Introduction to the Finite Element Method

Dr. C. S. Jog

Dept. of Mechanical Engineering

Indian Institute of Science

Bangalore-560012

Chapter 1

Calculus of Variations

1.1 Introduction

In 1696, the Swiss mathematician Johann Bernoulli challenged his colleagues to solve an unresolved issue called the brachistochrone problem which is as follows: Find the shape of a frictionless chute between points (1) and (2) in a vertical plane such that a body sliding under the action of its own weight goes from point (1) to point (2) in the shortest time. Bernoulli originally specified a deadline of six months, but extended it to a year and half at the request of Leibniz, one of the leading scholars of the time, and the man who had, independently of Newton, invented the differential and integral calculus. The challenge was delivered to Newton at four p.m. on January 29, 1697. Before leaving for work the next morning, he had invented an entire new branch of mathematics called the calculus of variations, used it to solve the brachistochrone problem and sent off the solution, which was published at Newton's request, anonymously. But the brilliance and originality of the work betrayed the identity of its author. When Bernoulli saw the solution, he commented, "We recognize the lion by his claw". Newton was then fifty-five years old. Much of the formulation of this discipline was also developed by the Swiss mathematician Leonhard Euler.

What is the relevance of the calculus of variations to the finite element method? For one, the finite element formulation can be derived in a more direct way from a variational principle than from the corresponding energy functional. But perhaps more importantly, there are several problems such as, for example, inelastic deformations, where the governing differential equations can be cast in variational form but for which no corresponding energy functional exists. In such cases, developing the finite element formulation from the variational principle is the only alternative. We start by presenting the principles of the calculus of variations. The treatment presented here follows closely the one given in Shames and Dym [2].

1.2 Principles of the Calculus of Variations

While dealing with a function $y = f(x)$ in the ordinary calculus, we often want to determine the values of x for which y is a maximum or minimum. To establish the conditions of local extrema, we expand $f(x)$ as a Taylor series about a position $x = a$:

$$f(x) = f(a) + \left.\frac{df}{dx}\right|_{x=a} (x-a) + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x=a} (x-a)^2 + \ldots, \tag{1.1}$$

or alternatively,

$$f(x) - f(a) = \left.\frac{df}{dx}\right|_{x=a} \Delta x + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x=a} (\Delta x)^2 + \ldots,$$

where $\Delta x = x - a$. Since Δx can be positive or negative, the necessary condition for a local maxima or minima at an interior point (i.e., point not lying on the boundary of the domain) $x = a$ is

$$\left.\frac{df}{dx}\right|_{x=a} = 0.$$

The point $x = a$ is called an extremal point.

In the calculus of variations, we are concerned with extremizing (minimizing or maximizing) *functionals* which, roughly speaking, are 'functions of functions'. An example of a functional is

$$I = \int_{x_1}^{x_2} F(x, y, y')dx,$$

where y' denotes the derivative of y with respect to x.[1] Note that I is a function of a function $y(x)$, and x is the independent variable. We now present examples of some functionals.

1.2.1 Examples of functionals

(i) The Brachistochrone Problem:
This is the problem which was mentioned in the introduction. Referring to Fig. 1.1, the problem can be stated as: Find the shape, $y = f(x)$, of a frictionless chute between points 1 and 2 in a vertical plane so that a body starting from rest and sliding under the action of its own weight goes from 1 to 2 in the shortest possible time.

If v denotes the velocity of the body, and ds denotes an element of length along the curve, the time for descent is given by

$$I = \int_1^2 \frac{ds}{v} = \int_1^2 \frac{\sqrt{dx^2 + dy^2}}{v} = \int_{x_1}^{x_2} \frac{1}{v}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

[1] We shall follow the convention of denoting derivatives by primes throughout the remainder of these notes.

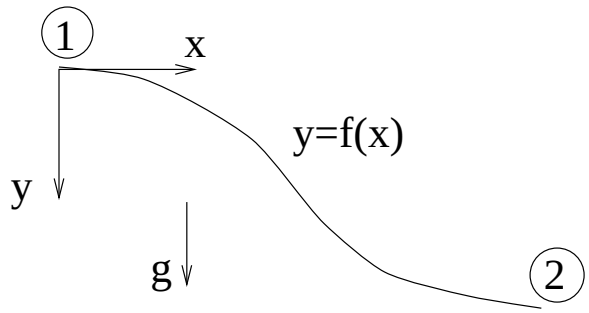


Figure 1.1: The brachistochrone problem.

The velocity v at a height y can be found using the equation of conservation of energy:

$$\frac{1}{2}mv_1^2 - mgy_1 = \frac{1}{2}mv^2 - mgy,$$

where the sign on the potential energy terms is negative because we have assumed our y-axis as positive in the direction of the gravitational acceleration. Since we have fixed our origin at point 1, and since the body starts from rest, $y_1 = v_1 = 0$, and we get $v = \sqrt{2gy}$. Hence, the time for descent is

$$I = \int_{x_1}^{x_2} \sqrt{\frac{1 + (y')^2}{2gy}}\, dx. \tag{1.2}$$

We shall prove later that the curve $y = f(x)$ which minimizes I is a *cycloid*.

(ii) The Geodesic Problem:
This problem deals with the determination of a curve on a given surface $g(x, y, z) = 0$, having the shortest length between two points 1 and 2 on this surface. Such curves are called *geodesics*. Hence, the problem is:
Find the extreme values of the functional

$$I = \int_{x_1}^{x_2} \sqrt{1 + (y')^2 + (z')^2}\, dx, \tag{1.3}$$

subject to the constraint

$$g(x, y, z) = 0. \tag{1.4}$$

Hence, this is a constrained optimization problem.

(iii) The Isoperimetric Problem:
The isoperimetric problem is: Of all the closed non-intersecting curves having a fixed length L, which curve encloses the greatest area A? From calculus, we know that the answer can be given in terms of the following contour integral:

$$I = A = \frac{1}{2}\oint (x\, dy - y\, dx) = \frac{1}{2}\oint_{\tau_1}^{\tau_2} \left(x\frac{\partial y}{\partial \tau} - y\frac{\partial x}{\partial \tau} \right) d\tau, \tag{1.5}$$

subject to the constraint

$$L = \oint ds = \oint_{\tau_1}^{\tau_2} \left[\left(\frac{\partial x}{\partial \tau} \right)^2 + \left(\frac{\partial y}{\partial \tau} \right)^2 \right]^{1/2} d\tau, \tag{1.6}$$

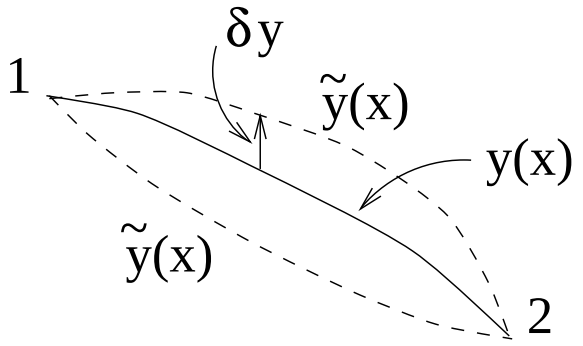


Figure 1.2: Extremal path and the varied paths.

where τ is a parameter. We shall show later that, as expected, the curve which solves the isoperimetric problem is a circle.

Note that each of the above three problems are different in character. The brachistochrone problem is an *unconstrained* problem since there are no constraints while performing the minimization. In the geodesic problem, we have a constraint on a *function* of the independent variables (see Eqn. 1.4), whereas in the isoperimetric problem we have a constraint on a *functional* (see Eqn. 1.6).

1.2.2 The first variation

Consider the functional

$$I = \int_{x_1}^{x_2} F(x, y, y')\, dx, \tag{1.7}$$

where F is a known function. Assume that $y(x_1)$ and $y(x_2)$ are given (e.g., the starting and ending points in the brachistochrone problem). Let $y(x)$ be the optimal path which extremizes (minimizes or maximizes) I, and let $\tilde{y}(x)$ be varied paths as shown in Fig. 1.2. Note that $\tilde{y}(x)$ agrees with $y(x)$ at the end points $y(x_1)$ and $y(x_2)$ since these are known. Our goal is to expand the functional I in the form of a Taylor series similar to Eqn. 1.1 so that the usual principles of calculus can be used for obtaining the path which minimizes or maximizes I. Towards this end, we relate $\tilde{y}(x)$ and $y(x)$ by using a scalar parameter ϵ as follows:

$$\tilde{y}(x) = y(x) + \epsilon\eta(x), \tag{1.8}$$

where $\eta(x)$ is a differentiable function having the requirement $\eta(x_1) = \eta(x_2) = 0$, so that $\tilde{y}(x_1) = y(x_1)$ and $\tilde{y}(x_2) = y(x_2)$. We see that an infinite number of paths can be generated by varying ϵ. In terms of the varied paths $\tilde{y}(x)$, the functional in Eqn. 1.7 can be written as

$$\tilde{I} = \int_{x_1}^{x_2} F(x, \tilde{y}, \tilde{y}')dx = \int_{x_1}^{x_2} F(x, y + \epsilon\eta, y' + \epsilon\eta')dx. \tag{1.9}$$

For $\epsilon = 0$, $\tilde{y}$ coincides with the extremizing function y, and $\tilde{I}$ reduces to the extreme value of the functional. Since $\tilde{I}$ is a function of the scalar parameter

ϵ, we can expand it as a power series similar to the one in Eqn. 1.1. Thus,

$$\tilde{I} = \tilde{I}\Big|_{\epsilon=0} + \left.\frac{d\tilde{I}}{d\epsilon}\right|_{\epsilon=0} \epsilon + \left.\frac{d^2\tilde{I}}{d\epsilon^2}\right|_{\epsilon=0} \frac{\epsilon^2}{2!} + \ldots, \tag{1.10}$$

or, alternatively,

$$\tilde{I} - \tilde{I}\Big|_{\epsilon=0} = \tilde{I} - I = \left.\frac{d\tilde{I}}{d\epsilon}\right|_{\epsilon=0} \epsilon + \left.\frac{d^2\tilde{I}}{d\epsilon^2}\right|_{\epsilon=0} \frac{\epsilon^2}{2!} + \ldots.$$

The necessary condition for I to be extreme is

$$\left.\frac{d\tilde{I}}{d\epsilon}\right|_{\epsilon=0} = 0,$$

or in other words,

$$\left[\int_{x_1}^{x_2} \left(\frac{\partial F}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial \epsilon} + \frac{\partial F}{\partial \tilde{y}'}\frac{\partial \tilde{y}'}{\partial \epsilon}\right) dx\right]_{\epsilon=0} = 0.$$

Since $\tilde{y} = y + \epsilon\eta$, we have $\partial\tilde{y}/\partial\epsilon = \eta$, $\partial\tilde{y}'/\partial\epsilon = \eta'$, $\tilde{y}|_{\epsilon=0} = y$ and $\tilde{y}'|_{\epsilon=0} = y'$. Hence, the above equation reduces to

$$\int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right) dx = 0 \quad \forall\eta(x).$$

Integrating by parts, we get

$$\int_{x_1}^{x_2} \frac{\partial F}{\partial y}\eta\, dx + \left.\frac{\partial F}{\partial y'}\eta\right|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\eta\, dx = 0 \quad \forall\eta(x).$$

Since $\eta(x_1) = \eta(x_2) = 0$, the above equation reduces to

$$\int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right]\eta\, dx = 0 \quad \forall\eta(x),$$

which implies that

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0. \tag{1.11}$$

Equation 1.11 is the famous Euler-Lagrange equation.

1.2.3 The 'delta' operator

We saw that the process of deriving the Euler-Lagrange equation was quite straightforward, but cumbersome. In order to make the procedure more 'mechanical', we introduce the 'delta' operator. Again referring to Fig. 1.2, we define the variation of $y(x)$ by

$$\delta y(x) = \tilde{y}(x) - y(x). \tag{1.12}$$

Thus, δy is simply the vertical distance between points on different curves at the same value of x. Since

$$\delta\left(\frac{dy}{dx}\right) = \frac{d\tilde{y}}{dx} - \frac{dy}{dx} = \frac{d}{dx}(\tilde{y} - y) = \frac{d(\delta y)}{dx},$$

the delta-operator is commutative with the derivative operator. The delta-operator is also commutative with the integral operator:

$$\delta \int y\,dx = \int \tilde{y}\,dx - \int y\,dx = \int (\tilde{y} - y)\,dx = \int (\delta y)\,dx.$$

From Eqn. 1.12, $\tilde{y} = y + \delta y$ and $\tilde{y}' = y' + \delta y'$. Assuming the existence of an extremal function, $y(x)$, and its derivative, $y'(x)$, we can expand F as a Taylor series about the extremal path and its derivative, using increments δy and $\delta y'$. That is

$$F(x, y+\delta y, y'+\delta y') = F(x, y, y') + \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y'\right) + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right),$$

or, equivalently,

$$F(x, y+\delta y, y'+\delta y') - F(x, y, y') = \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y'\right) + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right). \tag{1.13}$$

The left hand side of the above equation is the total variation of F, which we denote as $\delta^{(T)}F$. We call the bracketed term on the right-hand side of the equation as the *first variation*, $\delta^{(1)}F$. Thus,

$$\delta^{(T)}F = F(x, y + \delta y, y' + \delta y') - F(x, y, y'), \tag{1.14}$$

$$\delta^{(1)}F = \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y'. \tag{1.15}$$

Integrating Eqn. 1.13 between the limits x_1 and x_2, we get

$$\int_{x_1}^{x_2} F(x, y+\delta y, y'+\delta y')\,dx - \int_{x_1}^{x_2} F(x, y, y')\,dx = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y'\right) dx + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right).$$

This, may be written as

$$\tilde{I} - I = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y'\right) dx + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right).$$

Note that I represents the extreme value of the functional occurring at $\tilde{y} = y$. We call the term on the left hand side of the above equation as the total variation of I, and denote it by $\delta^{(T)}I$, while we call the first term on the right as the first variation and denote it by $\delta^{(1)}I$. Thus,

$$\delta^{(T)}I = \tilde{I} - I,$$

$$\delta^{(1)} I = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y} \delta y + \frac{\partial F}{\partial y'} \delta y' \right) dx.$$

Writing $\delta y'$ as $d(\delta y)/dx$, and integrating the expression for $\delta^{(1)} I$ by parts, we get

$$\delta^{(1)} I = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx} \left(\frac{\partial F}{\partial y'} \right) \right] \delta y \, dx + \left(\frac{\partial F}{\partial y'} \delta y \right) \Bigg|_{x_1}^{x_2}. \tag{1.16}$$

Since $\delta y(x_1) = \delta y(x_2) = 0$, the above equation reduces to

$$\delta^{(1)} I = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx} \left(\frac{\partial F}{\partial y'} \right) \right] \delta y \, dx,$$

so that we have

$$\begin{aligned} \delta^{(T)} I &= \delta^{(1)} I + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right) \\ &= \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx} \left(\frac{\partial F}{\partial y'} \right) \right] \delta y \, dx + \mathrm{o}\left((\delta y)^2\right) + \mathrm{o}\left((\delta y')^2\right). \end{aligned}$$

In order for I to be a maximum or minimum, $\delta^{(T)} I$ must have the same sign for all possible variations δy over the interval (x_1, x_2). Note that δy at any given position x can be either positive or negative. Hence, the only way $\delta^{(T)} I$ can have the same sign for all possible variations is if the term in parenthesis vanishes, which leads us again to the Euler-Lagrange equation given by Eqn. 1.11. Thus, we can conclude that the a necessary condition for extremizing I is that the first variation of I vanishes, i.e.,

$$\delta^{(1)} I \equiv \left(\frac{\partial \tilde{I}}{\partial \epsilon} \right)_{\epsilon = 0} \epsilon = 0 \quad \forall \delta y.$$

We are now in a position to solve some example problems.

1.2.4 Examples

In this subsection, we shall solve the brachistochrone and geodesic problems which were presented earlier.

The Brachistochrone Problem:

We have seen that the functional for this problem is given by Eqn. 1.2. We thus have

$$I = \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \sqrt{\frac{1 + (y')^2}{y}} \, dx,$$

or in other words

$$F = \sqrt{\frac{1 + (y')^2}{y}}.$$

Substituting for F in the Euler-Lagrange equation given by Eqn. 1.11, we get the differential equation

$$2yy'' + 1 + (y')^2 = 0.$$

To solve the above differential equation, we make the substitution $y' = u$ to get

$$u\frac{du}{dy} = -\frac{1+u^2}{2y},$$

or

$$-\int \frac{udu}{1+u^2} = \int \frac{dy}{2y}.$$

Integrating, we get

$$y(1+u^2) = c_1,$$

or, alternatively

$$y(1+(y')^2) = c_1.$$

The above equation can be written as

$$x = \int \frac{\sqrt{y}dy}{\sqrt{c_1 - y}} + x_0.$$

If we substitute $y = c_1 \sin^2(\tau/2)$, we get $x = c_1(\tau - \sin\tau)/2 + x_0$. Since $x = y = 0$ at $\tau = 0$, we have $x_0 = 0$. Hence, the solution for the extremizing path is

$$\begin{aligned} x &= \frac{c_1}{2}(\tau - \sin\tau), \\ y &= \frac{c_1}{2}(1 - \cos\tau), \end{aligned}$$

which is the parametric equation of a cycloid.

Geodesics on a Sphere:

We shall find the geodesics on the surface of a sphere, i.e., curves which have the shortest length between two given points on the surface of a sphere. Working with spherical coordinates, the coordinates of a point on the sphere are given by

$$\begin{aligned} x &= R\sin\theta\cos\phi, \\ y &= R\sin\theta\sin\phi, \\ z &= R\cos\theta. \end{aligned} \tag{1.17}$$

Hence, the length of an element ds on the surface of the sphere is

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2 + dz^2} \\ &= R\sqrt{d\theta^2 + \sin^2\theta d\phi^2} \\ &= Rd\theta\left[1 + \sin^2\theta\left(\frac{d\phi}{d\theta}\right)^2\right]^{1/2}. \end{aligned}$$

Since

$$I = \int ds = \int R\left[1 + \sin^2\theta\left(\frac{d\phi}{d\theta}\right)^2\right]^{1/2} d\theta,$$

we have

$$F = R\left[1 + \sin^2\theta\left(\frac{d\phi}{d\theta}\right)^2\right]^{1/2}.$$

Noting that F is a function of ϕ' but not of ϕ, the Euler-Lagrange equation when integrated gives the solution

$$\cos c_1 \sin\phi \sin\theta - \sin c_1 \sin\theta \cos\phi = c_2 \cos\theta,$$

where c_1 and c_2 are constants. Using Eqn. 1.17, the above equation can also be written as

$$(\cos c_1)y - (\cos c_1)x = c_2 z,$$

which is nothing but the equation of a plane surface going through the origin. Thus, given any two points on the sphere, the intersection of the plane passing through the two points and the origin, with the sphere gives the curve of the shortest length amongst all curves lying on the surface of the sphere and joining the two points. Such a curve is known as a 'Great Circle'.

In this example, the constraint that the curve lie on the surface of the sphere was treated implicitly by using spherical coordinates, and thus did not have to be considered explicitly. This procedure might not be possible for all problems where constraints on functions are involved. In such cases, we have to devise a procedure for handling constraints, both of the functional and function type. We shall do so in Sections 1.2.6 and 1.2.7.

1.2.5 First variation with several dependent variables

We now extend the results of the previous sections to finding the Euler-Lagrange equations for a functional with several dependent variables but only one independent variable. Consider the functional

$$I = \int_{x_1}^{x_2} F(y_1, y_2, ..., y_n, y_1', y_2', ..., y_n', x)\, dx$$

Setting the first variation of the above functional to zero, i.e., $\delta^{(1)} I = 0$, we get

$$\int_{x_1}^{x_2}\left[\frac{\partial F}{\partial y_1}\delta y_1 + \cdots + \frac{\partial F}{\partial y_n}\delta y_n + \frac{\partial F}{\partial y_1'}\delta y_1' + \cdots + \frac{\partial F}{\partial y_n'}\delta y_n'\right] dx = 0.$$

Since the variations can be arbitrary, choose $\delta y_2 = \delta y_3 = \ldots = \delta y_n = 0$. Then since δy_1 is arbitrary, we get the first Euler-Lagrange equation as

$$\frac{\partial F}{\partial y_1} - \frac{d}{dx}\left(\frac{\partial F}{\partial y_1'}\right) = 0.$$

Similarly, by taking δy_2 as the only nonzero variation, we get the second Euler-Lagrange equation with the subscript 1 replaced by 2. In all, we get n Euler-Lagrange equations corresponding to the n dependent variables:

$$\frac{\partial F}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial F}{\partial y_i'}\right) = 0, \quad i = 1, 2, \ldots, n. \tag{1.18}$$

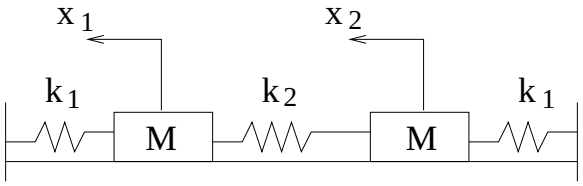


Figure 1.3: Spring mass system with two degrees of freedom.

To illustrate the above formulation, consider the problem shown in Fig. 1.3. Our goal is to write the equations of motion of the two mass particles. In order to do that, we use Hamilton's principle (see Section 6.2) which states that for a system of particles acted on by conservative forces, the paths taken from a configuration at time t_1 to a configuration at time t_2 are those that extremize the following functional:

$$I = \int_{t_1}^{t_2} (T - V)\, dt,$$

where T is the kinetic energy of the system, while V is the potential energy of the system. For the problem under consideration, we have

$$T = \frac{1}{2} M \dot{x}_1^2 + \frac{1}{2} M \dot{x}_2^2,$$
$$V = \frac{1}{2} k_1 x_1^2 + \frac{1}{2} k_2 (x_2 - x_1)^2 + \frac{1}{2} k_1 x_2^2.$$

Thus,

$$I = \int_{t_1}^{t_2} \left\{ \frac{1}{2} M \dot{x}_1^2 + \frac{1}{2} M \dot{x}_2^2 - \frac{1}{2} \left[k_1 x_1^2 + k_2 (x_2 - x_1)^2 + k_1 x_2^2 \right] \right\} dt.$$

Note that x_1, x_2 are the two dependent variables, and t is the independent variable in the above equation. From Eqn. 1.18, we have

$$\frac{\partial F}{\partial x_1} - \frac{d}{dt} \left(\frac{\partial F}{\partial \dot{x}_1} \right) = 0,$$
$$\frac{\partial F}{\partial x_2} - \frac{d}{dt} \left(\frac{\partial F}{\partial \dot{x}_2} \right) = 0.$$

Substituting for F, we get

$$M \ddot{x}_1 + k_1 x_1 - k_2 (x_2 - x_1) = 0,$$
$$M \ddot{x}_2 + k_1 x_2 + k_2 (x_2 - x_1) = 0,$$

which are the same as Newton's laws applied to the two masses. By integrating the above equations using the initial conditions $x_1(0)$, $x_2(0)$, $\dot{x}_1(0)$, $\dot{x}_2(0)$, we can find the subsequent motions of the masses $x_1(t)$ and $x_2(t)$.

In this example, we could have more easily employed Newton's laws directly. However, there are many problems where it is easier to proceed by the variational approach to arrive at the governing equations of motion. In addition, variational formulations also yield the appropriate boundary conditions to be imposed as we shall see in future sections.

1.2.6 Functional constraint

In the isoperimetric problem, we saw that we have to minimize a functional subject to a constraint on a functional. So far, we have dealt only with the minimization of unconstrained functionals. We now show how to carry out the extremization process in the presence of a functional constraint.

Consider the problem of minimizing

$$I = \int_{x_1}^{x_2} F(x, y, y')\, dx,$$

subject to the restriction

$$J = \int_{x_1}^{x_2} G(x, y, y')\, dx = \text{constant}.$$

If we still use a one-parameter family of admissible functions of the form $\tilde{y} = y(x) + \delta y$, where $\delta y = \epsilon\eta$, then we might no longer be able to satisfy the constraint

$$\tilde{J} = \int_{x_1}^{x_2} G(x, \tilde{y}, \tilde{y}')\, dx = \text{constant}. \tag{1.19}$$

Hence, we now need to introduce additional flexibility into the system by using a two-parameter family of admissible functions of the form

$$\tilde{y}(x) = y(x) + \delta y_1 + \delta y_2.$$

The variation δy_1 is arbitrary, while the variation δy_2 is such that Eqn. (1.19) is satisfied. For extremizing I, we set $\delta^{(1)} I$ to zero, while $\delta^{(1)} J$ is zero since J is a constant. Hence, using the method of Lagrange multipliers, we have

$$\delta^{(1)} I + \lambda \delta^{(1)} J = 0,$$

where λ is a constant known as the Lagrange multiplier. We can write the above condition as

$$\int_{x_1}^{x_2} \left\{ \left[\frac{\partial F^*}{\partial y} - \frac{d}{dx}\frac{\partial F^*}{\partial y'} \right] \delta y_1 + \left[\frac{\partial F^*}{\partial y} - \frac{d}{dx}\frac{\partial F^*}{\partial y'} \right] \delta y_2 \right\} dx = 0,$$

where $F^* = F + \lambda G$. We choose λ such that the second integral term vanishes. Then we have

$$\int_{x_1}^{x_2} \left[\frac{\partial F^*}{\partial y} - \frac{d}{dx}\frac{\partial F^*}{\partial y'} \right] \delta y_1\, dx = 0,$$

which, by virtue of the arbitrariness of δy_1, yields the required Euler-Lagrange equations:

$$\frac{\partial F^*}{\partial y} - \frac{d}{dx}\left(\frac{\partial F^*}{\partial y'} \right) = 0. \tag{1.20}$$

So far we have considered only one constraint equation. If we have n constraint equations

$$\int_{x_1}^{x_2} G_k(x, y, y')\, dx = c_k \quad k = 1, 2, \ldots, n,$$

then we need to use an $n+1$-parameter family of varied paths

$$\tilde{y} = y + \delta y_1 + \delta y_2 + \cdots + \delta y_{n+1},$$

out of which, say δy_1 is arbitrary, and the remaining are adjusted so as to satisfy the n constraint equations. The n Lagrange multipliers are chosen such that

$$\sum_{i=1}^{n} \int_{x_1}^{x_2} \left[\frac{\partial F^*}{\partial y} - \frac{d}{dx} \frac{\partial F^*}{\partial y'} \right] \delta y_{i+1}\, dx = 0,$$

where $F^* = F + \sum_{k=1}^{n} \lambda_k G_k$. The arbitrariness of δy_1 finally leads to the Euler-Lagrange equations

$$\frac{\partial F^*}{\partial y} - \frac{d}{dx}\left(\frac{\partial F^*}{\partial y'} \right) = 0.$$

Finally, for p independent variables, i.e.,

$$I = \int_{x_1}^{x_2} F(x, y_1, y_2, \ldots, y_p, y_1', y_2', \ldots, y_p')\, dx,$$

and n constraints

$$\int_{x_1}^{x_2} G_k(x, y_1, y_2, \ldots, y_p, y_1', y_2', \ldots, y_p')\, dx = c_k, \quad k = 1, 2, \ldots, n,$$

the Euler-Lagrange equations are given by

$$\frac{\partial F^*}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial F^*}{\partial y_i'} \right) = 0, \quad i = 1, 2, \ldots, p,$$

where $F^* = F + \sum_{k=1}^{n} \lambda_k G_k$.

We are now in a position to tackle the isoperimetric problem. We are interested in finding a curve $y(x)$, which for a given length

$$L = \int_{\tau_1}^{\tau_2} \sqrt{\left(\frac{\partial x}{\partial \tau} \right)^2 + \left(\frac{\partial y}{\partial \tau} \right)^2}\, d\tau,$$

encloses the greatest area

$$A = \frac{1}{2} \int_{\tau_1}^{\tau_2} \left(x \frac{\partial y}{\partial \tau} - y \frac{\partial x}{\partial \tau} \right) d\tau.$$

Note that τ is the independent variable, and x and y are the dependent variables. Denoting $\partial x/\partial \tau$ by $\dot{x}$ and $\partial y/\partial \tau$ by $\dot{y}$, we have

$$F^* = \frac{1}{2}(x\dot{y} - y\dot{x}) + \lambda(\dot{x}^2 + \dot{y}^2)^{1/2}.$$

The Euler-Lagrange equations are

$$\frac{\dot{y}}{2} - \frac{d}{d\tau}\left[-\frac{y}{2} + \frac{\lambda \dot{x}}{(\dot{x}^2 + \dot{y}^2)^{1/2}} \right] = 0,$$

$$-\frac{\dot{x}}{2} - \frac{d}{d\tau}\left[\frac{x}{2} + \frac{\lambda\dot{y}}{(\dot{x}^2+\dot{y}^2)^{1/2}}\right] = 0.$$

Integrating with respect to τ, we get

$$y - \frac{\lambda\dot{x}}{(\dot{x}^2+\dot{y}^2)^{1/2}} = c_1,$$
$$x + \frac{\lambda\dot{y}}{(\dot{x}^2+\dot{y}^2)^{1/2}} = c_2.$$

Eliminating λ from the above two equations, we get

$$(x - c_2)\,dx + (y - c_1)\,dy = 0.$$

Integrating the above equation, we obtain

$$(x - c_2)^2 + (y - c_1)^2 = c_3^2,$$

which is the equation of a circle with center at (c_2, c_1) and radius c_3. Thus, as expected, the curve which encloses the maximum area for a given perimeter is a circle.

1.2.7 Function constraints

We have seen in the geodesic problem that we can have a constraint on functions of the independent variables. In this subsection, we merely present the Euler-Lagrange equations for function constraints without going into the derivation which is quite similar to the derivation for functional constraints presented in the previous subsection.

Suppose we want to extremize

$$I = \int_{x_1}^{x_2} F(x, y_1, y_2, \ldots, y_n, y_1', y_2', \ldots, y_n')\,dx,$$

with the following m constraints

$$G_k(x, y_1, y_2 \ldots, y_n, y_1', y_2', \ldots, y_n') = 0, \quad k = 1, 2, \ldots, m.$$

Then the Euler-Lagrange equations are

$$\frac{\partial F^*}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial F^*}{\partial y_i'}\right) = 0, \quad i = 1, 2, \ldots, n,$$

where $F^* = F + \sum_{k=1}^{m} \lambda_k(x) G_k$. Note that the Lagrange multipliers, $\lambda_k(x)$, are *functions* of x, and are not constants as in the case of functional constraints.

1.2.8 A note on boundary conditions

So far, while extremizing I, we assumed that the value of the extremizing function $y(x)$ was specified at the endpoints x_1 and x_2. This resulted in $\delta y(x_1) = \delta y(x_2) = 0$, and hence to the elimination of the boundary terms in Eqn. 1.16. Such a boundary condition where $y(x)$ is specified at a point is known as a *kinematic* boundary condition. The question naturally arises: Are there other types of boundary conditions which are also valid? The answer is that there are.

Let

$$I = \int_{x_1}^{x_2} F(x, y, y')\, dx - g_0 y|_{x=x_2}\,,$$

where g_0 is a constant. Rewrite using Eqn. 1.16 the necessary condition for extremizing the functional I as

$$\delta^{(1)} I = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] \delta y\, dx + \left[\left(\frac{\partial F}{\partial y'} - g_0\right)\delta y\right]_{x_2} - \left.\frac{\partial F}{\partial y'}\delta y\right|_{x_1} = 0\ \forall \delta y. \tag{1.21}$$

Since the variations δy are arbitrary, we can choose variations δy such that

$$\delta y|_{x=x_1} = \delta y|_{x=x_2} = 0.$$

Then we get

$$\int_{x_1}^{x_2} \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] \delta y\, dx = 0,$$

for all δy which satisfy $\delta y|_{x=x_1} = \delta y|_{x=x_2} = 0$. This in turn leads us to the Euler-Lagrange equations

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0.$$

Hence, Eqn. 1.21 can be written as

$$\left[\left(\frac{\partial F}{\partial y'} - g_0\right)\delta y\right]_{x_2} - \left.\frac{\partial F}{\partial y'}\delta y\right|_{x_1} = 0 \quad \forall \delta y.$$

Consider variations such that $\delta y|_{x=x_2} = 0$. From the above equation, we get

$$\left.\frac{\partial F}{\partial y'}\delta y\right|_{x_1} = 0,$$

which in turn implies that

$$\left.\frac{\partial F}{\partial y'}\right|_{x=x_1} = 0 \quad \text{or} \quad y \text{ prescribed at } x_1.$$

We have already encountered the second boundary condition which we have called the kinematic boundary condition. The first condition is something we have not encountered so far, and is known as the *natural* boundary condition.

Similar to the above process, by considering variations such that $\delta y|_{x=x_1} = 0$, we get the boundary conditions at $x = x_2$ as

$$\left.\frac{\partial F}{\partial y'}\right|_{x=x_2} = g_0 \quad \text{or} \quad y \text{ prescribed at } x_2.$$

Summarizing, the variational process has resulted in not only yielding the governing differential equation but also the boundary conditions which can be specified. We can either specify

1. kinematic boundary conditions at both endpoints,
2. natural boundary boundary conditions at both endpoints, or
3. a natural boundary condition at one of the endpoints and a kinematic one at the other.

Note that physical considerations rule out specifying both natural and kinematic boundary conditions at the same boundary point. To give some examples, in the boundary value problems of elasticity, the kinematic conditions would correspond to specified displacements, while the natural boundary conditions would correspond to specified tractions. In problems of heat transfer, the kinematic and natural boundary conditions would correspond to specified temperature and specified normal heat flux, respectively.

1.2.9 Functionals involving higher-order derivatives

So far, we have considered only first-order derivatives in the functionals. We now extend the work of finding the Euler-Lagrange equations and the appropriate boundary conditions for functionals involving higher-order derivatives. Consider the functional

$$I = \int_{x_1}^{x_2} F(x, y, y', y'', y''')\, dx - M_0 y'|_{x=x_2} - V y''|_{x=x_2}\,.$$

For finding the extremizing function $y(x)$, we set the first variation of I to zero, i.e.,

$$\delta^{(1)} I = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y'' + \frac{\partial F}{\partial y'''}\delta y''' \right) dx - M_0 \delta y'|_{x=x_2} + V \delta y''|_{x=x_2} = 0.$$

Integrating by parts, we get

$$\begin{aligned} &- \int_{x_1}^{x_2} \left[\frac{d^3}{dx^3}\left(\frac{\partial F}{\partial y'''}\right) - \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) + \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} \right] \delta y\, dx \\ &\qquad + \left[\frac{\partial F}{\partial y'''} - V \right] \delta y'' \Big|_{x_1}^{x_2} - \left[\frac{d}{dx}\left(\frac{\partial F}{\partial y'''}\right) - \frac{\partial F}{\partial y''} \right] \delta y' \Big|_{x_1}^{x_2} \\ &\quad + \left[\frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y'''}\right) - \frac{d}{dx}\left(\frac{\partial F}{\partial y''}\right) + \frac{\partial F}{\partial y'} \right] \delta y \Big|_{x_1}^{x_2} - M_0 \delta y'|_{x=x_2} = 0 \quad \forall \delta y. \end{aligned}$$

Since the variations, δy, are arbitrary, the above equation yields the Euler-Lagrange equation

$$\frac{d^3}{dx^3}\left(\frac{\partial F}{\partial y'''}\right) - \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) + \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0,$$

and the boundary conditions

$$\begin{array}{lll} y'' \text{ specified} & \text{or} & \dfrac{\partial F}{\partial y'''} = V, \\ y' \text{ specified} & \text{or} & \dfrac{d}{dx}\left(\dfrac{\partial F}{\partial y'''}\right) - \dfrac{\partial F}{\partial y''} = -M_0, \\ y \text{ specified} & \text{or} & \dfrac{d^2}{dx^2}\left(\dfrac{\partial F}{\partial y'''}\right) - \dfrac{d}{dx}\left(\dfrac{\partial F}{\partial y''}\right) + \dfrac{\partial F}{\partial y'} = 0. \end{array}$$

In each row, the first condition corresponds to the kinematic boundary condition, while the second one corresponds to the natural boundary condition. We shall consider the extension of the above concepts to more than one independent variable directly by dealing with problems in elasticity and heat transfer in three-dimensional space. Before we do that, we consider some basic concepts in functional analysis. We shall need these concepts in formulating the variational principles and in presenting the abstract formulation.

1.3 Some Concepts from Functional Analysis

If V is a Hilbert space, then we say that L is a linear form on V if $L : V \to \Re$, i.e., $L(v) \in \Re \ \forall v \in V$, and if for all $u, v \in V$ and $\alpha, \beta \in \Re$,

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v).$$

$a(.,.)$ is a bilinear form on $V \times V$ if $a : V \times V \to \Re$, i.e., $a(u, v) \in \Re$ for all $u, v \in V$, and if $a(.,.)$ is linear in each argument, i.e., for all $u, v, w \in V$ and $\alpha, \beta \in \Re$, we have

$$\begin{aligned} a(u, \alpha v + \beta w) &= \alpha a(u, v) + \beta a(u, w), \\ a(\alpha u + \beta v, w) &= \alpha a(u, w) + \beta a(v, w). \end{aligned}$$

A bilinear form $a(.,.)$ on $V \times V$ is symmetric if

$$a(u, v) = a(v, u) \quad \forall u, v \in V.$$

A symmetric bilinear form $a(.,.)$ on $V \times V$ is said to be a scalar product on V if

$$a(v, v) > 0 \quad \forall v \in V.$$

The norm associated with the scalar product $a(.,.)$ is defined by

$$\|v\|_a = [a(v, v)]^{1/2} \quad \forall v \in V.$$

Now we define some Hilbert spaces which arise naturally in variational formulations. If $I = (a, b)$ is an interval, then the space of 'square-integrable' functions is

$$L_2(I) = \left\{ v : \int_a^b v^2 \, dx < \infty \right\}.$$

the scalar product associated with L_2 is

$$(v, w)_{L_2(I)} = \int_a^b vw \, dx,$$

and the corresponding norm (called the L_2-norm) is

$$\|v\|_2 = \left[\int_a^b v^2 \, dx \right]^{1/2} = (v, v)^{1/2}.$$

Note that functions belonging to L_2 can be unbounded, e.g., $x^{-1/4}$ lies in $L_2(0, 1)$, but x^{-1} does not. In general, $v(x) = x^{-\beta}$ lies in $L_2(0, 1)$ if $\beta < 1/2$.

Now we introduce the Hilbert space

$$H^1(I) = \{ v : v \in L_2(I), v' \in L_2(I) \}.$$

This space is equipped with the scalar product

$$(v, w)_{H^1(I)} = \int_a^b (vw + v'w') \, dx,$$

and the corresponding norm

$$\|v\|_{H^1(I)} = (v, v)_{H^1(I)}^{1/2} = \left[\int_a^b (v^2 + (v')^2) \, dx \right]^{1/2}.$$

For boundary value problems, we also need to define the space

$$H_0^1(I) = \{ v \in H_1(I) : v(a) = v(b) = 0 \}.$$

Extending the above definitions to three dimensions

$$\begin{aligned}
L_2(\Omega) &= \left\{ v : \int_\Omega v^2 \, d\Omega < \infty \right\}, \\
H^1(\Omega) &= \left\{ v \in L_2(\Omega); \frac{\partial v}{\partial x_i} \in L_2(\Omega) \right\}, \\
(v, w)_{L_2(\Omega)} &= \int_\Omega vw \, d\Omega, \\
\|v\|_{L_2(\Omega)} &= \left[\int_\Omega v^2 \, d\Omega \right]^{1/2}, \\
(v, w)_{H^1(\Omega)} &= \int_\Omega (vw + \boldsymbol{\nabla} v \cdot \boldsymbol{\nabla} w) \, d\Omega,
\end{aligned}$$

$$\|v\|_{H^1(\Omega)} = \left[\int_\Omega (v^2 + \boldsymbol{\nabla} v \cdot \boldsymbol{\nabla} v)\, d\Omega\right]^{1/2},$$
$$H_0^1(\Omega) = \{v \in H_1(\Omega) : v = 0 \text{ on } \Gamma\}.$$

For vector fields, we have

$$\|\boldsymbol{v}\|_{L_2} = \left[\int_\Omega (v_1^2 + v_2^2 + v_3^2)\, d\Omega\right]^{1/2},$$
$$\|\boldsymbol{v}\|_{H^1}^2 = \|\boldsymbol{v}\|_{L_2}^2 + \int_\Omega \left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial v_i}{\partial x_j}\right)^2\right] d\Omega.$$

1.4 Variational Principles in Elasticity

In this course, we shall restrict ourself to *single-field* or *irreducible* variational statements, i.e., variational statements which have only one independent solution field. The dependent solution fields are found by pointwise application of the governing differential equations and boundary conditions. For example, in the case of elasticity, one can have a variational principle with the displacement field as the independent solution field, and the strain and stress fields as the dependent solution fields, while in the case of heat transfer problems, we can have temperature as the independent, and heat flux as the dependent solution field.

We start by presenting the strong form of the governing equations for linear elasticity. In what follows, scalars are denoted by lightface letters, while vectors and higher-order tensors are denoted by boldface letters. The dot symbol is used to denote contraction over one index, while the colon symbol is used to denote contraction over two indices. When indicial notation is used summation over repeated indices is assumed with the indices ranging from 1 to 3. For example, $\boldsymbol{t}\cdot\boldsymbol{u} = t_i u_i = t_1u_1 + t_2u_2 + t_3u_3$, $(\boldsymbol{C} : \boldsymbol{\epsilon})_{ij} = C_{ijkl}\epsilon_{kl}$, where i and j are free indices, and k and l are dummy indices over which summation is carried out.

1.4.1 Strong form of the governing equations

Let Ω be an open domain whose boundary Γ is composed of two open, disjoint regions, $\Gamma = \overline{\Gamma_u \cup \Gamma_t}$. We consider the following boundary value problem:
Find the displacements $\boldsymbol{u}$, stresses $\boldsymbol{\tau}$, strains $\boldsymbol{\epsilon}$, and tractions $\boldsymbol{t}$, such that

$$\boldsymbol{\nabla}\cdot\boldsymbol{\tau} + \rho\boldsymbol{b} = \boldsymbol{0} \text{ on } \Omega, \tag{1.22}$$
$$\boldsymbol{\tau} = \boldsymbol{C}(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^0) + \boldsymbol{\tau}^0 \text{ on } \Omega, \tag{1.23}$$
$$\boldsymbol{\epsilon} = \frac{1}{2}\left[(\boldsymbol{\nabla}\boldsymbol{u}) + (\boldsymbol{\nabla}\boldsymbol{u})^t\right] \text{ on } \Omega, \tag{1.24}$$
$$\boldsymbol{t} = \boldsymbol{\tau}\boldsymbol{n} \text{ on } \Gamma, \tag{1.25}$$
$$\boldsymbol{t} = \bar{\boldsymbol{t}} \text{ on } \Gamma_t, \tag{1.26}$$

$$\boldsymbol{u} = \bar{\boldsymbol{u}} \ \text{ on } \Gamma_u, \tag{1.27}$$

where $\boldsymbol{\epsilon}^0$ and $\boldsymbol{\tau}^0$ are the initial strain and initial stress tensors, $\boldsymbol{n}$ is the outward normal to Γ, and $\bar{\boldsymbol{t}}$ are the prescribed tractions on Γ_t. In a single-field displacement-based variational formulation which we shall describe shortly, we enforce Eqns. 1.23-1.25, and Eqn. 1.27 in a strong sense, i.e., at each point in the domain, while the equilibrium equation given by Eqn. 1.22, and the prescribed traction boundary condition given by Eqn. 1.26 (which is the natural boundary condition) are enforced weakly or in an integral sense. For the subsequent development, we find it useful to define the following function spaces:

$$V_u = \{\boldsymbol{v} \in H^1(\Omega) : \ \boldsymbol{v} = \boldsymbol{0} \text{ on } \Gamma_u\},$$
$$L_u = \tilde{\boldsymbol{u}} + V_u \text{ where } \tilde{\boldsymbol{u}} \in H^1(\Omega) \text{ satisfies } \tilde{\boldsymbol{u}} = \bar{\boldsymbol{u}} \text{ on } \Gamma_u.$$

In what follows, we shall use direct tensorial notation. The corresponding formulation in indicial notation is given in [2].

1.4.2 Principle of virtual work

We obtain the variational formulation using the method of weighted residuals whereby we set the sum of the weighted residuals based on the equilibrium equation and the natural boundary condition to zero:

$$\int_\Omega \boldsymbol{v} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{\tau} + \rho \boldsymbol{b})\, d\Omega + \int_{\Gamma_t} \boldsymbol{v} \cdot (\bar{\boldsymbol{t}} - \boldsymbol{t})\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u. \tag{1.28}$$

In the above equation, $\boldsymbol{v}$ represent the variation of the displacement field, i.e., $\boldsymbol{v} \equiv \delta \boldsymbol{u}$. Using the tensor identity

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\tau}^t \boldsymbol{v}) = \boldsymbol{v} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{\tau}) + \boldsymbol{\nabla} \boldsymbol{v} : \boldsymbol{\tau}, \tag{1.29}$$

we get

$$\int_\Omega \left[\boldsymbol{\nabla} \cdot (\boldsymbol{\tau}^t \boldsymbol{v}) - \boldsymbol{\nabla} \boldsymbol{v} : \boldsymbol{\tau} + \rho \boldsymbol{v} \cdot \boldsymbol{b}\right]\, d\Omega + \int_{\Gamma_t} \boldsymbol{v} \cdot (\bar{\boldsymbol{t}} - \boldsymbol{t})\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u.$$

Now applying the divergence theorem to the first term of the above equation, we get

$$\int_\Gamma (\boldsymbol{\tau}^t \boldsymbol{v}) \cdot \boldsymbol{n}\, d\Gamma - \int_\Omega \left[\boldsymbol{\nabla} \boldsymbol{v} : \boldsymbol{\tau} - \rho \boldsymbol{v} \cdot \boldsymbol{b}\right]\, d\Omega + \int_{\Gamma_t} \boldsymbol{v} \cdot (\bar{\boldsymbol{t}} - \boldsymbol{t})\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u. \tag{1.30}$$

Using Eqn. 1.25 and the fact that $\boldsymbol{v} = \boldsymbol{0}$ on Γ_u, we have

$$\int_\Gamma (\boldsymbol{\tau}^t \boldsymbol{v}) \cdot \boldsymbol{n}\, d\Gamma = \int_\Gamma \boldsymbol{v} \cdot \boldsymbol{t}\, d\Gamma = \int_{\Gamma_t} \boldsymbol{v} \cdot \boldsymbol{t}\, d\Gamma \quad \forall \boldsymbol{v} \in V_u.$$

Hence, Eqn. 1.30 simplifies to

$$\int_\Omega \boldsymbol{\nabla} \boldsymbol{v} : \boldsymbol{\tau}\, d\Omega = \int_\Omega \rho \boldsymbol{v} \cdot \boldsymbol{b}\, d\Omega + \int_{\Gamma_t} \boldsymbol{v} \cdot \bar{\boldsymbol{t}}\, d\Gamma \quad \forall \boldsymbol{v} \in V_u.$$

Using the symmetry of the stress tensor, the above equation can be written as

$$\int_\Omega \frac{1}{2}\left[\boldsymbol{\nabla v} + (\boldsymbol{\nabla v})^t\right] : \boldsymbol{\tau}\, d\Omega = \int_\Omega \rho \boldsymbol{v}\cdot\boldsymbol{b}\, d\Omega + \int_{\Gamma_t} \boldsymbol{v}\cdot\bar{\boldsymbol{t}}\, d\Gamma \quad \forall \boldsymbol{v} \in V_u.$$

Using Eqn. 1.24, the single-field displacement-based variational formulation can be stated as

Find $\boldsymbol{u} \in L_u$ such that

$$\int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{\tau}\, d\Omega = \int_\Omega \rho \boldsymbol{v}\cdot\boldsymbol{b}\, d\Omega + \int_{\Gamma_t} \boldsymbol{v}\cdot\bar{\boldsymbol{t}}\, d\Gamma \quad \forall \boldsymbol{v} \in V_u, \tag{1.31}$$

where $\boldsymbol{\tau}$ is given by Eqn. 1.23. What we have shown is that if $\boldsymbol{u}$ is a solution to the strong form of the governing equations, it is also a solution of the weak form of the governing equations given by Eqn. 1.31. Note that the variational formulation has resulted in a weakening of the differentiability requirement on $\boldsymbol{u}$. While in the strong form of the governing equations, we require $\boldsymbol{u}$ to be twice differentiable, in the variational formulation, $\boldsymbol{u}$ just needs to be differentiable. This is reflected by the fact that $\boldsymbol{u}$ lies in the space L_u. The variational formulation presented above is often known as the 'principle of virtual work'.

We have shown that the governing equations in strong form can be stated in variational form. The converse if also true provided the solution $\boldsymbol{u}$ is assumed to be twice differentiable, i.e., a solution $\boldsymbol{u} \in L_u$ of the variational formulation is also a solution of the strong form of the governing equations provided it has sufficient regularity. To demonstrate this, we start with Eqn. 1.31. A large part of the proof is a reversal of the steps used in the preceding proof. Using the symmetry of the stress tensor, we can write $\int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{\tau}\, d\Omega$ as $\int_\Omega \boldsymbol{\nabla v} : \boldsymbol{\tau}\, d\Omega$. Using the tensor identity given by Eqn. 1.29 to eliminate $\int_\Omega \boldsymbol{\nabla v} : \boldsymbol{\tau}\, d\Omega$, we get

$$\int_\Omega \rho \boldsymbol{v}\cdot\boldsymbol{b}\, d\Omega + \int_{\Gamma_t} \boldsymbol{v}\cdot\bar{\boldsymbol{t}}\, d\Gamma - \int_\Omega \boldsymbol{\nabla}\cdot(\boldsymbol{\tau}^t\boldsymbol{v})\, d\Omega + \int_\Omega \boldsymbol{v}\cdot(\boldsymbol{\nabla}\cdot\boldsymbol{\tau})\, d\Omega = 0 \quad \forall \boldsymbol{v} \in V_u.$$

Now using the divergence theorem, and the fact that $\boldsymbol{v} = \boldsymbol{0}$ on Γ_u, we get

$$\int_\Omega \boldsymbol{v}\cdot(\boldsymbol{\nabla}\cdot\boldsymbol{\tau} + \rho\boldsymbol{b})\, d\Omega + \int_{\Gamma_t} \boldsymbol{v}\cdot(\bar{\boldsymbol{t}} - \boldsymbol{\tau}\boldsymbol{n})\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u. \tag{1.32}$$

Since the variations $\boldsymbol{v}$ are arbitrary, assume them to be zero on Γ_t, but not zero inside the body. Then

$$\int_\Omega \boldsymbol{v}\cdot(\boldsymbol{\nabla}\cdot\boldsymbol{\tau} + \rho\boldsymbol{b})\, d\Omega = 0,$$

which in turn implies that Eqn. 1.22 holds. Now Eqn. 1.32 reduces to

$$\int_{\Gamma_t} \boldsymbol{v}\cdot(\bar{\boldsymbol{t}} - \boldsymbol{\tau}\boldsymbol{n})\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u.$$

Again using the fact that the variations are arbitrary yields Eqn. 1.25.

Note that the principle of virtual work is valid for *any arbitrary* constitutive relation since we have not made use of Eqn. 1.23 so far. Hence, we can use the principle of virtual work as stated above for nonlinear elasticity, plasticity etc., provided the strains are 'small' (since we have used Eqn. 1.24 which is valid only for small displacement gradients). The variational statement for a *linear* elastic material is obtained by substituting Eqn. 1.23 in Eqn. 1.31. We get the corresponding problem statement as

Find $\boldsymbol{u} \in L_u$ such that

$$\int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{C}\boldsymbol{\epsilon}(\boldsymbol{u})\, d\Omega = \int_{\Gamma_t} \boldsymbol{v} \cdot \bar{\boldsymbol{t}}\, d\Gamma + \int_\Omega \left[\rho \boldsymbol{v} \cdot \boldsymbol{b} + \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{C}\boldsymbol{\epsilon}^0 - \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{\tau}^0\right] d\Omega \quad \forall \boldsymbol{v} \in V_u. \quad (1.33)$$

For elastic bodies, we can formulate the method of total potential energy which is equivalent to the variational formulation already presented as we now show

1.4.3 Principle of minimum potential energy

For a hyperelastic body (not necessarily *linear* elastic), we have the existence of a strain-energy density function $\mathcal{U}$ such that

$$\boldsymbol{\tau} = \frac{\partial \mathcal{U}}{\partial \boldsymbol{\epsilon}}, \quad (1.34)$$

or, in component form,

$$\tau_{ij} = \frac{\partial \mathcal{U}}{\partial \epsilon_{ij}}.$$

As an example, for a linear elastic body with the constitutive equation given by Eqn. 1.23, we have

$$\mathcal{U} = \frac{1}{2}\boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}(\boldsymbol{u}) + \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{\tau}^0 - \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}^0.$$

Using Eqn. 1.34, Eqn. 1.31 can be written as

$$\begin{aligned}\int_\Omega \rho \boldsymbol{v} \cdot \boldsymbol{b}\, d\Omega + \int_{\Gamma_t} \boldsymbol{v} \cdot \bar{\boldsymbol{t}}\, d\Gamma &= \int_\Omega \frac{\partial \mathcal{U}}{\partial \boldsymbol{\epsilon}} : \boldsymbol{\epsilon}(\boldsymbol{v})\, d\Omega \\ &= \int_\Omega \delta^{(1)}\mathcal{U}\, d\Omega \\ &= \delta^{(1)} \int_\Omega \mathcal{U}\, d\Omega \quad \forall \boldsymbol{v} \in V_u,\end{aligned}$$

where we have used the fact that the delta-operator and integral commute. Using the same fact on the left hand side of the above equation, we get

$$\delta^{(1)} \left[\int_\Omega \rho \boldsymbol{b} \cdot \boldsymbol{u}\, d\Omega + \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{u}\, d\Gamma\right] = \delta^{(1)} U.$$

or, alternatively,

$$\delta^{(1)}\left[U - V\right] = \delta^{(1)}\Pi = 0,$$

where Π is the *potential energy* defined as

$$\Pi = U - V, \tag{1.35}$$

with

$$U = \int_{\Omega} \mathcal{U}\, d\Omega,$$
$$V = \int_{\Omega} \rho \boldsymbol{b} \cdot \boldsymbol{u}\, d\Omega + \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{u}\, d\Gamma.$$

As an example, the potential energy for a linear elastic material is

$$\Pi = \frac{1}{2}\int_{\Omega} \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}(\boldsymbol{u})\, d\Omega - \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{u}\, d\Gamma - \int_{\Omega} \rho \boldsymbol{b} \cdot \boldsymbol{u}\, d\Omega + \int_{\Omega} \left[\boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{\tau}^0 - \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}^0\right]\, d\Omega \tag{1.36}$$

We have shown that the principle of virtual work is equivalent to the vanishing of the first variation of the potential energy. We now show that the converse is also true. i.e., $\delta^{(1)}\Pi = 0$ implies the principle of virtual work. Thus, assume that $\delta^{(1)}\Pi = \delta^{(1)}\left[U - V\right] = 0$ holds, or alternatively,

$$\int_{\Omega} \delta^{(1)}\mathcal{U}\, d\Omega - \int_{\Omega} \rho \boldsymbol{b} \cdot \boldsymbol{v}\, d\Omega - \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{v}\, d\Gamma = 0 \quad \forall \boldsymbol{v} \in V_u.$$

Using the fact that

$$\delta^{(1)}\mathcal{U} = \frac{\partial \mathcal{U}}{\partial \boldsymbol{\epsilon}} : \delta\boldsymbol{\epsilon} = \boldsymbol{\tau} : \boldsymbol{\epsilon}(\boldsymbol{v}),$$

we get the principle of virtual work given by Eqn. 1.31.

We have shown that the principle of virtual work is equivalent to extremizing Π. In fact, more is true. By using the positive-definiteness of the strain-energy density function, $\mathcal{U}$, one can show that the principle of virtual work is equivalent to *minimizing* the potential energy with respect to the displacement field (see [2]).

1.5 Variational Formulation for Heat Transfer Problems

As another example of how to formulate the variational problem for a given set of differential equations, we consider the problem of steady state heat transfer with only the conduction and convection modes (the radiation mode makes the problem nonlinear). Analogous to the presentation for elasticity, we first present the strong forms of the governing equations, and then show how the variational formulations can be derived from it. Note that in contrast to the elasticity problem, the independent solution field (the temperature) is a scalar, so that its variation is also a scalar field.

1.5.1 Strong form of the governing equations

Let Ω be an open domain whose boundary Γ is composed of two open, disjoint regions, $\Gamma = \overline{\Gamma_T \cup \Gamma_q}$. The governing equations are

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{q} &= Q \quad \text{on } \Omega, \\
\boldsymbol{q} &= -\boldsymbol{k}\boldsymbol{\nabla} T \quad \text{on } \Omega, \\
T &= \bar{T} \quad \text{on } \Gamma_T, \\
h(T - T_\infty) + (\boldsymbol{k}\boldsymbol{\nabla} T) \cdot \boldsymbol{n} &= \bar{q} \quad \text{on } \Gamma_q,
\end{aligned}$$

where T is the absolute temperature, $\boldsymbol{q}$ is the heat flux, Q is the heat generated per unit volume, $\boldsymbol{k}$ is a symmetric second-order tensor known as the thermal conductivity tensor, $\boldsymbol{n}$ is the outward normal to Γ, T_∞ is the absolute temperature of the surrounding medium (also known as the ambient temperature), $\bar{T}$ is the prescribed temperature on Γ_T, and $\bar{q}$ is the externally applied heat flux on Γ_q. Note that $\bar{q}$ is taken as positive when heat is supplied to the body. Typical units of the field variables and material constants in the SI system are

T : °C, k_{ij} : W/(m-°C), Q : W/m^3, q : W/m^2, h : W/(m^2-°C).

For presenting the variational formulation, we define the function spaces

$$\begin{aligned}
V_T &= \{v \in H^1(\Omega);\ v = 0 \text{ on } \Gamma_T\}, \\
L_T &= \tilde{T} + V_T \text{ where } \tilde{T} \in H^1(\Omega) \text{ satisfies } \tilde{T} = \bar{T} \text{ on } \Gamma_T.
\end{aligned}$$

1.5.2 Variational formulation

Similar to the procedure for elasticity, we set the weighted residual of the governing equation and the natural boundary condition to zero:

$$\int_\Omega v\left[\boldsymbol{\nabla} \cdot (\boldsymbol{k}\boldsymbol{\nabla} T) + Q\right] d\Omega + \int_{\Gamma_q} v\left[\bar{q} - h(T - T_\infty) - (\boldsymbol{k}\boldsymbol{\nabla} T) \cdot \boldsymbol{n}\right] d\Gamma = 0 \quad \forall v \in V_T.$$

Using the identity

$$v(\boldsymbol{\nabla} \cdot (\boldsymbol{k}\boldsymbol{\nabla} T)) = \boldsymbol{\nabla} \cdot (v\boldsymbol{k}\boldsymbol{\nabla} T) - \boldsymbol{\nabla} v \cdot (\boldsymbol{k}\boldsymbol{\nabla} T),$$

the above equation can be written as

$$\begin{aligned}
\int_\Omega \boldsymbol{\nabla} \cdot (v\boldsymbol{k}\boldsymbol{\nabla} T)\, d\Omega - \int_\Omega \left[\boldsymbol{\nabla} v \cdot (\boldsymbol{k}\boldsymbol{\nabla} T) - vQ\right] d\Omega + \\
\int_{\Gamma_q} v\left[\bar{q} - h(T - T_\infty) - (\boldsymbol{k}\boldsymbol{\nabla} T) \cdot \boldsymbol{n}\right] d\Gamma = 0 \quad \forall v \in V_T.
\end{aligned}$$

Applying the divergence theorem to the first term on the left hand side, and noting that $v = 0$ on Γ_T, we get

$$\int_\Omega \left[\boldsymbol{\nabla} v \cdot (\boldsymbol{k}\boldsymbol{\nabla} T) - vQ\right] d\Omega + \int_{\Gamma_q} v\left[-\bar{q} + h(T - T_\infty)\right] d\Gamma = 0 \quad \forall v \in V_T.$$

Taking the known terms to the right hand side, we can write the variational statement as

Find $T \in L_T$ such that

$$\int_\Omega \boldsymbol{\nabla} v \cdot (\boldsymbol{k} \boldsymbol{\nabla} T)\, d\Omega + \int_{\Gamma_q} vhT\, d\Gamma = \int_\Omega vQ\, d\Omega + \int_{\Gamma_q} v(\bar{q} + hT_\infty)\, d\Gamma \quad \forall v \in V_T. \tag{1.37}$$

We have shown how the variational form can be obtained from the strong form of the governing equations. By reversing the steps in the above proof, we can prove the converse, i.e., the governing differential equations can be obtained starting from the variational formulation, provided the solution has sufficient regularity. The energy functional corresponding to the weak form is

$$\Pi = \frac{1}{2} \int_\Omega \left[\boldsymbol{\nabla} T \cdot (\boldsymbol{k} \boldsymbol{\nabla} T) - QT \right] d\Omega + \int_{\Gamma_q} \left[\frac{1}{2} hT^2 - hTT_\infty - \bar{q}T \right] d\Gamma. \tag{1.38}$$

One can easily verify that $\delta^{(1)}\Pi = 0$ yields the variational formulation, and vice versa. By using the positive-definiteness of $\boldsymbol{k}$, we can show that the variational formulation is equivalent to minimizing the above energy functional.

We have seen that it is possible to give an 'energy formulation' which is equivalent to the variational formulation, both in the case of the elasticity and heat transfer problems. So the following questions arise

1. Is it possible to provide an abstract framework for representing problems in different areas such as elasticity and heat transfer?

2. Can one find the appropriate energy functional from the variational statement in a systematic way?

The answer to these questions is that an abstract formulation is possible for *self-adjoint* and *positive-definite* bilinear operators. The elasticity and heat transfer problems already considered are examples which fall under this category. There are several other problems from fluid mechanics and elasticity which also fall in this category. An abstract formulation provides a unified treatment of such apparently diverse problems. With the help of the abstract formulation, one can not only derive the energy functional in a systematic way, but also prove uniqueness of solutions. In addition, it is also possible to give an abstract formulation for the finite element method based on such an abstract formulation which makes it possible to derive valuable information such as error estimates without having to treat each of the problems, whether they be from elasticity, fluid mechanics or heat transfer, individually. We now discuss the details of such a formulation.

1.6 An Abstract Formulation

Given that V is a Hilbert space (for details on Hilbert spaces, see any book on functional analysis), suppose the differential operator $Lu = f$ can be

expressed in the form

Problem (V): Find u such that

$$a(u,v) = L(v) \quad \forall v \in V$$

where $a(.,.)$ is a bilinear operator, and $L(.)$ is a linear operator with the following properties:

1. $a(.,.)$ is symmetric (or self-adjoint).
2. $a(.,.)$ is continuous, i.e.,
$$a(v,w) \leq \gamma \|v\|_V \|w\|_V ,$$
where γ is a positive constant.
3. $a(.,.)$ is V-elliptic (or positive-definite), i.e., there is a constant $\alpha > 0$ such that
$$a(v,v) \geq \alpha \|v\|_V^2 .$$
4. $L(v)$ is continuous, i.e., there is a constant λ such that
$$L(v) \leq \lambda \|v\| .$$

Consider the following minimization problem

Problem (M): Find u such that

$$F(u) = \min_v F(v)$$

where

$$F(v) = \frac{1}{2}a(v,v) - L(v).$$

We have the following result relating problems (V) and (M):

Theorem 1.6.1. *The minimization problem (M) is equivalent to problem (V). The solution to Problems (M) and (V) is unique.*

Proof. First we prove that M $\implies$ V. Let $v \in V$ and $\epsilon \in \Re$. Since u is a minimum,

$$F(u) \leq F(u + \epsilon v) \quad \forall \epsilon \in \Re, \ \forall v \in V.$$

Using the notation $g(\epsilon) \equiv F(u + \epsilon v)$, we have

$$g(0) \leq g(\epsilon) \quad \forall \epsilon \in \Re,$$

or, alternatively, $g'(0) = 0$. The function $g(\epsilon)$ can be written as

$$\begin{aligned} g(\epsilon) &= \frac{1}{2}a(u+\epsilon v, u+\epsilon v) - L(u+\epsilon v) \\ &= \frac{1}{2}a(u,u) + \frac{\epsilon}{2}a(u,v) + \frac{\epsilon}{2}a(v,u) + \frac{\epsilon^2}{2}a(v,v) - L(u) - \epsilon L(v) \\ &= \frac{1}{2}a(u,u) - L(u) + \epsilon a(u,v) - \epsilon L(v) + \frac{\epsilon^2}{2}a(v,v), \end{aligned}$$

where we have used the symmetry of $a(.,.)$ in obtaining the last step. Now $g'(0) = 0$ implies $a(u, v) = L(v)$ for all $v \in V$.

Now we prove that V $\implies$ M. Let $\bar{v} = u + \epsilon v$. Then

$$F(\bar{v}) = \frac{1}{2}a(u,u) - L(u) + \epsilon\left[a(u,v) - L(v)\right] + \frac{\epsilon^2}{2}a(v,v).$$

Since $a(u, v) = L(v)$, the above equation reduces to

$$F(\bar{v}) = F(u) + \frac{\epsilon^2}{2}a(v,v) \geq F(u).$$

Noting that $F(\bar{v}) = F(u)$ when $\epsilon = 0$, we can write the above equation as

$$F(u) = \min_{\bar{v}\in V} F(\bar{v}),$$

which is nothing but Problem (M).

To prove uniqueness, assume that u_1 and u_2 are two distinct solutions of problem (V). Then

$$\begin{aligned} a(u_1, v) &= L(v) \quad \forall v \in V, \\ a(u_2, v) &= L(v) \quad \forall v \in V. \end{aligned}$$

Using the fact that $a(.,.)$ is a bilinear operator, we get

$$a(u_1 - u_2, v) = 0 \quad \forall v \in V.$$

Choosing $v = u_1 - u_2$, we get

$$a(u_1 - u_2, u_1 - u_2) = \|u_1 - u_2\|_a = 0.$$

By the V-ellipticity condition, this implies that

$$\|u_1 - u_2\|_V = 0,$$

which by the definition of a norm implies that $u_1 - u_2 = 0$, or that $u_1 = u_2$. □

Note that if $a(.,.)$ is non-symmetric, we still have uniqueness of solutions, but no corresponding minimization problem (M). We now consider some examples to illustrate the above concepts.

Examples

Consider the governing equation of a beam

$$\frac{d^2}{dx^2}\left[EI\frac{d^2v}{dx^2}\right] - q = 0,$$

with the boundary conditions

$$v(0) = \left.\frac{dv}{dx}\right|_{x=0} = 0; \; EI\left.\frac{d^2v}{dx^2}\right|_{x=L} = M_0; \; \left.\frac{d}{dx}\left(EI\frac{d^2v}{dx^2}\right)\right|_{x=L} = V.$$

The variational form is obtained as follows:
First define the appropriate function space in which v lies in:

$$V_u = \left\{\phi \in H^2(0,L), \phi = 0, \phi' = 0 \text{ at } x = 0\right\},$$

where

$$H^2(0,L) = \left\{\phi \in L_2(0,L), \phi' \in L_2(0,L), \phi'' \in L_2(0,L)\right\}.$$

Multiplying the governing differential equation by a test function ϕ and integrating, and adding the residues of the boundary terms, we get

$$\int_0^L \phi \left[\frac{d^2}{dx^2}\left(EI\frac{d^2v}{dx^2}\right) - q\right] dx + \left[\left(EI\frac{d^2v}{dx^2} - M_0\right)\frac{d\phi}{dx}\right]_{x=L} + \left[V - \frac{d}{dx}\left(EI\frac{d^2v}{dx^2}\right)\right]\phi\Big|_{x=L} = 0 \quad \forall \phi \in V_u.$$

Integrating by parts twice, we get

$$\int_0^L EI\frac{d^2v}{dx^2}\frac{d^2\phi}{dx^2}\,dx + \frac{d}{dx}\left(EI\frac{d^2v}{dx^2}\right)\phi\Big|_0^L - EI\frac{d^2v}{dx^2}\frac{d\phi}{dx}\Big|_0^L + \left[\left(EI\frac{d^2v}{dx^2} - M_0\right)\frac{d\phi}{dx}\right]_{x=L} + \left[V - \frac{d}{dx}\left(EI\frac{d^2v}{dx^2}\right)\right]\phi\Big|_{x=L} = \int_0^L q\phi\,dx \quad \forall \phi \in V_u. \tag{1.39}$$

Since $\phi|_{x=0} = d\phi/dx|_{x=0} = 0$, we can write Eqn. 1.39 as

$$a(v,\phi) = L(\phi) \quad \forall \phi \in V_u,$$

where

$$a(v,\phi) = \int_0^L EI\frac{d^2v}{dx^2}\frac{d^2\phi}{dx^2}\,dx,$$
$$L(\phi) = \int_0^L q\phi\,dx + M_0\,\frac{d\phi}{dx}\Big|_{x=L} - V\,\phi|_{x=L}.$$

Clearly $a(.,.)$ is symmetric. To prove V-ellipticity, note that

$$a(v,v) = \int_0^L EI\left(\frac{d^2v}{dx^2}\right)^2 dx > 0 \quad \text{for } v \neq 0.$$

If $a(v,v) = 0$, then $d^2v/dx^2 = 0$, which implies that $v = c_1x + c_2$. For a simply supported beam or a beam cantilevered at one or both ends, we get $c_1 = c_2 = 0$, leading to $v = 0$. Hence, $a(.,.)$ is symmetric and positive-definite. The functional corresponding to the variational form is

$$\Pi = \frac{1}{2}a(v,v) - L(v)$$

$$= \frac{1}{2}\int_0^L EI\left(\frac{d^2v}{dx^2}\right)^2 dx - \int_0^L qv\,dx - M_0\left.\frac{dv}{dx}\right|_{x=L} + V\ v|_{x=L}\,. \qquad (1.40)$$

Now consider the variational formulation for the elasticity problem given by Eqn. 1.33. The bilinear operator, $a(.,.)$ and the linear operator $L(.)$ are given by

$$a(\boldsymbol{u},\boldsymbol{v}) = \int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{C}\boldsymbol{\epsilon}(\boldsymbol{u})\,d\Omega,$$
$$L(\boldsymbol{v}) = \int_{\Gamma_t} \boldsymbol{v}\cdot\bar{\boldsymbol{t}}\,d\Gamma + \int_\Omega \left[\rho\boldsymbol{v}\cdot\boldsymbol{b} + \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{C}\boldsymbol{\epsilon}^0 - \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{\tau}^0\right]\,d\Omega.$$

$a(.,.)$ is clearly symmetric. It is positive definite if $\boldsymbol{C}$ is positive definite. The functional to be minimized is

$$\begin{aligned}\Pi &= \frac{1}{2}a(\boldsymbol{u},\boldsymbol{u}) - L(\boldsymbol{u})\\ &= \frac{1}{2}\int_\Omega \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}(\boldsymbol{u})\,d\Omega - \int_{\Gamma_t}\bar{\boldsymbol{t}}\cdot\boldsymbol{u}\,d\Gamma - \int_\Omega \rho\boldsymbol{b}\cdot\boldsymbol{u}\,d\Omega +\\ &\quad \int_\Omega \left[\boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{\tau}^0 - \boldsymbol{\epsilon}(\boldsymbol{u}) : \boldsymbol{C}\boldsymbol{\epsilon}^0\right]\,d\Omega,\end{aligned}$$

which is the same expression as given by Eqn 1.36.

Finally, consider the variational formulation for the heat transfer problem given by Eqn. 1.37. We now have

$$a(T,v) = \int_\Omega \boldsymbol{\nabla}v\cdot(\boldsymbol{k}\boldsymbol{\nabla}T)\,d\Omega + \int_{\Gamma_q} vhT\,d\Gamma,$$
$$L(v) = \int_\Omega vQ\,d\Omega + \int_{\Gamma_q} v(\bar{q}+hT_\infty)\,d\Gamma.$$

The functional to be minimized is

$$\begin{aligned}\Pi &= \frac{1}{2}a(T,T) - L(T)\\ &= \frac{1}{2}\int_\Omega \left[\boldsymbol{\nabla}T\cdot(\boldsymbol{k}\boldsymbol{\nabla}T) - QT\right]\,d\Omega + \int_{\Gamma_q}\left[\frac{1}{2}hT^2 - hTT_\infty - \bar{q}T\right]\,d\Gamma,\end{aligned}$$

which is the same expression as given by Eqn. 1.38.

1.7 Traditional Approximation Techniques

We discuss two popular traditional approximation techniques, namely, the Rayleigh-Ritz method and the Galerkin method.

1.7.1 Rayleigh-Ritz method

In this method, we find an approximate solution using the weak form of the governing equations, i.e., using either Problem (V) or Problem (M) in

section 1.6. We seek an approximate solution of the form

$$u = \phi_0 + \sum_{j=1}^{n} c_j \phi_j,$$

where ϕ_0 satisfies the essential boundary condition. The functions ϕ_j, $j = 1, \ldots, n$ are zero on that part of the boundary. For example, if $u(x_0) = u_0$ where x_0 is a boundary point then $\phi_0(x_0) = u_0$ and $\phi_i(x_0) = 0$ for $i = 1, \ldots, n$. If all the boundary conditions are homogeneous then $\phi_0 = 0$.

The variation of u is given by $v = \sum_{i=1}^{n} \bar{c}_i \phi_i$. Substituting for u and v in the variational formulation, we get

$$a(\phi_0 + \sum_{j=1}^{n} c_j \phi_j, \sum_{i=1}^{n} \bar{c}_i \phi_i) = \sum_{i=1}^{n} \bar{c}_i L(\phi_i), \quad \forall \bar{c}_i.$$

Simplifying using the bilinearity of $a(.,.)$, we get

$$\sum_{i=1}^{n} \bar{c}_i a(\phi_0, \phi_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{c}_i c_j a(\phi_j, \phi_i) = \sum_{i=1}^{n} \bar{c}_i L(\phi_i), \qquad \forall \bar{c}_i.$$

Choosing first $\bar{c}_1 = 1$ and the remaining $\bar{c}_i$ as zero, then choosing $\bar{c}_2 = 1$ and the remaining $\bar{c}_i$ as zero, and so on, we get

$$\sum_{j=1}^{n} A_{ij} c_j = L(\phi_i) - a(\phi_0, \phi_i).$$

In matrix form, the above equation can be written as

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{f}, \tag{1.41}$$

where

$$A_{ij} = a(\phi_j, \phi_i), \tag{1.42}$$

is a symmetric matrix (since $a(.,.)$ is symmetric), and

$$f_i = L(\phi_i) - a(\phi_0, \phi_i). \tag{1.43}$$

One obtains exactly the same set of equations by using minimization of the functional I given by

$$\begin{aligned} I(u) &= \frac{1}{2} a \left(\phi_0 + \sum_{j=1}^{n} c_j \phi_j, \phi_0 + \sum_{i=1}^{n} c_i \phi_i \right) - L(\phi_0) - \sum_{j=1}^{n} c_j L(\phi_j) \\ &= \frac{1}{2} a(\phi_0, \phi_0) + a(\phi_0, \sum_{i=1}^{n} c_i \phi_i) + \frac{1}{2} a \left(\sum_{j=1}^{n} c_j \phi_j, \sum_{i=1}^{n} c_i \phi_i \right) - L(\phi_0) - \sum_{j=1}^{n} c_j L(\phi_j). \end{aligned}$$

Now setting the partial derivatives of I with respect to the undetermined coefficients, i.e.,

$$\frac{\partial I}{\partial c_1} = 0, \ \frac{\partial I}{\partial c_2} = 0, \ldots, \frac{\partial I}{\partial c_n} = 0,$$

we get Eqn. 1.41. Note however, that when $a(,.,)$ is not symmetric, we do not have the existence of the functional I. Hence, deriving the matrix equations directly from the variational form (V) is preferred.

The approximation functions ϕ_i should satisfy the following conditions:

1. They should be such that $a(\phi_i, \phi_j)$ should be well-defined.
2. The set ϕ_i, $i = 1, \ldots, n$ should be linearly independent.
3. $\{\phi_i\}$ must be complete, e.g., when ϕ_i are algebraic polynomials, completeness requires that the set $\{\phi_i\}$ should contain all terms of the lowest-order order admissible, and upto the highest order desired.

Example:
We shall consider the same example of a beam that we considered in section 1.6. Recall that the variational formulation was

$$a(v, w) = L(w) \quad \forall w \in V_u,$$

where

$$\begin{aligned} a(v, w) &= \int_0^L EI \frac{d^2v}{dx^2} \frac{d^2w}{dx^2}\, dx, \\ L(w) &= \int_0^L qw\, dx + M_0 \left.\frac{dw}{dx}\right|_{x=L}. \end{aligned}$$

Since the specified essential boundary conditions are $v = dv/dx = 0$ at $x = 0$, we can take $\phi_0 = 0$, and select ϕ_i such that $\phi_i(0) = \phi'(0) = 0$. Choosing $\phi_1 = x^2$, $\phi_2 = x^3$ and so on, we can write $v(x)$ as

$$v(x) = \sum_{j=1}^{n} c_j \phi_j(x); \quad \phi_j = x^{j+1}.$$

The governing matrix equations are

$$\boldsymbol{Ac} = \boldsymbol{f},$$

where $\boldsymbol{A}$ and $\boldsymbol{f}$ are obtained using Eqns. 1.42 and 1.43:

$$\begin{aligned} A_{ij} &= \int_0^L EI(i+1)ix^{i-1}(j+1)jx^{j-1}\, dx \\ &= EI\frac{ij(i+1)(j+1)L^{i+j-1}}{i+j-1} \\ f_i &= \frac{qL^{i+2}}{i+2} + M_0(i+1)L^i. \end{aligned}$$

For $n = 2$, we get

$$EI \begin{bmatrix} 4L & 6L^2 \\ 6L^2 & 12L^3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{qL^3}{12} \begin{bmatrix} 4 \\ 3L \end{bmatrix} + M_0 L \begin{bmatrix} 2 \\ 3L \end{bmatrix}.$$

On solving for c_1 and c_2 and substituting in the expression for $v(x)$, we get

$$v(x) = \left[\frac{5qL^2 + 12M_0}{24EI}\right] x^2 - \frac{qL}{12EI}x^3.$$

For $n = 3$, we get

$$EI \begin{bmatrix} 4 & 6L & 8L^2 \\ 6L & 12L^2 & 18L^3 \\ 8L^2 & 18L^3 & 28.8L^4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}qL^2 + 2M_0 \\ \frac{1}{4}qL^3 + 3M_0L \\ \frac{1}{5}qL^4 + 4M_0L^2 \end{bmatrix}.$$

which leads to

$$v(x) = \frac{qx^2}{24EI}(6L^2 - 4Lx + x^2) + \frac{M_0x^2}{2EI}.$$

1.7.2 Galerkin method

The Galerkin method falls under the category of the method of weighted residuals. In the method of weighted residuals, the sum of the weighted residuals of the governing differential equation and the natural boundary conditions is set to zero, i.e.,

$$\int (L\tilde{u} - f)w + \int_\Gamma (B\tilde{u} - t)w = 0,$$

where $\tilde{u}$ is the approximate solution, w is a weighting function, and Γ denotes the boundary over which the natural boundary condition is applied. In the Galerkin method, the weighting function is constructed using the same basis functions as are used for the approximate solution $\tilde{u}$. Thus, if

$$\tilde{u} = \phi_0 + \sum_{i=1}^{n} c_j\phi_j,$$

then w is taken to be of the form

$$w = \sum_{i=1}^{n} \bar{c}\phi_i.$$

If the governing equation permits, one can carry out integration by parts, and transfer the differentiation from u to ϕ as in the Rayleigh-Ritz method. Note, however, that the weighted residual method is more general than the Rayleigh-Ritz method since no variational form is required. Also note that the approximating functions in the Galerkin method are of a higher order than in the Rayleigh-Ritz method. The Rayleigh-Ritz and Galerkin methods yield identical results for a certain class of problems, e.g., self-adjoint, second-order ordinary differential equations with homogeneous or non-homogeneous boundary conditions or the linear elasticity problem. But, in general, they can yield different results.

As an example, consider the ordinary differential equation:

$$x^2\frac{d^2y}{dx^2} + 2x\frac{dy}{dx} - 6x = 0,$$

with the boundary conditions $y(1) = y(2) = 0$. The exact solution of the above equation is

$$y = \frac{6}{x} + 3x - 9.$$

In the Galerkin method, we choose an approximate solution in the form of a polynomial which satisfies the two essential boundary conditions:

$$\tilde{y} = c_1\phi_1 = c_1(x^2 - 3x + 2).$$

Corresponding to this solution, we have

$$w = \bar{c}_1(x^2 - 3x + 2).$$

The weighted residual statement for the choice $\bar{c}_1 = 1$ is given by

$$\int_1^2 \left(x^2\frac{d^2\tilde{y}}{dx^2} + 2x\frac{d\tilde{y}}{dx} - 6x\right)\phi_1\, dx = 0,$$

which in turn yields

$$\int_1^2 \left(x^2\phi_1\frac{d^2\phi_1}{dx^2}c_1 + 2x\phi_1\frac{d\phi_1}{dx}c_1 - 6x\phi_1\right)\, dx = 0.$$

Solving the above equation, we get $c_1 = 1.875$. To improve the accuracy, we can take more terms in the expression for $\tilde{y}$, e.g.,

$$\tilde{y} = (x-1)(x-2)(c_1 + c_2x).$$

One can compute c_1 and c_2 using the Galerkin method and confirm that the Rayleigh-Ritz method also gives the same results.

1.8 Drawbacks of the Traditional Variational Formulations

We have studied two representative traditional variational methods for finding approximate solutions to differential equations, namely the Rayleigh-Ritz method and the Galerkin method. However, all such traditional methods suffer from drawbacks which we now discuss. Later, we shall see how the finite element method overcomes these drawbacks.

The main difficulty in the traditional methods is in constructing the approximation functions. When the given domain is geometrically complex, the selection becomes even more difficult. An effective computational method should have the following features:

1. The method should be convergent.

2. It should be easy to implement, even with complex geometries.

3. The formulation procedure should be independent of the shape of the domain and the specific form of the boundary conditions.

4. It should involve a systematic procedure which can be easily implemented.

The finite element method meets these requirements. In this method, we divide the domain into geometrically simple domains. The approximation functions (also known as basis functions or shape functions) are often algebraic polynomials that are derived using interpolation theory. Once the approximation functions have been derived, the method of obtaining the unknown coefficients is exactly the same as in the Rayleigh-Ritz or Galerkin method.

Chapter 2

Formulation of the Finite Element Method

In this chapter, we discuss the formulation of the finite element method, first in an abstract setting, and then applied to problems in elasticity and heat transfer.

2.1 Steps Involved in the Finite Element Method

The steps involved in a finite element formulation are as follows:

1. Discretize the domain into a collection of elements. Number the nodes of the elements and generate the geometric properties such as coordinates needed for the problem.

2. Assemble the element level matrices to form the global equation. For example, for a linear statics problem, we get a matrix equation of the form
$$\boldsymbol{K}\boldsymbol{u} = \boldsymbol{f}.$$

3. Impose the boundary conditions.

4. Solve the assembled equations.

5. Carry out the postprocessing of the results.

2.2 Abstract Formulation

Before showing how the finite element formulation is carried out for elasticity or heat transfer, we show how it is formulated in an abstract setting, so that later it is clear how the various special formulations fit into this more general framework.

First choose a finite-dimensional subspace V_h of V with dimension $m+n$, where m is the prescribed degrees of freedom, and n is the free degrees of

freedom to be solved for. Let $(\phi_1, \phi_2, \ldots, \phi_{n+m})$ be the basis for V_h so that any $v \in V_h$ has the unique representation

$$v = \sum_{i=1}^{n+m} \eta_i \phi_i, \quad \eta_i \in \Re.$$

The discrete analogue of problems (M) and (V) are
Find $u_h \in V_h$ such that

$$F(u_h) \leq F(v) \quad \forall v \in V_h.$$

or, equivalently,

$$a(u_h, v) = L(v) \quad \forall v \in V_h. \tag{2.1}$$

The approximate solution, u_h, is of the form

$$u_h = \sum_{i=1}^{m} (\phi_0)_i \bar{\xi}_i + \sum_{i=1}^{n} \phi_i \xi_i,$$

and the corresponding variation is of the form

$$v = \sum_{i=1}^{n} \phi_i \eta_i.$$

Substituting for u_h and v in the finite dimensional version of Problem (V) given by Eqn. 2.1, and successively choosing $\boldsymbol{\eta}$ as $(1, 0, 0, \ldots)$, $(0, 1, 0, \ldots)$ etc., we get

$$a(\sum_{i=1}^{n} \phi_i \xi_i, \phi_j) + a(\sum_{i=1}^{m} (\phi_0)_i \bar{\xi}_i, \phi_j) = L(\phi_j) \quad j = 1, n.$$

Using the bilinearity of $a(.,.)$, we can write the above equation as

$$\sum_{i=1}^{n} a(\phi_i, \phi_j) \xi_i + \sum_{i=1}^{m} a((\phi_0)_i, \phi_j) \bar{\xi}_i = L(\phi_j) \quad j = 1, n,$$

or in matrix form as

$$\boldsymbol{A\xi} = \boldsymbol{b} - \bar{\boldsymbol{A}}\bar{\boldsymbol{\xi}},$$

where $\bar{\boldsymbol{\xi}}$ is an $m \times 1$ vector, $\boldsymbol{\xi}$ is an $n \times 1$ vector, and

$$\begin{aligned} A_{ji} &= a(\phi_i, \phi_j) && \text{(n×n)}, \\ \bar{A}_{ji} &= a((\phi_0)_i, \phi_j) && \text{(n×m)}, \\ b_j &= L(\phi_j) && \text{(n×1)}. \end{aligned}$$

Note that the stiffness matrix $\boldsymbol{A}$ is symmetric and positive definite since $a(.,.)$ is symmetric and V-elliptic:

$$\boldsymbol{\eta}^t \boldsymbol{A\eta} = a(v, v) \geq \alpha \|v\|^2 \geq 0 \quad \forall \boldsymbol{\eta}.$$

2.3 Finite Element Formulation for Elasticity

In this section, we consider the finite element formulation for the linear elasticity problem based on Eqn. 1.33. Observe that this variational formulation involves dealing with fourth-order tensors such as the elasticity tensor, $\boldsymbol{C}$, whose matrix representation is given by a $3 \times 3 \times 3 \times 3$ matrix. Thus, carrying out the assembly of the stiffness matrix and force vector in a computer implementation based on Eqn. 1.33 would be quite cumbersome. What we need is an equivalent form of Eqn. 1.33 which would lead to simpler matrix manipulations in a computer implementation.

Towards this end, using the symmetry of the strain and stress tensors, we replace them by vectors such that the strain energy density expression in terms of the stress and strain components remains the same. This leads to the use of 'engineering strain' components instead of the tensorial components. For example, if we consider two-dimensional plane stress/plane strain problems, then we have

$$\begin{aligned} \boldsymbol{\epsilon} : \boldsymbol{\tau} &= \tau_{xx}\epsilon_{xx} + \tau_{xy}\epsilon_{xy} + \tau_{yx}\epsilon_{yx} + \tau_{yy}\epsilon_{yy} \\ &= \tau_{xx}\epsilon_{xx} + 2\tau_{xy}\epsilon_{xy} + \tau_{yy}\epsilon_{yy} \\ &= \boldsymbol{\epsilon}_c^t \boldsymbol{\tau}_c, \end{aligned}$$

where (with $\gamma_{xy} = 2\epsilon_{xy}$)

$$\boldsymbol{\epsilon}_c = \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \gamma_{xy} \end{pmatrix}; \qquad \boldsymbol{\tau}_c = \begin{pmatrix} \tau_{xx} \\ \tau_{yy} \\ \tau_{xy} \end{pmatrix}.$$

Note that $\boldsymbol{\epsilon}_c$ is the vector of engineering strain components. The subscript c on $\boldsymbol{\epsilon}$ and $\boldsymbol{\tau}$ in the above equation, denotes that the strain and stress components are written in 'column' form. The elasticity tensors for an isotropic material in the case of plane stress/plane strain are 3×3 matrices given by

$$\boldsymbol{C}_{\text{p. stress}} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix}; \quad \boldsymbol{C}_{\text{p. strain}} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix}.$$

Note that we now simply have

$$\boldsymbol{\tau}_c = \boldsymbol{C}(\boldsymbol{\epsilon}_c - \boldsymbol{\epsilon}_c^0) + \boldsymbol{\tau}_c^0, \tag{2.2}$$

instead of the earlier relation $\boldsymbol{\tau} = \boldsymbol{C}(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^0) + \boldsymbol{\tau}^0$. The virtual work statement given by Eqn. 1.33 written in terms of $\boldsymbol{\epsilon}_c$ and $\boldsymbol{\tau}_c$ is

Find $\boldsymbol{u} \in L_u$ such that

$$\int_\Omega [\boldsymbol{\epsilon}_c(\boldsymbol{v})]^t \boldsymbol{C} \boldsymbol{\epsilon}_c(\boldsymbol{u})\, d\Omega = \int_{\Gamma_t} \boldsymbol{v}^t \bar{\boldsymbol{t}}\, d\Gamma +$$

$$\int_\Omega \left[\rho \boldsymbol{v}^t \boldsymbol{b} + [\boldsymbol{\epsilon}_c(\boldsymbol{v})]^t \boldsymbol{C} \boldsymbol{\epsilon}_c^0 - [\boldsymbol{\epsilon}_c(\boldsymbol{v})]^t \boldsymbol{\tau}_c^0\right] \, d\Omega \quad \forall \boldsymbol{v} \in V_u. \quad (2.3)$$

To establish the finite element formulation, we introduce the *interpolation functions* or *shape functions*, $\boldsymbol{N}$, and write the displacement field, $\boldsymbol{u}$, as

$$\boldsymbol{u}_h = \boldsymbol{N}\hat{\boldsymbol{u}} + \bar{\boldsymbol{N}}\bar{\hat{\boldsymbol{u}}}.$$

In the above equation, $\boldsymbol{N}$ are the shape functions associated with the vector of *unknown* displacement degrees of freedom $\hat{\boldsymbol{u}}$, while $\bar{\boldsymbol{N}}$ are the shape functions associated with the vector of *known* displacement degrees of freedom, $\bar{\hat{\boldsymbol{u}}}$, on Γ_u. Typically, the prescribed displacement, $\bar{\boldsymbol{u}}$, is zero, and then $\bar{\hat{\boldsymbol{u}}} = \boldsymbol{0}$. Note that the shape functions are formulated directly over the entire domain, though for *convenience* we shall assemble matrices formulated at an element level in an actual implementation (see Section 2.5 for an example).

Since the displacement is prescribed on Γ_u, the variation $\boldsymbol{v}$ there is zero. Hence, the variation field, $\boldsymbol{v}$, expressed using the same shape functions as those used for the displacement field, is

$$\boldsymbol{v} = \boldsymbol{N}\hat{\boldsymbol{v}}.$$

Substituting the displacement and variation fields in the strain-displacement relation, we get

$$\begin{aligned} \boldsymbol{\epsilon}_c(\boldsymbol{u}) &= \boldsymbol{B}\hat{\boldsymbol{u}} + \bar{\boldsymbol{B}}\bar{\hat{\boldsymbol{u}}}, \\ \boldsymbol{\epsilon}_c(\boldsymbol{v}) &= \boldsymbol{B}\hat{\boldsymbol{v}}, \end{aligned}$$

where $\boldsymbol{B}$ is known as the *strain-displacement matrix*. $\bar{\boldsymbol{B}}$ is the strain-displacement matrix associated with the shape functions $\bar{\boldsymbol{N}}$. The stress tensor is obtained by substituting for $\boldsymbol{\epsilon}_c$ in Eqn. 2.2:

$$\boldsymbol{\tau}_c = \boldsymbol{C}\boldsymbol{B}\hat{\boldsymbol{u}} - \boldsymbol{C}\boldsymbol{\epsilon}^0 + \boldsymbol{\tau}^0 + \boldsymbol{C}\bar{\boldsymbol{B}}\bar{\hat{\boldsymbol{u}}}.$$

Let V_h be a finite-dimensional subspace of V_u. Substituting the above quantities in Eqn. 2.3, the finite element problem can be stated as

Find $\boldsymbol{u}_h \in V_h$ such that

$$\hat{\boldsymbol{v}}^t \left[\int_\Omega \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, d\Omega\right] \hat{\boldsymbol{u}} = \hat{\boldsymbol{v}}^t \Big[\int_{\Gamma_t} \boldsymbol{N}^t \bar{\boldsymbol{t}} \, d\Gamma + \int_\Omega (\rho \boldsymbol{N}^t \boldsymbol{b} + \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{\epsilon}_c^0 - \boldsymbol{B}^t \boldsymbol{\tau}_c^0) \, d\Omega - \int_\Omega \boldsymbol{B}^t \boldsymbol{C} \bar{\boldsymbol{B}} \, d\Omega \, \bar{\hat{\boldsymbol{u}}}\Big] \quad \forall \hat{\boldsymbol{v}}. \quad (2.4)$$

Since the above equation is valid for *all* vectors $\hat{\boldsymbol{v}}$, we can write the finite element formulation as

Find $\hat{\boldsymbol{u}}$ which satisfies

$$\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}, \quad (2.5)$$

where

$$\boldsymbol{K} = \int_\Omega \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, d\Omega,$$

$$\boldsymbol{f} = \int_{\Gamma_t} \boldsymbol{N}^t \bar{\boldsymbol{t}}\, d\Gamma + \int_{\Omega} (\rho \boldsymbol{N}^t \boldsymbol{b} + \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{\epsilon}_c^0 - \boldsymbol{B}^t \boldsymbol{\tau}_c^0)\, d\Omega - \left(\int_{\Omega} \boldsymbol{B}^t \boldsymbol{C} \bar{\boldsymbol{B}}\, d\Omega \right) \bar{\hat{\boldsymbol{u}}}.$$

Note that the stiffness matrix, $\boldsymbol{K}$, is symmetric and positive-definite since the elasticity matrix, $\boldsymbol{C}$, is symmetric and positive-definite. Note also that a nonzero prescribed displacement contributes to the load vector via the last term in $\boldsymbol{f}$. The force vector $\boldsymbol{f}$ is known as the *work-equivalent* load vector, since it is derived from the virtual work principle. The above way of formulating the load vector is the variationally correct way in which a distributed load on the actual structure should be converted to point loads at the nodes in the finite element model. If n is the number of free degrees of freedom and m is the number of prescribed degrees of freedom in a plane-stress/plane-strain problem, then the order of the various matrices is as follows: $\boldsymbol{K} : n \times n$, $\hat{\boldsymbol{u}} : n \times 1$, $\boldsymbol{f} : n \times 1$, $\bar{\hat{\boldsymbol{u}}} : m \times 1$, $\bar{\boldsymbol{B}} : 3 \times m$, $\boldsymbol{B} : 3 \times n$, $\boldsymbol{C} : 3 \times 3$, $\boldsymbol{N} : 2 \times n$.

It is also possible to derive Eqn. 2.5 using the potential energy functional, Π. First we write the expression for Π given by Eqn. 1.36 in terms of $\boldsymbol{\epsilon}_c$. We get

$$\Pi = \frac{1}{2} \int_{\Omega} [\boldsymbol{\epsilon}_c(\boldsymbol{u})]^t \boldsymbol{C} \boldsymbol{\epsilon}_c(\boldsymbol{u}) - \int_{\Gamma_t} \boldsymbol{u}^t \bar{\boldsymbol{t}}\, d\Gamma - \int_{\Omega} \rho \boldsymbol{u}^t \boldsymbol{b}\, d\Omega + \int_{\Omega} \left[[\boldsymbol{\epsilon}_c(\boldsymbol{u})]^t \boldsymbol{\tau}_c^0 - [\boldsymbol{\epsilon}_c(\boldsymbol{u})]^t \boldsymbol{C} \boldsymbol{\epsilon}_c^0 \right] d\Omega. \quad (2.6)$$

Substituting for $\boldsymbol{u}$ in Eqn. 2.6, we get

$$\Pi = \frac{1}{2} \hat{\boldsymbol{u}}^t \boldsymbol{K} \hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}^t \boldsymbol{f}. \quad (2.7)$$

We carry out the minimization of Π with respect to $\hat{\boldsymbol{u}}$ using indicial notation. We have

$$\Pi = \frac{1}{2} \hat{u}_i K_{ij} \hat{u}_j - \hat{u}_i f_i,$$

so that

$$\begin{aligned} \frac{\partial \Pi}{\partial \hat{u}_m} &= \frac{1}{2} \delta_{im} K_{ij} \hat{u}_j + \frac{1}{2} \hat{u}_i K_{ij} \delta_{jm} - \delta_{im} f_i \\ &= \frac{1}{2} K_{mj} \hat{u}_j + \frac{1}{2} K_{im} \hat{u}_i - f_m \\ &= K_{im} \hat{u}_i - f_m. \quad \text{(since } \boldsymbol{K} \text{ is symmetric)} \end{aligned}$$

For minimizing Π, we set $\partial \Pi / \partial \hat{u}_m = 0$ which yields Eqn. 2.5. Note that by using the variational formulation, we obtained Eqn. 2.5 directly without having to carry out any differentiation.

2.4 Finite Element Formulation for Heat-Transfer Problems

The procedure for deriving the finite element formulation for the heat transfer problem is analogous to that of elasticity. Note that for the purposes of

deriving the finite element formulation, we can directly make use of the variational statement given by Eqn. 1.37 since the thermal conductivity tensor is a second-order tensor (as opposed to its counterpart, the elasticity tensor, which is a fourth-order tensor).

The temperature T written in terms of the interpolation functions is

$$T = \boldsymbol{N}\hat{\boldsymbol{w}} + \bar{\boldsymbol{N}}\bar{\hat{\boldsymbol{w}}},$$

where $\hat{\boldsymbol{w}}$ is the vector of the unknown temperatures at the nodes, $\boldsymbol{N}$ are the shape functions associated with the nodes where the temperature is unknown, $\bar{\hat{\boldsymbol{w}}}$ is the vector of the prescribed temperatures at the nodes lying on Γ_T, and $\bar{\boldsymbol{N}}$ are the corresponding shape functions. Noting that the variation, v, of the temperature vanishes on Γ_T, we have

$$v = \boldsymbol{N}\hat{\boldsymbol{v}}.$$

The gradient of T is given by

$$\boldsymbol{\nabla} T = \boldsymbol{B}\hat{\boldsymbol{w}} + \bar{\boldsymbol{B}}\bar{\hat{\boldsymbol{w}}},$$

where

$$B_{ij} = \frac{\partial N_j}{\partial x_i}$$
$$\bar{B}_{ij} = \frac{\partial \bar{N}_j}{\partial x_i}.$$

Substituting for these quantities in Eqn. 1.37, and using the arbitrariness of $\hat{\boldsymbol{v}}$, we get

Find $\hat{\boldsymbol{w}}$ such that

$$\boldsymbol{K}\hat{\boldsymbol{w}} = \boldsymbol{f},$$

where

$$\boldsymbol{K} = \int_\Omega \boldsymbol{B}^t \boldsymbol{k} \boldsymbol{B}\, d\Omega + \int_{\Gamma_q} h \boldsymbol{N}^t \boldsymbol{N}\, d\Gamma,$$
$$\boldsymbol{f} = \int_\Omega \boldsymbol{N}^t Q\, d\Omega + \int_{\Gamma_q} \boldsymbol{N}^t (\bar{q} + hT_\infty)\, d\Gamma - \left(\int_\Omega \boldsymbol{B}^t \boldsymbol{k} \bar{\boldsymbol{B}}\, d\Omega \right) \bar{\hat{\boldsymbol{w}}}.$$

Note that $\boldsymbol{K}$ is symmetric and positive definite by virtue of the symmetry and positive-definiteness of $\boldsymbol{k}$. Also note that the second term in the 'stiffness' matrix, $\boldsymbol{K}$, involves an integral over part of the surface of the domain. There is no corresponding term in the expression for the stiffness matrix in the elasticity problem. If n is the number of free degrees of freedom and m the number of prescribed degrees of freedom, then the order of the various matrices in a three-dimensional problem is as follows: $\boldsymbol{K} : n \times n$, $\hat{\boldsymbol{w}} : n \times 1$, $\boldsymbol{f} : n \times 1$, $\boldsymbol{N} : 1 \times n$, $\boldsymbol{B} : 3 \times n$, $\bar{\boldsymbol{B}} : 3 \times m$, $\bar{\hat{\boldsymbol{w}}} : m \times 1$, $\bar{\boldsymbol{N}} : 1 \times m$.

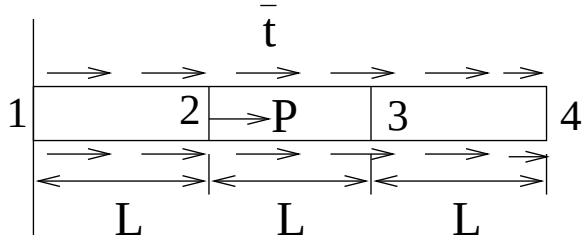


Figure 2.1: Bar subjected to uniaxial loading.

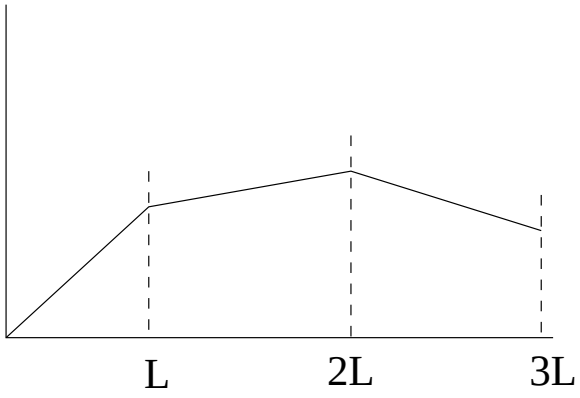


Figure 2.2: Example of a piecewise linear function.

2.5 A Model Problem

To illustrate the 'direct' formulation of the finite element equations over the entire domain, we consider the simple one-dimensional problem shown in Fig. 2.1. A bar of length $3L$, with Young's modulus E and cross-sectional area A is subjected to a surface force per unit length $\bar{t}$, a body force per unit mass, b, along its length, and a point force P at $x = L/3$. Assuming that E and A are constant, the governing differential equation is

$$E\frac{d^2u}{dx^2} + \rho b = 0.$$

The stress-strain and strain-displacement relations are $\tau = E\epsilon$ and $\epsilon = du/dx$.

We discretize the domain into 3 elements, each of length L, as shown in Fig. 2.1. We are interested in finding the approximate solution u_h in the space of *piecewise linear functions*, i.e.,

$$V_h \equiv \text{Space of piecewise linear functions.}$$

An example of a function in V_h is as shown in Fig. 2.2. Note that we need to satisfy $V_h \subset H^1(\Omega)$ or $V_h \subset H^2(\Omega)$ corresponding to second-order or fourth-order boundary value problems. When V_h is the space of piecewise polynomials then we have

$$V_h \subset H^1(\Omega) \Longleftrightarrow V_h \subset C^0(\Omega),$$

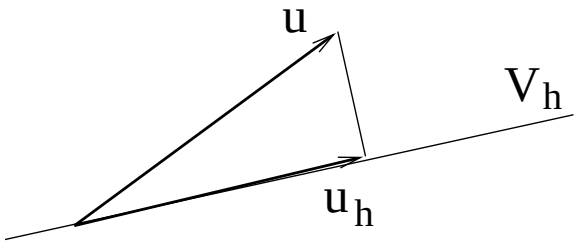


Figure 2.3: Finite element solution as a projection of the exact solution on V_h.

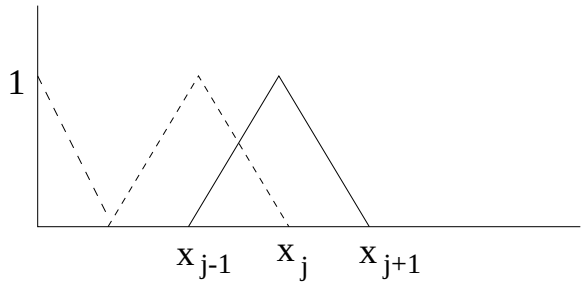


Figure 2.4: A typical shape function.

$$V_h \subset H^2(\Omega) \Longleftrightarrow V_h \subset C^1(\Omega),$$

where

$$C^0(\Omega) = \{v : v \text{ is a continuous function on } \Omega\}$$
$$C^1(\Omega) = \left\{ v \in C^0(\Omega) : \frac{\partial v}{\partial x_i} \in C^0(\Omega) \right\}.$$

The solution, u_h, that we seek is the projection of u on V_h as shown in Fig. 2.3.

To describe any function in V_h, we introduce the basis functions

$$N_j(x_i) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

A typical basis function (or shape function) is shown in Fig. 2.4. Since there are 4 nodes in our finite element model, the basis functions associated with the nodes, $\{N_i\}_{i=1}^4$, form a basis of V_h for this particular problem. Hence, any function $u_h \in V_h$ can be represented as

$$u_h = \sum_{i=1}^{4} N_i \hat{u}_i,$$

where $\hat{u}_i = u_h(x_i)$ is the value of the displacement at node x_i. The variational problem is (see Eqn. 1.31)

Find u_h which satisfies

$$\int_0^{3L} \epsilon(v)\tau A \, dx = \int_0^{3L} \bar{t}v \, dx + \int_0^{3L} \rho b v A \, dx + P \, v|_{x=L} \quad \forall v.$$

Substituting for τ and ϵ, we get the problem statement as

Find u_h which satisfies

$$\int_0^{3L} EA\frac{dv}{dx}\frac{du_h}{dx}\,dx = \int_0^{3L} \bar{t}v\,dx + \int_0^{3L} \rho bvA\,dx + P\,v|_{x=L} \quad \forall v.$$

Note that we need to satisfy the boundary conditions, $u(0) = 0$ and $v(0) = 0$. Hence, substituting $u_h = \sum_{i=2}^{4} N_i\hat{u}_i$ and $v = \sum_{i=2}^{4} N_i\hat{v}_i$, the above problem statement becomes

Find $\hat{\boldsymbol{u}}$ such that

$$\hat{\boldsymbol{v}}^t\left[\int_0^{3L} EA\frac{d\boldsymbol{N}^t}{dx}\frac{d\boldsymbol{N}}{dx}\,dx\right]\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}}^t\int_0^{3L}\boldsymbol{N}^t\bar{t}\,dx + \hat{\boldsymbol{v}}^t\int_0^{3L}\boldsymbol{N}^t(\rho bA)\,dx + P\hat{v}_2 \quad \forall\hat{\boldsymbol{v}},$$

where

$$\boldsymbol{N} = \begin{pmatrix} N_2 \\ N_3 \\ N_4 \end{pmatrix}, \quad \hat{\boldsymbol{u}} = \begin{pmatrix} \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \end{pmatrix}, \quad \hat{\boldsymbol{v}} = \begin{pmatrix} \hat{v}_2 \\ \hat{v}_3 \\ \hat{v}_4 \end{pmatrix}.$$

Using the fact that $\hat{v}_2$, $\hat{v}_3$ and $\hat{v}_4$ are arbitrary, we get the usual finite element equation

$$\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}, \tag{2.8}$$

where

$$\boldsymbol{K} = EA\begin{bmatrix} \int_0^{3L}\frac{dN_2}{dx}\frac{dN_2}{dx}\,dx & \int_0^{3L}\frac{dN_2}{dx}\frac{dN_3}{dx}\,dx & \int_0^{3L}\frac{dN_2}{dx}\frac{dN_4}{dx}\,dx \\ \int_0^{3L}\frac{dN_3}{dx}\frac{dN_2}{dx}\,dx & \int_0^{3L}\frac{dN_3}{dx}\frac{dN_3}{dx}\,dx & \int_0^{3L}\frac{dN_3}{dx}\frac{dN_4}{dx}\,dx \\ \int_0^{3L}\frac{dN_4}{dx}\frac{dN_2}{dx}\,dx & \int_0^{3L}\frac{dN_4}{dx}\frac{dN_3}{dx}\,dx & \int_0^{3L}\frac{dN_4}{dx}\frac{dN_4}{dx}\,dx \end{bmatrix},$$

$$\boldsymbol{f} = \int_0^{3L}\begin{bmatrix} N_2 \\ N_3 \\ N_4 \end{bmatrix}\bar{t}\,dx + \int_0^{3L}\begin{bmatrix} N_2 \\ N_3 \\ N_4 \end{bmatrix}\rho bA\,dx + \begin{bmatrix} P \\ 0 \\ 0 \end{bmatrix}.$$

Note that

$$\frac{dN_2}{dx} = \frac{1}{L} \quad 0 \le x \le L; \qquad \frac{dN_3}{dx} = \frac{1}{L} \quad L \le x \le 2L; \qquad \frac{dN_4}{dx} = 0 \quad 0 \le x \le 2L;$$
$$= -\frac{1}{L} \quad L \le x \le 2L; \qquad = -\frac{1}{L} \quad 2L \le x \le 3L; \qquad = \frac{1}{L} \quad 2L \le x \le 3L;$$
$$= 0 \quad \text{elsewhere}; \qquad = 0 \quad \text{elsewhere}.$$

Assuming that the tractions and body force, $\bar{t}$ and ρb, are constant, the stiffness matrix and force vector are given by

$$\begin{aligned} \boldsymbol{K} &= \frac{EA}{L}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \\ \boldsymbol{f} &= \frac{\bar{t}L}{2}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \frac{\rho bAL}{2}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} P \\ 0 \\ 0 \end{bmatrix} \end{aligned} \tag{2.9}$$

Note that the stiffness matrix is 'banded' since any shape function interacts only with its neighboring shape functions. Thus, for example N_2 and N_3 are both nonzero on $L \leq x \leq 2L$, but N_2 and N_4 are not both nonzero on any part of the domain. This leads to the $K(1,2)$ term being nonzero, but the $K(1,3)$ term being zero.

Once Eqn. 2.8 has been solved for $\hat{\boldsymbol{u}}$, the strains can be recovered from the displacement field by using

$$\begin{aligned} \epsilon = \frac{du_h}{dx} &= \begin{bmatrix} \dfrac{dN_1}{dx} & \dfrac{dN_2}{dx} & \dfrac{dN_3}{dx} & \dfrac{dN_4}{dx} \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \end{bmatrix} \\ &= \boldsymbol{B}\hat{\boldsymbol{u}}, \end{aligned}$$

where $\boldsymbol{B}$ is the strain-displacement matrix. We get

$$\begin{aligned} \epsilon &= \frac{1}{L}(\hat{u}_2 - \hat{u}_1); \quad 0 \leq x \leq L, \\ &= \frac{1}{L}(\hat{u}_3 - \hat{u}_2); \quad L \leq x \leq 2L, \\ &= \frac{1}{L}(\hat{u}_4 - \hat{u}_3); \quad 2L \leq x \leq 3L. \end{aligned}$$

Note that as a result of assuming a piecewise-linear displacement field the strain field is constant over each element. Thus, in general, it can be discontinuous across element boundaries as shown in Fig. 2.5. If we had assumed a piecewise-quadratic displacement field then we would have obtained a piecewise-linear (again, not necessarily continuous) strain field. The stress is simply obtained by using $\tau = E\epsilon$. Since the strain is just multiplied by a constant, the qualitative behavior of the stress field is the same as that of the strain field.

We would like to caution the reader though, that the procedure of directly working on the entire domain as just shown is *not* followed while implementing a finite element program. In an actual implementation, the stiffness and

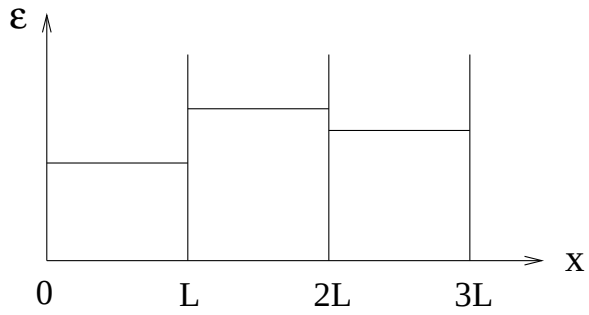


Figure 2.5: The strain field is, in general, discontinuous across element boundaries.

force matrices are formulated at an element level and then assembled to form the global stiffness matrix and force vector as shown in the following chapter. The intention behind carrying out the computations at the global level was to provide insights into issues which get obscured otherwise (e.g., the banded nature of the stiffness matrix), and also to provide a justification for the element-level formulation which is followed in practice.

Chapter 3

Implementation of the Finite Element Method

We have seen in the previous chapter that the procedure of formulating the stiffness and load vectors can be quite cumbersome. It is more convenient to form the element stiffness and load vectors, and assemble them to form the global stiffness and load vectors.

3.1 Formulation of the Stiffness Matrix and the Load Vector from Element-Level Matrices

Referring to Fig. 2.4, we see that the restriction of the shape function to an element is as shown in Fig. 3.1. A shape function associated with a particular node has a value of one at that node and a value of zero at other nodes in that element. Though it is trivial to satisfy continuity of the independent field across an element boundary in a one-dimensional problem, it is quite difficult to achieve this in two or three-dimensional problems when the elements are not rectangular or parallelepipeds (try it!). To achieve this objective, it is helpful to introduce a natural or intrinsic coordinate system. We shall first carry out this exercise for one-dimensional elements, and later extend it to higher dimensions.

As shown in Fig. 3.2, the element with endpoint coordinates given by

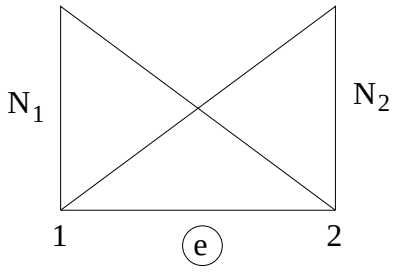


Figure 3.1: Restriction of a shape function to an element.

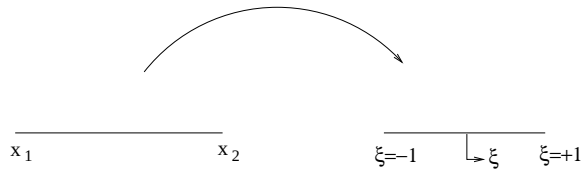


Figure 3.2: Mapping of an element to a master element.

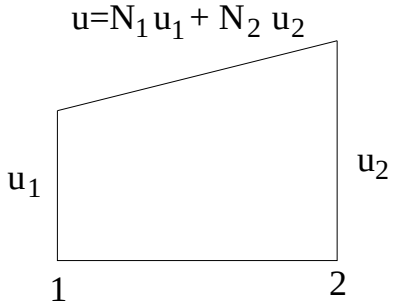


Figure 3.3: Displacement interpolation

(x_1, x_2) is mapped to a natural coordinate system with endpoint coordinates ξ given by $(-1, 1)$. The shape functions on the element shown in Fig. 3.1 are now formulated in terms of ξ. Let us consider the formulation of N_1. Since, N_1 has to satisfy two boundary conditions, namely $N_1(-1) = 1$ and $N_1(1) = 0$ at the end nodes, we assume it to be of the form

$$N_1(\xi) = a + b\xi,$$

where a and b are constants which are determined from the boundary conditions. We get

$$N_1(\xi) = \frac{1}{2}(1 - \xi).$$

Following a similar procedure, we get

$$N_2(\xi) = \frac{1}{2}(1 + \xi).$$

The coordinate x of a point in the element is given by

$$x = N_1(\xi)x_1 + N_2(\xi)x_2,$$

where ξ is the natural coordinate of that point. If we interpolate the displacements also in the same way, we get

$$u = N_1(\xi)u_1 + N_2(\xi)u_2.$$

If the same shape functions are used to interpolate both the displacements and the geometry as above, the mapping is known as an isoparametric mapping. Note that $u(\xi = -1) = u_1$, $u(\xi = +1) = u_2$ with a linear interpolation in between as shown in Fig. 3.3.

The strain is given by

$$\begin{aligned}\epsilon = \frac{du}{dx} &= \frac{du}{d\xi}\frac{d\xi}{dx} \\ &= \frac{1}{x_2 - x_1}(u_2 - u_1) \\ &= \frac{1}{l_e}(u_2 - u_1),\end{aligned}$$

where $l_e = (x_2 - x_1)$ is the length of the element. The above equation can be written as

$$\epsilon = \boldsymbol{B}\hat{\boldsymbol{u}},$$

where

$$\boldsymbol{B} = \frac{1}{l_e}[-1 \quad 1]$$

$$\hat{\boldsymbol{u}} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

The stress is given by

$$\sigma = E\epsilon = E\boldsymbol{B}\hat{\boldsymbol{u}}.$$

Note that the stresses and strains are constant within an element.

The element stiffness matrix and load vector can be derived using either the potential energy functional or the virtual work principle. We shall follow the latter approach. We know from Eqn. 2.3 that the virtual work principle involves finding $\boldsymbol{u} \in L_u$ such that

$$\int_\Omega [\boldsymbol{\epsilon}_c(\boldsymbol{v})]^t \boldsymbol{\tau}_c \, d\Omega = \int_{\Gamma_t} \boldsymbol{v}^t \bar{\boldsymbol{t}} \, d\Gamma + \int_\Omega \rho \boldsymbol{v}^t \boldsymbol{b} \, d\Omega + \sum P_i v_i \quad \forall \boldsymbol{v} \in V_u.$$

where, now, we have added the contribution of the point loads P_i explicitly instead of considering it as part of the prescribed tractions $\bar{\boldsymbol{t}}$. Writing the above equation as a summation over the number of elements, we get

$$\sum_e \int_{\Omega_e} [\boldsymbol{\epsilon}_c(\boldsymbol{v})]^t \boldsymbol{\tau}_c \, d\Omega = \sum_e \left[\int_{\Omega_e} \rho \boldsymbol{v} \cdot \boldsymbol{b} \, d\Omega + \int_{\Gamma_e} \boldsymbol{v} \cdot \bar{\boldsymbol{t}} \, d\Gamma \right] + \sum P_i v_i \quad \forall \boldsymbol{v} \in V_u,$$

Note that while computing the surface integral term in the above equation, we need to consider only the those elements which have a surface on Γ_t. The reason is that the net contribution of the internal tractions at any shared edge between two elements to the global load vector is zero since the traction vector on the edge of one element is equal and opposite to the traction vector on the adjacent element.

For the one-dimensional problem that we are considering, we have $v = \boldsymbol{N}\hat{\boldsymbol{v}}$, $\epsilon(v) = \boldsymbol{B}\hat{\boldsymbol{v}}$, $\tau = E\boldsymbol{B}\hat{\boldsymbol{u}}$, $d\Omega = Adx = Al_e \, d\xi/2$ and $d\Gamma = dx = l_e d\xi/2$. Substituting these various quantities in the above equation, we get

$$\sum_e \int_{-1}^{1} \boldsymbol{v}^t \boldsymbol{B}^t EA\boldsymbol{B} \frac{l_e}{2}\, d\xi \hat{\boldsymbol{u}} = \sum_e \left[\int_{-1}^{1} \boldsymbol{v}^t \boldsymbol{N}^t \bar{t} \frac{l_e}{2}\, d\xi + \int_{-1}^{1} \hat{\boldsymbol{v}}^t \boldsymbol{N}^t \rho b A \frac{l_e}{2}\, d\xi \right] + \hat{\boldsymbol{v}}^t \{N_i(\xi_j) P_j\},$$

where in the last term ξ_j is the position at which load P_j is applied.

From the above equation, the element stiffness matrix is given by

$$\boldsymbol{K}^{(e)} = \frac{1}{2} \int_{-1}^{1} \boldsymbol{B}^t EA\boldsymbol{B} l_e\, d\xi.$$

Assuming that EA is constant over the element, we get

$$\boldsymbol{K}^{(e)} = \frac{E_e A_e}{l_e} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The element load vectors due to the traction, body forces and point loads are

$$\boldsymbol{f}_t^{(e)} = \frac{1}{2} \int_{-1}^{1} \boldsymbol{N}^t \bar{t} l_e\, d\xi = \begin{bmatrix} \frac{l_e}{2} \int_{-1}^{1} N_1 \bar{t}\, d\xi \\ \frac{l_e}{2} \int_{-1}^{1} N_2 \bar{t}\, d\xi \end{bmatrix},$$

$$\boldsymbol{f}_b^{(e)} = \frac{1}{2} \int_{-1}^{1} \boldsymbol{N}^t A \rho b l_e\, d\xi = \begin{bmatrix} \frac{l_e}{2} \int_{-1}^{1} N_1 \rho b A\, d\xi \\ \frac{l_e}{2} \int_{-1}^{1} N_2 \rho b A\, d\xi \end{bmatrix},$$

$$\boldsymbol{f}_P^{(e)} = \begin{bmatrix} \sum_i N_1(\xi_i) P_i \\ \sum_i N_2(\xi_i) P_i \end{bmatrix}.$$

If $\bar{t}$ and ρb are constant over the element then we get

$$\boldsymbol{f}_t^{(e)} = \frac{\bar{t} l_e}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad \boldsymbol{f}_b^{(e)} = \frac{A \rho b l_e}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

On the other hand if the body force varies as $b = b_0 x = b_0(N_1 x_1 + N_2 x_2)$, with ρ constant, then

$$\begin{aligned} \boldsymbol{f}_b^{(e)} &= \frac{\rho b_0 l_e A}{2} \begin{bmatrix} \int_{-1}^{1} N_1 N_1\, d\xi & \int_{-1}^{1} N_1 N_2\, d\xi \\ \int_{-1}^{1} N_2 N_1\, d\xi & \int_{-1}^{1} N_2 N_2\, d\xi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{\rho b_0 l_e A}{2} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{\rho b_0 l_e A}{6} \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix}. \end{aligned}$$

3.2 Assembly of the Element Stiffness and Load Vectors into the Global Stiffness and Load Vectors

Consider the contribution of the second element to the term $\hat{\boldsymbol{v}}^t \boldsymbol{K} \hat{\boldsymbol{u}}$ in Eqn. 2.4. We have

$$
\begin{aligned}
(\hat{\boldsymbol{v}}^t \boldsymbol{K} \hat{\boldsymbol{u}})_2 &= \begin{bmatrix} v_2 & v_3 \end{bmatrix} \frac{E_2 A_2}{l_2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \end{bmatrix} \\
&= \begin{bmatrix} \hat{v}_1 & \hat{v}_2 & \hat{v}_3 & \hat{v}_4 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{E_2 A_2}{l_2} & -\frac{E_2 A_2}{l_2} & 0 \\ 0 & -\frac{E_2 A_2}{l_2} & \frac{E_2 A_2}{l_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \end{bmatrix}.
\end{aligned}
$$

Thus, we see that the elements of $\boldsymbol{K}^{(2)}$ occupy the second and third rows and columns of the global stiffness matrix $\boldsymbol{K}$. Since two and three are the global degrees of freedom associated with element 2, we conclude that the element stiffness matrix should be assembled based on the degrees of freedom associated with the element. Overlapping elements are added. Similar to the process for the stiffness matrix, we assemble the global load vector using the individual element load vectors. For the model 1-d problem considered in Section 2.5, we get

$$
\begin{aligned}
\boldsymbol{K} &= \frac{EA}{L} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \\
\boldsymbol{f} &= \frac{\bar{t} L}{2} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + \frac{\rho b A L}{2} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} R \\ P \\ 0 \\ 0 \end{bmatrix},
\end{aligned}
\tag{3.1}
$$

where R is the reaction at the wall. On incorporating the boundary conditions into the above matrices as shown in the next section, the above matrices will reduce to those in Eqn. 2.9. Notice that this procedure of assembling element-level matrices is considerably simpler compared to the 'direct' method that we followed in Section 2.5.

The stiffness matrix has the following properties:

1. The dimension of the stiffness matrix is $n \times n$ where n is the number of free degrees of freedom.

2. $\boldsymbol{K}$ is symmetric.

3. $\boldsymbol{K}$ is banded since basis functions have compact local support.

4. Using the symmetry and bandedness properties, $\boldsymbol{K}$ is stored in compact form. For example, one of the ways that the stiffness matrix in Eqn. 3.1 is stored is

$$K_{\text{banded}} = \frac{EA}{l_e} \begin{bmatrix} 1 & -1 \\ 2 & -1 \\ 2 & -1 \\ 1 & 0 \end{bmatrix}$$

Thus, the storage space is reduced from $n \times n$ to $n \times n_{bw}$, where n_{bw} is the half-bandwidth of $\boldsymbol{K}$, and is given by

$$n_{bw} = N_{\max} * n_{\text{ndf}} + n_{\text{ndf}}, \tag{3.2}$$

where $N_{\max}$ is the maximum difference between node numbers connecting an element and n_{ndf} is the number of degrees of freedom per node. In our model 1-d problem, $n_{\text{ndf}} = 1$, and the maximum difference between node numbers connecting an element is 1. Hence $n_{bw} = 2$. Arbitrary numbering of nodes can result in higher bandwidth and hence increased storage. For example, if in our model problem, we had numbered the nodes as $(1, 4, 3, 2)$, then

$$n_{bw} = \max(4 - 1, 4 - 3, 3 - 2) + 1 = 4,$$

whereas we have seen that with sequential node-numbering we have seen that $n_{bw} = 2$.

3.3 Treatment of Boundary Conditions

In the previous section, we have seen how to assemble the element level matrices to form the global stiffness matrix. However, the stiffness matrix thus obtained is singular since rigid body modes have not been eliminated. We need to incorporate the boundary conditions which prevent rigid body motion. We discuss two methods of incorporating the boundary conditions, the elimination approach and the penalty approach. The elimination approach is more exact, but is more difficult to implement than the penalty approach.

3.3.1 The elimination approach

Suppose that we have obtained the matrix equations $\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}$, and suppose that it is given that $\hat{u}_1 = a_1$, where a_1 is a specified value. Also suppose that

the external force applied and the reaction corresponding to this degree of freedom are F_1 and R_1, respectively. Then we have

$$\begin{bmatrix} K_{11} & K_{12} & \dots & K_{1n} \\ K_{21} & K_{22} & \dots & K_{2n} \\ \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & K_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ \hat{u}_2 \\ . \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} F_1 + R_1 \\ F_2 \\ . \\ F_n \end{bmatrix}.$$

Writing out the first equation:

$$R_1 = K_{11}a_1 + K_{12}\hat{u}_2 + \cdots + K_{1n}\hat{u}_n - F_1 \tag{3.3}$$

Writing out the second equation, we get

$$K_{21}a_1 + K_{22}\hat{u}_2 + \cdots + K_{2n}\hat{u}_n = F_2,$$

or, alternatively,

$$K_{22}\hat{u}_2 + K_{23}\hat{u}_3 + \cdots + K_{2n}\hat{u}_n = F_2 - K_{21}a_1,$$

Similarly we can write the third equation as

$$K_{32}\hat{u}_2 + K_{33}\hat{u}_3 + \cdots + K_{3n}\hat{u}_n = F_3 - K_{31}a_1,$$

and so on. In matrix form, all the equations excluding the first can be written as

$$\begin{bmatrix} K_{22} & K_{23} & \dots & K_{2n} \\ K_{32} & K_{33} & \dots & K_{3n} \\ \dots & \dots & \dots & \dots \\ K_{n2} & K_{n3} & \dots & K_{nn} \end{bmatrix}_{(n-1)\times(n-1)} \begin{bmatrix} \hat{u}_2 \\ \hat{u}_3 \\ \dots \\ \hat{u}_n \end{bmatrix}_{(n-1)\times 1} = \begin{bmatrix} F_2 - K_{21}a_1 \\ F_3 - K_{31}a_1 \\ \dots \\ F_n - K_{n1}a_1 \end{bmatrix}_{(n-1)\times 1}.$$

Thus, the reduced $\boldsymbol{K}$ matrix is obtained by eliminating the row and column corresponding to the specified or 'support' degree of freedom. Note that the force matrix also gets modified when the prescribed quantity is non-zero. Since the rigid-body modes have been eliminated, the reduced $\boldsymbol{K}$ matrix is no longer singular. Hence we can solve for the unknown vector $\hat{\boldsymbol{u}}$.

In order to compute the strains the element-level quantities are extracted from the global $\hat{\boldsymbol{u}}$ vector. We then use $\boldsymbol{\epsilon} = \boldsymbol{B}\hat{\boldsymbol{u}}^{(e)}$. The stress is then obtained using the appropriate constitutive relation, e.g., in our model problem $\tau = E\boldsymbol{B}\hat{\boldsymbol{u}}^{(e)}$. In order to compute the reactions, we use Eqn. 3.3 since all the quantities on the right-hand side of that equation are known. Thus, if we need to compute the reactions, then we need to store the rows of $\boldsymbol{K}$ which are eliminated for the purposes of getting the reduced stiffness matrix.

The above procedure can be easily generalized for the case when there are r prescribed degrees of freedom, say $\hat{u}_{p_1} = a_1$, $\hat{u}_{p_2} = a_2$, ..., $\hat{u}_{p_r} = a_r$:

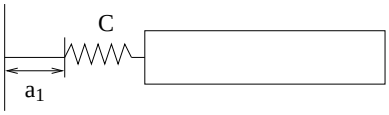


Figure 3.4: Penalty approach to handle boundary conditions

- Store the p_1, p_2, ..., p_r row of the global stiffness matrix and the load vector $\boldsymbol{f}$.
- Delete the p_1, p_2, ..., p_r'th row and column from the $\boldsymbol{K}$ matrix. The resulting stiffness matrix size is $(n-r)\times(n-r)$. Delete the p_1, p_2, ..., p_r'th row from the load vector. Modify the load components as

$$f_i \to f_i - \sum_j K_{i,p_j} a_j.$$

- Solve $\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}$.
- Extract the element displacement fields $\hat{\boldsymbol{u}}^{(e)}$ from $\hat{\boldsymbol{u}}$, and determine the strains and stresses.
- Evaluate the reactions at each support degree of freedom:

$$R_{p_1} = K_{p_1 1}\hat{u}_1 + K_{p_1 2}\hat{u}_2 + \ldots + K_{p_1 n}\hat{u}_n - f_{p_1} = \sum_i K_{p_1 i}\hat{u}_i - f_{p_1},$$
$$R_{p_2} = K_{p_2 1}\hat{u}_1 + K_{p_2 2}\hat{u}_2 + \ldots + K_{p_2 n}\hat{u}_n - f_{p_2} = \sum_i K_{p_2 i}\hat{u}_i - f_{p_2},$$

  and so on. In general

$$R_{p_j} = \sum_i K_{p_j i}\hat{u}_i - f_{p_j}.$$

  One good debugging check is to see that the (vectorial) sum of the reactions is equal to the load applied on the structure.

In the model problem considered in Section 2.5, the reaction at the wall is given by

$$R = -\frac{1}{L}EA\hat{u}_2 - \frac{1}{2}AL\rho b - \frac{1}{2}L\bar{t}.$$

3.3.2 The penalty approach

Again consider a boundary condition of the type $\hat{u}_1 = a_1$. Now we introduce a spring with large stiffness, C, to model the support as shown in Fig. 3.4. One end of the spring is displaced by a_1, while the other end is displaced by $\hat{u}_1$. Hence, the net extension of the spring is $\hat{u}_1 - a_1$. Note that if the spring is infinitely rigid, then the net extension of the spring would be zero, and we would get $\hat{u}_1 = a_1$. The strain energy in the spring is

$$U_s = \frac{1}{2}C(\hat{u}_1 - a_1)^2.$$

Hence, the expression for the potential energy gets modified to

$$\Pi = \frac{1}{2}\hat{\boldsymbol{u}}^t \boldsymbol{K} \hat{\boldsymbol{u}} + \frac{1}{2} C(\hat{u}_1 - a_1)^2 - \hat{\boldsymbol{u}}^t \boldsymbol{f}.$$

To derive the matrix equations, we follow the usual procedure and set the partial derivatives of Π with respect to the degrees of freedom to zero, i.e.,

$$\frac{\partial \Pi}{\partial \hat{u}_1} = 0; \ \frac{\partial \Pi}{\partial \hat{u}_2} = 0; \ \ldots; \frac{\partial \Pi}{\partial \hat{u}_n} = 0.$$

We get

$$\begin{bmatrix} K_{11}+C & K_{12} & \ldots & K_{1n} \\ K_{21} & K_{22} & \ldots & K_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ K_{n1} & K_{n2} & \ldots & K_{nn} \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \ldots \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} f_1 + Ca_1 \\ f_2 \\ \ldots \\ f_n \end{bmatrix}. \tag{3.4}$$

Thus, a large number gets added to the first diagonal term of $\boldsymbol{K}$, and Ca_1 gets added to f_1. The reaction force is the force exerted by the spring on the structure. Since the net extension of the spring is $\hat{u}_1 - a_1$, the reaction force is

$$R_1 = -C(\hat{u}_1 - a_1).$$

Generalizing the above discussion to the case when we have r prescribed boundary conditions of the form $\hat{u}_{p_1} = a_1$, $\hat{u}_{p_2} = a_2$, ..., $\hat{u}_{p_r} = a_r$, the procedure of treating the boundary conditions is now as follows:

- Modify the stiffness matrix $\boldsymbol{K}$ by adding a large number C to each of the p_1, p_2, ..., p_r'th diagonal elements of $\boldsymbol{K}$. Also modify the global load vector $\boldsymbol{f}$ by adding Ca_1 to f_{p_1}, Ca_2 to f_{p_2}, ..., Ca_r to f_{p_r}.
- Solve $\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}$.
- Extract the element displacement field $\hat{\boldsymbol{u}}^{(e)}$ from the global displacement vector $\hat{\boldsymbol{u}}$ using the element connectivity, and determine the strains and stresses.
- Evaluate the reaction forces at each support using the equation

$$R_{p_i} = -C(\hat{u}_{p_i} - a_i) \quad i = 1, 2, \ldots, r.$$

As already mentioned, the penalty approach is an approximate approach, and the accuracy of the solution depends on the choice of C. We now give a rough guideline for choosing C. The first equation in Eqn. 3.4 is

$$(K_{11} + C)\hat{u}_1 + K_{12}\hat{u}_2 + \ldots + K_{1n}\hat{u}_n = F_1 + Ca_1,$$

which in turn implies that

$$\left(\frac{K_{11}}{C} + 1\right)\hat{u}_1 + \frac{K_{12}}{C}\hat{u}_2 + \ldots + \frac{K_{1n}}{C}\hat{u}_n = \frac{F_1}{C} + a_1.$$

Thus, if C is large compared to K_{11}, K_{12},..., K_{1n}, then $\hat{u}_1 \approx a_1$. Hence, choose C as

$$C = \max |K_{ij}| \times 10^4 \quad 1 \leq i, j \leq n.$$

3.3.3 Multi-point constraints

Multi-point constraints are constraints involving several displacement degrees of freedom, as, for example, occurs when a node is constrained to move along an inclined surface. We consider only linear multi-point constraints in what follows. Let the constraints be of the form

$$E\boldsymbol{C}\hat{\boldsymbol{u}} = E\boldsymbol{a}, \tag{3.5}$$

where $\boldsymbol{C}$ is an $m \times n$ matrix corresponding to m constraint equations, $\hat{\boldsymbol{u}}$ is the $n \times 1$ vector containing the displacement degrees of freedom, and the factor E (Young modulus) has been introduced to alleviate subsequent ill-conditioning that can occur. We use a Lagrange muliplier technique to impose these constraints, and modify the potential energy expression in Eqn. (2.7) to read

$$\Pi = \frac{1}{2}\hat{\boldsymbol{u}} \cdot \boldsymbol{K}\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{f}} - E\boldsymbol{\lambda} \cdot [\boldsymbol{C}\hat{\boldsymbol{u}} - \boldsymbol{a}]\,,$$

where $\boldsymbol{\lambda}$ is a $m \times 1$ vector of the Lagrange multipliers corresponding to the m constraints in Eqn. (3.5). The corresponding variational formulation gets modified to
Find $(\hat{\boldsymbol{u}}, \boldsymbol{\lambda})$ such that

$$\hat{\boldsymbol{v}} \cdot \Big[\boldsymbol{K}\hat{\boldsymbol{u}} - \hat{\boldsymbol{f}} - E\boldsymbol{C}^T\boldsymbol{\lambda}] - E\boldsymbol{\lambda}_\delta \cdot [\boldsymbol{C}\hat{\boldsymbol{u}} - \boldsymbol{a}\Big] = 0 \quad \forall(\hat{\boldsymbol{v}}, \boldsymbol{\lambda}_\delta),$$

First setting $\boldsymbol{\lambda}_\delta$ to zero, and using the arbitrariness of $\hat{\boldsymbol{v}}$ yields

$$\boldsymbol{K}\hat{\boldsymbol{u}} - E\boldsymbol{C}^T\boldsymbol{\lambda} = \hat{\boldsymbol{f}}. \tag{3.6}$$

Next using the arbitrariness of $\boldsymbol{\lambda}_\delta$, we recover the constraint equations given by Eqn. (3.5). The system of equations given by Eqns. (3.5) and (3.6) can be written as

$$\begin{bmatrix} \boldsymbol{K} & -E\boldsymbol{C}^T \\ E\boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{u}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{f}} \\ E\boldsymbol{a} \end{bmatrix}. \tag{3.7}$$

The above set of equations is to be solved for $(\hat{\boldsymbol{u}}, \boldsymbol{\lambda})$. The Lagrange multipliers can be given the physical interpretation of reaction forces.

3.3.4 Example

To illustrate the elimination and penalty approaches, consider the example shown in Fig. 3.5. The Young's modulus of the two bars are given by $E_1 = 70 \times 10^9$ N/m^2 and $E_2 = 200 \times 10^9$ N/m^2. The cross-sectional areas are $A_1 = 2400$ mm^2 and $A_2 = 600$ mm^2. The applied load is $P = 200 \times 10^3$ N.

The element stiffness matrices are (with units N/mm)

$$K^{(1)} = \frac{70 \times 10^3 \times 2400}{300} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

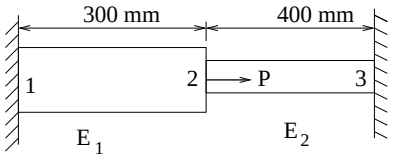


Figure 3.5: Analysis of two coaxial bars

$$K^{(2)} = \frac{200 \times 10^3 \times 600}{400} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The global stiffness matrix and load vector are

$$\boldsymbol{K} = 10^6 \begin{bmatrix} 0.56 & -0.56 & 0 \\ -0.56 & 0.86 & -0.3 \\ 0 & -0.3 & 0.3 \end{bmatrix}; \quad \boldsymbol{f} = \begin{bmatrix} R_1 \\ 200 \times 10^3 \\ R_2 \end{bmatrix}.$$

Using the elimination approach, we the 1'st and 3'rd rows and columns from $\boldsymbol{K}$. We get

$$10^6(0.86)\hat{u}_2 = 200 \times 10^3,$$

which yields $\hat{u}_2 = 0.23255$ mm. The element stresses are

$$\tau_1 = E_1 \boldsymbol{B}_1 \hat{\boldsymbol{u}}^{(1)} = \frac{E_1}{(x_2 - x_1)} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = 54.77 \text{ MPa},$$

$$\tau_2 = E_2 \boldsymbol{B}_2 \hat{\boldsymbol{u}}^{(2)} = \frac{E_1}{(x_2 - x_1)} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = -116.29 \text{ MPa}.$$

Note the discontinuity in the stresses. The reactions are

$$R_1 = 10^6[0.56(0) - 0.56(0.23255) + 0] = -130.23 \times 10^3 \text{ N},$$
$$R_2 = 10^6[0 - 0.3(0.23255) + 0] = -69.77 \times 10^3 \text{ N}.$$

Note that the sum of the reactions is equal to the applied load.

The results for $(\hat{u}_1, \hat{u}_2, \hat{u}_3, R_1, R_2)$ could directly have been obtained using Eqns. (3.7), with (R_1, R_2) playing the role of the Lagrange multipliers.

Now let us carry out the same computations using the penalty approach. Using the suggested guidelines, we have $C = 0.8 \times 10^{10}$. The stiffness matrix is modified by adding C to the appropriate diagonal terms. The matrix equations are given by

$$10^6 \begin{bmatrix} 8600.56 & -0.56 & 0 \\ -0.56 & 0.86 & -0.3 \\ 0 & -0.3 & 8600.3 \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 200 \times 10^3 \\ 0 \end{bmatrix}.$$

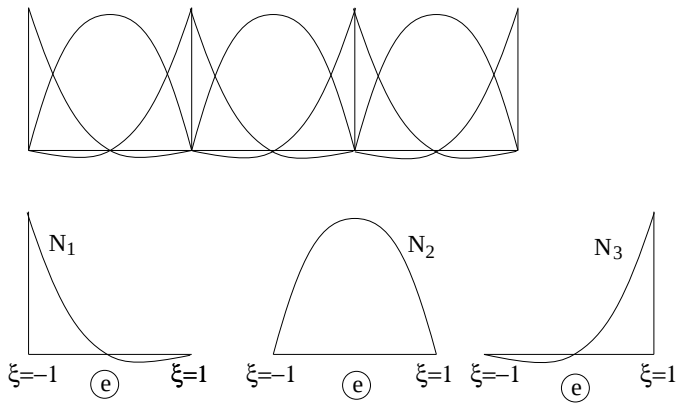


Figure 3.6: Quadratic shape functions: Global and element level pictures

On solving these equations, we get

$$\begin{bmatrix}\hat{u}_1 & \hat{u}_2 & \hat{u}_3\end{bmatrix} = \begin{bmatrix}15.14\times10^{-6} & 0.23257 & 8.1127\times10^{-6}\end{bmatrix} \text{ mm.}$$

Notice that $\hat{u}_1, \hat{u}_3 \approx 0$. The stresses are

$$\tau_1 = E_1\boldsymbol{B}_1\hat{\boldsymbol{u}}^{(1)} = \frac{70\times10^3}{300}\begin{bmatrix}-1 & 1\end{bmatrix}\begin{bmatrix}15.14\times10^{-6}\\ 0.23257\end{bmatrix} = 54.27 \text{ MPa},$$

$$\tau_2 = E_2\boldsymbol{B}_2\hat{\boldsymbol{u}}^{(2)} = \frac{200\times10^3}{400}\begin{bmatrix}-1 & 1\end{bmatrix}\begin{bmatrix}0.23257\\ 8.1127\times10^{-6}\end{bmatrix} = -116.27 \text{ MPa}.$$

The reactions are

$$R_1 = -C\hat{u}_1 = -[0.86\times10^10]\times(15.143\times10^{-6}) = -130.23\times10^3 \text{ N}$$
$$R_2 = -C\hat{u}_3 = -[0.86\times10^10]\times(8.1127\times10^{-6}) = -69.77\times10^3 \text{ N}.$$

3.4 Higher-order shape functions

The accuracy of a finite element analysis can be increased by reducing the typical length of an element 'h' or by increasing the order of the shape functions 'p' (or a combination of the two). As an example, consider the quadratic shape functions

$$\begin{aligned} N_1(\xi) &= -\frac{1}{2}\xi(1-\xi),\\ N_2(\xi) &= (1+\xi)(1-\xi),\\ N_3(\xi) &= \frac{1}{2}\xi(1+\xi). \end{aligned} \tag{3.8}$$

The 'global' and 'element' level pictures of the shape function are shown in Fig. 3.6. In order to obtain the expressions given in Eqn. 3.8, consider the

expression for N_1. Since N_1 is zero at $\xi = 0$ and $\xi = 1$, we assume $N_1 = c\xi(1-\xi)$. The constant c is determined by using the condition $N_1|_{\xi=-1} = 1$. A similar procedure is followed for determining N_2 and N_3.

Note that the shape functions formulated above give C^0 continuity for the function being interpolated, but not, in general C^1 continuity. Such shape functions are called *Lagrange* shape functions[1]. In general, a function ϕ of degree $(n-1)$ and defined by n values ϕ_i at corresponding ξ_i has the form

$$\phi = N_1\phi_1 + N_2\phi_2 + \ldots + N_n\phi_n,$$

where

$$\begin{aligned}
N_1 &= \frac{(\xi_2-\xi)(\xi_3-\xi)(\xi_4-\xi)\ldots(\xi_n-\xi)}{(\xi_2-\xi_1)(\xi_3-\xi_1)(\xi_4-\xi_1)\ldots(\xi_n-\xi_1)},\\
N_2 &= \frac{(\xi_1-\xi)(\xi_3-\xi)(\xi_4-\xi)\ldots(\xi_n-\xi)}{(\xi_1-\xi_2)(\xi_3-\xi_2)(\xi_4-\xi_2)\ldots(\xi_n-\xi_2)},\\
&\ldots\\
N_n &= \frac{(\xi_1-\xi)(\xi_2-\xi)(\xi_3-\xi)\ldots(\xi_{n-1}-\xi)}{(\xi_1-\xi_n)(\xi_2-\xi_n)(\xi_3-\xi_n)\ldots(\xi_{n-1}-\xi_n)}.
\end{aligned}$$

For $n = 3$, we can easily verify that we get the expressions in Eqn. 3.8.

For an isoparametric element, we have

$$\begin{aligned}
u &= N_1u_1 + N_2u_2 + N_3u_3,\\
x &= N_1x_1 + N_2x_2 + N_3x_3,
\end{aligned}$$

or, in matrix form

$$\begin{aligned}
u &= \boldsymbol{N}\hat{\boldsymbol{u}}\\
x &= \boldsymbol{N}\hat{\boldsymbol{x}}.
\end{aligned}$$

Thus, both u and x are piecewise quadratic polynomials. In what follows, we shall assume that the midnode is located at the center of the element, i.e., $2x_2 = x_1 + x_3$. Thus, we have

$$\begin{aligned}
\frac{dx}{d\xi} &= \frac{1}{2}(x_3 - x_1) + \xi(x_1 + x_3 - 2x_2)\\
&= \frac{l_e}{2},
\end{aligned}$$

where l_e is the length of the element. The strain is

$$\begin{aligned}
\epsilon &= \frac{du}{dx}\\
&= \frac{du}{d\xi}\frac{d\xi}{dx}
\end{aligned}$$

[1]For providing C^1 continuity, we use *Hermitian* shape functions which we shall discuss later.

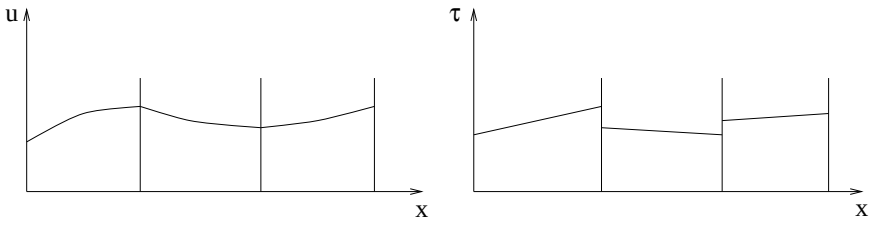


Figure 3.7: Variation of the displacement and stress field

$$= \frac{2}{(x_3 - x_1)} \begin{bmatrix} \frac{dN_1}{d\xi} & \frac{dN_2}{d\xi} & \frac{dN_3}{d\xi} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$= \boldsymbol{B}\hat{\boldsymbol{u}}^{(e)},$$

where

$$\boldsymbol{B} = \frac{2}{l_e} \begin{bmatrix} -\frac{1-2\xi}{2} & -2\xi & \frac{1+2\xi}{2} \end{bmatrix}.$$

The stress is given by

$$\tau = E\boldsymbol{B}\hat{\boldsymbol{u}}^{(e)}.$$

The qualitative variation of the displacement and stress is shown in Fig. 3.7. Note that a quadratic variation of the displacement implies that the stresses and strains can vary linearly within the element though they can still be discontinuous across element boundaries as shown.

Assuming that E and A are constant, the element stiffness matrix is given by

$$\begin{aligned} \boldsymbol{K}^{(e)} &= \int_{\Omega_e} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B}\, d\Omega \\ &= \int_{x_1}^{x_2} \boldsymbol{B}^t E \boldsymbol{B} A\, dx \\ &= \frac{EAl_e}{2} \int_{-1}^{1} \boldsymbol{B}^t \boldsymbol{B}\, d\xi \\ &= \frac{EA}{3l_e} \begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix}. \end{aligned} \tag{3.9}$$

Note that the size of the element stiffness matrix is $n_{\text{dof}} \times n_{\text{dof}}$, where n_{dof} is the number of element degrees of freedom.

The element force vector due to a constant body force b is

$$\begin{aligned} \boldsymbol{f}^{(e)} &= \int_{x_1}^{x_2} \boldsymbol{N}^t \rho b A\, dx \\ &= \int_{-1}^{1} \boldsymbol{N}^t \rho b A \frac{l_e}{2}\, d\xi \end{aligned}$$

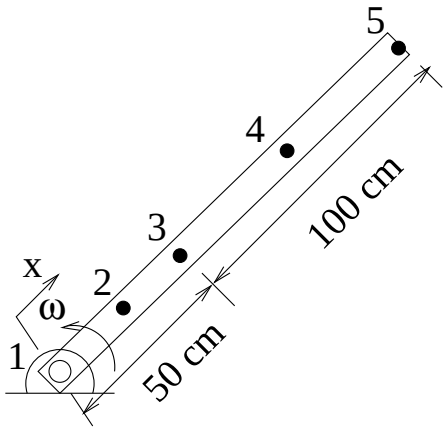


Figure 3.8: Analysis of a rotating bar

$$= \frac{\rho b A l_e}{2} \int_{-1}^{1} \boldsymbol{N}^t \, d\xi$$

$$= \frac{\rho b A l_e}{6} \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} .$$

Similarly, the element-force vector due to a constant force per unit length $\bar{t}$ is

$$\boldsymbol{f}_t^{(e)} = \frac{l_e \bar{t}}{6} \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} .$$

3.5 Examples

Example 1:
A rod of length 150 cm rotates about a point with a constant angular velocity as shown in Fig. 3.8. The density, Young's modulus, cross sectional area, and angular velocity are given by $\rho = 7500$ kg/m^3, $E = 60$ GPa, $A = 4 \times 10^{-4}$ m^2, and $\omega = 20^c$/s. Assume $\boldsymbol{u} = u(x)\boldsymbol{e}_x$, and the loading to be only due to the centrifugal body force (given by $\rho b = \rho x \omega^2$). Formulate the element stiffness matrix, and the consistent load vector for a *quadratic* (3-node) element. Then using the two-element (5-node) model shown in the figure, determine the displacements and the stress distribution in the rod using two quadratic elements. Plot the stress distribution as a function of x. What do you observe about the stress distribution at node 3?

Solution:

The body force is given by

$$\rho b = \rho x \omega^2 = 7500(400) \begin{bmatrix} N_1 & N_2 & N_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Hence, the element force vector is

$$\begin{aligned}
\boldsymbol{f}_b^{(e)} &= \int \boldsymbol{N}^t \rho b A \, dx \\
&= 1.2 \times 10^3 \int_{-1}^{1} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix} \begin{bmatrix} N_1 & N_2 & N_3 \end{bmatrix} \frac{l_e}{2} \, d\xi \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= 0.6 \times 10^3 l_e \int_{-1}^{1} \begin{bmatrix} N_1^2 & N_1 N_2 & N_1 N_3 \\ N_1 N_2 & N_2^2 & N_2 N_3 \\ N_1 N_3 & N_2 N_3 & N_3^2 \end{bmatrix} d\xi \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= 40 l_e \begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= 40 l_e \begin{bmatrix} 4x_1 + 2x_2 - x_3 \\ 2x_1 + 16x_2 + 2x_3 \\ -x_1 + 2x_2 + 4x_3 \end{bmatrix}
\end{aligned}$$

For element 1, we have $x_1 = 0$, $x_2 = 0.25$, $x_3 = 0.5$, $l_e = 0.5$, while for element 2, we have $x_1 = 0.5$, $x_2 = 1.0$, $x_3 = 1.4$, $l_e = 1.0$, which yields

$$f_b^{(1)} = \begin{bmatrix} 0 \\ 100 \\ 50 \end{bmatrix}; \; f_b^{(2)} = \begin{bmatrix} 100 \\ 800 \\ 300 \end{bmatrix}.$$

The global force vector obtained after assembling the two load vectors is

$$\boldsymbol{f} = \begin{bmatrix} 0 \\ 100 \\ 150 \\ 800 \\ 300 \end{bmatrix}.$$

As a check, we verify that the total load obtained by summing the entries in the load vector (1350 N) is the same as obtained from the continuum problem, i.e.,

$$\int_0^{1.5} 1.2 \times 10^3 x \, dx = 1350 \text{ N}.$$

The element stiffness matrix is given by Eqn. 3.9. For the two elements, we have

$$\boldsymbol{K}^{(1)} = 0.8 \times 10^7 \begin{bmatrix} 14 & -16 & 2 \\ -16 & 32 & -16 \\ 2 & -16 & 14 \end{bmatrix}; \; \boldsymbol{K}^{(2)} = 0.8 \times 10^7 \begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix}.$$

Assembling the two stiffness matrices, we get

$$\boldsymbol{K} = 0.8 \times 10^7 \begin{bmatrix} 14 & -16 & 2 & 0 & 0 \\ -16 & 32 & -16 & 0 & 0 \\ 2 & -16 & 21 & -8 & 1 \\ 0 & 0 & -8 & 16 & -8 \\ 0 & 0 & 1 & -8 & 7 \end{bmatrix}$$

Incorporating the boundary condition $\hat{u}_1 = 0$, we get

$$\begin{bmatrix} 32 & -16 & 0 & 0 \\ -16 & 21 & -8 & 1 \\ 0 & -8 & 16 & -8 \\ 0 & 1 & -8 & 7 \end{bmatrix} \begin{bmatrix} \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \end{bmatrix} = 1.25 \times 10^{-7} \begin{bmatrix} 100 \\ 150 \\ 800 \\ 300 \end{bmatrix}$$

Solving, we get $\hat{u}_1 = 0$, $\hat{u}_2 = 1.39323 \times 10^{-5}$ m, $\hat{u}_3 = 2.7083 \times 10^{-5}$ m, $\hat{u}_4 = 4.791667 \times 10^{-5}$ m, $\hat{u}_5 = 5.625 \times 10^{-5}$ m. The reaction is

$$R_1 = 0.8 \times 10^7 [-16\hat{u}_2 + 2\hat{u}_3] = -1350 \text{ N}.$$

The stresses in the two elements are

$$\begin{aligned} \tau^{(1)} &= E\boldsymbol{b}\hat{\boldsymbol{u}}^{(1)} \\ &= \frac{2E}{l_1} \begin{bmatrix} -\frac{1-2\xi}{2} & -2\xi & \frac{1-2\xi}{2} \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} \\ \tau^{(2)} &= \frac{2E}{l_2} \begin{bmatrix} -\frac{1-2\xi}{2} & -2\xi & \frac{1-2\xi}{2} \end{bmatrix} \begin{bmatrix} \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \end{bmatrix}. \end{aligned}$$

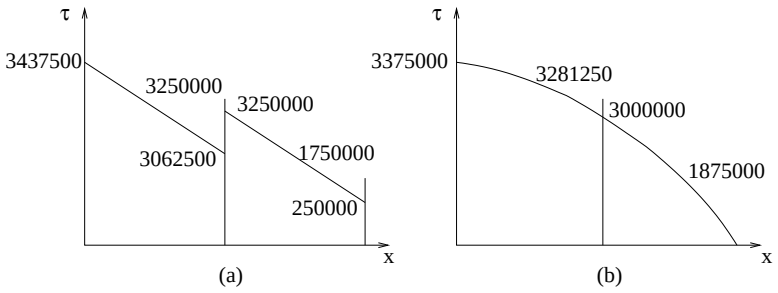


Figure 3.9: Stress field in rotating bar: (a) Finite element & (b) Exact solutions

The exact solution is

$$u = \frac{\rho\omega^2}{2E}\left(L^2 x - \frac{x^3}{3}\right),$$
$$\tau = \frac{1}{2}\rho\omega^2(L^2 - x^2).$$

Note that the FEM displacement solution is exact at the nodes! The variation of the stress field as per the finite element solution and the exact solution are shown in Fig. 3.9.

Example 2:

Problems with thermal stresses can be treated as initial strain problems. For example, for a one-dimensional problem, we can take $\epsilon^0 = \alpha\Delta T$. In general, the load vector is given by

$$F_{\text{th}} = \int \boldsymbol{B}_u^t \boldsymbol{C}\boldsymbol{\epsilon}_c^0 \, d\Omega.$$

For a one-dimensional linear bar element, with constant E, ΔT and A, and with $\boldsymbol{B} = \frac{1}{l_e}[-1\ 1]$, we have

$$F_{\text{th}} = \frac{1}{2l_e} E\alpha\Delta T \int_{-1}^{1} \begin{bmatrix} -1 \\ 1 \end{bmatrix} A l_e \, d\xi$$
$$= EA\alpha\Delta T \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \tag{3.10}$$

We shall use this load vector to solve the statically indeterminate problem shown in Fig. 3.10. The material and geometric properties are $E_1 = 70$ GPa, $A_1 = 900$ mm^2, $\alpha_1 = 23 \times 10^{-6}/^{\circ}$C, $E_2 = 200$ GPa, $A_2 = 1200$ mm^2, $\alpha_2 = 11.7 \times 10^{-6}/^{\circ}$C. The point load and change in temperature are $P = 3 \times 10^5$ N and $\Delta T = 40^{\circ}$C.

The element stiffness matrices are

$$\boldsymbol{K}^{(1)} = \frac{E_1 A_1}{l_1} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{70 \times 10^3 \times 900}{200} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

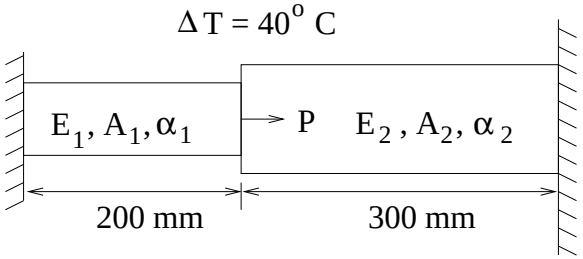


Figure 3.10: Stepped bar subjected to thermal loading

$$\boldsymbol{K}^{(2)} = \frac{E_2 A_2}{l_2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{200 \times 10^3 \times 1200}{300} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The global stiffness matrix after assembling the above matrices is

$$\boldsymbol{K} = 10^3 \begin{bmatrix} 315 & -315 & 0 \\ -315 & 1115 & -800 \\ 0 & -800 & 800 \end{bmatrix} \text{ N/mm}.$$

The global load vector due to the thermal loads (obtained using Eqn. 3.10) and the point load P is given by

$$\boldsymbol{f} = 10^3 \begin{bmatrix} -57.96 + R_1 \\ 245.64 \\ 112.32 + R_2 \end{bmatrix}.$$

Incorporating the boundary conditions $\hat{u}_1 = \hat{u}_3 = 0$, we get

$$(1115 \times 10^3)\hat{u}_2 = 245.64 \times 10^3,$$

which on solving yields $\hat{u}_2 = 0.22$ mm. The global displacement vector is, thus, $[0\ 0.22\ 0]^t$ mm. The stress in the bars is

$$\tau_1 = E_1 \boldsymbol{B}_1 \hat{\boldsymbol{u}}^{(1)} - E_1 \alpha_1 \Delta T = 12.60 \text{ MPa},$$
$$\tau_2 = E_2 \boldsymbol{B}_2 \hat{\boldsymbol{u}}^{(2)} - E_2 \alpha_2 \Delta T = -240.27 \text{ MPa}.$$

Note from the above example that statically indeterminate problems can be handled in a routine way by the finite element method.

3.6 Truss structures

Since elements of a truss have various orientations, we introduce local and global degrees of freedom for each member as shown. The local coordinate system is along the length of the member, while the global one is along the

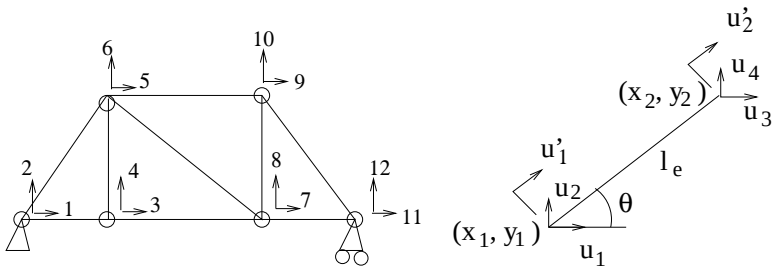


Figure 3.11: Global and local degrees of freedom for a truss member.

global coordinate axes. Thus, for example, each member of a two-dimensional truss of length l_e, and oriented at an angle θ has the global degrees of freedom u_1, u_2, u_3 and u_4, and the local degrees of freedom u'_1 and u'_2 as shown in Fig. 3.11. Note that

$$u'_1 = u_1 l + u_2 m,$$
$$u'_2 = u_3 l + u_4 m,$$

where $l = \cos\theta$ and $m = \sin\theta$ are the direction cosines which are given by

$$l = \frac{x_2 - x_1}{l_e};\ m = \frac{y_2 - y_1}{l_e};\ l_e = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

In matrix form, we have

$$\begin{bmatrix} u'_1 \\ u'_2 \end{bmatrix} = \boldsymbol{Q} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}, \tag{3.11}$$

where

$$\boldsymbol{Q} = \begin{bmatrix} l & m & 0 & 0 \\ 0 & 0 & l & m \end{bmatrix}.$$

In the local coordinate system, the element stiffness is given by

$$\boldsymbol{K'} = \frac{E_e A_e}{l_e} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The element stiffness matrix in global coordinates is derived using the fact that by virtue of the tensorial nature of the quantities involved, the virtual work contribution is the same irrespective of the coordinate system that we work in. Thus,

$$(\boldsymbol{v'})^t \boldsymbol{K'} \boldsymbol{u'} = \boldsymbol{v}^t (\boldsymbol{Q}^t \boldsymbol{K'} \boldsymbol{Q}) \boldsymbol{u},$$

leading to

$$\boldsymbol{K}^{(e)} = \boldsymbol{Q}^t \boldsymbol{K'} \boldsymbol{Q}$$

$$= \frac{E_e A_e}{l_e} \begin{bmatrix} l^2 & lm & -l^2 & -lm \\ lm & m^2 & -lm & -m^2 \\ -l^2 & -lm & l^2 & lm \\ -lm & -m^2 & lm & m^2 \end{bmatrix}.$$

One gets the same result by equating the strain energy in the local and global coordinate systems.

The stress in a member is given by

$$\tau = \frac{E_e}{l_e} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} u_1' \\ u_2' \end{bmatrix} - E_e \alpha_e \Delta T$$

$$= \frac{E_e}{l_e} \begin{bmatrix} -l & -m & l & m \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} - E_e \alpha_e \Delta T.$$

The thermal load vector is obtained by equating the virtual work expressions in the local and global coordinate systems. We get

$$(\boldsymbol{v}')^t \boldsymbol{f}'_{\text{th}} = (\boldsymbol{v})^t \boldsymbol{Q}^t \int_{-1}^{1} \boldsymbol{B}^t E_e \alpha_e \Delta T A_e \frac{l_e}{2} \, d\xi,$$

leading us to the expression

$$\boldsymbol{f}_{\text{th}} = E_e \alpha_e \Delta T A_e l_e \boldsymbol{Q}^t \boldsymbol{B}^t = E_e \alpha_e \Delta T A_e \begin{bmatrix} -l \\ -m \\ l \\ m \end{bmatrix}.$$

It is easy to extend the above discussion for two-dimensional trusses to three-dimensional ones. The transformation matrix between the local and global coordinate systems is now given by

$$\boldsymbol{Q} = \begin{bmatrix} l & m & n & 0 & 0 & 0 \\ 0 & 0 & 0 & l & m & n \end{bmatrix},$$

where the direction cosines, l, m, and n, are

$$l = \frac{x_2 - x_1}{l_e};\ m = \frac{y_2 - y_1}{l_e};\ n = \frac{z_2 - z_1}{l_e}.$$

3.7 One-Dimensional Heat Transfer Problems

Since the procedure for formulating the stiffness and load vectors is exactly analogous to the elasticity problem, we shall illustrate it directly by means of an example.

Problem: Find the temperature distribution in a bar of length L subject to the following boundary conditions:

1. $T(0) = T_1$; $T(L) = T_2$.
2. $-k\,\frac{dT}{dx}\big|_{x=0} = \bar{q}$; $\left[k\frac{dT}{dx} + h(T - T_\infty)\right]_{x=L} = 0$.
3. $T(0) = T_1$; $k\frac{dT}{dx}\big|_{x=L} = \bar{q}$.

Solution: If we assume linear elements, then

$$T = N_1 w_1 + N_2 w_2,$$

where $N_1 = (1-\xi)/2$ and $N_2 = (1+\xi)/2$. Thus,

$$\frac{dT}{dx} = \begin{bmatrix} \frac{dN_1}{dx} & \frac{dN_2}{dx} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

1. For the boundary conditions given, we see that $\Gamma = \Gamma_T$. Thus,

$$\boldsymbol{K}^{(e)} = \int_{-1}^{1} k\boldsymbol{B}^t\boldsymbol{B}\frac{l}{2}A\,d\xi = \frac{kA}{l}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

where l denotes the length of the element. If we divide the domain into 3 elements of equal length ($l = L/3$), then the finite element equations prior to incorporating the boundary conditions are

$$\frac{kA}{l}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} T_1 \\ w_2 \\ w_3 \\ T_2 \end{bmatrix} = \frac{QAl}{2}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + A\begin{bmatrix} q_1 \\ 0 \\ 0 \\ q_4 \end{bmatrix},$$

where q_1 and q_4 are the unknown heat fluxes at the two ends. The equations for w_2 and w_3 can be written as

$$\frac{kA}{l}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} w_2 \\ w_3 \end{bmatrix} = QAl\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{kA}{l}\begin{bmatrix} T_1 \\ T_2 \end{bmatrix},$$

while those for q_1 and q_4 can be written as

$$\begin{bmatrix} q_1 \\ q_4 \end{bmatrix} = \frac{k}{l}\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} T_1 \\ w_2 \\ w_3 \\ T_2 \end{bmatrix} - \frac{Ql}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Solving for w_2 and w_3, we get

$$w_2 = \frac{QL^2}{9k} + \frac{1}{3}(2T_1 + T_2),$$
$$w_3 = \frac{QL^2}{9k} + \frac{1}{3}(T_1 + 2T_2),$$

which on substituting in the equations for q_1 and q_4 yields

$$q_1 = -\frac{1}{2}QL + \frac{k}{3}(T_1 - T_2),$$
$$q_4 = -\frac{1}{2}QL + \frac{k}{3}(T_2 - T_1).$$

The exact solution is

$$T(x) = \frac{QL^2}{2k}\left[\frac{x}{L} - \left(\frac{x}{L}\right)^2\right] + \frac{1}{L}(T_2 - T_1)x + T_1,$$
$$q(x) = -k\frac{dT}{dx} = -\frac{QL}{2}\left(1 - \frac{2x}{L}\right) - \frac{k}{L}(T_2 - T_1).$$

Note that the solution for the temperature field is exact at the nodes. However, since we have assumed a linear interpolation for the temperature, while the real variation is quadratic, there is an error in the finite element solution at points other than the nodes. Since the exact solution is quadratic, we can get the exact solution simply using one quadratic element. We now have

$$\frac{kA}{3l}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix}\begin{bmatrix} T_1 \\ w_2 \\ T_2 \end{bmatrix} = \frac{AQl}{6}\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} + \begin{bmatrix} q_1 \\ 0 \\ q_3 \end{bmatrix},$$

where $l = L$. The solution is

$$w_2 = \frac{1}{8}\frac{QL^2}{k} + \frac{1}{2}(T_1 + T_2).$$

The heat flux at the ends is

$$q_1 = -\frac{1}{2}QL + \frac{k}{L}(T_1 - T_2),$$
$$q_3 = -\frac{1}{2}QL + \frac{k}{L}(T_2 - T_1).$$

2. The two ends of the bar constitute Γ_q. The finite element equations are

$$\frac{kA}{l}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 + \frac{hl}{k} \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \frac{AQl}{2}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} A\bar{q} \\ 0 \\ 0 \\ AhT_\infty \end{bmatrix}$$

On solving, we get

$$
\begin{aligned}
w_1 &= \frac{QL+\bar{q}}{h} + \frac{L(QL+2\bar{q})}{2k} + T_\infty, \\
w_2 &= \frac{QL+\bar{q}}{h} + \frac{2L(2QL+3\bar{q})}{9k} + T_\infty, \\
w_3 &= \frac{QL+\bar{q}}{h} + \frac{L(5QL+6\bar{q})}{18k} + T_\infty, \\
w_4 &= \frac{QL+\bar{q}}{h} + T_\infty.
\end{aligned}
$$

The exact solution is

$$
T(x) = \frac{QL^2}{2k}\left(1 + \frac{2k}{hL} - \frac{x^2}{L^2}\right) + \frac{\bar{q}L}{k}\left(1 + \frac{k}{hL} - \frac{x}{L}\right) + T_\infty.
$$

Note again that the temperature solution is exact at the nodes.

3. The right end of the bar constitutes Γ_q. The finite element equations are

$$
\frac{kA}{l}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} T_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \frac{AQl}{2}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} Aq_1 \\ 0 \\ 0 \\ A\bar{q} \end{bmatrix},
$$

which on incorporating the boundary conditions becomes

$$
\frac{kA}{l}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} w_2 \\ w_3 \\ w_4 \end{bmatrix} = \frac{AQl}{2}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ A\bar{q} \end{bmatrix} + \frac{kA}{l}\begin{bmatrix} T_1 \\ 0 \\ 0 \end{bmatrix}.
$$

On solving, we get

$$
\begin{aligned}
w_2 &= \frac{L(5QL+6\bar{q})}{18k} + T_1, \\
w_3 &= \frac{2L(2QL+3\bar{q})}{9k} + T_1, \\
w_4 &= \frac{L(QL+2\bar{q})}{2k} + T_1.
\end{aligned}
$$

The flux at the left end is

$$
q_1 = \frac{k}{l}(T_1 - w_2) - \frac{1}{2}Ql = -(QL+\bar{q}).
$$

The exact solution is

$$
\begin{aligned}
T(x) &= \frac{L^2}{k}\left[\frac{Q}{2}\left(\frac{2x}{L} - \frac{x^2}{L^2}\right) + \frac{\bar{q}x}{L^2}\right] + T_1, \\
q(x) &= -k\frac{dT}{dx} = -\left[QL\left(1 - \frac{x}{L}\right) + \bar{q}\right].
\end{aligned}
$$

Note that the temperature solution is exact at the nodes.

3.8 Beams and Frames: An Introduction to C^1 elements

We shall consider only the classical or slender beam theory here, and not the Timoshenko beam theory which includes the effect of shear deformation. We restrict our attention to beams with symmetric cross-sections.

Assuming that the x axis is directed along the length, and the y-axis is oriented downward with $y = 0$ at the neutral plane, we have

$$\tau_{xx} = -\frac{My}{I} = E\epsilon_{xx}.$$

The remaining stress components are assumed to be zero. The governing differential equation for the displacement is

$$EI\frac{d^2v}{dx^2} = M(x),$$

where $M(x)$ is the bending moment along the beam.

The expression for the potential energy is

$$\begin{aligned}\Pi &= \frac{1}{2}\int_\Omega \boldsymbol{\epsilon}_c^t \boldsymbol{\tau}\, d\Omega - \int_{\Gamma_t} \bar{\boldsymbol{t}}\cdot\boldsymbol{u}\, d\Gamma - \int_\Omega \rho\boldsymbol{b}\cdot\boldsymbol{u}\, d\Omega - \sum P_i u_i \\ &= \frac{1}{2}\int_0^L \frac{M^2}{EI^2}\left(\int_A y^2\, dA\right) dx - \int qv\, dx - \sum p_i v_i - \sum M_i v_i' \\ &= \frac{1}{2}\int_0^L EI\left(\frac{d^2v}{dx^2}\right)^2 dx - \int_0^L qv\, dx - \sum P_i v_i - \sum M_i v_i'.\end{aligned}$$

The variational formulation is obtained by setting the first variation of Π to zero. We get

$$\int_0^L EI\frac{d^2v}{dx^2}\frac{d^2\phi}{dx^2}\, dx - \int_0^L q\phi - \sum P_i\phi_i - \sum M_i\phi_i' = 0 \quad \forall\phi.$$

The variational formulation can also be obtained from the governing equations

$$\frac{dM}{dx} = -V; \quad \frac{dV}{dx} = -q; \quad EI\frac{d^2v}{dx^2} = M,$$

where V is the shear force. The above equations lead to the governing equation

$$\frac{d^2}{dx^2}\left(EI\frac{d^2v}{dx^2}\right) - q = 0.$$

Multiplying this governing equation by a test function ϕ, and integrating by parts leads us to the variational formulation as has already been demonstrated in Section 1.6.

In order that d^2v/dx^2 is well-defined, we need $v \in C^1$, i.e, both v and v' must be continuous. Now we formulate *Hermite* shape functions which

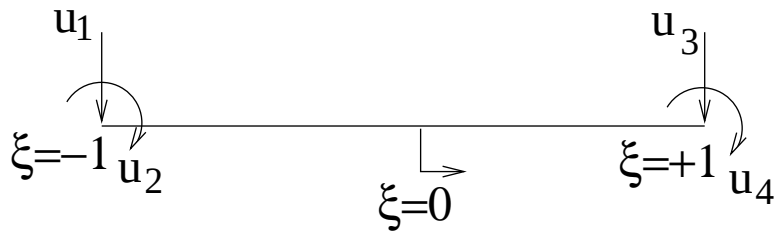


Figure 3.12: Degrees of freedom for a beam.

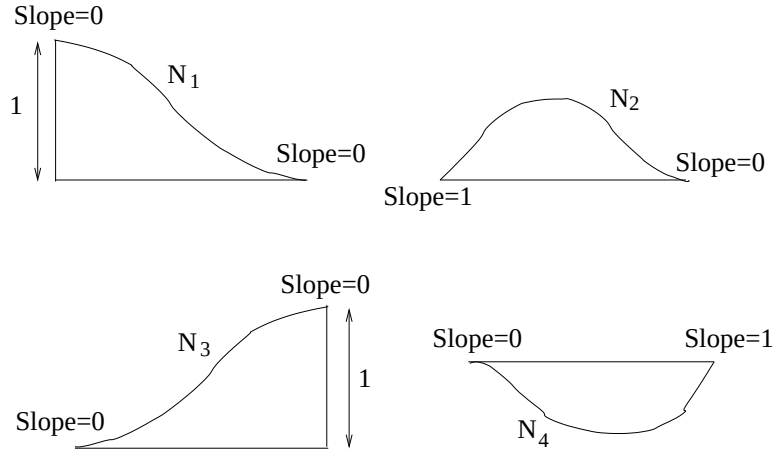


Figure 3.13: Hermite shape functions.

ensure that $v \in C^1$. Let u_1, u_2, u_3 and u_4 be the local degrees of freedom as shown in Fig. 3.12, where

$$(u_1, u_2, u_3, u_4) \equiv \left(v_1, \left.\frac{dv}{dx}\right|_1, v_2, \left.\frac{dv}{dx}\right|_2 \right).$$

The vertical displacement v is given by

$$v(\xi) = N_1 v_1 + N_2 v_1' + N_3 v_2 + N_4 v_2',$$

where

$$v_1' = \left.\frac{dv}{d\xi}\right|_{\xi=-1}; \quad v_2' = \left.\frac{dv}{d\xi}\right|_{\xi=+1}.$$

In order that $v(-1) = v_1$, $v'(-1) = v_1'$, $v(+1) = v_2$ and $v'(+1) = v_2'$, the shape functions should be as shown in Fig. 3.13. We assume each of the N_i's to be of the form

$$N_i(\xi) = a_i + b_i\xi + c_i\xi^2 + d_i\xi^3,$$

and find the constants a_i-d_i based on the boundary conditions that each shape function has to satisfy. For example, $N_1(-1) = 1$, $N_1'(-1) = 0$, $N_1(+1) = 0$, and $N_1'(+1) = 0$. We get

$$N_1 = \frac{1}{4}(1-\xi)^2(2+\xi) = \frac{1}{4}(2 - 3\xi + \xi^3),$$

$$N_2 = \frac{1}{4}(1-\xi)^2(1+\xi) = \frac{1}{4}(1-\xi-\xi^2+\xi^3),$$
$$N_3 = \frac{1}{4}(1+\xi)^2(2-\xi) = \frac{1}{4}(2+3\xi-\xi^3),$$
$$N_4 = \frac{1}{4}(1+\xi)^2(\xi-1) = \frac{1}{4}(-1-\xi+\xi^2+\xi^3).$$

We assume that the coordinate interpolation is still given by

$$x = \frac{1}{2}(1-\xi)x_1 + \frac{1}{2}(1+\xi)x_2,$$

leading to $dx = l_e\, d\xi/2$. Using the chain rule

$$\frac{dv}{d\xi} = \frac{dv}{dx}\frac{dx}{d\xi} = \frac{l_e}{2}\frac{dv}{dx}.$$

Since $(dv/dx)_1 = u_2$ and $(dv/dx)_2 = u_4$, we have

$$v(\xi) = N_1 u_1 + \frac{l_e}{2}N_2 u_2 + N_3 u_3 + \frac{l_e}{2}N_4 u_4.$$

In matrix form, the above equation can be written as

$$\boldsymbol{v} = \boldsymbol{N}\hat{\boldsymbol{u}}^{(e)}, \tag{3.12}$$

where

$$\boldsymbol{N} \equiv \begin{bmatrix} N_1 & \frac{l_e}{2}N_2 & N_3 & \frac{l_e}{2}N_4 \end{bmatrix},$$
$$\hat{\boldsymbol{u}}^{(e)} \equiv \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \end{bmatrix}^t.$$

Using the chain rule again, we get

$$\frac{d^2v}{dx^2} = \frac{4}{l_e^2}\frac{d^2v}{d\xi^2}.$$

Using the same interpolation for the variations and substituting in the variational formulation, we get

$$\hat{\boldsymbol{\phi}}^t \int_{-1}^{1} \frac{16EI}{l_e^4}\frac{d^2\boldsymbol{N}^t}{d\xi^2}\frac{d^2\boldsymbol{N}}{d\xi^2}\frac{l_e}{2}\, d\xi \hat{\boldsymbol{u}} = \hat{\boldsymbol{\phi}}^t \int_{-1}^{1} \boldsymbol{N}^t q \frac{l_e}{2}\, d\xi + \sum P_i\phi(x_i) + \sum M_i\phi'(x_i).$$

Since $\hat{\boldsymbol{\phi}}$ is arbitrary, we get the familiar equation

$$\boldsymbol{K}^{(e)}\hat{\boldsymbol{u}}^{(e)} = \hat{\boldsymbol{f}}^{(e)},$$

where

$$\boldsymbol{K}^{(e)} = \int_{-1}^{1} \boldsymbol{B}^t(EI)\boldsymbol{B}\frac{l_e}{2}\, d\xi$$

$$\hat{\boldsymbol{f}}^{(e)} = \int_{-1}^{1} \boldsymbol{N}^t q \frac{l_e}{2}\, d\xi + \{N_i(\xi_j)P_j\} + \{N_i'(\xi_j)M_j\},$$

with $\boldsymbol{B}$ given by

$$\boldsymbol{B} = \frac{4}{l_e^2}\frac{d^2\boldsymbol{N}}{d\xi^2} = \frac{4}{l_e^2}\left[\tfrac{3}{2}\xi \quad \tfrac{l_e}{4}(3\xi-1) \quad -\tfrac{3}{2}\xi \quad \tfrac{l_e}{4}(3\xi+1)\right].$$

If EI is assumed to be constant over the element, then we get

$$\boldsymbol{K}^{(e)} = \frac{EI}{l_e^3}\begin{bmatrix} 12 & 6l_e & -12 & 6l_e \\ 6l_e & 4l_e^2 & -6l_e & 2l_e^2 \\ -12 & -6l_e & 12 & -6l_e \\ 6l_e & 2l_e^2 & -6l_e & 4l_e^2 \end{bmatrix}$$

Assuming that q is uniform over the element, and that there are no point loads and moments, the element load vector is

$$\boldsymbol{f}^{(e)} = \begin{bmatrix} \frac{ql_e}{2} \\ \frac{ql_e^2}{12} \\ \frac{ql_e}{2} \\ -\frac{ql_e^2}{12} \end{bmatrix}.$$

The element level matrices are assembled into the global ones as usual, and we solve the system of equations $\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}$. The moment and shear force are found as follows:

$$\begin{aligned} M &= EI\frac{d^2v}{dx^2} \\ &= EI\frac{d^2\boldsymbol{N}}{dx^2}\hat{\boldsymbol{u}}^{(e)} \\ &= \frac{4EI}{l_e^2}\frac{d^2\boldsymbol{N}}{d\xi^2}\hat{\boldsymbol{u}}^{(e)} \\ &= \frac{EI}{l_e^2}\left[6\xi u_1 + (3\xi-1)l_e u_2 - 6\xi u_3 + (3\xi+1)l_e u_4\right], \\ V &= -\frac{dM}{dx} \\ &= -\frac{2}{l_e}\frac{dM}{d\xi} \\ &= -\frac{6EI}{l_e^3}\left[2u_1 + l_e u_2 - 2u_3 + l_e u_4\right]. \end{aligned}$$

As an example, consider the beam shown in Fig. 3.14. The Young's modulus and moment of inertia are $E = 200$ GPa and $I = 4 \times 10^{-6}$ m^4. The

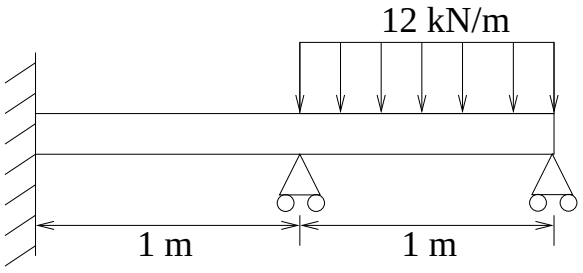


Figure 3.14: Beam subjected to distributed load.

element stiffness matrices are

$$\boldsymbol{K}^{(1)} = \boldsymbol{K}^{(2)} = 8 \times 10^5 \begin{bmatrix} 12 & 6 & -12 & 6 \\ 6 & 4 & -6 & 2 \\ -12 & -6 & 12 & -6 \\ 6 & 2 & -6 & 4 \end{bmatrix}$$

The global stiffness matrix is

$$\boldsymbol{K} = 8 \times 10^5 \begin{bmatrix} 12 & 6 & -12 & 6 & 0 & 0 \\ 6 & 4 & -6 & 2 & 0 & 0 \\ -12 & -6 & 24 & 0 & -12 & 6 \\ 6 & 2 & 0 & 8 & -6 & 2 \\ 0 & 0 & -12 & -6 & 12 & -6 \\ 0 & 0 & 6 & 2 & -6 & 4 \end{bmatrix}.$$

The global load vector is

$$\boldsymbol{f} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 + 6000 \\ 1000 \\ R_5 + 6000 \\ -1000 \end{bmatrix}.$$

The prescribed degrees of freedom are $\hat{u}_1 = \hat{u}_2 = \hat{u}_3 = \hat{u}_5 = 0$. The equations for the unknowns $\hat{u}_4$ and $\hat{u}_6$ are

$$8 \times 10^5 \begin{bmatrix} 8 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \hat{u}_4 \\ \hat{u}_6 \end{bmatrix} = \begin{bmatrix} 1000 \\ -1000 \end{bmatrix},$$

which on solving yields $\hat{u}_4 = 2.679 \times 10^{-4}$ radians and $\hat{u}_6 = -4.464 \times 10^{-4}$ radians. The reactions are given by

$$
\begin{aligned}
R_1 &= 8 \times 10^5[6\hat{u}_4] = 1285.92 \text{ N}, \\
R_2 &= 8 \times 10^5[2\hat{u}_4] = 428.64 \text{ N-m}, \\
R_3 &= 8 \times 10^5[6\hat{u}_6] - 6000 = -8142.72 \text{ N}, \\
R_5 &= 8 \times 10^5[-6\hat{u}_4 - 6\hat{u}_6] - 6000 = -5143.2 \text{ N}.
\end{aligned}
$$

The positive sign on R_1 indicates that the reaction at the wall is directed downwards, while the negative signs on R_3 and R_5 indicate that the reactions at the roller supports are directed upwards. The positive sign on R_2 indicates that the moment is clockwise at the wall.

The deflection at the midpoint of the second element is

$$
\begin{aligned}
v &= \boldsymbol{N}\hat{\boldsymbol{u}}^{(2)}\Big|_{\xi=0} \\
&= \begin{bmatrix} 0.5 & 0.5(0.25) & 0.5 & -0.5(0.25) \end{bmatrix} \begin{bmatrix} 0 \\ \hat{u}_4 \\ 0 \\ \hat{u}_6 \end{bmatrix} \\
&= 0.0893 \text{ mm}.
\end{aligned}
$$

The moment and shear force at this point are

$$
\begin{aligned}
M &= \frac{EI}{l_e^2}[-l_e\hat{u}_4 + l_e\hat{u}_6] \\
&= \frac{EI}{l_e}[\hat{u}_6 - \hat{u}_4], \\
V &= -12EI(\hat{u}_4 + \hat{u}_6).
\end{aligned}
$$

The shear force in element 1 is

$$
V = -6EI\hat{u}_4 = -R_1,
$$

as expected.

3.9 Plane Frames

The element-level matrices for a frame element are derived using the element-level matrices for the bar and beam elements in a manner analogous to the derivation of the truss element in Section 3.6. For the plane frame element shown in Fig. 3.15, the relation between the displacement degrees of freedom in the local and global coordinate systems is

$$
\hat{\boldsymbol{u}}' = \boldsymbol{Q}\hat{\boldsymbol{u}},
$$

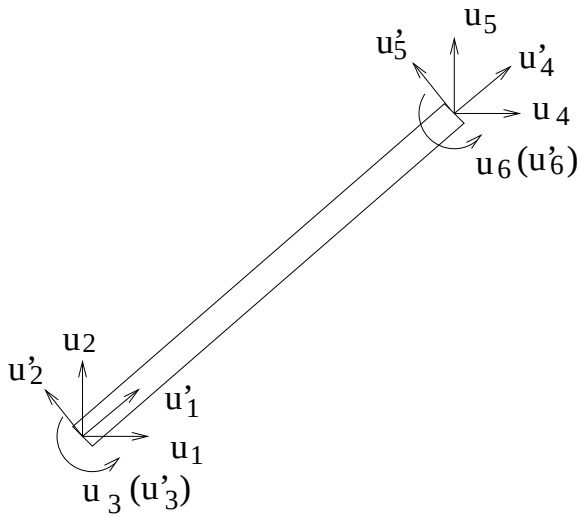


Figure 3.15: Frame element.

where

$$\boldsymbol{Q} = \begin{bmatrix} l & m & 0 & 0 & 0 & 0 \\ -m & l & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l & m & 0 \\ 0 & 0 & 0 & -m & l & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Similar to the truss element, the element stiffness matrix in global coordinates is given by

$$\boldsymbol{K}^{(e)} = \boldsymbol{Q}^t \boldsymbol{K}' \boldsymbol{Q},$$

where

$$\boldsymbol{K}' = \begin{bmatrix} \frac{EA}{l} & 0 & 0 & -\frac{EA}{l} & 0 & 0 \\ 0 & \frac{12EI}{l^3} & \frac{6EI}{l^2} & 0 & -\frac{12EI}{l^3} & \frac{6EI}{l^2} \\ 0 & \frac{6EI}{l^2} & \frac{4EI}{l} & 0 & -\frac{6EI}{l^2} & \frac{2EI}{l} \\ -\frac{EA}{l} & 0 & 0 & \frac{EA}{l} & 0 & 0 \\ 0 & -\frac{12EI}{l^3} & -\frac{6EI}{l^2} & 0 & \frac{12EI}{l^3} & -\frac{6EI}{l^2} \\ 0 & \frac{6EI}{l^2} & \frac{2EI}{l} & 0 & -\frac{6EI}{l^2} & \frac{4EI}{l} \end{bmatrix}.$$

The load vector in global coordinates is

$$\boldsymbol{f}^{(e)} = \boldsymbol{Q}^t \boldsymbol{f}',$$

where $\boldsymbol{f}'$ is the load vector in the local coordinate system.

3.10 A Note on Equation Solving

We have seen that our finite element equations are a system of algebraic equations of the form

$$\boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}.$$

There are two types of methods available for solving these equations, namely
(i) Direct methods which are based on some variant of Gauss elimination,
(ii) Iterative methods.

In the direct methods, $\boldsymbol{K}$ is decomposed into a lower triangular matrix, $\boldsymbol{L}$ with unit diagonals, and an upper triangular matrix $\boldsymbol{U}$, i.e.,

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{U}.$$

The solution to the equations is obtained by solving the equations

$$\boldsymbol{L}\boldsymbol{x} = \boldsymbol{f} \text{ and } \boldsymbol{U}\hat{\boldsymbol{u}} = \boldsymbol{x}.$$

Thus,

$$\begin{aligned} x_1 &= f_1 \\ x_i &= f_i - \sum_{j=1}^{i-1} L_{ij} x_j \quad i = 2, \ldots, n. \end{aligned} \tag{3.13}$$

$$\begin{aligned} \hat{u}_n &= \frac{x_n}{U_{nn}}, \\ \hat{u}_i &= \frac{1}{U_{ii}} \left[x_i - \sum_{j=i+1}^{n} U_{ij} x_j \right] \quad i = n-1, n-2, \ldots, 1. \end{aligned} \tag{3.14}$$

Eqn. 3.13 is called 'forward elimination', while Eqn. 3.14 is called 'back substitution'. The actual decomposition $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{U}$ is carried out using some variation of Gauss elimination.

Once the *L-U* decomposition is computed, several solutions for different load vectors can be computed using Eqns. 3.13 and 3.14. This process is called 'resolution' since it is not necessary to recompute $\boldsymbol{L}$ and $\boldsymbol{U}$. Note that the *L-U* decomposition is very costly compared to the resolution ($\mathrm{o}\,(n^3)$ versus $\mathrm{o}\,(n^2)$).

Several strategies exist for taking advantage of the banded and symmetric form of the stiffness matrix. Even further storage reductions are possible by using a 'profile' of the $\boldsymbol{K}$ matrix. We then need to use a profile equation solving program. See [4] for details of such solvers and of iterative techniques such as the Gauss-Seidel method.

Chapter 4

Modeling of Two-Dimensional Problems

In this chapter, we shall consider problems which can be modeled as two-dimensional problems such as plane stress, plane strain or axisymmetric problems. We shall first consider the constant strain triangle (CST) element, and later quadrilateral and higher-order triangular elements.

4.1 Constant Strain Triangle Element (CST element)

The domain is discretized into triangular 'elements' as shown in Fig. 4.1. The corners of the elements are known as nodes[1]. Note that the curved portion of the body is approximated by piecewise linear segments corresponding to the edges of the elements.

Since we are modeling a two-dimensional domain, each node has two degrees of freedom. Thus, the degrees of freedom corresponding to node j (before incorporating the boundary conditions) are $\hat{u}_{2j-1}$ and $\hat{u}_{2j}$ The geometric input data which is used by a finite element code for computing the element level matrices is the set of nodal coordinates and the connectivity

[1]Nodes need not always be located at the corners of an element as we have seen even in the case of a quadratic one-dimensional element, but in the CST element they are.

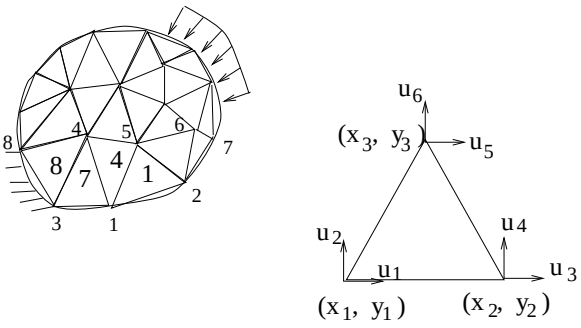


Figure 4.1: Discretization of a two-dimensional domain; a typical element

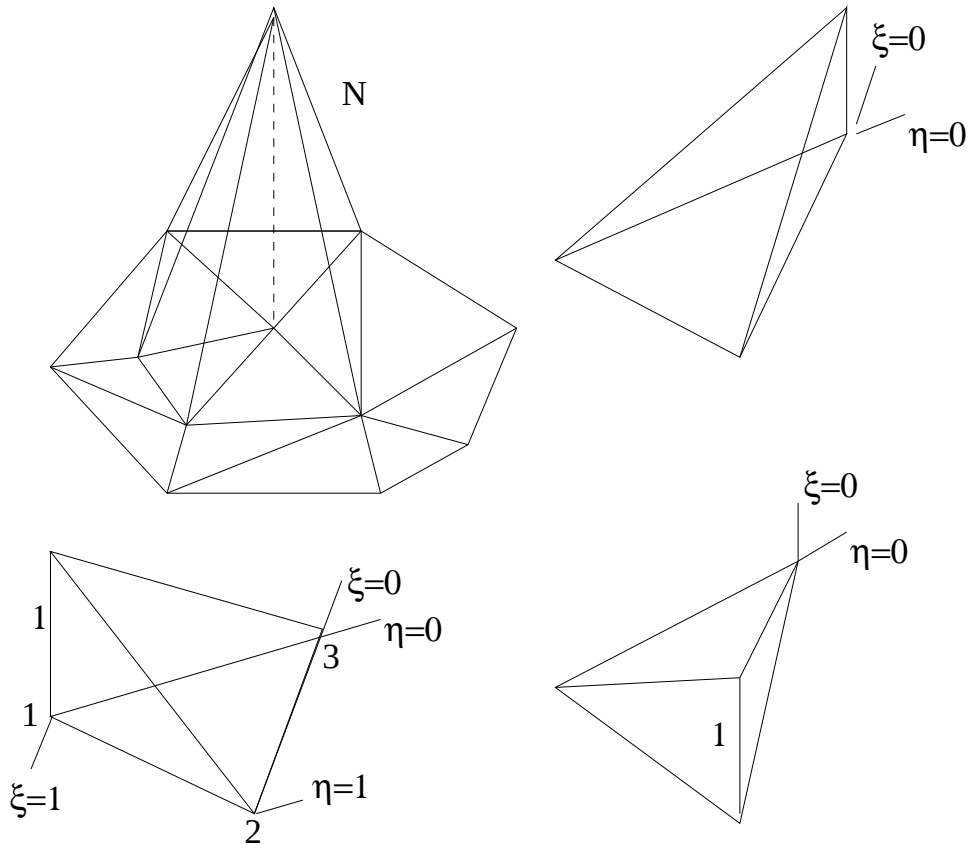


Figure 4.2: Shape functions at a node; restriction of shape function to an element

data. The connectivity of a given element is the node numbers associated with that element. For example, in Fig. 4.1, the connectivity of element 1 is $(1, 2, 5)$, that of element 4 is $(1, 5, 4)$ that of element 8 is $(3, 4, 8)$ and so on.

The shape functions are 'hat functions' as shown in Fig. 4.2 assuming a value of one at the node associated with it and zero elsewhere. Due to this property they can be represented by area coordinates as shown in Fig. 4.3. Mathematically, they are given by

$$N_1 = \frac{A_1}{A};\ N_2 = \frac{A_2}{A};\ N_3 = \frac{A_3}{A}. \tag{4.1}$$

Note that $N_1 + N_2 + N_3 = 1$. We define natural coordinates as

$$\xi = \frac{A_1}{A};\ \eta = \frac{A_2}{A}. \tag{4.2}$$

With this definition, we have $\xi = 0$ on side 2-3, and $\xi = 1$ at point 1, and $\eta = 0$ on side 1-3, and $\eta = 1$ at point 2, as shown in Fig. 4.3. From Eqns. 4.1 and 4.2, we get

$$N_1 = \xi;\ N_2 = \eta;\ N_3 = 1 - \xi - \eta. \tag{4.3}$$

For developing an isoparametric formulation, we use the same shape functions for interpolating both, the displacements and the geometry. We thus

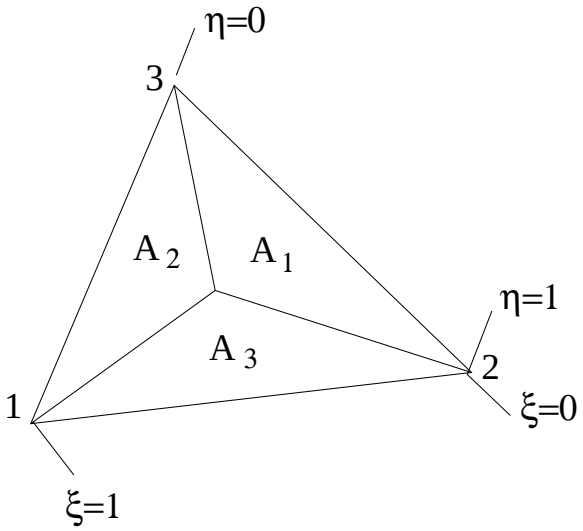


Figure 4.3: Representation of shape functions using area coordinates

have

$$\begin{aligned} u &= N_1 u_1 + N_2 u_3 + N_3 u_5, \\ v &= N_1 u_2 + N_2 u_4 + N_3 u_6, \end{aligned} \tag{4.4}$$

or, alternatively,

$$\boldsymbol{u} = \boldsymbol{N}\hat{\boldsymbol{u}},$$

where

$$\boldsymbol{N} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 \end{bmatrix}$$
$$\hat{\boldsymbol{u}} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 \end{bmatrix}^t.$$

Using the same interpolation functions for the geometry, we have

$$\begin{aligned} x &= N_1 x_1 + N_2 x_2 + N_3 x_3, \\ y &= N_1 y_1 + N_2 y_2 + N_3 y_3. \end{aligned} \tag{4.5}$$

In a sub-parametric formulation, the geometry is represented by lower-order elements than those used to approximate the displacements, e.g., Euler-Bernoulli beam element where Hermite shape functions are used to approximate the deflection while the geometry is interpolated using linear interpolation functions.

In superparametric formulations, the geometry is represented by higher-order interpolations than the displacement. Such a formulation is seldom used in practice.

Using Eqn. 4.3, Eqns. 4.4 and 4.5 can be written as

$$u = (u_1 - u_5)\xi + (u_3 - u_5)\eta + u_5,$$

$$
\begin{aligned}
v &= (u_2 - u_6)\xi + (u_4 - u_6)\eta + u_6, \\
x &= (x_1 - x_3)\xi + (x_2 - x_3)\eta + x_3, \\
y &= (y_1 - y_3)\xi + (y_2 - y_3)\eta + y_3,
\end{aligned}
$$

We now turn to the evaluation of the $\boldsymbol{B}$ matrix. Using the chain rule, we have

$$
\begin{aligned}
\frac{\partial u}{\partial \xi} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \xi}, \\
\frac{\partial u}{\partial \eta} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial \eta} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \eta},
\end{aligned}
$$

or, in matrix form

$$
\begin{aligned}
\begin{bmatrix} \frac{\partial u}{\partial \xi} \\ \frac{\partial u}{\partial \eta} \end{bmatrix} &= \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{bmatrix} \\
&= \boldsymbol{J} \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{bmatrix},
\end{aligned}
$$

where (with $x_{ij} \equiv x_i - x_j$)

$$
\boldsymbol{J} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} x_{13} & y_{13} \\ x_{23} & y_{23} \end{bmatrix},
$$

is the Jacobian. Since

$$
\det \boldsymbol{J} = x_{13}y_{23} - x_{23}y_{13}
$$

is twice the area of the triangle, it is nonzero, and hence the Jacobian matrix is invertible. Thus,

$$
\begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} \frac{\partial u}{\partial \xi} \\ \frac{\partial u}{\partial \eta} \end{bmatrix}.
$$

where

$$
\boldsymbol{J}^{-1} = \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23} & -y_{13} \\ -x_{23} & x_{13} \end{bmatrix}
$$

If the points 1, 2 and 3 are ordered in a counterclockwise manner, $\det \boldsymbol{J}$ is positive in sign.

As an example, let the coordinates of the nodes of an element be given by $(1.5, 2)$, $(7, 3.5)$ and $(4, 7)$, and let P be a point inside the triangle with coordinates $(3.85, 4.8)$. Then the natural coordinates of P are obtained using

$$
3.85 = 1.5N_1 + 7N_2 + 4N_3,
$$

$$4.8 = 2N_1 + 3.5N_2 + 7N_3,$$
$$1 = N_1 + N_2 + N_3.$$

Solving the above set of equations, we get $N_1 = \xi = 0.3$ and $N_2 = \eta = 0.2$, $N_3 = 0.5$. The Jacobian matrix is

$$\boldsymbol{J} = \begin{bmatrix} x_{13} & y_{13} \\ x_{23} & y_{23} \end{bmatrix} = \begin{bmatrix} -2.5 & -5.0 \\ 3.0 & -3.5 \end{bmatrix},$$

with $\det \boldsymbol{J} = 23.75$.

To find the strain-displacement matrix, note that

$$\boldsymbol{\epsilon}_c = \begin{bmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial v}{\partial y} \\ \dfrac{\partial u}{\partial y} + \dfrac{\partial v}{\partial x} \end{bmatrix}$$
$$= \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23}(u_1 - u_5) - y_{13}(u_3 - u_5) \\ -x_{23}(u_2 - u_6) + x_{13}(u_4 - u_6) \\ -x_{23}(u_1 - u_5) + x_{13}(u_3 - u_5) + y_{23}(u_2 - u_6) - y_{13}(u_4 - u_6) \end{bmatrix}.$$

Now using the fact that $y_{13} = y_{31}$, $y_{12} = -y_{21}$ and so on, we get

$$\boldsymbol{B} = \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23} & 0 & y_{31} & 0 & y_{12} & 0 \\ 0 & x_{32} & 0 & x_{13} & 0 & x_{21} \\ x_{32} & y_{23} & x_{13} & y_{31} & x_{21} & y_{12} \end{bmatrix}$$

Since $\boldsymbol{B}$ is a matrix of constants which are independent of the position, the strain is constant over the element. This is as expected since the displacement field is a linear combination of $(1, x, y)$ leading to constant strain over the element. The stiffness matrix is

$$\begin{aligned} \boldsymbol{K} &= \int_\Omega \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, d\Omega \\ &= \int_{A_e} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} t \, dA \\ &= t \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} A_e, \end{aligned}$$

where the last step follows from the fact that $\boldsymbol{B}$ and $\boldsymbol{C}$ are constant over the element. Using Eqn. 3.2, the half-bandwidth is given by

$$n_{bw} = 2 \left[\max_{1 \le e \le n_e} m_e + 1 \right].$$

In the above equation m_e is given by

$$m_e = \max\{|i_1 - i_2|\,,\ |i_2 - i_3|\,,\ |i_1 - i_3|\}$$

where i_1, i_2 and i_3 are the node numbers of an element.

The consistent force vector due to body forces is given by

$$\boldsymbol{f}_b = \int_{A_e} \boldsymbol{N}^t \rho b t_e \, dA.$$

Assuming that the density and body force are constant, we get

$$\begin{aligned}
\boldsymbol{f}_b &= \rho t_e \int_{A_e} \begin{bmatrix} N_1 & 0 \\ 0 & N_1 \\ N_2 & 0 \\ 0 & N_2 \\ N_3 & 0 \\ 0 & N_3 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \end{bmatrix} dA \\
&= \rho t_e \int \begin{bmatrix} N_1 b_x \\ N_1 b_y \\ N_2 b_x \\ N_2 b_y \\ N_3 b_x \\ N_3 b_y \end{bmatrix} dA \\
&= \frac{\rho t_e A_e}{3} \begin{bmatrix} b_x \\ b_y \\ b_x \\ b_y \\ b_x \\ b_y \end{bmatrix},
\end{aligned}$$

where the last step follows from either the physical interpretation of $\int N_i \, dA$ as the volume of a tetrahedron of unit height over a triangular base (Volume=one-third of area of base multiplied by height), or by carrying out the integration explicitly as

$$\begin{aligned}
\int N_1 \, dA &= \int_0^1 \int_0^{1-\xi} N_1 \det \boldsymbol{J} \, d\eta d\xi \\
&= 2A_e \int_0^1 \int_0^{1-\xi} \xi \, d\eta d\xi
\end{aligned}$$

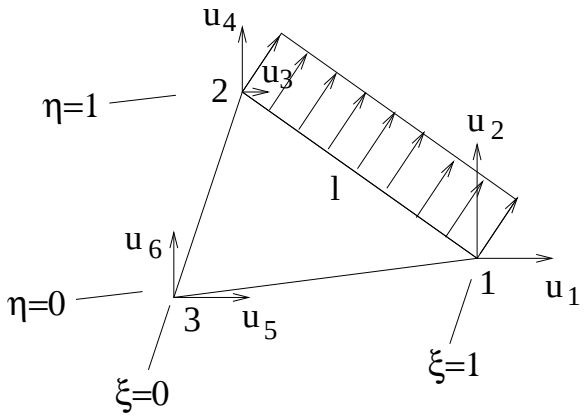


Figure 4.4: Computation of the load vector due to traction

$$= \frac{A_e}{3}.$$

The contribution to the load vector due to the tractions acting on an edge of the element is given by

$$\boldsymbol{f}_t = \int \boldsymbol{N}^t \bar{\boldsymbol{t}} t_e \, dA = t_e \int \begin{bmatrix} N_1 t_x \\ N_1 t_y \\ N_2 t_x \\ N_2 t_y \\ N_3 t_x \\ N_3 t_y \end{bmatrix} ds.$$

As an example consider the case shown in Fig. 4.4. On edge 1-2, we have

$$x = \xi x_1 + \eta x_2,$$
$$y = \xi y_1 + \eta y_2,$$

leading to

$$dx = d\xi x_1 + d\eta x_2,$$
$$dy = d\xi y_1 + d\eta y_2.$$

Since $N_3 = 0$ on edge 1-2, we have $\xi + \eta = 1$, and hence $d\xi = -d\eta$. Thus,

$$ds = \sqrt{dx^2 + dy^2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}\, d\xi = l \, d\xi.$$

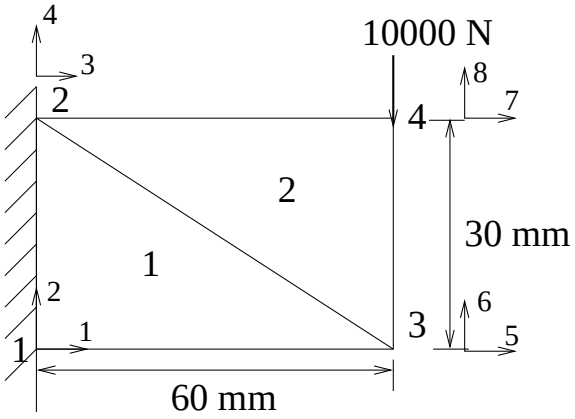


Figure 4.5: Example problem

Therefore

$$\boldsymbol{f}_t = t_e \begin{bmatrix} t_x \int_0^1 N_1 \, l_e d\xi \\ t_y \int_0^1 N_1 \, l_e d\xi \\ t_x \int_0^1 N_2 \, l_e d\xi \\ t_y \int_0^1 N_2 \, l_e d\xi \\ t_x \int_0^1 N_3 \, l_e d\xi \\ t_y \int_0^1 N_3 \, l_e d\xi \end{bmatrix}.$$

Using the fact that one edge 1-2, $N_1 = \xi$, $N_2 = \eta = 1 - \xi$ and $N_3 = 0$, we get

$$\boldsymbol{f}_t = \frac{t_e l}{2} \begin{bmatrix} t_x \\ t_y \\ t_x \\ t_y \\ 0 \\ 0 \end{bmatrix}.$$

Loads due to thermal effects are calculated by substituting $\boldsymbol{\epsilon}_0 = [\alpha\Delta T, \ \alpha\Delta T, \ 0]^t$ in the case of plane stress, and $\boldsymbol{\epsilon}_0 = (1+\nu)\,[\alpha\Delta T, \ \alpha\Delta T, \ 0]^t$ in the case of plane strain in the expression

$$\boldsymbol{f}_{\text{th}} = t_e \int_{A_e} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{\epsilon}_0 \, dA.$$

As an example, consider the problem shown in Fig. 4.5. The thickness, Young modulus and Poisson ratio are $t = 10$ mm, $E = 70 \times 10^9$ Pa and $\nu = 1/3$. The connectivity of the elements is

$$1 \rightarrow 1 \quad 3 \quad 2$$

$$2 \to 3 \quad 4 \quad 2.$$

The coordinates of the nodes 1-4 are $(0.0, 0.0)$, $(0.0, 0.03)$, $(0.06, 0.0)$ and $(0.06, 0.03)$. For element 1, we have $x_{13} = 0$, $y_{23} = -0.03$, $x_{23} = 0.06$, $y_{13} = -0.03$, $y_{12} = 0$, $x_{21} = 0.06$, $\det \boldsymbol{J} = 0.06(0.03) = 0.0018$. Thus,

$$\boldsymbol{B}^{(1)} = \frac{1}{0.0018} \begin{bmatrix} -0.03 & 0 & 0.03 & 0 & 0 & 0 \\ 0 & -0.06 & 0 & 0 & 0 & 0.06 \\ -0.06 & -0.03 & 0 & 0.03 & 0.06 & 0 \end{bmatrix}.$$

For element 2, we have $x_{13} = 0.06$, $y_{23} = 0$, $x_{23} = 0.06$, $y_{13} = -0.03$, $y_{12} = -0.02$, $x_{21} = 0$, $\det \boldsymbol{J} = 0.06(0.03) = 0.0018$, yielding

$$\boldsymbol{B}^{(2)} = \frac{1}{0.0018} \begin{bmatrix} 0 & 0 & 0.03 & 0 & -0.03 & 0 \\ 0 & -0.06 & 0 & 0.06 & 0 & 0 \\ -0.06 & 0 & 0.06 & 0.03 & 0 & -0.03 \end{bmatrix}.$$

The elasticity matrix is given by

$$\boldsymbol{C} = \frac{9 \times 70 \times 10^9}{8} \begin{bmatrix} 1 & 1/3 & 0 \\ 1/3 & 1 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

The matrices $\boldsymbol{B}^t\boldsymbol{C}\boldsymbol{B}$ for the two elements are given by

$$[\boldsymbol{B}^t\boldsymbol{C}\boldsymbol{B}]_1 = 2.43 \times 10^{13} \begin{bmatrix} 2.1 & 1.2003 & -0.9 & -0.6 & -1.2 & -0.6 \\ 1.2003 & 3.9 & -0.6 & -0.3 & -0.6 & -3.6 \\ -0.9 & -0.6 & 0.9 & 0 & 0 & 0.6 \\ -0.6 & -0.3 & 0 & 0.3 & 0.6 & 0 \\ -1.2 & -0.6 & 0 & 0.6 & 1.2 & 0 \\ -0.6 & -3.6 & 0.6 & 0 & 0 & 3.6 \end{bmatrix},$$

$$[\boldsymbol{B}^t\boldsymbol{C}\boldsymbol{B}]_2 = 2.43 \times 10^{13} \begin{bmatrix} 1.2 & 0 & -1.2 & -0.6 & 0 & 0.6 \\ 0 & 3.6 & -0.6 & -3.6 & 0.6 & 0 \\ -1.2 & -0.6 & 2.1 & 1.2 & -0.9 & -0.6 \\ -0.6 & -3.6 & 1.2 & 3.9 & -0.6 & -0.3 \\ 0 & 0.6 & -0.9 & -0.6 & 0.9 & 0 \\ 0.6 & 0 & -0.6 & -0.3 & 0 & 0.3 \end{bmatrix}.$$

The global degrees of freedom associated with the two elements are $(1, 2, 5, 6, 3, 4)$ and $(5, 6, 7, 8, 3, 4)$. Assembling the element level matrices into the global stiffness matrix and incorporating the boundary conditions, we get

$$9 \times 2.43 \times 10^7 \begin{bmatrix} 2.1 & 0 & -1.2 & -0.6 \\ 0 & 3.9 & -0.6 & -3.6 \\ -1.2 & -0.6 & 2.1 & 1.2 \\ -0.6 & -3.6 & 1.2 & 3.9 \end{bmatrix} \begin{bmatrix} \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -10000 \end{bmatrix}.$$

Solving the above set of equations, we get $\hat{u}_5 = -2.1 \times 10^{-5}$ m, $\hat{u}_6 = -1.208 \times 10^{-4}$ m, $\hat{u}_7 = 3.123 \times 10^{-5}$ m, $\hat{u}_8 = -1.361 \times 10^{-4}$ m, and $R_1 = 20000.0$ N, $R_2 = 10690.0$ N, $R_3 = -20000.0$ N, $R_4 = -690.42$ N. Note that $R_1 + R_3 \approx 0$ and $R_2 + R_4 \approx 10000$ N as expected.

The stresses are given by

$$\boldsymbol{\tau}^{(1)} = \boldsymbol{C}\boldsymbol{B}^{(1)} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_3 \\ \hat{u}_4 \end{bmatrix} = -4.375 \times 10^6 \begin{bmatrix} 6.3 \\ 2.1 \\ 12.08 \end{bmatrix},$$

$$\boldsymbol{\tau}^{(2)} = \boldsymbol{C}\boldsymbol{B}^{(2)} \begin{bmatrix} \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_3 \\ \hat{u}_4 \end{bmatrix} = -4.375 \times 10^6 \begin{bmatrix} -6.3 \\ 6.07 \\ 3.167 \end{bmatrix}.$$

4.2 The Four-Node Quadrilateral Element (Bilinear Element)

An easy way to ensure compatibility of the field variable across the element edges and also simplify computations of the element matrices is to map the element into a master element as shown in Fig. 4.6. The shape functions are now formulated in terms of ξ and η. As an example consider finding the expression for N_1. N_1 should have a value of one at node 1 and zero at the other nodes. We assume an expression for N_1 of the form

$$N_1 = c(1 - \xi)(1 - \eta),$$

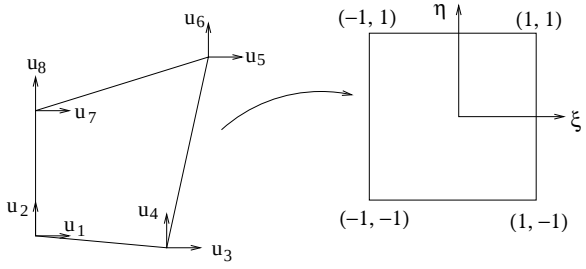


Figure 4.6: The bilinear element: Mapping the actual element into a master element

which ensures that it is zero on the edges 2-3 and 3-4. The constant c is determined by using the fact that N_1 is one at $(\xi, \eta) = (-1, -1)$. A similar procedure is followed for finding N_2, N_3 and N_4. We get

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta),$$
$$N_2 = \frac{1}{4}(1+\xi)(1-\eta),$$
$$N_3 = \frac{1}{4}(1+\xi)(1+\eta),$$
$$N_4 = \frac{1}{4}(1-\xi)(1+\eta).$$

A general way of representing the above functions is

$$N_i = \frac{1}{4}(1+\xi\xi_i)(1+\eta\eta_i),$$

where (ξ_i, η_i) are the coordinates of node i.

The displacement field can be written as

$$u = N_1 u_1 + N_2 u_3 + N_3 u_5 + N_4 u_7,$$
$$v = N_1 u_2 + N_2 u_4 + N_3 u_6 + N_4 u_8,$$

or alternatively in matrix form as $\boldsymbol{u} = \boldsymbol{N}\hat{\boldsymbol{u}}$, where

$$\boldsymbol{N} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 \end{bmatrix}.$$

In an isoparametric formulation, the geometry is also interpolated using the same shape functions, i.e.,

$$x = N_1 x_1 + N_2 x_2 + N_3 x_3 + N_4 x_4,$$
$$v = N_1 y_1 + N_2 y_2 + N_3 y_3 + N_4 y_4.$$

By using the chain rule, we get

$$\begin{bmatrix} \frac{\partial f}{\partial \xi} \\ \frac{\partial f}{\partial \eta} \end{bmatrix} = \boldsymbol{J} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix},$$

where

$$\boldsymbol{J} \equiv \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix}$$

$$= \frac{1}{4} \begin{bmatrix} -(1-\eta) & (1-\eta) & (1+\eta) & -(1+\eta) \\ -(1-\xi) & -(1+\xi) & (1+\xi) & (1-\xi) \end{bmatrix} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \xi} \\ \frac{\partial f}{\partial \eta} \end{bmatrix},$$

where

$$\Gamma_{11} = \frac{J_{22}}{|\boldsymbol{J}|}; \quad \Gamma_{12} = -\frac{J_{12}}{|\boldsymbol{J}|}; \quad \Gamma_{21} = -\frac{J_{21}}{|\boldsymbol{J}|}; \quad \Gamma_{22} = \frac{J_{11}}{|\boldsymbol{J}|},$$

where $|\boldsymbol{J}| \equiv \det \boldsymbol{J}$.

The strain-displacement matrix is found as follows:
We have (using the comma notation for denoting partial derivatives)

$$\boldsymbol{\epsilon}_c = \begin{bmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \gamma_{xy} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} u_{,x} \\ u_{,y} \\ v_{,x} \\ v_{,y} \end{bmatrix} = \boldsymbol{R}_1 \begin{bmatrix} u_{,x} \\ u_{,y} \\ v_{,x} \\ v_{,y} \end{bmatrix}. \tag{4.6}$$

Next, we have

$$\begin{bmatrix} u_{,x} \\ u_{,y} \\ v_{,x} \\ v_{,y} \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & 0 & 0 \\ \Gamma_{21} & \Gamma_{22} & 0 & 0 \\ 0 & 0 & \Gamma_{11} & \Gamma_{12} \\ 0 & 0 & \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} u_{,\xi} \\ u_{,\eta} \\ v_{,\xi} \\ v_{,\eta} \end{bmatrix} = \boldsymbol{R}_2 \begin{bmatrix} u_{,\xi} \\ u_{,\eta} \\ v_{,\xi} \\ v_{,\eta} \end{bmatrix}. \tag{4.7}$$

Finally, we have

$$\begin{bmatrix} u_{,\xi} \\ u_{,\eta} \\ v_{,\xi} \\ v_{,\eta} \end{bmatrix} = \begin{bmatrix} N_{1,\xi} & 0 & N_{2,\xi} & 0 & N_{3,\xi} & 0 & N_{4,\xi} & 0 \\ N_{1,\eta} & 0 & N_{2,\eta} & 0 & N_{3,\eta} & 0 & N_{4,\eta} & 0 \\ 0 & N_{1,\xi} & 0 & N_{2,\xi} & 0 & N_{3,\xi} & 0 & N_{4,\xi} \\ 0 & N_{1,\eta} & 0 & N_{2,\eta} & 0 & N_{3,\eta} & 0 & N_{4,\eta} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \end{bmatrix}$$

$$= \frac{1}{4} \begin{bmatrix} -(1-\eta) & 0 & (1-\eta) & 0 & (1+\eta) & 0 & -(1+\eta) & 0 \\ -(1-\xi) & 0 & -(1+\xi) & 0 & (1+\xi) & 0 & (1-\xi) & 0 \\ 0 & -(1-\eta) & 0 & (1-\eta) & 0 & (1+\eta) & 0 & -(1+\eta) \\ 0 & -(1-\xi) & 0 & -(1+\xi) & 0 & (1+\xi) & 0 & (1-\xi) \end{bmatrix} \hat{\boldsymbol{u}}$$

$$= \boldsymbol{R}_3 \hat{\boldsymbol{u}}. \qquad (4.8)$$

From Eqns. 4.6, 4.7 and 4.8, we get

$$\boldsymbol{\epsilon}_c = \boldsymbol{R}_1 \boldsymbol{R}_2 \boldsymbol{R}_3 \hat{\boldsymbol{u}},$$

thus yielding

$$\boldsymbol{B} = \boldsymbol{R}_1 \boldsymbol{R}_2 \boldsymbol{R}_3.$$

The stiffness matrix is given by

$$\boldsymbol{K} = \int_\Omega \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, d\Omega = \int_{-1}^{1} \int_{-1}^{1} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} t \, |\boldsymbol{J}| \, d\xi \, d\eta.$$

Since the above integral is complicated to evaluate in closed form, we use a numerical technique such as Gaussian quadrature (see Section 4.4).

The consistent load vector is given by

$$\boldsymbol{f}^{(e)} = \int_{\Gamma_t} \boldsymbol{N}^t \bar{\boldsymbol{t}} \, d\Gamma + \int_{-1}^{1} \int_{-1}^{1} \rho \boldsymbol{N}^t \boldsymbol{b} t \, |\boldsymbol{J}| \, d\xi \, d\eta.$$

The body force term is most easily computed using Gaussian quadrature. As an example, consider the computation of the consistent load vector due to a uniform traction on the edge 2-3 in Fig. 4.6. Since $x = N_2 x_2 + N_3 x_3$ and $y = N_2 y_2 + N_3 y_3$ with $N_2 = (1-\eta)/2$ and $N_3 = (1+\eta)/2$ along this edge, we have $dx = (x_3 - x_2)\, d\eta/2$ and $dy = (y_3 - y_2)\, d\eta/2$ leading to

$$d\Gamma = t \, ds = t\sqrt{dx^2 + dy^2} = \frac{t l_{23}}{2} \, d\eta,$$

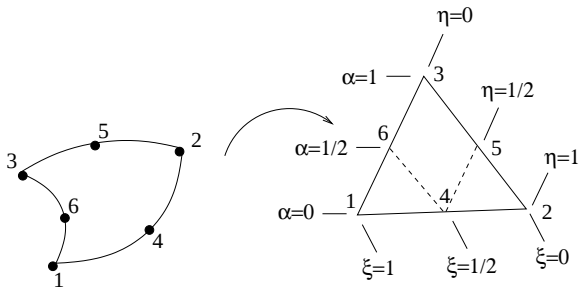


Figure 4.7: Six-node triangular element

where $l_{23} = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}$ is the length of the side 2-3. Hence,

$$\boldsymbol{f}^{(e)} = \int_{-1}^{1} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1-\eta}{2} & 0 \\ 0 & \frac{1-\eta}{2} \\ \frac{1+\eta}{2} & 0 \\ 0 & \frac{1+\eta}{2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix} t\frac{l_{23}}{2}\, d\eta,$$

$$= \frac{tl_{23}}{2} \begin{bmatrix} 0 & 0 & t_x & t_y & t_x & t_y & 0 & 0 \end{bmatrix}^t,$$

which agrees with the intuitively expected answer.

4.3 Higher-Order Elements

We now consider higher-order elements such as the 6-node triangle (quadratic triangle), and the 8-node and 9-node quadrilateral elements. Note that all these elements can model curved boundaries more accurately than the linear elements.

4.3.1 Six-node triangular element

The master element is as shown in Fig. 4.7. The shape functions are

$$N_1 = \xi(2\xi - 1), \qquad N_4 = 4\xi\eta,$$
$$N_2 = \eta(2\eta - 1), \qquad N_5 = 4\alpha\eta,$$
$$N_3 = \alpha(2\alpha - 1), \qquad N_6 = 4\xi\alpha,$$

where $\alpha = (1 - \xi - \eta)$. N_1 is obtained by assuming it to be of the form $c\xi(2\xi - 1)$ (so that it is zero on the lines $\xi = 0$ and $\xi = 1/2$) and then finding c so that $N_1 = 1$ at $\xi = 1$. The other shape functions are obtained similarly.

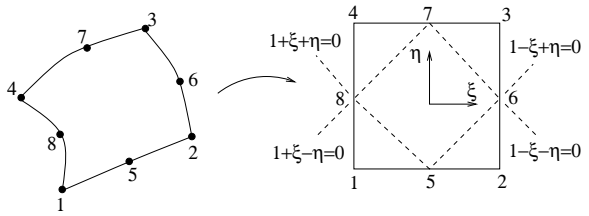


Figure 4.8: Eight-node quadrilateral element

The displacements are interpolated using $\boldsymbol{u} = \sum_{i=1}^{6} N_i \hat{u}_i$. The geometry is interpolated using the same shape functions in an isoparametric formulation. The $\boldsymbol{B}$ matrix is found in the same way as for the four-node quadrilateral element. The stiffness matrix is given by

$$\boldsymbol{K}^{(e)} = \int_0^1 \int_0^{1-\eta} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} t \, |\boldsymbol{J}| \; d\xi \, d\eta.$$

4.3.2 Eight-node quadrilateral (serendipity) element

The mapping is as shown in Fig. 4.8. The shape function N_1 is found by assuming it to be of the form

$$c(1-\xi)(1-\eta)(1+\xi+\eta),$$

so that it is zero at all nodes except node 1. Since $N_1 = 1$ at $(-1,-1)$, we get $c = -1/4$. Similarly, N_5 is found assuming it to be of the form

$$k(1-\xi)(1-\eta)(1+\xi),$$

and then determining k. The final expressions for the shape functions are

$$\begin{aligned}
N_1 &= -\frac{1}{4}(1-\xi)(1-\eta)(1+\xi+\eta), & N_5 &= \frac{1}{2}(1-\xi^2)(1-\eta),\\
N_2 &= -\frac{1}{4}(1+\xi)(1-\eta)(1-\xi+\eta), & N_6 &= \frac{1}{2}(1+\xi)(1-\eta^2),\\
N_3 &= -\frac{1}{4}(1+\xi)(1+\eta)(1-\xi-\eta), & N_7 &= \frac{1}{2}(1-\xi^2)(1+\eta),\\
N_4 &= -\frac{1}{4}(1-\xi)(1+\eta)(1+\xi-\eta), & N_8 &= \frac{1}{2}(1-\xi)(1-\eta^2).
\end{aligned}$$

The stiffness matrix is given by

$$\boldsymbol{K}^{(e)} = \int_{-1}^1 \int_{-1}^1 \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} t \, |\boldsymbol{J}| \; d\xi \, d\eta,$$

where $\boldsymbol{B}$ is formulated in the same way as for the four-node quadrilateral element.

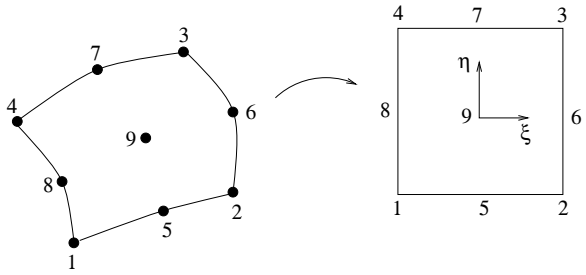


Figure 4.9: Nine-node quadrilateral element

4.3.3 Nine-node quadrilateral element

The shape functions for the nine node-quadrilateral element shown in Fig. 4.9 are found by using products of quadratic shape functions in one-dimension along the ξ and η directions. The shape functions are given by

$$
\begin{aligned}
N_1 &= L_1(\xi)L_1(\eta), & N_5 &= L_2(\xi)L_1(\eta),\\
N_2 &= L_3(\xi)L_1(\eta), & N_6 &= L_3(\xi)L_2(\eta),\\
N_3 &= L_3(\xi)L_3(\eta), & N_7 &= L_2(\xi)L_3(\eta),\\
N_4 &= L_1(\xi)L_3(\eta), & N_8 &= L_1(\xi)L_2(\eta),\\
N_9 &= L_2(\xi)L_2(\eta), &&
\end{aligned}
$$

where

$$
\begin{aligned}
L_1(\xi) &= -\frac{1}{2}\xi(1-\xi), & L_1(\eta) &= -\frac{1}{2}\eta(1-\eta),\\
L_2(\xi) &= (1+\xi)(1-\xi), & L_2(\eta) &= (1+\eta)(1-\eta),\\
L_3(\xi) &= \frac{1}{2}\xi(1+\xi), & L_3(\eta) &= \frac{1}{2}\eta(1+\eta).
\end{aligned}
$$

The stiffness matrix is formed in the same way as for the eight-node quadrilateral element.

4.3.4 Work-equivalent loads for the eight-node and nine-node quadrilateral elements

As an example of the computation of work-equivalent loads, consider that a linearly varying pressure is applied to edge 2-3 in Fig. 4.8 (or Fig. 4.9; the same work-equivalent loads are obtained in either case). Assume that node 6 is located midway between nodes 2 and 3, i.e., $x_6 = (x_2 + x_3)/2$ $y_6 = (y_2 + y_3)/2$. Let the magnitude of the pressure be p_i at node 2 and p_j at node 3. Since edge 2-3 is mapped into the edge $\xi = 1$, we have

$$\bar{t} = \frac{1}{2}(1-\eta)p_i + \frac{1}{2}(1+\eta)p_j.$$

Also

$$N_2|_{\xi=1} = -\frac{1}{2}\eta(1-\eta),$$

$$N_3|_{\xi=1} = \frac{1}{2}\eta(1+\eta),$$
$$N_6|_{\xi=1} = 1-\eta^2.$$

The element of area is given by

$$d\Gamma = t\,ds = t\sqrt{dx^2+dy^2} = \frac{tl_{23}}{2}\,d\eta.$$

The elements of the consistent load vector corresponding to the degrees of freedom associated with edge 2-3 are

$$\begin{aligned}
\boldsymbol{f} &= \int \boldsymbol{N}^t \bar{\boldsymbol{t}}\, d\Gamma \\
&= \frac{tl_{23}}{8}\int_{-1}^{1} \begin{bmatrix} -\eta(1-\eta)[p_i+p_j-\eta(p_i-p_j)] \\ 0 \\ 2(1-\eta^2)[p_i+p_j-\eta(p_i-p_j)] \\ 0 \\ \eta(1+\eta)[p_i+p_j-\eta(p_i-p_j)] \\ 0 \end{bmatrix} d\eta \\
&= \frac{tl_{23}}{8} \begin{bmatrix} \frac{4p_i}{3} \\ 0 \\ \frac{8(p_i+p_j)}{3} \\ 0 \\ \frac{4p_j}{3} \\ 0 \end{bmatrix} \\
&= t \begin{bmatrix} \frac{p_i l_{23}}{6} \\ 0 \\ \frac{2(p_i+p_j)l_{23}}{6} \\ 0 \\ \frac{p_j l_{23}}{6} \end{bmatrix}
\end{aligned}$$

In the special case when $p_i = p_j = p$, the total load $P = ptl_{23}$ is distributed as $(P/6, 4P/6, P/6)$ at the three nodes 2, 6 and 3, something which is not intuitively obvious.

4.4 Gauss Quadrature

Gauss quadrature is an approximate method of computing integrals. We first consider the gauss quadrature for one-dimensional integrals

4.4.1 One-dimensional problems

Consider an integral of the form

$$I = \int_{-1}^{1} f(\xi)\, d\xi.$$

We approximate it as

$$I = w_1 f(\xi_1) + w_2 f(\xi_2) + \ldots + w_n f(\xi_n), \tag{4.9}$$

where w_1, w_2,...,w_n are the weights, and ξ_1, ξ_2,..., ξ_n are the *sampling points* or *Gauss points*. We select the Gauss points and weights such that Eqn. 4.9 provides an exact answer for polynomials of as large a degree as possible. The idea is that if n-point formula is exact for polynomials of as high a degree as possible, then the formula will 'work well' even if $f(\xi)$ is not a polynomial.

One-point formula

We approximate the integral just using the first term, i.e.,

$$\int_{-1}^{1} f(\xi)\, d\xi \approx w_1 f(\xi_1).$$

The two parameters w_1 and ξ_1 are found such that the integral is exact when $f(\xi)$ is of the form $a_0 + a_1\xi$. Hence, we get

$$\int_{-1}^{1} (a_0 + a_1\xi)\, d\xi = w_1(a_0 + a_1\xi_1),$$

or, alternatively,

$$2a_0 = w_1 a_0 + w_1\xi_1 a_1.$$

Comparing the coefficients of a_0 and a_1, we get $w_1 = 2$ and $\xi_1 = 0$. Thus, the one-point approximation formula for any arbitrary $f(\xi)$ is

$$\int_{-1}^{1} f(\xi)\, d\xi \approx 2f(0).$$

Two-point formula:

Now we have the approximation

$$\int_{-1}^{1} f(\xi)\, d\xi = w_1 f(\xi_1) + w_2 f(\xi_2).$$

Since there are four parameters, w_1, ξ_1, w_2 and ξ_2, we can integrate a cubic polynomial exactly. Thus, we determine these four parameters such that

$$\int_{-1}^{1} (a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3)\, d\xi = w_1(a_0 + a_1\xi_1 + a_2\xi_1^2 + a_3\xi_1^3) + w_2(a_0 + a_1\xi_2 + a_2\xi_2^2 + a_3\xi_2^3).$$

Comparing the coefficients of a_0, a_1, a_2 and a_3, we get

$$\begin{aligned} w_1 + w_2 &= 2, \\ w_1\xi_1 + w_2\xi_2 &= 0, \\ w_1\xi_1^2 + w_2\xi_2^2 &= \frac{2}{3}, \\ w_1\xi_1^3 + w_2\xi_2^3 &= 0. \end{aligned}$$

Solving the above equations, we get $w_1 = w_2 = 1$, and $-\xi_1 = \xi_2 = 1/\sqrt{3}$. Thus, for any arbitrary function $f(\xi)$, we have

$$\int_{-1}^{1} f(\xi)\, d\xi \approx f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right).$$

In general, an n-point Gauss-quadrature formula will provide an exact answer if $f(\xi)$ is a polynomial of order $(2n-1)$ or less.

Sampling points and weights for Gaussian quadrature		
Order n	Location ξ_i	Weight w_i
1	0	2
2	$\pm\frac{1}{\sqrt{3}}$	1
3	$\pm\sqrt{0.6}, 0$	$\frac{5}{9}, \frac{5}{9}, \frac{8}{9}$
4	$\pm\sqrt{\frac{3+2\sqrt{1.2}}{7}}$ $\pm\sqrt{\frac{3-2\sqrt{1.2}}{7}}$	$\frac{1}{2} - \frac{1}{6\sqrt{1.2}}$ $\frac{1}{2} + \frac{1}{6\sqrt{1.2}}$

Note that the Gauss points are located symmetrically with respect to the origin,and that symmetrically placed points have the same weight.

Example:
Consider the approximate evaluation of

$$I = \int_{-1}^{1} \left[3e^x + x^2 + \frac{1}{x+2}\right] dx = 8.8165.$$

Using a one-point quadrature, we get $I = 2f(0) = 7.0$. Using a two-point quadrature, we get $I = f(-1/\sqrt{3}) + f(1/\sqrt{3}) = 8.7857$. Note that the result converges to the exact solution.

4.4.2 Two-dimensional Gauss quadrature

An integral with two independent variables such as

$$I = \int_{-1}^{1}\int_{-1}^{1} f(\xi, \eta)\, d\xi\, d\eta,$$

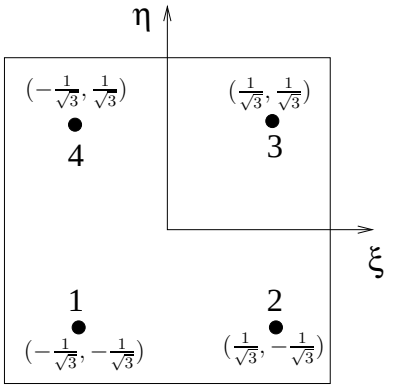


Figure 4.10: Gauss points for a 2×2 integration rule

is approximated as

$$\begin{aligned} I &= \int_{-1}^{1} \left[\sum_{i=1}^{n} w_i f(\xi_i, \eta) \right] d\eta \\ &= \sum_{j=1}^{n} w_j \left[\sum_{i=1}^{n} w_i f(\xi_i, \eta_j) \right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j f(\xi_i, \eta_j). \end{aligned}$$

As an example, consider the evaluation of the stiffness matrix given by

$$\boldsymbol{K}^{(e)} = \int_{-1}^{1} \int_{-1}^{1} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, |\boldsymbol{J}| \, t \, d\xi \, d\eta.$$

Let $\phi(\xi, \eta) = t \, |\boldsymbol{J}| \, (\boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B})_{ij}$. For a four-node quadrilateral element, if we use a 2×2 integration rule (see Fig. 4.10), we get

$$\begin{aligned} K_{ij} &= w_1^2 \phi(\xi_1, \eta_1) + w_1 w_2 \phi(\xi_1, \eta_2) + w_2 w_1 \phi(\xi_2, \eta_1) + w_2^2 \phi(\xi_2, \eta_2) \\ &= \phi(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}) + \phi(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}) + \phi(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}) + \phi(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}). \end{aligned}$$

4.4.3 Order of integration for quadrilateral elements

We first calculate the required integration order for a four-node rectangular element. Let the coordinates of the four nodes be (x_1, y_1), (x_2, y_1), (x_2, y_2) and (x_1, y_2). Then

$$\begin{aligned} x &= N_1 x_1 + N_2 x_2 + N_3 x_2 + N_4 x_1 \\ &= \frac{1}{2}(1 - \xi) x_1 + \frac{1}{2}(1 + \xi) x_2 \\ y &= \frac{1}{2}(1 - \eta) y_1 + \frac{1}{2}(1 + \eta) y_2. \end{aligned}$$

The Jacobian is

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(x_2 - x_1) & 0 \\ 0 & \frac{1}{2}(y_2 - y_1) \end{bmatrix} = \text{constant}.$$

Hence, the determinant of the Jacobian, $|\boldsymbol{J}|$ is also constant at all points in the element. Therefore,

$$\boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} \, |\boldsymbol{J}| = f(\xi^2, \eta^2, \xi\eta).$$

Since the highest order of the polynomial in one direction is 2, the order of Gaussian integration which is needed is 2×2 ($2n - 1 \geq 2 \implies n = 2$). In a similar way, the required integration order for evaluating the mass matrix or the load vector can be assessed.

If the element is distorted, i.e., it is not rectangular or a parralelogram, then the Jacobian matrix is *not* a constant over the element. In such a case, the integrand in the stiffness matrix is a ratio of polynomials which in general is not a polynomial. 'Full numerical integration' is the order of integration that gives exact matrices (i.e., same as analytical integration) when the elements are undistorted. Using this integration order for a distorted element will not yield the exactly integrated element matrices. However, the numerical integration errors are small if the geometric distortions are small. If the element is highly distorted (a situation which should be avoided as far as possible), one should use a higher order of quadrature than the one used in full integration. The recommended full gauss numerical integration orders for the 4, 8, 9 and 16 node isoparametric quadrilateral elements are 2×2, 3×3, 3×3 and 4×4, respectively.

If an integration order lower than the 'full' order is used it is called 'reduced' integration. Using reduced integration is, in general, unreliable. It can result in an instability known variously as a mechanism, kinematic mode, hourglass mode or zero-energy mode. For example, the nodal displacement in the setup shown in Fig. 4.11 where 2×2 integration is used to compute the stiffness matrix are very large due to the zero-energy mode. An eigenvalue analysis of the element stiffness matrices for a plane [solid] element should yield three [six] zero eigenvalues corresponding to the three [six] rigid body modes (2 translations and a rotation [3 translations and 3 rotations]). However, in the problem shown above, the element stiffness computed using a 2×2 integration rule has one spurious zero-energy mode, i.e., four zero eigenvalues, due to the use of reduced integration.

4.4.4 Numerical integration for triangular elements

Let $\phi(\xi, \eta, \alpha)$ be a function to be integrated. Then

$$\int \phi \, |\boldsymbol{J}| \; d\xi \, d\eta \approx \frac{1}{2} \sum_{i=1}^{n} w_i \, |J|_i \, \phi_i, \tag{4.10}$$

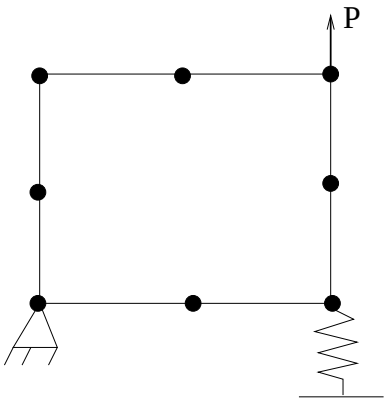


Figure 4.11: An example of zero-energy modes due to reduced integration

where w_i, ϕ_i and $|\boldsymbol{J}|_i$ are the values of the weight, function and determinant of the Jacobian at the sampling point. The factor 1/2 occurs in the formula because for an undistorted triangle of unit area, $|\boldsymbol{J}| = 2$ throughout the triangle. Since $\sum w_i = 1$, we get $\int dA = 1$ when $\phi = 1$. The area coordinates and the weights for various orders of integration are presented in the following table.

Area coordinates and weights for triangular elements

Order n	Area coordinates	Weight w_i
1	$\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$	1
2	$\frac{2}{3}, \frac{1}{6}, \frac{1}{6}$	$\frac{1}{3}$
	$\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$	$\frac{1}{3}$
	$\frac{1}{6}, \frac{1}{6}, \frac{2}{3}$	$\frac{1}{3}$
2	$\frac{1}{2}, \frac{1}{2}, 0$	$\frac{1}{3}$
	$0, \frac{1}{2}, \frac{1}{2}$	$\frac{1}{3}$
	$\frac{1}{2}, 0, \frac{1}{2}$	$\frac{1}{3}$
3	$\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$	-0.5625
	$\frac{3}{5}, \frac{1}{5}, \frac{1}{5}$	0.5208333..
	$\frac{1}{5}, \frac{3}{5}, \frac{1}{5}$	0.5208333..
	$\frac{1}{5}, \frac{1}{5}, \frac{3}{5}$	0.5208333..

(4.11)

The location of the Gauss points and the associated weights in Table (4.11) are found as follows. If $f(\xi, \eta) = a_0 + a_1\xi + a_2\eta$, then using Eqn. (4.10), we get

$$a_0 + \frac{1}{3}(a_1 + a_2) = w_1(a_0 + a_1\xi_1 + a_2\eta_1),$$

which yields $w_1 = 1$ and $\xi_1 = \eta_1 = 1/3$. If $f(\xi, \eta) = a_0 + a_1\xi + a_2\eta + a_3\xi^2 +$

$a_4\xi\eta + a_5\eta^2$, then using Eqn. (4.10), we get

$$a_0+\frac{1}{3}(a_1+a_2)+\frac{1}{6}(a_3+a_5)+\frac{1}{12}a_4 = w_1(a_0+a_1\xi_1+a_2\eta_1+a_3\xi_1^2+a_4\xi_1\eta_1+a_5\eta_1^2) + w_2(a_0 + a_1\xi_2 + a_2\eta_2 + a_3\xi_2^2 + a_4\xi_2\eta_2 + a_5\eta_2^2).$$

However, the Gauss points are *not* symmetrically located! If we choose a 3-point quadrature rule, i.e.,

$$a_0+\frac{1}{3}(a_1+a_2)+\frac{1}{6}(a_3+a_5)+\frac{1}{12}a_4 = w_1(a_0+a_1\xi_1+a_2\eta_1+a_3\xi_1^2+a_4\xi_1\eta_1+a_5\eta_1^2) + w_2(a_0+a_1\xi_2+a_2\eta_2+a_3\xi_2^2+a_4\xi_2\eta_2+a_5\eta_2^2)+w_3(a_0+a_1\xi_3+a_2\eta_3+a_3\xi_3^2+a_4\xi_3\eta_3+a_5\eta_3^2),$$

then we have more unknowns than equations. The Gauss points can now be chosen to be symmetrically located, but as seen in Table (4.11), the choice is not unique.

We have seen that for a three-node triangle, the stiffness matrix can be constructed explicitly, and there is no need of numerical integration. For a quadratic triangle,

$$\boldsymbol{B}^t\boldsymbol{C}\boldsymbol{B}\,|\boldsymbol{J}| = f(\xi^2, \xi\eta, \eta^2).$$

Hence, a three point integration is exact when the triangle is an equilateral triangle. As in the case of the quadrilateral elements, the integration formula is approximate when the triangular element is distorted.

4.4.5 Stress Computation

If stresses are directly computed using

$$\boldsymbol{\tau}_c = \boldsymbol{C}(\boldsymbol{B}\hat{\boldsymbol{u}} - \boldsymbol{\epsilon}_c^0) + \boldsymbol{\tau}_c^0,$$

then the results obtained can be quite inaccurate since the process of differentiation results in a loss of accuracy. For isoparametric elements, the stresses are most accurate at Gauss points of a quadrature rule one order lower than that required for full integration of the element stiffness. Such points are known as Barlow points.

Consider the examples shown in Fig. 4.12. In example (a), the displacements are of the form $u = -a\xi\eta$, and $v = 0$, where a is a positive constant. Thus, the shear strain γ_{xy} is proportional to $(-\xi)$ yielding the variation shown in the figure. Since the exact shear strain is zero, we see that the finite element solution matches the exact solution at $\xi = 0$ and $\eta = 0$, which is also the Gauss point location for a order-1 integration rule. Similarly, as shown in Fig. 4.12b, for an 8-node (or even 9-node) element, the Barlow points (or 'superconvergent' points) are the Gauss-point locations for the 2×2 integration rule ($\xi = \eta = \pm 1/\sqrt{3}$). These conclusions are true for rectangular elements. For distorted elements, Gauss points may not be optimal locations, but, nonetheless, they remain good choices.

To obtain more accurate stresses than those obtained from direct computation, the strategy used is to extrapolate the stress values from the Gauss

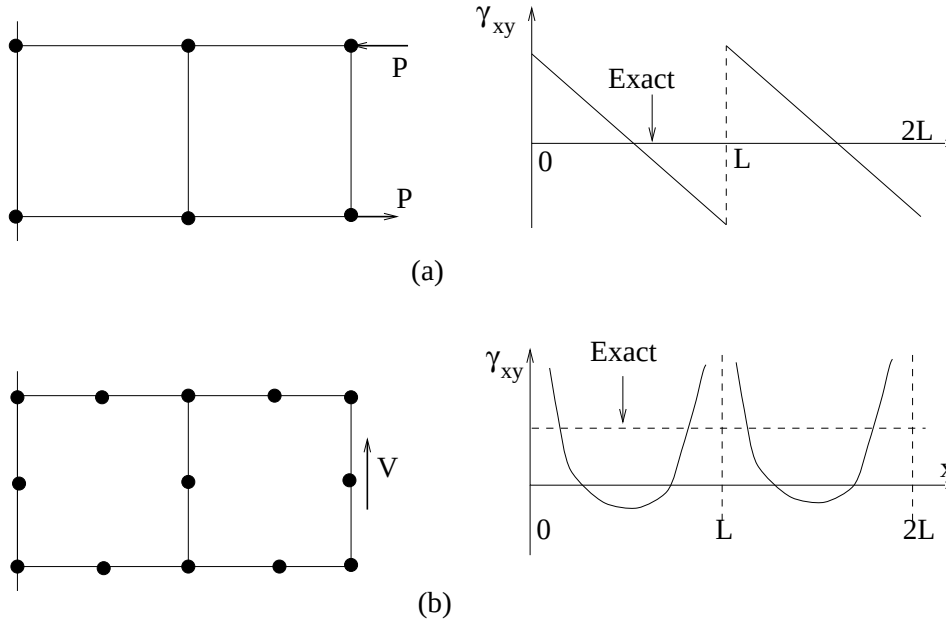


Figure 4.12: Stress variation within an element

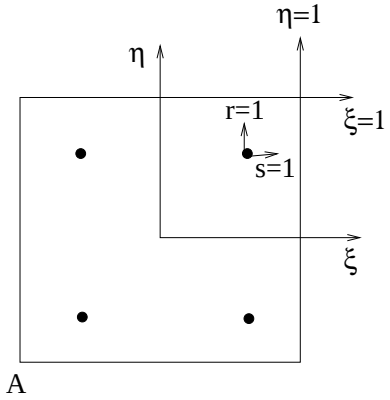


Figure 4.13: Extrapolation of stress

point to the nodes, and then average the values obtained from all the elements sharing that node to obtain the nodal stresses. An example of an extrapolation strategy is shown in Fig. 4.13. Since the element shown is a four-node element, the 2×2 Gauss points are located at $\xi = \eta = \pm 1/\sqrt{3}$. Let r and s be a scaled version of the ξ-η system such that they have the same origin but with $r = s = 1$ when $\xi = \eta = 1/\sqrt{3}$. In other words, let $r = \sqrt{3}\xi$ and $s = \sqrt{3}\eta$. The extrapolated stress components are given by (τ is a typical stress component)

$$\tau = \sum_{i=1}^{4} \bar{N}_i \sigma_i,$$

where σ_i are the values of the stress components at the Gauss points, and

$$\bar{N}_i = \frac{1}{4}(1 \pm r)(1 \pm s).$$

Thus, the stress components at point A in Fig. 4.13 are obtained by substituting $r = s = -\sqrt{3}$ in the above formula. The four values obtained at each node are averaged, and this mean value is taken as the nodal value. The improved stress field is given by

$$\tau = \sum_{i=1}^{n} N_i (\tau_{\text{mean}})_i,$$

where n is the number of nodes, and N_i are the displacement shape functions.

4.5 Heat transfer Problems

The procedure for formulating the various matrices is similar to that for elasticity. We shall just consider the details for the CST element. The temperature is interpolated as

$$T = N_1 T_1 + N_2 T_2 + N_3 T_3,$$

where $N_1 = \xi$, $N_2 = \eta$ and $N_3 = 1 - \xi - \eta$. Thus, the temperature gradient is

$$\begin{bmatrix} \frac{\partial T}{\partial \xi} \\ \frac{\partial T}{\partial \eta} \end{bmatrix} = \boldsymbol{J} \begin{bmatrix} \frac{\partial T}{\partial x} \\ \frac{\partial T}{\partial y} \end{bmatrix},$$

where (with $x_{ij} \equiv x_i - x_j$)

$$\boldsymbol{J} = \begin{bmatrix} x_{13} & y_{13} \\ x_{23} & y_{23} \end{bmatrix}.$$

Inverting the above relation, we get

$$\begin{aligned} \begin{bmatrix} \frac{\partial T}{\partial x} \\ \frac{\partial T}{\partial y} \end{bmatrix} &= \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23} & -y_{13} \\ -x_{23} & x_{13} \end{bmatrix} \begin{bmatrix} \frac{\partial T}{\partial \xi} \\ \frac{\partial T}{\partial \eta} \end{bmatrix} \\ &= \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23} & -y_{13} \\ -x_{23} & x_{13} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} \\ &= \boldsymbol{B} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}, \end{aligned}$$

where

$$\boldsymbol{B} = \frac{1}{\det \boldsymbol{J}} \begin{bmatrix} y_{23} & y_{31} & y_{12} \\ x_{32} & x_{13} & x_{21} \end{bmatrix}.$$

Assuming that $\boldsymbol{K}$ is constant and isotropic over each element, i.e., $\boldsymbol{K} = k\boldsymbol{I}$, the volume term of the stiffness matrix is

$$\boldsymbol{K}_1^{(e)} = k_e \int \boldsymbol{B}^t \boldsymbol{B} t \, dA = t k_e A_e \boldsymbol{B}^t \boldsymbol{B}.$$

Assuming that convection takes place across edge 2-3 ($\xi = 0$) of the element, the shape function matrix can be written as

$$\boldsymbol{N} = \begin{bmatrix} 0 & \eta & 1-\eta \end{bmatrix}.$$

The surface area element is given by $t\, ds = t l_{23}\, d\eta$. Hence, the surface term of the stiffness matrix is

$$\begin{aligned} \boldsymbol{K}_2^{(e)} &= \int_0^1 h \boldsymbol{N}^t \boldsymbol{N} t l_{23} \, d\eta \\ &= \frac{h t l_{23}}{6} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}. \end{aligned}$$

The load term due to the heat source Q is

$$\boldsymbol{f}^{(1)} = \int \boldsymbol{N}^t \boldsymbol{Q} t \, dA.$$

We get

$$\begin{aligned} \boldsymbol{f}^{(1)} &= \frac{t Q_e A_e}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{if } Q \text{ is constant,} \\ &= \frac{t A_e}{12} \begin{bmatrix} 2Q_1 + Q_2 + Q_3 \\ Q_1 + 2Q_2 + Q_3 \\ Q_1 + Q_2 + 2Q_3 \end{bmatrix} \quad \text{if } Q = N_1 Q_1 + N_2 Q_2 + N_3 Q_3, \\ &= t Q_0 N_0(\xi_0, \eta_0) = t Q_0 \begin{bmatrix} \xi_0 \\ \eta_0 \\ 1 - \xi_0 - \eta_0 \end{bmatrix} \quad \text{if } Q \text{ is a point source.} \end{aligned}$$

Assuming that the edge 2-3 constitutes part of Γ_q, and assuming that $(\bar{q} + hT_\infty) =$ constant, the load term due to the prescribed heat flux and convection is

$$\boldsymbol{f}^{(2)} = \int_0^1 \boldsymbol{N}^t (\bar{q} + hT_\infty) t l_{23} \, d\eta$$

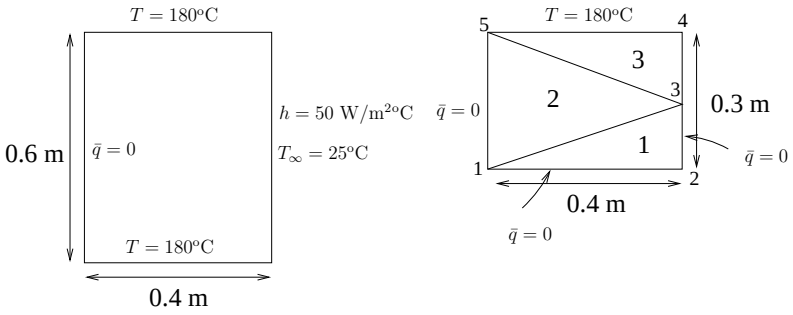


Figure 4.14: Example problem

$$= \frac{1}{2}(\bar{q} + hT_\infty)tl_{23} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

As an example, consider the problem shown in Fig. 4.14 where the domain's conductivity is $\boldsymbol{K} = k\boldsymbol{I}$, where $k = 1.5$ W/m °C. Due to the symmetry of the domain and boundary conditions, we can model only the upper half of the domain as shown. Due to symmetry, the heat flux at the bottom edge of the finite element analysis domain is $\bar{q} = \frac{\partial T}{\partial n} = 0$. On the right edge, since there is no prescribed flux, we have $\frac{\partial T}{\partial n} = -h(T - T_\infty)$. The connectivity is given by

$$\begin{array}{llll} 1 \to 1 & 2 & 3 \\ 2 \to 5 & 1 & 3 \\ 3 \to 5 & 3 & 4 \end{array}$$

The $\boldsymbol{B}$ matrices are

$$\boldsymbol{B}^{(1)} = \frac{1}{0.06} \begin{bmatrix} -0.15 & 0.15 & 0 \\ 0 & -0.4 & 0.4 \end{bmatrix},$$

$$\boldsymbol{B}^{(2)} = \frac{1}{0.12} \begin{bmatrix} -0.15 & -0.15 & 0.3 \\ 0.4 & -0.4 & 0 \end{bmatrix},$$

$$\boldsymbol{B}^{(3)} = \frac{1}{0.06} \begin{bmatrix} -0.15 & 0 & 0.15 \\ 0 & -0.4 & 0.4 \end{bmatrix}.$$

The stiffness matrix due to the conductivity part is

$$\boldsymbol{K}_T^{(e)} = kA\boldsymbol{B}^t\boldsymbol{B},$$

while that due to the convection part (assuming that the convection is taking

place at edge 2-3 of the element) is

$$\boldsymbol{K}_h^{(e)} = \frac{hl_{23}}{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix},$$

The global stiffness matrix is

$$\boldsymbol{K} = \begin{bmatrix} 1.42125 & -0.28125 & -0.28125 & 0 & -0.86 \\ -0.28125 & 4.78125 & -0.75 & 0 & 0 \\ -0.28125 & -0.75 & 9.5625 & -0.75 & -0.28125 \\ 0 & 0 & -0.75 & 4.78125 & -0.28125 \\ -0.86 & 0 & -0.28125 & -0.28125 & 1.42125 \end{bmatrix}.$$

The load vector due to convection is

$$\boldsymbol{f}_h^{(e)} = \frac{hT_\infty l_{23}}{2}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$$

The global load vector is

$$\boldsymbol{f} = 93.75\begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

The prescribed degrees of freedom are $T_4 = T_5 = 180$. Solving for T_1, T_2 and T_3 we get

$$\begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix} = \begin{bmatrix} 124.5 & 34.0 & 45.4 \end{bmatrix}.$$

Note that a large number of nodes are required to capture accurately the large temperature gradient along the edge 2-4 ($T_2 = 34$, $T_4 = 180$).

4.6 Axisymmetric Problems

We now consider problems with axisymmetric geometry and axisymmetric loading. Such problems can be treated as two-dimensional problems since there is no dependence of any of the variables on θ. We denote the components (u_r, u_θ, u_z) by (u, v, w). Because of the symmetry about the z-axis,

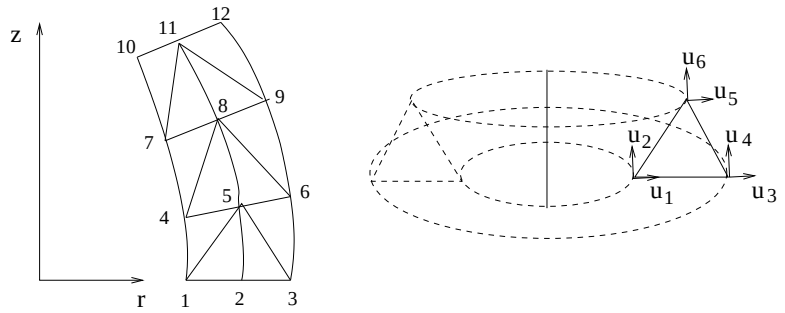


Figure 4.15: Meshing for an axisymmetric problem

$v = 0$. The engineering strains are

$$\boldsymbol{\epsilon}_c = \begin{bmatrix} \epsilon_{rr} \\ \epsilon_{zz} \\ \gamma_{rz} \\ \epsilon_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial w}{\partial z} \\ \frac{\partial u}{\partial z} + \frac{\partial w}{\partial r} \\ \frac{u}{r} \end{bmatrix}.$$

If $\boldsymbol{\tau} \equiv [\tau_{rr}, \tau_{zz}, \tau_{rz}, \tau_{\theta\theta}]^t$, the stress-strain relations are

$$\boldsymbol{\tau}_c = \boldsymbol{C}(\boldsymbol{\epsilon}_c - \boldsymbol{\epsilon}_c^0),$$

where (with $f = \nu/(1-\nu)$ and $g = (1-2\nu)/(2(1-\nu))$)

$$\boldsymbol{C} = \frac{(1-\nu)E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1 & f & 0 & f \\ f & 1 & 0 & f \\ 0 & 0 & g & 0 \\ f & f & 0 & 1 \end{bmatrix}.$$

For thermal loading $\boldsymbol{\epsilon}_c^0 = \alpha \Delta T [1, 1, 0, 1]^t$.

The meshing for an axisymmetric problem using CST elements, and the degrees of freedom for a single element are shown in Fig. 4.15. The displacement is given by

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{N}\hat{\boldsymbol{u}} \\ &= \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 \end{bmatrix}^t. \end{aligned}$$

For an isoparametric element,

$$\begin{aligned} r &= \xi r_1 + \eta r_2 + (1-\xi-\eta) r_3, \\ z &= \xi z_1 + \eta z_2 + (1-\xi-\eta) z_3. \end{aligned}$$

In the usual manner, we have

$$\begin{bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} \frac{\partial u}{\partial \xi} \\ \frac{\partial u}{\partial \eta} \end{bmatrix}; \quad \begin{bmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial z} \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} \frac{\partial w}{\partial \xi} \\ \frac{\partial w}{\partial \eta} \end{bmatrix},$$

where (with $r_{ij} \equiv r_i - r_j$ and $z_{ij} \equiv z_i - z_j$)

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial r}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial r}{\partial \eta} & \frac{\partial z}{\partial \eta} \end{bmatrix} = \begin{bmatrix} r_{13} & z_{13} \\ r_{23} & z_{23} \end{bmatrix}$$

Note that $\det \boldsymbol{J} = r_{13}z_{23} - r_{23}z_{13} = 2A$. The engineering strain vector is given by

$$\boldsymbol{\epsilon}_c = \boldsymbol{B}\hat{\boldsymbol{u}},$$

where

$$\boldsymbol{B} = \begin{bmatrix} \frac{z_{23}}{|\boldsymbol{J}|} & 0 & \frac{z_{31}}{|\boldsymbol{J}|} & 0 & \frac{z_{12}}{|\boldsymbol{J}|} & 0 \\ 0 & \frac{r_{32}}{|\boldsymbol{J}|} & 0 & \frac{r_{13}}{|\boldsymbol{J}|} & 0 & \frac{r_{21}}{|\boldsymbol{J}|} \\ \frac{r_{32}}{|\boldsymbol{J}|} & \frac{z_{23}}{|\boldsymbol{J}|} & \frac{r_{13}}{|\boldsymbol{J}|} & \frac{z_{31}}{|\boldsymbol{J}|} & \frac{r_{21}}{|\boldsymbol{J}|} & \frac{z_{12}}{|\boldsymbol{J}|} \\ \frac{N_1}{r} & 0 & \frac{N_2}{r} & 0 & \frac{N_3}{r} & 0 \end{bmatrix}$$

The element stiffness matrix is given by

$$\begin{aligned} \boldsymbol{K}^{(e)} &= 2\pi \int \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} r \, dr dz \\ &= 2\pi \int_0^1 \int_0^{1-\eta} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} r (2A_e) \, d\xi d\eta \\ &= 4\pi A_e \int_0^1 \int_0^{1-\eta} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B} r \, d\xi d\eta, \end{aligned}$$

where $r = \xi r_1 + \eta r_2 + (1 - \xi - \eta) r_3$. The element force vectors are (assuming that the traction acts on edge 2-3 of the element)

$$\begin{aligned} \boldsymbol{f}^{(e)} &= 2\pi \int_0^1 \int_0^{1-\eta} \rho \boldsymbol{N}^t \boldsymbol{b} r (2A_e) \, d\xi d\eta + 2\pi \int_0^1 \boldsymbol{N}^t \bar{\boldsymbol{t}} r l_{23} \, d\eta \\ &= 4\pi \rho A_e \int_0^1 \int_0^{1-\eta} \begin{bmatrix} N_1 r b_r \\ N_1 r b_z \\ N_2 r b_r \\ N_2 r b_z \\ N_3 r b_r \\ N_3 r b_z \end{bmatrix} d\xi d\eta + 2\pi l_{23} \int_0^1 \begin{bmatrix} 0 \\ 0 \\ N_2 r t_r \\ N_2 r t_z \\ N_3 r t_r \\ N_3 r t_z \end{bmatrix} d\eta, \end{aligned}$$

where $r = \eta r_2 + (1 - \eta) r_3$ in the second integral.

Since the 2π factor is common in both, the stiffness and the force matrices, one can cancel it out, or alternatively, carry out the formulation using a one radian sector in the θ direction. An example of a axisymmetric problem is a rotating flywheel with $b_r = \rho r\omega^2$ and $b_z = -\rho g$.

For the four-node bilinear element, the strain-displacement matrix is found in the same manner as in the plane stress/plane strain case. We have

$$\begin{bmatrix} \epsilon_{rr} \\ \epsilon_{zz} \\ \gamma_{rz} \\ \epsilon_{\theta\theta} \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \\ \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial z} \\ u \end{bmatrix}, \tag{4.12}$$

where

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{r} \end{bmatrix}.$$

Also

$$\begin{bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \\ \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial z} \\ u \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \frac{\partial u}{\partial \xi} \\ \frac{\partial u}{\partial \eta} \\ \frac{\partial w}{\partial \xi} \\ \frac{\partial w}{\partial \eta} \\ u \end{bmatrix}, \tag{4.13}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{J}^{-1} & 0 & 0 \\ 0 & \boldsymbol{J}^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Finally,

$$\begin{bmatrix} \frac{\partial u}{\partial \xi} \\ \frac{\partial u}{\partial \eta} \\ \frac{\partial w}{\partial \xi} \\ \frac{\partial w}{\partial \eta} \\ u \end{bmatrix} = \boldsymbol{Q}\hat{\boldsymbol{u}}, \tag{4.14}$$

where

$$\boldsymbol{Q} = \begin{bmatrix} N_{1,\xi} & 0 & N_{2,\xi} & 0 & N_{3,\xi} & 0 & N_{4,\xi} & 0 \\ N_{1,\eta} & 0 & N_{2,\eta} & 0 & N_{3,\eta} & 0 & N_{4,\eta} & 0 \\ 0 & N_{1,\xi} & 0 & N_{2,\xi} & 0 & N_{3,\xi} & 0 & N_{4,\xi} \\ 0 & N_{1,\eta} & 0 & N_{2,\eta} & 0 & N_{3,\eta} & 0 & N_{4,\eta} \\ N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \end{bmatrix}.$$

From Eqns. 4.12-4.14, we get

$$\boldsymbol{B} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{Q}.$$

The stiffness matrix is given by

$$\boldsymbol{K}^{(e)} = 2\pi \int_{-1}^{1}\int_{-1}^{1} \boldsymbol{B}^t\boldsymbol{C}\boldsymbol{B}\,|\boldsymbol{J}|\,r\,d\xi d\eta,$$

where $r = N_1 r_1 + N_2 r_2 + N_3 r_3 + N_4 r_4$. Assuming that the tractions act on edge 2-3, the element load vector is

$$\boldsymbol{f}^{(e)} = 2\pi \int_{-1}^{1}\int_{-1}^{1} \rho \boldsymbol{N}^t \boldsymbol{b} r\,|\boldsymbol{J}|\;d\xi d\eta + 2\pi l_{23}\int_{-1}^{1} [\boldsymbol{N}^t \bar{\boldsymbol{t}} r]_{\xi=1}\,d\eta.$$

Some terms in the $\boldsymbol{K}$ matrix are of the form $1/r$. With Gauss quadrature, these terms remain finite since there are no Gauss points at $r = 0$. For stress computation, the indeterminate form $\epsilon_{\theta\theta} = u/r = 0/0$ arises for points on the z-axis. Hence, it is convenient to calculate the stresses at the Gauss points and extrapolate them to the axis. Another way is to use the fact that $\epsilon_{\theta\theta} = \epsilon_{rr}$ at $r = 0$. Thus, for $r = 0$, replace the $\epsilon_{\theta\theta}$ row of the $\boldsymbol{B}$ matrix by the ϵ_{rr} row. The radial displacement $u = 0$ has to be prescribed at all nodes that lie on the z-axis. Rigid body motion is restrained by prescribing w on a single nodal circle.

Chapter 5

Convergence of the Finite Element Method

In this chapter, we discuss some of the important theoretical properties of the finite element method. Recall that V_h is a subspace of the Hilbert space V in which the finite element solution lies, and that h is the generic element size.

5.1 Some Properties of the Finite Element Method

The abstract formulation of the finite element method is
Find $u_h \in V_h$ such that

$$a(u_h, v_h) = L(v_h) \quad \forall v_h \in V_h.$$

Assume $a(.,.)$ to be symmetric. Then we have the following properties:
Property 1 (Error equation): Let the error between the exact solution and the finite element solution be e_h, i.e., $e_h = u - u_h$. Then

$$a(e_h, v_h) = 0 \quad \forall v_h \in V_h.$$

Proof: Since $V_h \subset V$, we have

$$\begin{aligned} a(u, v_h) &= L(v_h) \quad \forall v_h \in V_h, \\ a(u_h, v_h) &= L(v_h) \quad \forall v_h \in V_h. \end{aligned}$$

Subtracting the second equation from the first and using the bilinearity of $a(.,.)$, we get the desired result. This result can be interpreted as saying that the error is orthogonal to all elements of V_h in $a(.,.)$. Thus, as V_h increases, with the larger space containing the smaller space, the solution accuracy increases continuously.
Property 2:

$$a(u_h, u_h) \leq a(u, u).$$

Proof:

$$a(u, u) = a(u_h + e_h, u_h + e_h)$$

$$= a(u_h, u_h) + a(e_h, e_h) + 2a(u_h, e_h)$$
$$= a(u_h, u_h) + a(e_h, e_h), \tag{5.1}$$

since $a(u_h, e_h) = 0$ by property 1 with $v_h = u_h$. Since $a(e_h, e_h) > 0$ for any $e_h \neq 0$, we get the desired relation. The physical interpretation of this statement is that the strain energy corresponding to the finite element solution is always smaller than or equal to the strain energy corresponding to the exact solution, or in other words, the finite element approximation makes the structure 'stiffer' than it actually is.

Property 3:

$$a(e_h, e_h) \leq a(u - v_h, u - v_h) \quad \forall v_h \in V_h.$$

Proof: Since $a(e_h, w_h) = 0$ for all $w_h \in V_h$,

$$a(u - u_h + w_h, u - u_h + w_h) = a(e_h, e_h) + a(w_h, w_h),$$

which in turn implies that

$$a(e_h, e_h) \leq a(e_h + w_h, e_h + w_h).$$

Choose $w_h = u_h - v_h$ to get the desired result. Thus, the finite element solution u_h is chosen from all possible displacement patterns $v_h \in V_h$ such that the 'energy distance' between u and the elements in V_h is minimized. Thus, the solution u_h is the projection of u onto V_h as shown in Fig. 2.3.

5.2 Superconvergence of the Finite Element Method in One-dimensional Problems

We show that in certain types of one-dimensional problems, the finite element solution is exact at the nodes. In two or three dimensions however, we do not have this property. We first introduce the Dirac-Delta function $\delta_y(x) = \delta(y - x)$ which is defined by means of the equation

$$\int_0^1 w(x)\delta_y(x)\, dx = w(y).$$

Integration of $\delta_y(x)$ yields the Heaviside (or step function)

$$H_y(x) = H(x - y) = \begin{cases} 0 & \text{if } x < y \\ 1 & \text{if } x > y \end{cases}.$$

The integral of $H_y(x)$ is the Macaulay bracket:

$$\langle x - y \rangle = \begin{cases} 0 & \text{if } x \leq y \\ x - y & \text{if } x > y \end{cases}.$$

The pictorial representation of the three functions discussed above is shown in Fig. 5.1.

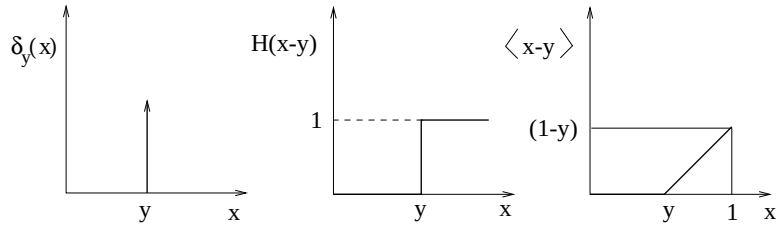


Figure 5.1: Dirac-Delta, Heaviside and Macaulay bracket functions

Consider the one-dimensional problem

$$\frac{d^2u}{dx^2} + f = 0 \quad \text{on } [0,1], \tag{5.2}$$

with the boundary conditions $u(0) = u(1) = 0$. The variational formulation is

$$\int_0^1 \frac{du}{dx}\frac{dv}{dx}\,dx = \int_0^1 fv\,dx \quad \forall v \in H_0^1(I). \tag{5.3}$$

The solution to Eqn. 5.2 when $f = \delta_y(x)$ (unit point load at $x = y$) is known as the Greens' function (denoted by G):

$$\frac{d^2G}{dx^2} + \delta_y(x) = 0 \quad \text{on } [0,1], \tag{5.4}$$

with $G(0) = G(1) = 0$. Integrating Eqn. 5.4, we get

$$G(x) + \langle x - y \rangle = c_1 x + c_2,$$

where c_1 and c_2 are determined from the boundary conditions on G. We get $c_1 = (1-y)$ and $c_2 = 0$, thus yielding

$$G(x) = \begin{cases} (1-y)x & \text{if } 0 \le x \le y \\ y(1-x) & \text{if } y \le x \le 1. \end{cases}$$

Note that G is piecewise linear and hence belongs to V_h when y is a nodal point. Substituting $u = G$ and $f = \delta_y(x)$ in Eqn. 5.3, we get

$$a(G,v) = \int_0^1 \frac{dG}{dx}\frac{dv}{dx} = \int_0^1 \delta_y(x)v\,dx = v(y) \quad \forall v \in H_0^1(I).$$

The error in the finite element solution at a nodal point x_i is

$$\begin{aligned} e(x_i) &= u(x_i) - u_h(x_i) \\ &= \int_0^1 (u - u_h)\delta_{x_i}(x)\,dx \\ &= \int_0^1 \frac{d}{dx}(u - u_h)\frac{dG}{dx}\,dx \quad (\text{since } u - u_h \in V) \\ &= a(e_h, v_h) \quad (\text{letting } v_h = G) \end{aligned}$$

$$= 0,$$

where the last step follows from the error equation. Note that we have used the fact that G belongs to V_h in the penultimate step. Thus,

$$u_h(x_i) = u(x_i) \quad i = 1, 2, \ldots, n,$$

and we get the result that the finite element solution is exact at the nodes. In some other kinds of one-dimensional problems, and in two and three-dimensional problems, we do not, in general, have this property since the Green's function does not necessarily belong to V_h.

5.3 Convergence of the Finite Element Solutions to the Exact Solution

We define convergence as

$$a(u - u_h, u - u_h) \to 0 \text{ as } h \to 0,$$

or, equivalently, using Eqn. 5.1 as

$$a(u_h, u_h) \to a(u, u) \text{ as } h \to 0.$$

Physically, the above statement means that the strain energy calculated by the finite element solution converges to the exact strain energy.

5.3.1 Criteria for Monotonic Convergence

For monotonic convergence, the elements must be complete and compatible. If these conditions are fulfilled, the accuracy of the solution will increase continuously as the mesh is refined. Mesh refinement is performed by subdividing a previously used element into two or more elements. Thus, the old mesh is 'embedded' in the new mesh.

Completeness means that the displacement functions of the element must be able to represent rigid body displacements and constant strains. Rigid-body displacements are those displacement modes that the element must undergo as a rigid body without stresses being developed in it. As an example, consider the shaded element in Fig. 5.2. This element should remain stress-free even after the application of the shown distributed load.

The straining and rigid-body modes can be found by solving the eigenvalue problem

$$\boldsymbol{K}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}.$$

The zero eigenvalues correspond to either rigid-body modes or spurious modes (such as those generated by using reduced integration).

An element should also be able to model a constant strain state because in the limit of mesh refinement, the strain in each element approaches a constant value, and any complex variation of strain can be approximated.

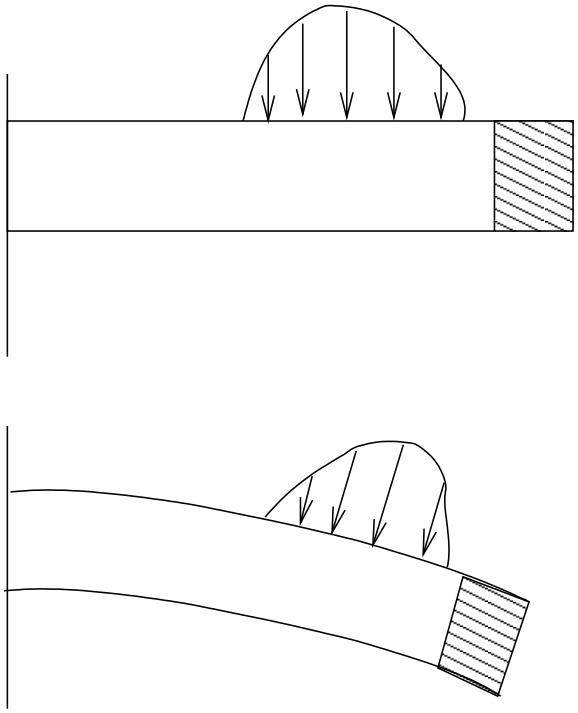


Figure 5.2: Element undergoing rigid-body displacement.

For example, a plane stress element should be able to represent constant τ_{xx}, τ_{xy} and τ_{yy} conditions.

The second requirement for monotonic convergence, viz. compatibility, means that the displacements within an element and across element boundaries must be continuous. Physically, compatibility ensures that no gaps occur between elements when the assemblage is loaded. For example, in a 3-D element, the u, v and w displacements must be continuous, in a beam element v and dv/dx must be continuous, and in a plate element based on the Kirchoff theory w, $\partial w/\partial x$, $\partial w/\partial y$ must be continuous.

In general if the potential $\Pi = \Pi(\phi)$ contains derivatives of ϕ through order m, then

1. Within each element, the assumed field must contain a complete polynomial of degree m,

2. Across element boundaries, ϕ must be continuous through order $m-1$.

A polynomial is complete if it is of high-enough degree and if no terms are omitted. For example, in one-dimension, a complete polynomial of degree 2 is $a_0 + a_1 x + a_2 x^2$. In two-dimensions, a polynomial is of degree n if it contains a term of the form $x^l y^m$, where l and m are nonnegative integers, and $l + m = n$, and is complete if all combinations of l and m for which $l + m \leq n$ are included. For example, a complete polynomial of degree 2 in two-dimensions is

$$a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2.$$

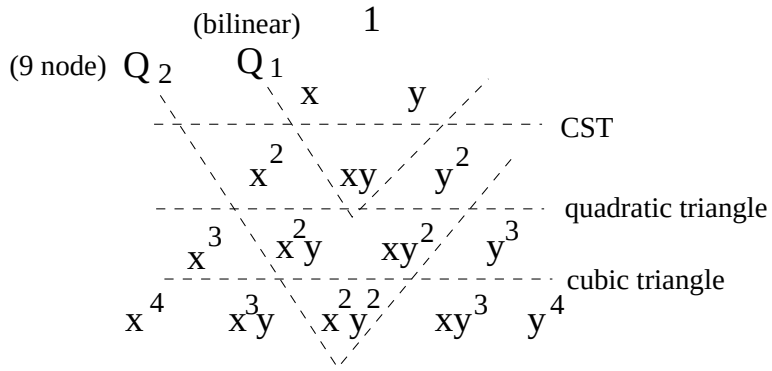


Figure 5.3: Pascal triangle.

In two dimensions, a complete polynomial of degree n contains $(n+1)(n+2)/2$ terms which can be seen from the Pascal triangle shown in Fig. 5.3. Geometric isotropy requires that there are no 'preferred directions'. Hence, the displacement expansions for u and v in a two-dimensional problem should be the same, and should include terms symmetric about the axis of the Pascal triangle. As an example

$$u = a_1 + a_2 x + a_3 y + a_4 x^2,$$
$$v = a_5 + a_6 x + a_7 y + a_8 y^2,$$

is not geometrically isotropic, whereas

$$u = a_1 + a_2 x + a_3 y + a_4 xy,$$
$$v = a_5 + a_6 x + a_7 y + a_8 xy,$$

is geometrically isotropic. Similar remarks apply for three-dimensions.

5.4 Rate of Convergence

The rate of convergence depends on the order of the polynomials used in the displacement assumptions. Let us assume that we employ elements with complete polynomials of degree k, and that the exact solution $\boldsymbol{u}$ to our elasticity problem is 'smooth' in the sense that $\boldsymbol{u} \in H^{k+1}(\Omega)$, so that $\|\boldsymbol{u}\|_{H^{k+1}(\Omega)} < \infty$, where

$$\|\boldsymbol{u}\|^2_{H^{k+1}(\Omega)} = \int_\Omega \left[\sum_{i=1}^{3} u_i^2 + \sum_{i=1}^{3}\sum_{j=1}^{3} \left(\frac{\partial u_i}{\partial x_j} \right)^2 + \sum_{n=2}^{k+1}\sum_{i=1}^{3} \sum_{r+s+t=n} \left(\frac{\partial^n u_i}{\partial x_1^r \partial x_2^s \partial x_3^t} \right)^2 \right] d\Omega.$$

Then the rate of convergence of $\boldsymbol{u}_h$ to $\boldsymbol{u}$ is given by

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{H^1(\Omega)} \leq c h^k \|u\|_{H^{k+1}(\Omega)}, \tag{5.5}$$

where c is a constant independent of h, but dependent on the material properties. From Eqn. 5.5, we can say that the rate of convergence is $\mathrm{o}\left(h^k\right)$.

The dependence of c on the material properties is detrimental when (almost) incompressible materials are considered because c becomes very large, and the order of convergence k results in good accuracy only at very small (impractical) values of h. The constant c depends on the type of element being used. For example, the CST and Q4 elements which both belong to the $k = 1$ group give different magnitudes of error for a given h, while the order with which the error decreases as the mesh is refined is the same. Eqn. 5.5 gives in essence an error estimate for the displacement gradient (and hence for the stresses and strains) because the primary contribution to the $H^1(\Omega)$ norm is due to errors in the derivatives of displacements. The error in the displacements is

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L_2(\Omega)} \leq ch^{k+1} \|\boldsymbol{u}\|_{H^{k+1}(\Omega)} .$$

Thus, the order of convergence for the displacements is one order higher than for the strains.

These results are intuitively reasonable. If we think in terms of a Taylor series analysis, a finite element of dimension h with a complete displacement field expansion of order k can represent displacement variations upto that order exactly. Hence, the local error in representing arbitrary displacements with a uniform mesh should be o $\left(h^{k+1}\right)$. The stresses are obtained by differentiating the displacements, and hence have error of o $\left(h^k\right)$.

In the above convergence study, it is assumed that uniform discretizations are used (e.g., square elements in 2-D), and that the exact solution is smooth. If the solution is not smooth, and a uniform mesh is used, the order of convergence decreases. In practice, graded meshes are used with small elements in the areas of stress concentration and larger elements away from those regions. In general, a refined mesh is required in places where acute changes in geometry, boundary conditions, loading, material properties or solution occur. A 'good' mesh is one for which the error is almost equal for all the elements, and below a certain tolerance.

The method just discussed where we increase the number of elements keeping the interpolation functions fixed is known as 'h-adaptivity'. An alternative method of increasing the accuracy of the solution is the 'p' method where the order of the interpolation functions in 'high-error' elements is increased using hierarchical shape functions.

A very high rate of convergence in the solution can be obtained if we increase the number of elements and at the same time increase the order of interpolation functions. This approach is known as the h-p method. In all of the above approaches, the mesh grading should be such that the error in each element is almost the same.

Mesh distortion and numerical integration will not reduce the order of convergence provided the distortion is reasonable and full numerical integration is used.

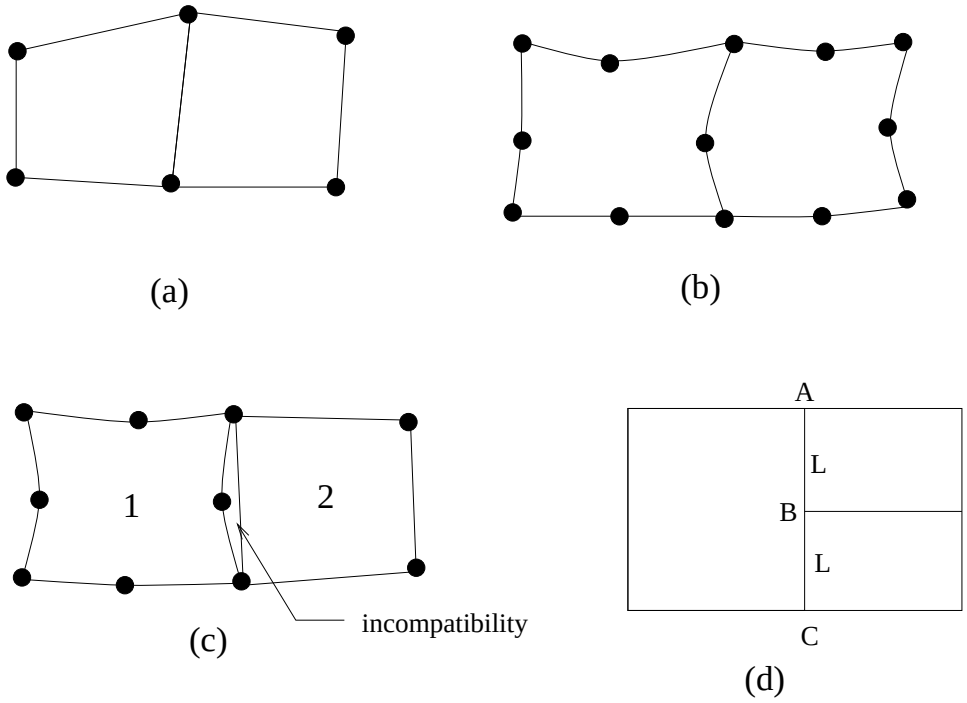


Figure 5.4: Compatibility of an element assemblage.

5.5 Convergence of Isoparametric Elements

As we have seen, the two requirements for convergence are compatibility and completeness. To investigate the compatibility of an element assemblage, we need to consider each face between adjacent elements. For compatibility, it is necessary that the coordinates and the displacements of the elements at the face must be the same. For example, consider the cases shown in Fig. 5.4. For the bilinear element shown in (a), the coordinates and displacements are interpolated as $x = \sum_i N_i x_i$ and $u = \sum_i N_i u_i$. At the shared edge $\xi = 1$, we have $x \sim \eta$. Since, a linear interpolation between two points can be carried out only in one way, we have compatibility of the geometry and displacements along the edge. Similarly, for the serendipity element shown in (b), we have compatibility along the shared edge $\xi = 1$, since a quadratic interpolation through three points is unique. In contrast, the element assemblage shown in (c) is incompatible, since the geometry and displacements can vary quadratically along the edge in element 1, but only linearly in element 2. Compatibility can be achieved as shown in (d) by imposing the constraint equations $u_B = (u_A + u_C)/2$ and $v_B = (v_A + v_C)/2$.

Mesh grading can be achieved easily using isoparametric elements. A one-to-two layer transition is shown in Fig. 5.5.

We now analyze the criterion of completeness. First, we prove that in an isoparametric formulation, $\sum_i N_i = 1$. Observe that the relations $x = \sum_i N_i x_i$, $y = \sum_i N_i y_i$ and $z = \sum_i N_i z_i$ can be used regardless of the location

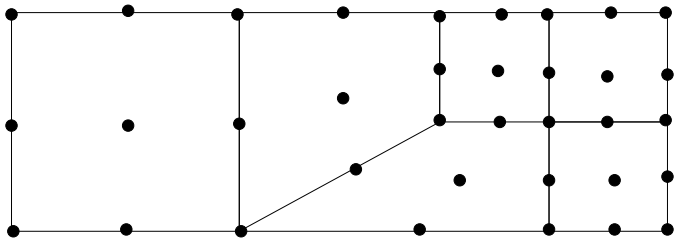


Figure 5.5: Mesh grading using isoparametric elements.

of the coordinate system. Let $x_i = X_i + h$. Then

$$x = \sum_i N_i x_i = \sum_i N_i X_i + h \sum_i N_i = X + h \sum_i N_i = x - h + h \sum_i N_i,$$

which implies that $\sum_i N_i = 1$ since h is nonzero.

Completeness requires that rigid body displacements and constant strain states should be possible, i.e., the lowest possible interpolation should be

$$\begin{aligned} u &= a_1 + b_1 x + c_1 y + d_1 z, \\ v &= a_2 + b_2 x + c_2 y + d_2 z, \\ w &= a_3 + b_3 x + c_3 y + d_3 z. \end{aligned} \tag{5.6}$$

The nodal displacements are

$$\begin{aligned} u_i &= a_1 + b_1 x_i + c_1 y_i + d_1 z_i, \\ v_i &= a_2 + b_2 x_i + c_2 y_i + d_2 z_i, \\ w_i &= a_3 + b_3 x_i + c_3 y_i + d_3 z_i, \end{aligned} \tag{5.7}$$

at each node i. The test of completeness is that the displacement field given by Eqn. 5.6 should be obtained within the element when the nodal displacements are given by Eqn. 5.7. Since $u = \sum_i N_i u_i$, $v = \sum_i N_i v_i$ and $w = \sum_i N_i w_i$, we have

$$\begin{aligned} u &= a_1 \sum_i N_i + b_1 \sum_i N_i x_i + c_1 \sum_i N_i y_i + d_1 \sum_i N_i z_i, \\ v &= a_2 \sum_i N_i + b_2 \sum_i N_i x_i + c_2 \sum_i N_i y_i + d_2 \sum_i N_i z_i, \\ w &= a_3 \sum_i N_i + b_3 \sum_i N_i x_i + c_3 \sum_i N_i y_i + d_3 \sum_i N_i z_i. \end{aligned} \tag{5.8}$$

For an isoparametric formulation $\sum_i N_i x_i = x$, $\sum_i N_i y_i = y$ and $\sum_i N_i z_i = z$. Also, we have proved that $\sum_i N_i = 1$. Thus, Eqn. 5.8 agrees with Eqn. 5.6, and completeness is established.

5.6 Effect of Element Distortion on the Order of Convergence

We saw that the rate of convergence of $\boldsymbol{u}_h$ to $\boldsymbol{u}$ is given by

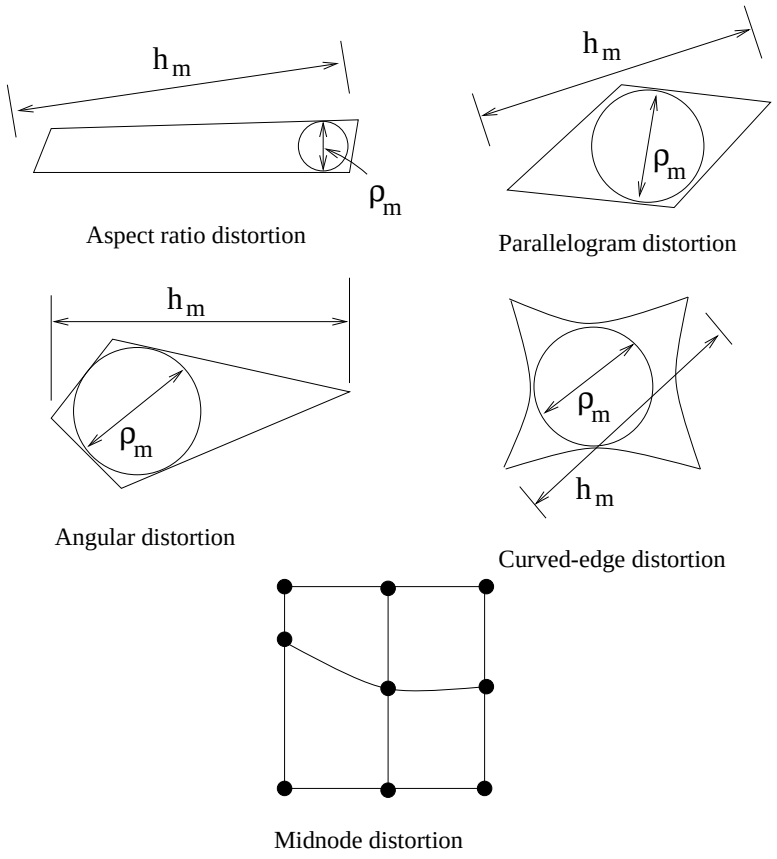


Figure 5.6: Types of element distortion.

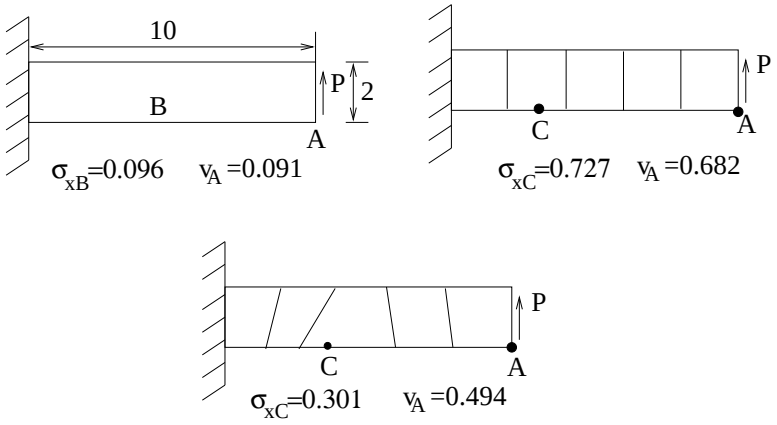


Figure 5.7: Effect of mesh distortion in bilinear elements.

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{H^1(\Omega)} \leq ch^k \|\boldsymbol{u}\|_{H^{k+1}(\Omega)}\,, \tag{5.9}$$

where k is the order of polynomial, and c is a constant independent of h but dependent on material properties. We introduce a measure of mesh regularity

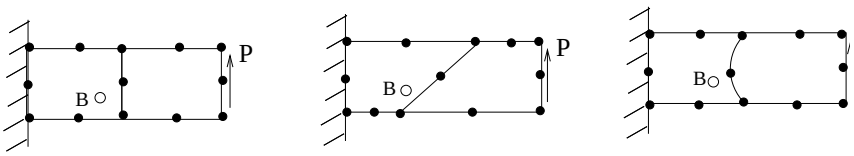
$$\sigma_m = \frac{h_m}{\rho_m},$$

where h_m is the largest dimension of the element, and ρ_m is the diameter of the largest circle or sphere that can be inscribed in m (see Fig. 5.6). A sequence of meshes is regular if $\sigma_m \leq \sigma_0$, where σ_0 is a fixed positive value. We also require that for each element, the ratio of the largest to the smallest side lengths is smaller than a reasonable positive number.

Since all the distortions (except the midnode distortion which is used only in fracture mechanics problems) shown in Fig. 5.6 are used in mesh designs (to achieve grading), we need to study the effect of these distortions on the order of convergence. The rate of convergence given by Eqn. 5.9 is maintained if the distortion compared to the size of the element is small.

When element sizes are large, geometric distortions can affect the predictive capabilities to a significant degree as shown in Fig. 5.7 and 5.8. The values shown in these figures are finite element solutions which have been normalized with respect to the exact solution. One can see from the results that geometric distortion makes the results less accurate. From Fig. 5.8, we see that the 8-node and 9-node results are comparable for undistorted elements, but that the 9-node element is vastly superior when the elements are distorted. In fact, the 8-node element even predicts a wrong sign for the stress at point B with 2×2 integration.

An ideal element is compact, straight-sided and has equal corner angles. Hence, one should try to minimize the distortion of elements to as large an extent as possible. The value of the determinant of the Jacobian matrix at the Gauss points is a good indicator of the amount of distortion. We need $|\boldsymbol{J}| > 0$ at all the Gauss points. The loss of predictive capability is due to the element no longer being able to represent the same order of polynomials



		σ_B	v_A	σ_B	v_A	σ_B	v_A
8-node	2x2	1.0	0.968	0.051	0.362	-0.048	0.430
	3x3	1.129	0.93	0.048	0.161	0.050	0.221
9-node	2x2	1.00	1.006	1.125	1.109	0.958	0.955
	3x3	1.141	0.954	0.687	0.791	0.705	0.737

Figure 5.8: Effect of mesh distortion in 8-node and 9-node quadrilateral elements.

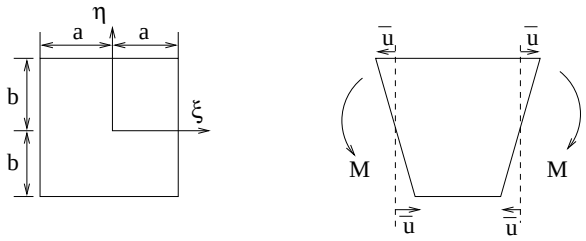


Figure 5.9: Parasitic shear in a bilinear element.

in the physical coordinates x, y and z after the geometric distortion as it did without the distortion. For example, the 8-node element is able to represent only the terms 1, x and y after geometric distortion, while before distortion, it can represent 1, x, y, x^2, xy, y^2, x^2y and xy^2. Similarly, the 9-node element can represent only 1, x, y, x^2, xy and y^2 after distortion, while it can represent 1, x, y, x^2, xy, y^2, x^2y, xy^2 and x^2y^2 without distortion.

5.7 Incompatible Elements and the Patch Test

Bilinear elements are attractive because of their simplicity. Unfortunately, they are too stiff in bending. Consider a bilinear element subjected to pure bending as shown in Fig. 5.9. The finite element solution is

$$u = \xi\eta\bar{u}; \quad v = 0,$$

leading to the strains

$$\epsilon_{xx} = \frac{\eta\bar{u}}{a}; \quad \epsilon_{yy} = 0; \quad \gamma_{xy} = \frac{\xi\bar{u}}{b}.$$

The exact displacements and strains are

$$u = \xi\eta\bar{u}; \quad v = (1-\xi^2)\frac{a\bar{u}}{2b} + (1-\eta^2)\nu\frac{b\bar{u}}{2a},$$
$$\epsilon_{xx} = \frac{\eta\bar{u}}{a}; \quad \epsilon_{yy} = -\nu\eta\frac{\bar{u}}{a}; \quad \gamma_{xy} = 0.$$

Comparing the finite element solution with the exact solution, we see that there is a spurious shear strain γ_{xy}. For large a/b, the mesh 'locks'.

To overcome the problem of parasitic shear, we improve the $Q4$ element by adding additional degrees of freedom:

$$u = \sum_i N_i u_i + (1-\xi^2)a_1 + (1-\eta^2)a_2,$$
$$v = \sum_i N_i v_i + (1-\xi^2)a_3 + (1-\eta^2)a_4,$$

where N_i are the usual bilinear functions. The above element represents bending exactly, but now the compatibility of displacements at the element edges is lost. Such elements are called incompatible or nonconforming elements.

The results obtained using incompatible elements are of high quality. However, the exact potential energy of the system is not necessarily an upper bound on the calculated potential energy. Hence, monotonic convergence is not ensured. But we can establish conditions so that there is at least non-monotonic convergence. The conditions are

- The element completeness condition must always be satisfied, i.e., it should be able to represent rigid body modes and the constant strain states
- As the finite element mesh is refined, each element should approach a constant strain condition. Hence, a assemblage of incompatible finite elements should be able to represent constant strain conditions. Note that this condition is on an assemblage of elements and not on a single element.

To test the second condition, the patch test has been proposed. In this test a patch of elements is subjected to nodal loads consistent with constant τ_{xx}, τ_{yy} and τ_{xy} in a two-dimensional problem (or all the six stress components in a three-dimensional one). If for any patch of elements, the element stresses actually represent the constant stress conditions, then the element passes the patch test. If an incompatible element passes the patch-test then convergence is ensured (although it may not be monotonic). In the case of a compatible element, we know that the convergence is monotonic. In such a case, the patch test can be used as a 'debugging' aid.

Two examples of the patch test are shown in Fig. 5.10. The roller supports are provided since a fixed support would prevent the Poisson contraction in the y-direction. Actually, we need to test all possible patches. But testing just one patch is what is practiced. This practice is generally found to be quite reliable.

One can design variations of the patch test. For example, if an element fails to display the constant stress state in a patch of elements, but displays the constant stress state as the mesh is repeatedly subdivided, then a 'weak patch test' can be designed with a larger number of elements in the assemblage being tested. Similarly a higher-order patch test can be designed for

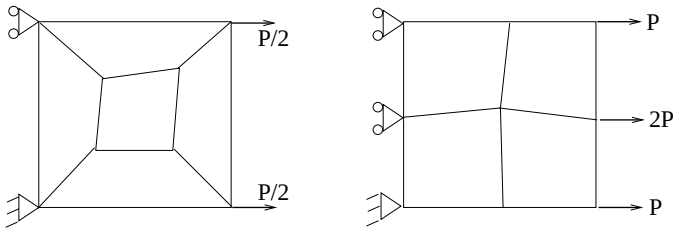


Figure 5.10: Two examples of the patch test.

a 9-node element subjected to pure bending to test whether it gives a linear stress distribution within the element.

Chapter 6

Eigenvalue and Time-Dependent Problems

6.1 Eigenvalue problems

An eigenvalue problem deals with the determination of the values of the parameter λ such that the equation

$$A(u) = \lambda B(u),$$

where $A(u)$ and $B(u)$ are differential operators, has nontrivial solutions. The values of λ are called the eigenvalues and the associated functions u are called the eigenfunctions. As an example, in the equation

$$-\frac{d^2u}{dx^2} = \lambda u,$$

λ denotes the square of the frequency of vibration ω.

In structural problems, eigenvalues denote natural frequencies or buckling loads. In fluid mechanics and heat transfer, eigenvalue problems arise in connection with the determination of the homogeneous part of the solution. Eigenvalues denote the amplitudes of the Fourier components making up the solution. Eigenvalues are also useful in determining the stability characteristics of temporal schemes.

Consider the axial motion of a bar

$$\rho\frac{\partial^2u}{\partial t^2} - \frac{\partial}{\partial x}\left(E\frac{\partial u}{\partial x}\right) = b, \tag{6.1}$$

where b is the axial force per unit length. The natural (i.e., when $b = 0$) axial oscillations of the bar are periodic and can be determined by assuming a solution of the form

$$u(x,t) = U(x)e^{i\omega t},\ i = \sqrt{-1}, \tag{6.2}$$

where ω is the natural frequency and $U(x)$ is the mode shape. Substituting Eqn. 6.2 into the homogeneous form of Eqn. 6.1, we get

$$\left[-\rho\omega^2U - \frac{d}{dx}\left(E\frac{dU}{dx}\right)\right]e^{i\omega t} = 0,$$

or

$$-\frac{d}{dx}\left(E\frac{dU}{dx}\right) - \lambda\rho U = 0, \tag{6.3}$$

where $\lambda \equiv \omega^2$. Eqn. 6.3 is an eigenvalue problem which involves finding λ and U.

For constant E and ρ, the solution is given by

$$U(x) = c_1 \sin\sqrt{\frac{\lambda\rho}{E}}x + c_2 \cos\sqrt{\frac{\lambda\rho}{E}}x,$$

yielding

$$u(x,t) = \left[c_1 \sin\sqrt{\frac{\lambda\rho}{E}}x + c_2 \cos\sqrt{\frac{\lambda\rho}{E}}x\right] e^{i\omega t}.$$

The constants c_1, c_2 and λ are determined from the initial and boundary conditions.

Now consider the parabolic partial differential equation

$$\rho c\frac{\partial T}{\partial t} - \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) = Q. \tag{6.4}$$

We seek a homogeneous solution to the above equation in the form

$$T = U(x)e^{-\lambda t}. \tag{6.5}$$

Substituting Eqn. 6.5 in the homogeneous part of Eqn. 6.4, we get

$$-\frac{d}{dx}\left(k\frac{\partial U}{\partial x}\right) - \lambda\rho c U = 0,$$

which is of the same form as Eqn. 6.3. Hence, we obtain

$$T = e^{-\lambda t}\left[c_1 \sin\sqrt{\frac{\rho c\lambda}{k}}x + c_2 \cos\sqrt{\frac{\rho c\lambda}{k}}x\right].$$

Similarly, the transverse free vibration of a beam using the Euler-Bernoulli beam theory is governed by

$$\rho A\frac{\partial^2 v}{\partial t^2} + \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 v}{\partial x^2}\right) = 0.$$

Assuming the solution to be of the form

$$v = U(x)e^{i\omega t},$$

where ω is the frequency of vibration, we get

$$\frac{d^2}{dx^2}\left(EI\frac{d^2U}{dx^2}\right) = \lambda\rho A U,$$

which is an eigenvalue problem.

The study of buckling also leads to an eigenvalue problem. The equation of equilibrium of a beam subjected to an axial force P is

$$\frac{d^2}{dx^2}\left(EI\frac{d^2v}{dx^2}\right) + P\frac{d^2v}{dx^2} = 0,$$

which is an eigenvalue problem with $\lambda = P$ as the eigenvalue, which represents the buckling load. The smallest value of λ is the critical buckling load.

6.2 Hamilton's principle for dynamics problems

The equation of motion is given by

$$\rho\frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \boldsymbol{\nabla}\cdot\boldsymbol{\tau} + \rho\boldsymbol{b}.$$

The virtual work principal is derived in the usual way. Starting from

$$\int_\Omega \left(\boldsymbol{\nabla}\cdot\boldsymbol{\tau} + \rho\boldsymbol{b} - \rho\frac{\partial^2\boldsymbol{u}}{\partial t^2}\right)\cdot\boldsymbol{v}\,d\Omega + \int_{\Gamma_t}(\bar{\boldsymbol{t}}-\boldsymbol{t})\cdot\boldsymbol{v}\,d\Gamma = 0 \quad \forall\boldsymbol{v}.$$

Integrating by parts, we get

$$-\int_\Omega \rho\frac{\partial^2\boldsymbol{u}}{\partial t^2}\cdot\boldsymbol{v}\,d\Omega + \int_\Omega \rho\boldsymbol{b}\cdot\boldsymbol{v}\,d\Omega + \int_{\Gamma_t}\bar{\boldsymbol{t}}\cdot\boldsymbol{v}\,d\Gamma - \int_\Omega \boldsymbol{\tau}:\boldsymbol{\epsilon}(\boldsymbol{v})\,d\Omega = 0 \quad \forall\boldsymbol{v}. \tag{6.6}$$

We shall use Eqn. 6.6 to develop the finite element formulation. Just for the sake of completeness, however, we shall derive Hamilton's principle which we had used in Section 1.2.5.

Assuming the existence of a strain-energy density function W for the body so that

$$\boldsymbol{\tau} = \frac{\partial W}{\partial \boldsymbol{\epsilon}},$$

the above expression can be written as

$$\int_\Omega \rho\frac{\partial^2\boldsymbol{u}}{\partial t^2}\cdot\boldsymbol{v}\,d\Omega + \delta^{(1)}\Pi = 0, \tag{6.7}$$

where

$$\Pi = \int_\Omega W\,d\Omega - \int_\Omega \rho\boldsymbol{b}\cdot\boldsymbol{u}\,d\Omega - \int_{\Gamma_t}\bar{\boldsymbol{t}}\cdot\boldsymbol{u}\,d\Gamma.$$

Now we enforce Eqn. 6.7 which holds at every instant of time t in a weak sense with respect to time:

$$\int_{t_1}^{t_2}\int_\Omega \left[\rho\frac{\partial^2\boldsymbol{u}}{\partial t^2}\cdot\boldsymbol{v}\,d\Omega\right]\,dt + \int_{t_1}^{t_2}\delta^{(1)}\Pi\,dt = 0,$$

which on interchanging the order of integration yields

$$\int_\Omega \left[\int_{t_1}^{t_2} \rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} \cdot \boldsymbol{v} \, dt \right] d\Omega + \delta^{(1)} \int_{t_1}^{t_2} \Pi \, dt = 0.$$

Integrating by parts, we get

$$\int_\Omega \rho \, \frac{\partial \boldsymbol{u}}{\partial t} \cdot \boldsymbol{v} \Big|_{t_1}^{t_2} - \int_\Omega \int_{t_1}^{t_2} \rho \frac{\partial \boldsymbol{u}}{\partial t} \cdot \frac{\partial \boldsymbol{v}}{\partial t} \, dt + \delta^{(1)} \int_{t_1}^{t_2} \Pi \, dt = 0. \tag{6.8}$$

Assuming that $\boldsymbol{u}$ is prescribed or $\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{0}$ at t_1 and t_2, Eqn. 6.8 reduces to

$$- \int_{t_1}^{t_2} \int_\Omega \frac{\rho}{2} \delta^{(1)} \frac{\partial \boldsymbol{u}}{\partial t} \cdot \frac{\partial \boldsymbol{u}}{\partial t} \, d\Omega \, dt + \delta^{(1)} \int_{t_1}^{t_2} \Pi \, dt = 0.$$

Assuming ρ to be constant, the above equation can be written as

$$\delta^{(1)} \int_{t_1}^{t_2} (T - \Pi) \, dt = 0,$$

where

$$T = \int_\Omega \frac{\rho}{2} \frac{\partial \boldsymbol{u}}{\partial t} \cdot \frac{\partial \boldsymbol{u}}{\partial t} \, d\Omega$$

is the kinetic energy of the body at time t. Alternatively, we can write the above equation as

$$\delta^{(1)} \int_{t_1}^{t_2} L \, dt = 0,$$

where $L = T - \Pi$, which is nothing but Hamilton's principle.

6.2.1 Finite element formulation

As mentioned, we use Eqn. 6.6 to develop the finite element formulation. Discretizing the space variables (but not the time variable; if only partial discretization is carried out, the process is known as semi-discretization, if time is also discretized, then it is known as full-discretization), we get

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{N}\hat{\boldsymbol{u}} + \bar{\boldsymbol{N}}\bar{\hat{\boldsymbol{u}}}, & \boldsymbol{\epsilon}_c(\boldsymbol{u}) &= \boldsymbol{B}\hat{\boldsymbol{u}} + \bar{\boldsymbol{B}}\bar{\hat{\boldsymbol{u}}}, \\ \boldsymbol{v} &= \boldsymbol{N}\hat{\boldsymbol{v}}, & \boldsymbol{\epsilon}_c(\boldsymbol{v}) &= \boldsymbol{B}\hat{\boldsymbol{v}}, \\ \dot{\boldsymbol{u}} &= \boldsymbol{N}\dot{\hat{\boldsymbol{u}}} + \bar{\boldsymbol{N}}\dot{\bar{\hat{\boldsymbol{u}}}}, & \boldsymbol{\tau}_c(\boldsymbol{u}) &= \boldsymbol{C}\boldsymbol{B}\hat{\boldsymbol{u}} - \boldsymbol{C}\boldsymbol{\epsilon}_c^0 + \boldsymbol{\tau}_c^0 + \boldsymbol{C}\bar{\boldsymbol{B}}\bar{\hat{\boldsymbol{u}}}, \end{aligned} \tag{6.9}$$

where the shape function $\boldsymbol{N}$ are functions of the space variables, and the vectors $\hat{\boldsymbol{u}}$ are functions of time. It may appear from our semi-discrete approximate form given by Eqn. 6.9 that this approximation is valid only when the exact solution is 'separable' in $\boldsymbol{x}$ and t. This is not so. Even nonseparable solutions can be approximated by our separable approximation. Substituting Eqn. 6.9 in Eqn. 6.6, we get

$$\boldsymbol{M}\ddot{\hat{\boldsymbol{u}}} + \boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{f}, \tag{6.10}$$

where

$$\boldsymbol{M} = \int_{\Omega} \rho \boldsymbol{N}^t \boldsymbol{N}\, d\Omega,$$
$$\boldsymbol{K} = \int_{\Omega} \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{B}\, d\Omega,$$
$$\boldsymbol{f} = \int_{\Omega} \left[\boldsymbol{N}^t \rho \boldsymbol{b} + \boldsymbol{B}^t \boldsymbol{C} \boldsymbol{\epsilon}_c^0 - \boldsymbol{B}^t \boldsymbol{\tau}_c^0 - \boldsymbol{B}^t \boldsymbol{C} \bar{\boldsymbol{B}} \bar{\hat{\boldsymbol{u}}} - \rho \boldsymbol{N}^t \bar{\boldsymbol{N}} \ddot{\bar{\hat{\boldsymbol{u}}}} \right] d\Omega + \int_{\Gamma_t} \boldsymbol{N}^t \bar{\boldsymbol{t}}\, d\Gamma.$$

Since any system is not perfectly elastic, the effects of energy dissipation have to be incorporated in Eqn. (6.10). These effects are incorporated by modifying Eqn. (6.10) to

$$\boldsymbol{M} \ddot{\hat{\boldsymbol{u}}} + \boldsymbol{C} \dot{\hat{\boldsymbol{u}}} + \boldsymbol{K} \hat{\boldsymbol{u}} = \boldsymbol{f}, \tag{6.11}$$

where $\boldsymbol{C}$ is the damping matrix. Eqn. 6.11 is solved subject to the initial conditions $\hat{\boldsymbol{u}}(0) = \hat{\boldsymbol{u}}_0$ and $\dot{\hat{\boldsymbol{u}}}(0) = \dot{\hat{\boldsymbol{u}}}_0$.

The damping matrix, $\boldsymbol{C}$, is positive definite. To see this in the purely mechanical theory where thermal effects are ignored, consider the first law of thermodynamics

$$\frac{d}{dt} \int_{\Omega} \left(\frac{1}{2} \rho \boldsymbol{v} \cdot \boldsymbol{v} + e \right) d\Omega = \int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{v}\, d\Gamma + \int_{\Omega} \rho \boldsymbol{b} \cdot \boldsymbol{v}\, d\Omega.$$

If one considers the loading to be periodic with period T, then

$$\begin{aligned} \int_{t_1}^{t_1+T} \left[\int_{\Gamma_t} \bar{\boldsymbol{t}} \cdot \boldsymbol{v}\, d\Gamma + \int_{\Omega} \rho \boldsymbol{b} \cdot \boldsymbol{v}\, d\Omega \right] dt &= \left[\int_{\Omega} \left(\frac{1}{2} \rho \boldsymbol{v} \cdot \boldsymbol{v} + e \right) d\Omega \right]_{t_1}^{t_1+T} \\ &= \left[\int_{\Omega} e\, d\Omega \right]_{t_1}^{t_1+T} \\ &> 0, \end{aligned}$$

since the kinetic energy has the same value at times t_1 and $t_1 + T$, and since the internal energy increases due to dissipation (the contribution to e due to the elastic strain energy, however, is the same at times t_1 and $t_1 + T$). Now taking the dot product of the finite element equations

$$\begin{aligned} \boldsymbol{f} \cdot \dot{\hat{\boldsymbol{u}}} &= \left(\boldsymbol{M} \ddot{\hat{\boldsymbol{u}}} + \boldsymbol{C} \dot{\hat{\boldsymbol{u}}} + \boldsymbol{K} \hat{\boldsymbol{u}} \right) \cdot \dot{\hat{\boldsymbol{u}}} \\ &= \frac{1}{2} \frac{d}{dt} \left(\dot{\hat{\boldsymbol{u}}} \cdot \boldsymbol{M} \dot{\hat{\boldsymbol{u}}} + \hat{\boldsymbol{u}} \cdot \boldsymbol{K} \hat{\boldsymbol{u}} \right) + \dot{\hat{\boldsymbol{u}}} \cdot \boldsymbol{C} \dot{\hat{\boldsymbol{u}}}. \end{aligned}$$

Integrating the above equation between the limits t_1 and $t_1 + T$ and using the fact that the kinetic and strain energies are the same at times t_1 and $t_1 + T$, we get

$$\int_{t_1}^{t_1+T} \dot{\hat{\boldsymbol{u}}} \cdot \boldsymbol{C} \dot{\hat{\boldsymbol{u}}}\, dt = \int_{t_1}^{t_1+T} \boldsymbol{f} \cdot \dot{\hat{\boldsymbol{u}}}\, dt > 0.$$

Since the above equation holds for all choices of $\dot{\hat{\boldsymbol{u}}}$, t_1 and T, we conclude that $\boldsymbol{C}$ is positive definite.

There is no systematic technique to generate the damping matrix unlike the mass and stiffness matrices. A simple but popular model that is usually used is *Rayleigh* or *proportional* damping in which we assume

$$\boldsymbol{C} = \alpha \boldsymbol{M} + \beta \boldsymbol{K},$$

where α and β are constants which are determined experimentally. We shall discuss more about the damping matrix in Section 6.4.1.

6.2.2 Free vibrations

Consider the special case of undamped harmonic motion.

$$\boldsymbol{M}\ddot{\hat{\boldsymbol{u}}} + \boldsymbol{K}\hat{\boldsymbol{u}} = \boldsymbol{0}. \tag{6.12}$$

Assume a solution for $\hat{\boldsymbol{u}}$ of the form $\hat{\boldsymbol{u}} = \tilde{\boldsymbol{u}}e^{i\omega t}$, where $\tilde{\boldsymbol{u}}$ is the set of constant values at the nodes called as the modal vector, and ω is known as the natural frequency. Substituting this solution in Eqn. 6.12, we get

$$(\boldsymbol{K} - \omega^2 \boldsymbol{M})\tilde{\boldsymbol{u}} = \boldsymbol{0}.$$

For a nontrivial solution to the above equation, it is necessary that the determinant of the coefficients of $\tilde{\boldsymbol{u}}$ be zero, i.e.,

$$\det(\boldsymbol{K} - \omega^2 \boldsymbol{M}) = 0.$$

The eigenvalues ω are real. To see this, let ω and the corresponding mode shape be complex-valued. Then

$$\begin{aligned} \tilde{\boldsymbol{u}}^* \cdot \boldsymbol{K}\tilde{\boldsymbol{u}} &= \omega^2 \tilde{\boldsymbol{u}}^* \cdot \boldsymbol{M}\tilde{\boldsymbol{u}}, \\ \tilde{\boldsymbol{u}} \cdot \boldsymbol{K}\tilde{\boldsymbol{u}}^* &= (\omega^2)^* \tilde{\boldsymbol{u}} \cdot \boldsymbol{M}\tilde{\boldsymbol{u}}^*, \end{aligned}$$

where * denotes complex conjugate. Subtracting the first equation from the second, and using the symmetry of $\boldsymbol{K}$ and $\boldsymbol{M}$, we get

$$\tilde{\boldsymbol{u}}^* \cdot \boldsymbol{M}\tilde{\boldsymbol{u}}[\omega^2 - (\omega^2)^*] = 0.$$

Since $\boldsymbol{M} = \int_\Omega \rho \boldsymbol{N}^T \boldsymbol{N}\, d\Omega$, it follows that $\tilde{\boldsymbol{u}}^* \cdot \boldsymbol{M}\tilde{\boldsymbol{u}} = \int_\Omega \rho \boldsymbol{u}^* \cdot \boldsymbol{u}\, d\Omega > 0$, where $\boldsymbol{u} = \boldsymbol{N}\tilde{\boldsymbol{u}}$. Thus,

$$\omega^2 = (\omega^2)^*,$$

i.e., ω^2 is real-valued. Since now $\tilde{\boldsymbol{u}} \cdot \boldsymbol{K}\tilde{\boldsymbol{u}} = \omega^2 \tilde{\boldsymbol{u}} \cdot \boldsymbol{M}\tilde{\boldsymbol{u}}$, and since $\boldsymbol{M}$ is positive definite and $\boldsymbol{K}$ is positive semi-definite, we have $\omega^2 \geq 0$. If the order of $\boldsymbol{K}$ and $\boldsymbol{M}$ is n, then there are n natural frequencies and correspondingly n natural mode shapes or eigenvectors.

Assuming that all eigenvalues are distinct, the eigenvectors are orthogonal in the following sense:

$$\begin{aligned} \tilde{\boldsymbol{u}}_i \cdot \boldsymbol{K}\tilde{\boldsymbol{u}}_j &= 0, \quad i \neq j, \\ \tilde{\boldsymbol{u}}_i \cdot \boldsymbol{M}\tilde{\boldsymbol{u}}_j &= 0, \quad i \neq j. \end{aligned}$$

The proof is as follows. For the i'th and j'th modes, we have

$$\boldsymbol{K}\tilde{\boldsymbol{u}}_i = \omega_i^2 \boldsymbol{M}\tilde{\boldsymbol{u}}_i,$$
$$\boldsymbol{K}\tilde{\boldsymbol{u}}_j = \omega_j^2 \boldsymbol{M}\tilde{\boldsymbol{u}}_j.$$

Hence,

$$\tilde{\boldsymbol{u}}_j^t \boldsymbol{K}\tilde{\boldsymbol{u}}_i = \omega_i^2 \tilde{\boldsymbol{u}}_j^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i, \tag{6.13}$$
$$\tilde{\boldsymbol{u}}_i^t \boldsymbol{K}\tilde{\boldsymbol{u}}_j = \omega_j^2 \tilde{\boldsymbol{u}}_i^t \boldsymbol{M}\tilde{\boldsymbol{u}}_j. \tag{6.14}$$

Since $\boldsymbol{K}$ and $\boldsymbol{M}$ are symmetric matrices, we have $\tilde{\boldsymbol{u}}_j^t \boldsymbol{K}\tilde{\boldsymbol{u}}_i = \tilde{\boldsymbol{u}}_i^t \boldsymbol{K}\tilde{\boldsymbol{u}}_j$ and $\tilde{\boldsymbol{u}}_j^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i = \tilde{\boldsymbol{u}}_i^t \boldsymbol{M}\tilde{\boldsymbol{u}}_j$. Subtracting Eqn. 6.14 from Eqn. 6.13, we get

$$0 = (\omega_i^2 - \omega_j^2)\tilde{\boldsymbol{u}}_j^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i.$$

For $i \neq j$, $\omega_i \neq \omega_j$, we get

$$\tilde{\boldsymbol{u}}_j^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i = 0.$$

Hence, from Eqn. 6.13 $\tilde{\boldsymbol{u}}_j^t \boldsymbol{K}\tilde{\boldsymbol{u}}_i = 0$. For $i = j$, we get

$$\tilde{\boldsymbol{u}}_i^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i = E_M,$$
$$\tilde{\boldsymbol{u}}_i^t \boldsymbol{K}\tilde{\boldsymbol{u}}_i = \omega_i^2 E_M,$$

where E_M is a constant. Usually, $\tilde{\boldsymbol{u}}$ is normalized so that $E_M = 1$. Then, we have

$$\tilde{\boldsymbol{u}}_i^t \boldsymbol{M}\tilde{\boldsymbol{u}}_i = 1,$$
$$\tilde{\boldsymbol{u}}_i^t \boldsymbol{K}\tilde{\boldsymbol{u}}_i = \omega_i^2.$$

6.2.3 The Rayleigh quotient

The Rayleigh quotient is defined as

$$Q(\boldsymbol{v}) = \frac{\boldsymbol{v}^t \boldsymbol{K}\boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{M}\boldsymbol{v}},$$

where $\boldsymbol{v}$ is an arbitrary vector. One of the important properties of the Rayleigh quotient is

$$\omega_1^2 \leq Q(\boldsymbol{v}) \leq \omega_n^2,$$

where the frequencies ω_1, ω_2,..., ω_n are assumed to be ordered such that

$$\omega_1 \leq \omega_2 \leq \ldots \leq \omega_n.$$

The proof is as follows. Any vector $\boldsymbol{v}$ can be expressed in terms of the normalized eigenvector $\tilde{\boldsymbol{u}}_i$ as

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \tilde{\boldsymbol{u}}_i.$$

Substituting this expansion in the expression for $Q(\boldsymbol{v})$, and using the orthogonality property of $\tilde{\boldsymbol{u}}$, we get

$$\begin{aligned} Q(\boldsymbol{v}) &= \frac{\alpha_1^2\omega_1^2 + \alpha_2^2\omega_2^2 + \cdots + \alpha_n^2\omega_n^2}{\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2} \\ &= \omega_1^2 \left[\frac{\alpha_1^2 + \alpha_2^2 \left(\frac{\omega_2}{\omega_1}\right)^2 + \cdots + \alpha_n^2 \left(\frac{\omega_n}{\omega_1}\right)^2}{\alpha 1^2 + \alpha_2^2 + \cdots + \alpha_n^2} \right] \\ &\geq \omega_1^2. \end{aligned}$$

Similarly,

$$\begin{aligned} Q(\boldsymbol{v}) &= \omega_n^2 \left[\frac{\alpha_1^2 \left(\frac{\omega_1}{\omega_n}\right)^2 + \alpha_2^2 \left(\frac{\omega_2}{\omega_n}\right)^2 + \cdots + \alpha_n^2}{\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2} \right] \\ &\leq \omega_n^2. \end{aligned}$$

Thus, we get $\omega_1^2 \leq Q(\boldsymbol{v}) \leq \omega_n^2$. Choosing $\boldsymbol{v} = \alpha_1\tilde{\boldsymbol{u}}_1$ and $\boldsymbol{v} = \alpha_n\tilde{\boldsymbol{u}}_n$, we get

$$\begin{aligned} \omega_1^2 &= \min_{\boldsymbol{v}\in\Re^n} Q(\boldsymbol{v}), \\ \omega_n^2 &= \max_{\boldsymbol{v}\in\Re^n} Q(\boldsymbol{v}). \end{aligned}$$

The maximum eigenvalue can be estimated from the result

$$\lambda_n^h \equiv \omega_n^2 \leq \max_e (\omega_{\max}^2)_e,$$

where $(\omega_{\max}^2)_e$ is the maximum eigenvalue of element e found by solving the eigenvalue problem for that element:

$$(\boldsymbol{K}^{(e)} - \lambda^e \boldsymbol{M}^{(e)})\hat{\boldsymbol{u}}^{(e)} = 0.$$

6.2.4 Consistent mass matrices for various element types

In this section, we shall evaluate the consistent element mass matrix given by

$$\boldsymbol{M}^{(e)} = \int_{\Omega_e} \rho \boldsymbol{N}^t \boldsymbol{N}\, d\Omega, \tag{6.15}$$

for the various element types that we have studied so far. We shall assume that $\rho =$ constant.

One-dimensional bar element

The shape function matrix for a linear element is given by

$$\boldsymbol{N} = \begin{bmatrix} \frac{1-\xi}{2} & \frac{1+\xi}{2} \end{bmatrix}.$$

Hence, the consistent element mass matrix is

$$\begin{aligned}\boldsymbol{M}^{(e)} &= \rho_e \int_{\Omega_e} \boldsymbol{N}^t \boldsymbol{N} A \, dx \\ &= \frac{\rho_e A_e l_e}{2} \int_{-1}^{1} \boldsymbol{N}^t \boldsymbol{N} \, d\xi \\ &= \frac{\rho_e A_e l_e}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.\end{aligned}$$

For a quadratic bar element, we get

$$\boldsymbol{M}^{(e)} = \frac{\rho_e A_e l_e}{30} \begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4 \end{bmatrix}.$$

Truss element

If u_1-u_4 represent the degrees of freedom of a truss element, then

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} N_1 & 0 & N_2 & 0 \\ 0 & N_1 & 0 & N_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix},$$

thus yielding

$$\boldsymbol{N} = \begin{bmatrix} N_1 & 0 & N_2 & 0 \\ 0 & N_1 & 0 & N_2 \end{bmatrix},$$

where $N_1 = (1 - \xi)/2$ and $N_2 = (1 + \xi)2$. Substituting in Eqn. 6.15, we get

$$\boldsymbol{M}^{(e)} = \frac{\rho_e A_e l_e}{6} \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}.$$

Beam element

Using the Hermite shape function matrix, i.e.,

$$\boldsymbol{N} = \begin{bmatrix} N_1 & \frac{l_e}{2} N_2 & N_3 & \frac{l_e}{2} N_4 \end{bmatrix},$$

where

$$N_1 = \frac{1}{4}(2 - 3\xi + \xi^3),$$
$$N_2 = \frac{1}{4}(1 - \xi - \xi^2 + \xi^3),$$
$$N_3 = \frac{1}{4}(2 + 3\xi - \xi^3),$$
$$N_4 = \frac{1}{4}(-1 - \xi + \xi^2 + \xi^3).$$

Substituting in Eqn. 6.15, we get

$$\begin{aligned}\boldsymbol{M}^{(e)} &= \int_{-1}^{1} \boldsymbol{N}^t \boldsymbol{N} \frac{\rho_e A_e l_e}{2}\, d\xi \\ &= \frac{\rho_e A_e l_e}{420} \begin{bmatrix} 156 & 22l_e & 54 & -13l_e \\ 22l_e & 4l_e^2 & 13l_e & -3l_e^2 \\ 54 & 13l_e & 156 & -22l_e \\ -13l_e & -3l_e^2 & -22l_e & 4l_e^2 \end{bmatrix}\end{aligned}$$

Frame element

In the body coordinate system, the mass matrix is a combination of the bar element and the beam element:

$$\boldsymbol{M}_b^{(e)} = \begin{bmatrix} 2a & 0 & 0 & a & 0 & 0 \\ 0 & 156b & 22l_e b & 0 & 54b & -13l_e b \\ 0 & 22l_e b & 4l_e^2 b & 0 & 13l_e b & -3l_e^2 b \\ a & 0 & 0 & 2a & 0 & 0 \\ 0 & 54b & 13l_e b & 0 & 156b & -22l_e b \\ 0 & -13l_e b & -3l_e^2 b & 0 & -22l_e b & 4l_e^2 b \end{bmatrix}$$

where $a = \rho_e A_e l_e/6$ and $b = \rho_e A_e l_e/420$. In the global coordinate system, the element mass matrix is given by

$$\boldsymbol{M}^{(e)} = \boldsymbol{L}^t \boldsymbol{M}_b^{(e)} \boldsymbol{L},$$

where

$$\boldsymbol{L} = \begin{bmatrix} l & m & 0 & 0 & 0 & 0 \\ -m & l & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l & m & 0 \\ 0 & 0 & 0 & -m & l & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

CST element

The shape function matrix is given by

$$\boldsymbol{N} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 \end{bmatrix},$$

where $N_1 = \xi = A_1/A$, $N_2 = \eta = A_2/A$ and $N_3 = 1 - \xi - \eta = A_3/A$. The mass matrix is

$$\boldsymbol{M}^{(e)} = \rho_e t_e \int_A \boldsymbol{N}^t \boldsymbol{N}\, dA = \frac{\rho t_e A_e}{12} \begin{bmatrix} 2 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 2 \end{bmatrix}.$$

Axisymmetric triangular element

The mass matrix is

$$\begin{aligned} \boldsymbol{M}^{(e)} &= 2\pi\rho_e \int_A \boldsymbol{N}^t \boldsymbol{N} r\, dA \\ &= 2\pi\rho_e \int_A (N_1 r_1 + N_2 r_2 + N_3 r_3) \boldsymbol{N}^t \boldsymbol{N}\, dA \\ &= \frac{\pi\rho_e A_e}{10} \begin{bmatrix} \frac{4r_1}{3} + 2\bar{r} & 0 & 2\bar{r} - \frac{r_3}{3} & 0 & 2\bar{r} - \frac{r_2}{3} & 0 \\ 0 & \frac{4r_1}{3} + 2\bar{r} & 0 & 2\bar{r} - \frac{r_3}{3} & 0 & 2\bar{r} - \frac{r_2}{3} \\ 2\bar{r} - \frac{r_3}{3} & 0 & \frac{4r_2}{3} + 2\bar{r} & 0 & 2\bar{r} - \frac{r_1}{3} & 0 \\ 0 & 2\bar{r} - \frac{r_3}{3} & 0 & \frac{4r_2}{3} + 2\bar{r} & 0 & 2\bar{r} - \frac{r_1}{3} \\ 2\bar{r} - \frac{r_2}{3} & 0 & 2\bar{r} - \frac{r_1}{3} & 0 & \frac{4r_3}{3} + 2\bar{r} & 0 \\ 0 & 2\bar{r} - \frac{r_2}{3} & 0 & 2\bar{r} - \frac{r_1}{3} & 0 & \frac{4r_3}{3} + 2\bar{r} \end{bmatrix}, \end{aligned}$$

where $\bar{r} = (r_1 + r_2 + r_3)/3$.

Q4 element

The shape function matrix is given by

$$\boldsymbol{N} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 \end{bmatrix},$$

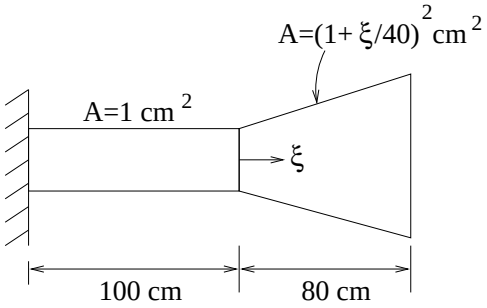


Figure 6.1: Calculation of the natural frequencies of a structure

where $N_i = (1 + \xi\xi_i)(1 + \eta\eta_i)/4$. The mass matrix is

$$\boldsymbol{M}^{(e)} = \rho_e t_e \int_{-1}^{1}\int_{-1}^{1} \boldsymbol{N}^t \boldsymbol{N} (\det \boldsymbol{J})\, d\xi\, d\eta.$$

The integration has to be performed numerically. Usually the same order of integration as for the stiffness matrix is appropriate.

Note: The sum of the elements of the mass matrix for the bar element is equal to the mass of the element. For the truss, CST, axisymmetric and Q4 elements the sum of the elements of the mass matrix is equal to twice the mass of the element since the linear momentum equation is a vectorial equation with components along each of the coordinate axes.

Example

Consider the structure shown in Fig. 6.1 with the material properties $E = 60 \times 10^9$ Pa and $\rho = 7500$ kg/m^3. Using two linear elements to model the structure, the mass and stiffness matrices are

$$\boldsymbol{M} = 10^2 \begin{bmatrix} 0.0025 & 0.00125 & 0 \\ 0.00125 & 0.0073 & 0.0042 \\ 0 & 0.0042 & 0.0128 \end{bmatrix}$$

$$\boldsymbol{K} = 10^6 \begin{bmatrix} 6 & -6 & 0 \\ -6 & 38.5 & -32.5 \\ 0 & -32.5 & 32.5 \end{bmatrix}$$

We have $\hat{u}_1 = 0$. Hence, the eigenvalue problem corresponding to $\hat{u}_2$ and $\hat{u}_3$ is solved using

$$\begin{vmatrix} 38.5 - \lambda(7.3 \times 10^{-7}) & -32.5 - \lambda(4.2 \times 10^{-7}) \\ -32.5 - \lambda(4.2 \times 10^{-7}) & 32.5 - \lambda(1.28 \times 10^{-6}) \end{vmatrix} = 0.$$

On solving the above equation, we get $\lambda_1 = 1.3035 \times 10^8$ and $\lambda_2 = 1.97 \times 10^6$. Thus, $\omega_1 = \sqrt{\lambda_1} = 11417.3$ rad/s and $\omega_2 = \sqrt{\lambda_2} = 1404.854$ rad/s. The

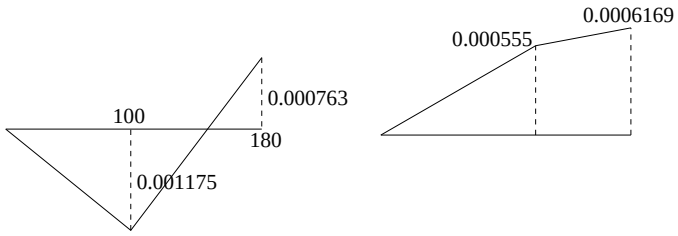


Figure 6.2: Mode shapes

frequency in Hertz is obtained by dividing ω by 2π. We get $f_1 = 1817$ Hz and $f_2 = 223.6$ Hz. Substituting λ_1 in the eigenvalue problem, we get

$$10^8 \begin{bmatrix} -0.56659 & -0.87249 \\ -0.87249 & -1.2835 \end{bmatrix} \begin{bmatrix} \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which on solving yields $\hat{u}_2 = -1.5399\hat{u}_3$. As mentioned, the normalization of the mode shape is done with respect to the mass matrix, i.e., $\tilde{\boldsymbol{u}}^t \boldsymbol{M} \tilde{\boldsymbol{u}} = 1$, or

$$\begin{bmatrix} 0 & -1.5399\hat{u}_3 & \hat{u_3} \end{bmatrix} \begin{bmatrix} 0.25 & 0.125 & 0 \\ 0.125 & 0.73 & 0.42 \\ 0 & 0.42 & 1.28 \end{bmatrix} \begin{bmatrix} 0 \\ -1.5399\hat{u}_3 \\ \hat{u_3} \end{bmatrix} = 1.$$

On solving the above equation, we get $\hat{u}_3 = \pm 0.76304$ m. Thus, $\hat{u}_2 = \mp 1.175$ m. Similarly, corresponding to λ_2, we get $\hat{u}_2 = 0.5549$ m and $\hat{u}_3 = 0.6171$ m. The two mode shapes are shown in Fig. 6.2.

6.3 Unsteady Heat Conduction

In this section, we consider the transient heat conduction equation (compare with the steady state treatment in Section 1.5)

$$\rho c \frac{\partial T}{\partial t} = \boldsymbol{\nabla} \cdot (\boldsymbol{k} \boldsymbol{\nabla} T) + Q \text{ in } \Omega,$$

where c is the specific heat (J/kgC) and ρ is the density, subject to the boundary conditions

$$T = \bar{T} \text{ on } \Gamma_T,$$
$$h(T - T_\infty) + (\boldsymbol{k} \boldsymbol{\nabla} T) \cdot \boldsymbol{n} = \bar{q} \text{ on } \Gamma_q,$$

and subject to the initial condition

$$T(\boldsymbol{x}) = T_0(\boldsymbol{x}) \text{ at } t = 0.$$

The weak form of this equation is given by
Find $T \in L_T$ such that

$$\int_\Omega \rho c \frac{\partial T}{\partial t} v \, d\Omega + \int_\Omega \boldsymbol{\nabla} v^t \boldsymbol{k} \boldsymbol{\nabla} T \, d\Omega + \int_{\Gamma_q} v h T \, d\Gamma = \int_\Omega v Q \, d\Omega + \int_{\Gamma_q} v(\bar{q} + hT_\infty) \, d\Gamma \quad \forall v \in V_T,$$

subject to the initial condition $T(\boldsymbol{x}, 0) = T_0(\boldsymbol{x})$. The finite element formulation is carried out similar to that in the elasticity case. Discretizing the temperature as

$$T = \sum \boldsymbol{N}(\boldsymbol{x}) \boldsymbol{u}(t) + \bar{\boldsymbol{N}}(\boldsymbol{x}) \bar{\boldsymbol{u}}(t), \tag{6.16}$$

and substituting in the variational formulation, we get

$$\boldsymbol{M}\dot{\boldsymbol{u}} + \boldsymbol{K}\boldsymbol{u} = \boldsymbol{f}$$
$$\text{subject to } \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

where (with $\boldsymbol{\nabla} T = \boldsymbol{B}\boldsymbol{u}$)

$$\boldsymbol{M} = \int_\Omega \rho c \boldsymbol{N}^t \boldsymbol{N} \, d\Omega,$$
$$\boldsymbol{K} = \int_\Omega \boldsymbol{B}^t \boldsymbol{K} \boldsymbol{B} \, d\Omega + \int_{\Gamma_q} h \boldsymbol{N}^t \boldsymbol{N} \, d\Gamma,$$
$$\boldsymbol{f} = \int_\Omega \boldsymbol{N}^t Q \, d\Omega + \int_{\Gamma_q} \boldsymbol{N}^t (\bar{q} + hT_\infty) \, d\Gamma - \left(\int_\Omega \boldsymbol{B}^t \boldsymbol{k} \bar{\boldsymbol{B}} \, d\Omega \right) \bar{\boldsymbol{u}} - \int_\Omega \rho c \boldsymbol{N}^t \bar{\boldsymbol{N}} \dot{\bar{\boldsymbol{u}}} \, d\Omega,$$

Note the similarity of the form of $\boldsymbol{M}$ with the form of the mass matrix (also denoted by $\boldsymbol{M}$) in the dynamics formulation. In particular it is symmetric and positive definite. In practice, diagonal or 'lumped' mass matrices are used instead of the consistent mass matrix derived above, because they lead to economical time-integration schemes (so called 'explicit methods').

The eigenvalue problem can be derived by assuming

$$\boldsymbol{u} = \tilde{\boldsymbol{u}} e^{-\lambda t},$$

and substituting in $\boldsymbol{M}\dot{\boldsymbol{u}} + \boldsymbol{K}\boldsymbol{u} = \boldsymbol{0}$, to get

$$(\boldsymbol{K} - \lambda \boldsymbol{M})\tilde{\boldsymbol{u}} = \boldsymbol{0},$$

which is similar in form to the eigenvalue problem for the dynamics problem.

Example:
Consider a plane wall initially at uniform temperature T_0, which is suddenly exposed to a fluid at temperature T_∞. The governing equation is

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t},$$

where $\alpha = k/(\rho c)$ is the diffusion coefficient. The initial condition is $T(x, 0) = T_0$. Consider two sets of boundary conditions:

1. $T(0, t) = T_\infty$, $T(L, t) = T_\infty$ for $t > 0$.

2. $T(0,t) = T_\infty,\ kA\frac{\partial T}{\partial x} + hA(T - T_\infty)\big|_{x=L} = 0.$

To find the solution, we define some nondimensional parameters:

$$\bar{x} = \frac{x}{L};\ \bar{t} = \frac{\alpha t}{L^2};\ u = \frac{T - T_\infty}{T_0 - T_\infty}.$$

The differential equations and boundary conditions in terms of these parameters are (dropping the 'bars' on x and t for notational convenience)

$$-\frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial t} = 0,$$

1. $u(0,t) = 0,\ u(1,t) = 0,\ u(x,0) = 1.$
2. $u(0,t) = 0,\ \frac{\partial u}{\partial x} + Hu\big|_{x=1} = 0,\ u(x,0) = 1,$ where $H = hL/k.$

Assume $u = e^{-\lambda t}U(x)$. Substituting in the governing differential equation, we get

$$-\frac{d^2 u}{dx^2} - \lambda U = 0.$$

Solving the above equation, we get

$$u = \left[c_1 \sin\sqrt{\lambda}x + c_2 \cos\sqrt{\lambda}x\right] e^{-\lambda t}.$$

1. For the first set of boundary conditions, we get $c_2 = 0,\ \sqrt{\lambda} = n\pi$. Thus, the exact solution is

$$u = \sum_{n=1}^{\infty} \left[c_n \sin n\pi x\right] e^{-n^2\pi^2 t}.$$

The constants c_n can be determined from the initial condition, and the orthogonality of the sine functions:

$$1 = \sum_{n=1}^{\infty} c_n \sin n\pi x$$

For a mesh of 2 linear elements, we have

$$\left\{ 2\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} - \frac{\lambda}{12}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 2 \end{bmatrix} \right\} \hat{\boldsymbol{u}} = \begin{bmatrix} Q_1 \\ 0 \\ Q_2 \end{bmatrix},$$

subject to the boundary conditions $u(0) = u(1) = 0$. Solving, we get

$$(4 - \frac{4}{12}\lambda)\hat{u}_2 = 0,$$

or $\lambda_1 = 12$.

For a mesh of one quadratic element, we get

$$\frac{16}{3} - \frac{16\lambda}{30} = 0,$$

or $\lambda_1 = 10$. Summarizing, we have

- for 2 linear elements, $\lambda_1 = 12$.
- for 1 quadratic element, $\lambda_1 = 10$.
- exact, $\lambda_1 = \pi^2$.

The corresponding eigenfunctions are

- for 2 linear elements,

$$U(x) = \begin{cases} \frac{x}{h} & 0 \le x \le 0.5 \\ 1 - \frac{x}{h} & 0.5 \le x \le 1 \end{cases}$$

- for 1 quadratic element,

$$U(x) = 4\frac{x}{h}\left(1 - \frac{x}{h}\right).$$

- exact, $U_n(x) = \sin n\pi x$.

2. Assuming $H = 1$, the second set of boundary conditions yields $U_n(x) = \sin\sqrt{\lambda_n}x$, where λ_n satisfies

$$1 + \sqrt{\lambda_n}\cot\sqrt{\lambda_n} = 0.$$

For a mesh of two linear elements, we get

$$\left\{ \begin{bmatrix} 4 & -2 \\ -2 & 3 \end{bmatrix} - \frac{\lambda}{12} \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \right\} \begin{bmatrix} \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

On solving, we get $\lambda_1 = 4.49$, $\lambda_2 = 36.65$. The first eigenmode is $U^{(1)} = (0, 0.6881, 0.7256)$. The exact eigenvalues are $\lambda_1 = 4.1159$ and $\lambda_2 = 24.1393$.

Note that as the mesh is refined, not only do we increase the number of eigenvalues, but also improve the accuracy of the preceding ones. Note also that the convergence of the numerical eigenvalues to the exact ones is from above, i.e., the finite element solution provides an upper bound to the exact eigenvalues because the finite element system is stiffer resulting in higher eigenvalues or frequencies.

6.3.1 Algorithms for parabolic problems

In this section, we discuss algorithms for solving the governing equation

$$\boldsymbol{M}\dot{\boldsymbol{u}} + \boldsymbol{K}\boldsymbol{u} = \boldsymbol{f}, \tag{6.17}$$

subject to the initial condition $\boldsymbol{u}(0) = \boldsymbol{u}_0$. First we state the exact solution to these set of equations. Multiplying Eqn. (6.17) by $e^{\boldsymbol{M}^{-1}\boldsymbol{K}t}\boldsymbol{M}^{-1}$, we get

$$\frac{d}{dt}\left[e^{\boldsymbol{M}^{-1}\boldsymbol{K}t}\boldsymbol{u}\right] = e^{\boldsymbol{M}^{-1}\boldsymbol{K}t}\boldsymbol{M}^{-1}\boldsymbol{f}.$$

Integrating this equation either between the limits $[0, t]$ or $[t, t+\Delta t]$, we get

$$\boldsymbol{u}(t) = e^{-\boldsymbol{M}^{-1}\boldsymbol{K}t}\boldsymbol{u}_0 + \int_0^t e^{\boldsymbol{M}^{-1}\boldsymbol{K}(\xi-t)}\boldsymbol{M}^{-1}\boldsymbol{f}(\xi)\,d\xi, \tag{6.18a}$$

$$\boldsymbol{u}(t+\Delta t) = e^{-\boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}\boldsymbol{u}(t) + \int_t^{t+\Delta t} e^{\boldsymbol{M}^{-1}\boldsymbol{K}(\xi-t-\Delta t)}\boldsymbol{M}^{-1}\boldsymbol{f}(\xi)\,d\xi. \tag{6.18b}$$

The trapezoidal family of methods to solve Eqn. 6.17 is given by

$$\boldsymbol{M}\boldsymbol{v}_{n+1} + \boldsymbol{K}\boldsymbol{u}_{n+1} = \boldsymbol{f}_{n+1}, \tag{6.19}$$

$$\boldsymbol{M}\boldsymbol{v}_n + \boldsymbol{K}\boldsymbol{u}_n = \boldsymbol{f}_n, \tag{6.20}$$

$$\frac{\boldsymbol{u}_{n+1} - \boldsymbol{u}_n}{\Delta t} = (1-\alpha)\boldsymbol{v}_n + \alpha\boldsymbol{v}_{n+1}. \tag{6.21}$$

where $\boldsymbol{u}_n$ and $\boldsymbol{v}_n$ are approximations to $\boldsymbol{u}(t_n)$ and $\dot{\boldsymbol{u}}(t_n)$, $\boldsymbol{f}_{n+1} = \boldsymbol{f}(t_{n+1})$, Δt is the time step and $\alpha \in [0, 1]$.

α	Method
0	Forward differences; forward Euler
$\frac{1}{2}$	Trapezoidal rule; midpoint rule; Crank-Nicolson
1	Backward differences; backward Euler

The problem is to determine $\boldsymbol{u}_{n+1}$ and $\boldsymbol{v}_{n+1}$ given $\boldsymbol{u}_n$ and $\boldsymbol{v}_n$. The initial value $\boldsymbol{v}_0$ is determined from

$$\boldsymbol{M}\boldsymbol{v}_0 = \boldsymbol{f}_0 - \boldsymbol{K}\boldsymbol{u}_0.$$

By multiplying Eqn. (6.19) by α, Eqn. (6.20) by $1-\alpha$, adding the two resulting equations, and using Eqn. (6.21), we get

$$(\boldsymbol{M} + \alpha\Delta t\boldsymbol{K})\boldsymbol{u}_{n+1} - (\boldsymbol{M} - (1-\alpha)\Delta t\boldsymbol{K})\boldsymbol{u}_n = \Delta t\boldsymbol{f}_{n+\alpha}, \tag{6.22}$$

where $\boldsymbol{f}_{n+\alpha} = \alpha\boldsymbol{f}_{n+1} + (1-\alpha)\boldsymbol{f}_n$. After obtaining $\boldsymbol{u}_{n+1}$ from Eqn. (6.22), $\boldsymbol{v}_{n+1}$ is recovered in the post-processing step using Eqn. (6.21).

Note

1. When $\alpha = 0$, the method is said to be explicit. In this case, Eqn. (6.22) reduces to
$$\boldsymbol{M}\boldsymbol{u}_{n+1} = \Delta t\boldsymbol{f}_n + (\boldsymbol{M} - \Delta t\boldsymbol{K})\boldsymbol{u}_n.$$
If $\boldsymbol{M}$ is 'lumped' (i.e., diagonal), the solution for $\boldsymbol{u}_{n+1}$ can be found without the necessity of equation solving.

2. If $\alpha \neq 0$, the method is said to be implicit. In this case, a system of equations needs to be solved. If Δt is constant, only one factorization of $(\boldsymbol{M} + \alpha\Delta t\boldsymbol{K})$ needs to be carried out at $t = 0$.

We now show how a form close to Eqn. (6.22) can be obtained as an approximation to the exact solution given by Eqn. (6.18b). In the derivation, we will use the property that $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}}$ if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ commute. First multiply Eqn. (6.18b) by $e^{\alpha \boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}$ to get

$$e^{\alpha \boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}\boldsymbol{u}_{n+1} = e^{-(1-\alpha)\boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}\boldsymbol{u}_n + \int_t^{t+\Delta t} e^{\boldsymbol{M}^{-1}\boldsymbol{K}(\xi - t - (1-\alpha)\Delta t)}\boldsymbol{M}^{-1}\boldsymbol{f}(\xi)\, d\xi.$$

Make the substitution $h = \xi - t - (1-\alpha)\Delta t$ in the integral to get

$$e^{\alpha \boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}\boldsymbol{u}_{n+1} = e^{-(1-\alpha)\boldsymbol{M}^{-1}\boldsymbol{K}\Delta t}\boldsymbol{u}_n + \int_{-(1-\alpha)\Delta t}^{\alpha\Delta t} e^{\boldsymbol{M}^{-1}\boldsymbol{K}h}\boldsymbol{M}^{-1}\boldsymbol{f}(h)\, dh. \quad (6.23)$$

Assume next that $\boldsymbol{f}$ is approximated by a linear interpolation between the values at t and $t + \Delta t$, i.e.,

$$\boldsymbol{f}(h) = \frac{1}{\Delta t}\left[(\alpha\Delta t - h)\boldsymbol{f}_n + [h + (1-\alpha)\Delta t]\boldsymbol{f}_{n+1}\right]. \quad (6.24)$$

Since Δt and h are small, we approximate the exponential matrices as

$$\begin{aligned} e^{\alpha \boldsymbol{M}^{-1}\boldsymbol{K}\Delta t} &\approx \boldsymbol{I} + \alpha\Delta t\boldsymbol{M}^{-1}\boldsymbol{K}, \\ e^{-(1-\alpha)\boldsymbol{M}^{-1}\boldsymbol{K}\Delta t} &\approx \boldsymbol{I} - (1-\alpha)\Delta t\boldsymbol{M}^{-1}\boldsymbol{K}, \\ e^{\boldsymbol{M}^{-1}\boldsymbol{K}h} &\approx \boldsymbol{I}, \end{aligned}$$

where in the last approximation, we have retained only the first term to get a linear approximation in Δt in the final result. Substituting Eqn. (6.24) and the above approximations into Eqn. (6.23), and carrying out the integration, we get

$$\left[\boldsymbol{I} + \alpha\Delta t\boldsymbol{M}^{-1}\boldsymbol{K}\right]\boldsymbol{u}_{n+1} = \left[\boldsymbol{I} - (1-\alpha)\Delta t\boldsymbol{M}^{-1}\boldsymbol{K}\right]\boldsymbol{u}_n + \frac{1}{2}\Delta t\boldsymbol{M}^{-1}\left[\boldsymbol{f}_n + \boldsymbol{f}_{n+1}\right].$$

Multiplying the above equation by $\boldsymbol{M}$, we obtain a form close to Eqn. (6.22).

6.3.2 Analysis of the generalized trapezoidal method

An algorithm is convergent if for t_n fixed and $\Delta t = t_n/n$

$$\boldsymbol{u}_n \to \boldsymbol{u}(t_n) \text{ as } \Delta t \to 0.$$

To establish the convergence of an algorithm, two additional notions must be considered: stability and consistency. We shall prove that stability and consistency implies convergence. We use the 'modal' approach (also known as spectral or Fourier analysis) in which the problem is decomposed into n uncoupled scalar equations. The first step in the analysis is to perform the reduction to single-degree-of-freedom (SDOF) form.

Modal reduction to SDOF form

The essential property used in reducing to SDOF form is the orthogonality of the eigenvectors of the associated eigenvalue problem

$$(\boldsymbol{K} - \lambda_i \boldsymbol{M})\boldsymbol{\Psi}_i = 0; \quad i = 1, 2, \ldots, n$$

where n is the number of degrees of freedom and the λ_i are assumed to be ordered such that $0 \leq \lambda_1 \leq \ldots \leq \lambda_n$. By the orthonormality properties proved in Section 6.2.2, we have

$$\begin{aligned} \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{\Psi}_j &= \delta_{ij}, \\ \boldsymbol{\Psi}_i^t \boldsymbol{K} \boldsymbol{\Psi}_j &= \lambda_i \delta_{ij}, \end{aligned} \tag{6.25}$$

with no sum on i. The eigenvectors $\boldsymbol{\Psi}_i|_{i=1}^n$ constitute a basis for $\Re^n$, i.e., any element of $\Re^n$ can be written as a linear combination of $\boldsymbol{\Psi}_i|_{i=1}^n$. Therefore,

$$\begin{aligned} \boldsymbol{u}(t) &= \sum_{j=1}^{n} u_j(t) \boldsymbol{\Psi}_j, \\ \dot{\boldsymbol{u}}(t) &= \sum_{j=1}^{n} \dot{u}_j(t) \boldsymbol{\Psi}_j, \end{aligned} \tag{6.26}$$

where

$$\begin{aligned} u_i(t) &= \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}(t), \\ \dot{u}_i(t) &= \boldsymbol{\Psi}_i^t \boldsymbol{M} \dot{\boldsymbol{u}}(t). \end{aligned}$$

Substituting Eqn. 6.26 in Eqn. 6.17, and premultiplying by $\boldsymbol{\Psi}_i^t$ yields

$$\sum_{j=1}^{n} (\dot{u}_j \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{\Psi}_j + u_j \boldsymbol{\Psi}_i^t \boldsymbol{K} \boldsymbol{\Psi}_j) = \boldsymbol{\Psi}_i^t \boldsymbol{f} \quad i = 1, 2, \ldots, n.$$

Using Eqns. 6.25, we get

$$\dot{u}_i + \lambda_i u_i = \boldsymbol{\Psi}_i^t \boldsymbol{f}(t), \quad i = 1, 2, \ldots, n.$$

We shall denote $\boldsymbol{\Psi}_i^t \boldsymbol{f}(t)$ by f_i. The initial condition is

$$u_i(0) = \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}(0) = \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}_0.$$

We shall denote $\boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}_0$ by $u_{0(i)}$. Omitting the subscript i, the typical modal initial-value problem is

$$\dot{u} + \lambda u = f \tag{6.27}$$

subject to $u(0) = u_0$.

Similar to the decomposition of the semi-discrete equation, we now decompose the generalized trapezoidal algorithm using

$$\boldsymbol{u}_n = \sum_{j=1}^{n} u_{n(j)} \boldsymbol{\Psi}_j,$$

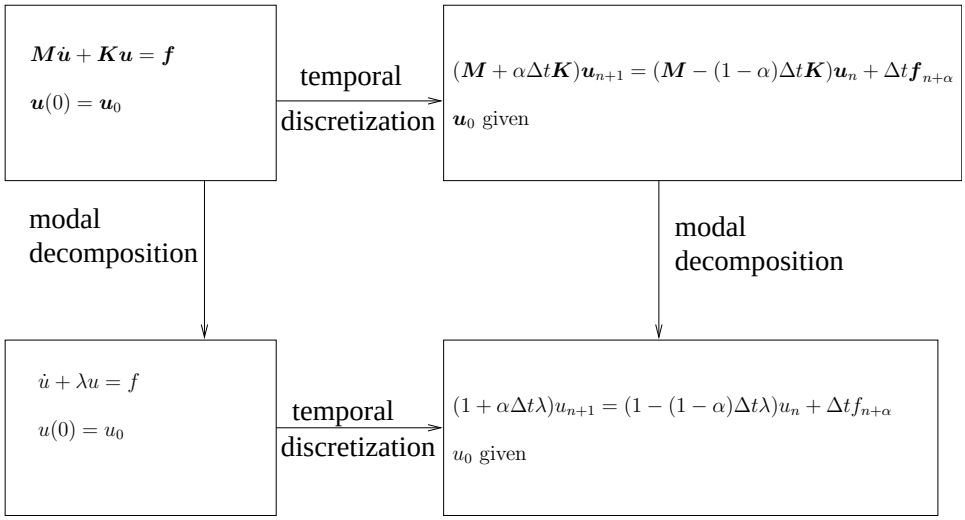


Figure 6.3: Temporal discretization and modal decomposition for parabolic problems

$$\boldsymbol{u}_{n+1} = \sum_{j=1}^{n} u_{n+1(j)} \boldsymbol{\Psi}_j,$$

where

$$\begin{aligned} u_{n(i)} &= \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}_n, \\ u_{n+1(i)} &= \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}_{n+1}. \end{aligned}$$

Substituting for $\boldsymbol{u}_n$ and $\boldsymbol{u}_{n+1}$ in Eqn. 6.22, and premultiplying by $\boldsymbol{\Psi}_i^t$, we get

$$\sum_{j=1}^{n} \left[u_{n+1(j)} \boldsymbol{\Psi}_i^t (\boldsymbol{M} + \alpha\Delta t \boldsymbol{K}) \boldsymbol{\Psi}_j - u_{n(j)} \boldsymbol{\Psi}_i^t (\boldsymbol{M} - (1-\alpha)\Delta t \boldsymbol{K}) \boldsymbol{\Psi}_j \right] = \Delta t \boldsymbol{\Psi}_i^t \boldsymbol{f}_{n+\alpha},$$

where $\boldsymbol{f}_{n+\alpha} = (1-\alpha)\boldsymbol{f}_n + \alpha_{\mathbf{n+1}}$. Using Eqn. 6.25, we get the modal equation as

$$(1 + \alpha\Delta t\lambda_i) u_{n+1(i)} = (1 - (1-\alpha)\Delta t\lambda_i) u_{n(i)} + \Delta t f_{n+\alpha(i)},$$

subject to the given initial condition $u_{0(i)} = \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{u}_0$. Omitting the subscript i, we get

$$(1 + \alpha\Delta t\lambda) u_{n+1} = (1 - (1-\alpha)\Delta t\lambda) u_n + \Delta t f_{n+\alpha}, \tag{6.28}$$

with u_0 given.

Note that applying the trapezoidal rule to the SDOF equation given by Eqn. 6.27 also results in the temporally discretized SDOF problem (see Fig. 6.3).

The convergence of the original problem is equivalent to the convergence of the SDOF problem, i.e.,

$$\boldsymbol{u}_n \to \boldsymbol{u}(t_n) \Longleftrightarrow u_{n(i)} \to u_i(t_n) \quad \text{for } i = 1, 2, \ldots, n.$$

The proof is as follows. The errors in the original and SDOF system at time t_n are given by

$$\begin{aligned}\boldsymbol{e}(t_n) &= \boldsymbol{u}(t_n) - \boldsymbol{u}_n,\\ e_i(t_n) &= u_i(t_n) - u_{n(i)}.\end{aligned}$$

Hence,

$$\begin{aligned}\boldsymbol{e}^t(t_n)\boldsymbol{M}\boldsymbol{e}(t_n) &= \sum_{i,j=1}^{n}(e_i(t_n)\boldsymbol{\Psi}_i)^t\boldsymbol{M}(e_j(t_n)\boldsymbol{\Psi}_j)\\ &= \sum_{i,j=1}^{n} e_i(t_n)e_j(t_n)\boldsymbol{\Psi}_i^t\boldsymbol{M}\boldsymbol{\Psi}_j\\ &= \sum_{i,j=1}^{n} e_i(t_n)e_j(t_n)\delta_{ij}\\ &= \sum_{i=1}^{n}\left[e_i(t_n)\right]^2.\end{aligned}$$

Hence,

$$\boldsymbol{e}^t(t_n)\boldsymbol{M}\boldsymbol{e}(t_n) \to 0 \Longleftrightarrow |e_i(t_n)| \to 0 \quad \forall i.$$

Since $\boldsymbol{M}$ is assumed to be positive definite,

$$\boldsymbol{e}^t(t_n)\boldsymbol{M}\boldsymbol{e}(t_n) \to 0 \Longleftrightarrow \boldsymbol{e}(t_n) \to 0,$$

and the theorem is proved.

Stability

If u now denotes a perturbation of the solution, then noting that Eqn. (6.27) is linear, we get

$$\dot{u} + \lambda u = 0.$$

Since the solution of this equation is given by $ce^{-\lambda t}$, we have $u(t_n) = ce^{-\lambda t_n}$ and $u(t_{n+1}) = ce^{-\lambda t_{n+1}}$. Thus,

$$u(t_{n+1}) = e^{-\lambda(t_{n+1}-t_n)}u(t_n).$$

For the solution to remain stable, i.e., the condition that the perturbation not blow up, we need

$$\begin{aligned}|u(t_{n+1})| &< |u(t_n)|, \quad \lambda > 0,\\ u(t_{n+1}) &= u(t_n), \quad \lambda = 0.\end{aligned} \tag{6.29}$$

These are the conditions that we wish to mimic in the temporally discrete case.

If u_n and u_{n+1} denotes the perturbations in the finite element solution at times t_n and t_{n+1}, then using Eqn. (6.28), we have

$$(1 + \alpha\Delta t\lambda)u_{n+1} = (1 - (1-\alpha)\Delta t\lambda)u_n.$$

Note that $(1 + \alpha\Delta t\lambda) > 0$ since α and Δt are both greater than zero. We can write the above equation as

$$u_{n+1} = Au_n, \tag{6.30}$$

where

$$A = \frac{1 - (1 - \alpha)\Delta t\lambda}{1 + \alpha\Delta t\lambda},$$

is the amplification factor. To mimic Eqn. 6.29, we need

$$|u_{n+1}| < |u_n|\,, \quad \lambda > 0,$$
$$u_{n+1} = u_n, \quad \lambda = 0.$$

From the definition of A, we see that the second condition is automatically satisfied. The first condition implies $|A| < 1$, or $-1 < A < 1$, i.e.,

$$-1 < \frac{1 - (1 - \alpha)\Delta t\lambda}{1 + \alpha\Delta t\lambda} < 1.$$

The right-hand inequality is automatically satisfied. The left-hand inequality is satisfied whenever $\alpha \geq 0.5$. If $\alpha < 0.5$, then

$$\lambda\Delta t < \frac{2}{1 - 2\alpha}.$$

An algorithm for which stability imposes a time-step restriction is called *conditionally stable*. An algorithm for which there is no time-step restriction imposed by stability is called *unconditionally stable*. Even if the amplification factor, A, is slightly greater than 1, disastrous growth can occur. For example, for $|A| = 1.01$ and $n = 1000$, we have

$$\frac{u_n}{u_0} = A^n = 1.01^{1000} = 2.09 \times 10^4.$$

Note the following points:

1. In the conditionally stable case, the stability condition

   $$\Delta t < \frac{2}{(1 - 2\alpha)\lambda},$$

   must hold for all modes λ_i, $i = 1, 2, \ldots, n$. Hence, the greatest λ_i, i.e., λ_n imposes the most stringent restriction on the time step, i.e.,

   $$\Delta t < \frac{2}{(1 - 2\alpha)\lambda_n}.$$

   In heat transfer problems, it can be shown that $\lambda_n = \mathrm{o}\,(1/h^2)$, where h is the typical element size. Thus, we need $\Delta t < ch^2$, where c is some constant. This, is a very severe constraint on the step size; thus, unconditionally stable algorithms are generally preferred.

2. For $\alpha = 0.5$ and $\lambda\Delta t \gg 1$, $A \approx -1$. Thus, high modal components will behave like $(-1)^n$ ('sawtooth' pattern). These spurious modes are filtered out by reporting step-to-step averages , $(u_{n+1} + u_n)/2$.

Convergence

The temporally discrete model problem can be written as

$$u_{n+1} - Au_n - L_n = 0, \tag{6.31}$$

where

$$L_n = \frac{\Delta t f_{n+\alpha}}{1 + \alpha \Delta t \lambda}.$$

If we replace u_n and u_{n+1} by the exact values, we get

$$u(t_{n+1}) - Au(t_n) - L_n = \Delta t \tau(t_n), \tag{6.32}$$

where $\tau(t_n)$ is the local truncation error. If

$$|\tau(t)| \leq c\Delta t^k \quad \forall t \in [0, T],$$

where c is a constant independent of Δt, and $k > 0$, then the algorithm defined by Eqn. 6.31 is *consistent*, and k is called the *order of accuracy* or *rate of convergence*.

The generalized trapezoidal methods are consistent. For $\alpha \in [0, 1]$, $k = 1$, except $\alpha = 0.5$, when $k = 2$. The expression for τ is

$$\tau = (1 - 2\alpha)\mathrm{o}\left(\Delta t\right) + \mathrm{o}\left(\Delta t^2\right).$$

Thus, the trapezoidal rule ($\alpha = 0.5$) is the only member of the family of methods that is second-order accurate.

We have the following result:
Theorem: Consider Eqns. 6.31 and 6.32. Let t_n be fixed (n, and hence $\Delta t = t_n/n$ are allowed to vary). Assume that the following conditions hold

1. $|A| \leq 1$ (stability)
2. $|\tau(t)| \leq c\Delta t^k$, $t \in [0, T]$, $k > 0$ (consistency).

Then $e(t_n) \to 0$ as $\Delta t \to 0$.
Proof: Subtract Eqn. 6.32 from 6.31 to get

$$e(t_{n+1}) = Ae(t_n) - \Delta t \tau(t_n). \tag{6.33}$$

Form Eqn. 6.33 (known as the error equation) it follows that

$$e(t_n) = Ae(t_{n-1}) - \Delta t \tau(t_{n-1}).$$

Thus,

$$\begin{aligned} e(t_{n+1}) &= A^2 e(t_{n-1}) - \Delta t A \tau(t_{n-1}) - \Delta t \tau(t_n) \\ &= A^3 e(t_{n-2}) - \Delta t A^2 \tau(t_{n-2}) - \Delta t A \tau(t_{n-1}) - \Delta t \tau(t_n) \\ &= A^{n+1} e(0) - \Delta t \sum_{i=0}^{n} A^i \tau(t_{n-i}). \end{aligned}$$

The first term on the right vanishes since $e(0) = u_0 - u(0) = 0$. Thus,

$$\begin{aligned}
|e(t_n)| &= \Delta t \left| \sum_{i=0}^{n-1} A^i \tau(t_{n-1-i}) \right| \\
&\leq \Delta t \sum_{i=0}^{n-1} |A|^i \, |\tau(t_{n-1-i})| \\
&\leq \Delta t \sum_{i=0}^{n-1} |\tau(t_{n-1-i})| && \text{(by stability)} \\
&\leq t_n \max |\tau(t)| && (\Delta t = t_n/n) \\
&\leq t_n c \Delta t^k && \text{(consistency)}
\end{aligned}$$

Hence, $e(t_n) \to 0$ as $\Delta t \to 0$, and furthermore the rate of convergence is k, i.e., $e(t_n) = \mathrm{o}\left(\Delta t^k\right)$. The maximal rate of convergence is 2.

The theorem just proved is a particular example of the Lax equivalence theorem which states that consistency plus stability is necessary and sufficient for convergence.

6.3.3 Modal Analysis

Modal analysis is an alternative technique to the step-by-step integration procedure already described. The steps involved are

1. Solve the eigenvalue problem

   $$(\boldsymbol{K} - \lambda_i \boldsymbol{M})\boldsymbol{\Psi}_i = 0 \quad i = 1, 2, \ldots, n_{\text{eq}},$$

   where $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and

   $$\begin{aligned}
   \boldsymbol{\Psi}_i^t \boldsymbol{M} \boldsymbol{\Psi}_j &= \delta_{ij} \\
   \boldsymbol{\Psi}_i^t \boldsymbol{K} \boldsymbol{\Psi}_j &= \lambda_i \delta_{ij} \quad \text{(no sum on } i\text{)}.
   \end{aligned}$$

   for the pairs $\{\lambda_i, \boldsymbol{\Psi}_i\}$, $1 \leq i \leq n_{\text{modes}}$, where $n_{\text{modes}} \leq n_{\text{eq}}$ is the desired number of modes to participate in the subsequent calculations. Two important factors to be kept in mind are

   (a) λ_i and $\boldsymbol{\Psi}_i$ should be good approximations of the exact eigenvalues and eigenvectors.

   (b) The spatial variation of $\boldsymbol{u}$ and $\boldsymbol{f}$ must be adequately represented by the expansion upto n_{modes}.

2. Solve the ordinary differential equations

   $$\dot{u}_i + \lambda_i u_i = f_i \quad 1 \leq i \leq n_{\text{modes}},$$

   where $u_i(0) = u_{0(i)}$. Solutions to the above equations can be obtained easily if $\boldsymbol{f}(t)$ is a simple function of time, or by a step-by-step method using a very small time step. Since the problem in step 2 is of a scalar nature, the computational cost in step 2 is negligible compared to step 1.

3. The approximation to $\boldsymbol{u}(t)$ is synthesized from the modes, viz.,

$$\boldsymbol{u}(t) \approx \sum_{i=1}^{n_{\text{modes}}} u_i(t)\boldsymbol{\Psi}_i.$$

The calculation in the above equation must be performed for each time t at which the solution is required.

To summarize, the features of the step-by-step method are

1. they are easily coded
2. they are more efficient for short-time calculations
3. they are generalizable to nonlinear problems

while those of the modal methods are

1. they are efficient if many analyses of the same configuration are required.
2. for long-time calculations
3. if only a small number of modes are participating in the solution.

6.4 Algorithms for Hyperbolic Problems

We have seen in Section 6.2.1 that the semi-discrete equation of motion is

$$\boldsymbol{M}\ddot{\hat{\boldsymbol{u}}} + \boldsymbol{C}\dot{\hat{\boldsymbol{u}}} + \boldsymbol{K}\hat{\boldsymbol{u}} = \hat{\boldsymbol{f}}, \tag{6.34}$$

where $\boldsymbol{M}$ is the mass matrix, $\boldsymbol{C}$ is the viscous damping matrix, $\boldsymbol{K}$ is the stiffness matrix, and $\boldsymbol{u}$, $\dot{\boldsymbol{u}}$, and $\ddot{\boldsymbol{u}}$ represent the displacement, velocity and acceleration vectors. The initial conditions are given to be $\hat{\boldsymbol{u}}(0) = \boldsymbol{u}_0$ and $\hat{\boldsymbol{v}}(0) = \boldsymbol{v}_0$. The exact solution can be found by converting Eqn. (6.34) into two sets of first-order equations. Let $\hat{\boldsymbol{v}} = \dot{\hat{\boldsymbol{u}}}$, so that Eqn. (6.34) can be written as $M\dot{\hat{\boldsymbol{v}}} + \boldsymbol{C}\hat{\boldsymbol{v}} + \boldsymbol{K}\hat{\boldsymbol{u}} = \hat{\boldsymbol{f}}$. These two sets of first-order equations can be written as

$$\tilde{\boldsymbol{M}}\dot{\tilde{\boldsymbol{u}}} + \tilde{\boldsymbol{K}}\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{f}},$$

where

$$\tilde{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{bmatrix}, \quad \tilde{\boldsymbol{K}} = \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{I} \\ \boldsymbol{K} & \boldsymbol{C} \end{bmatrix}, \quad \tilde{\boldsymbol{u}} = \begin{bmatrix} \hat{\boldsymbol{u}} \\ \hat{\boldsymbol{v}} \end{bmatrix}, \quad \tilde{\boldsymbol{f}} = \begin{bmatrix} \boldsymbol{0} \\ \hat{\boldsymbol{f}} \end{bmatrix}.$$

From Eqns. (6.18), it directly follows that

$$\tilde{\boldsymbol{u}}(t) = e^{-\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{K}}t}\tilde{\boldsymbol{u}}_0 + \int_0^t e^{\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{K}}(\xi - t)}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{f}}(\xi)\, d\xi, \tag{6.35a}$$

$$\tilde{\boldsymbol{u}}(t+\Delta t) = e^{-\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{K}}\Delta t}\tilde{\boldsymbol{u}}(t) + \int_t^{t+\Delta t} e^{\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{K}}(\xi-t-\Delta t)}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{f}}(\xi)\,d\xi, \quad (6.35b)$$

where

$$\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{K}} = \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{I} \\ \boldsymbol{M}^{-1}\boldsymbol{K} & \boldsymbol{M}^{-1}\boldsymbol{C} \end{bmatrix}, \quad \tilde{\boldsymbol{u}}_0 = \begin{bmatrix} \boldsymbol{u}_0 \\ \boldsymbol{v}_0 \end{bmatrix}.$$

We now discuss the modal analysis and direct methods.

6.4.1 Modal analysis

Modal analysis is preferred for problems of structural dynamics, since frequencies and mode shapes usually need to be computed for design purposes. Hence, the most costly constituent in modal analysis, which is the solution of the eigenproblem is effectively taken care of. If one were to calculate all the eigenvectors, then the modal analysis method would be prohibitively expensive. The reason why, often, modal analysis is cost-effective is that in many problems, only a small number of low-frequency modes participate in the structural response. Hence, only the first few eigenvectors need to be computed. Thus, modal analysis effectively damps out the higher modes.

Let the normalized eigenvectors $\boldsymbol{\phi}_i$ in the eigenvalue problem

$$\boldsymbol{K}\boldsymbol{\phi}_i = \omega_i^2 \boldsymbol{M}\boldsymbol{\phi}_i$$

be arranged as columns in the matrix $\boldsymbol{\Phi}$. Then as discussed in Section 6.2.2,

$$\boldsymbol{\Phi}^t\boldsymbol{M}\boldsymbol{\Phi} = \boldsymbol{I},$$
$$\boldsymbol{\Phi}^t\boldsymbol{K}\boldsymbol{\Phi} = \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}$ is a matrix with the eigenvectors ω_i^2 along the diagonal and all other elements zero. Since the vectors $\boldsymbol{\phi}_i$ form a basis, we can write the displacement vector, $\hat{\boldsymbol{u}}$ as

$$\hat{\boldsymbol{u}} = \boldsymbol{\Phi}\boldsymbol{u}.$$

Since $\boldsymbol{\Phi}$ is independent of time, we can write Eqn. (6.34) as

$$\boldsymbol{M}\boldsymbol{\Phi}\ddot{\boldsymbol{u}} + \boldsymbol{C}\boldsymbol{\Phi}\dot{\boldsymbol{u}} + \boldsymbol{K}\boldsymbol{\Phi}\boldsymbol{u} = \hat{\boldsymbol{f}}.$$

If we premultiply the above equation by $\boldsymbol{\Phi}^t$, then we get

$$\ddot{\boldsymbol{u}} + \boldsymbol{\Phi}^t\boldsymbol{C}\boldsymbol{\Phi}\dot{\boldsymbol{u}} + \boldsymbol{\Omega}\boldsymbol{u} = \boldsymbol{f},$$

where $\boldsymbol{f} = \boldsymbol{\Phi}^t\hat{\boldsymbol{f}}$. A considerable simplification, namely, uncoupling of the above equations takes place if we assume the damping matrix to be orthogonal, i.e.,

$$\boldsymbol{\Phi}^t\boldsymbol{C}\boldsymbol{\Phi} = \boldsymbol{\Xi},$$

where $\boldsymbol{\Xi}$ is a diagonal matrix. Now one has a set of n uncoupled ordinary differential equations

$$\ddot{\boldsymbol{u}} + \boldsymbol{\Xi}\dot{\boldsymbol{u}} + \boldsymbol{\Omega}\boldsymbol{u} = \boldsymbol{f},$$

which are to be solved subject to the initial conditions $\boldsymbol{u}(0) = \boldsymbol{\Phi}^t \boldsymbol{M} \hat{\boldsymbol{u}}_0$ and $\dot{\boldsymbol{u}}(0) = \boldsymbol{\Phi}^t \boldsymbol{M} \dot{\hat{\boldsymbol{u}}}_0$. Based on the analogy with a sping-mass-damper system, each diagonal element of $\boldsymbol{\Xi}$ is conventionally written as $2\xi_i\omega_i$, where $\xi_i > 0$ is known as the modal damping parameter. The damping ratio $\xi = 1$ marks the transition between oscillatory and nonoscillatory response, and is hence known as the *critical damping ratio*. Damping ratios are found experimentally by observing the vibratory response of the structure. Typically, ξ ranges between 0.005 to 0.15.

If the modal damping parameters (and hence $\boldsymbol{\Xi}$) is known, then the damping matrix can be constructed explicitly as

$$\boldsymbol{C} = \boldsymbol{M}\boldsymbol{\Phi}\boldsymbol{\Xi}\boldsymbol{\Phi}^t\boldsymbol{M}.$$

One can easily verify from the above equation that $\boldsymbol{\Phi}^t\boldsymbol{C}\boldsymbol{\Phi} = \boldsymbol{\Xi}$. However, since computing the $\boldsymbol{\Phi}$ matrix is quite an expensive proposition, one usually assumes Rayleigh or proportional damping, i.e.,

$$\boldsymbol{C} = \alpha\boldsymbol{M} + \beta\boldsymbol{K},$$

where α and β are determined as follows. Premultiplying and postmultiplying the above equation by $\boldsymbol{\Phi}^t$ and $\boldsymbol{\Phi}$ respectively, we get

$$\alpha + \beta\omega_i^2 = 2\xi_i\omega_i \quad i = 1, 2, ..., n. \tag{6.36}$$

Obviously the above equation cannot hold, in general, for all values of i since there are only two undetermined constants α and β. Hence, the above equations are solved for two given damping ratios (which are determined experimentally) corresponding to two distinct natural frequencies ω_1 and ω_2. We get

$$\alpha = \frac{2\omega_1\omega_2(\xi_1\omega_2 - \xi_2\omega_1)}{\omega_2^2 - \omega_1^2},$$
$$\beta = \frac{2(\xi_2\omega_2 - \xi_1\omega_1)}{\omega_2^2 - \omega_1^2}.$$

Once α and β are determined, the remaining modal damping parameters are determined from Eqn. (6.36), though these will in general, not match with the actual damping parameters of the system. Note that in this approach $\boldsymbol{C}$ is banded, whereas in other approaches, it might be fully populated.

One usually chooses ω_1 to be the lowest natural frequency and ω_2 to be the maximum frequency of interest in the loading or response. Choosing ω_1 to be the lowest natural frequency guarantees that $\boldsymbol{C}$ is positive definite. To see this, let $\boldsymbol{a}$ be some arbitrary vector. We want to prove that $\boldsymbol{a}^t\boldsymbol{C}\boldsymbol{a} > 0$. Since the columns of $\boldsymbol{\Phi}$ act as a basis, we can express $\boldsymbol{a}$ as $\boldsymbol{\Phi}\boldsymbol{b}$, where $\boldsymbol{b}$ is another vector. Thus, if $\boldsymbol{b} \equiv (b_1, b_2, ..., b_n)$ then

$$\begin{aligned}\boldsymbol{a}^t\boldsymbol{C}\boldsymbol{a} &= \boldsymbol{b}^t\boldsymbol{\Phi}^t(\alpha\boldsymbol{M} + \beta\boldsymbol{K})\boldsymbol{\Phi}\boldsymbol{b} \\ &= \boldsymbol{b}^t(\alpha\boldsymbol{I} + \beta\boldsymbol{\Omega})\boldsymbol{b}\end{aligned}$$

$$
\begin{aligned}
&= (\alpha + \beta\omega_1^2)b_1^2 + \cdots + (\alpha + \beta\omega_n^2)b_n^2 \\
&> 0,
\end{aligned}
$$

where the last step follows from the fact that

$$\alpha + \beta\omega_n^2 \geq \alpha + \beta\omega_{n-1}^2 \geq \cdots \geq \alpha + \beta\omega_1^2 = 2\xi_1\omega_1 > 0.$$

Now we describe the modal analysis method. The steps are

- Solve the eigenvalue problem

  $$(\boldsymbol{K} - \omega_i^2 \boldsymbol{M})\boldsymbol{\phi}_i = \boldsymbol{0} \quad i = 1, 2, \ldots, n_{\text{eq}},$$

  where $0 \leq \omega_1 \leq \omega_2 \leq \ldots \leq \omega_n$ and

  $$
  \begin{aligned}
  \boldsymbol{\phi}_i^t \boldsymbol{M} \boldsymbol{\phi}_j &= \delta_{ij} \\
  \boldsymbol{\phi}_i^t \boldsymbol{K} \boldsymbol{\phi}_j &= \omega_i^2 \delta_{ij} \quad \text{(no sum on } i\text{)}.
  \end{aligned}
  $$

  for the pairs $\{\omega_i, \boldsymbol{\phi}_i\}$, $1 \leq i \leq n_{\text{modes}}$, where $n_{\text{modes}} \leq n_{\text{eq}}$ is the desired number of modes to participate in the subsequent calculations.

- Solve the ordinary differential equations

  $$\ddot{u}_i + 2\xi_i\omega_i\dot{u}_i + \omega_i^2 u_i = f_i \quad 1 \leq i \leq n_{\text{modes}},$$

  where $u_i(0) = u_{0(i)}$ and $\dot{u}_i(0) = \dot{u}_{0(i)}$. Solutions to the above equations can be obtained easily if $\boldsymbol{f}(t)$ is a simple function of time, or by a step-by-step method using a very small time step. Since the problem in step 2 is of a scalar nature, the computational cost in step 2 is negligible compared to step 1.

- The approximation to $\boldsymbol{u}(t)$ is synthesized from the modes, viz.,

  $$\boldsymbol{u}(t) \approx \sum_{i=1}^{n_{\text{modes}}} u_i(t)\boldsymbol{\phi}_i.$$

  The calculation in the above equation must be performed for each time t at which the solution is required.

6.4.2 Direct methods

We now describe an energy-momentum conserving method for the direct solution of Eqn. (6.34), which is identical to the trapezoidal scheme in the Newmark method. We first ignore damping ($\boldsymbol{C} = \boldsymbol{0}$). The balance of linear and angular momenta, and the balance of energy (in the context of linear elasticity) are given by

$$\frac{d}{dt}\int_V \rho\boldsymbol{v}\, dV = \int_S \boldsymbol{t}\, dS + \int_V \rho\boldsymbol{b}\, dV,$$

$$\frac{d}{dt}\int_V \rho(\boldsymbol{X}\times\boldsymbol{v})\,dV = \int_S \boldsymbol{X}\times\boldsymbol{t}\,dS + \int_V \rho\boldsymbol{X}\times\boldsymbol{b}\,dV,$$
$$\frac{d}{dt}\int_V \left[\rho\frac{\boldsymbol{v}\cdot\boldsymbol{v}}{2} + W(\boldsymbol{\epsilon})\right]\,dV = \int_S \boldsymbol{t}\cdot\boldsymbol{v}\,dS + \int_V \rho\boldsymbol{b}\cdot\boldsymbol{v}\,dV,$$

where $W(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon} : \mathbb{C}\boldsymbol{\epsilon}/2$ is the strain-energy density function. In deriving the last equation, we have used the fact that $\boldsymbol{D} \approx \dot{\boldsymbol{\epsilon}}$ in the context of linear elasticity, so that $\boldsymbol{\tau} : \boldsymbol{D} = \partial W/\partial t$. Thus, in the absence of any displacement constraints and in the absence of loading, i.e., when $\boldsymbol{t} = \boldsymbol{b} = \boldsymbol{0}$, the linear momentum, angular momentum, and total energy (i.e., kinetic+strain energy) are conserved, i.e.,

$$\int_V \rho\boldsymbol{v}\,dV = \text{constant},$$
$$\int_V \rho(\boldsymbol{X}\times\boldsymbol{v})\,dV = \text{constant},$$
$$\int_V \left[\rho\frac{\boldsymbol{v}\cdot\boldsymbol{v}}{2} + W(\boldsymbol{\epsilon})\right]\,dV = \text{constant},$$

at all times.

We want to devise a numerical strategy that mimics the above continuum behavior. Let $\hat{\boldsymbol{u}}_n$, $\hat{\boldsymbol{v}}_n$ and $\hat{\boldsymbol{f}}_n$ denote the nodal displacement, velocity and load vectors, respectively, at time t_n, and let $t_\Delta = t_{n+1} - t_n$. We propose the following time-stepping strategy over the time-interval $[t_n, t_{n+1}]$:

$$\boldsymbol{M}\left(\frac{\hat{\boldsymbol{v}}_{n+1} - \hat{\boldsymbol{v}}_n}{t_\Delta}\right) + \boldsymbol{K}\left(\frac{\hat{\boldsymbol{u}}_n + \hat{\boldsymbol{u}}_{n+1}}{2}\right) = \frac{\hat{\boldsymbol{f}}_n + \hat{\boldsymbol{f}}_{n+1}}{2}, \tag{6.37}$$

where

$$\frac{\hat{\boldsymbol{u}}_{n+1} - \hat{\boldsymbol{u}}_n}{t_\Delta} = \frac{\hat{\boldsymbol{v}}_n + \hat{\boldsymbol{v}}_{n+1}}{2}. \tag{6.38}$$

In the absence of loading, i.e., when $\hat{\boldsymbol{f}}_n = \hat{\boldsymbol{f}}_{n+1} = \boldsymbol{0}$, Eqn. (6.37) can be written as

$$\int_V \rho\boldsymbol{u}_\delta \cdot \left(\frac{\boldsymbol{v}_{n+1} - \boldsymbol{v}_n}{t_\Delta}\right)\,dV + \int_V \boldsymbol{\epsilon}(\boldsymbol{u}_\delta) : \mathbb{C}\left[\frac{\boldsymbol{\epsilon}(\boldsymbol{u}_n) + \boldsymbol{\epsilon}(\boldsymbol{u}_{n+1})}{2}\right]\,dV = 0 \quad \forall \boldsymbol{u}_\delta. \tag{6.39}$$

We now make special choices of $\boldsymbol{u}_\delta$ in Eqn. (6.39) in order to prove the conservation properties of the algorithm. First choose $\boldsymbol{u}_\delta = \boldsymbol{c}$ over the entire domain, where $\boldsymbol{c}$ is a constant vector. This choice is permissible since the entire boundary is assumed to be free of displacement constraints. For this choice, $\boldsymbol{\epsilon}(\boldsymbol{u}_\delta) = \boldsymbol{0}$, so that we get

$$\boldsymbol{c}\cdot\int_V \rho\left(\frac{\boldsymbol{v}_{n+1} - \boldsymbol{v}_n}{t_\Delta}\right)\,dV = 0.$$

Since $\boldsymbol{c}$ is arbitrary, we get

$$\int_V \rho\boldsymbol{v}_{n+1}\,dV = \int_V \rho\boldsymbol{v}_n\,dV,$$

which proves that the linear momentum is conserved.

Now choose $\boldsymbol{u}_\delta = \boldsymbol{c} \times \boldsymbol{X}$ over the entire domain, where $\boldsymbol{c}$ is a constant vector. If $\boldsymbol{W}$ is the skew tensor whose axial vector is $\boldsymbol{c}$, then we have $\boldsymbol{u}_\delta = \boldsymbol{W}\boldsymbol{X}$, so that $\boldsymbol{\nabla} \boldsymbol{u}_\delta = \boldsymbol{W}$, again resulting in $\boldsymbol{\epsilon}(\boldsymbol{u}_\delta) = \boldsymbol{0}$. Using the property $(\boldsymbol{p} \times \boldsymbol{q}) \cdot \boldsymbol{r} = \boldsymbol{p} \cdot (\boldsymbol{q} \times \boldsymbol{r})$, Eqn. (6.39) reduces to

$$\boldsymbol{c} \cdot \int_V \rho \boldsymbol{X} \times (\boldsymbol{v}_{n+1} - \boldsymbol{v}_n)\, dV = 0,$$

which, by virtue of the arbitrariness of $\boldsymbol{c}$, leads to

$$\int_V \rho \boldsymbol{X} \times \boldsymbol{v}_{n+1}\, dV = \int_V \rho \boldsymbol{X} \times \boldsymbol{v}_n\, dV,$$

i.e., the angular momentum is conserved.

Finally to show that the total energy is conserved, choose $\boldsymbol{u}_\delta = \boldsymbol{u}_{n+1} - \boldsymbol{u}_n$, so that $\boldsymbol{\epsilon}(\boldsymbol{u}_\delta) = \boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n$. Eqn. (6.39) reduces to

$$\int_V \rho(\boldsymbol{u}_{n+1} - \boldsymbol{u}_n) \cdot \left(\frac{\boldsymbol{v}_{n+1} - \boldsymbol{v}_n}{t_\Delta} \right) dV + \int_V (\boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n) : \mathbb{C} \left[\frac{\boldsymbol{\epsilon}_{n+1} + \boldsymbol{\epsilon}_n}{2} \right] dV = 0.$$

Using Eqn. (6.38), the above equation reduces to

$$\frac{1}{2} \int_V \rho(\boldsymbol{v}_{n+1} + \boldsymbol{v}_n) \cdot (\boldsymbol{v}_{n+1} - \boldsymbol{v}_n)\, dV + \frac{1}{2} \int_V (\boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n) : \mathbb{C}(\boldsymbol{\epsilon}_{n+1} + \boldsymbol{\epsilon}_n)\, dV = 0,$$

which, by virtue of the symmetry of the material constitutive tensor $\mathbb{C}$, reduces further to

$$\frac{1}{2} \int_V \rho \left[\boldsymbol{v}_{n+1} \cdot \boldsymbol{v}_{n+1} - \boldsymbol{v}_n \cdot \boldsymbol{v}_n \right] dV + \frac{1}{2} \int_V \left[\boldsymbol{\epsilon}_{n+1} : \mathbb{C}\boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n : \mathbb{C}\boldsymbol{\epsilon}_n \right] dV = 0. \tag{6.40}$$

Thus, the total energy is conserved:

$$\left[\text{K.E.} + \text{Strain energy} \right]_{n+1} = \left[\text{K.E.} + \text{Strain energy} \right]_n .$$

Note that the choices for $\boldsymbol{u}_\delta$ made above fall within the finite element space for $\boldsymbol{u}$, and hence the conservation properties hold in the fully-discrete setting.

By eliminating $\boldsymbol{v}_{n+1}$ from Eqns. (6.37) and (6.38), we get

$$\left[\frac{2\boldsymbol{M}}{t_\Delta^2} + \frac{\boldsymbol{K}}{2} \right] \hat{\boldsymbol{u}}_{n+1} = \frac{2}{t_\Delta^2} \boldsymbol{M} \hat{\boldsymbol{u}}_n + \frac{2}{t_\Delta} \boldsymbol{M} \hat{\boldsymbol{v}}_n - \frac{1}{2} \boldsymbol{K} \hat{\boldsymbol{u}}_n + \frac{1}{2} \left(\hat{\boldsymbol{f}}_n + \hat{\boldsymbol{f}}_{n+1} \right). \tag{6.41}$$

Using the initial displacement and velocity vectors, $\hat{\boldsymbol{u}}_0$ and $\hat{\boldsymbol{v}}_0$, we first solve for the nodal displacement vector $\hat{\boldsymbol{u}}_{t_1}$ at time t_1 using Eqn. (6.41). Next, we substitute for $\hat{\boldsymbol{u}}_{t_1}$ in Eqn. (6.38) to find $\hat{\boldsymbol{v}}_{t_1}$. Using $\hat{\boldsymbol{u}}_{t_1}$ and $\hat{\boldsymbol{v}}_{t_1}$, we find $(\hat{\boldsymbol{u}}_{t_2}, \hat{\boldsymbol{v}}_{t_2})$, and so on. Thus, starting from the initial displacement and velocity fields, we march forward in time, until the time instant at which the response is desired is reached.

Now consider a nonzero damping matrix $\boldsymbol{C}$, which is assumed to be positive semi-definite (e.g., $\boldsymbol{C} = \alpha \boldsymbol{M} + \beta \boldsymbol{K}$). Using the approximation

$\dot{\hat{\boldsymbol{u}}} \approx (\hat{\boldsymbol{u}}_{n+1} - \hat{\boldsymbol{u}}_n)/t_\Delta$ in Eqn. (6.34), we now get instead of Eqn. (6.40) the relation

$$\frac{1}{2}\int_V \rho\left[\boldsymbol{v}_{n+1}\cdot\boldsymbol{v}_{n+1} - \boldsymbol{v}_n\cdot\boldsymbol{v}_n\right]\,dV + \frac{1}{2}\int_V \left[\boldsymbol{\epsilon}_{n+1} : \mathbb{C}\boldsymbol{\epsilon}_{n+1} - \boldsymbol{\epsilon}_n : \mathbb{C}\boldsymbol{\epsilon}_n\right]\,dV + \frac{1}{t_\Delta}(\hat{\boldsymbol{u}}_{n+1} - \hat{\boldsymbol{u}}_n)\cdot\boldsymbol{C}(\hat{\boldsymbol{u}}_{n+1} - \hat{\boldsymbol{u}}_n) = 0,$$

which by virtue of the positive definiteness of $\boldsymbol{C}$ shows that

$$\left[\text{K.E.} + \text{Strain energy}\right]_{n+1} \leq \left[\text{K.E.} + \text{Strain energy}\right]_n.$$

Thus, in the presence of damping, similar to the continuum dynamics, the linear and angular momenta are conserved, and the total energy in non-increasing. Instead of Eqn. (6.41), we now have

$$\left[\frac{2\boldsymbol{M}}{t_\Delta^2} + \frac{\boldsymbol{C}}{t_\Delta} + \frac{\boldsymbol{K}}{2}\right]\hat{\boldsymbol{u}}_{n+1} = \frac{2}{t_\Delta^2}\boldsymbol{M}\hat{\boldsymbol{u}}_n + \frac{2}{t_\Delta}\boldsymbol{M}\hat{\boldsymbol{v}}_n + \frac{1}{t_\Delta}\boldsymbol{C}\hat{\boldsymbol{u}}_n - \frac{1}{2}\boldsymbol{K}\hat{\boldsymbol{u}}_n + \frac{1}{2}\left(\hat{\boldsymbol{f}}_n + \hat{\boldsymbol{f}}_{n+1}\right). \tag{6.42}$$

Since the energy is non-increasing, the above time-stepping scheme is unconditionally stable, i.e., there are no restrictions on the time step t_Δ. If t_Δ is chosen to be a constant, then the matrix on the left hand side of Eqn. (6.42) can be decomposed right at the outset, and one merely needs to use back-substitution for all the subsequent time steps, making the whole process extremely efficient.

Another direct method for solving Eqn. (6.34) is the Newmark method which we now describe. If $\boldsymbol{u}_n$, $\boldsymbol{v}_n$ and $\boldsymbol{a}_n$ are approximations to $\boldsymbol{u}(t_n)$, $\dot{\boldsymbol{u}}(t_n)$ and $\ddot{\boldsymbol{u}}(t_n)$ respectively, then the Newmark scheme is given by

$$\boldsymbol{M}\boldsymbol{a}_{n+1} + \boldsymbol{C}\boldsymbol{v}_{n+1} + \boldsymbol{K}\boldsymbol{u}_{n+1} = \boldsymbol{f}_{n+1}, \tag{6.43}$$

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \Delta t\boldsymbol{v}_n + \frac{\Delta t^2}{2}\left[(1-2\beta)\boldsymbol{a}_n + 2\beta\boldsymbol{a}_{n+1}\right], \tag{6.44}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + \Delta t\left[(1-\gamma)\boldsymbol{a}_n + \gamma\boldsymbol{a}_{n+1}\right]. \tag{6.45}$$

The constants, β and γ, determine the stability and accuracy of the algorithm. The implementation to find $\boldsymbol{u}_{n+1}$, $\boldsymbol{v}_{n+1}$ and $\boldsymbol{a}_{n+1}$ given $\boldsymbol{u}_n$, $\boldsymbol{v}_n$ and $\boldsymbol{a}_n$ is described next.

6.4.3 Implementation

Define predictors

$$\tilde{\boldsymbol{u}}_{n+1} = \boldsymbol{u}_n + \Delta t\boldsymbol{v}_n + \frac{\Delta t^2}{2}(1-2\beta)\boldsymbol{a}_n, \tag{6.46}$$

$$\tilde{\boldsymbol{v}}_{n+1} = \boldsymbol{v}_n + (1-\gamma)\Delta t\boldsymbol{a}_n. \tag{6.47}$$

Using the above relations, Eqns. 6.44 and 6.45 can be written as

$$\boldsymbol{u}_{n+1} = \tilde{\boldsymbol{u}}_{n+1} + \beta\Delta t^2\boldsymbol{a}_{n+1}, \tag{6.48}$$

$$\boldsymbol{v}_{n+1} = \tilde{\boldsymbol{v}}_{n+1} + \gamma \Delta t \boldsymbol{a}_{n+1}. \tag{6.49}$$

To start the process $\boldsymbol{a}_0$ is calculated from

$$\boldsymbol{M}\boldsymbol{a}_0 = \boldsymbol{f} - \boldsymbol{C}\boldsymbol{v}_0 - \boldsymbol{K}\boldsymbol{u}_0,$$

or specified directly. The following relation obtained by substituting Eqns. 6.48 and 6.49 in Eqn. 6.43 allows us to solve for $\boldsymbol{a}_{n+1}$:

$$\left(\boldsymbol{M} + \gamma \Delta t \boldsymbol{C} + \beta \Delta t^2 \boldsymbol{K}\right) \boldsymbol{a}_{n+1} = \boldsymbol{f}_{n+1} - \boldsymbol{C}\tilde{\boldsymbol{v}}_{n+1} - \boldsymbol{K}\tilde{\boldsymbol{u}}_{n+1}.$$

After finding $\boldsymbol{a}_{n+1}$, we use Eqns. 6.48 and 6.49 to obtain $\boldsymbol{u}_{n+1}$ and $\boldsymbol{v}_{n+1}$, respectively.

6.4.4 Convergence Analysis

The convergence analysis is similar to that for parabolic problems. The steps are

1. Reduction to SDOF model
2. Definition of a suitable notion of stability which is shown to hold under certain circumstances
3. Determination of the order of accuracy
4. Use of (ii) and (iii) to prove convergence.

If we assume Rayleigh damping, i.e.,

$$\boldsymbol{C} = a\boldsymbol{M} + b\boldsymbol{K},$$

then the symmetry of $\boldsymbol{M}$ and $\boldsymbol{K}$ make the decomposition shown in Fig. 6.4 possible. The parameter ω in Fig. 6.4 is the undamped frequency of vibration, while ξ is known as the damping ratio, and is given by

$$\xi = \frac{1}{2}\left(\frac{a}{\omega} + b\omega\right).$$

The SDOF model problem can be written as

$$\boldsymbol{y}_{n+1} = \boldsymbol{A}\boldsymbol{y}_n + \boldsymbol{L}_n,$$

where $\boldsymbol{y}_n \equiv [\boldsymbol{u}_n, \boldsymbol{v}_n]^t$, and $\boldsymbol{A}$ is the *amplification matrix*. The stability conditions for the Newmark method are

Unconditional:

$$2\beta \geq \gamma \geq 0.5.$$

Conditional:

$$\gamma \geq 0.5; \quad \beta < \frac{\gamma}{2}; \quad \omega \Delta t \leq \Omega_{\text{crit}},$$

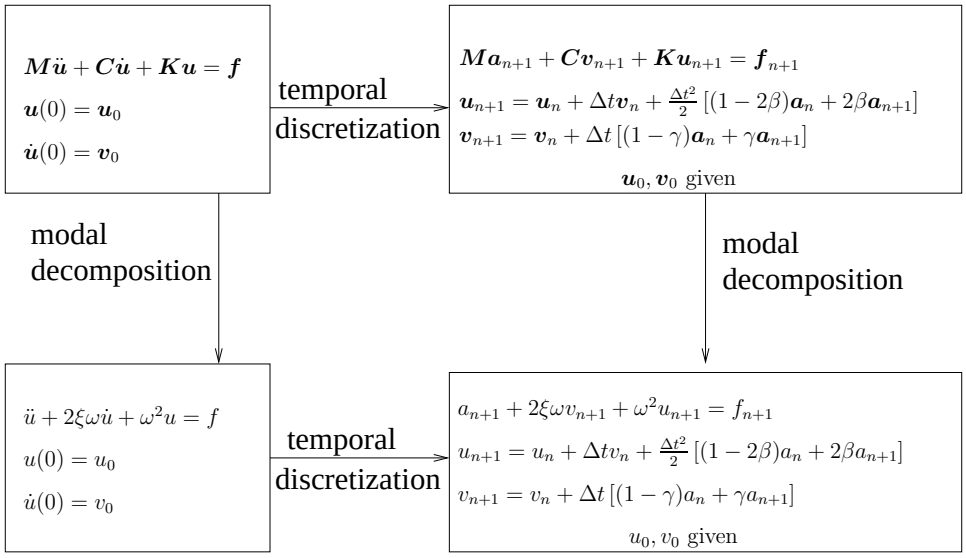


Figure 6.4: Temporal discretization and modal decomposition for hyperbolic problems

where

$$\Omega_{\text{crit}} = \frac{\xi(\gamma - 0.5) + \left[\frac{\gamma}{2} - \beta + \xi^2(\gamma - 0.5)^2\right]^{1/2}}{\left(\frac{\gamma}{2} - \beta\right)},$$

is known as the critical sampling frequency. The stability condition must be satisfied for each mode. Hence, the maximum natural frequency ω_n is critical for determining the time step. We have

$$\Delta t \leq \frac{\Omega_{\text{crit}}}{\omega_n}$$

As usual ω_n is bounded by the maximum frequency of the individual elements.

Observe that when $\gamma = 0.5$, viscous damping has no effect on the stability, since we have

$$\Omega_{\text{crit}} = \left(\frac{\gamma}{2} - \beta\right)^{-0.5}.$$

When $\gamma > 0.5$, the effect of viscous damping is to increase the critical time step of conditionally stable Newmark methods. Thus, the undamped critical frequency $(\gamma/2 - \beta)^{-0.5}$ serves as a conservative estimate when an estimate of the modal damping coefficient is not available.

6.4.5 Special cases of the Newmark method

Method	Type	β	γ	Stability condition	Order of accuracy
Average acceleration (trapezoidal rule)	implicit	$\frac{1}{4}$	$\frac{1}{2}$	unconditional	2
Linear acceleration	implicit	$\frac{1}{6}$	$\frac{1}{2}$	$\Omega_{\text{crit}} = 2\sqrt{3}$	2
Fox-Goodwin (royal road)	implicit	$\frac{1}{12}$	$\frac{1}{2}$	$\Omega_{\text{crit}} = \sqrt{6}$	2
Central difference	explicit	0	$\frac{1}{2}$	$\Omega_{\text{crit}} = 2$	2

Note: For the central difference method $\boldsymbol{M}$ and $\boldsymbol{C}$ need to be diagonal. Stability is based on the undamped case in which $\xi = 0$. Second-order accuracy is achieved if and only if $\gamma = 0.5$. In elastic wave propagation problems, the time step restriction is not too restrictive. Hence, the central difference method is preferred.

6.4.6 Finding the $\boldsymbol{A}$ matrix

We have

$$u_{n+1} = u_n + \Delta t v_n + \frac{\Delta t^2}{2}\left[(1-2\beta)a_n + 2\beta a_{n+1}\right], \tag{6.50}$$

$$v_{n+1} = v_n + \Delta t\left[(1-\gamma)a_n + \gamma a_{n+1}\right], \tag{6.51}$$

$$a_n + 2\xi\omega v_n + \omega^2 u_n = f_n, \tag{6.52}$$

$$a_{n+1} + 2\xi v_{n+1} + \omega^2 u_{n+1} = f_{n+1} \tag{6.53}$$

Eliminate a_n and a_{n+1} from Eqns. 6.50 and 6.51, and substitute in Eqns. 6.52 and 6.53 to get

$$\boldsymbol{A} = \boldsymbol{A}_1^{-1}\boldsymbol{A}_2,$$

$$\boldsymbol{L}_n = \boldsymbol{A}_1^{-1}\begin{bmatrix} \frac{\Delta t^2}{2}\left[(1-2\beta)f_n + 2\beta f_{n+1}\right] \\ \Delta t\left[(1-\gamma)f_n + \gamma f_{n+1}\right] \end{bmatrix},$$

where

$$\boldsymbol{A}_1 = \begin{bmatrix} 1+\Delta t\beta\omega^2 & 2\Delta t^2\beta\omega \\ \Delta t\gamma\omega^2 & 1+2\Delta t\gamma\xi\omega \end{bmatrix}$$

$$\boldsymbol{A}_2 = \begin{bmatrix} 1-\frac{\Delta t^2}{2}(1-2\beta)\omega^2 & \Delta t\left[1-\Delta t(1-2\beta)\xi\omega\right] \\ -\Delta t(1-\gamma)\omega^2 & 1-2\Delta t(1-\gamma)\xi\omega \end{bmatrix}$$

6.4.7 Convergence and stability

The local truncation error $\boldsymbol{\tau}$ is defined by

$$\boldsymbol{y}(t_{n+1}) = \boldsymbol{A}\boldsymbol{y}(t_n) + \boldsymbol{L}_n + \Delta t\boldsymbol{\tau}(t_n),$$

where $\boldsymbol{y}(t_n) = [u(t_n)\ \dot{u}(t_n)]^t$. Using an analysis similar to that for parabolic problems, we get

$$\boldsymbol{\tau}(t) = \mathrm{o}\left(\Delta t^k\right) \quad \forall t \in [0, T],$$

where $k = 2$ if $\gamma = 0.5$ and $k = 1$ otherwise.

The stability of the Newmark methods is determined by the properties of the amplification matrix. The error equation is given by

$$\boldsymbol{e}(t_n) = \boldsymbol{A}^n \boldsymbol{e}(0) - \Delta t \sum_{i=0}^{n-1} \boldsymbol{A}^i \boldsymbol{\tau}(t_{n-1-i}).$$

Once again, stability plus consistency implies convergence. The rate of convergence is k.

Let $\lambda_i(\boldsymbol{A})$ denote the eigenvalues of $\boldsymbol{A}$. The modulus of $\lambda_i(\boldsymbol{A})$ is written as

$$|\lambda_i(\boldsymbol{A})| = \left[\lambda_i(\boldsymbol{A})\bar{\lambda}_i(\boldsymbol{A})\right]^{1/2},$$

where $\bar{\lambda}_i(\boldsymbol{A})$ is the complex conjugate of $\lambda_i(\boldsymbol{A})$. The spectral radius of $\boldsymbol{A}$, $\rho(\boldsymbol{A})$ is defined by

$$\rho(\boldsymbol{A}) = \max_i |\lambda_i(\boldsymbol{A})|.$$

The stability condition is needed to prevent the amplification of $\boldsymbol{A}^n$ as n becomes large. The conditions for stability are

1. $\rho(\boldsymbol{A}) \leq 1$
2. The eigenvalues of $\boldsymbol{A}$ of multiplicity greater than 1 are strictly less than 1 in modulus.

The above conditions define a spectrally stable $\boldsymbol{A}$. Since the stability condition depends only on the eigenvalues of $\boldsymbol{A}$, it can be expressed in terms of the principal invariants of $\boldsymbol{A}$. Thus, λ is given by the solution to $\lambda^2 - I_1\lambda + I_2 = 0$, i.e.,

$$\lambda = \frac{I_1 \pm \sqrt{I_1^2 - 4I_2}}{2},$$

where

$$\begin{aligned} I_1 &= \mathrm{tr}\, \boldsymbol{A} = A_{11} + A_{22}, \\ I_2 &= \det \boldsymbol{A} = A_{11}A_{22} - A_{12}A_{21}. \end{aligned}$$

It is necessary to have $\gamma > 0.5$ to introduce high-frequency dissipation, since viscous damping damps only an intermediate band of frequencies without significant effect on the all important high modes. However, $\gamma > 0.5$ results in a drop to first-order accuracy. It is necessary to dissipate the higher modes since they are artifacts of the discretization process, and not representative of the behavior of the governing partial differential equations.

6.4.8 Superconvergence in one-dimensional problems

Consider the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{E}{\rho}\frac{\partial^2 u}{\partial x^2} = c^2 \frac{\partial^2 u}{\partial x^2},$$

where $c = \sqrt{E/\rho}$ is the characteristic velocity or the speed of sound. $\Delta t = h/c$ is called the characteristic time step. It is equal to the transit time for a 'wave' moving at speed c to traverse one element.

The element stiffness and mass matrices are given by

$$\boldsymbol{K}^{(e)} = \frac{EA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix};\ \boldsymbol{M}^{(e)} = \rho A h \begin{bmatrix} 0.5 - r & r \\ r & 0.5 - r \end{bmatrix}.$$

Particular values of r yield the commonly used mass matrices, e.g.,

1. consistent mass ($r = 1/6$)

$$\boldsymbol{M}^{(e)} = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

2. lumped mass ($r = 0$)

$$\boldsymbol{M}^{(e)} = \frac{h}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3. higher-order mass ($r = 1/12$)

$$\boldsymbol{M}^{(e)} = \frac{h}{12}\begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$$

If Δt is equal to the characteristic time step given by h/c, and $\beta = r$, $\gamma = 0.5$ and $\xi = 0$, then $\omega_{\text{approx}} = \omega_{\text{exact}}$, i.e., the errors introduced by the finite element discretization, the particular mass matrix and temporal algorithm all cancel to yield the exact results at the nodes ('superconvergence') no matter how few elements are employed.

The characteristic time step also represents the stability limit (i.e., the critical time step) of the lumped mass, central difference method. Hence, it is considered advantageous to compute at a time step as close to critical as possible. Also 'matched methods' ($\beta = r$) are preferred, e.g., central difference and lumped mass ($\beta = r = 0$).

6.4.9 Time estimates for some simple finite elements

Consider the two-node linear rod element. If we assume a lumped mass matrix then

$$\boldsymbol{K}^{(e)} = \frac{EA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}; \quad \boldsymbol{M}^{(e)} = \frac{\rho A h}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Computing the eigenvalues using $\det(\boldsymbol{K}^{(e)} - \omega^2 \boldsymbol{M}^{(e)}) = 0$, i.e.

$$\det \begin{bmatrix} \frac{2c^2}{h^2} - \omega^2 & -\frac{2c^2}{h^2} \\ -\frac{2c^2}{h^2} & \frac{2c^2}{h} - \omega^2 \end{bmatrix} = 0,$$

where $c = \sqrt{E/\rho}$. The above equation yields

$$\omega^2 = \frac{4c^2}{h^2} \text{ or } \omega = 0.$$

Thus, $\omega_{\max} = 2c/h$. For $\beta = 0$, $\gamma = 0.5$,

$$\Delta t_{\text{critical}} = \frac{2}{\omega_{\max}} = \frac{h}{c} = \Delta t_{\text{characteristic}}.$$

For a consistent mass matrix $\omega_{\max} = 2\sqrt{3}c/h$, which results in a reduced critical time step:

$$\Delta t \leq \frac{h}{\sqrt{3}c}.$$

In general, consistent-mass matrices tend to yield smaller critical time steps than lumped-mass matrices.

For the heat conduction problem with $\alpha = 0$, the condition $\lambda \Delta t \leq 2(1 - 2\alpha)$ implies that

$$\Delta t \leq \frac{2}{\lambda_{\max}}.$$

Since $\lambda_{\max} = \omega_{\max}^2 = 4k/\rho c h^2$, we get

$$\Delta t \leq \frac{\rho c h^2}{2k}.$$

For a 3-node quadratic element,

$$\boldsymbol{M}^{(e)}_{\text{lumped}} = \frac{\rho h A}{6}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \boldsymbol{K}^{(e)} = \frac{EA}{h}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix},$$

we get $\omega_{\max} = 2\sqrt{6}c/h$, yielding

$$\Delta t \leq \frac{h}{\sqrt{6}c}.$$

Thus, the allowable time step is about 0.4082 that for linear elements with lumped mass.

For a four-node quadrilateral element

$$\omega_{\max} \leq c_d \sqrt{g}, \tag{6.54}$$

where $c_d = \sqrt{(\lambda + 2\mu)/\rho}$ is the dilatational wave velocity, λ and μ are Lame constants, and g is a geometric parameter. For example, for a quadrilateral

$$g = \frac{4}{A^2} \sum_{i=1}^{2} \sum_{j=1}^{4} B_{ij} B_{ij},$$

where A is the area, and

$$B_{ij} = \frac{1}{2} \begin{bmatrix} y_2 - y_4 & y_3 - y_1 & y_4 - y_2 & y_1 - y_3 \\ x_4 - x_2 & x_1 - x_3 & x_2 - x_4 & x_3 - x_1 \end{bmatrix}.$$

For the central-difference method, Eqn. 6.54 leads to

$$\Delta t \leq \frac{2}{c_d \sqrt{g}}.$$

For a rectangular element with sides h_1 and h_2, the above formula reduces to

$$\Delta t \leq \frac{1}{c_d \left(\frac{1}{h_1^2} + \frac{1}{h_2^2} \right)^{1/2}}.$$

Appendix A

Stiffness and mass matrices for the Timoshenko beam element

In what follows, A, I, J, and L denote the area of cross-section, second moment of the cross-sectional area, polar moment of inertia, and length of the beam, and E, G, C, ρ, k denote the Young modulus, shear modulus, torsional rigidity, density and shear correction factor (usually taken to be 5/6), respectively. We assume all geometric and material properties to be uniform along the length of the beam element, though, of course, they can differ from element to element.

Following the notation of Bathe [1], let $(w_1, \theta_x^1, \theta_y^1, w_2, \theta_x^2, \theta_y^2)$, and $(w_1, \theta_x^1, \theta_y^1, w_2, \theta_x^2, \theta_y^2, w_3, \theta_x^3, \theta_y^3)$ be the degrees of freedom associated with a linear and quadratic beam element, respectively, and let $\alpha = EI/(GAk)$ and $\beta = C/(GAk)$.

Linear beam element:

The mass matrix for a beam along the x-axis is given by

$$M = \frac{\rho L}{6} \begin{bmatrix} 2A & 0 & 0 & A & 0 & 0 \\ 0 & 2J & 0 & 0 & J & 0 \\ 0 & 0 & 2I & 0 & 0 & I \\ A & 0 & 0 & 2A & 0 & 0 \\ 0 & J & 0 & 0 & 2J & 0 \\ 0 & 0 & I & 0 & 0 & 2I \end{bmatrix}.$$

For a beam along the y-axis, interchange I and J in the above matrix.

The stiffness matrices for beams along the x and y-axes are given by

$$\boldsymbol{K}_x = GAk \begin{bmatrix} \frac{1}{L} & 0 & -\frac{1}{2} & -\frac{1}{L} & 0 & -\frac{1}{2} \\ 0 & \frac{\beta}{L} & 0 & 0 & -\frac{\beta}{L} & 0 \\ -\frac{1}{2} & 0 & \frac{L}{4}+\frac{\alpha}{L} & \frac{1}{2} & 0 & \frac{L}{4}-\frac{\alpha}{L} \\ -\frac{1}{L} & 0 & \frac{1}{2} & \frac{1}{L} & 0 & \frac{1}{2} \\ 0 & -\frac{\beta}{L} & 0 & 0 & \frac{\beta}{L} & 0 \\ -\frac{1}{2} & 0 & \frac{L}{4}-\frac{\alpha}{L} & \frac{1}{2} & 0 & \frac{L}{4}+\frac{\alpha}{L} \end{bmatrix},$$

$$\boldsymbol{K}_y = GAk \begin{bmatrix} \frac{1}{L} & \frac{1}{2} & 0 & -\frac{1}{L} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{L}{4}+\frac{\alpha}{L} & 0 & -\frac{1}{2} & \frac{L}{4}-\frac{\alpha}{L} & 0 \\ 0 & 0 & \frac{\beta}{L} & 0 & 0 & -\frac{\beta}{L} \\ -\frac{1}{L} & -\frac{1}{2} & 0 & \frac{1}{L} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{L}{4}-\frac{\alpha}{L} & 0 & -\frac{1}{2} & \frac{L}{4}+\frac{\alpha}{L} & 0 \\ 0 & 0 & -\frac{\beta}{L} & 0 & 0 & \frac{\beta}{L} \end{bmatrix}.$$

Quadratic element:

The mass matrix for a beam along the x-axis is

$$\boldsymbol{M} = \frac{\rho L}{30} \begin{bmatrix} 4A & 0 & 0 & 2A & 0 & 0 & -A & 0 & 0 \\ 0 & 4J & 0 & 0 & 2J & 0 & 0 & -J & 0 \\ 0 & 0 & 4I & 0 & 0 & 2I & 0 & 0 & -I \\ 2A & 0 & 0 & 16A & 0 & 0 & 2A & 0 & 0 \\ 0 & 2J & 0 & 0 & 16J & 0 & 0 & 2J & 0 \\ 0 & 0 & 2I & 0 & 0 & 16I & 0 & 0 & 2I \\ -A & 0 & 0 & 2A & 0 & 0 & 4A & 0 & 0 \\ 0 & -J & 0 & 0 & 2J & 0 & 0 & 4J & 0 \\ 0 & 0 & -I & 0 & 0 & 2I & 0 & 0 & 4I \end{bmatrix}.$$

The mass matrix for a beam along the y-axis is obtained by interchanging I and J in the above equation.

The stiffness matrices for beams along the x and y-axes are given by

$$\boldsymbol{K}_x = GAk \begin{bmatrix} \frac{7}{3L} & 0 & -\frac{1}{2} & -\frac{8}{3L} & 0 & -\frac{2}{3} & \frac{1}{3L} & 0 & \frac{1}{6} \\ 0 & \frac{7\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{\beta}{3L} & 0 \\ -\frac{1}{2} & 0 & \frac{7\alpha}{3L}+\frac{L}{9} & \frac{2}{3} & 0 & -\frac{8\alpha}{3L}+\frac{L}{9} & -\frac{1}{6} & 0 & \frac{\alpha}{3L}-\frac{L}{18} \\ -\frac{8}{3L} & 0 & \frac{2}{3} & \frac{16}{3L} & 0 & 0 & -\frac{8}{3L} & 0 & -\frac{2}{3} \\ 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{16\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} & 0 \\ -\frac{2}{3} & 0 & -\frac{8\alpha}{3L}+\frac{L}{9} & 0 & 0 & \frac{16\alpha}{3L}+\frac{4L}{9} & \frac{2}{3} & 0 & -\frac{8\alpha}{3L}+\frac{L}{9} \\ \frac{1}{3L} & 0 & -\frac{1}{6} & -\frac{8}{3L} & 0 & \frac{2}{3} & \frac{7}{3L} & 0 & \frac{1}{2} \\ 0 & \frac{\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{7\beta}{3L} & 0 \\ \frac{1}{6} & 0 & \frac{\alpha}{3L}-\frac{L}{18} & -\frac{2}{3} & 0 & -\frac{8\alpha}{3L}+\frac{L}{9} & \frac{1}{2} & 0 & \frac{7\alpha}{3L}+\frac{L}{9} \end{bmatrix},$$

$$\boldsymbol{K}_y = GAk \begin{bmatrix} \frac{7}{3L} & \frac{1}{2} & 0 & -\frac{8}{3L} & \frac{2}{3} & 0 & \frac{1}{3L} & -\frac{1}{6} & 0 \\ \frac{1}{2} & \frac{7\alpha}{3L}+\frac{L}{9} & 0 & -\frac{2}{3} & -\frac{8\alpha}{3L}+\frac{L}{9} & 0 & \frac{1}{6} & \frac{\alpha}{3L}-\frac{L}{18} & 0 \\ 0 & 0 & \frac{7\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{\beta}{3L} \\ -\frac{8}{3L} & -\frac{2}{3} & 0 & \frac{16}{3L} & 0 & 0 & -\frac{8}{3L} & \frac{2}{3} & 0 \\ \frac{2}{3} & -\frac{8\alpha}{3L}+\frac{L}{9} & 0 & 0 & \frac{16\alpha}{3L}+\frac{4L}{9} & 0 & -\frac{2}{3} & -\frac{8\alpha}{3L}+\frac{L}{9} & 0 \\ 0 & 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{16\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} \\ \frac{1}{3L} & \frac{1}{6} & 0 & -\frac{8}{3L} & -\frac{2}{3} & 0 & \frac{7}{3L} & -\frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{\alpha}{3L}-\frac{L}{18} & 0 & \frac{2}{3} & -\frac{8\alpha}{3L}+\frac{L}{9} & 0 & -\frac{1}{2} & \frac{7\alpha}{3L}+\frac{L}{9} & 0 \\ 0 & 0 & \frac{\beta}{3L} & 0 & 0 & -\frac{8\beta}{3L} & 0 & 0 & \frac{7\beta}{3L} \end{bmatrix}.$$

Bibliography

[1] Bathe, K. J., *Finite Element Procedures*, Prentice-Hall, New Jersey, 1997.

[2] Shames, I. H. and C. Dym, *Energy and Finite Element Methods in Structural Mechanics*, Hemisphere Publishing Co., 1985.

[3] Hughes, T. J. R., *Linear, Static and Dynamic Finite Element Analysis*, Prentice Hall, 1987.

[4] Zienkiewicz, O. C. and R. L. Taylor, *The Finite-Element Method*, Vol. 1. McGraw-Hill International, 1989