Hybrid and Mixed Finite Element Methods
Edited by S. N. Atluri, R. H. Gallagher, and O. C. Zienkiewicz

Chapter 22

Selection of Finite Element Methods

D. N. Arnold, I. Babuška, and J. Osborn

22.1 INTRODUCTION

The goal of engineering computations is to obtain quantitative information about engineering problems. This goal is usually achieved by the approximate solution of a mathematically formulated problem. Although a relevant mathematical formulation of the problem and its approximate solution are closely related (see, for example, [1, 2]), here we shall suppose that a mathematical formulation has already been determined and is amenable to an approximate treatment. We shall discuss a broad class of approaches based on variational methods of discretization which allow one to find the approximate solution within a desired range of accuracy.

Let H denote the linear space of possible solutions and $u \in H$ the exact solution of the problem. A (*linear, consistent*) *variational method of discretization* consists of a finite dimensional linear subspace $S \subset H$ called the *trial space* in which the approximate solution is sought, a *test space* V (of the same dimension as the trial space S), and a *bilinear form* $B(u, v)$ defined on $H \times V$. The approximate solution, denoted by Pu, is then determined by the conditions

$$Pu \in S \tag{22.1a}$$

$$B(Pu, v) = B(u, v) \qquad \text{for all } v \in V \tag{22.1b}$$

In order that Pu be computable, the following two conditions should be satisfied:

For any $v \in V$, $B(u, v)$ is computable from the data of the problem (without knowing u). (22.2a)

For any $s \in S$ there is some $v \in V$ such that $B(s, v) \neq 0$. (22.2b)

It follows that (22.1) leads to a system of linear equations which is uniquely solvable. It is obvious that $Pu = u$ for any $u \in S$. The approximate solution Pu obviously depends on the selection of S, V, and B.

The acceptability of the approximate solution is stated in terms of a norm $\|\cdot\|$ of the difference of u and Pu, i.e. we accept Pu if

$$\|Pu - u\| \leqslant \tau \|u\| \tag{22.3}$$

where τ is an *a priori* given tolerance. (An absolute error criterion or other variant is equally possible.) Thus, given $\|\cdot\|$ and τ, the goal is to select S, V, and B so that (22.3) is achieved in the most effective way. (We do not give here an exact meaning to the word 'effective'.)

For each mathematical problem *there exists a wide variety of possible variational methods of discretization.* In this paper we shall discuss properties of these methods which enable us to distinguish among them and which aid in the selection or design of a method which is effective in achieving the given goals of the computation. In Section 22.2 some general considerations are discussed. The remainder of the paper is devoted to specific illustrative results. We conclude this introduction with a very simple example in terms of which some of the main ideas will be explained.

Let us consider a longitudinally loaded bar on an elastic support. For $0<x<l$ denote by $u(x)$ and $\sigma(x)$ the longitudinal displacement and normal stress respectively. A classical formulation consists of the boundary value problem

$$E(x)u'(x)=\sigma(x) \tag{22.4a}$$

$$-(F(x)\sigma(x))'+c(x)u(x)=p(x) \tag{22.4b}$$

$$u(0)=0, \qquad u(l)=0 \tag{22.4c}$$

Here $E(x)$ denotes the modulus of elasticity, $F(x)$ the cross-section, $c(x)$ the spring constant of the elastic support, and $p(x)$ the longitudinal load. We can cast this problem in a variational form in various ways. For example, define the bilinear form

$$B_1(u,\sigma;v,\rho)=\int_0^l (Eu'\rho-\sigma\rho+F\sigma v'+cuv)\,\mathrm{d}x \tag{22.5}$$

Then

$$B_1(u,\sigma;v,\rho)=\int_0^l pv\,\mathrm{d}x \tag{22.6}$$

for any $v\in\mathring{H}^1$ and $\rho\in H^0$, where

$$\mathring{H}^1=\left\{v\,\middle|\,\int_0^l [v^2+(v')^2]\,\mathrm{d}x<\infty,\ v(0)=v(l)=0\right\}$$

and

$$H^0=\left\{\rho\,\middle|\,\int_0^l \rho^2\,\mathrm{d}x<\infty\right\}$$

and so $B_1(u,\sigma;v,\rho)$ is computable without explicitly knowing the exact solution (u,σ). Note that $B_1(u,\sigma;v,\rho)$ is defined for all $(u,\sigma)\in\mathring{H}^1\times H^0=H$ and all $(v,\rho)\in H$.

There are many other bilinear forms which could be used in a variational formulation of (22.4). For example, let

$$B_2(u, \sigma; v, \rho) = \int_0^l \left(u'\rho - \frac{\sigma\rho}{E} + F\sigma v' + cuv\right) dx \tag{22.7}$$

Then

$$B_2(u, \sigma; v, \rho) = \int_0^l pv \, dx$$

for (u, σ), $(v, \rho) \in H$ (with H defined as above). Integrating by parts in (22.7) we get

$$B_3(u, \sigma; v, \rho) = \int_0^l \left[-u\rho' - \frac{\sigma\rho}{E} - (F\sigma)'v + cuv\right] dx \tag{22.8}$$

so

$$B_3(u, \sigma, v, \rho) = \int_0^l pv \, dx$$

Here we assume $u, v \in H^0$ and

$$\sigma, \rho \in H^1 = \left\{v \,\middle|\, \int_0^l [v^2 + (v')^2] \, dx < \infty\right\}$$

In both cases the bilinear form is obviously computable from data. For a final example set

$$B_4(u, v) = \int_0^l (EFu'v' + cuv) \, dx \tag{22.9}$$

By eliminating σ from (22.4) we see that

$$B_4(u, v) = \int_0^l pv \, dx$$

for $u, v \in \mathring{H}^1$. This is the usual form used in displacement finite element methods. Of course many other variational formulations of (22.4) are possible (in fact an infinite number).

To complete the specification of a discretization method we must select, in addition to the bilinear form, the trial and test spaces S and V. For example, in the case of B_1 we may select any finite dimensional subspaces S and V of H which are of the same dimension and satisfy (22.2b).

Let us now define some of the norms which we will consider. For k a

non-negative integer set $u^{[k]} = \dfrac{d^k u}{dx^k}$ and let

$$\|u\|_{H^k} = \left[\int_0^l \sum_{j=0}^{k} u^{[j]2}(x)\,dx\right]^{1/2}$$

$$\|u\|_{H^k} = \sum_{j=0}^{k} \operatorname{ess}\sup_{0\leq x\leq l} |u^{[j]}(x)|$$

We also define analogous norms for k a negative integer. For such k define $v = u^{[k]}$ by $u(x) = v^{[-k]}(x)$, choosing the constants of integration so that $\int_0^l v^{[2m]}\,dx = 0$, $0 \leq m \leq (-k-1)/2$; $v^{[2m+1]}(0) = v^{[2m+1]}(l) = 0$, $0 \leq m \leq (-k-2)/2$. Then we will write

$$\|u\|_{H^k} = \|v\|_{H^0}.$$

These negatively indexed norms emphasize the effect of oscillations of u less than do the positively indexed norms. Analogous norms can be defined when u depends on two variables.

22.2 GENERAL CONSIDERATIONS

To achieve the acceptance criterion (22.3) it is certainly necessary that

$$Z(u, S) = \inf_{s\in S} \|u - s\| \leq \tau \|u\| \tag{22.10}$$

The quantity $Z(u, S)$ measures the error in best possible approximation of u by elements of S with respect to the chosen norm $\|\cdot\|$, i.e. the *approximability* of u by S.

The choice of S is clearly essential to the effectiveness of the discretization method. The solution u is unknown *a priori*, and often only the information that $u \in H$ is available. In such cases S has to be selected so that *every* element in H can be approximated well. More information about u allows more effective choice of S. Such information can be achieved through a learning process during the computation and thus S can be selected adaptively (see, for example, [3, 4]).

That the trial functions approximate the solution well, i.e. that the magnitude of $Z(u, S)$ is small, does not alone insure that the approximate solution Pu is close to the exact solution u. Therefore it is reasonable to ask that the method be *quasi-optimal*. This means that

$$\|u - Pu\| \leq KZ(u, S) = K \inf_{s\in S} \|u - s\| \qquad \text{for all } u \in H \tag{22.11}$$

where K is a constant which is not too large. The smallest value of K for which (22.11) holds is called the *quasi-optimality constant*.

Condition (22.11) is equivalent to another condition called the *stability condition.* This states that

$$\|Pu\| \leq K^* \|u\| \qquad \text{for all } u \in H \tag{22.12}$$

The smallest value of K^* for which (22.12) holds is called the *stability constant.* To see that quasi-optimality and stability are equivalent, assume that (22.12) holds. Now, if $s \in S$, then

$$\|u - Pu\| = \|(u-s) - P(u-s)\| \leq \|u-s\| + \|P(u-s)\|$$
$$\leq \|u-s\| + K^* \|u-s\| \leq (K^*+1)\|u-s\|$$

(Here we used the fact that $Ps = s$, as mentioned earlier.) Thus (22.11) holds with $K \leq K^*+1$. On the other hand, assuming (22.11), we have that

$$\|Pu\| \leq \|Pu-u\| + \|u\| \leq KZ(u,S) + \|u\| \leq (K+1)\|u\|$$

and so (22.12) holds with $K^* \leq K+1$. The importance of (22.12) is that it is often easier to verify than (22.11).

Note that while approximability is affected only by the choice of the trial space S, stability (or quasi-optimality) depends on the interplay between B, H, S, and V. Because the test functions are not needed for approximation purposes, *the main goal in the selection of V is to achieve stability* with the smallest possible constant K. Let us remark that for certain bilinear forms and certain norms, the choice $V = S$ leads to the stability constant 1. In such cases, the performance of the method depends solely on the selection of S.

Note, further, that both approximability and stability depend heavily on the norm under consideration. Changing the norm can violate quasi-optimality although the computational algorithm remains the same. Because of this, the method must be investigated in close relation to the given acceptance criterion.

Although approximability and stability are essential and of primary interest for the method, there are other important features to be considered in the rational selection of discretization procedures.

22.2.1 Robustness

The bilinear form B and the solution u may depend on various parameters, e.g. in the above example of the bar problem, E, F, and c may play a significant role. Both approximability and stability will depend on such parameters. A method is called robust when its performance is relatively uninfluenced by the variation of the parameters within a large range.

22.2.2 A *posteriori* estimates and adaptive approaches

A typical acceptance criterion, as mentioned above, is

$$\|u - Pu\| \leq \tau \|u\|$$

where τ is a given tolerance. Although we have

$$\|u - Pu\| \leq KZ(u, s)$$

this estimate may have no direct practical importance. In the first place we will in general not know precise values for the quasi-optimality constant K or for $Z(u, S)$. Moreover, even when these are known the resulting estimate may be very pessimistic. The reason is that the quasi-optimality constant K is based on the worst case (since 22.11 must hold for all $u \in H$), while the true solution may have special properties unknown to us. The only general ways to implement the acceptance criterion reliably are based on *a posteriori* analysis of the approximate solution Pu. Thus a computable error estimator ε is introduced, which depends solely on input data and Pu and satisfies

$$\varepsilon \sim \|u - Pu\|$$

in the sense that

$$C_1\varepsilon \leq \|u - Pu\| \leq C_2\varepsilon$$

and

$$\theta = \frac{\varepsilon}{\|u - Pu\|} \to 1 \qquad \text{as } \varepsilon \to 0$$

This can be achieved (see, for example, [3 to 6]), but not every selection of H, S, V, B, and $\|\cdot\|$ allows for estimators with the same effectiveness and reliability. Feasibility of adaptive selection of test and trial spaces may also be an important feature to be considered in the selection of the form B.

22.3 ILLUSTRATIVE RESULTS

In this section we discuss some concrete mathematical results illustrating the ideas introduced above.

22.3.1 Approximability

First we consider some questions related to approximation. In engineering computations the solution we are interested in approximating usually has special properties. For example, it may be smooth except for some singular behaviour in the neighbourhood of a known point such as a crack tip, corner, or concentrated load. Moreover, the qualitative nature of such singularities is known.

22.3.1.1 The one-dimensional case

The one-dimensional analogue of 'corner' behaviour in two and three dimensions is given by functions $u_\gamma(x) = x^\gamma$, γ a real number. This function

has the property that

$$\sum_{l=0}^{k} \int_0^1 x^{2l-2k+\alpha} \left(\frac{d^l u}{dx^l}\right)^2 dx < \infty \tag{22.13}$$

for any integer $k>0$ and any real number $\alpha > 2k-1-2\gamma$. Suppose we are interested in approximating in the H^0 norm a function satisfying (22.13). It can be shown that there exists a sequence of subspaces $S^{(n)}$ of H^0 such that $S^{(n)}$ has dimension n and the $S^{(n)}$ satisfy the following approximation property: if $u \in H^0$ is any function satisfying (22.13) for a non-negative integer k and any real number α such that $2k > \alpha \geq 0$, or if $u \in H^k$, then

$$Z(u, S^{(n)}) \leq C(k, \alpha) n^{-k} \tag{22.14}$$

Moreover, (22.14) exhibits the best rate of convergence achievable by any subspaces $S^{(n)}$ of dimension n (see [7]). This is a very robust approximation property. In particular, all the functions u_γ, with $\gamma > -\frac{1}{2}$, can be approximated with this rate. (For $\gamma \leq -\frac{1}{2}$, $u_\gamma \notin H^0$ so such a result cannot apply.)

In fact, since the functions u_γ have additional properties, even better approximation than indicated by (22.14), namely an exponential rate of convergence, may be obtained for them by another choice of spaces. Thus there exists a sequence of subspaces $\bar{S}^{(n)}$ of dimension n such that if $\gamma > -\frac{1}{2}$, then

$$Z(u_\gamma, \bar{S}^{(n)}) \leq C \exp(-\beta\sqrt{n}) \tag{22.15}$$

for some $\beta > 0$ (see [8] for details).

The two results quoted above relate to the existence of a sequence of subspaces of H^0 with good approximation properties. We now consider the quality of approximation achieved by some concrete choices of the sequences $S^{(n)}$ suitable for computation. First let $P^{(n)}$ be the space of polynomials of degree less than n. Then any function satisfying (22.13) can be approximated with the error

$$Z(u, P^{(n)}) = \inf_{S \in P^{(n)}} \|u - s\| \leq C(\alpha, k) n^{-\min(k, 2k-\alpha)} \tag{22.16}$$

Applying (22.6) to the functions u_γ ($\gamma > -\frac{1}{2}$) we get the estimate

$$Z(u_\gamma, P^{(n)}) \leq C(\gamma, \varepsilon)^{-(1+2\gamma)+\varepsilon} \tag{22.17}$$

with $\varepsilon > 0$ arbitrarily small. It can also be shown that the estimate (22.17) is essentially the best possible one.

Let us now select $S^{(n)} = S_p^{(n)}$, the space of all piecewise polynomials of degree less than p on a quasi-uniform partition $[0, 1]$ into n elements. This space has dimensions roughly proportional to n. The results analogous to

(22.16) and (22.17) in this case are

$$Z(u, S_p^{(n)}) \leq C(k, \alpha, p) n^{-\min(p, k-\alpha/2)} \tag{22.18}$$

and

$$Z(u_\gamma, S_p^{(n)}) \leq C(\gamma, \varepsilon) n^{-\min[p, (1+2\gamma)/2]+\varepsilon} \tag{22.19}$$

and these rates are essentially unimprovable. Comparing (22.16) and (22.18) we see that for functions u only assumed to satisfy (22.13) the rate of approximation achieved by the polynomials is certainly not worse and may be better than that achieved by the piecewise polynomials. For the functions u_γ, (22.17) and (22.19) show that the rate achieved by the polynomials is at least twice that achieved by the piecewise polynomials (for more details, see [9]).

The estimate (22.18) is in essence the classical estimate

$$Z(u, S_p^{(n)}) \leq C(k, p) n^{-k} \|u\|_{H^k}$$

when $p > k$ and $\alpha = 0$ (see, for example, [10]). The question arises whether under these conditions an expression for $C(k, p)$ can be given which explicitly characterizes the behaviour with respect to p. In [11] such an expression is given in both the one- and two-dimensional cases (and for approximation in H^1). It is shown there that on the right-hand side we can have $\hat{C}(k) n^{-k+\varepsilon}$ with $\hat{C}$ independent of p.

Neither the piecewise polynomial spaces nor the polynomial spaces achieve the optimal rate of convergence characterized by (22.14). For example, for $\gamma > -\frac{1}{2}$ sufficiently small, the function u_γ is not approximated at the optimal rate of n^{-k} for either of these cases. Such a rate can be achieved in the first case by a proper refinement of the mesh in a neighbourhood of the origin and in the second case by changing the polynomials to some other system of functions (see [11, 12]). The importance of this observation is that for engineering computations it appears likely that approximation spaces can be created which yield a rate of convergence which is better than polynomial and is probably exponential.

Thus far we have considered approximation in the H^0 norm. Similar results are available for all the H^l norms, l both positive and negative. An interesting fact is that when l decreases the rate of convergence furnished by either $p^{(n)}$ or $S_p^{(n)}$ increases linearly with l (for more details, for example, see [13]).

22.3.1.2 The multidimensional case

So far we have discussed only the one-dimensional case. Analogous results exist in more than one dimension, but these are far from complete. We will not go into details here, but refer the reader, for example, to [9, 11, 14].

22.3.1.3 The h, p, and h–p versions of the finite element method

As was stated above, there are important cases when selecting the same trial and test spaces leads to a stability constant of 1; thus approximability by the trial space determines the performance of the method. The classical finite element method uses piecewise polynomials of fixed degree p on meshes which are refined to achieve accuracy. Because the size of the elements is usually denoted by h, this method is called the *h version* and the approximability properties (22.18) and (22.19) for such spaces over quasi-uniform meshes are used. The *p version* achieves accuracy by fixing the mesh and increasing the degree p of the polynomial. In this case the approximation results (22.16) and (22.17) are applicable. The p version has been implemented in the program COMET X. We refer the reader to [9] and [15] and references therein for detailed information. Finally, the *h–p version* combines both of these approaches. The exponential convergence rate given in (22.15) can be realized in the h–p version.

22.3.2 Finite element methods

We turn now to a discussion of finite element methods. As discussed in Section 22.2, the quality of approximation yielded by such a method is assured by stability in conjunction with approximability. The stability of a method depends on the interplay between the spaces S and V, the bilinear form B, and the norm $\|\cdot\|$. This is illustrated in the first example.

22.3.2.1 An example illustrating the role of the trial and test spaces in stability

First we consider a one-dimensional problem with the simplest possible bilinear form. Setting $l=1$, $EF=1$, and $c=0$ in (22.9), we get the form

$$B(u, v)=\int_0^1 u'v'\,dx \tag{22.20}$$

The solution $u\in\mathring{H}^1$ satisfies $B(u, v)=\int_0^1 pv\,dx$ for all $v\in\mathring{H}^1$, and the related two-point boundary value problem is

$$-u''=p$$
$$u(0)=u(1)=0$$

For discretization we define spaces of smooth splines. Let $\Delta=\{0=x_0<x_1<\cdots<x_n=1\}$ be a mesh of $[0, 1]$ and set $h_i=x_i-x_{i-1}$. For $\gamma\geq 1$,

the mesh is called γ *quasi-uniform* if

$$\max_{i,j} \frac{h_i}{h_j} \leq \gamma$$

A weaker restriction is that the mesh be γ *locally quasi-uniform*, i.e. that

$$\max_{i} \frac{h_i}{h_{i\pm 1}} \leq \gamma$$

Given any mesh Δ we define for $r = 0, 1, 2, \ldots$ the *smooth splines of degree r* subordinate to Δ to be the piecewise polynomials of degree r with $r-1$ continuous derivatives. The space of all such splines is denoted $M^r(\Delta)$. In particular $M^0(\Delta)$ is the space of piecewise constant functions and $M^1(\Delta)$ is the space of continuous piecewise linear functions. We also denote by $\dot{M}^1(\Delta)$ the space of piecewise linear functions with zero boundary values and by $\dot{M}^3(\Delta)$ the space of natural cubic splines, that is $\dot{M}^3(\Delta) = \{v \in M^3(\Delta) \mid v = v'' = 0 \text{ when } x = 0 \text{ or } 1\}$.

We consider the use of such spline spaces for S and V in conjunction with the bilinear form B defined in (22.20). It is possible to show that if S and V are taken to be spaces of smooth splines of degree r_1 and r_2 respectively (with appropriate boundary conditions), and if r_1 and r_2 are either both odd or both even, then condition (22.2b) is satisfied. Hence $Pu \in S$ is uniquely defined by (22.1).

The most standard case occurs with $S = V = \dot{M}^1(\Delta)$. The stability properties of this method in several norms are summarized in the following theorem.

Theorem 22.1 *Let $S = V = \dot{M}^1(\Delta)$ for an arbitrary partition Δ. Then*

$$\|Pu - u\|_{H^1} \leq K \inf_{s \in S} \|u - s\|_{H^1} \tag{22.21a}$$

$$\|Pu - u\|_{H^1_\infty} \leq K \inf_{s \in S} \|u - s\|_{H^1_\infty} \tag{2r.21b}$$

$$\|\mathrm{Pu} - \mathrm{u}\|_{\mathrm{H}^0_\infty} \leq K \inf_{s \in S} \|u - s\|_{H^0_\infty} \tag{22.21c}$$

for all $u \in \mathring{H}^1$, with K independent of Δ. However, for any $C > 0$ and any mesh Δ there exists $u \in H^1$ such that

$$\|Pu - u\|_{H^0} \geq C \inf_{s \in S} \|u - s\|_{H^0}$$

For more details see [16].

The case where $S = \dot{M}^1(\Delta)$, $V = \dot{M}^3(\Delta)$ is less familiar and more involved.

Theorem 22.2 *Let* $S = \dot{M}^1(\Delta)$, $V = \dot{M}^3(\Delta)$. *Then*:

(*a*) *For an arbitrary partition* Δ,

$$\|Pu - u\|_{H^0} \leq K \inf_{s \in S} \|u - s\|_{H^0}$$

with K *independent of* Δ *and* $u \in \mathring{H}^1$.

(*b*) *For any* $\gamma \geq 1$ *there exists a constant* $K(\gamma)$ *such that for all* γ *quasi-uniform partitions and all* $u \in \mathring{H}^1$,

$$\|Pu - u\|_{H^1} \leq K(\gamma) \inf_{s \in S} \|u - s\|_{H^1}$$

and

$$\|Pu - u\|_{H^{-1}} \leq K(\gamma) \inf_{s \in S} \|u - s\|_{H^{-1}}$$

(*c*) *However, for any* $C > 0$ *and any partition* Δ, *there exists* $u \in H^1$ *such that*

$$\|Pu - u\|_{H^{-2}} \geq C \inf_{s \in S} \|u - s\|_{H^{-2}}$$

(*d*) *If* $1 \leq \gamma < \gamma_0 = 1 + \sqrt{3} + \sqrt{(3 + 2\sqrt{3})} = 5.2745\ldots$, *there exists a constant* $\bar{K}(\gamma)$ *such that*

$$\|Pu - u\|_{H^1} \leq \bar{K}(\gamma) \inf_{s \in s} \|u - s\|_{H^1}$$

and

$$\|Pu - u\|_{H^{-1}} \leq \bar{K}(\gamma) \inf_{s \in S} \|u - s\|_{H^{-1}}$$

for any γ *locally quasi-uniform partition* Δ *and* u. *However,* $\lim_{\gamma \to \gamma_0} \bar{K}(\gamma) = \infty$.

Thus we see that there is a very fine interplay between the trial and test spaces, even for the simplest bilinear form.

So far we have analysed the form (22.20). We now consider the form

$$B(u, v) = \int_0^1 Eu'v' \, dx \tag{22.22}$$

where $0 < e_0 \leq E(x) \leq e_1 < \infty$. The question arises whether Theorems 22.1 and 22.2 remain true as stated. It is possible to show that if $E(x)$ is sufficiently smooth, then Theorems 22.1 and 22.2 hold without change. The

requirement of smoothness means that K may also depend on the maximum of the first few derivatives of E as well as on e_0 and e_1. It can be shown that (22.21a) holds with K depending only on e_0 and e_1, but (22.21c) is not true when no differentiability restrictions are made on E. Thus we may say that the performance of that method is more robust with respect to the coefficient E in the norm $\|\cdot\|_{H^1}$ than it is in the norm $\|\cdot\|_{H^0_\infty}$.

22.3.2. The bilinear form and robustness

We continue to consider the simple problem (22.4) but now consider the effect of the choice of the bilinear form on the robustness of the method. For simplicity assume that $F=1$ and $c=0$. Let $E(x)$ be given satisfying $0<e_0\leqslant E(x)\leqslant e_1$ and consider the bilinear forms (22.5) and (22.7). The bilinear form (22.5) clearly stems from the system of equations

$$\begin{aligned} Eu' &= \sigma \\ \sigma' &= p \end{aligned} \tag{22.23}$$

while (22.7) comes from

$$\begin{aligned} u' &= \frac{\sigma}{E} \\ \sigma' &= p \end{aligned} \tag{22.24}$$

Assume now that we take

$$S = V = \dot{M}^1(\Delta)\times M^0(\Delta) \tag{22.25}$$

in both cases. It is easy to see that σ can be eliminated from the system of linear equations arising from (22.5) with the choice of spaces given in (22.13), and we then get the same method as when (22.10) is used with $S = V = \dot{M}^1(\Delta)$. The properties of this method were summarized in Theorem 22.1. The form (22.7), however, gives different results, which we now consider in detail.

Letting $(\bar{u}, \bar{\sigma}) = P(u, \sigma)$ we get:

Theorem 22.3 *Let Δ be an arbitrary mesh. Define P by the bilinear form (22.7) with S and V defined by (22.25). Then there exists a constant K depending only on e_0 and e_1 such that*

(*a*) $$\|u-\bar{u}\|_{H^1}+\|\sigma-\bar{\sigma}\|_{H^0}\leqslant K \inf_{(\chi,\psi)\in S}[\|u-\chi\|_{H^1}+\|\sigma-\psi\|_{H^0}]$$

(*b*) $$\|u-\bar{u}\|_{H_\infty}+\|\sigma-\bar{\sigma}\|_{H_\infty}\leqslant K\left[\inf_{(\chi,\psi)\in S}\|u-\chi\|_{H^0_\infty}+\|\sigma-\psi\|_{H^0_\infty}\right]$$

(*c*) $$\|\sigma-\bar{\sigma}\|_{H^0}\leqslant K \inf_{\psi\in M^0(\Delta)}\|\sigma-\psi\|_{H^0}$$

(*d*) $$\|\sigma-\bar{\sigma}\|_{H^0_\infty}\leqslant K \inf_{\psi\in M^0(\Delta)}\|\sigma-\psi\|_{H^0_\infty}$$

The statements analogous to (c) and (d) for the error $\|u-\bar{u}\|$ are not true. In order to elaborate this point let us introduce a further notation. For $\chi \in \dot{M}^1(\Delta)$ let $\tilde{\chi}$ be defined by

$$\tilde{\chi}(x_j) = \chi(x_j) \qquad \text{for } j = 0, 1, \ldots, n$$
$$(E\tilde{\chi}')' = 0 \text{ on } (x_{j-1}, x_j) \qquad \text{for } j = 1, 2, \ldots, n$$

Then we have:

Theorem 22.4 *Let $\bar{u}$ be defined as in Theorem 22.3 (i.e. using the form* (22.7) *and spaces of* (22.25). *Then*

$$\|u-\bar{u}\|_{H_0^0} \leq K(e_0, e_1, V(E)) \inf_{\chi \in \dot{M}^1(\Delta)} [\|u-\chi\|_{H_0^0} + \|u-\tilde{\chi}\|_{H_0^0}]$$

and

$$\max_j |u(x_j) - \bar{u}(x_j)| \leq K(e_0, e_1, V(E)) \inf_{\chi \in \dot{M}^1(\Delta)} \|u-\tilde{\chi}\|_{H_0^0}$$

with K depending on e_0, e_1, and the variation $V(E)$ of the function E.

We remark that this theorem is not valid when the dependence of K on $V(E)$ is suppressed. Moreover, while the term $\inf \|u-\tilde{\chi}\|_{H_0^0}$ in the second estimate is necessary, it is usually smaller than the first term. Comparing Theorem 22.4 with the previous results we see that the form (22.7) is much more robust than (22.5) with respect to all the norms we have considered except the H^1 norm, and should be preferred in most situations.

Let us comment on the system of linear equations which the approximate solution led to when $\bar{\sigma}$ is eliminated in either of the two methods discussed in this section. In both cases $Pu \in \dot{M}^1(\Delta)$ is defined by a system of the form

$$\int_0^1 E_\Delta (Pu)' v' \, dx = \int_0^1 pv \, dx \qquad \text{for all } v \in \dot{M}^1(\Delta)$$

When the form (22.5) is used

$$E_\Delta = \frac{1}{h_i} \int_{x_{i-1}}^{x_i} E(x) \, dx \qquad \text{on } (x_{i-1}, x_i)$$

while when (22.7) is used

$$E_\Delta = \left[\frac{1}{h_i} \int_{x_{i-1}}^{x_i} \frac{1}{E(x)} dx \right]^{-1} \qquad \text{on } (x_{i-1}, x_i)$$

$i = 1, 2, \ldots, n$. Thus in the former case E is replaced by its piecewise average and in the latter by its piecewise harmonic average. The above-stated results show that the usual finite element method does not have as good stability properties when $E(x)$ changes significantly over an element,

and so should not be used. The change can be measured by the ratio of the average to the harmonic average.

22.3.2.3 Changing the dependent variable to improve approximability

Now we turn to the analysis of the form defined in (22.8) where for simplicity we take $F=1$ and $c=0$. If we choose $S=V=M^0(\Delta)\times M^1(\Delta)$ and set $(\bar{u},\bar{\sigma})=P(u,\sigma)$, it is easy to prove that

$$\|u-\bar{u}\|_{H^0}+\|\sigma-\bar{\sigma}\|_{H^1}\leq C\inf_{(\chi,\psi)\in S}[\|u-\chi\|_{H^0}+\|\sigma-\psi\|_{H^1}] \tag{22.26}$$

with C independent of Δ but depending on E. Now for any $g\in H^1$ consider the variational formulation

$$B_3(u_g,\sigma_g;v,\rho)=\int_0^l (p-g')v\,dx-\int_0^l \frac{pg}{E}dx \tag{22.27}$$

instead of the method just considered:

$$B_3(u,\sigma;v,\rho)=\int_0^l pv\,dx$$

The new variables are related to the old by

$$\sigma_g=\sigma-g,\qquad u_g=u$$

Thus we may compute σ_g and then take $\sigma=\sigma_g+g$ for the stress component of the approximate solution. Because the same bilinear form occurs in both these approaches, the stability is unchanged and we get:

Theorem 22.5 *For the method associated with (22.27)*

$$\|u-\bar{u}\|_{H^0}+\|\sigma_g-\bar{\sigma}_g\|_{H^1}\leq C\inf_{(\chi,\rho)\in S}\left[\|u-\chi\|_{H^0}+\|\sigma_g-\rho\|_{H^1}\right] \tag{22.28}$$

where the C is the same constant which appears in (22.26) (and therefore is independent of g). Moreover, $\sigma-\bar{\sigma}=\sigma_g-\bar{\sigma}_g$.

Now we note that the proper choice of g can increase the smoothness of σ_g, thereby increasing its approximability and so improving the accuracy of the method. In this simple one-dimensional case the best choice is $g=\int p\,dx$ so σ_g is constant. (In some situations it is as easy or easier for the user to input g as p.) In this case the last term in (22.28) will disappear and $\bar{\sigma}$ will exactly equal σ. A similar idea may be fruitfully applied to related mixed methods in more than one dimension.

22.3.2.4 A robust method for a parameter-dependent problem

In Subsection 22.3.2.2 we discussed a simple case in which changing the bilinear form significantly increased the robustness of the method. We now discuss another example in which a parameter enters in a direct fashion. The problem to be considered models the deflection of a beam allowing for the effect of shear stress. In the simplest case this model can be described by the system of equations

$$\begin{aligned} -\phi_d'' + d^{-2}(\phi_d - \omega_d') &= 0 \qquad 0<x<1 \\ d^{-2}(\phi_d - \omega_d')' &= g \qquad 0<x<1 \end{aligned} \tag{22.29}$$

with the boundary conditions

$$\phi_d(0) = \phi_d(1) = \omega_d(0) = \omega_d(1) = 0$$

Physically $\phi_d(x)$ represents the displacement, $\omega_d(x)$ the rotation of the cross-section, and $g(x)$ the transverse load. The solution depends on the beam thickness d. We associate to the problem (22.29) the bilinear form

$$B_d(\phi, \omega; \psi, \nu) = \int_0^1 [\phi'\psi' + d^{-2}(\phi - \omega')(\psi - \nu')]\,dx$$

Let $S = V = \dot{M}^1(\Delta) \times \dot{M}^1(\Delta)$. Then we have

$$B_d(\phi_d, \omega_d; \psi, \nu) = \int_0^1 g\nu\,dx \tag{22.30}$$

and so the bilinear form is computable. Denoting $P(\phi_d, \omega_d)$ by $(\bar{\phi}_d, \bar{\omega}_d)$ we get

Theorem 22.6

$$\|\phi_d - \bar{\phi}_d\|_{H^1} + \|\omega_d - \bar{\omega}_d\|_{H^1} \leq C(d) \inf_{(\chi,\rho)\in S} [\|\phi_d - \chi\|_{H^1} + \|\omega_d - \rho\|_{H^1}]$$

The constant $C(d)$ is independent of ϕ, $\omega \in \mathring{H}^1$, but $C(d) \to \infty$ as $d \to 0$.

A corollary of this theorem is:

Theorem 22.7 *Let any non-zero load g be given and let $0<\sigma<1$ be arbitrary. Then for any partition Δ there exists a value of d depending on Δ such that*

$$\|\phi_d - \bar{\phi}_d\|_{H^1} \geq \sigma \|\phi\|_{H^1}$$
$$\|\omega_d - \bar{\omega}\|_{H^1} \geq \sigma \|\omega_d\|_{H^1}$$

Theorem 22.7 shows that for small d the method based on (22.30) is virtually useless.

We now associate to our problem another bilinear form, in which we introduce a new variable ξ, representing the shear stress.

$$\hat{B}_d(\phi, \omega, \xi; \psi, \nu, \eta) = \int_0^1 [\phi'\psi' + \xi(\psi - \nu') + \eta(\phi - \omega') - d^2\xi\eta]\,dx \tag{22.31}$$

The functions ϕ_d, ω_d, and $\xi_d = d^{-2}(\phi_d - \omega_d')$ satisfy

$$\hat{B}_d(\phi_d, \omega_d, \xi_d; \psi, \nu, \eta) = \int_0^1 g\nu\,dx \tag{22.32}$$

for any $\psi \in \mathring{H}^1$, $\nu \in \mathring{H}^1$, $\eta \in H^0$. Select now $S = V = \dot{M}^1(\Delta) \times \dot{M}^1(\Delta) \times M^0(\Delta)$, and let $(\hat{\phi}_d, \hat{\omega}_d, \hat{\xi}_d) = P(\phi_d, \omega_d, \xi_d)$. The robustness of the new method with respect to the parameter d is evidenced by the following result.

Theorem 22.8 *For the method associated with (22.32)*

$$\|\phi_d - \hat{\phi}_d\|_{H^1} + \|\omega_d - \hat{\omega}_d\|_{H^1} + \|\xi_d - \hat{\xi}_d\|_{H^0} \leq C \inf_{(\chi,\rho,\lambda)\in S} [\|\phi_d - \chi\|_{H^1} + \|\omega_d - \rho\|_{H^1} + \|\xi_d - \lambda\|_{H^0}]$$

with C independent of Δ and d.

When g in (22.29) is smooth, then ϕ_d, ω_d, and ξ_d are smooth also, and may well be approximated independently of d. It follows that computations based on (22.31) and (22.32) give very good results while we have seen that computations based on (22.30) yield extremely poor results for small d. This difference in the robustness of the two methods with respect to d is very striking in practice. It is also worth noting that the additional variable $\hat{\xi}_d$ may be eliminated from (22.32). The resulting method is identical with the method based on (22.30) except that the integrals are calculated by the composite mid-point rule. By employing this reduced integration implementation, the mixed method entails no extra expense whatever (for more details, see [17]).

22.3.2.5 Robust methods in two dimensions

So far we have discussed various ideas concerning the proper selection of S, V, and B in the context of simple one-dimensional examples. Analogous ideas can be used in more dimensions also, although much less is known at present. Nevertheless we will briefly consider some examples. Consider the problem

$$\frac{\partial}{\partial x} a \frac{\partial u}{\partial x} + \frac{\partial}{\partial y} a \frac{\partial u}{\partial y} = f$$

on

$$\Omega = \{(x, y) \mid |x| < 1, |y| < 1\}$$

with the condition $u = 0$ on $\partial\Omega$. Assume that $a = a_0$ for $x < 0$ and $a = a_1$ for $x \geq 0$, where a_0 and a_1 are distinct positive constants. Having selected a triangulation $\mathcal{T}$ (with minimal angle condition), the usual method employs the bilinear form

$$B(u, v) = \int_\Omega \left(a \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} + a \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} \right) dx\, dy \tag{22.33}$$

and equal trial and test spaces consisting of functions which are continuous, linear on every triangle, and zero on $\partial\Omega$. If the interface $x = 0$ does not coincide with the boundaries of the triangles, then the solution, which is not smooth at the interface, will be approximated less accurately than for a problem with a smooth coefficient. We shall show how to proceed in a slightly different fashion which will avoid this problem.

We select for V the space of continuous piecewise linear functions. Thus the restriction of a test function to a triangle is a linear combination of the three functions 1, x, and y and each test function is continuous at the nodes. The trial functions are taken to restrict on each triangle to a linear combination of the functions 1, $\int_0^x dt\, a$, and y and to be continuous at the nodes. The trial functions, unlike the test functions, need not be continuous on element boundaries except at the nodes (and so are 'non-conforming'). Now interpret (22.33) as a sum of integrals over the individual triangles and replace the norm $\|\cdot\|_{H^1}$ by the norm $\|\cdot\|_{H^1(\mathcal{T})}$ defined as the square root of a sum of integrals over the triangles. This is a common approach in non-conforming finite element methods.

Theorem 22.9 *For this method*

$$\|u - Pu\|_{H^1(\mathcal{T})} \leq C \inf_{\chi \in S} \|u - \chi\|_{H^1(\mathcal{T})}$$

with C independent of $\mathcal{T}$ and u. The value of the unusual trial space S used here is that while stability still holds (as stated in Theorem 22.9), these trial functions mimic the behaviour of the solution u and thus greatly improve the approximability. That is, $Z(u, S)$ is generally much smaller for this choice of S than if S is taken equal to V (the usual choice).

The resulting method therefore gives superior results. An important observation to be made here is that the difficulties encountered with non-conforming methods generally arise from the non-conformity of the test space. Non-conforming trial functions cause no such problems.

A similar idea can be applied to corner problems. Consider solving Laplace's equation on a domain with a corner angle of $\frac{3}{2}\pi$, and zero boundary conditions. The solution then has a singular component of the

type $r^{2/3} \sin 2\pi\theta/3$. For test functions we use the usual linear elements but for trial functions we use elements based on the functions 1, $r^{2/3} \sin 2\pi\theta/3$, and $r^{2/3} \cos 2\pi\theta/3$, instead of 1, x, and y. Using this approach, the loss of accuracy due to the singular behaviour of the solution in a neighbourhood of the corner is prevented. We remark that this procedure need not entail any computational difficulties because it can be implemented in a way which preserves the symmetry of the linear equations, and one may work in the usual way with the microstiffness matrices and nodal variables.

22.4 CONCLUSIONS

We summarize here the main ideas we have presented. As we have seen, there is virtually an unlimited variety of possible variational discretization methods. Such a method is characterized by the bilinear form and the trial and test space. In selecting a method it is of paramount importance to consider the goals of the computation, in particular the norm with which the error is to be assessed. The goals of the computations are best achieved by a method which has good approximability and stability properties with respect to the desired norm. The method should be robust in the sense that these properties apply uniformly over the relevant class of problems. We note that often the obvious method is not best and various variations can lead to strikingly improved results.

REFERENCES

1. M. Vogelius and I. Babuška, 'On a dimensional reduction method: Parts I, II, III', *Math. of Comput.*, **37,** 31–46 (Part I), 47–68 (Part II), to appear (Part III) (1981).
2. I. Babuška and W. Rheinboldt, 'Computational aspects of the finite element method', in *Math. Software III*, pp. 225–255, Academic Press, 1977.
3. I. Babuška and W. Rheinboldt, 'Reliable error estimation and mesh adaptation for the finite element method', in *Computational Methods in Nonlinear Mechanics* (Ed. J. T. Oden), pp. 67–108, North Holland, (1980).
4. I. Babuška, 'Analysis of optimal finite element meshes in R^1', *Math. of Comp.*, **1979,** 435–463 (1979).
5. I. Babuška and W. Rheinboldt, '*A-posteriori* error analysis of finite element solutions for one dimensional problems', *SIAM J. Num. Anal.*, **1981,** 365–389 (1981).
6. I. Babuška and A. Miller, '*A-posteriori* error estimates and adaptive techniques for a finite element method', Inst. for Phys. Sci. & Tech., Lab. for Num. Anal., University of Maryland Tech. Note BN-968, June 1981.
7. H. Triebel, *Interpolation Theory Function Spaces, Differential Operators*, North Holland, Amsterdam, 1978.
8. R. De Vore and K. Scherer, 'Variable knot variable degree spline approximation to x quantitative approximation', *Proc. Bonn Conference*, North Holland, 1978.
9. I. Babuška, B. A. Szabo, and I. N. Katz, 'The *P*-version of the finite element method', *SIAM J. Num. Anal.*, **1981,** 515–545 (1981).

10. P. G. Ciarlet, *The Finite Element Methods for Elliptic Problems*, North Holland, Amsterdam, 1978.
11. I. Babuška, and M. R. Dorr, 'Error estimates for the combined h and p versions of the finite element method', *Num. Math.*, **1981,** 257–277 (1981).
12. M. R. Dorr, Dissertation (in preparation).
13. I. Babuška and A. K. Aziz, 'Survey Lectures on the mathematical foundations of the finite element method', in *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (Ed. A. K. Aziz), Academic Press, 1972.
14. I. Babuška, R. B. Kellogg, and J. Pitkäranta, 'Direct and inverse estimates for finite elements with mesh refinement', *Num. Math.*, **1979,** 447–471 (1979).
15. I. Babuška and B. A. Szabo, 'On the rates of convergence of the finite element method', Center for Comp. Mechanics, Washington Univ., St. Louis, Rept. WU/CCM-80/2. To appear in *Int. J. Num. Meth. in Eng.*, 1981.
16. I. Babuška and J. Osborn, 'Analysis of finite element methods for second order boundary value problems using mesh dependent norms', *Num. Math.*, **34,** 41–62 (1980).
17. D. N. Arnold, 'Discretization by finite elements of a model parameter dependent problem', *Num. Math.*, **37,** 405–421 (1981).