomial, (ii) vanish when $x \notin \Omega_e$, and (iii) be continuous. These conditions imply that $N^2_{j-1/2,e}$ is proportional to the quadratic Lagrange shape function (2.4.5b) and we will take it to be identical; thus,

$$N^2_{j-1/2,e}(x) = \begin{cases} 1 - 4(\frac{x - x_{j-1/2}}{h_j})^2, & \text{if } x \in \Omega_e \\ 0, & \text{otherwise} \end{cases} . \tag{2.5.1e}$$

The normalization $N^2_{j-1/2,e}(x_{j-1/2}) = 1$ is not necessary, but seems convenient.

Like the quadratic Lagrange approximation, the quadratic hierarchical polynomial has three nontrivial shape functions per element; however, two of them are linear and only one is quadratic (Figure 2.5.1). The basis, however, still spans quadratic polynomials. Examining (2.5.1), we see that $c_{j-1} = U(x_{j-1})$ and $c_j = U(x_j)$; however,

$$U(x_{j-1/2}) = \frac{c_{j-1} + c_j}{2} + c_{j-1/2}.$$

Differentiating (2.5.1a) twice with respect to x gives an interpretation to $c_{j-1/2}$ as

$$c_{j-1/2} = -\frac{h^2}{8} U''(x_{j-1/2}).$$

This interpretation may be useful but is not necessary.

A basis may be constructed from the shape functions in the manner described for Lagrange polynomials. With a mesh having the structure used for the piecewise-quadratic Lagrange polynomials (Figure 2.4.1), the piecewise-quadratic hierarchical functions have the form

$$U(x) = \sum_{j=0}^{N} c_j \phi^1_j(x) + \sum_{j=1}^{N} c_{j-1/2} \phi^2_{j-1/2}(x) \tag{2.5.2}$$

where $\phi^1_j(x)$ is the hat function basis (2.4.1a) and $\phi^2_j(x) = N^2_{j,e}(x)$.

Higher-degree hierarchical polynomials are obtained by adding more correction terms to the lower-degree polynomials. It is convenient to construct and display these polynomials on the canonical $[-1, 1]$ element used in Section 2.4. The linear transformation

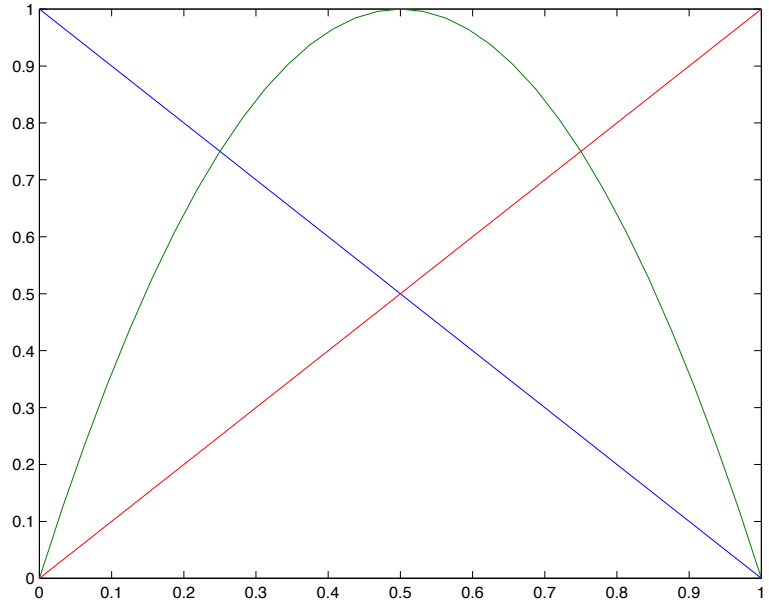

Figure 2.5.1: Quadratic hierarchical shape on $[x_{j-1}, x_j]$. When comparing with (2.5.1), set $x_{j-1} = 0$ and $x_j = 1$.

(2.4.6) is again used to map an arbitrary element $[x_{j-1}, x_j]$ onto $-1 \le \xi \le 1$. The vertex nodes at $\xi = -1$ and 1 are associated with the linear shape functions and, for simplicity, we will index them as -1 and 1. The remaining $p-1$ shape functions are on the element interior. They need not be associated with any nodes but, for convenience, we will associate all of them with a single node indexed by 0 at the center ($\xi = 0$) of the element. The restriction of the finite element solution $U(\xi)$ to the canonical element has the form

$$U(\xi) = c_{-1}N^1_{-1}(\xi) + c_1 N^1_1(\xi) + \sum_{i=2}^{p} c_i N^i_0(\xi), \qquad \xi \in [-1, 1]. \tag{2.5.3}$$

(We have dropped the elemental index e on $N^i_{j,e}$ since we are only concerned with approximations on the canonical element.) The vertex shape functions N^1_{-1} and N^1_1 are the hat function segments (2.4.10) on the canonical element

$$N^1_{-1}(\xi) = \frac{1-\xi}{2}, \qquad N^1_1(\xi) = \frac{1+\xi}{2}, \qquad \xi \in [-1, 1]. \tag{2.5.4}$$

Once again, the higher-degree shape functions $N^i_0(\xi)$, $i = 2, 3, \ldots, p$, are required to have the proper degree and vanish at the element's ends $\xi = -1, 1$ to maintain continuity. Any normalization is arbitrary and may be chosen to satisfy a specified condition, *e.g.*, $N^2_0(0) = 1$. We use a normalization of Szabó and Babuška [7] which relies on Legendre polynomials. The Legendre polynomial $P_i(\xi)$, $i \ge 0$, is a polynomial of degree i in ξ satisfying [1]:

1. the differential equation

$$(1-\xi^2)P_i'' - 2\xi P_i' + i(i+1)P_i = 0, \qquad -1 < \xi < 1, \qquad i \geq 0; \tag{2.5.5a}$$

2. the normalization

$$P_i(1) = 1, \qquad i \geq 0; \tag{2.5.5b}$$

3. the orthogonality relation

$$\int_{-1}^{1} P_i(\xi)P_j(\xi)d\xi = \frac{2}{2i+1}\begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}; \tag{2.5.5c}$$

4. the symmetry condition

$$P_i(-\xi) = (-1)^i P_i(\xi), \qquad i \geq 0; \tag{2.5.5d}$$

5. the recurrence relation

$$(i+1)P_{i+1}(\xi) = (2i+1)\xi P_i(\xi) - iP_{i-1}(\xi), \qquad i \geq 1; \tag{2.5.5e}$$

and

6. the differentiation formula

$$P_{i+1}'(\xi) = (2i+1)P_i(\xi) + P_{i-1}'(\xi), \qquad i \geq 1. \tag{2.5.5f}$$

The first six Legendre polynomials are

$$\begin{aligned} P_0(\xi) &= 1, & P_1(\xi) &= \xi, \\ P_2(\xi) &= \frac{3\xi^2-1}{2}, & P_3(\xi) &= \frac{5\xi^3-3\xi}{2}, \\ P_4(\xi) &= \frac{35\xi^4-30\xi^2+3}{2}, & P_5(\xi) &= \frac{63\xi^5-70\xi^3+15\xi}{8}. \end{aligned} \tag{2.5.6}$$

With these preliminaries, we define the shape functions

$$N_0^i(\xi) = \sqrt{\frac{2i-1}{2}}\int_{-1}^{\xi} P_{i-1}(\sigma)d\sigma, \qquad i \geq 2. \tag{2.5.7a}$$

Using (2.5.5d,f), we readily show that

$$N_0^i(\xi) = \frac{P_i(\xi) - P_{i-2}(\xi)}{\sqrt{2(2i-1)}}, \qquad i \geq 2. \tag{2.5.7b}$$

Use of the normalization and symmetry properties (2.5.5b,d) further reveal that

$$N_0^i(-1) = N_0^i(1) = 0, \qquad i \geq 2, \tag{2.5.7c}$$

and use of the orthogonality property (2.5.5c) indicates that

$$\int_{-1}^{1} \frac{dN_0^i(\xi)}{d\xi} \frac{dN_0^j(\xi)}{d\xi} d\xi = \delta_{ij}, \qquad i, j \geq 2. \tag{2.5.7d}$$

Substituting (2.5.6) into (2.5.7b) gives

$$\begin{aligned} N_0^2(\xi) &= \frac{3}{2\sqrt{6}}(\xi^2 - 1), \qquad N_0^3(\xi) = \frac{5}{2\sqrt{10}}\xi(\xi^2 - 1), \\ N_0^4(\xi) &= \frac{7}{8\sqrt{14}}(5\xi^4 - 6\xi^2 + 1), \qquad N_0^5(\xi) = \frac{9}{8\sqrt{18}}(7\xi^5 - 10\xi^3 + 3\xi). \end{aligned} \tag{2.5.8}$$

Shape functions $N_0^i(\xi)$, $i = 2, 3, \ldots, 6$, are shown in Figure 2.5.2.

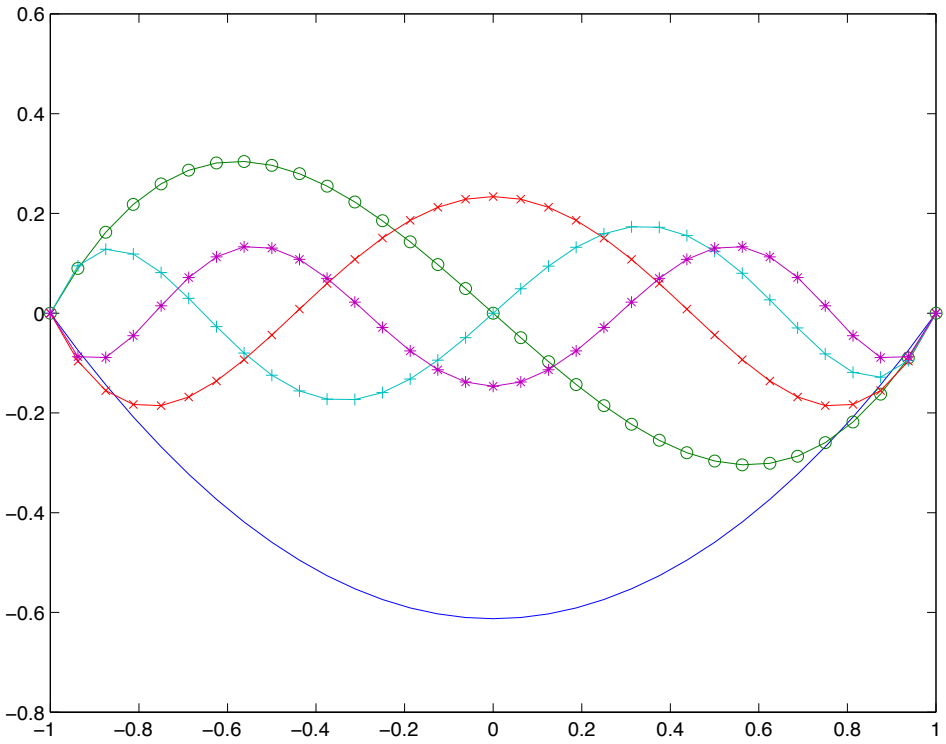

Figure 2.5.2: One-dimensional hierarchical shape functions of degrees 2 (solid), 3(∘), 4 (×), 5 (+), and 6 (*) on the canonical element $-1 \leq \xi \leq 1$.

The representation (2.5.3) with use of (2.5.5b,d) reveals that the parameters c_{-1} and c_1 correspond to the values of $U(-1)$ and $U(1)$, respectively; however, the remaining parameters c_i, $i \geq 2$, do not correspond to solution values. In particular, using (2.5.3),

(2.5.5d), and (2.5.7b) yields

$$U(0) = \frac{c_{-1} + c_1}{2} + \sum_{i=2,4}^{p} c_i N_0^i(0).$$

Hierarchical bases can be constructed so that c_i is proportional to $d^i U(0)/d\xi^i$, $i \geq 2$ (*cf.* [3], Section 2.8); however, the shape functions (2.5.8) based on Legendre polynomials reduce sensitivity of the basis to round-off error accumulation. This is very important when using high-order finite element approximations.

Example 2.5.1. Let us solve the two-point boundary value problem

$$-pu'' + qu = f(x), \qquad 0 < x < 1, \qquad u(0) = u(1) = 0, \tag{2.5.9}$$

using the finite element method with piecewise-quadratic hierarchical approximations. As in Chapter 1, we simplify matters by assuming that $p > 0$ and $q \geq 0$ are constants.

By now we are aware that the Galerkin form of this problem is given by (2.2.2). As in Chapter 1, introduce (*cf.* (1.3.9))

$$A_j^S(v, u) = \int_{x_{j-1}}^{x_j} pv'u'dx.$$

We use (2.4.6) to map $[x_{j-1}, x_j]$ to the canonical $[-1, 1]$ element as

$$A_j^S(v, u) = \frac{2}{h_j} \int_{-1}^{1} p \frac{dv}{d\xi} \frac{du}{d\xi} d\xi. \tag{2.5.10}$$

Using (2.5.3), we write the restriction of the piecewise-quadratic trial and text functions to $[x_{j-1}, x_j]$ as

$$U(\xi) = [c_{j-1}, c_j, c_{j-1/2}] \begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix}, \qquad V(\xi) = [d_{j-1}, d_j, d_{j-1/2}] \begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix}. \tag{2.5.11}$$

Substituting (2.5.11) into (2.5.10)

$$A_j^S(V, U) = [d_{j-1}, d_j, d_{j-1/2}] \mathbf{K}_j \begin{bmatrix} c_{j-1} \\ c_j \\ c_{j-1/2} \end{bmatrix} \tag{2.5.12a}$$

where $\mathbf{K}_j$ is the element stiffness matrix

$$\mathbf{K}_j = \frac{2p}{h_j} \int_{-1}^{1} \frac{d}{d\xi} \begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix} \frac{d}{d\xi} [N_{-1}^1, N_1^1, N_0^2] d\xi.$$

Substituting for the basis definitions (2.5.4, 2.5.8)

$$\mathbf{K}_j = \frac{2p}{h_j}\int_{-1}^{1}\begin{bmatrix} -1/2 \\ 1/2 \\ \xi\sqrt{\frac{3}{2}} \end{bmatrix}[-1/2, 1/2, \xi\sqrt{\frac{3}{2}}]d\xi.$$

Integrating

$$\mathbf{K}_j = \frac{2p}{h_j}\int_{-1}^{1}\begin{bmatrix} 1/4 & -1/4 & -\xi\sqrt{3/8} \\ -1/4 & 1/4 & \xi\sqrt{3/8} \\ -\xi\sqrt{3/8} & \xi\sqrt{3/8} & 3\xi^2/2 \end{bmatrix} d\xi = \frac{p}{h_j}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \qquad (2.5.12b)$$

The orthogonality relation (2.5.7d) has simplified the stiffness matrix by uncoupling the linear and quadratic modes.

In a similar manner,

$$A_j^M(V,U) = \int_{x_{j-1}}^{x_j} qVU dx = \frac{qh_j}{2}\int_{-1}^{1} VU d\xi. \qquad (2.5.13a)$$

Using (2.5.11)

$$A_j^M(V,U) = [d_{j-1}, d_j, d_{j-1/2}]\mathbf{M}_j\begin{bmatrix} c_{j-1} \\ c_j \\ c_{j-1/2} \end{bmatrix} \qquad (2.5.13b)$$

where, upon use of (2.5.4, 2.5.8), the element mass matrix $\mathbf{M}_j$ satisfies

$$\mathbf{M}_j = \frac{qh_j}{2}\int_{-1}^{1}\begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix}[N_{-1}^1, N_1^1, N_0^2]d\xi = \frac{qh_j}{6}\begin{bmatrix} 2 & 1 & -\sqrt{3/2} \\ 1 & 2 & -\sqrt{3/2} \\ -\sqrt{3/2} & -\sqrt{3/2} & 6/5 \end{bmatrix}. \qquad (2.5.13c)$$

The higher and lower order terms of the element mass matrix have not decoupled. Comparing (2.5.12b) and (2.5.13c) with the forms developed in Section 1.3 for piecewise-linear approximations, we see that the piecewise linear stiffness and mass matrices are contained as the upper 2×2 portions of these matrices. This will be the case for linear problems; thus, each higher-degree polynomial will add a "border" to the lower-degree stiffness and mass matrices.

Finally, consider

$$(V,f)_j = \int_{x_{j-1}}^{x_j} Vf dx = \frac{h_j}{2}\int_{-1}^{1} Vf d\xi. \qquad (2.5.14a)$$

Using (2.5.11)

$$(V,f)_j = [d_{j-1}, d_j, d_{j-1/2}]\mathbf{l}_j \qquad (2.5.14b)$$

where

$$\mathbf{l}_j = \frac{h_j}{2}\int_{-1}^{1}\begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix} f(x(\xi))d\xi. \tag{2.5.14c}$$

As in Section 1.3, we approximate $f(x)$ by piecewise-linear interpolation, which we write as

$$f(x) \approx N_{-1}^1(\xi)f_{j-1} + N_1^1(\xi)f_j$$

with $f_j := f(x_j)$. The manner of approximating $f(x)$ should clearly be related to the degree p and we will need a more careful analysis. Postponing this until Chapters 6 and 7, we have

$$\mathbf{l}_j = \frac{h_j}{2}\int_{-1}^{1}\begin{bmatrix} N_{-1}^1 \\ N_1^1 \\ N_0^2 \end{bmatrix}[N_{-1}^1, N_1^1]d\xi \begin{bmatrix} f_{j-1} \\ f_j \end{bmatrix} = \frac{h_j}{6}\begin{bmatrix} 2f_{j-1}+f_j \\ f_{j-1}+2f_j \\ -\sqrt{3/2}(f_{j-1}+f_j) \end{bmatrix} \tag{2.5.14d}$$

Using (2.2.2a) with (2.5.12a), (2.5.13a), and (2.5.14a), we see that assembly requires evluating the sum

$$\sum_{j=1}^{N}[A_j^S(V,U) + A_j^M(V,U) - (V,f)_j] = 0.$$

Following the strategy used for the piecewise-linear solution of Section 1.3, the local stiffness and mass matrices and load vectors are added into their proper locations in their global counterparts. Imposing the condition that the system be satisfied for all choices of d_j, $j = 1/2, 1, 3/2, \ldots, N-1$, yields the linear algebraic system

$$(\mathbf{K}+\mathbf{M})\mathbf{c} = \mathbf{l}. \tag{2.5.15}$$

The structure of the stiffness and mass matrices $\mathbf{K}$ and $\mathbf{M}$ and load vector $\mathbf{l}$ depend on the ordering of the unknowns $\mathbf{c}$ and virtual coordinates $\mathbf{d}$. One possibility is to order them by increasing index, *i.e.*,

$$\mathbf{c} = [c_{1/2}, c_1, c_{3/2}, c_2, \ldots, c_{N-1}, c_{N-1/2}]^T. \tag{2.5.16}$$

As with the piecewise-linear basis, we have assumed that the homogeneous boundary conditions have explicitly eliminated $c_0 = c_N = 0$. Assembly for this ordering is similar to the one used in Section 1.3 (*cf.* Problem 2 at the end of this section). This is a natural ordering and the one most used for this approximation; however, for variety, let us order the unknowns by listing the vertices first followed by those at element midpoints, *i.e.*,

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_L \\ \mathbf{c}_Q \end{bmatrix}, \qquad \mathbf{c}_L = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{N-1} \end{bmatrix}, \qquad \mathbf{c}_Q = \begin{bmatrix} c_{1/2} \\ c_{3/2} \\ \vdots \\ c_{N-1/2} \end{bmatrix}. \tag{2.5.17}$$

In this case, $\mathbf{K}$, $\mathbf{M}$, and $\mathbf{l}$ have a block structure and may be partitioned as

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_Q \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} \mathbf{M}_L & \mathbf{M}_{LQ} \\ \mathbf{M}_{LQ}^T & \mathbf{M}_Q \end{bmatrix}, \qquad \mathbf{l} = \begin{bmatrix} \mathbf{l}_L \\ \mathbf{l}_Q \end{bmatrix} \tag{2.5.18}$$

where, for uniform mesh spacing $h_j = h$, $j = 1, 2, \ldots, N$, these matrices are

$$\mathbf{K}_L = \frac{p}{h}\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}, \qquad \mathbf{K}_Q = \frac{p}{h}\begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & \ddots & & \\ & & & 2 & \\ & & & & 2 \end{bmatrix}, \tag{2.5.19}$$

$$\mathbf{M}_L = \frac{qh}{6}\begin{bmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{bmatrix}, \qquad \mathbf{M}_{LQ} = -\frac{qh}{6}\sqrt{\frac{3}{2}}\begin{bmatrix} 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & 1 & 1 & \\ & & & & 1 & 1 \end{bmatrix},$$

$$\mathbf{M}_Q = \frac{qh}{5}\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}, \tag{2.5.20}$$

$$\mathbf{l}_L = \frac{h}{6}\begin{bmatrix} f_0 + 4f_1 + f_2 \\ f_1 + 4f_2 + f_3 \\ \vdots \\ f_{N-2} + 4f_{N-1} + f_N \end{bmatrix}, \qquad \mathbf{l}_Q = -\frac{h}{\sqrt{24}}\begin{bmatrix} f_0 + f_1 \\ f_1 + f_2 \\ \vdots \\ f_{N-1} + f_N \end{bmatrix}. \tag{2.5.21}$$

With $N-1$ vertex unknowns $\mathbf{c}_L$ and N elemental unknowns $\mathbf{c}_Q$, the matrices $\mathbf{K}_L$ and $\mathbf{M}_L$ are $(N-1)\times(N-1)$, $\mathbf{K}_Q$ and $\mathbf{M}_Q$ are $N \times N$, and $\mathbf{M}_{LQ}$ is $(N-1)\times N$. Similarly, $\mathbf{l}_L$ and $\mathbf{l}_Q$ have dimension $N-1$ and N, respectively. The indicated ordering implies that the 3×3 element stiffness and mass matrices (2.5.12b) and (2.5.13c) for element j are added to rows and columns $j-1$, j, and $N-1+j$ of their global counterparts. The first row and column of the element stiffness and mass matrices are deleted when $j = 1$ to satisfy the left boundary condition. Likewise, the second row and column of these matrices are deleted when $j = N$ to satisfy the right boundary condition.

The structure of the system matrix $\mathbf{K} + \mathbf{M}$ is

$$\mathbf{K} + \mathbf{M} = \begin{bmatrix} \mathbf{K}_L + \mathbf{M}_L & \mathbf{M}_{LQ} \\ \mathbf{M}_{LQ}^T & \mathbf{K}_Q + \mathbf{M}_Q \end{bmatrix}. \tag{2.5.22}$$

The matrix $\mathbf{K}_L + \mathbf{M}_L$ is the same one used for the piecewise-linear solution of this problem in Section 1.3. Thus, an assembly and factorization of this matrix done during a prior piecewise-linear finite element analysis could be reused. A solution procedure using this factorization is presented as Problem 3 at the end of this section. Furthermore, if $q \equiv 0$ then $\mathbf{M}_{LQ} = 0$ (*cf.* (2.5.20b)) and the linear and quadratic portions of the system uncouple.

In Example 1.3.1, we solved (2.5.9) with $p = 1$, $q = 1$, and $f(x) = x$ using piecewise-linear finite elements. Let us solve this problem again using piecewise-quadratic hierarchical approximations and compare the results. Recall that the exact solution of this problem is

$$u(x) = x - \frac{\sinh x}{\sinh 1}.$$

Results for the error in the L^2 norm are shown in Table 2.5.1 for solutions obtained with piecewise-linear and quadratic approximations. The results indicate that solutions with piecewise-quadratic approximations are converging as $O(h^3)$ as opposed to $O(h^2)$ for piecewise-linear approximations. Subsequently, we shall show that smooth solutions generally converge as $O(h^{p+1})$ in the L^2 norm and as $O(h^p)$ in the strain energy (or H^1) norm.

N	Linear			Quadratic		
	DOF	$\|\|e\|\|_0$	$\|\|e\|\|_0/h^2$	DOF	$\|\|e\|\|_0$	$\|\|e\|\|_0/h^3$
4	3	0.265(-2)	0.425(-1)	7	0.126(-3)	0.807(-2)
8	7	0.656(-3)	0.426(-1)	15	0.158(-4)	0.809(-2)
16	15	0.167(-3)	0.427(-1)	31	0.198(-5)	0.809(-2)
32	31	0.417(-4)	0.427(-1)			

Table 2.5.1: Errors in L^2 and degrees of freedom (DOF) for piecewise-linear and piecewse-quadratic solutions of Example 2.5.1.

The number of elements N is not the only measure of computational complexity. With higher-order methods, the number of unknowns (degrees of freedom) provides a better index. Since the piecewise-quadratic solution has approximately twice the number of unknowns of the linear solution, we should compare the linear solution with spacing h and the quadratic solution with spacing $2h$. Even with this analysis, the superiority of the higher-order method in Table 2.5.1 is clear.

Problems

1. Consider the approximation in strain energy of a given function $u(\xi)$, $-1 < \xi < 1$, by a polynomial $U(\xi)$ in the hierarchical form (2.5.3). The problem consists of

determining $U(\xi)$ as the solution of the Galerkin problem

$$A(V,U) = A(V,u), \qquad \forall V \in S^p,$$

where S^p is a space of $p\,th$-degree polynomials on $[-1,1]$. For simplicity, let us take the strain energy as

$$A(v,u) = \int_{-1}^{1} v_\xi u_\xi d\xi.$$

With $c_{-1} = u(-1)$ and $c_1 = u(1)$, find expressions for determining the remaining coefficients c_i, $i = 2,3,\ldots,p$, so that the approximation satisfies the specified Galerkin projection.

2. Show how to generate the global stiffness and mass matrices and load vector for Example 2.5.1 when the equations and unknowns are written in order of increasing index (2.5.16).

3. Suppose $\mathbf{K}_L + \mathbf{M}_L$ have been assembled and factored by Gaussian elimination as part of a finite element analysis with piecewise-linear approximations. Devise an algorithm to solve (2.5.15) for $\mathbf{c}_L$ and $\mathbf{c}_Q$ that utilizes the given factorization.

2.6 Interpolation Errors

Errors of finite element solutions can be measured in several norms. We have already introduced pointwise and global metrics. In this introductory section on error analysis, we'll define some basic principles and study interpolation errors. As we shall see shortly, errors in interpolating a function u by a piecewise polynomial approximation U will provide bounds on the errors of finite element solutions.

Once again, consider a Galerkin problem for a second-order differential equation: find $u \in H_0^1$ such that

$$A(v,u) = (v,f), \qquad \forall v \in H_0^1. \tag{2.6.1}$$

Also consider its finite element counterpart: find $U \in S_0^N$ such that

$$A(V,U) = (V,f), \qquad \forall V \in S_0^N. \tag{2.6.2}$$

Let the approximating space $S_0^N \subset H_0^1$ consist of piecewise-polynomials of degree p on N-element meshes. We begin with two fundamental results regarding Galerkin's method and finite element approximations.

Theorem 2.6.1. *Let $u \in H_0^1$ and $U \in S_0^N \subset H_0^1$ satisfy (2.6.1) and (2.6.2), respectively, then*

$$A(V, u - U) = 0, \qquad \forall V \in S_0^N. \tag{2.6.3}$$

Proof. Since $V \in S_0^N$ it also belongs to H_0^1. Thus, it may be used to replace v in (2.6.1). Doing this and subtracting (2.6.2) yields the result. □

We shall subsequently show that the strain energy furnishes an inner product. With this interpretation, we may regard (2.6.3) as an orthogonality condition in a "strain energy space" where $A(v, u)$ is an inner product and $\sqrt{A(u, u)}$ is a norm. Thus, the finite element solution error

$$e(x) := u(x) - U(x) \tag{2.6.4}$$

is orthogonal in strain energy to all functions V in the subspace S_0^N. We use this orthogonality to show that solutions obtained by Galerkin's method are optimal in the sence of minimizing the error in strain energy.

Theorem 2.6.2. *Under the conditions of Theorem 2.6.1,*

$$A(u - U, u - U) = \min_{V \in S_0^N} A(u - V, u - V). \tag{2.6.5}$$

Proof. Consider

$$A(u - U, u - U) = A(u, u) - 2A(u, U) + A(U, U).$$

Use (2.6.3) with V replaced by U to write this as

$$A(u - U, u - U) = A(u, u) - 2A(u, U) + A(U, U) + 2A(u - U, U)$$

or

$$A(u - U, u - U) = A(u, u) - A(U, U).$$

Again, using (2.6.3) for any $V \in S_0^N$

$$A(u - U, u - U) = A(u, u) - A(U, U) + A(V, V) - A(V, V) - 2A(u - U, V)$$

or

$$A(u - U, u - U) = A(u - V, u - V) - A(U - V, U - V).$$

Since the last term on the right is non-negative, we can drop it to obtain

$$A(u - U, u - U) \leq A(u - V, u - V), \qquad \forall V \in S_0^N.$$

We see that equality is attained when $V = U$ and, thus, (2.6.5) is established. □

With optimality of Galerkin's method, we may obtain estimates of finite element discretization errors by bounding the right side of (2.6.5) for particular choices of V. Convenient bounds are obtained by selecting V to be an interpolant of the exact solution u. Bounds furnished in this manner generally provide the exact order of convergence in the mesh spacing h. Furthermore, results similar to (2.6.5) may be obtained in other norms. They are rarely as precise as those in strain energy and typically indicate that the finite element solution differs by no more than a constant from the optimal solution in the considered norm.

Thus, we will study the errors associated with interpolation problems. This can be done either on a physical or a canonical element, but we will proceed using a canonical element since we constructed shape functions in this manner. For our present purposes, we regard $u(\xi)$ as a known function that is interpolated by a p *th*-degree polynomial $U(\xi)$ on the canonical element $[-1, 1]$. Any form of the interpolating polynomial may be used. We use the Lagrange form (2.4.8), where

$$U(\xi) = \sum_{k=0}^{p} c_k N_k(\xi) \tag{2.6.6}$$

with $N_k(\xi)$ given by (2.4.7b). (We have omitted the elemental index e on N_k for clarity since we are concerned with one element.) An analysis of interpolation errors whith hierarchical shape functions may also be done (*cf.* Problem 1 at the end of this section). Although the Lagrangian and hierarchical shape functions differ, the resulting interpolation polynomials $U(\xi)$ and their errors are the same since the interpolation problem has a unique solution [2, 6].

Selecting $p+1$ distinct points $xi_i \in [-1, 1]$, $i = 0, 1, \ldots, p$, the interpolation conditions are

$$U(\xi_i) = u(\xi_i) := u_i = c_i, \qquad j = 0, 1, \ldots, p, \tag{2.6.7}$$

where the rightmost condition follows from (2.4.7a).

There are many estimates of pointwise interpolation errors. Here is a typical result.

Theorem 2.6.3. *Let $u(\xi) \in C^{p+1}[-1, 1]$ then, for each $\xi \in [-1, 1]$, there exists a point $\zeta(\xi) \in (-1, 1)$ such that the error in interpolating $u(\xi)$ by a p th-degree polynomial $U(\xi)$ is*

$$e(\xi) = \frac{u^{(p+1)}(\zeta)}{(p+1)!} \prod_{i=0}^{p} (\xi - \xi_i). \tag{2.6.8}$$

Proof. Although the proof is not difficult, we'll just sketch the essential details. A complete analysis is given in numerical analysis texts such as Burden and Faires [2], Chapter 3, and Isaacson and Keller [6], Chapter 5.

Since

$$e(\xi_0) = e(\xi_1) = \ldots = e(\xi_p) = 0$$

the error must have the form

$$e(\xi) = g(\xi) \prod_{i=0}^{p} (\xi - \xi_i).$$

The error in interpolating a polynomial of degree p or less is zero; thus, $g(\xi)$ must be proportional to $u^{(p+1)}$. We may use a Taylor's series argument to infer the existence of $\zeta(\xi) \in (-1, 1)$ and

$$e(\xi) = C u^{(p+1)}(\zeta) \prod_{i=0}^{p} (\xi - \xi_i).$$

Selecting u to be a polynomial of degree $p+1$ and differentiating this expression $p+1$ times yields C as $1/(p+1)!$ and (2.6.8). □

The pointwise error (2.6.8) can be used to obtain a variety of global error estimates. Let us estimate the error when interpolating a smooth function $u(\xi)$ by a linear polynomial $U(\xi)$ at the vertices $\xi_0 = -1$ and $\xi_1 = 1$ of an element. Using (2.6.8) with $p = 1$ reveals

$$e(\xi) = \frac{u''(\zeta)}{2}(\xi + 1)(\xi - 1), \qquad \xi \in (-1, 1). \tag{2.6.9}$$

Thus,

$$|e(\xi)| \leq \frac{1}{2} \max_{-1 \leq \xi \leq 1} |u''(\xi)| \max_{-1 \leq \xi \leq 1} |\xi^2 - 1|.$$

Now,

$$\max_{-1 \leq \xi \leq 1} |\xi^2 - 1| = 1.$$

Thus,

$$|e(\xi)| \leq \frac{1}{2} \max_{-1 \leq \xi \leq 1} |u''(\xi)|.$$

Derivatives in this expression are taken with respect to ξ. In most cases, we would like results expressed in physical terms. The linear transformation (2.4.6) provides the necessary conversion from the canonical element to element j: $[x_{j-1}, x_j]$. Thus,

$$\frac{d^2 u(\xi)}{d\xi^2} = \frac{h_j^2}{4} \frac{d^2 u(\xi)}{dx^2}$$

with $h_j = x_j - x_{j-1}$. Letting

$$\|f(\cdot)\|_{\infty,j} := \max_{x_{j-1} \leq x \leq x_j} |f(x)| \tag{2.6.10}$$

denote the local "maximum norm" of $f(x)$ on $[x_{j-1}, x_j]$, we have

$$\|e(\cdot)\|_{\infty,j} \leq \frac{h_j^2}{8}\|u''(\cdot)\|_{\infty,j}. \tag{2.6.11}$$

(Arguments have been replaced by a $\cdot$ to emphasize that the actual norm doesn't depend on x.)

If $u(x)$ were interpolated by a piecewise-linear function $U(x)$ on N elements $[x_{j-1}, x_j]$, $j = 1, 2, \ldots, N$, then (2.6.11) could be used on each element to obtain an estimate of the maximum error as

$$\|e(\cdot)\|_\infty \leq \frac{h^2}{8}\|u''(\cdot)\|_\infty, \tag{2.6.12a}$$

where

$$\|f(\cdot)\|_\infty := \max_{1\leq j\leq N} \|f(\cdot)\|_{\infty,j}, \tag{2.6.12b}$$

and

$$h := \max_{1\leq j\leq N} (x_j - x_{j-1}). \tag{2.6.12c}$$

As a next step, let us use (2.6.9) and (2.4.6) to compute an error estimate in the L^2 norm; thus,

$$\int_{x_{j-1}}^{x_j} e^2(x)dx = \frac{h_j}{2}\int_{-1}^{1} [\frac{u''(\zeta(\xi))}{2}(\xi^2 - 1)]^2 d\xi.$$

Since $|\xi^2 - 1| \leq 1$, we have

$$\int_{x_{j-1}}^{x_j} e^2(x)dx \leq \frac{h_j}{8}\int_{-1}^{1} [u''(\zeta(\xi))]^2 d\xi.$$

Introduce the "local L^2 norm" of a function $f(x)$ as

$$\|f(\cdot)\|_{0,j} := \left(\int_{x_{j-1}}^{x_j} f^2(x)dx\right)^{1/2}. \tag{2.6.13}$$

Then,

$$\|e(\cdot)\|_{0,j}^2 \leq \frac{h_j}{8}\int_{-1}^{1} [u''(\zeta(\xi))]^2 d\xi,$$

It is tempting to replace the integral on the right side of our error estimate by $\|u''\|_{0,j}^2$. This is almost correct; however, $\zeta = \zeta(\xi)$. We would have to verify that ζ varies smoothly with ξ. Here, we will assume this to be the case and expand u'' using Taylor's theorem to obtain

$$u''(\zeta) = u''(\xi) + u'''(\theta)(\zeta - \xi) = u''(\xi) + O(|\zeta - \xi|), \qquad \theta \in (\xi, \zeta),$$

or

$$|u''(\zeta)| \le C|u''(\xi)|.$$

The constant C absorbs our careless treatment of the higher-order term in the Taylor's expansion. Thus, using (2.4.6), we have

$$\|e(\cdot)\|_{0,j}^2 \le C^2 \frac{h_j}{8} \int_{-1}^{1} [u''(\xi)]^2 d\xi = C^2 \frac{h_j^4}{64} \int_{x_{j-1}}^{x_j} [u''(x)]^2 dx,$$

where derivatives in the rightmost expression are with respect to x. Using (2.6.13)

$$\|e(\cdot)\|_{0,j}^2 \le C^2 \frac{h_j^4}{64} \|u''(\cdot)\|_{0,j}^2. \tag{2.6.14}$$

If we sum (2.6.14) over the N finite elements of the mesh and take a square root we obtain

$$\|e(\cdot)\|_0 \le Ch^2 \|u''(\cdot)\|_0, \tag{2.6.15a}$$

where

$$\|f(\cdot)\|_0^2 = \sum_{j=1}^{N} \|f(\cdot)\|_{0,j}^2. \tag{2.6.15b}$$

(The constant C in (2.6.15a) replaces the constant $C/8$ of (2.6.14), but we won't be precise about identifying different constants.)

With a goal of estimating the error in H^1, let us examine the error $u'(\xi) - U'(\xi)$. Differentiating (2.6.9) with respect to ξ

$$e'(\xi) = u''(\zeta)\xi + \frac{u'''(\zeta)}{2} \frac{d\zeta}{d\xi} (\xi^2 - 1).$$

Assuming that $d\zeta/d\xi$ is bounded, we use (2.6.13) and (2.4.6) to obtain

$$\|e'\|_{0,j}^2 = \int_{x_{j-1}}^{x_j} [\frac{de(x)}{dx}]^2 dx = \frac{2}{h_j} \int_{-1}^{1} [u''(\zeta)\xi + \frac{u'''(\zeta)}{2} \frac{d\zeta}{d\xi} (\xi^2 - 1)]^2 d\xi.$$

Following the arguments that led to (2.6.14), we find

$$\|e'(\cdot)\|_{0,j}^2 \le Ch_j^2 \|u''(\cdot)\|_{0,j}^2.$$

Summing over the N elements

$$\|e'(\cdot)\|_0^2 \le Ch^2 \|u''(\cdot)\|_0. \tag{2.6.16}$$

To obtain an error estimate in the H^1 norm, we combine (2.6.15a) and (2.6.16) to get

$$\|e(\cdot)\|_1 \leq Ch\|u''(\cdot)\|_0 \tag{2.6.17a}$$

where

$$\|f(\cdot)\|_1^2 := \sum_{j=1}^{N} [\|f'(\cdot)\|_{0,j}^2 + \|f(\cdot)\|_{0,j}^2]. \tag{2.6.17b}$$

The methodology developed above may be applied to estimate interpolation errors of higher-degree polynomial approximations. A typical result follows.

Theorem 2.6.4. *Introduce a mesh $a \leq x_0 < x_1 < \ldots < x_N \leq b$ such that $U(x)$ is a polynomial of degree p or less on every subinterval (x_{j-1}, x_j) and $U \in H^1(a,b)$. Let $U(x)$ interpolate $u(x) \in H^{p+1}[a,b]$ such that no error results when $u(x)$ is any polynomial of degree p or less. Then, there exists a constant $C_p > 0$, depending on p, such that*

$$\|u - U\|_0 \leq C_p h^{p+1} \|u^{(p+1)}\|_0 \tag{2.6.18a}$$

and

$$\|u - U\|_1 \leq C h_p^p \|u^{(p+1)}\|_0. \tag{2.6.18b}$$

where h satisfies (2.6.12c).

Proof. The analysis follows the one used for linear polynomials. □

Problems

1. Choose a hierarchical polynomial (2.5.3) on a canonical element $[-1,1]$ and show how to determine the coefficients c_j, $j = -1, 1, 2, \ldots, p$, to solve the interpolation problem (2.6.7).

Bibliography

[1] M. Abromowitz and I.A. Stegun. *Handbook of Mathematical Functions*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, Gathersburg, 1964.

[2] R.L. Burden and J.D. Faires. *Numerical Analysis*. PWS-Kent, Boston, fifth edition, 1993.

[3] G.F. Carey and J.T. Oden. *Finite Elements: A Second Course*, volume II. Prentice Hall, Englewood Cliffs, 1983.

[4] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Wiley-Interscience, New York, 1953.

[5] C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

[6] E. Isaacson and H.B. Keller. *Analysis of Numerical Methods*. John Wiley and Sons, New York, 1966.

[7] B. Szabó and I. Babuška. *Finite Element Analysis*. John Wiley and Sons, New York, 1991.