CHAPTER

3 Weak Forms and Finite Element Approximation: 1-D Problems

3.1 Weak forms

In the preceding chapter we stated the governing equations for elasticity and field problems in terms of partial differential equations along with their boundary and, where appropriate, initial conditions. These describe the *continuous problem*. General solutions to problems formulated in this manner are seldom possible for situations of interest in science and engineering. Thus, one seeks to find other approaches that may be used. One approach is to recast the problem in an alternate form from which accurate approximate solutions may be achieved. In this chapter we consider one such approach called a *weak form* of the differential equations. In the following chapter we consider a related but alternate approach based on a *variational approach*. Later we show that the two approaches are closely related and often may be used interchangeably.

A weak form to a set of differential equations is obtained using the following steps:

- Multiply each equation by an appropriate *arbitrary* function.
- Integrate this product over the space domain of the problem.
- Use integration by parts to reduce the order of derivatives to a minimum.
- Introduce boundary conditions if possible.

(3.1)

In this volume we only integrate over the space domain and, thus, it will be necessary to express time dependency and initial conditions explicitly. It is also possible to consider integrations over the time domain [1–3]; however, this is outside the scope of the present volume. The above steps to construct a weak form are direct and may be used to reduce any set of differential equations to a weak form. An arbitrary function is one that can take any value we can imagine. It can be a polynomial, a trigonometric function, a Heaviside function, a Dirac delta function, or any other function. A special case of an arbitrary function is a virtual displacement which may already be known to the reader.

To illustrate the above process we first consider a weak form for the one-dimensional equilibrium equation of elasticity.

The Finite Element Method: Its Basis and Fundamentals. http://dx.doi.org/10.1016/B978-1-85617-633-0.00003-4

3.2 One-dimensional form of elasticity

We first consider a one-dimensional problem in a Cartesian coordinate system with a single component x. The equation for linear elasticity may be deduced by writing the u-displacement as a function of the x-coordinate and time. We assume that our problem is defined in a domain Ω, which is some interval $a < x < b$, and time $t \geq 0$. Accordingly, we write $u(x, t)$ for the displacement and observe that all strain components defined in the previous chapter are zero except

$$\varepsilon_x = \frac{\partial u}{\partial x} \tag{3.2}$$

For simplicity, to describe the stress-strain behavior we assume a Poisson ratio of zero. For this value all components of stress are zero except σ_x; thus, the equilibrium equations reduce to the single equation

$$\frac{\partial \sigma_x}{\partial x} + b_x = \rho \frac{\partial^2 u}{\partial t^2} \tag{3.3}$$

and the constitutive equation to

$$\sigma_x = E\, \varepsilon_x \tag{3.4}$$

where E is Young's modulus of elasticity. The problem is completed by introducing boundary conditions as

$$u = \bar{u} \quad \text{or} \quad t_x = \bar{t}_x = n_x \sigma_x \quad \text{on } x = a, b \tag{3.5}$$

and initial conditions

$$u(x, 0) = \bar{d}(x) \quad \text{and} \quad \dot{u}(x, 0) = \bar{v}(x) \quad \text{in } \Omega \tag{3.6}$$

In Eq. (3.5) n_x is the unit outward normal that is positive unity at b and negative unity at a.

The above equations may be combined by substituting the constitutive equation and strain-displacement equation into the equilibrium equation. This gives the equilibrium equation expressed in terms of displacement as

$$\frac{\partial}{\partial x}\left(E \frac{\partial u}{\partial x}\right) + b_x = \rho \frac{\partial^2 u}{\partial t^2} \tag{3.7}$$

where dependence on x of E is permitted. We note this again is an irreducible form of the differential equation since only a single dependent variable exists.

The set of Eqs. (3.2)–(3.6) is called the *strong form* of the one-dimensional elasticity problem. It can be used to write a set of discrete equations using finite difference methods [4–7]. However, a more powerful approach is to introduce an integral *weak form* of the equations using the steps described in (3.1) from which an alternate approximate method may be introduced.

3.2.1 Weak form of equilibrium equation

We start by introducing an arbitrary function $w(x)$ that is defined in the domain Ω described by the interval $a < x < b$. Multiplying the equilibrium equation (3.3) by this function we may write

$$g(w,\, u,\, \sigma_x) = w(x) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_x - \frac{\partial \sigma_x}{\partial x} \right) = 0 \tag{3.8}$$

It is convenient, though not necessary, that we write a form that is equal to zero. In this way we observe that if $w(x)$ is zero everywhere except at $x = x_c$ then for (3.8) to be zero the differential equation must be satisfied at $x = x_c$. Repeating for all points the expression says the differential equation is satisfied at all points in the domain provided $w(x)$ is arbitrary everywhere. When, in the second step of (3.1), we integrate over the domain we obtain an *integral form* that is also zero and is expressed by

$$\mathrm{G}(w,\, u,\, \sigma_x) = \int_\Omega w(x) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_x - \frac{\partial \sigma_x}{\partial x} \right) \mathrm{d}x = 0 \tag{3.9}$$

This form is a function of functions called a *functional*. Repeating the argument on $w(x)$ being nonzero only at individual points again provides the necessary argument that the differential equation must be zero at all points in the domain. We call this the *fundamental lemma* for an integral form. As a third step, we can integrate the stress term by parts as

$$-\int_\Omega w(x) \frac{\partial \sigma_x}{\partial x} \,\mathrm{d}x = \int_\Omega \frac{\partial w}{\partial x} \,\sigma_x \,\mathrm{d}x - w(x)\, n_x\, \sigma_x \Big|_\Gamma$$

where Γ is the boundary of Ω and n_x is the outward pointing normal to the boundary. The boundary term may be expressed in terms of the traction as

$$\begin{aligned} w(x)\, n_x\, \sigma_x|_\Gamma &= w(b)\, \sigma_x(b) - w(a)\, \sigma_x(a) \\ &= w(b)\, t_x(b) + w(a)\, t_x(a) \end{aligned}$$

where we have noted $n_x(b) = 1$ and $n_x(a) = -1$. We also again introduce the notation that Γ_u is a boundary where $u = \bar{u}$, Γ_t is a boundary where $t_x = \bar{t}_x$, and the total boundary is $\Gamma = \Gamma_u \cup \Gamma_t$. With this notation we can write the *weak form* for equilibrium as

$$\begin{aligned} \mathrm{G}(w,\, u,\, \sigma_x) = &\int_\Omega w(x) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_x \right) \mathrm{d}x \\ &+ \int_\Omega \frac{\partial w}{\partial x} \,\sigma_x \,\mathrm{d}x - w\, t_x \,\big|_{\Gamma_u} - w\, \bar{t}_x \big|_{\Gamma_t} = 0 \end{aligned} \tag{3.10}$$

Note that the time derivative is not reduced as we have not integrated over time. If we wish to show what equations are satisfied by the weak form, we merely remove all derivatives from the weight function w using integrations by parts and recover the governing differential equation along with some boundary terms. The recovered

differential equation is called the *Euler equation* of the weak form and the recovered boundary conditions are called *natural boundary conditions*. Boundary conditions that are not natural are called *essential*. The reader can verify that (3.10) yields the result

$$\mathrm{G}(w,\, u,\, \sigma_x) = \int_\Omega w(x) \left(\rho \frac{\partial^2 u}{\partial t^2} - \frac{\partial \sigma_x}{\partial x} - b_x \right) \mathrm{d}x - w \left(\bar{t}_x - t_x \right)\big|_{\Gamma_t} = 0 \quad (3.11)$$

From this we conclude that a weak form may be written as a sum of all the governing equations in which appropriate weight functions are introduced. This obviously includes equations both in the domain, Ω, and those on the boundary, Γ.

We can introduce the constitutive equation and strain displacement equation to obtain a purely *displacement or "irreducible" form* of the equations. Accordingly, we obtain

$$\mathrm{G}(w,\, u) = \int_\Omega w(x) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_x \right) \mathrm{d}x + \int_\Omega \frac{\partial w}{\partial x} E \frac{\partial u}{\partial x} \mathrm{d}x - w \bar{t}_x\big|_{\Gamma_t} = 0 \quad (3.12)$$

For simplicity, where displacements are specified we set $w = 0$ to eliminate the need to consider t_x. This form is known as a *displacement model* or an *irreducible form*.[1] A *static* problem is one for which the inertia term is omitted. Integrating (3.12) by parts to recover the Euler differential equations gives

$$\mathrm{G}(w,\, u) = \int_\Omega w(x) \left[\rho \frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x} \left(E \frac{\partial u}{\partial x} \right) - b_x \right] \mathrm{d}x - w \left(\bar{t}_x - t_x \right)\big|_{\Gamma_t} = 0 \quad (3.13)$$

Thus, we can observe that the traction boundary conditions are natural ones and the displacement ones are essential.

Comparing Eq. (3.12) with Eq. (3.7) or (3.13) we note that the weak form only requires a specification of first derivatives for u and w with respect to the spatial coordinate x, whereas the strong form needs second derivatives. The weak form also provides a basis to construct approximate solutions.

3.2.1.1 Adjoint forms

If one considers the static problem without load and boundary terms and continues to integrate the weak form by parts to remove all derivatives on u, a differential equation acting on the arbitrary weight function is created. This differential equation is called the *adjoint* to the original equation. For the static problem expressed in (3.12) the adjoint equation is given by

$$\mathrm{G}_{\mathrm{stat}}(w,\, u) = -\int_\Omega u \frac{\partial}{\partial x} \left(E \frac{\partial w}{\partial x} \right) \mathrm{d}x$$

and since it is exactly the same as the original static differential equation on u [viz. (3.13)], the equation is called *self-adjoint*.

[1] When more variables are included in the weak form we shall call the method *mixed*.

Indeed, any linear differential equation $\mathcal{L}(u)$ with weak form written as

$$G_{\text{stat}}(w,\, u) = \int_\Omega w\, \mathcal{L}(u)\, \mathrm{d}x = \int_\Omega u\, \hat{\mathcal{L}}(w)\, \mathrm{d}x \quad \text{with} \quad \mathcal{L} = \hat{\mathcal{L}}$$

is called self-adjoint. Self-adjoint differential equations have useful properties that we will describe later.

3.3 Approximation to integral and weak forms: The weighted residual (Galerkin) method

To construct an approximate solution we express the displacement $u(x,\, t)$ in terms of a set of specified functions multiplied by unknown parameters. In a similar way, we write the arbitrary function in terms of an equal number of specified functions and arbitrary parameters. These may be expressed as

$$\begin{aligned} u(x,\, t) \approx \hat{u}(x,\, t) &= \sum_{n=1}^{N} \phi_n(x)\, a_n(t) + u_{\bar{b}}(x,\, t) \\ w(x,\, t) \approx \hat{w}(x) &= \sum_{m=1}^{N} \psi_m(x)\, w_m \end{aligned} \tag{3.14}$$

In the above form we assume that both $\phi_n(x)$ and $\psi_m(x)$ are zero at all locations where the boundary displacement is specified. The function $u_{\bar{b}}(x,\, t)$ is then specified as any function that satisfies the displacement boundary condition. For example, if the displacement must satisfy $u(L,\, t) = \bar{d}(t)$ on the domain $0 < x < L$, this function may be taken as

$$u_{\bar{b}}(x,\, t) = \frac{x}{L}\, \bar{d}(t) = \phi_{\bar{b}}(x)\, \bar{d}(t)$$

There are many ways that approximations may be assumed. These include:

1. *Point collocation* [8]. $\psi_m = \delta(x_m)$ in (3.12), where $\delta(x_m)$ is such that for $x \neq x_m$, $\psi_m = 0$ but $\int_\Omega \psi_m \mathrm{d}x = 1$. This procedure is equivalent to simply making the weak form zero at N points within the domain and integration is "nominal". Finite difference methods are particular cases of this weighting. For point approximation of $w(x)$ the approximation for $u(x,\, t)$ must have continuous derivatives for all terms in the integral form. For (3.13) this requires that second derivatives be continuous at every x_m.
2. *Subdomain collocation* [9]. $\psi_m = 1$ in subdomains $x_m < x < x_{m+1}$ and zero elsewhere. This essentially makes the integral of the error zero over the specified subdomains. When used with (3.9) this is one of the many finite volume methods [10]. The requirements for approximations of $u(x,\, t)$ permit use of functions whose highest derivatives in the integral form are piecewise continuous. For (3.13) this implies that the $\phi_n(x)$ have continuous second derivatives.

3. *The Galerkin method* (Bubnov-Galerkin) [9,11]. $\psi_n = \phi_n$. This form allows use of approximating functions with piecewise continuous highest derivatives in the *weak form*. For (3.12) we only need the ϕ_n to be continuous. This method, as we shall see, frequently (but by no means always) leads to symmetric matrices and for this and other reasons will be adopted in this volume almost exclusively.

The name "weighted residuals" is clearly much older than that of the "finite element method." The latter uses mainly locally based (element) functions in the expansion of Eq. (3.14) but the general procedures are identical. As the process always leads to equations which, being of integral form, can be obtained by summation of contributions from various subdomains, we choose to embrace all weighted residual approximations under the name *generalized finite element method*. On occasion, simultaneous use of both local and "global" trial functions may be found to be useful.

In the literature the names Petrov and Galerkin [11] are often associated with the use of weighting functions such that $\psi_n \neq \phi_n$. It is important to remark that the well-known *finite difference method* of approximation is a particular case of collocation with locally defined basis functions and is thus a case of a Petrov-Galerkin scheme.

3.3.1 Galerkin solution of elasticity equation

An approximate Galerkin solution for the elasticity problem may be obtained by inserting (3.14) into (3.12) to obtain

$$\begin{aligned}\hat{\mathrm{G}}(\hat{w},\,\hat{u}) = &\sum_{m=1}^{N} w_m \sum_{n=1}^{N} \int_\Omega \psi_m\, \rho \left[\phi_n \frac{\mathrm{d}^2 a_n}{\mathrm{d}t^2} + \phi_{\bar{b}} \frac{\mathrm{d}^2 \bar{d}}{\mathrm{d}t^2}\right] \mathrm{d}x \\ &+ \sum_{m=1}^{N} w_m \sum_{n=1}^{N} \int_\Omega \frac{\mathrm{d}\psi_m}{\mathrm{d}x}\, E \left[\frac{\mathrm{d}\phi_n}{\mathrm{d}x}\, a_n(t) + \frac{\mathrm{d}\phi_{\bar{b}}}{\mathrm{d}x}\, \bar{d}(t)\right] \mathrm{d}x \\ &- \sum_{m=1}^{N} w_m \int_\Omega \psi_m\, b_x \,\mathrm{d}x - \sum_{m=1}^{N} w_m\, \psi_m(x)\, \bar{t}_x\big|_{\Gamma_t} = 0\end{aligned} \tag{3.15}$$

Since the functions ϕ_n, ψ_n, and $\phi_{\bar{b}}$ are all known functions, the integrals may be evaluated as

$$\begin{aligned}M_{mn} &= \int_\Omega \psi_m\, \rho\, \phi_n \,\mathrm{d}x \\ K_{mn} &= \int_\Omega \frac{\mathrm{d}\psi_m}{\mathrm{d}x}\, E\, \frac{\mathrm{d}\phi_n}{\mathrm{d}x}\,\mathrm{d}x \\ f_m(t) &= \int_\Omega \psi_m\, b_x \,\mathrm{d}x + \psi_m\, \bar{t}_x\big|_{\Gamma_t} - \int_\Omega \psi_m\, \rho\, \phi_{\bar{b}} \,\mathrm{d}x\, \frac{\mathrm{d}^2 \bar{d}}{\mathrm{d}t^2} \\ &\quad - \int_\Omega \frac{\mathrm{d}\psi_m}{\mathrm{d}x}\, E\, \frac{\mathrm{d}\phi_{\bar{b}}}{\mathrm{d}x}\,\mathrm{d}x\, \bar{d}(t)\end{aligned} \tag{3.16}$$

The array M_{mn} is called the *mass matrix*, the array K_{mn} is the *stiffness matrix*, and f_m is the *specified load matrix*. Since the parameters w_m are arbitrary, the expression multiplying each one must be zero. This leads to the set of equations

$$\sum_{n=1}^{N}\left[M_{mn}\,\frac{\mathrm{d}^2 a_n}{\mathrm{d}t^2}+K_{mn}\,a_n(t)\right]=f_m(t),\quad m=1,2,\ldots,N \tag{3.17}$$

The original problem of partial differential equations has been reduced to a set of ordinary differential equations in time. For the present we will consider problems in which inertial effects are not important; transient problems are considered in Section 3.8. Problems without inertia effects are called *static*. The equations are now purely algebraic and given by

$$\sum_{n=1}^{N}K_{mn}\,a_n=f_m,\quad m=1,2,\ldots,N \tag{3.18}$$

where the load matrix reduces to

$$f_m=\int_\Omega \psi_m\,b_x\,\mathrm{d}x+\psi(x)\,\bar{t}_x\big|_{\Gamma_t}-\int_\Omega\frac{\mathrm{d}\psi_m}{\mathrm{d}x}\,E\,\frac{\mathrm{d}\phi_{\bar{b}}}{\mathrm{d}x}\,\mathrm{d}x\,\bar{d} \tag{3.19}$$

Expressing the problem in matrix form

$$\mathbf{K}\,\mathbf{a}=\mathbf{f} \tag{3.20}$$

where

$$\mathbf{K}=\begin{bmatrix}K_{11}&K_{12}&\cdots&K_{1N}\\K_{21}&K_{22}&\cdots&K_{2N}\\\vdots&&&\vdots\\K_{N1}&K_{N2}&\cdots&K_{NN}\end{bmatrix}\quad\text{and}\quad\mathbf{f}=\begin{Bmatrix}f_1\\f_2\\\vdots\\f_N\end{Bmatrix}$$

The formal solution to the problem is now given by

$$\mathbf{a}=\mathbf{K}^{-1}\mathbf{f} \tag{3.21}$$

However, in practical situations the inverse of $\mathbf{K}$ is never determined. Instead the solution is obtained using either a direct solution of (3.20) or by an iterative method (see Appendix D).

Once the parameters a_n are known, the approximation for displacement may be computed from (3.14) and that for stresses from

$$\sigma(x)=E\sum_{n=1}^{N}\frac{\mathrm{d}\phi_n}{\mathrm{d}x}\,a_n$$

Example 3.1. Static solution

As an example we consider a static problem with length 10 units and $E=1000$. Figure 3.1 shows the problem to solve. The loading is given by

$$b_x=\begin{cases}10 & \text{for } 0<x<5\\ 0 & \text{for } 5<x<10\end{cases}$$

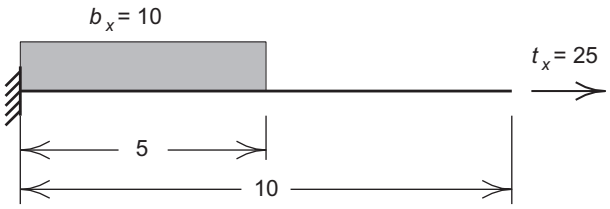


FIGURE 3.1
One-dimensional elasticity example.

and the boundary conditions by

$$\bar{t}_x(10) = 25 \quad \text{and} \quad u(0) = 0$$

For this problem the weak form is given by

$$\int_0^{10} \frac{\partial w}{\partial x}\, 1000\, \frac{\partial u}{\partial x}\, \mathrm{d}x - \int_0^5 w(x)\, 10\, \mathrm{d}x - w(10)\, 25 = 0$$

We consider an approximate solution given by

$$\hat{u}(x) = \sum_{n=1}^{N} \left(\frac{x}{10}\right)^n a_n \quad \text{and} \quad \hat{w}(x) = \sum_{m=1}^{N} \left(\frac{x}{10}\right)^m w_m$$

Note that $u_{\bar{b}} = 0$ for this problem. The solution to the weak form is obtained from

$$\sum_{n=1}^{N} \left[\int_0^{10} 10\, mn \left(\frac{x}{10}\right)^{m-1} \left(\frac{x}{10}\right)^{n-1} \mathrm{d}x \right] a_n = \int_0^5 \left(\frac{x}{10}\right)^m 10\, \mathrm{d}x + 25$$

where $m = 1, 2, \ldots, N$. The stiffness matrix and load vector are given by

$$K_{mn} = \int_0^{10} 10\, mn \left(\frac{x}{10}\right)^{m+n-2} \mathrm{d}x = \frac{100\, mn}{m+n-1}$$

and (3.22)

$$f_m = \int_0^5 \left(\frac{x}{10}\right)^m 10\, \mathrm{d}x + 25 = \frac{100}{m+1} \left(\frac{1}{2}\right)^{m+1} + 25$$

The solution for the parameters using one to five terms in the solution is shown in Table 3.1.

The results from the above problem indicate some important aspects in constructing approximate solutions. First it is noticed that the solutions using three and four terms are identical. This implies that the weak form is able to discard terms that do not contribute to the solution. A second conclusion, however, is that constructing solutions in a global manner does not lead to convergent behavior in the individual parameters. Thus, it is essential to use approximation procedures that have more stable properties. We will observe that finite element methods do possess such good properties. The solution for displacement and stress obtained using one to three terms is

Table 3.1 Parameters for Bar Example

N-terms	a_1	a_2	a_3	a_4	a_5
1	0.37500				
2	0.62500	−0.25000			
3	0.78125	−0.71875	0.31250		
4	0.78125	−0.71875	0.31250	0.00000	
5	0.73437	−0.25000	−1.09275	1.64063	0.65625

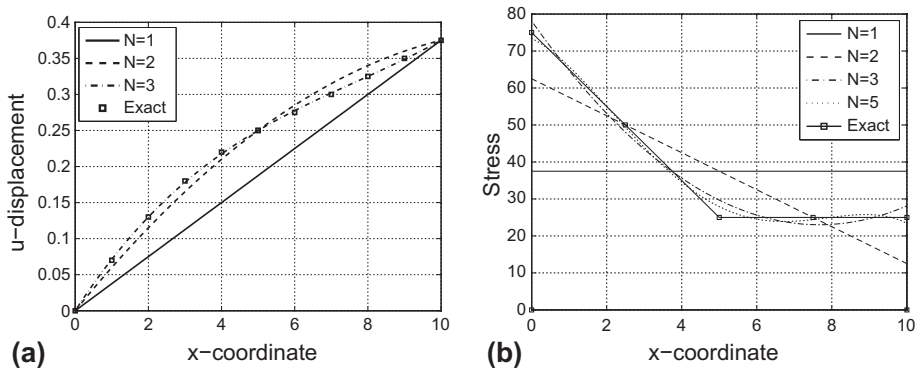


FIGURE 3.2
Displacement and stress for bar using *N*-term solutions: (a) displacement and (b) stress.

shown in Fig. 3.2. From this figure we observe that the displacement at the free end is the same no matter how many terms we use. This often happens in one-dimensional static problems but, unfortunately, is seldom true in higher dimensional problems. The solution for stress converges more slowly than that for the displacements; however, once again we observe that some points are more accurate than others. These we shall call *super-convergent* points and these points play an important role in our later discussion on error estimates and adaptive refinement of solutions.

3.4 Finite element solution

A more convenient method to construct the approximating functions ψ_m and ϕ_n is obtained by dividing the domain to be analyzed into small regular shaped regions. For example, we can divide the one-dimensional region between a and b into a set of "M" small finite segments by defining a set of N points x_i such that

$$x_1 = a, \quad x_i < x_{i+1} \quad \text{and} \quad x_N = b$$

For a one-dimensional problem we can let each increment define a *finite element* domain (or more simply, an *element*) and the set of points define the *nodes*.

The division into elements and nodes is a fundamental part of the finite element method and describes what we will refer to as the *finite element mesh* or simply the *mesh* for the problem.

Using the above subdivision, a simple set of continuous polynomial approximating functions may be defined by

$$\phi_i = \begin{cases} 0, & x < x_{i-1} \\ \dfrac{x - x_{i-1}}{x_i - x_{i-1}}, & x_{i-1} \le x \le x_i \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i}, & x_i < x \le x_{i+1} \\ 0, & x > x_{i+1} \end{cases} \tag{3.23}$$

A plot for these functions and their first derivative is shown in Fig. 3.3. Such functions are called C_0 since only the function is continuous in x, whereas the first derivative is only piecewise continuous with the discontinuities located at the nodes. Development of appropriate C_0 functions for higher spatial dimensions will be considered in subsequent chapters. Here we shall focus on how to describe general forms in *one dimension*.

From the figure we can note that, if required, the end functions can serve as the spatial form for the $\phi_{\bar{b}}(x)$ functions.

If we let $\psi_i = \phi_i$ then we observe that all the integrals defined in the weak form functional can be integrated over the length. Indeed, we usually can evaluate the integrals individually over each element by noting

$$\int_\Omega (\cdot)\, \mathrm{d}x = \sum_{i=1}^{M} \int_{x_i}^{x_{i+1}} (\cdot)\, \mathrm{d}x \equiv \sum_e \int_{\Omega_e} (\cdot)\, \mathrm{d}x \tag{3.24}$$

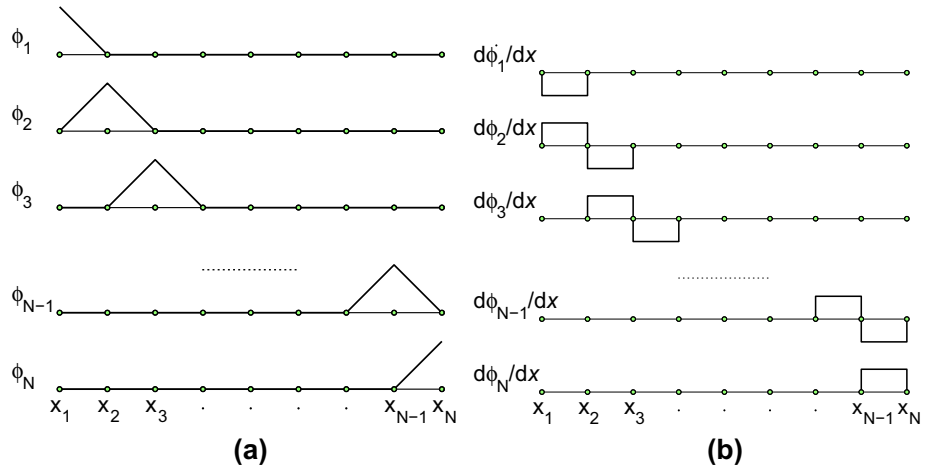


FIGURE 3.3
One-dimensional finite element approximation for ϕ_i: (a) functions and (b) derivatives.

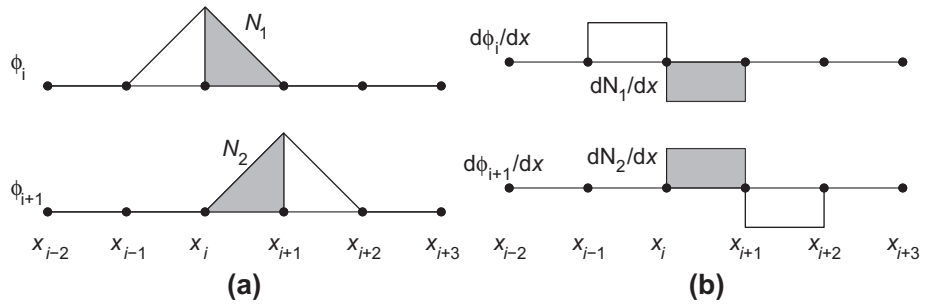


FIGURE 3.4
One-dimensional finite element shape functions: (a) functions and (b) derivatives.

Table 3.2 Local to Global Node Numbering for Two-Node Elements

Local Node	Element Number				
	1	2	3	···	M
1	1	2	3	···	$N-1$
2	2	3	4	···	N

Considering any interval $[x_i,\ x_{i+1}]$ shown in Fig. 3.4 we note that each interval is defined by the same two local functions N_1 and N_2. We call these the *shape functions* for the element. To simplify the notation we also define local nodal coordinates on each element as x_1^e and x_2^e. Then for each element we define the relationship of the local nodes to the global node number as indicated in Table 3.2.

With this notation we can always write the displacements and arbitrary weight function as

$$\begin{aligned} \hat{u}^e &= N_1(x')\,\tilde{u}_1^e + N_2(x')\,\tilde{u}_2^e \\ \hat{w}^e &= N_1(x')\,\tilde{w}_1^e + N_2(x')\,\tilde{w}_2^e \end{aligned} \tag{3.25}$$

If we define a local coordinate system in each element as $x' = x - x_1^e$ and the element length as $h_e = x_2^e - x_1^e$, the shape functions are given by

$$N_1(x') = 1 - \frac{x'}{h_e} \quad \text{and} \quad N_2(x') = \frac{x'}{h_e} \tag{3.26a}$$

and are the same for every element. The derivatives of the shape functions are given by

$$\frac{\mathrm{d}N_1}{\mathrm{d}x} = \frac{\mathrm{d}N_1}{\mathrm{d}x'} = -\frac{1}{h_e} \quad \text{and} \quad \frac{\mathrm{d}N_2}{\mathrm{d}x} = \frac{\mathrm{d}N_2}{\mathrm{d}x'} = \frac{1}{h_e} \tag{3.26b}$$

Using the above we can write the approximation to the weak form as

$$\widehat{\mathrm{G}}(\hat{w},\ \hat{u}) = \widehat{\mathrm{G}}_i(\hat{w},\ \hat{u}) + \widehat{\mathrm{G}}_\sigma(\hat{w},\ \hat{u}) - \widehat{\mathrm{G}}_f(\hat{w},\ \hat{u}) - \hat{w}(x)\bar{t}_x\big|_{\Gamma_t} \tag{3.27a}$$

where $\widehat{\mathrm{G}}_i$ denotes the inertial contribution, $\widehat{\mathrm{G}}_\sigma$ denotes the internal stress contribution, and $\widehat{\mathrm{G}}_f$ denotes the contribution to the force from the body loading. Using (3.12) each

term is defined by

$$\widehat{\mathrm{G}}_i(\hat{w},\,\hat{u}) = \sum_{e=1}^{M} \begin{bmatrix} \tilde{w}_1^e & \tilde{w}_2^e \end{bmatrix} \int_0^{h_e} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \rho_e \begin{bmatrix} N_1 & N_2 \end{bmatrix} \mathrm{d}x' \begin{Bmatrix} \ddot{\tilde{u}}_1^e \\ \ddot{\tilde{u}}_2^e \end{Bmatrix}$$

$$\widehat{\mathrm{G}}_\sigma(\hat{w},\hat{u}) = \sum_{e=1}^{M} \begin{bmatrix} \tilde{w}_1^e & \tilde{w}_2^e \end{bmatrix} \int_0^{h_e} \begin{Bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x'} \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}x'} \end{Bmatrix} E_e \begin{bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x'} & \dfrac{\mathrm{d}N_2}{\mathrm{d}x'} \end{bmatrix} \mathrm{d}x' \begin{Bmatrix} \tilde{u}_1^e \\ \tilde{u}_2^e \end{Bmatrix} \tag{3.27b}$$

$$\widehat{\mathrm{G}}_f(\hat{w},\hat{u}) = \sum_{e=1}^{M} \begin{bmatrix} \tilde{w}_1^e & \tilde{w}_2^e \end{bmatrix} \int_0^{h_e} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} b_x \,\mathrm{d}x'$$

where $\dot{u} = \mathrm{d}u/\mathrm{d}t$.

Each element can be evaluated as

$$\mathbf{M}^e = \int_0^{h_e} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \rho_e \begin{bmatrix} N_1 & N_2 \end{bmatrix} \mathrm{d}x' = \begin{bmatrix} M_{11}^e & M_{12}^e \\ M_{21}^e & M_{22}^e \end{bmatrix}$$

$$\mathbf{K}^e = \int_0^{h_e} \begin{Bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x'} \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}x'} \end{Bmatrix} E \begin{bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x'} & \dfrac{\mathrm{d}N_2}{\mathrm{d}x'} \end{bmatrix} \mathrm{d}x' = \begin{bmatrix} K_{11}^e & K_{12}^e \\ K_{21}^e & K_{22}^e \end{bmatrix} \tag{3.28}$$

$$\mathbf{f}^e = \int_0^{h_e} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} b_x \,\mathrm{d}x' = \begin{Bmatrix} f_1^e \\ f_2^e \end{Bmatrix}$$

For the shape functions given in Eq. (3.26a) and assuming E_e, ρ_e, and b_x are constant within each element the matrices are given by

$$\mathbf{M}^e = \frac{1}{6}\rho_e h_e \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{K}^e = \frac{E_e}{h_e} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{f}^e = \frac{1}{2} b_x h_e \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \tag{3.29}$$

The summation indicated in (3.24) and subsequent expressions leads to a *standard linear problem* with the final equations given by

$$\mathbf{M}\,\ddot{\tilde{\mathbf{u}}} + \mathbf{K}\,\tilde{\mathbf{u}} = \mathbf{f} \tag{3.30}$$

For a static problem the inertia term may be ignored and the expanded form of the matrices is given by

$$\begin{bmatrix} K_{11}^1 & K_{12}^1 & 0 & & \\ K_{21}^1 & (K_{22}^1 + K_{11}^2) & K_{12}^2 & 0 & \\ 0 & K_{21}^2 & (K_{22}^2 + K_{11}^3) & K_{12}^3 & \\ & & & \ddots & \\ & 0 & K_{21}^{M-1} & (K_{22}^{M-1} + K_{11}^M) & K_{12}^M \\ & & 0 & K_{21}^M & K_{22}^M \end{bmatrix} \begin{Bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \\ \vdots \\ \tilde{u}_{N-1} \\ \tilde{u}_N \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{N-1} \\ f_N \end{Bmatrix} \tag{3.31}$$

Up to this point we have not considered boundary conditions other than noting the end functions can serve as the boundary approximation. With the above finite element form it is very simple to impose the displacement boundary conditions since the parameters are now all physical values. That is they obey the property

$$\hat{u}(x_a) \equiv \tilde{u}_a \tag{3.32}$$

Thus, if we wish to impose a displacement condition $u(0) = \bar{u}$, we set $\tilde{u}_1 = \bar{u}$ and rewrite (3.31) as

$$\begin{bmatrix} 1 & 0 & 0 & & \\ 0 & (K_{22}^1 + K_{11}^2) & K_{12}^2 & 0 & \\ 0 & K_{21}^2 & (K_{22}^2 + K_{11}^3) & K_{12}^3 & \\ & & & \ddots & \\ & 0 & K_{21}^{M-1} & (K_{22}^{M-1} + K_{11}^M) & K_{12}^M \\ & & 0 & K_{21}^M & K_{22}^M \end{bmatrix} \begin{Bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \\ \vdots \\ \tilde{u}_{N-1} \\ \tilde{u}_N \end{Bmatrix}$$

$$= \begin{Bmatrix} \bar{u} \\ f_2 - K_{21}^1 \bar{u} \\ f_3 \\ \vdots \\ f_{N-1} \\ f_N \end{Bmatrix} \tag{3.33}$$

This is equivalent to setting $\tilde{w}_1 = 0$. Similarly, imposing a traction condition at $x = L$ only requires the modification

$$f_N \rightarrow f_N + \bar{t}_x(L) \tag{3.34}$$

These physical properties point out the distinct advantage of using a finite element form to approximate the variables.

Example 3.2. Static solution by finite element method

We again consider the solution of the example problem previously considered and shown in Fig. 3.1. We divide the domain into four equal elements as shown in Fig. 3.5 with the solution shown in Fig. 3.6. The solution using eight elements is also shown in the figure. Comparing the two solutions, we note that the stresses are converging more slowly than the displacements (this was true for the first example also) but again have super-convergent points.

3.4.1 Requirements for finite element approximations

The construction of finite element approximations for the basis functions, $\phi_i(x)$, is quite easy to express. Indeed there are only three basic requirements:

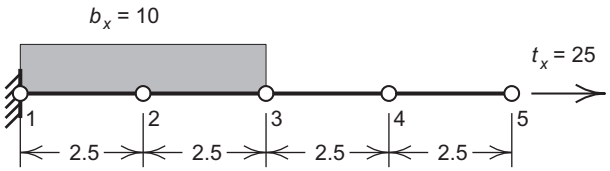


FIGURE 3.5
Four-element mesh for one-dimensional elasticity example.

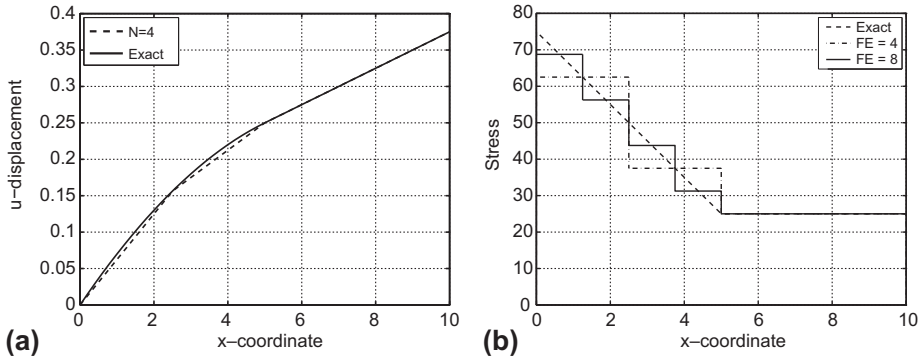


FIGURE 3.6
One-dimensional finite element solution to bar problem: (a) displacements and (b) stress.

1. *Continuity:* Each function must be continuous in the problem domain up to one less than the highest derivative in the functional $G(w, u)$.
2. *Completeness:* Adding terms to the solutions (i.e., increasing the number of ϕ_i and ψ_i) must be such that a suitable choice of the parameters be able to match any bounded solution and its derivative more closely as terms are added. That is,

$$\left|u - \hat{u}|_N\right| < O(h_e^p) \quad \text{and} \quad \left|\frac{\mathrm{d}u}{\mathrm{d}x} - \frac{\mathrm{d}\hat{u}}{\mathrm{d}x}\,|_N\right| < O(h_e^{p-1})$$

   as N is increased.
3. *Linear independence:* The functions must be linearly independent, that is, no combination of some functions may give another of the functions.

It is clear that the simple functions defined in (3.23) satisfy all the above requirements for the one-dimensional elasticity problem. They are continuous and linearly independent as seen in Figs. 3.3 and 3.4. It is also clear that short piecewise continuous straight segments can match any bounded curve as closely as we want by reducing the segment lengths. In particular, linear functions can satisfy the condition $u = const$, which is a rigid translation or rigid body mode of the solution. This condition is crucial to all problems formulated in Cartesian coordinates since any constant displacement causes no strain. If strains were present, we would not be able to match the above

derivative condition. Indeed by using polynomials complete to degree p to define the shape functions we can always satisfy the completeness condition for problems requiring C_0 continuity provided p is at least of degree 1.

3.5 Isoparametric form

The element displacement function for the two-node element was obtained as

$$\hat{u}^e(x,\ t) = N_1(x')\,\tilde{u}_1^e(t) + N_2(x')\,\tilde{u}_2^e(t)$$

where $x = x_1^e + x'$ and the shape functions are expressed by

$$N_1(x') = 1 - \frac{x'}{h^e} \quad \text{and} \quad N_2(x') = \frac{x'}{h^e}$$

If we let

$$\frac{x'}{h^e} = \xi \quad \text{where} \quad 0 \le \xi \le 1$$

the shape functions may be rewritten as

$$N_1(\xi) = 1 - \xi \quad \text{and} \quad N_2(\xi) = \xi, \quad 0 \le \xi \le 1 \tag{3.35a}$$

This permits the writing of the element coordinate, displacement, and arbitrary function in the parametric form

$$\begin{aligned} x^e &= N_1(\xi)x_1^e + N_2(\xi)x_2^e \\ \hat{u}^e &= N_1(\xi)\tilde{u}_1^e + N_2(\xi)\tilde{u}_2^e \\ \hat{w}^e &= N_1(\xi)\tilde{w}_1^e + N_2(\xi)\tilde{w}_2^e \end{aligned} \tag{3.35b}$$

This is called an *isoparametric* form since all three expressions are identical. In this form the coordinate ξ is a parameter called the *parent coordinate*.

The computation of x derivatives now requires use of the chain rule which for any function $f(\xi)$ may be written as

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial \xi}\,\frac{\partial \xi}{\partial x} \tag{3.36}$$

If we let $f(\xi) = x$ then we obtain

$$\frac{\partial x}{\partial x} = 1 = \frac{\partial x}{\partial \xi}\,\frac{\partial \xi}{\partial x}$$

The term $\partial x/\partial \xi$ is called the *Jacobian* and is easily computed from the shape functions as

$$j_e = \frac{\partial x^e}{\partial \xi} = x_2^e - x_1^e = h_e$$

and, thus, for the linear shape functions

$$\frac{\partial \xi}{\partial x} = \frac{1}{h_e}$$

The derivative of the shape functions may now be computed as

$$\frac{\partial N_a}{\partial x} = \frac{\partial N_a}{\partial \xi}\frac{\partial \xi}{\partial x} = \frac{1}{h_e}\frac{\partial N_a}{\partial \xi}, \quad a = 1,\ 2$$

3.5.1 Higher order elements: Lagrange interpolation

Using the isoparametric concept the development of higher order approximations for an element now can be easily constructed. In general we shall write the approximation as

$$u^e \approx \hat{u}^e = \sum_a^n N_a(\xi)\,\tilde{u}_a \tag{3.37}$$

The completeness condition then requires that u contain any constant c which then yields

$$u(\xi) = \sum_a N_a(\xi)\, c = c$$

or

$$\sum_a N_a(\xi) = 1 \tag{3.42}$$

Thus, for any isoparametric shape functions that we may devise, it is only necessary to check whether (3.42) is satisfied and that the functions are continuous across each point on the element boundary.

3.5.1.1 Linear shape functions

Setting $\xi_1 = -1$ and $\xi_2 = 1$, from (3.39) we obtain for $p = 1$ the two shape functions

$$N_1(\xi) = l_1^1(\xi) = \frac{\xi - 1}{-1 - 1} = \frac{1}{2}(1 - \xi) \quad \text{and} \quad N_2(\xi) = l_2^1(\xi) = \frac{\xi + 1}{1 + 1} = \frac{1}{2}(1 + \xi)$$

The Jacobian for this two-node linear element is given by

$$j_e(\xi) = \frac{\partial x}{\partial \xi} = \frac{1}{2}\,(x_2^e - x_1^e) = \frac{1}{2}\,h_e$$

The derivatives of the shape functions are given by

$$\frac{\partial N_a}{\partial x} = \frac{2}{h_e}\,\frac{\partial N_a}{\partial \xi}, \quad a = 1,\ 2$$

3.5.1.2 Quadratic shape functions

For quadratic shape functions we let the nodes be placed at

$$\xi_1 = -1, \quad \xi_2 = 1, \quad \text{and} \quad \xi_3 = 0$$

and obtain the three shape functions as

$$\begin{aligned}
N_1(\xi) = l_1^2(\xi) &= \frac{(\xi - 1)(\xi - 0)}{(-1 - 1)(-1 - 0)} = \frac{1}{2}\xi\,(\xi - 1) \\
N_2(\xi) = l_2^2(\xi) &= \frac{(\xi + 1)(\xi - 0)}{(1 + 1)(1 - 0)} = \frac{1}{2}\xi\,(\xi + 1) \\
N_3(\xi) = l_3^2(\xi) &= \frac{(\xi + 1)(\xi - 1)}{(0 + 1)(0 - 1)} = 1 - \xi^2
\end{aligned} \tag{3.43}$$

Note we have placed the local node 3 in the interior of the element. It is sometimes convenient to let node 1 and node 2 be the boundary nodes of the element with any interior nodes numbered as node 3 and higher.

If we let the coordinates for the element be given by $x_1^e,\ x_2^e,\ x_3^e$, the parametric form for the physical element is given by

$$x = N_1(\xi)\, x_1^e + N_2(\xi)\, x_2^e + N_3(\xi)\, x_3^e$$

The Jacobian is now given by

$$\begin{aligned} j_e(\xi) = \frac{\partial x}{\partial \xi} &= \left(\xi - \frac{1}{2}\right) x_1^e + \left(\xi + \frac{1}{2}\right) x_2^e - 2\,\xi\, x_3^e \\ &= \frac{1}{2} h_e + \xi\, \left(x_1^e + x_2^e - 2\, x_3^e\right) \end{aligned} \tag{3.44}$$

and we note that the Jacobian is not constant unless the coordinate for node 3 is placed at the middle of the element. Furthermore, for a valid isoparametric mapping, the Jacobian must be positive for all values of ξ. This restricts the location for the center node to satisfy

$$x_3^e - x_1^e = \alpha\, (x_2^e - x_1^e), \quad \frac{1}{4} < \alpha < \frac{3}{4} \tag{3.45}$$

A general form for the one-dimensional shape function derivative is given by the rational expression

$$\frac{\partial N_a}{\partial x} = \frac{1}{j_e(\xi)} \frac{\partial N_a}{\partial \xi}, \quad a = 1,\ 2,\ \ldots,\ n$$

3.5.2 Integrals on the parent element: Numerical integration

Since the shape functions are expressed in terms of the parent coordinate, it also is convenient to express the integrals in terms of the parent coordinate. This is easily obtained upon noting the differential is given in terms of the Jacobian by

$$\mathrm{d}x = \frac{\partial x}{\partial \xi}\, \mathrm{d}\xi = j_e(\xi)\, \mathrm{d}\xi \tag{3.46}$$

Thus, our integrals may be expressed as

$$\int_0^{h_e} f(x)\, \mathrm{d}x = \int_{-1}^{1} \hat{f}(\xi)\ j_e(\xi)\, \mathrm{d}\xi \tag{3.47}$$

where $\hat{f}(\xi) = f(x(\xi))$.

Often the integrals involve rational polynomial forms that are difficult to integrate in closed form. In this case it is expedient to use quadrature (numerical integration) to approximate the integrals. While quadrature may not produce exact results for the integrals, used correctly the integrals are sufficiently accurate for use in a finite element calculation. An efficient formula for polynomial forms is Gauss-Legendre quadrature that is defined on the interval $-1 \le \xi \le 1$ (this is the reason we chose this interval above!). The integral is replaced by the form

$$I = \int_{-1}^{1} f(\xi)\, \mathrm{d}\xi \approx \sum_{j=1}^{n} f(\xi_j)\, w_j \tag{3.48}$$

Table 3.3 Gaussian Quadrature Points and Weights for $\int_{-1}^{1} f(\xi)\mathrm{d}\xi = \sum_{j=1}^{n} f(\xi_j)\, w_j$

n	ξ_j	w_j
1	0	2.0
2	$\pm 1/\sqrt{3}$	1.0
3	$\pm\sqrt{0.6}$	5/9
	0.0	8/9
4	$\dfrac{\pm\sqrt{3+\sqrt{4.8}}}{7}$	$\dfrac{1}{2} - \dfrac{1}{3\sqrt{4.8}}$
	$\dfrac{\pm\sqrt{3-\sqrt{4.8}}}{7}$	$\dfrac{1}{2} + \dfrac{1}{3\sqrt{4.8}}$

With appropriate choice of the points ξ_j and weights w_j the formula integrates a polynomial in ξ of degree $2n-1$ exactly. Table 3.3 gives the locations of the points and weights for the first four formulas. For $f(\xi)$ that are rational polynomials use of (3.48) results in an error for the integral; however, by using a sufficient number of points the error will be smaller than the error in the approximation of u using polynomial shape functions.

Example 3.3. Stiffness for two-node element

The stiffness for the two-node element is computed from

$$\mathbf{K}^e = \int_{-1}^{1} \begin{Bmatrix} \dfrac{\partial N_1}{\partial x} \\ \dfrac{\partial N_2}{\partial x} \end{Bmatrix} E_e \begin{bmatrix} \dfrac{\partial N_1}{\partial x} & \dfrac{\partial N_2}{\partial x} \end{bmatrix} \frac{1}{2} h_e \,\mathrm{d}\xi = \frac{E_e}{h_e} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

which is identical to that obtained previously.

Numerical integration is performed using the one-point quadrature formula since the shape function derivatives are constant. The numerical integral is just

$$\int_{-1}^{1} \mathrm{d}\xi = w_1 = 2$$

which is obviously the correct answer. For this problem quadrature was not needed; however, for higher order elements quadrature greatly simplifies the calculations.

Example 3.4. Stiffness for three-node element

The stiffness for the three-node element involves the integral of

$$\mathbf{K}^e = \int_{-1}^{1} \begin{Bmatrix} \dfrac{\partial N_1}{\partial x} \\ \dfrac{\partial N_2}{\partial x} \\ \dfrac{\partial N_3}{\partial x} \end{Bmatrix} E_e \begin{bmatrix} \dfrac{\partial N_1}{\partial x} & \dfrac{\partial N_2}{\partial x} & \dfrac{\partial N_3}{\partial x} \end{bmatrix} j_e(\xi)\,\mathrm{d}\xi$$

where the shape functions are given by (3.43) and the Jacobian by (3.44). For the case where the coordinate for node 3 is centered between nodes 1 and 2 the derivatives are linear in ξ and the Jacobian is constant. In this case the exact stiffness is recovered using a two-point quadrature formula. The reader can verify that the final result is

$$\mathbf{K}^e = \frac{E_e}{3h_e}\begin{bmatrix} 7 & 1 & -8 \\ 1 & 7 & -8 \\ -8 & -8 & 16 \end{bmatrix}$$

If node 3 is not centered the derivatives become rational functions and quadrature yields an approximate answer for the integrals. Numerical experiment shows that two points are still accurate enough to yield good solutions and for convergence to occur as more elements are added.

Example 3.5. Mass matrix for three-node element

The element mass matrix for the three-node element is computed from

$$\mathbf{M}^e = \int_{-1}^{1} \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \end{Bmatrix} \rho_e \begin{bmatrix} N_1 & N_2 & N_3 \end{bmatrix} j_e(\xi)\, \mathrm{d}\xi$$

Since each shape function is a quadratic polynomial and the Jacobian can be a linear polynomial, the highest order polynomial is of order 5. Thus, a three-point quadrature is needed to evaluate the matrix exactly. In solution of transient problems the mass matrix must always be positive definite. Use of a lower order quadrature than three points will make the mass matrix indefinite—that is, possess a zero determinant—and should therefore never be used. For the case where node 3 is centered in the element and the density is constant the mass matrix is given by

$$\mathbf{M}^e = \frac{1}{30}\rho_e h_e \begin{bmatrix} 4 & -1 & 2 \\ -1 & 4 & 2 \\ 2 & 2 & 16 \end{bmatrix}$$

3.6 Hierarchical interpolation

The essence of the finite element method stated above is in approximating the unknown (displacement) in each element by an expansion given in (3.37). We have explicitly chosen to identify such variables with the values of the unknown function at element nodes, thus making

$$\tilde{u}_a = \hat{u}(x_a) \tag{3.49}$$

The shape functions so defined will be referred to as "standard" ones and are the basis of most finite element programs. If polynomial expansions are used and the element satisfies the completeness convergence condition (which specifies that rigid

body displacements cause no strain), it is clear that a constant value of $\tilde{u}_a$ specified at all nodes must result in a constant value of $\hat{u}$,

$$\hat{u} = \left(\sum_{a=1}^{n} N_a \right) u_0 = u_0 \tag{3.50}$$

when $\tilde{u}_a = u_0$. It follows that

$$\sum_{a=1}^{n} N_a = 1 \tag{3.51}$$

at all points of the domain. This important property is known as a *partition of unity* [12].

A drawback exists with "standard" functions; when element refinement occurs, totally new shape functions have to be generated and hence all calculations repeated. It is sometimes advantageous to avoid this difficulty by considering the expression (3.37) as a *series* in which the shape function N_a does not depend on the number of nodes in the mesh n. This indeed is achieved with *hierarchic shape functions*.

The hierarchic concept is well illustrated by the one-dimensional (elastic bar) problem of Fig. 3.7. Here for simplicity the elastic property is taken as constant (E) and the body force b is assumed to vary in such a manner as to produce the exact solution shown in the figure (with zero displacements at both ends).

Two meshes are shown and a linear interpolation between nodal points assumed. For both standard and hierarchic forms the coarse mesh gives

$$K_{11}^{c}\, \tilde{u}_1^{c} = f_1 \tag{3.52}$$

For a finer mesh two additional nodes are added and with the standard shape function the equations requiring solution are

$$\begin{bmatrix} K_{11}^{F} & K_{12}^{F} & 0 \\ K_{21}^{F} & K_{22}^{F} & K_{23}^{F} \\ 0 & K_{32}^{F} & K_{33}^{F} \end{bmatrix} \begin{Bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \end{Bmatrix} \tag{3.53}$$

In this form the zero matrices have been automatically inserted due to element interconnection which is here obvious, and we note that as no coefficients are the same, the new equations have to be resolved.

With the "hierarchic" form using the shape functions shown, a similar form of equation arises and an identical approximation is achieved (being simply given by a series of straight segments). The *final* solution is identical but the meaning of the parameters $\tilde{u}_a^{\star}$ is now different, as shown in Fig. 3.7.

Quite generally,

$$K_{11}^{F} = K_{11}^{c} \tag{3.54}$$

as an identical shape function is used for the first variable. Further, in this particular case the off-diagonal coefficients are zero and the final equations become, for the

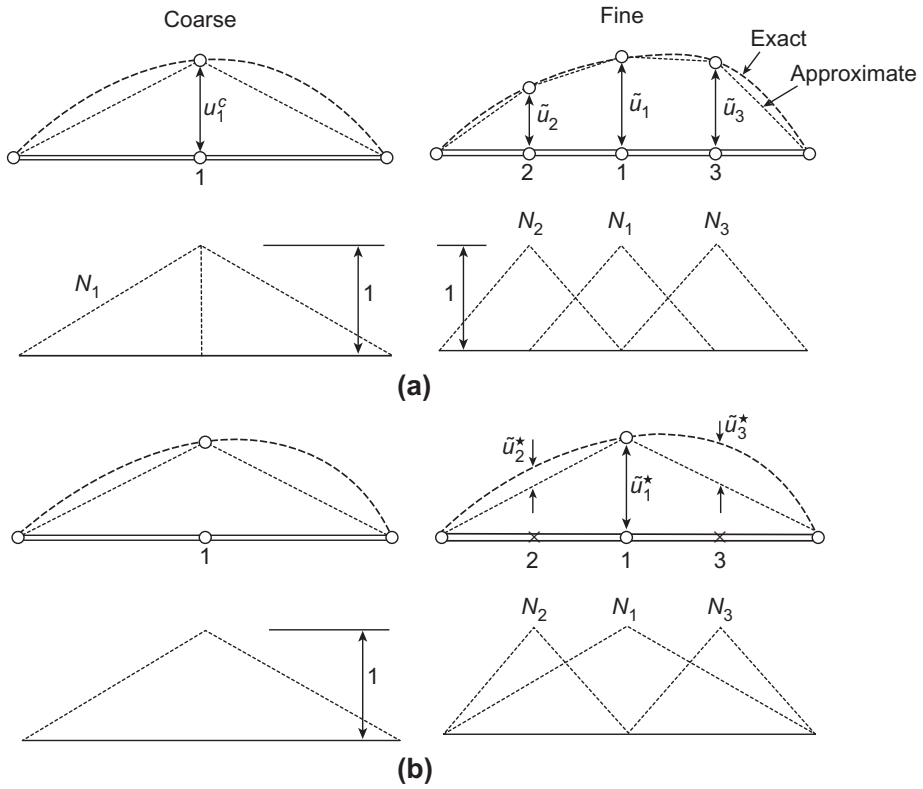


FIGURE 3.7
A one-dimensional problem of stretching of a uniform elastic bar by prescribed body forces.

fine mesh,

$$\begin{bmatrix} K_{11}^c & 0 & 0 \\ 0 & K_{22}^F & 0 \\ 0 & 0 & K_{33}^F \end{bmatrix} \begin{Bmatrix} \tilde{u}_1^\star \\ \tilde{u}_2^\star \\ \tilde{u}_3^\star \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \end{Bmatrix} \tag{3.55}$$

The "diagonality" feature is only true in the one-dimensional problem, but in higher dimensions it will be found that the matrices obtained using hierarchic shape functions are more nearly diagonal and hence usually imply better conditioning than those with standard shape functions. Improved conditioning is advantageous when iterative methods are used to solve the assembled stiffness equations.

Although the variables are now not subject to the obvious interpretation (as local displacement values), they can be easily transformed to those if desired. Though it is not usual to use hierarchic forms in linearly interpolated elements, their derivation in polynomial form is simple and very advantageous.

The reader should note that with hierarchic forms it is convenient to consider the finer mesh as still using the same, coarse, elements but now adding additional refining functions.

Hierarchic forms provide a link with other approximate (orthogonal) series solutions. Many problems solved in classical literature by trigonometric Fourier series expansion are indeed particular examples of this approach.

Example 3.6. Stiffness for three-node hierarchical element

For a three-node element with hierarchical shape functions

$$N_1(\xi) = l_1^1(\xi) = \frac{1}{2}(1-\xi)$$
$$N_2(\xi) = l_2^1(\xi) = \frac{1}{2}(1+\xi)$$
$$N_3(\xi) = l_3^2(\xi) = 1-\xi^2$$

the stiffness matrix for an element with node 3 placed at the center of the element is given by

$$\mathbf{K}^e = \frac{E_e}{3h_e}\begin{bmatrix} 3 & -3 & 0 \\ -3 & 3 & 0 \\ 0 & 0 & 16 \end{bmatrix}$$

Example 3.7. Mass for three-node hierarchical element

Using the shape functions in the previous example the mass matrix for the three-node element with the mid-side node centered is given by

$$\mathbf{M}^e = \frac{1}{30}\rho_e h_e \begin{bmatrix} 10 & 5 & 10 \\ 5 & 10 & 10 \\ 10 & 10 & 16 \end{bmatrix}$$

Thus, while the hierarchical formulation helps in defining the stiffness matrix it does not have the same advantage in describing the mass behavior.

3.7 Axisymmetric one-dimensional problem

The axisymmetric elasticity problem described in Chapter 2 also can be reduced to a one-dimensional problem in terms of the radial direction. Here we shall assume an infinitely long cylinder in which the displacement field is given by

$$u(r,\, z) = u(r) \quad \text{and} \quad v(r,\, z) = 0 \tag{3.56}$$

The nonzero strains for this case may be determined from (2.10) as

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial u}{\partial r} \\ \dfrac{u}{r} \end{Bmatrix} \tag{3.57}$$

The equilibrium equation given in (2.23) simplifies to the single form

$$\frac{\partial \sigma_r}{\partial r} + \frac{\sigma_r - \sigma_\theta}{r} + b_r = \rho \frac{\partial^2 u}{\partial t^2} \tag{3.58}$$

For an isotropic material the stress-strain relations, now including temperature effects, are given by

$$\begin{Bmatrix} \sigma_r \\ \sigma_\theta \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} (1-\nu) & \nu \\ \nu & (1-\nu) \end{bmatrix} \begin{Bmatrix} \varepsilon_r - \alpha\, \Delta T \\ \varepsilon_\theta - \alpha\, \Delta T \end{Bmatrix} \tag{3.59a}$$

and

$$\sigma_z = \nu\, (\sigma_r + \sigma_\theta) - E\, \alpha\, \Delta T \tag{3.59b}$$

where $\alpha\, \Delta T$ is the thermal strain effect.

3.7.1 Weak form for axisymmetric problem

The construction of the weak form for the axisymmetric problem proceeds in an identical manner as described in Section 3.1. We always use Eq. (3.1) to do the construction. The first step in the construction gives

$$g(w,\, u,\, \sigma_r,\, \sigma_\theta) = w(r) \left(\rho \frac{\partial^2}{} \right.$$

We note that the factor 2π appears in all terms and can be omitted if desired. The boundary term may be split into the traction and displacement parts as

$$w(r)\, r\, t_r|_a^b = w\, r\, t_r|_{\Gamma_u} + w\, r\, \bar{t}_r\big|_{\Gamma_t}$$

and again for simplicity we will let $w = 0$ and $u = \bar{u}$ on Γ_u to write the final weak form as

$$\begin{aligned} \mathrm{G}(w,\, u,\, \sigma_r,\, \sigma_\theta) = 2\pi \int_a^b w(r) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_r\right) r\, \mathrm{d}r \\ + 2\pi \int_a^b \left(\frac{\partial w}{\partial r}\sigma_r + \frac{w}{r}\sigma_\theta\right) r\, \mathrm{d}r - 2\pi w(r)\, r\, \bar{t}_r\big|_{\Gamma_t} = 0 \end{aligned} \tag{3.61}$$

3.7.2 A variational notation

By now the reader should have observed that the multiplier of each stress has an identical form as that for the associated strain. Indeed, if we replace the arbitrary function w by a *variational* term δu (read as "variation of u"), the association is complete. Readers familiar with "virtual work" concepts will observe that the variational notation for displacements is identical to "virtual displacement" and we will use the terms interchangeably in this text. Inserting this variational form the axisymmetric variational strains (or virtual strains) are given by

$$\begin{aligned} \frac{\partial w}{\partial r} \to \frac{\partial \delta u}{\partial r} = \delta\varepsilon_r \\ \frac{w}{r} \to \frac{\delta u}{r} = \delta\varepsilon_\theta \end{aligned}$$

3.7.3 Irreducible form for axisymmetric problem

If we substitute the constitutive equation (3.59a) into the weak form (3.62) we obtain

$$\begin{aligned} \mathrm{G}(\delta u, u) = {} & 2\pi \int_a^b \delta u(r) \left(\rho \frac{\partial^2 u}{\partial t^2} - b_r \right) r \, \mathrm{d}r \\ & + 2\pi \int_a^b \begin{bmatrix} \delta\varepsilon_r & \delta\varepsilon_\theta \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{Bmatrix} \varepsilon_r - \alpha\Delta T \\ \varepsilon_\theta - \alpha\Delta T \end{Bmatrix} r \, \mathrm{d}r \\ & - 2\pi \, \delta u(r) \, r \, \bar{t}_r \big|_{\Gamma_t} = 0 \end{aligned} \tag{3.63}$$

where all strains are expressed in terms of the displacement u or δu and the D_{ij} are defined by (3.59a). This is obviously an irreducible or displacement form for the axisymmetric problem.

3.7.4 Finite element solution

Again we divide the region of a cylinder defined by $a \le r \le b$ into a set of M elements defined by nodes r_i and express the integrals as

$$2\pi \int_a^b (\cdot) \, r \, \mathrm{d}r = 2\pi \sum_{i=1}^{M} \int_{r_i}^{r_{i+1}} (\cdot) \, r \, \mathrm{d}r = 2\pi \sum_{i=1}^{M} \int_{-1}^{1} (\cdot) \, r j_e \, \mathrm{d}\xi \tag{3.64a}$$

The strain terms in the weak form (3.63) may be expressed in a matrix form as

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial u}{\partial r} \\ \dfrac{u}{r} \end{Bmatrix} = \begin{bmatrix} \dfrac{\partial N_1}{\partial r} & \dfrac{\partial N_2}{\partial r} \\ \dfrac{N_1}{r} & \dfrac{N_2}{r} \end{bmatrix} \begin{Bmatrix} \tilde{u}_i \\ \tilde{u}_{i+1} \end{Bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \tilde{\mathbf{u}}_e \tag{3.64b}$$

where

$$\mathbf{B}_i = \begin{Bmatrix} \dfrac{\partial N_i}{\partial r} \\ \dfrac{N_i}{r} \end{Bmatrix}, \quad i = 1, 2 \tag{3.64c}$$

The moduli and shape functions are grouped as

$$\begin{aligned} \tilde{\mathbf{u}}_e &= \begin{Bmatrix} \tilde{u}_i \\ \tilde{u}_{i+1} \end{Bmatrix} \\ \mathbf{D} &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \\ \mathbf{N} &= \begin{bmatrix} N_1 & N_2 \end{bmatrix} \quad \text{and} \\ \mathbf{B} &= \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \end{aligned} \tag{3.64d}$$

This permits the weak form to be expressed by

$$\begin{aligned} \mathrm{G}(\delta\mathbf{u},\,\mathbf{u}) = 2\,\pi \sum_{i=1}^{M} \delta\tilde{\mathbf{u}}_e^{\mathrm{T}} \Bigg(& \int_{r_i}^{r_{i+1}} \mathbf{N}^{\mathrm{T}}\,\rho\mathbf{N}\,r\,\mathrm{d}r\,\ddot{\tilde{\mathbf{u}}}_e \\ & + \int_{r_i}^{r_{i+1}} \mathbf{B}^{\mathrm{T}}\,\mathbf{D}\mathbf{B}\,r\,\mathrm{d}r\,\tilde{\mathbf{u}}_e - \int_{r_i}^{r_{i+1}} \mathbf{B}^{\mathrm{T}}\,\mathbf{D}\,\boldsymbol{\alpha}\Delta T\,r\,\mathrm{d}r \\ & - \int_{r_i}^{r_{i+1}} \mathbf{N}^{\mathrm{T}}\,b_r\,r\,\mathrm{d}r \Bigg) - 2\pi\delta u(r)\,\bar{t}_r\,r\big|_{\Gamma_t} = 0 \end{aligned} \tag{3.65}$$

where $\boldsymbol{\alpha}^{\mathrm{T}} = [\alpha\;\alpha]$. In matrix notation the discretized weak form may be written as

$$\mathrm{G}(\delta\mathbf{u},\,\mathbf{u}) = \sum_{e} \delta\tilde{\mathbf{u}}_e^{\mathrm{T}} \left[\mathbf{M}^e\ddot{\tilde{\mathbf{u}}}_e + \mathbf{K}^e\tilde{\mathbf{u}} - \mathbf{f}^e \right] = 0 \tag{3.66}$$

where

$$\begin{aligned} \mathbf{M}^e &= 2\pi \int_{r_i}^{r_{i+1}} \mathbf{N}^{\mathrm{T}}\rho\mathbf{N}\,r\,\mathrm{d}r \\ \mathbf{K}^e &= 2\pi \int_{r_i}^{r_{i+1}} \mathbf{B}^{\mathrm{T}}\mathbf{D}\,\mathbf{B}\,r\,\mathrm{d}r \\ \mathbf{f}^e &= 2\pi \int_{r_i}^{r_{i+1}} \mathbf{N}^{\mathrm{T}} b_r\,r\,\mathrm{d}r + 2\pi \int_{r_i}^{r_{i+1}} \mathbf{B}^{\mathrm{T}}\,\mathbf{D}\,\boldsymbol{\alpha}\Delta T\,r\,\mathrm{d}r \\ &\quad + 2\pi\delta u(r)\,\bar{t}_r\,r\big|_{\Gamma_t} \end{aligned} \tag{3.67}$$

The contribution of the last term of $\mathbf{f}^e$ applies only to those elements at the boundary of the solution domain. After summing over all elements we obtain the assembled form as

$$\mathrm{G}(\delta\mathbf{u},\mathbf{u}) = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \left(\mathbf{M}\,\ddot{\tilde{\mathbf{u}}} + \mathbf{K}\,\tilde{\mathbf{u}} - \mathbf{f} \right) = 0$$

Since the $\delta\tilde{\mathbf{u}}$ is arbitrary, the final set of equations is again in the form of the standard linear problem given in (3.30).

Example 3.8. Thick-walled cylinder subjected to internal pressure

As an example of *static* axisymmetric problem we consider a thick-walled cylinder subjected to an internal pressure and a traction-free external pressure. For the axisymmetric one-dimensional problem there are no rigid body modes. Hence, no essential displacement boundary conditions are needed. We ignore temperature and intertia effects. The properties of the cylinder are given as

$$a = 5,\; b = 10,\; b_r = 0,\; \bar{t}_r(a) = -2,\; E = 10{,}000 \text{ and } \nu = 0.3$$

The weak form for the above static finite element problem is given by

$$\mathrm{G}(\delta\mathbf{u},\,\mathbf{u}) = \sum_{i=1}^{M} \delta\tilde{\mathbf{u}}_e^{\mathrm{T}} \int_{r_i}^{r_{i+1}} \mathbf{B}^{\mathrm{T}}\,\mathbf{D}\mathbf{B}\,r\,\mathrm{d}r\,\tilde{\mathbf{u}}_e - \delta u(r)\,\bar{t}_r\,r\big|_{\Gamma_t} = 0 \tag{3.68}$$

where we also drop the 2π factor. The stiffness matrix for each element may be computed in the parent domain as

$$\mathbf{K}^e = \int_{-1}^{1} \mathbf{B}^{\mathrm{T}}\mathbf{D}\mathbf{B}\, r\, j_e(r)\, \mathrm{d}\xi$$

and is approximated by two-point Gauss-Legendre quadrature. An eight-element problem of linear, two-node elements is considered with element coordinates and displacements given by

$$r^e = N_1(\xi)\, r_1 + N_2(\xi)\, r_2 \quad \text{and} \quad \hat{u}^e = N_1^e(\xi)\, \tilde{u}_1^e + N_2(\xi)\, \tilde{u}_2^e$$

The nodes for the problem are placed at $r_1 = 5$ and $r_9 = 10$ with the other nodes equally spaced in between. The nodal force from the internal boundary traction is given by

$$f_1 = a\, \bar{t}_r(a) = -2 \cdot 5 = -10$$

The results for the displacement and stresses are shown in Fig 3.8. The reader should note the unusual behavior for σ_r but again also note that super-convergent points exist for all stress components.

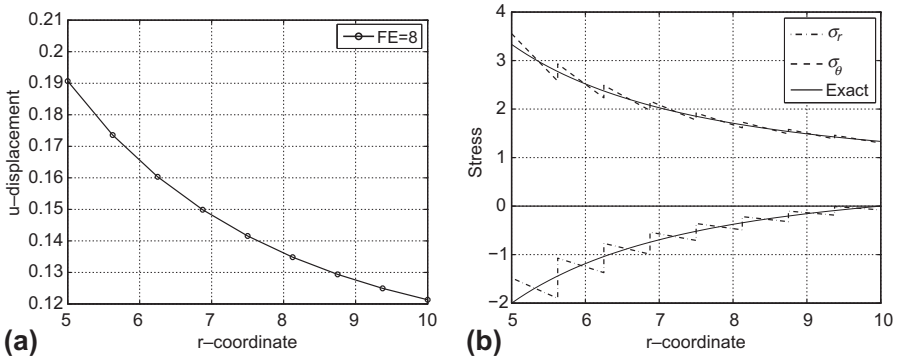


FIGURE 3.8
One-dimensional finite element solution to axisymmetric problem: (a) displacements and (b) stresses.

3.8 Transient problems

When loading is applied with time variation such that the inertial term is significant the full weak form given in (3.30) must be considered. Two methods of solution to the problem may be considered: discrete approximation in time and semi-discretization in which an explicit form of the time behavior is assumed. Both methods reduce the problem to one of algebraic form. Here we summarize the two approaches when applied to the finite element equations.

3.8.1 Discrete time methods

In discrete time methods the transient solution for the time duration of interest is described at a set of discrete time points t_n and approximations to the solution are determined at these points. Eq. (3.30) contains time derivatives up to second degree; thus, we need three different discrete time values which we denote as

$$\left(\tilde{\mathbf{u}}(t_n),\ \dot{\tilde{\mathbf{u}}}(t_n),\ \ddot{\tilde{\mathbf{u}}}(t_n)\right) \approx \left(\tilde{\mathbf{u}}_n,\ \tilde{\mathbf{v}}_n,\ \tilde{\mathbf{a}}_n\right) \tag{3.69}$$

There are many different discrete forms; one which is frequently used is the *Newmark method* [14]. The Newmark method is a *one step* procedure that describes how a solution known at time t_n is determined at time t_{n+1}. Since there are three discrete quantities given in (3.30) we need three independent time equations to obtain a solution at each discrete time. Two of the equations are the Newmark formulas and the third is the discrete equilibrium equation. These are written as

$$\begin{aligned}
\tilde{\mathbf{u}}_{n+1} &= \tilde{\mathbf{u}}_n + \Delta t\, \tilde{\mathbf{v}}_n + \left(\frac{1}{2} - \beta\right) \Delta t^2 \tilde{\mathbf{a}}_n + \beta\, \Delta t^2 \tilde{\mathbf{a}}_{n+1} \\
\tilde{\mathbf{v}}_{n+1} &= \tilde{\mathbf{v}}_n + (1-\gamma)\ \Delta t\, \tilde{\mathbf{a}}_n + \gamma\, \Delta t\, \tilde{\mathbf{a}}_{n+1} \\
\mathbf{R}_{n+1} &= \mathbf{f}_{n+1} - \mathbf{M}\, \tilde{\mathbf{a}}_{n+1} - \mathbf{K}\, \tilde{\mathbf{u}}_{n+1} = \mathbf{0}
\end{aligned} \tag{3.70}$$

where $\Delta t = t_{n+1} - t_n$. The quantity $\mathbf{R}_{n+1}$ is called the *residual* of the equilibrium equation. The result $\mathbf{R}_{n+1} = \mathbf{0}$ describes the requirement that the discrete solution must be satisfied at each time t_{n+1}. In these equations the two parameters β and γ control the overall numerical response for *stability* and *numerical dissipation*. A stable solution is one in which the response after a load is removed does not become large at later discrete times. Numerical dissipation in the above equations implies the solution approaches zero at later discrete times after the loading is removed.

To solve (3.70) we may select any one of the discrete time quantities as the primary dependent time variable and solve for the other two from the Newmark formulas. The simplest form is to select $\tilde{\mathbf{a}}_{n+1}$ as the primary variable and directly substitute the two Newmark formulas into the residual equation, $\mathbf{R}_{n+1}$. The result gives

$$\left[\mathbf{M} + \beta\, \Delta t^2 \mathbf{K}\right] \tilde{\mathbf{a}}_{n+1} = \mathbf{f}_{n+1} - \mathbf{K}\, \widehat{\mathbf{u}}_{n+1} \tag{3.71}$$

where

$$\widehat{\mathbf{u}}_{n+1} = \tilde{\mathbf{u}}_n + \Delta t\, \tilde{\mathbf{v}}_n + \left(\frac{1}{2} - \beta\right) \Delta t^2 \tilde{\mathbf{a}}_n$$

collects all the values at t_n together. The acceleration parameters are given formally as

$$\tilde{\mathbf{a}}_{n+1} = \left[\mathbf{M} + \beta\, \Delta t^2 \mathbf{K}\right]^{-1} \left(\mathbf{f}_{n+1} - \mathbf{K}\, \widehat{\mathbf{u}}_{n+1}\right) \tag{3.72}$$

and we note solving the transient problem involves a repeated solution of the type indicated in (3.21) but with a modified stiffness matrix. Two types of methods exist: (a) $\beta > 0$ which is called an *implicit* method and (b) $\beta = 0$ which is called an *explicit* method. Explicit methods are always conditionally stable, meaning a limit on the size of Δt exists beyond which the solution grows in an uncontrolled manner. If the mass matrix $\mathbf{M}$ is diagonal (see Appendix H), explicit methods are very cheap for each time step and are used extensibly in crash analysis of vehicles. Their main drawback is the small time step which must be used to remain stable. Implicit methods for linear problems can be unconditionally stable. Using $\beta = 1/4$ and $\gamma = 1/2$ results in an unconditionally stable solution without any numerical dissipation for our linear problem [15,16]. The Newmark procedure with these parameters is often called a *trapezoidal rule* or a *constant acceleration* method.

After the acceleration is determined from (3.72) the displacement and velocity at t_{n+1} may be determined from the two Newmark formulas and, thus, all the values at t_{n+1} are known and may be used to advance the solution to the next time. Repeating the process provides a discrete solution to the problem for any specified loading history, $\mathbf{f}_{n+1}$. We note that the solution for the time response is obtained from the solution of a set of *algebraic equations* in the same manner as for static problems.

3.8.1.1 Stability and dissipation

It was noted that the parameters β and γ control the stability and numerical dissipation in the algorithm. To observe how this may be assessed we consider the case where $\beta = 1/4$ and $\gamma = 1/2$. For this case we can write the two Newmark formulas as

$$\begin{aligned}
\tilde{\mathbf{u}}_{n+1} - \tilde{\mathbf{u}}_n &= \frac{1}{2}\, \Delta t\, \left(\tilde{\mathbf{v}}_n + \tilde{\mathbf{v}}_{n+1}\right) \\
\tilde{\mathbf{v}}_{n+1} - \tilde{\mathbf{v}}_n &= \frac{1}{2} \Delta t\, \left(\tilde{\mathbf{a}}_n + \tilde{\mathbf{a}}_{n+1}\right)
\end{aligned}$$

Adding together the equilibrium equations evaluated at t_n and t_{n+1} and multiplying by the transpose of $(\tilde{\mathbf{u}}_{n+1} - \tilde{\mathbf{u}}_n)/2$ gives

$$\begin{aligned}
&\frac{1}{2} \left(\tilde{\mathbf{u}}_{n+1} - \tilde{\mathbf{u}}_n\right)^{\mathrm{T}} \mathbf{M} \left(\tilde{\mathbf{a}}_{n+1} + \tilde{\mathbf{a}}_n\right) + \frac{1}{2} \left(\tilde{\mathbf{u}}_{n+1} - \tilde{\mathbf{u}}_n\right)^{\mathrm{T}} \mathbf{K} \left(\tilde{\mathbf{u}}_{n+1} + \tilde{\mathbf{u}}_n\right) \\
&\quad = \frac{1}{2} \left(\tilde{\mathbf{u}}_{n+1} - \tilde{\mathbf{u}}_n\right)^{\mathrm{T}} \left(\mathbf{f}_{n+1} + \mathbf{f}_n\right)
\end{aligned}$$

Using the Newmark formulas on the mass term we obtain (after dropping the common factor Δt)

$$\begin{aligned}\frac{1}{2}\left(\tilde{\mathbf{u}}_{n+1}-\tilde{\mathbf{u}}_n\right)^{\mathrm{T}}\mathbf{M}\left(\tilde{\mathbf{a}}_{n+1}+\tilde{\mathbf{a}}_n\right) &= \frac{1}{2}\left(\tilde{\mathbf{v}}_{n+1}+\tilde{\mathbf{v}}_n\right)^{\mathrm{T}}\mathbf{M}\left(\tilde{\mathbf{v}}_{n+1}-\tilde{\mathbf{v}}_n\right)\\ &= \frac{1}{2}\tilde{\mathbf{v}}_{n+1}^{\mathrm{T}}\mathbf{M}\,\tilde{\mathbf{v}}_{n+1}-\frac{1}{2}\tilde{\mathbf{v}}_n^{\mathrm{T}}\mathbf{M}\,\tilde{\mathbf{v}}_n\\ &= \mathcal{K}_{n+1}-\mathcal{K}_n\end{aligned}$$

where $\mathcal{K}_n$ is the kinetic energy at time t_n. Similarly, for the stiffness term we recover

$$\begin{aligned}\frac{1}{2}\left(\tilde{\mathbf{u}}_{n+1}-\tilde{\mathbf{u}}_n\right)^{\mathrm{T}}\mathbf{K}\left(\tilde{\mathbf{u}}_{n+1}+\tilde{\mathbf{u}}_n\right) &= \frac{1}{2}\tilde{\mathbf{u}}_{n+1}^{\mathrm{T}}\mathbf{K}\,\tilde{\mathbf{u}}_{n+1}-\frac{1}{2}\tilde{\mathbf{u}}_n^{\mathrm{T}}\mathbf{K}\,\tilde{\mathbf{u}}_n\\ &= \mathcal{U}_{n+1}-\mathcal{U}_n\end{aligned}$$

where $\mathcal{U}_n$ is the elastic potential energy at time t_n. In the above we assume $\mathbf{K}$ and $\mathbf{M}$ are symmetric. Denoting the incremental work of the applied loads as

$$\frac{1}{2}\left(\tilde{\mathbf{u}}_{n+1}-\tilde{\mathbf{u}}_n\right)^{\mathrm{T}}\left(\mathbf{f}_{n+1}+\mathbf{f}_n\right)=\Delta\mathcal{W}$$

and collecting all the terms together gives the energy equation

$$\begin{aligned}\mathcal{K}_{n+1}+\mathcal{U}_{n+1} &= \mathcal{K}_n+\mathcal{U}_n+\Delta\mathcal{W} \quad \text{or}\\ \mathcal{E}_{n+1} &= \mathcal{E}_n+\Delta\mathcal{W}\end{aligned}$$

where $\mathcal{E}_n$ is the total energy at time t_n. For times when no incremental work is added or removed we observe that the *energy is conserved*. Indeed, the energy only changes if forces are applied to the system and in these cases the amount of energy change is a finite quantity. Solution methods with these properties are termed *unconditionally stable*. Of course the solution may not be accurate if the time increments are too large; thus, additional assessments are needed.

3.8.2 Semi-discretization of the problem

An alternative approach to discrete time stepping uses a specified form for the time behavior. For linear ordinary differential equations with constant coefficients the exact time behavior is given in terms of exponentials in time. For the second-order problem we can substitute the form

$$\tilde{\mathbf{u}}=\bar{\mathbf{u}}\,\exp(i\omega t)\quad\text{and}\quad\mathbf{f}=\bar{\mathbf{f}}\,\exp(i\omega t)\quad\text{where } i=\sqrt{-1} \tag{3.73}$$

into (3.30) and divide by the exponential term. The governing equation becomes

$$\left(\mathbf{K}-\omega^2\,\mathbf{M}\right)\bar{\mathbf{u}}=\bar{\mathbf{f}} \tag{3.74}$$

In this form the amplitude of the solution $\bar{\mathbf{u}}$ depends upon the frequency of the excitation ω; however, the solution procedure is again one with an algebraic equation that is identical in form to that for a static problem. Considering all the frequencies we will find that near certain frequencies the amplitude becomes very large. To understand

this phenomenon we consider a simple one degree of freedom problem given by

$$\left(k - \omega^2 m\right) \bar{u} = \bar{f}$$

The solution to the problem is given by

$$\bar{u} = \frac{\bar{f}}{\left(k - \omega^2 m\right)}$$

We can immediately observe that the frequency $\omega = \sqrt{k/m}$ will result in an infinite value of $\bar{u}$. Such frequencies are called the *natural frequencies* of the system. In real physical systems an infinite solution does not occur due to either material dissipation or large displacement effects. However, the determination of the natural frequencies of an elastic body can provide useful information about the properties of linear systems. It is also used as a solution procedure for linear problems. The natural frequencies are independent of the loading and may be computed from the homogeneous equation

$$\left(\mathbf{K} - \omega_i^2 \mathbf{M}\right) \bar{\mathbf{u}}_i = \mathbf{0}$$

The subscript is added to emphasize that a solution to the homogeneous equations is possible only at a finite number of natural frequencies, ω_i. Thus, in general we will need to solve the problem

$$\mathbf{K}\,\bar{\mathbf{U}} = \mathbf{M}\,\bar{\mathbf{U}}\,\boldsymbol{\Lambda} \tag{3.75}$$

where

$$\bar{\mathbf{U}} = \left[\, \bar{\mathbf{u}}_1 \quad \bar{\mathbf{u}}_2 \quad \cdots \quad \bar{\mathbf{u}}_n \,\right]^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\Lambda} = \begin{bmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & 0 & \cdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \omega_n^2 \end{bmatrix}$$

The above problem is an *eigenvalue problem* (see Appendix A) that has the solution pairs $\bar{\mathbf{u}}_i,\ \omega_i^2$ for $i = 1,\ 2,\ \ldots,\ n$. The $\bar{\mathbf{u}}_i$ is the *eigenvector* and ω_i is the associated eigenvalue.

For each mode i of the problem we can compute a scalar mass and stiffness from

$$m_i = \bar{\mathbf{u}}_i^{\mathrm{T}} \mathbf{M} \bar{\mathbf{u}}_i = 1 \quad \text{and} \quad k_i = \bar{\mathbf{u}}_i^{\mathrm{T}} \mathbf{K} \bar{\mathbf{u}}_i = \omega_i^2$$

where the unity condition on m_i is enforced to scale the eigenvector. Next, considering two separate pairs we can write

$$\mathbf{K}\,\bar{\mathbf{u}}_i = \mathbf{M}\,\bar{\mathbf{u}}_i\,\omega_i^2 \quad \text{and} \quad \mathbf{K}\,\bar{\mathbf{u}}_j = \mathbf{M}\,\bar{\mathbf{u}}_j\,\omega_j^2$$

Premultiplying the first by $\bar{\mathbf{u}}_j$, the second by $\bar{\mathbf{u}}_i$, noting symmetry for $\mathbf{K}$ and $\mathbf{M}$, and taking the difference gives

$$\bar{\mathbf{u}}_j^{\mathrm{T}} \mathbf{M} \bar{\mathbf{u}}_i \left(\omega_i^2 - \omega_j^2\right) = 0$$

For situations where $\omega_i \neq \omega_j$ we then obtain the result

$$\bar{\mathbf{u}}_j^{\mathrm{T}} \mathbf{M} \bar{\mathbf{u}}_i = 0 \tag{3.76}$$

For cases in which $\omega_i = \omega_j$ we will enforce (3.76) and, thus, the eigenvectors are constructed to be *mass orthonormal*.

A solution to (3.30) may now be constructed by assuming

$$\mathbf{u}(t) = \sum_{i=1}^{m} \bar{\mathbf{u}}_i \, y_i(t) \tag{3.77}$$

and obtaining

$$\sum_{i=1}^{m} \left(\mathbf{M} \bar{\mathbf{u}}_i \, \ddot{y}_i + \mathbf{K} \, \bar{\mathbf{u}}_i \, y_i \right) = \mathbf{f}$$

Premultiplying by the transpose of eigenvector $\bar{\mathbf{u}}_j$ gives m scalar equations

$$\ddot{y}_j + \omega_j^2 \, y_j = f_j, \quad j = 1, 2, \ldots, m \tag{3.78}$$

where

$$f_j = \bar{\mathbf{u}}_j^{\mathrm{T}} \mathbf{f}$$

The reader can easily solve this equation for simple forms of loading. However, for general time-varying loading a numerical approach is needed. We can, of course, apply the Newmark method to numerically solve each equation and simply add the solution together as indicated in (3.77).

3.8.2.1 Stability of modes

The stability of a discrete time method may also be established from the eigenmodes of the semi-discrete problem. For example letting $y_j = u$ in (3.78) and substituting into a scalar form of the Newmark formulas (3.70) we obtain

$$\begin{aligned} u_{n+1} &= u_n + \Delta t v_n + \left(\frac{1}{2} - \beta \right) \Delta t^2 [f_n - \omega^2 u_n] + \beta \Delta t^2 [f_{n+1} - \omega^2 u_{n+1}] \\ v_{n+1} &= v_n + (1 - \gamma) \Delta t [f_n - \omega^2 u_n] + \gamma \Delta t [f_{n+1} - \omega^2 u_{n+1}] \end{aligned} \tag{3.79}$$

Multiplying the second equation by Δt and collecting terms the above may be written in matrix form as

$$\begin{aligned} \begin{bmatrix} (1 + \beta \Omega^2) & 0 \\ \gamma \Omega^2 & 1 \end{bmatrix} \begin{Bmatrix} u_{n+1} \\ \Delta t v_{n+1} \end{Bmatrix} &= \begin{bmatrix} \left(1 - \left(\frac{1}{2} - \beta \right) \Omega^2 \right) & 1 \\ -(1 - \gamma) \Omega^2 & 1 \end{bmatrix} \begin{Bmatrix} u_n \\ \Delta t v_n \end{Bmatrix} \\ &+ \begin{Bmatrix} \left(\frac{1}{2} - \beta \right) g_n + \beta g_{n+1} \\ (1 - \gamma) g_n + \gamma g_{n+1} \end{Bmatrix} \end{aligned} \tag{3.80}$$

where $\Omega = \omega \Delta t$ and $g_n = \Delta t^2 f_n$. Solving for the solution at t_{n+1} gives

$$\begin{Bmatrix} u_{n+1} \\ \Delta t v_{n+1} \end{Bmatrix} = \frac{1}{1+\beta\Omega^2} \begin{bmatrix} 1 & 0 \\ -\gamma\Omega^2 & (1+\beta\Omega^2) \end{bmatrix} \begin{bmatrix} \left(1 - \left(\frac{1}{2} - \beta\right)\Omega^2\right) & 1 \\ -(1-\gamma)\Omega^2 & 1 \end{bmatrix} \begin{Bmatrix} u_n \\ \Delta t v_n \end{Bmatrix} + \frac{1}{1+\beta\Omega^2} \begin{bmatrix} 1 & 0 \\ -\gamma\Omega^2 & (1+\beta\Omega^2) \end{bmatrix} \begin{Bmatrix} \left(\frac{1}{2} - \beta\right) g_n + \beta g_{n+1} \\ (1-\gamma) g_n + \gamma g_{n+1} \end{Bmatrix} \tag{3.81}$$

The stability of the solution is evaluated from the homogeneous equation by assuming $u_{n+1} = \lambda u_n$ and $v_{n+1} = \lambda v_n$. This yields the algebraic eigenproblem

$$\begin{bmatrix} \left[1 - \left(\frac{1}{2} - \beta\right)\Omega^2 - \bar{\lambda}\right] & 1 \\ \left[-\Omega^2 + \left(\frac{1}{2}\gamma - \beta\right)\Omega^4\right] & [1 + (\beta - \gamma)\Omega^2 - \bar{\lambda}] \end{bmatrix} \begin{Bmatrix} u_n \\ \Delta t v_n \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{3.82}$$

where $\bar{\lambda} = (1 + \beta\Omega^2)\lambda$. A full analysis for all values of β and γ may be found in Refs. [15,17]. Here we consider only the special cases of $\gamma = 1/2$ with $\beta = 0$ or $\beta = 1/4$.

Example 3.9. Implicit integration: $\beta = 1/4$

We first consider the implicit problem with $\gamma = 1/2$ and $\beta = 1/4$. For this case (3.82) simplifies to

$$\begin{bmatrix} \left[\left(1 - \frac{1}{4}\Omega^2\right) - \bar{\lambda}\right] & 1 \\ -\Omega^2 & \left[\left(1 - \frac{1}{4}\Omega^2\right) - \bar{\lambda}\right] \end{bmatrix} \begin{Bmatrix} u_n \\ \Delta t v_n \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

where $\bar{\lambda} = \left(1 + \frac{1}{4}\Omega^2\right)\lambda$. The two values for the $\bar{\lambda}$ are obtained from the characteristic polynomial

$$\left[\left(1 - \frac{1}{4}\Omega^2\right) - \bar{\lambda}\right]^2 + \Omega^2 = 0$$

and yield the result

$$\lambda_j = \frac{\left(1 - \frac{1}{4}\Omega^2\right) \pm i\,\Omega}{\left(1 + \frac{1}{4}\Omega^2\right)} \quad \text{for } j = 1,\ 2$$

The solution is complex and has a modulus value $\rho = \max(|\lambda_1|, |\lambda_2|) = 1$ independent of the value of Ω (and Δt). Thus, again we have shown that the use of these parameters yields an unconditionally stable method for second-order, linear differential equations.

Example 3.10. Explicit integration: $\beta = 0$

We next consider the explicit problem with $\gamma = 1/2$ and $\beta = 0$. For this case (3.82) simplifies to

$$\begin{bmatrix} \left[\left(1 - \frac{1}{2}\Omega^2\right) - \lambda\right] & 1 \\ -\Omega^2 + \frac{1}{4}\Omega^4 & \left[\left(1 - \frac{1}{2}\Omega^2\right) - \lambda\right] \end{bmatrix} \begin{Bmatrix} u_n \\ \Delta t\, v_n \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

The two values of λ may be determined from

$$\left[\left(1 - \frac{1}{2}\Omega^2\right) - \lambda\right]^2 + \Omega^2 - \frac{1}{4}\Omega^4 = 0$$

For this case two possibilities for roots exist depending on the value of Ω:

$$\lambda_j = \begin{cases} 1 - \frac{1}{2}\Omega^2 \pm \left[\frac{1}{4}\Omega^4 - \Omega^2\right]^{1/2} & \text{for } \Omega > 2 \\ 1 - \frac{1}{2}\Omega^2 \pm i\left[\Omega^2 - \frac{1}{4}\Omega^4\right]^{1/2} & \text{for } \Omega < 2 \end{cases}$$

When $\Omega = 2$ the roots are both unity and, thus, are at a limit of stability. For $\Omega > 2$ the value of ρ defined above will be greater than unity and the method is unstable. Thus for stability we must always have

$$\Delta t < \frac{2}{\omega}$$

where ω is the *largest eigenvalue* from the problem being solved. This value may be estimated by computing the maximum frequency of individual elements [18]. Alternatively, the critical time step can be estimated by a few iterations of a Rayleigh quotient [19].

3.9 Weak form for one-dimensional quasi-harmonic equation

In the previous sections we have considered a procedure to solve one-dimensional elasticity problems using a weak form of the governing equations. Here we repeat the process for a one-dimensional quasi-harmonic equation. We consider a problem in Cartesian coordinates where

$$q_x = q_x(x) \quad \text{and} \quad q_y = q_z = 0$$

With this assumption the quasi-harmonic equation is given by

$$-\frac{\partial q_x}{\partial x} + Q = c\,\frac{\partial \phi}{\partial t} \tag{3.83}$$

The one-dimensional form of the material law (2.83) is given by

$$q_x = -k \frac{\partial \phi}{\partial x} \tag{3.84}$$

and the boundary and initial conditions are taken as one-dimensional forms of (2.81a), (2.81b), and (2.82).

3.9.1 Weak form

Following the rules given in (3.1) for constructing a weak form we may write the first step as

$$g(q_x,\ \phi,\ \delta\phi) = \delta\phi \left(c \frac{\partial \phi}{\partial t} - Q + \frac{\partial q_x}{\partial x} \right) = 0$$

Integrating over the domain $0 < x < L$ we obtain the integral form

$$\mathrm{G}(q_x,\ \phi,\ \delta\phi) = \int_\Omega \delta\phi \left(c \frac{\partial \phi}{\partial t} - Q + \frac{\partial q_x}{\partial x} \right) \mathrm{d}x = 0$$

Upon integrating the last term by parts the expression becomes

$$\mathrm{G}(q_x,\ \phi,\ \delta\phi) = \int_\Omega \left[\delta\phi \left(c \frac{\partial \phi}{\partial t} - Q \right) - \frac{\partial \delta\phi}{\partial x} q_x \right] \mathrm{d}x + \delta\phi q_x n_x|_\Gamma = 0$$

Finally, considering the boundary conditions and letting $\phi = \bar{\phi}$ and $\delta\phi = 0$ on Γ_ϕ and $q_x n_x = \bar{q}_n + H\,(\phi - \phi_0)$ on Γ_q we obtain

$$\begin{aligned} \mathrm{G}(q_x,\ \phi,\ \delta\phi) = \int_\Omega \left[\delta\phi \left(c \frac{\partial \phi}{\partial t} - Q \right) - \frac{\partial \delta\phi}{\partial x} q_x \right] \mathrm{d}x \\ + \delta\phi \left[\bar{q}_n + H\,(\phi - \phi_0) \right]\big|_{\Gamma_q} = 0 \end{aligned} \tag{3.85}$$

An *irreducible form* is obtained by inserting the one-dimensional form (3.84) into (3.85) to obtain

$$\begin{aligned} \mathrm{G}(\phi,\ \delta\phi) = \int_\Omega \left[\delta\phi \left(c \frac{\partial \phi}{\partial t} - Q \right) + \frac{\partial \delta\phi}{\partial x} k \frac{\partial \phi}{\partial x} \right] \mathrm{d}x \\ + \delta\phi \left[\bar{q}_n + H\,(\phi - \phi_0) \right]\big|_{\Gamma_q} = 0 \end{aligned} \tag{3.86}$$

3.9.2 Finite element solution of quasi-harmonic problem

A finite element solution to the irreducible one-dimensional quasi-harmonic problem may be constructed in an identical manner to that performed for the elasticity problem. To this end we define a set of nodes x_i on the domain and, similarly, a set of elements using the nodes. Using the isoparametric concept we can define approximations on each element using shape functions. Since only first derivatives in space are contained

in the weak form we may use any of the C_0 shape functions defined above for the elasticity problem. Accordingly, we can write for each element

$$x^e = \sum_a N_a(\xi)\, x_a^e,\ \phi^e = \sum_a N_a(\xi)\, \tilde{\phi}_a^e \quad \text{and} \quad \delta\phi^e = \sum_a N_a(\xi)\, \delta\tilde{\phi}_a^e \tag{3.87}$$

Dividing the domain into a set of elements the weak form (3.86) may be written as

$$\mathrm{G}(\phi,\ \delta\phi) = \sum_e \mathrm{G}^e(\phi,\ \delta\phi) + \mathrm{G}^b(\phi,\ \delta\phi) \tag{3.88}$$

which gives the element expression

$$\mathrm{G}^e(\phi,\ \delta\phi) = \sum_a \delta\tilde{\phi}_a^e \left[\int_{-1}^{1} N_a \left(c \sum_b N_b \frac{\partial \tilde{\phi}_b^e}{\partial t} - Q \right) + \frac{\partial N_a}{\partial x}\, k \sum_b \frac{\partial N_b}{\partial x}\, \tilde{\phi}_b^e \right] j_e \, \mathrm{d}\xi \tag{3.89a}$$

Similarly, the boundary term becomes

$$\mathrm{G}^b(\phi,\ \delta\phi) = \delta\tilde{\phi}_a N_b(x_a) \left[\bar{q}_n + H\,(N_b(x_a)\tilde{\phi}_b - \phi_0) \right]\Big|_{x_a \in \Gamma_q} \tag{3.89b}$$

The integrals may now be evaluated to give the element matrix expressions

$$\begin{aligned} C_{ab}^e &= \int_{-1}^{1} N_a\, c\, N_b\, j_e \, \mathrm{d}\xi \\ H_{ab}^e &= \int_{-1}^{1} \frac{\partial N_a}{\partial x}\, k\, \frac{\partial N_b}{\partial x}\, j_e \, \mathrm{d}\xi \quad \text{and} \\ s_a^e &= \int_{-1}^{1} N_a\, Q\, j_e \, \mathrm{d}\xi \end{aligned} \tag{3.90}$$

Similarly, considering the boundary term we obtain

$$\begin{aligned} H_{ab}^b &= N_a(x_a)\, H\, N_b(x_a)|_{x_a \in \Gamma_q} \quad \text{and} \\ s_a^b &= N_a(x_a)\left[-\bar{q}_n + H\, \phi_0 \right]\Big|_{x_a \in \Gamma_q} \end{aligned} \tag{3.91}$$

where we note the boundary radiation term contributes to both the coefficient matrix and the loading term. Inserting the matrices into (3.89a) and (3.89b) gives

$$\begin{aligned} &\mathrm{G}^e(\phi,\ \delta\phi) + \mathrm{G}^b(\phi,\ \delta\phi) \\ &\quad = \sum_a \delta\phi_a^e \left[\sum_b \left(C_{ab}^e \frac{\partial \tilde{\phi}_b^e}{\partial t} + \left(H_{ab}^e + H_{ab}^b \right) \tilde{\phi}_b^e \right) - \left(s_a^e + s_a^b \right) \right] \end{aligned} \tag{3.92}$$

Performing the summation indicated in (3.88) yields

$$\mathrm{G}(\boldsymbol{\phi},\ \delta\boldsymbol{\phi}) = \delta\tilde{\boldsymbol{\phi}}^{\mathrm{T}} \left[\mathbf{C} \frac{\partial \tilde{\boldsymbol{\phi}}}{\partial t} + \mathbf{H}\,\tilde{\boldsymbol{\phi}} - \mathbf{s} \right] = 0 \tag{3.93}$$

in which $\mathbf{s}$ and $\mathbf{H}$ contain also the contribution from the boundary terms on Γ_q and modification for any Γ_ϕ conditions are also imposed. Since the variation $\delta\tilde{\boldsymbol{\phi}}$ is arbitrary the above expression again leads to a standard linear problem

$$\mathbf{C}\frac{\partial\tilde{\boldsymbol{\phi}}}{\partial t} + \mathbf{H}\,\tilde{\boldsymbol{\phi}} = \mathbf{s} \tag{3.94}$$

If the loading $\mathbf{s}$ is applied very slowly, the transient term may be neglected and a *steady-state* solution evaluated from

$$\mathbf{H}\,\tilde{\boldsymbol{\phi}} = \mathbf{s} \tag{3.95}$$

Once a solution is known the flux in each element may be determined from (3.84).

3.9.3 Transient problems

Transient problems may be solved by discretizing the time derivative in (3.94) in a manner similar to that for the elasticity problem. Here, however, only a first derivative in time exists. An approximate solution at each discrete time t_n between $t = 0$ and $t = T$ is denoted by

$$(\tilde{\boldsymbol{\phi}}(t_n),\ \dot{\tilde{\boldsymbol{\phi}}}(t_n)) \approx (\tilde{\boldsymbol{\phi}}_n,\ \dot{\tilde{\boldsymbol{\phi}}}_n)$$

Similar to the Newmark formulas we consider the approximation

$$\tilde{\boldsymbol{\phi}}_{n+1} = \tilde{\boldsymbol{\phi}}_n + (1-\theta)\,\Delta t\,\dot{\tilde{\boldsymbol{\phi}}}_n + \theta\,\Delta t\,\dot{\tilde{\boldsymbol{\phi}}}_{n+1} \tag{3.96}$$

where θ is a parameter that may range between 0 and 1. A discrete approximation to (3.94) is given by

$$\mathbf{C}\,\dot{\tilde{\boldsymbol{\phi}}}_{n+1} + \mathbf{H}\,\tilde{\boldsymbol{\phi}}_{n+1} = \mathbf{s}_{n+1} \tag{3.97}$$

Substituting (3.96) into (3.97) yields the algebraic equations

$$\left[\mathbf{C} + \theta\,\Delta t\,\mathbf{H}\right]\dot{\tilde{\boldsymbol{\phi}}}_{n+1} = \mathbf{s}_{n+1} - \mathbf{H}\,\widehat{\boldsymbol{\phi}}_{n+1} \tag{3.98}$$

where

$$\widehat{\boldsymbol{\phi}}_{n+1} = \tilde{\boldsymbol{\phi}}_n + (1-\theta)\,\Delta t\,\dot{\tilde{\boldsymbol{\phi}}}_n$$

3.9.3.1 *Stability*

The stability of the discrete approximation may be determined from a single homogeneous scalar equation

$$c\dot{\phi} + h\phi = 0$$

Evaluating this equation at t_n and t_{n+1} and substituting into (3.96) gives the recurrence

$$\phi_{n+1} = \frac{[1-(1-\theta)\,\lambda\Delta t]}{[1+\theta\,\lambda\Delta t]}\,\phi_n$$

where $\lambda = h/c$. The exact solution to the scalar equation is $\exp(-\lambda t)$. Thus, the approximate solution should always produce solutions where $|\phi_{n+1}| < |\phi_n|$. A simple calculation reveals that this is always true for any Δt if $\theta > 0.5$ and thus the solution

is unconditionally stable. For values of $\theta < 0.5$ the solution is only stable for time increments less than a critical value. For example, if $\theta = 0$ we obtain

$$\phi_{n+1} = [1 - \lambda\,\Delta t]\,\phi_n$$

and $\Delta t_{cr} = 2/\lambda$.

For the multi-degree-of-freedom problem it is necessary to compute the *largest* λ from the eigenproblem

$$\mathbf{C\Phi} = \mathbf{H\Phi}\Lambda$$

Generally, however, we will find it advantageous to use unconditionally stable forms with $\theta \geq 1/2$ when solving quasi-harmonic equations forms.

Example 3.11. One-dimensional steady-state heat conduction

The problem here will be a one-dimensional representation of the heat conduction equation [Eq. (3.94)] with unit conductivity (this problem could equally well represent many other physical situations, e.g., deflection of a loaded string with unit tension). Here we let (see Fig. 3.9)

$$-\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} + Q(x) = 0 \quad (0 < x < L)$$

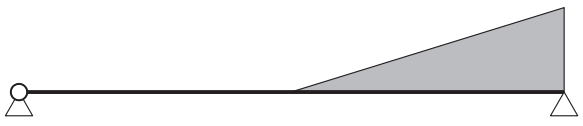
with $Q(x)$ given by

$$Q(x) = \begin{cases} 0 & 0 < x \leq L/2 \\ -2\,Q_0\,(x/L - 1/2) & L/2 < x < L \end{cases}$$

The boundary conditions assumed will be simply $\phi = 0$ at $x = 0$ and $x = L$.

The problem is first solved using two-node isoparametric elements to form $\mathbf{H}$ and $\mathbf{s}$. Assembly of four equal size elements results, after inserting the boundary conditions $\tilde{\phi}_1 = \tilde{\phi}_5 = 0$, in the equation set

$$\frac{4}{L}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{Bmatrix} \tilde{\phi}_2 \\ \tilde{\phi}_3 \\ \tilde{\phi}_4 \end{Bmatrix} = \frac{Q_0 L}{48}\begin{Bmatrix} 0 \\ 1 \\ 6 \end{Bmatrix}$$



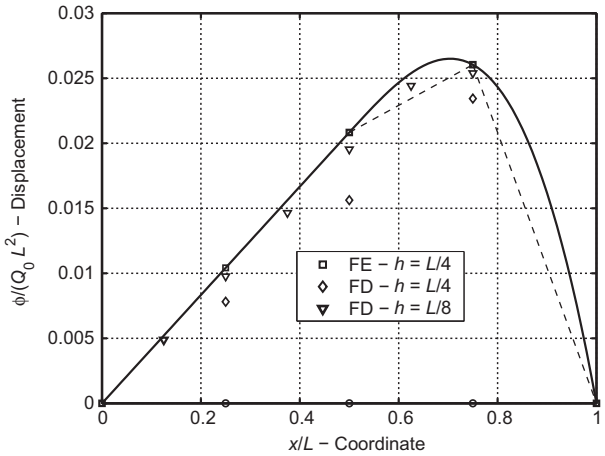


FIGURE 3.10
One-dimensional heat conduction. Solution by finite element method with linear elements and $h = L/4$; finite difference method with $h = L/4$ and $h = L/8$.

The solution using two-node elements is shown in Fig. 3.10 along with the exact solution to the problem. For comparison purposes we also show a finite difference solution in which simple collocation is used in a weighted residual equation together with the approximation for the second derivative given by a Taylor expansion

$$\left.\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}\right|_{x_a} \approx \frac{1}{h^2}\left(\tilde{\phi}_{a-1} - 2\tilde{\phi}_a + \tilde{\phi}_{a+1}\right)$$

which yields the approximation for each node point

$$\frac{1}{h^2}\left(-\tilde{\phi}_{a-1} + 2\tilde{\phi}_a - \tilde{\phi}_{a+1}\right) + Q_a = 0$$

After including the boundary conditions a set of three equations for the points 2, 3, and 4 is expressed as

$$\frac{16}{L^2}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{Bmatrix} \tilde{\phi}_2 \\ \tilde{\phi}_3 \\ \tilde{\phi}_4 \end{Bmatrix} = Q_0 \begin{Bmatrix} 0 \\ 0 \\ 1/2 \end{Bmatrix}$$

The reader will note that the coefficient matrix for the finite element and finite difference methods differs by only a constant multiplier (for the boundary conditions assumed in this one-dimensional problem); however, the right-hand sides differ significantly. Here we note that the nodal results for the finite element method are *exact* whereas those for the finite difference solution are all in error (although convergence can be observed for the finer subdivision). The nodal exactness is a property of the particular equation being solved and unfortunately does not carry over to general problems [20] (see also Appendix G). However, based on the above result

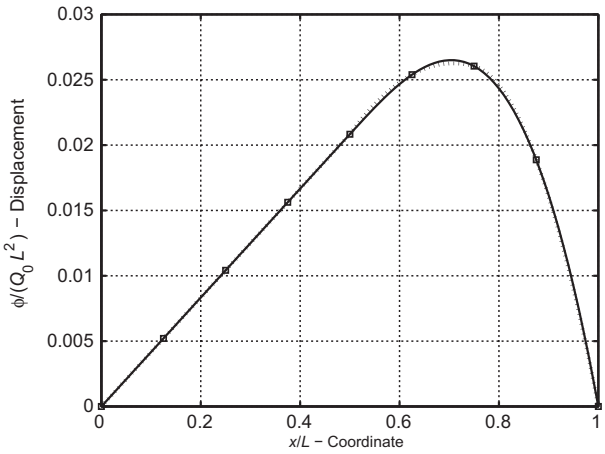


FIGURE 3.11

One-dimensional heat conduction. Solution by finite element method with quadratic elements and $h = L/4$.

and other experiences we can say that the finite element method always achieves (the same or) better results than classical finite difference methods. In addition, the finite element method permits an approximation of the solution *at all points in the domain* as indicated by the dashed lines in Fig. 3.10 for the one-dimensional problem.

The problem is repeated using 4-quadratic order finite elements and the results are shown in Fig. 3.11. It is evident that the use of quadratic order greatly increases the accuracy of the results obtained. Indeed, if cubic order elements were used results would be exact, since for linear varying Q the solution over the loaded portion will only contain polynomials up to cubic order.

3.10 Concluding remarks

In this chapter we have presented all the steps necessary to perform a finite element analysis of any problem for which a differential equation is known. The presentation has been simplified to equations with one coordinate for simplicity. However, as we shall observe in the later chapters, the basic steps used are general and apply equally to higher dimensional problems.

We also include in the development the possibility of considering transient problems. Using weak forms as the basis for a finite element discretization no additional complexity arises and we can consider both transient and static (steady-state) applications. The solution to transient problems may be addressed using discrete time methods, such as the Newmark method, or for linear forms by methods based on semi-discretization using modal methods.

The list of problem types that can be considered based on the presentation in this chapter is large and the examples presented cover only a small portion of possibilities.

In the next chapter we present an alternative to weak forms which may also be used with finite element approximation. Once both of these chapters are fully understood, the reader should be able to consider many additional classes of problems. In subsequent chapters we concentrate on problems in solid and structural mechanics and develop the methodology to construct the necessary finite element approximations for a wide class of applications.

3.11 Problems

3.1 Write weak forms for the following differential equations and boundary conditions. For each form state appropriate continuity conditions for approximations to the dependent variable u and the weighting function v. The domain for each one-dimensional differential equation is $0 < x < 1$.

(a)

$$a\frac{\mathrm{d}u}{\mathrm{d}x} + cu + q = 0, \quad u(0) = \bar{g}$$

(b)

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(a\frac{\mathrm{d}u}{\mathrm{d}x}\right) + q = 0, \quad u(0) = \bar{g} \quad \text{and} \quad a\frac{\mathrm{d}u}{\mathrm{d}x} + ku = \bar{g}, \ \text{at } x = 1$$

(c)

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(a\frac{\mathrm{d}u}{\mathrm{d}x}\right) + b\frac{\mathrm{d}u}{\mathrm{d}x} + q = 0, \quad u(0) = \bar{g}_0, \quad u(1) = \bar{g}_2$$

(d)

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(a\frac{\mathrm{d}^2u}{\mathrm{d}x^2}\right) + f = 0,$$
$$u(0) = \bar{g}_0, \ \left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_{x=0} = \bar{h}_0 \quad \text{and} \quad u(1) = \bar{g}_1$$

3.2 The differential equations for bending of a beam are given by

$$(1)\ \frac{\mathrm{d}V}{\mathrm{d}x} + q = 0 \qquad (2)\ \frac{\mathrm{d}M}{\mathrm{d}x} + V = 0$$
$$(3)\ \frac{\mathrm{d}\theta}{\mathrm{d}x} - \frac{M}{EI} = 0 \quad (4)\ \frac{\mathrm{d}w}{\mathrm{d}x} - \theta - \frac{V}{GA} = 0$$

in which V is shear force, M is moment, θ is section rotation, w is displacement, EI is bending stiffness, GA is shear stiffness, and q is load as shown in Fig. 3.12.

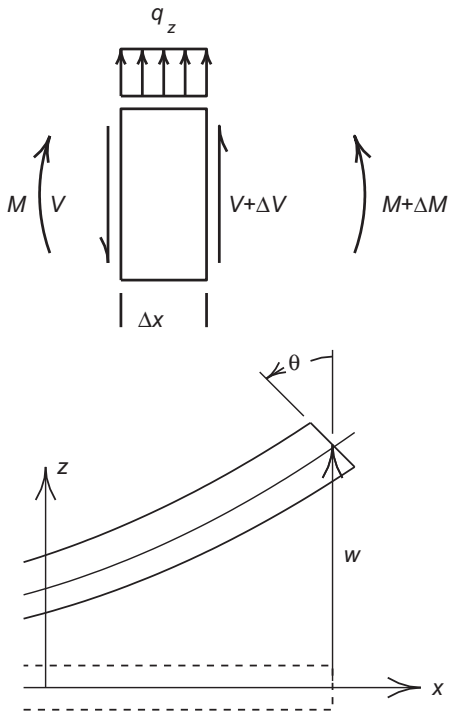


FIGURE 3.12
Beam bending description.

Boundary conditions are given by

$$(1)\ V = \overline{V} \quad \text{or } w = \bar{w}$$
$$(2)\ M = \overline{M} \quad \text{or } \theta = \bar{\theta}$$

Construct a weak form for the beam equations by multiplying (1) by $\pm\delta w$, (2) by $\pm\delta\theta$, (3) by δM, and (4) by δV.
Choose the correct sign for δw and $\delta\theta$ to give symmetry.

3.3 Add all boundary conditions to the weak form obtained in Problem 3.2.

3.4 For $GA = \infty$ (no shear deformation) deduce the irreducible differential equation in terms of w. Express all boundary conditions in terms of w.

3.5 Construct a weak form for Problem 3.4. What is the required continuity of the dependent variable needed for approximation by a finite element method? What are the natural and essential boundary conditions for the weak form?

3.6 For $GA = \infty$ (no shear deformation) deduce the differential equations in terms of w and M. Express all boundary conditions in terms of these variables.

3.7 Deduce a weak form for Problem 3.6 that permits approximation using C_0 functions to approximate w and M. Let

$$w = \sum_{a=1}^{2} N_a \tilde{w}_a \quad \text{and} \quad M = \sum_{a=1}^{2} N_a \tilde{M}_a$$

where N_a are given by (3.46). Ensure your weak form gives a symmetric coefficient matrix for these approximations.
Compute typical element matrices $\mathbf{K}$ and $\mathbf{f}$ for an element of length h with constant EI and q in the element.

3.8 For a simply supported beam of length 10 and constant cross-section $EI = 3$ compute the solution for a uniform load of $q = 1$. The boundary conditions at each end of the beam for a simple support are $w = M = 0$. Obtain a solution using 2, 4, and 8 elements. It is recommended that a small computer program be written using a high level language, e.g. MATLAB [21] or GNU Octave [22], to perform the numerical calculations. Compare your results to an exact solution.

3.9 The transient heat equation in one-dimension is given by

$$-\frac{\partial}{\partial x}\left(k\,\frac{\partial \phi}{\partial x}\right) + Q + c\,\frac{\partial \phi}{\partial t} = 0$$

where ϕ is temperature, k is thermal conductivity, Q is heat generation per unit length, and c is specific heat.
Boundary conditions may be given as

$$\phi = \bar{\phi} \quad \text{on} \quad \Gamma_1 \quad \text{or} \quad q = -k\,\frac{\partial \phi}{\partial x} = \bar{q} \quad \text{on} \quad \Gamma_2$$

where q is the heat flux and $\bar{\phi},\ \bar{q}$ are specified values. Initial conditions are given as $\phi(x,\ 0) = \bar{\phi}_0(x)$.

(a) Construct a weak form for the problem.

(b) Using the shape functions given in Eq. (3.26a) and the approximation

$$u^e = N_1(x)\tilde{u}_1(t) + N_2(x)\tilde{u}_2(t)$$
$$\delta u^e = N_1(x)\delta\tilde{u}_1 + N_2(x)\delta\tilde{u}_2$$

construct the semi-discrete form for a typical element of length h.

(c) Consider a region of length 10, with properties $k = 5, c = 1, Q = 0$. Divide the region into four equal length elements and establish the set of global semi-discrete equations.

(d) Consider a set of discrete times t_n. Approximate time derivatives of nodal values by $\mathrm{d}\phi/\mathrm{d}t(t_n) \approx (\phi_n - \phi_{n-1})/\Delta t$ where ϕ_n is the approximation to $\phi(t_n)$ and $\Delta t = t_n - t_{n-1}$ and write the fully discrete equations.
Write a computer program (e.g., using MATLAB or GNU Octave) to solve the problem. Assume the initial temperature of the region is zero and boundary conditions $\phi(0) = 0$ and $\phi(10) = 1$ are applied at time zero and held constant. Solve the problem using 10 steps with $\Delta t = 0.01$, followed by 9 steps with $\Delta t = 0.1$ and finally 9 steps with $\Delta t = 1$. Plot the finite element solution for ϕ vs. x at times 0.01, 0.1, 1.0, and 10.0.
Replace the element matrix associated with c by a diagonal (lumped) form with $c\,h/2$ on each diagonal ($h = x_2^e - x_1^e$). Repeat the above solution and compare results with the consistent form for the matrix.

References

[1] O.C. Zienkiewicz, W.L. Wood, N.W. Hine, R.L. Taylor, A unified set of single-step algorithms. Part 1: general formulation and applications, International Journal for Numerical Methods in Engineering 20 (1984) 1529–1552.

[2] W.L. Wood, A unified set of single-step algorithms. Part 2: theory, International Journal for Numerical Methods in Engineering 20 (1984) 2302–2309.

[3] M. Katona, O.C. Zienkiewicz, A unified set of single-step algorithms. Part 3: the beta-m method, a generalization of the Newmark scheme, International Journal for Numerical Methods in Engineering 21 (1985) 1345–1359.

[4] G.E. Forsythe, W.R. Wasow, Finite Difference Methods for Partial Differential Equations, John Wiley & Sons, New York, 1960.

[5] F.B. Hildebrand, Finite Difference Equations and Simulations, Prentice-Hall, Englewood Cliffs, N.J., 1968.

[6] F.B. Hildebrand. Introduction to Numerical Analysis, Dover Publishers, New York, second ed., 1987.

[7] A.R. Mitchell, D. Griffiths, The Finite Difference Method in Partial Differential Equations, John Wiley & Sons, London, 1980.

[8] R.A. Frazer, W.P. Jones, and S.W. Sken. Approximations to functions and to the solution of differential equations, Technical Report 1799, Aero. Research Committee, Report, 1937.

[9] C.B. Biezeno, R. Grammel, Technishe Dynamik, Springer-Verlag, Berlin, 1933, p. 142.

[10] O.C. Zienkiewicz, E. Oñate, Finite elements versus finite volumes. Is there a choice? in: P Wriggers, W Wagner (Eds.), Nonlinear Computational Mechanics, State of the Art, Springer, Berlin, 1991.

[11] B.G. Galerkin, Series solution of some problems in elastic equilibrium of rods and plates, Vestnik Inzhenerov i Tekhnikov 19 (1915) 897–908.

[12] W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, third ed., 1976.

[13] O.C. Zienkiewicz, R.L. Taylor, P. Nithiarasu, The Finite Element Method for Fluid Dynamics, Elsevier, Oxford, seventh ed., 2013.

[14] N. Newmark, A method of computation for structural dynamics, Journal of Engineering Mechanics ASCE 85 (1959) 67–94.

[15] G.L. Goudreau, R.L. Taylor, Evaluation of numerical integration methods in elastodynamics, Computer Methods in Applied Mechanics and Engineering 2 (1972) 69–97.

[16] T.J.R. Hughes, The Finite Element Method: Linear Static and Dynamic Analysis, Dover Publications, New York, 2000.

[17] T.J.R. Hughes, Analysis of transient algorithms with particular reference to stability behavior, in: T Belytschko, TJR Hughes (Eds.), Computational Methods for Transient Analysis, North-Holland, Amsterdam, 1983.

[18] T.J.R. Hughes, K.S. Pister, R.L. Taylor, Implicit-explicit finite elements in nonlinear transient analysis, Computer Methods in Applied Mechanics and Engineering 17–18 (1979) 159–182.
[19] D.J. Benson, Stable time step estimation for multi-material Eulerian hydrocodes, Computer Methods in Applied Mechanics and Engineering 167 (1998) 191–205.
[20] P. Tong, Exact solution of certain problems by the finite element method, Journal of AIAA 7 (1969) 179–180.
[21] MATLAB, 2012. <www.mathworks.com>.
[22] GNU Octave, 2012. <www.gnu.org/software/octave>.

CHAPTER

4 Variational Forms and Finite Element Approximation: 1-D Problems

4.1 Variational principles

In the previous chapter we described a weak form as an integral expression of a differential equation. For problems in which time derivative terms are not present an alternative integral expression known as a *variational principle* often exists. Variational principles can also be used as a basis to construct finite element solutions.

First, a definition: A "variational principle" specifies a scalar quantity (a functional) Π, which for the one-dimensional problems considered here is defined by an integral form

$$\Pi = \int_{\Omega} F\left(u, \frac{\mathrm{d}u}{\mathrm{d}x}, \dots\right)\mathrm{d}x + E\left(u, \frac{\mathrm{d}u}{\mathrm{d}x}, \dots\right)\Big|_{\Gamma} \tag{4.1}$$

in which u is the unknown function and F and E are specified differential operators. The solution to the continuous problem is a function u which makes Π *stationary* with respect to arbitrary changes δu. Thus, for a solution to the continuous problem, the "variation" is

$$\delta\Pi = 0 \tag{4.2}$$

for any δu, which defines the condition of stationarity [1].

If a "variational principle" can be found, then means are immediately established for obtaining approximate solutions in the standard, integral form suitable for finite element analysis.

Dividing the domain into elements and assuming a trial function expansion in the usual form [Eq. (3.37)]

$$u^e \approx \hat{u}^e = \sum_{a=1}^{n} N_a \tilde{u}_a = \mathbf{N}\tilde{\mathbf{u}}$$

we can insert this into (4.1) and write

$$\begin{aligned} \delta\Pi &= \sum_{e=1}^{M} \delta\Pi^e = 0 \\ \delta\Pi^e &= \frac{\partial\Pi^e}{\partial\tilde{u}_1}\delta\tilde{u}_1 + \frac{\partial\Pi^e}{\partial\tilde{u}_2}\delta\tilde{u}_2 + \cdots + \frac{\partial\Pi^e}{\partial\tilde{u}_n}\delta\tilde{u}_n \end{aligned} \tag{4.3}$$

The Finite Element Method: Its Basis and Fundamentals. http://dx.doi.org/10.1016/B978-1-85617-633-0.00004-6

This being true for any variations $\delta\tilde{\mathbf{u}}$ yields a set of equations for the system

$$\frac{\partial \Pi}{\partial \tilde{\mathbf{u}}} = \begin{Bmatrix} \dfrac{\partial \Pi}{\partial \tilde{u}_1} \\ \vdots \\ \dfrac{\partial \Pi}{\partial \tilde{u}_n} \end{Bmatrix} = \mathbf{0} \tag{4.4}$$

from which parameters $\tilde{u}_a$ are found. The equations are of an integral form necessary for the finite element approximation as the original specification of Π was given in terms of domain and boundary integrals.

The process of finding stationarity with respect to trial function parameters $\tilde{\mathbf{u}}$ is an old one and is associated with the names of Rayleigh [2] and Ritz [3].

If the functional Π is "quadratic," i.e., if the function u and its derivatives occur in powers not exceeding 2, then (4.4) reduces to a standard linear form similar to (3.20), i.e.,

$$\frac{\partial \Pi}{\partial \tilde{\mathbf{u}}} \equiv \mathbf{K}\tilde{\mathbf{u}} - \mathbf{f} = \mathbf{0} \tag{4.5}$$

It is easy to show that the matrix $\mathbf{K}$ will now always be symmetric. To do this let us consider a linearization of the vector $\partial\Pi/\partial\tilde{\mathbf{u}}$. This we can write as

$$\Delta\left(\frac{\partial \Pi}{\partial \tilde{\mathbf{u}}}\right) = \begin{Bmatrix} \frac{\partial}{\partial \tilde{u}_1}\left(\frac{\partial \Pi}{\partial \tilde{u}_1}\right)\Delta\tilde{u}_1 + \frac{\partial}{\partial \tilde{u}_2}\left(\frac{\partial \Pi}{\partial \tilde{u}_1}\right)\Delta\tilde{u}_2 + \cdots \\ \vdots \end{Bmatrix} \equiv \mathbf{K}_{\mathrm{T}}\Delta\tilde{\mathbf{u}} \tag{4.6}$$

in which $\mathbf{K}_{\mathrm{T}}$ is generally known as the tangent matrix, of significance in nonlinear analysis, and $\Delta\tilde{\mathbf{u}}$ are small incremental changes to $\tilde{\mathbf{u}}$. Now it is easy to see that

$$K_{\mathrm{T}ab} = \frac{\partial^2 \Pi}{\partial \tilde{u}_a \partial \tilde{u}_b} = K_{\mathrm{T}ba} \tag{4.7}$$

Hence $\mathbf{K}_{\mathrm{T}}$ is symmetric.

For a quadratic functional we have, from (4.5), $\mathbf{K}_{\mathrm{T}} = \mathbf{K}$, a constant matrix.

The fact that *symmetric matrices will arise whenever a variational principle exists is one of the most important merits of variational approaches for discretization*. However, symmetric forms will frequently arise directly from the Galerkin process. In such cases we simply conclude that the variational principle exists but we shall not need to use it directly since then it is automatically known that

$$\delta\Pi(u, \delta u) \equiv \mathrm{G}(u, \delta u) = 0 \tag{4.8}$$

Further, the discovery of symmetry from a weighted residual process leads directly to variational principles [4–6].

We note that frequently the physical aspects of a problem can be stated directly in a variational principle form. Theorems such as minimization of total potential energy

to achieve equilibrium in mechanical systems, least energy dissipation principles in viscous flow, etc., may be known to the reader and are considered by many as the basis of the formulation. Variational principles of this kind are called "natural" ones but unfortunately they do not exist for all problems for which well-defined differential equations can be formulated.

There is another category of variational principles which we may call "contrived" and these can always be constructed for any specified problem, either by extending the number of unknown functions u using additional variables known as Lagrange multipliers, or by procedures imposing a higher degree requirement on continuity, such as in least squares problems. In subsequent sections we shall discuss some "natural" and "contrived" variational principles.

Before proceeding further it is worth noting that, in addition to symmetry occurring in equations derived by variational means, sometimes further motivation arises. When "natural" variational principles exist the quantity Π may be of specific interest itself. If this arises a variational approach possesses the merit of easy evaluation of this quantity.

The reader will observe that if the functional is "quadratic" and yields (4.5), then we can write the approximate "functional" Π simply as

$$\Pi = \tfrac{1}{2}\tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{K}\tilde{\mathbf{u}} - \tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{f} \tag{4.9}$$

Hence, upon inserting the solution

$$\mathbf{K}\tilde{\mathbf{u}} - \mathbf{f} = \mathbf{0}$$

into (4.9) we obtain

$$\Pi = -\frac{1}{2}\tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{f} = -\frac{1}{2}\tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{K}\tilde{\mathbf{u}}$$

4.2 "Natural" variational principles and their relation to governing differential equations

4.2.1 Euler equations

If we consider the definitions of (4.1) and (4.2) we observe that for stationarity we can write, after performing some differentiations and integrations by parts,

$$\delta\Pi = \int_{\Omega} \delta u \mathcal{A}(u)\mathrm{d}x + \delta u \mathcal{B}(u)|_{\Gamma} = 0 \tag{4.10}$$

As the above has to be true for any variations δu, we must have

$$\mathcal{A}(u) = 0 \quad \text{in } \Omega \quad \text{and} \quad \mathcal{B}(u) = 0 \quad \text{on } \Gamma \tag{4.11}$$

If $\mathcal{A}$ corresponds precisely to the differential equations governing a problem of interest and $\mathcal{B}$ to its boundary conditions, then the variational principle is a *natural* one. Similar

to the result from a weak form, (4.11) are known as the *Euler equations* corresponding to the variational principle requiring the stationarity of Π. It is easy to show that for any variational principle a corresponding set of Euler equations can be established. The reverse is unfortunately not true, i.e., only certain forms of differential equations are Euler equations of a variational functional. In the next section we shall consider the conditions necessary for the existence of variational principles and give a prescription for establishing the Π from a set of suitable linear differential equations. In this section we shall continue to assume that the form of the variational principle is known.

Example 4.1. Heat equation in one dimension

To illustrate the process let us now consider a specific example. Suppose we specify a problem by requiring the stationarity of a functional

$$\Pi = \int_\Omega \left[\frac{1}{2}k\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)^2 - Q\phi\right]\mathrm{d}x + \left[\bar{q}_n\phi + H\left(\frac{1}{2}\phi^2 - \phi_0\phi\right)\right]\Big|_{\Gamma_q} \tag{4.12}$$

in which k and Q depend only on position and we assume $\phi = \bar{\phi}$ is satisfied on Γ_ϕ. We now perform the variation. This can be written following the rules of differentiation as [1]

$$\delta\Pi = \int_\Omega \left[k\frac{\mathrm{d}\phi}{\mathrm{d}x}\delta\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right) - Q\delta\phi\right]\mathrm{d}x + \delta\phi\left[\bar{q}_n + H(\phi - \phi_0)\right]\Big|_{\Gamma_q} = 0 \tag{4.13}$$

As

$$\delta\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right) = \frac{\mathrm{d}}{\mathrm{d}x}(\delta\phi)$$

we can integrate by parts (as in Section 3.9) and, since $\delta\phi = 0$ on Γ_ϕ, obtain

$$\delta\Pi = \int_\Omega \delta\phi\left[-\frac{\mathrm{d}}{\mathrm{d}x}\left(k\frac{\mathrm{d}\phi}{\mathrm{d}x}\right) - Q\right]\mathrm{d}x + \delta\phi\left(k\frac{\mathrm{d}\phi}{\mathrm{d}n} + \bar{q}_n + H(\phi - \phi_0)\right)\Big|_{\Gamma_q} = 0 \tag{4.14a}$$

This is of the form of (4.10) and we immediately observe that the Euler equations are

$$\begin{aligned} \mathcal{A}(\phi) &= -\frac{\mathrm{d}}{\mathrm{d}x}\left(k\frac{\mathrm{d}\phi}{\mathrm{d}x}\right) - Q = 0 \quad \text{in } \Omega \\ \mathcal{B}(\phi) &= k\frac{\mathrm{d}\phi}{\mathrm{d}n} + \bar{q}_n + H(\phi - \phi_0) = 0 \quad \text{on } \Gamma_q \end{aligned} \tag{4.14b}$$

If ϕ is prescribed so that $\phi = \bar{\phi}$ on Γ_ϕ and $\delta\phi = 0$ on that boundary, then the problem is precisely the one we have already discussed in Section 3.9 and the functional (4.12) specifies the one-dimensional *quasi-harmonic form* for the heat conduction problem in an alternative way.

In this case we have "guessed" the functional but the reader will observe that the variation operation could have been carried out for any functional specified and corresponding *Euler* equations could have been established.

Example 4.2. Variational finite element solution

Let us continue the process to obtain an approximate solution of the linear heat conduction problem for the case where the convection coefficient H is zero. Taking, as usual,

$$\phi^e \approx \hat{\phi}^e = \sum_a N_a \tilde{\phi}_a = \mathbf{N}\tilde{\boldsymbol{\phi}} \tag{4.15}$$

we substitute this approximation into the expression for the functional Π [Eq. (4.12)] and obtain

$$\Pi^e = \int_{\Omega_e} \frac{1}{2} k \left(\sum_a \frac{\mathrm{d}N_a}{\mathrm{d}x} \tilde{\phi}_a \right)^2 \mathrm{d}x - \int_{\Omega_e} Q \sum_a N_a \tilde{\phi}_a \mathrm{d}x + \bar{q}_n \sum_a N_a \tilde{\phi}_a \Big|_{\Gamma_{qe}} \tag{4.16}$$

On differentiation with respect to a typical parameter $\tilde{\phi}_b$ we have

$$\frac{\mathrm{d}\Pi^e}{\mathrm{d}\tilde{\phi}_b} = \int_{\Omega_e} k \left(\sum_a \frac{\mathrm{d}N_a}{\mathrm{d}x} \tilde{\phi}_a \right) \frac{\mathrm{d}N_b}{\mathrm{d}x} \mathrm{d}x - \int_{\Omega_e} Q N_b \mathrm{d}x + \bar{q}_n N_b |_{\Gamma_{qe}} \tag{4.17}$$

where after summing over all the elements and setting the $\delta\Pi = 0$ we obtain the system of equations for the solution of the problem

$$\mathbf{H}\tilde{\boldsymbol{\phi}} - \mathbf{s} = \mathbf{0} \tag{4.18}$$

with element arrays given by

$$\begin{aligned} H^e_{ab} &= \int_{\Omega_e} k \frac{\mathrm{d}N_a}{\mathrm{d}x} \frac{\mathrm{d}N_b}{\mathrm{d}x} \mathrm{d}x = H^e_{ba} \\ s^e_b &= \int_{\Omega_e} N_b Q \mathrm{d}x - N_b \bar{q}_n |_{\Gamma_{qe}} \end{aligned} \tag{4.19}$$

The reader will observe that the approximation equations are here identical to those obtained in Section 3.9.2 for the same problem using the Galerkin process. No special advantage accrues to the variational formulation here, and indeed we can predict now that *Galerkin and variational procedures must give the same answer for cases where natural variational principles exist*.

4.3 Establishment of natural variational principles for linear, self-adjoint differential equations

General rules for deriving natural variational principles from nonlinear differential equations are complicated and even the tests necessary to establish the existence of such variational principles are not simple. Much mathematical work has been done in this context by Vainberg [7], Tonti [6], Oden [8,9], and others.

For linear differential equations the situation is much simpler and a study is available in the works of Mikhlin [4,5], and in this section a brief presentation of such rules is given.

We shall consider here only the establishment of variational principles for a linear system of equations with *essential* (forced) boundary conditions, implying only variation of functions which yield $\delta\mathbf{u} = \mathbf{0}$ on their boundaries.

Writing a linear system of differential equations as

$$\mathcal{A}(u) \equiv \mathcal{L}u + b = 0 \tag{4.20}$$

in which $\mathcal{L}$ is a linear differential operator, natural variational principles require that the operator $\mathcal{L}$ be such that

$$\int_\Omega \psi(\mathcal{L}\gamma)\mathrm{d}x = \int_\Omega \gamma(\mathcal{L}\psi)\mathrm{d}x + \text{b.t.} \tag{4.21}$$

for any two function sets ψ and γ. In the above, "b.t." stands for boundary terms which we disregard in the present context. The property required in the above operator is called *self-adjointness* or *symmetry*, which we mentioned in Section 3.2.1.

If the operator $\mathcal{L}$ is self-adjoint, the variational principle can be written immediately as

$$\Pi = \int_\Omega \left[\frac{1}{2}u(\mathcal{L}u) + ub\right]\mathrm{d}x + \text{b.t.} \tag{4.22}$$

To prove the veracity of the last statement a variation needs to be considered. We thus write (omitting boundary terms)

$$\delta\Pi = \int_\Omega \left[\frac{1}{2}\delta u(\mathcal{L}u) + \frac{1}{2}u\delta(\mathcal{L}u) + \delta ub\right]\mathrm{d}x = 0 \tag{4.23}$$

Noting that for any linear operator

$$\delta(\mathcal{L}u) \equiv \mathcal{L}\delta u \tag{4.24}$$

and that u and δu can be treated as the independent functions, by identity (4.21) we can write (4.23) as

$$\delta\Pi = \int_\Omega \delta u[\mathcal{L}u + b]\mathrm{d}x = 0 \tag{4.25}$$

We observe immediately that the term in the brackets, i.e., the Euler equation of the functional, is identical with the original equation postulated, and therefore the variational principle is verified.

The above gives a very simple test and a prescription for the establishment of natural variational principles for differential equations of the problem.

Again self-adjointness of the operator can be tested and found to be satisfied. We now write a functional of *mixed* type as

$$\begin{aligned}\Pi &= \int_\Omega \left[\tfrac{1}{2} \begin{Bmatrix} q \\ \phi \end{Bmatrix}^{\mathrm{T}} \left(\begin{bmatrix} 1 & \frac{\mathrm{d}}{\mathrm{d}x} \\ -\frac{\mathrm{d}}{\mathrm{d}x} & 0 \end{bmatrix} \begin{Bmatrix} q \\ \phi \end{Bmatrix} \right) + \begin{Bmatrix} q \\ \phi \end{Bmatrix}^{\mathrm{T}} \begin{Bmatrix} 0 \\ Q \end{Bmatrix} \right] \mathrm{d}x \\ &= \int_\Omega \left[\tfrac{1}{2} \left(q^2 + q \tfrac{\mathrm{d}\phi}{\mathrm{d}x} - \phi \tfrac{\mathrm{d}q}{\mathrm{d}x} \right) + \phi Q \right] \mathrm{d}x \end{aligned} \tag{4.32}$$

The verification of the correctness of the above, by executing a variation, is left to the reader. This is called a *mixed variational form* since more than one variable type exists.

These two examples illustrate the simplicity of application of the general expressions. The reader will observe that self-adjointness of the operator will generally exist if even orders of differentiation are present. For odd orders self-adjointness is only possible if the operator is a "skew"-symmetric matrix such as occurs in the second example.

4.4 Maximum, minimum, or a saddle point?

In discussing variational principles so far we have assumed simply that at the solution point $\delta\Pi = 0$ and the functional is *stationary*. It is often desirable to know whether Π is at a maximum, minimum, or simply at a "saddle point." If a maximum or a minimum is involved, then the approximation to Π will always be "bounded," i.e., will provide approximate values of Π which are either smaller or larger than the correct ones.[1] The bound in itself may be of practical significance in some problems.

When, in elementary calculus, we consider a stationary point of a function Π of one variable u, we investigate the rate of change of $\mathrm{d}\Pi$ with $\mathrm{d}u$ and write

$$\mathrm{d}(\mathrm{d}\Pi) = \mathrm{d}\left(\frac{\partial \Pi}{\partial u} \mathrm{d}u \right) = \frac{\partial^2 \Pi}{\partial u^2} (\mathrm{d}u)^2 \tag{4.33}$$

The sign of the second derivative determines whether Π is a minimum, maximum, or simply stationary (saddle point), as shown in Fig. 4.1. By analogy in the calculus of variations we shall consider changes of $\delta\Pi$. Noting the general form of this quantity given by (4.3) and the notion of the second derivative of (4.6) we can write, in terms of discrete parameters,

$$\delta(\delta\Pi) \equiv \delta\left(\frac{\partial \Pi}{\partial \tilde{\mathbf{u}}} \right)^{\mathrm{T}} \delta\tilde{\mathbf{u}} = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \delta\left(\frac{\partial \Pi}{\partial \tilde{\mathbf{u}}} \right) = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \left(\frac{\partial^2 \Pi}{\partial \tilde{\mathbf{u}} \partial \tilde{\mathbf{u}}} \delta\tilde{\mathbf{u}} \right) = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \mathbf{K}_{\mathrm{T}} \delta\tilde{\mathbf{u}} \tag{4.34}$$

If, in the above, $\delta(\delta\Pi)$ is always negative then Π is obviously reaching a maximum, if it is always positive then Π is a minimum, but if the sign is indeterminate this shows only the existence of a saddle point.

[1] Provided all integrals are exactly evaluated.

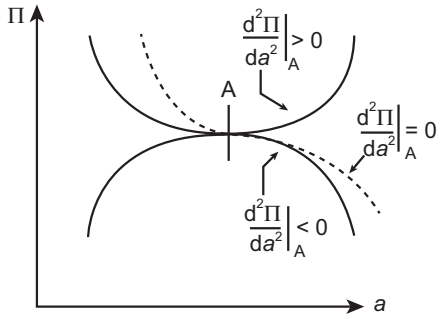


FIGURE 4.1
Maximum, minimum, and a "saddle" point for a functional Π of one variable.

As $\delta\tilde{\mathbf{u}}$ is an arbitrary vector this statement is equivalent to requiring the matrix $\mathbf{K}_T$ to be negative definite for a maximum *or* positive definite for a minimum. The form of the matrix $\mathbf{K}_T$ (or in linear problems of $\mathbf{K}$ which is identical to it) is thus of importance in the solution of variational problems. A matrix is positive definite if all its eigenvalues are positive. If the eigenvalues are positive or zero it is called positive semi-definite.

4.5 Constrained variational principles

4.5.1 Lagrange multipliers

Consider the problem of making a functional Π stationary, subject to the unknown u obeying an additional relationship

$$C(u) = 0 \quad \text{in } \Omega \tag{4.35}$$

We can introduce this constraint by forming another functional

$$\bar{\Pi}(u, \lambda) = \Pi(u) + \int_\Omega \lambda C(u) \mathrm{d}x \tag{4.36}$$

in which λ is some function of x known as a *Lagrange multiplier*. The variation of the new functional is now

$$\delta\bar{\Pi} = \delta\Pi + \int_\Omega \lambda \delta C(u) \mathrm{d}x + \int_\Omega \delta\lambda C(u) \mathrm{d}x = 0 \tag{4.37}$$

which immediately gives $C(u) = 0$ and, simultaneously, an added contribution to the original $\delta\Pi$ involving λ. We note that the above process could also be introduced into a weak form directly by merely appending the last two terms.

In a similar way, constraints can be introduced at some points of the domain. For instance, if we require that u obey

$$E(u) = 0 \quad \text{at } x = x_e \tag{4.38}$$

we would add to the original functional the term

$$\lambda E(u)|_{x=x_e} \tag{4.39}$$

with λ now being an unknown parameter defined only at $x = x_e$.

It appears, therefore, possible to always introduce a number of additional functions λ and modify a functional to include any prescribed constraint. In the "discretization" process we shall now have to use trial functions to describe both u and λ. Writing, for instance,

$$\hat{u}(x,t) = \sum_a \mathrm{N}_a \tilde{u}_a = \mathbf{N}(x)\tilde{\mathbf{u}}(t); \quad \hat{\lambda}(x,t) = \sum_b \bar{N}_b \tilde{\lambda}_b = \bar{\mathbf{N}}(x)\tilde{\boldsymbol{\lambda}}(t) \tag{4.40}$$

we shall obtain a set of equations

$$\frac{\partial \Pi}{\partial \mathbf{w}} = \begin{Bmatrix} \dfrac{\partial \Pi}{\partial \tilde{\mathbf{u}}} \\ \dfrac{\partial \Pi}{\partial \tilde{\boldsymbol{\lambda}}} \end{Bmatrix} = \mathbf{0} \quad \text{where} \quad \mathbf{w} = \begin{Bmatrix} \tilde{\mathbf{u}} \\ \tilde{\boldsymbol{\lambda}} \end{Bmatrix} \tag{4.41}$$

from which both sets of parameters $\tilde{\mathbf{u}}$ and $\tilde{\boldsymbol{\lambda}}$ can be obtained. It is somewhat paradoxical that the "constrained" problem has resulted in a larger number of unknown parameters than the original one and, indeed, has complicated the solution. We shall, nevertheless, find practical use for Lagrange multipliers in formulating some physical variational principles, and will make use of these in a more general context in Chapter 9.

Before proceeding further it is of interest to investigate the form of equations resulting from the modified functional Π of (4.36). If the original functional Π gave as its Euler equations a system

$$\boldsymbol{\mathcal{A}}(\mathbf{u}) = \mathbf{0} \tag{4.42}$$

then we have (omitting the boundary terms) for a system of constraints

$$\delta\bar{\Pi} = \int_\Omega \delta\mathbf{u}^{\mathrm{T}} \boldsymbol{\mathcal{A}}(\mathbf{u})\mathrm{d}x + \int_\Omega \delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\lambda}\mathrm{d}x + \int_\Omega \delta\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{C}(\mathbf{u})\mathrm{d}x = 0 \tag{4.43}$$

Substituting the trial functions (4.40) we can write for a linear set of constraints

$$\mathbf{C}(\mathbf{u}) = \boldsymbol{\mathcal{L}}_1\mathbf{u} + \mathbf{C}_1$$

that

$$\begin{aligned} \delta\bar{\Pi} = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \left[\int_\Omega \mathbf{N}^{\mathrm{T}} \boldsymbol{\mathcal{A}}(\hat{\mathbf{u}})\mathrm{d}x + \int_\Omega (\boldsymbol{\mathcal{L}}_1\mathbf{N})^{\mathrm{T}}\hat{\boldsymbol{\lambda}}\,\mathrm{d}x \right] \\ + \delta\tilde{\boldsymbol{\lambda}}^{\mathrm{T}} \int_\Omega \bar{\mathbf{N}}^{\mathrm{T}}(\boldsymbol{\mathcal{L}}_1\hat{\mathbf{u}} + \mathbf{C}_1)\mathrm{d}x = 0 \end{aligned} \tag{4.44}$$

As this has to be true for all variations $\delta\tilde{\mathbf{u}}$ and $\delta\tilde{\boldsymbol{\lambda}}$, we have a system of equations

$$\begin{aligned}\int_\Omega \mathbf{N}^{\mathrm{T}}\boldsymbol{\mathcal{A}}(\hat{\mathbf{u}})\mathrm{d}x + \int_\Omega (\boldsymbol{\mathcal{L}}_1\mathbf{N})^{\mathrm{T}}\hat{\boldsymbol{\lambda}}\,\mathrm{d}x &= \mathbf{0}\\ \int_\Omega \bar{\mathbf{N}}^{\mathrm{T}}(\boldsymbol{\mathcal{L}}_1\hat{\mathbf{u}} + \mathbf{C}_1)\mathrm{d}x &= \mathbf{0}\end{aligned} \tag{4.45}$$

For linear equations $\boldsymbol{\mathcal{A}}$, the first term of the first equation is precisely the ordinary, unconstrained, variational approximation

$$\mathbf{K}_{uu}\tilde{\mathbf{u}} + \mathbf{f}_u \tag{4.46}$$

and inserting again the trial functions (4.40) we can write the approximated (4.45) as a linear system:

$$\mathbf{K}_w\mathbf{w} = \begin{bmatrix}\mathbf{K}_{uu} & \mathbf{K}_{u\lambda}\\ \mathbf{K}_{u\lambda}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}\begin{Bmatrix}\tilde{\mathbf{u}}\\ \tilde{\boldsymbol{\lambda}}\end{Bmatrix} + \begin{Bmatrix}\mathbf{f}_u\\ \mathbf{f}_\lambda\end{Bmatrix} = \mathbf{0} \tag{4.47}$$

with

$$\mathbf{K}_{u\lambda}^{\mathrm{T}} = \int_\Omega \bar{\mathbf{N}}^{\mathrm{T}}(\boldsymbol{\mathcal{L}}_1\mathbf{N})\mathrm{d}x, \quad \mathbf{f}_\lambda = \int_\Omega \bar{\mathbf{N}}^{\mathrm{T}}\mathbf{C}_1\mathrm{d}x \tag{4.48}$$

Clearly the system of equations is symmetric but now possesses zeros on the diagonal, and therefore the variational principle Π is merely stationary. Further, computational difficulties may be encountered unless the solution process allows for zero diagonal terms.

4.5.2 Identification of Lagrange multipliers: Forced boundary conditions and modified variational principles

Although the Lagrange multipliers were introduced as a mathematical concept necessary for the enforcement of certain external constraints required to satisfy the original variational principle, we shall find that in many situations they can be identified with certain physical quantities of importance to the original mathematical model. Such an identification will follow immediately from the definition of the variational principle established in (4.36) and through the first of the Euler equations in (4.45) corresponding to it. The variation $\delta\bar{\Pi}$, written in (4.37), supplies through its third term the constraint equation. The first two terms can always be rewritten as

$$\int_\Omega \delta\mathbf{C}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}\,\mathrm{d}x + \int_\Omega \delta\mathbf{u}^{\mathrm{T}}\boldsymbol{\mathcal{A}}(\mathbf{u})\,\mathrm{d}x = \mathbf{0} \tag{4.49a}$$

and/or

$$\int_\Gamma \delta\mathbf{E}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}\,\mathrm{d}\Gamma + \int_\Gamma \delta\mathbf{u}^{\mathrm{T}}\boldsymbol{\mathcal{B}}(\mathbf{u})\,\mathrm{d}\Gamma = \mathbf{0} \tag{4.49b}$$

This supplies the identification of $\boldsymbol{\lambda}$.

In the literature of variational calculation such identification arises frequently and the reader is referred to the book by Washizu [10] for numerous examples.

Example 4.5. Identification of Lagrange multiplier for boundary condition

Here we shall introduce this identification by means of the example considered in Section 4.2.1. As we have noted, the variational principle of (4.12) established the governing equation and the natural boundary conditions of the heat conduction problem providing the essential boundary condition

$$E(\phi) = \phi - \bar{\phi} = 0 \tag{4.50}$$

was satisfied on Γ_ϕ in the choice of the trial function for ϕ.

The above forced boundary condition can, however, be considered as a constraint on the original problem. We can write the constrained variational principle as

$$\bar{\Pi} = \Pi + \lambda(\phi - \bar{\phi})\big|_{\Gamma_\phi} \tag{4.51}$$

where Π is given by (4.12).

Performing the variation we have

$$\delta\bar{\Pi} = \delta\Pi + \delta\phi\lambda|_{\Gamma_\phi} + \delta\lambda(\phi - \bar{\phi})\big|_{\Gamma_\phi} = 0 \tag{4.52}$$

$\delta\Pi$ is now given by expression (4.14a) augmented by a term

$$\delta\phi k \left.\frac{\mathrm{d}\phi}{\mathrm{d}n}\right|_{\Gamma_\phi} \tag{4.53}$$

which was previously disregarded (as we had assumed that $\delta\phi = 0$ on Γ_ϕ). In addition to the conditions of (4.14b), we now require that

$$\delta\lambda(\phi - \bar{\phi})\big|_{\Gamma_\phi} + \delta\phi\left(\lambda + k\frac{\mathrm{d}\phi}{\mathrm{d}n}\right)\Bigg|_{\Gamma_\phi} = 0 \tag{4.54}$$

which must be true for all variations $\delta\lambda$ and $\delta\phi$. The first simply reiterates the constraint

$$\phi - \bar{\phi} = 0 \quad \text{on } \Gamma_\phi \tag{4.55}$$

The second *defines* λ as

$$\lambda = -k\frac{\mathrm{d}\phi}{\mathrm{d}n} \tag{4.56}$$

Noting that $k(\partial\phi/\partial n)$ is the negative to the flux q_n on the boundary Γ_ϕ, the physical identification of the multiplier has been achieved—that is, $\lambda \equiv q_n$.

4.6 Constrained variational principles: Penalty function and perturbed Lagrangian methods

In the previous section we have seen how the process of introducing Lagrange multipliers allows constrained variational principles to be obtained at the expense of increasing the total number of unknowns. Further, we have shown that even in linear problems the algebraic equations which have to be solved are now complicated by having zero diagonal terms. In this section we shall consider alternative procedures of introducing constraints which do not possess these drawbacks.

4.6.1 Penalty functions

Considering once again the problem of obtaining stationarity of Π with a set of constraint equations $\mathbf{C}(\mathbf{u}) = \mathbf{0}$ in domain Ω, we note that the product

$$\mathbf{C}^{\mathrm{T}}\mathbf{C} = C_1^2 + C_2^2 + \cdots \tag{4.57}$$

where $\mathbf{C}^{\mathrm{T}} = [C_1, C_2, \ldots]$ must always be a quantity which is positive or zero. Clearly, the latter value is found when the constraints are satisfied and clearly the variation

$$\delta(\mathbf{C}^{\mathrm{T}}\mathbf{C}) = 0 \tag{4.58}$$

as the product reaches that minimum.

We can now write a new functional

$$\bar{\bar{\Pi}} = \Pi + \frac{1}{2}\alpha \int_\Omega \mathbf{C}^{\mathrm{T}}(\mathbf{u})\mathbf{C}(\mathbf{u})\mathrm{d}x \tag{4.59}$$

in which α is a "penalty number" and then require stationarity for the constrained solution. If Π is itself a minimum of the solution then α should be a positive number. The solution obtained by the stationarity of the functional $\bar{\bar{\Pi}}$ will satisfy the constraints only approximately. The larger the value of α the better will be the constraints achieved. Further, it seems obvious that the process is best suited to cases where Π is a minimum (or maximum) principle, but success can be obtained even with purely saddle point problems. The process is equally applicable to constraints applied on boundaries or simple discrete constraints. In this latter case integration is dropped.

Example 4.6. Linear constraint

Consider the linear constraint

$$C(u) = C_0 u + C_1$$

with the approximation for u given by (4.40). The added term to the variational theorem is given by

$$\int_\Omega [C(u)]^2 \mathrm{d}x = \tilde{\mathbf{u}}^{\mathrm{T}} \left[\int_\Omega \mathbf{N}^{\mathrm{T}} C_0^2 \mathbf{N}\, \mathrm{d}x \tilde{\mathbf{u}} + \int_\Omega \mathbf{N}^{\mathrm{T}} 2C_0 C_1\, \mathrm{d}x \right] + \int_\Omega C_1^2\, \mathrm{d}x$$

Upon variation the last term vanishes.

4.6.2 Perturbed Lagrangian

As an alternative to a penalty method we consider once again the problem of obtaining stationarity of Π with a constraint equation $C(u) = 0$ in domain Ω. The Lagrange multiplier form to embed the constraint is given in (4.36). Here we modify the expression by appending a quadratic term of the form λ^2 scaled by a parameter α. The form of the final equation is given by

$$\breve{\Pi}(u, \lambda) = \Pi(u) + \int_\Omega \lambda C(u)\mathrm{d}x - \frac{1}{2\alpha}\int_\Omega \lambda^2 \mathrm{d}x \tag{4.60}$$

We note that as the parameter α tends toward infinity the form approaches a Lagrange multiplier form. Accordingly, this form is called a *perturbed Lagrangian* functional. Taking the variation we obtain the result

$$\delta\breve{\Pi} = \delta\Pi + \int_\Omega \lambda\delta C(u)\mathrm{d}x + \int_\Omega \delta\lambda C(u)\mathrm{d}x - \frac{1}{\alpha}\int_\Omega \delta\lambda\lambda\,\mathrm{d}x = 0 \tag{4.61}$$

If the constraints are a linear form given by

$$C(u) = C_0\,u + C_1 = 0$$

we can introduce the approximations (4.40) into (4.61) along with the constant $\lambda = \tilde{\lambda}$ to obtain the set of equations

$$\begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{u\lambda} \\ \mathbf{K}_{\lambda u} & -\frac{1}{\alpha}K_{\lambda\lambda} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{u}} \\ \tilde{\lambda} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ f_\lambda \end{Bmatrix} \tag{4.62}$$

where $\mathbf{K}_{uu}$ is the coefficient array from $\delta\Pi$ and

$$\mathbf{K}_{u\lambda} = \int_\Omega \mathbf{N}^{\mathrm{T}} C_0 \mathrm{d}x = \mathbf{K}_{\lambda u}^{\mathrm{T}}, \quad K_{\lambda\lambda} = \int_\Omega \mathrm{d}x \quad \text{and} \quad f_\lambda = -\int_\Omega C_1\,\mathrm{d}x$$

The second equation of (4.62) may be solved for $\tilde{\lambda}$ in terms of $\tilde{\mathbf{u}}$ and substituted into the first equation to obtain

$$\bar{\mathbf{K}}_{uu}\tilde{\mathbf{u}} = \left[\mathbf{K}_{uu} + \alpha\mathbf{K}_{u\lambda}K_{\lambda\lambda}^{-1}\mathbf{K}_{\lambda u}\right]\tilde{\mathbf{u}} = \mathbf{f} + \alpha\mathbf{K}_{u\lambda}K_{\lambda\lambda}^{-1}f_\lambda$$

It is now apparent that the perturbed Lagrangian and penalty forms are closely related. The perturbed Lagrangian uses

$$\mathbf{K}_{u\lambda}K_{\lambda\lambda}^{-1}\mathbf{K}_{\lambda u}$$

to impose the constraint whereas the penalty method uses

$$\int_\Omega \mathbf{N}^{\mathrm{T}} C_0^2 \mathbf{N}\,\mathrm{d}x$$

When the constraint is a simple scalar relation at a single point the two methods are identical. However, when general forms $\mathbf{C}(\mathbf{u})$ are considered the methods will yield different approximations unless the shape functions for the set of $\boldsymbol{\lambda}$ include all the terms contained in $\delta\mathbf{C}(\mathbf{u})$.

In practical applications the method of penalty functions has proved to be quite effective [11], and indeed is often introduced intuitively.

Example 4.7. Boundary condition by penalty method

In the example presented next the forced boundary conditions are not introduced *a priori* and the problem gives, on assembly, a singular system of equations

$$\mathbf{K}\tilde{\mathbf{u}} - \mathbf{f} = \mathbf{0} \tag{4.63}$$

which can be obtained from the functional (providing $\mathbf{K}$ is symmetric)

$$\Pi = \frac{1}{2}\tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{K}\tilde{\mathbf{u}} - \tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{f} \tag{4.64}$$

Introducing a prescribed value of u_1, i.e., writing

$$C(u_1) = u_1 - \bar{u}_1 = 0 \tag{4.65}$$

the functional can be modified to

$$\bar{\bar{\Pi}} = \Pi + \frac{1}{2}\alpha(u_1 - \bar{u}_1)^2 \tag{4.66}$$

yielding

$$\bar{\bar{K}}_{11} = K_{11} + \alpha; \quad \bar{\bar{f}}_1 = f_1 - \alpha\bar{u}_1 \tag{4.67}$$

and giving no change in any of the other matrix coefficients.

The use of the penalty function in the finite element context presents certain difficulties.

Firstly, the constrained functional of (4.59) leads to equations of the form

$$(\mathbf{K}_1 + \alpha\mathbf{K}_2)\tilde{\mathbf{u}} - \bar{\bar{\mathbf{f}}} = \mathbf{0} \tag{4.68}$$

where $\mathbf{K}_1$ derives from the original functional and $\mathbf{K}_2$ from the constraints. As α increases $\mathbf{K}_1$ becomes insignificant compared to $\alpha\mathbf{K}_2$ and the above equation degenerates to

$$\mathbf{K}_2\tilde{\mathbf{u}} = \mathbf{f}/\alpha \to \mathbf{0}$$

and $\tilde{\mathbf{u}} = \mathbf{0}$ unless the matrix $\mathbf{K}_2$ is singular. The phenomenon where $\tilde{\mathbf{u}} \Rightarrow \mathbf{0}$ is known as *locking* and has often been encountered by researchers who failed to recognize its source. This singularity in the equations does not always arise and we shall discuss means of its introduction in Chapters 9 and 10.

Secondly, with large but finite values of α numerical difficulties will be encountered. Noting that discretization errors can be of comparable magnitude to those due to not *satisfying* the constraint, we can make

$$\alpha = \text{constant}(1/h)^n$$

(where h is an element size parameter) ensuring a limiting convergence to the correct answer. Fried [12,13] discusses this problem in detail.

A more general discussion of the whole topic is given in Ref. [14] and in Chapter 10 where the relationship between Lagrange constraints and penalty forms is made clear.

4.7 Least squares approximations

A general variational principle also may be constructed if the constraints described in the previous section are simply the governing equations of the problem

$$\mathbf{C}(\mathbf{u}) = \boldsymbol{\mathcal{A}}(\mathbf{u}) \tag{4.69}$$

Obviously the same procedure can be used in the context of the penalty function approach by setting $\Pi = 0$ in (4.59). We can thus write a "variational principle"

$$\bar{\bar{\Pi}} = \frac{1}{2}\int_\Omega \left(A_1^2 + A_2^2 + \cdots\right) \mathrm{d}x = \frac{1}{2}\int_\Omega \boldsymbol{\mathcal{A}}^{\mathrm{T}}(\mathbf{u})\boldsymbol{\mathcal{A}}(\mathbf{u})\mathrm{d}x \tag{4.70}$$

for any set of differential equations. In the above equation the boundary conditions are assumed to be satisfied by $\mathbf{u}$ (forced boundary condition) and the parameter α is dropped as it becomes a multiplier.

Clearly, the above statement is a requirement that the sum of the squares of the residuals of the differential equations should be a minimum at the correct solution. This minimum is obviously zero at that point, and the process is simply the well-known *least squares method* of approximation.

It is equally obvious that we could obtain the correct solution by minimizing any functional of the form

$$\bar{\bar{\Pi}} = \frac{1}{2}\int_\Omega \left(p_1 A_1^2 + p_2 A_2^2 + \cdots\right) \mathrm{d}x = \frac{1}{2}\int_\Omega \boldsymbol{\mathcal{A}}^{\mathrm{T}}(\mathbf{u})\mathbf{p}\boldsymbol{\mathcal{A}}(\mathbf{u})\mathrm{d}x \tag{4.71}$$

in which $p_1, p_2, \ldots$, etc., are positive valued weighting functions or constants and $\mathbf{p}$ is a diagonal matrix:

$$\mathbf{p} = \begin{bmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & p_3 & \\ 0 & & & \ddots \end{bmatrix} \tag{4.72}$$

The above alternative form is sometimes convenient as it places a different importance on the satisfaction of individual components of the equation set and allows additional freedom in the choice of the approximate solution. Once again this weighting function could be chosen so as to ensure a constant ratio of terms contributed by various equations.

A least squares method of the kind shown above is a very powerful alternative procedure for obtaining integral forms from which an approximate solution can be started, and has been used with considerable success [15–18]. As a least squares variational principle can be written for *any* set of differential equations without introducing additional variables, we may well inquire as to what the difference is between these and the *natural variational principles* discussed previously. On performing a variation in a specific case the reader will find that the Euler equations which are obtained

no longer give the original differential equations but give higher order derivatives of these. Thus, higher order continuity of trial functions is now generally needed. This may be a serious drawback but frequently can be bypassed by stating the original problem as a set of lower order equations. The appearance of higher order derivatives in the Euler equations also introduces the possibility of spurious solutions if incorrect boundary conditions are used.

Example 4.8. Least squares solution for Helmholtz equation

To illustrate the use of a least squares approach consider the Helmholtz problem governed by (4.26), for which we have already obtained a *natural* variational principle [Eq. (4.30)] in which only first derivatives were involved, requiring C_0 continuity for ϕ. Now, if we use the operator $\mathcal{L}$ and term b defined by (4.28), we have a set of approximating equations with

$$\begin{aligned} K_{ab} &= \int_\Omega \left(\frac{\mathrm{d}^2 N_a}{\mathrm{d}x^2} + cN_a\right)\left(\frac{\mathrm{d}^2 N_b}{\mathrm{d}x^2} + cN_b\right)\mathrm{d}x \\ f_a &= \int_\Omega \left(\frac{\mathrm{d}^2 N_a}{\mathrm{d}x^2} + cN_a\right) Q\,\mathrm{d}x \end{aligned} \tag{4.73}$$

The reader will observe that due to the presence of second derivatives C_1 continuity is now needed for the trial functions $\mathbf{N}$.

An alternative, avoiding the requirement of C_1 functions, is to write (4.26) as a first-order system. This can be written as

$$\mathcal{A}(\mathbf{u}) = \begin{Bmatrix} q + \dfrac{\mathrm{d}\phi}{\mathrm{d}x} \\ -\dfrac{\mathrm{d}q}{\mathrm{d}x} + c\phi + Q \end{Bmatrix} = 0 \tag{4.74}$$

or, introducing the vector $\mathbf{u}$,

$$\mathbf{u} = \begin{Bmatrix} q \\ \phi \end{Bmatrix} = (\mathbf{N}\tilde{\mathbf{u}}) \tag{4.75}$$

as the unknown we can write an approximation as

$$\mathbf{u} \approx \hat{\mathbf{u}} = \begin{bmatrix} \mathbf{N}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_\phi \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{q}} \\ \tilde{\boldsymbol{\phi}} \end{Bmatrix} = \mathbf{N}\tilde{\mathbf{u}} \tag{4.76}$$

where $\mathbf{N}_q$ and $\mathbf{N}_\phi$ are C_0 shape functions for the q and ϕ variables, respectively. The least squares approximation is now given by

$$\delta\bar{\bar{\Pi}} = \delta\tilde{\mathbf{u}}^{\mathrm{T}} \int_\Omega (\mathcal{L}\mathbf{N})^{\mathrm{T}}[(\mathcal{L}\mathbf{N})\tilde{\mathbf{u}} + \mathbf{b}]\mathrm{d}x = 0 \tag{4.77a}$$

where

$$\mathcal{L}\mathbf{N} = \begin{bmatrix} \mathbf{N}_q & \dfrac{\mathrm{d}\mathbf{N}_\phi}{\mathrm{d}x} \\ -\dfrac{\mathrm{d}\mathbf{N}_q}{\mathrm{d}x} & c\mathbf{N}_\phi \end{bmatrix}; \quad \mathbf{b} = \begin{Bmatrix} 0 \\ Q \end{Bmatrix} \tag{4.77b}$$

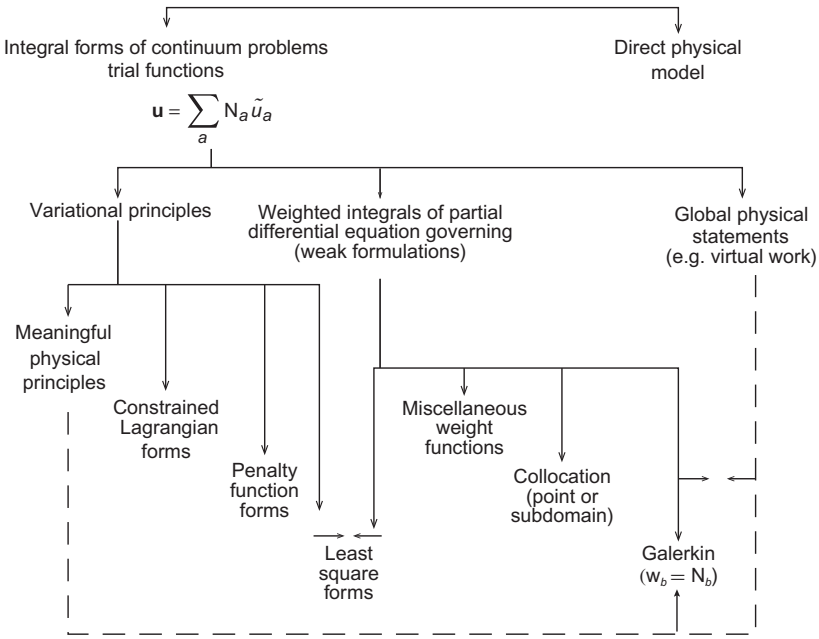
The reader can now perform the final steps to obtain the **K** and **f** matrices. The approximation equations in a form requiring only C_0 continuity are obtained, however, at the expense of additional variables. Use of such forms has been made extensively in the finite element context [15–22].

4.8 Concluding remarks: Finite difference and boundary methods

This chapter along with Chapter 3 present the general possibilities of using the finite element process in almost any mathematical or mathematically modeled physical problem. The essential approximation processes mostly have been given in a simple one-dimensional form. In the chapters that follow we shall apply many of these methods to multidimensional problems in mechanics and other fields. In some we shall show, however, that certain extensions of the process are possible. For example, in Chapter 8 we show how a violation of some of the rules here expounded can be accepted.

The numerous approximation procedures discussed fall into several categories. To remind the reader of these, we present in Table 4.1 a comprehensive catalog of the methods used here and in Chapter 3. The only aspect of the finite element process mentioned in this table that has not been discussed here is that of a *direct physical method*. In such models an "atomic" rather than continuum concept is the starting

Table 4.1 Finite Element Approximation



point. While much interest exists in the possibilities offered by such models, their discussion is outside the scope of this book.

In all the continuum processes discussed the first step is always the choice of suitable shape or trial functions. A few simple one-dimensional forms of such functions have been introduced as needs demanded and many new forms will be introduced in the next two chapters. Indeed, the reader who has mastered the essence of the last two chapters will have little difficulty in applying the finite element method to any suitably defined physical problem once appropriate shape functions are available.

The methods listed do not include specifically two well-known techniques, i.e., *finite difference* methods and *boundary solution* methods (sometimes known as boundary elements). In the general sense these belong to the category of a *generalized finite element* method [23]. Boundary solution methods choose the trial functions such that the governing equation is automatically satisfied in the domain Ω. In this class of problems only boundary terms remain to be satisfied. Finite difference procedures can be interpreted as an approximation based on local, discontinuous, shape functions with collocation weighting applied (although usually the derivation of the approximation algorithm is based on a Taylor expansion).

Many textbooks deal exclusively with these types of approximations. References [24–27] discuss finite difference approximation and Refs. [28–31] relate to boundary methods.

4.9 Problems

4.1 Deduce the Euler differential equation and boundary conditions for the variational principle expressed as

$$\Pi(u) = \int_a^b \left[EI \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 - Pu^2 \right] \mathrm{d}x - ug \Big|_{x=b}, \quad u(a) = 0$$

Classify Π as a minimum, maximum, or saddle point form.

4.2 Deduce the Euler differential equation and boundary conditions for the variational principle expressed as

$$\Pi(u) = \int_a^b \left[EA \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 + ku^2 - 2qu \right] \mathrm{d}x + \alpha[(u(a))^2 + (u(b))^2]$$

where EA and k are constant parameters and α is a penalty parameter.

4.3 Deduce the Euler equations and boundary conditions for the variational principle expressed as

$$\Pi(u, \lambda_a, \lambda_b) = \int_a^b \left[EA \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 + ku^2 - 2qu \right] \mathrm{d}x + \lambda_a u(a) + \lambda_b u(b)$$

where EA, k, and q are constant parameters and α is a penalty parameter.

4.4 Construct a variational theorem which gives the weak form obtained in Problems 3.2 and 3.3 as the first variation.

4.5 Construct a variational theorem which has the irreducible differential equation for the beam problem described by Problem 3.4 as its first variation.

4.6 Solve the one-dimensional heat equation given in Example 3.11 by enforcing the boundary conditions by the penalty formulation described by Example 4.7. How large must each penalty parameter be taken to make the boundary error less than $10^{-6}|\phi_{\max}|$?

References

[1] F.B. Hildebrand, Methods of Applied Mathematics, second ed., Prentice-Hall, 1965 (reprinted by Dover Publishers, 1992).

[2] Lord Rayleigh (J.W. Strutt), On the theory of resonance, Trans. Roy. Soc. (Lond.) A161 (1870) 77–118.

[3] W. Ritz, Über eine neue Methode zur Lösung gewisser variationsproblem der mathematischen physik, J. für die reine und angewandte Mathematik 135 (1908) 1–61.

[4] S.C. Mikhlin, Variational Methods in Mathematical Physics, Macmillan, New York, 1964.

[5] S.C. Mikhlin, The Problem of the Minimum of a Quadratic Functional, Holden-Day, San Francisco, 1966.

[6] E. Tonti, Variational formulation of non-linear differential equations, Bull. Acad. Roy. Belg. Classe Sci, 55 (1969) 137–165, 263–278.

[7] M.M. Vainberg, Variational Methods for the Study of Nonlinear Operators, Holden-Day Inc., San Francisco, CA, 1964.

[8] J.T. Oden, A general theory of finite elements. Part I. Topological considerations, Int. J. Numer. Methods Eng. 1 (1969) 205–246.

[9] J.T. Oden, A general theory of finite elements. Part II. Applications, Int. J. Numer. Methods Eng. 1 (1969) 247–254.

[10] K. Washizu, Variational Methods in Elasticity and Plasticity, third ed., Pergamon Press, New York, 1982.

[11] O.C. Zienkiewicz, Constrained variational principles and penalty function methods in finite element analysis, in: Lecture Notes in Mathematics, No. 363, Springer-Verlag, Berlin, 1974, pp. 207–214.

[12] I. Fried, Shear in c^0 and c^1 bending finite elements, Int. J. Solids Struct. 9 (1973) 449–460.

[13] I. Fried, Finite element analysis of incompressible materials by residual energy balancing, Int. J. Solids Struct. 10 (1974) 993–1002.

[14] O.C. Zienkiewicz, E. Hinton, Reduced integration, function smoothing and non-conformity in finite element analysis, J. Franklin Inst. 302 (1976) 443–461.

[15] P.P. Lynn, S.K. Arya, Finite elements formulation by the weighted discrete least squares method, Int. J. Numer. Methods Eng. 8 (1974) 71–90.

[16] O.C. Zienkiewicz, D.R.J. Owen, K.N. Lee, Least square finite element for elastostatic problems—use of reduced integration, Int. J. Numer. Methods Eng. 8 (1974) 341–358.
[17] B.-N. Jiang, Least Squares Finite Element Method: Theory and Applications in Computational Fluid Dynamics and Electromagnetics, Springer, New York, 1998.
[18] P.B. Bochev, M.D. Gunzburger, Least-Squares Finite Element Methods, Springer, New York, 2009.
[19] B.-N. Jiang, Optimal least-squares finite element method for elliptic problems, Comput. Methods Appl. Mechanics Eng. 102 (1993) 199–212.
[20] B.-N. Jiang, On the least-squares method, Computer Methods in Appl. Mech. Eng. 152 (1998) 239–257.
[21] B.-N. Jiang, The least-squares finite element method in elasticity. I. Plane stress or strain with drilling degrees of freedom, Int. J. Numer. Methods Eng. 53 (2002) 621–636.
[22] B.-N. Jiang, The least-squares finite element method in elasticity. II. Bending of thin plates, Int. J. Numer. Methods Eng. 54 (2002) 1459–1475.
[23] O.C. Zienkiewicz, K. Morgan, Finite Elements and Approximation, John Wiley & Sons, London, 1983.
[24] R.V. Southwell, Relaxation Methods in Theoretical Physics, Clarendon Press, Oxford, 1946.
[25] D.N. de G. Allen, Relaxation Methods, McGraw-Hill, London, 1955.
[26] F.B. Hildebrand, Introduction to Numerical Analysis, second ed., Dover Publishers, 1987.
[27] A.R. Mitchell, D. Griffiths, The Finite Difference Method in Partial Differential Equations, John Wiley & Sons, London, 1980.
[28] P.K. Banerjee, The Boundary Element Methods in Engineering, McGraw-Hill, London, 1994.
[29] Prem K. Kythe, An Introduction to Boundary Element Methods, CRC Press, 1995.
[30] G. Beer, J.O. Watson, Programming the Boundary Element Method: An Introduction for Engineers, John Wiley & Sons, Chichester, 2001.
[31] L. Gaul, Boundary Element Methods for Engineers and Scientists, Springer, Berlin, 2003.