Chapter 1

Review of the Finite Element Method

Chapter Points

- The ends of words and sentences are marked by spaces. It does not matter how many spaces you type; one is as good as 100. The end of a line counts as a space.
- The ends of words and sentences are marked by spaces. It does not matter how many spaces you type; one is as good as 100. The end of a line counts as a space.

In the early 1950s, the research conducted by Jon Turner and coworkers [18] during the Boeing Summer Faculty Program led to the development of what was later named by Ray Clough the finite element method (FEM) [19]. Fueled initially by engineers and applied scientists seeking to find solutions to real problems, the methodology received later in the 70s the attention of the mathematical rigor that led to a deeper understanding [20]. The development of FEM took place hand in hand with the development of computers, and its indisputable success, in part due to its robustness, was reflected in the vast number of finite element codes only a few years after its introduction (650 codes by the end of 1970s) [21].

FEM is a numerical procedure for finding approximate solutions to boundary and initial value problems. Initially proposed for structural analysis, today FEM is considered the standard procedure for solving most problems in solid mechanics. In a nutshell, the partial differential equations that describe a physical phenomenon are discretized by FEM, leading to a discrete algebraic system of equations that is solved to find the unknown primary field, *e.g.*, displacements or temperatures. This chapter aims at giving a short introduction to FEM. The material here should by no means be taken as a thorough treaty on the subject, for which the reader is referred to classic textbooks [22].

1.1 LINEAR ELASTOSTATICS IN ONE DIMENSION

In this first section we review the finite element method for a 1-D bar—a structural element that can only be subjected to axial loads. By studying this simple problem we will gather insight that will ease the presentation of the finite element formulation in higher dimensions.

Figure 1.1 shows a bar[1] $\Omega \subset \mathbb{R}$ of length l and cross-sectional area $A(x)$.

1. The bar is mathematically represented as an open set $\Omega \subset \mathbb{R}$. Given its closure $\overline{\Omega}$, *i.e.*, the

An infinitesimal bar segment of length Δx shows the forces acting on it, which include the load per unit length $b\,(x)$, a displacement-dependent force in direction opposite to the displacement $s\,(x)\,u\,(x)$ (where s is the stiffness per unit length), and the axial forces at the segment ends $N\,(x)$ and $N\,(x+\Delta x)$. The dependence on position x of these quantities is implied henceforth.

The static equilibrium of the segment is obtained by

$$N\,(x+\Delta x) - N\,(x) - su\Delta x + b\Delta x = 0,$$

and after divicing by Δx and taking the limit as $\Delta x \to 0$,

$$\lim_{\Delta x \to 0} \frac{N\,(x+\Delta x) - N\,(x)}{\Delta x} - su + b = 0,$$

$$\frac{\mathrm{d}N}{\mathrm{d}x} - su + b = 0,$$

We assume that the material of the bar behaves linearly according to Hooke's law, *i.e.*, the stress σ_x and strain $\varepsilon_x \equiv \frac{\mathrm{d}u}{\mathrm{d}x}\frac{\mathrm{d}u}{\mathrm{d}x}$ are related through $\sigma_x = E\varepsilon_x$, where $E\,(x)$ is Young's modulus. The axial force can thus be written as $N = \sigma_x A = EA\frac{\mathrm{d}u}{\mathrm{d}x}$. Therefore, the equilibrium at every point along the bar is given by

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right) - su + b = 0 \qquad \forall\, x \in \Omega, \tag{1.1}$$

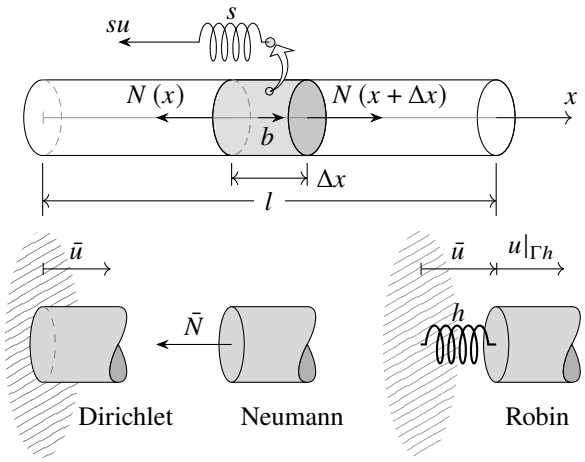

FIGURE 1.1 (top) Schematic of a 1-D bar used to set up the boundary value problem; (bottom) Boundary condition types.

open set plus its endpoints, the boundary (in this case the left and right ends) is then defined as $\partial\Omega = \Gamma \equiv \overline{\Omega} \setminus \Omega$.

Eq. (1.1) is a second-order ordinary differential equation, and thus its solution requires two integration constants that are determined through boundary conditions (BCs). Three types of BCs can be applied to our 1-D bar (*cf.* Figure **??**):

- *Essential* or *Dirichlet* BC on the *primal* variable, which in this case is the displacement of the bar u. Denoting as Γ_u the part of the boundary with prescribed displacement (either or both bar ends), this condition is stated as

$$u|_{\Gamma_\mathrm{u}} = \bar{u}.$$

- *Natural* or *Neumann* BC imposed on the *dual* variable, which for this case is the axial load N. With Γ_t denoting the part of the boundary with prescribed axial force, this BC is usually expressed as

$$EA\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_{\Gamma_\mathrm{t}} n = \bar{N},$$

  where n is either -1 or 1 for the left and right end, respectively. This is a consequence of our sign convention since a positive force on the left end would produce a negative strain (compression in the bar). Similarly, a positive force at the right end would produce tension and thus a positive strain.
- *Mixed* or *Robin* BC, which can be seen as a combination of the previous two. As before, we will denote Γ_h the part of the boundary with prescribed mixed condition, so

$$EA\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_{\Gamma_\mathrm{h}} n = h\left(\bar{u} - u|_{\Gamma_\mathrm{h}}\right),$$

  where for the case of the bar h denotes the spring that links the bar[2]. The same reasoning for the sign applies here, and $\bar{u} > u|_{\Gamma_h}$ produces a positive force that compresses the bar.

It is not possible to prescribe more than one type of BC at the same end of the bar. Thus, boundary conditions are *disjoint*, a condition usually expressed as $\Gamma_\mathrm{u} \cap \Gamma_\mathrm{t} = \emptyset$, or $\Gamma_\mathrm{u} \cap \Gamma_\mathrm{h} = \emptyset$, or $\Gamma_\mathrm{t} \cap \Gamma_\mathrm{h} = \emptyset$.

1.1.1 The strong form

At this point we have everything we need to formally state the *strong form* of the 1-D elastostatics boundary value problem: Given the Young's modulus $E : \overline{\Omega} \to \mathbb{R}$ in $[\mathrm{F/L}^2]$ (force per length squared) units, the cross sectional area $A : \overline{\Omega} \to \mathbb{R}$ in $[\mathrm{L}^2]$, the distributed spring $k : \Omega \to \mathbb{R}$ in $[\mathrm{F/L}^2]$, and the distributed axial load $b : \Omega \to \mathbb{R}$ in $[\mathrm{F/L}]$, find the displacement field $u \in C^2$ such that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right) - su + b = 0 \qquad \forall x \in \Omega,$$

2. The symbol h was chosen because this type of boundary condition is more common in heat transfer problems, where h refers to the heat transfer coefficient, as described later in this chapter.

with boundary conditions

$$u = \bar{u} \quad \text{on } \Gamma_{\mathrm{u}} \quad \text{and} \quad EA\frac{\mathrm{d}u}{\mathrm{d}x}n = \bar{N} \quad \text{on } \Gamma_{\mathrm{t}},$$

or

$$u = \bar{u} \quad \text{on } \Gamma_{\mathrm{u}} \quad \text{and} \quad EA\frac{\mathrm{d}u}{\mathrm{d}x}n = h\left(\bar{u} - u|_{\Gamma_{\mathrm{h}}}\right) \quad \text{on } \Gamma_{\mathrm{h}},$$

or

$$EA\frac{\mathrm{d}u}{\mathrm{d}x}n = \bar{N} \quad \text{on } \Gamma_{\mathrm{t}}, \quad \text{and} \quad EA\frac{\mathrm{d}u}{\mathrm{d}x}n = h\left(\bar{u} - u|_{\Gamma_{\mathrm{h}}}\right) \quad \text{on } \Gamma_{\mathrm{h}}.$$

Notice that if $\Gamma = \overline{\Omega} \setminus \Omega = \Gamma_{\mathrm{u}}$ we are dealing with a pure Dirichlet problem where the displacement is prescribed at both ends of the bar. Following a similar reasoning, we could also be dealing with a pure Neumann or pure Robin problem. For the pure Neumann problem, which results in a singular system of equations, the solution can only be determined up to a constant.

Example 1 (**Bi-material interface**). Consider a bar of length l clamped at the left end and subjected to a load P at the right end, as shown in Figure 1.2. A material interface at $w = x_\Gamma/l$ subdivides the bar in two parts Ω_i, $i = 1, 2$ with corresponding axial stiffness $k_i = E_i A_i$. State the strong form for this problem and obtain the exact displacement field analytically.

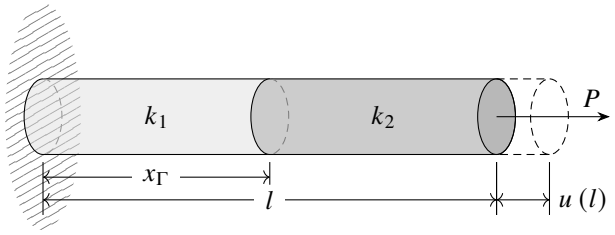

FIGURE 1.2 1-D bar composed of two materials with interface at x_Γ/l.

Solution: As long as the axial stiffness in each subdomain is different, the displacement field has a discontinuous gradient at the interface. In other words, the displacement field is C^0–continuous. The strong form for this problem can be stated as: Given the axial stiffness $k_i : \Omega_i \to \mathbb{R}$, $i = 1, 2$, the prescribed displacement $\bar{u}$ and the prescribed load P, find the displacement $u \in C^0$ such that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(k_i \frac{\mathrm{d}u_i}{\mathrm{d}x}\right) &= 0 \quad && \text{in } \Omega_i, \\ u_1 &= 0 \quad && \text{at } x = 0, \\ k_2 \frac{\mathrm{d}u_2}{\mathrm{d}x} &= P \quad && \text{at } x = l, \end{aligned}$$

with interface conditions at x_Γ (continuity of displacement and forces)

$$u_1 = u_2 \qquad \text{and} \qquad k_1 \frac{\mathrm{d}u_1}{\mathrm{d}x} = k_2 \frac{\mathrm{d}u_2}{\mathrm{d}x}.$$

The solution to this equation is simply obtained by integrating the equilibrium equation twice:

$$k_i \frac{\mathrm{d}u_i}{\mathrm{d}x} = C_i, \qquad u_i = \frac{1}{k_i}\left(C_i x + D_i\right). \tag{1.2}$$

The four constants of integration (two for each subdomain) are obtained by applying boundary and interface conditions:

$$\begin{aligned}
u_1(0) = \frac{D_1}{k_1} = 0 \qquad &\Rightarrow \quad D_1 = 0,\\
C_1 = k_1 \frac{\mathrm{d}u_1}{\mathrm{d}x}\bigg|_{x_\Gamma} = k_2 \frac{\mathrm{d}u_2}{\mathrm{d}x}\bigg|_{x_\Gamma} = C_2 \qquad &\Rightarrow \quad C_1 = C_2,\\
C_2 = k_2 \frac{\mathrm{d}u_2}{\mathrm{d}x}\bigg|_{l} = P \qquad &\Rightarrow \quad C_2 = P = C_1,\\
\frac{1}{k_1}\left(P x_\Gamma\right) = \frac{1}{k_2}\left(P x_\Gamma + D_2\right) \qquad &\Rightarrow \quad D_2 = \frac{x_\Gamma \llbracket k \rrbracket P}{k_1},
\end{aligned}$$

where $\llbracket k \rrbracket = k_2 - k_1$ represents the jump in axial stiffness at x_Γ. By replacing these constants in (1.2), the final displacement field is given by

$$u(x) = \begin{cases} \frac{Px}{k_1} & \text{for } x \le x_\Gamma, \\ \frac{Px_\Gamma}{k_1} + \frac{P(x - x_\Gamma)}{k_2} & \text{for } x \ge x_\Gamma. \end{cases} \tag{1.3}$$

The displacement field is shown in Figure 1.3 for $k_1/k_2 = 10$.

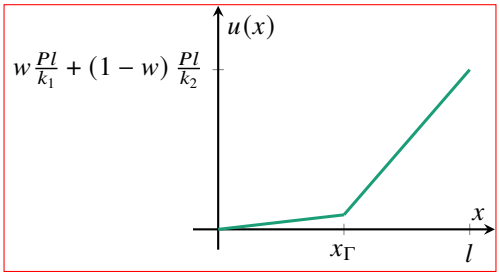

FIGURE 1.3 Displacement field given by Eq. 1.3 for $k_1/k_2 = 10$.

Example 2 (One-dimensional cracked bar with spring). Consider the 1-D bar of length l shown in Figure 1.4. The bar has axial stiffness $k = EA$, with A denoting cross-sectional area and E Young's modulus. The bar is composed by two parts joined at $x_\Gamma = l/2$ by a spring of constant $k_\Gamma > 0$. Both the material of the bar and the spring are assumed to be linear elastic. Obtain the exact solution to the problem.

Solution:

For this problem the strong form is: Given the axial stiffness k, the spring constant k_Γ, and prescribed displacement $u_1(0) = 0$ and force $k\frac{\mathrm{d}u_2}{\mathrm{d}x} = P$, find the displacement field $u \in C^{-1}$ such

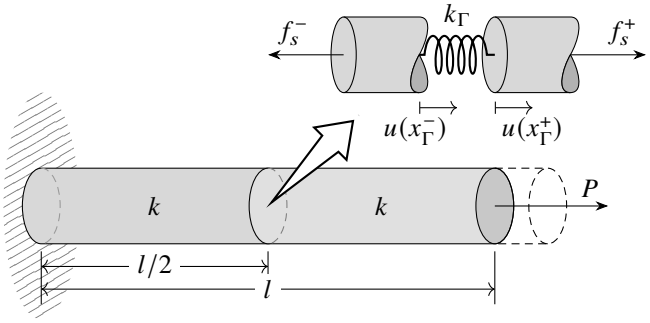

FIGURE 1.4 One-Dimensional cracked bar with spring.

that

$$\begin{aligned} k\frac{\mathrm{d}^2 u_i}{\mathrm{d}x^2} &= 0 && \text{in } \Omega_i, \\ u_1 &= 0 && \text{at } x = 0, \\ k\frac{\mathrm{d}u_2}{\mathrm{d}x} &= P && \text{at } x = l, \\ k\frac{\mathrm{d}u_1}{\mathrm{d}x} = k\frac{\mathrm{d}u_2}{\mathrm{d}x} &= k_\Gamma [\![u(x_\Gamma)]\!] && \text{at } x = x_\Gamma \end{aligned}$$

where $[\![u(x_\Gamma)]\!] = u(x_\Gamma^+) - u(x_\Gamma^-)$ denotes the displacement jump across discontinuity at $x = x_\Gamma$.

Similarly to Example 1, the solution is obtained by integrating the governing equation twice, obtaining a similar equation

$$k\frac{\mathrm{d}u_i}{\mathrm{d}x} = C_i, \qquad u_i = \frac{1}{k}\left(C_i x + D_i\right). \tag{1.4}$$

The integration constants are again found by using boundary and interface conditions:

$$\begin{aligned} u_1(0) = \frac{D_1}{k} = 0 \quad &\Rightarrow \quad D_1 = 0, \\ C_1 = k\frac{\mathrm{d}u_1}{\mathrm{d}x}\bigg|_{x_\Gamma} = k\frac{\mathrm{d}u_2}{\mathrm{d}x}\bigg|_{x_\Gamma} = C_2 \quad &\Rightarrow \quad C_1 = C_2, \\ C_2 = k\frac{\mathrm{d}u_2}{\mathrm{d}x}\bigg|_{l} = P \quad &\Rightarrow \quad C_2 = P = C_1, \\ P = k_\Gamma [\![u(x_\Gamma)]\!] \quad &\Rightarrow \quad [\![u(x_\Gamma)]\!] = \frac{P}{k_\Gamma} \\ [\![u(x_\Gamma)]\!] = \frac{1}{k}\left(P x_\Gamma + D_2\right) - \frac{P x_\Gamma}{k} = \frac{P}{k_\Gamma} \quad &\Rightarrow \quad D_2 = \frac{k}{k_\Gamma} P, \end{aligned}$$

The exact solution to this problem, which is shown in Figure 1.5, is

$$u(x) = \begin{cases} \frac{Px}{k} & \text{for } x < x_\Gamma, \\ \frac{Px}{k} + \frac{P}{k_\Gamma} & \text{for } x \geq x_\Gamma, \end{cases} \tag{1.5}$$

Example 3 (Constrained bar pullout). Consider a bar of constant axial stiffness EA (as shown on the right figure), constrained by a semi-infinite elastic foundation with stiffness k. Since there is no body force, the equilibrium equation (1.1)

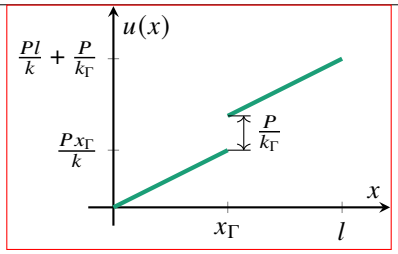

FIGURE 1.5 Displacement field given by Eq. 1.5.

simplifies to

$$EA\frac{\mathrm{d}u}{\mathrm{d}x}[2] - su = 0. \tag{1.6}$$

The boundary conditions include $u(\infty) = 0$ and $N(0) = P$. Find the exact solution of (1.6). With the exact solution, compute the strain energy as

$$U = \frac{1}{2}\int_0^\infty \left[EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 + su^2\right]\mathrm{d}x\,. \tag{1.7}$$

Solution: Eq. 1.6 can be rewritten as

$$\frac{\mathrm{d}u}{\mathrm{d}x}[2] = \alpha^2 u,$$

with $\alpha^2 = s/EA$ representing the foundation stiffness relative to that of the bar. This equation has solution $u = C_1 e^{\alpha x} + C_2 e^{-\alpha x}$ where the constants are determined by the boundary conditions. We see that for the Dirichlet boundary condition to hold, $C_1 = 0$ for only the second term decays as $x \to \infty$. To obtain C_2 we use the Neumann boundary condition:

$$EA\left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_{x=0} n = EAC_2\alpha = -P \qquad \Rightarrow \qquad C_2 = -\frac{P}{\sqrt{EAk}},$$

so the final solution is

$$u = -\frac{P}{EA\alpha}e^{-\alpha x}. \tag{1.8}$$

The solution can be easily verified by applying the derivative twice:

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{P}{EA}e^{-\alpha x} \qquad \Rightarrow \qquad \frac{\mathrm{d}u}{\mathrm{d}x}[2] = -\frac{P}{EA}\alpha e^{-\alpha x} = \alpha^2 u.$$

The solution for different values of α is shown in Figure 1.6. As expected, as the stiffness of the foundation relative to that of the bar decreases (lower values of α), the displacement along the bar is felt at deeper coordinates (increasing values of x). Also, the effect of a stiffer foundation relative to the bar translates into a sharper gradient in the displacement field close to $x = 0$.

With u given by (1.8), the energy in the system is

$$U = \lim_{x\to\infty}\frac{1}{2}\int_0^x \left[EA\left(\frac{du}{dx}\right)^2 + su^2\right]\mathrm{d}x = \lim_{x\to\infty}\frac{e^{-2\alpha x}\left(e^{2\alpha x} - 1\right)\alpha P^2}{2s} = \frac{\alpha P^2}{2s}. \tag{1.9}$$

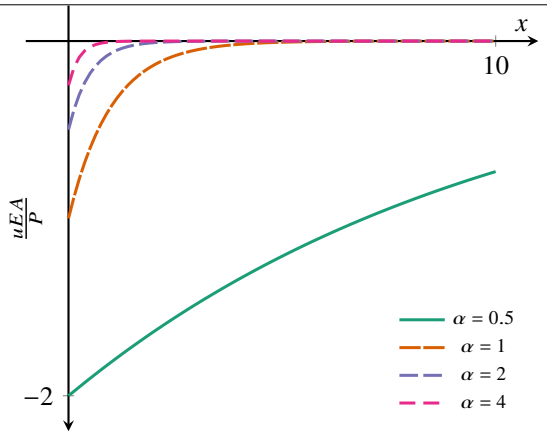

FIGURE 1.6 uEA/P for different values of α.

1.1.2 The weak form

Obtaining the solution of the strong form, that is, satisfying equilibrium and boundary conditions, is not straightforward even for simple load configurations. Thus we seek a *weak* formulation where equilibrium is satisfied on *average*.

The weak formulation can be obtained by either the principle of virtual work, by minimizing the potential energy of the bar, or by the method of weighted residuals. Here we adopt the latter as it is more general—not every differential equation can be derived from a potential. As mentioned above, the objective is to satisfy equilibrium on average or integral sense along the bar, thus we multiply Eq. (1.1) by a weight function w and set the integral over the domain $\Omega = [0, l]$ to zero:

$$\int_0^l w \left[\frac{\mathrm{d}}{\mathrm{d}x} \left(EA \frac{\mathrm{d}u}{\mathrm{d}x} \right) - ku + b \right] \mathrm{d}x = 0, \qquad \forall\, w \in \mathcal{E}(\Omega), \tag{1.10}$$

where $\mathcal{E}$ is the *energy space*, defined by the set of functions that have finite strain

energy $U(u)$[3]:

$$\mathcal{E}(\Omega) := \left\{ v |\ U(v) \equiv \frac{1}{2} \int_0^l \left[EA \left(\frac{\mathrm{d}v}{\mathrm{d}x} \right)^2 + kv^2 \right] \mathrm{d}x < \infty \right\}, \tag{1.11}$$

which induces the norm

$$\|v\|_{\mathcal{E}(\Omega)} := \sqrt{U(v)}. \tag{1.12}$$

In Eq. (1.10) u is not necessarily the "exact solution" anymore so it is called the trial function. The equation requires that the trial solution be differentiable twice, so we now perform integration by parts on the first term to lower this requirement. That is to say, we integrate by parts to balance the derivative orders, and thus to reduce the differentiability requirement on the trial function u:

$$\int_0^l w \frac{\mathrm{d}}{\mathrm{d}x} \left(EA \frac{\mathrm{d}u}{\mathrm{d}x} \right) \mathrm{d}x = \underbrace{wEA \frac{\mathrm{d}u}{\mathrm{d}x} \Big|_0^l}_{\text{boundary terms}} - \int_0^l EA \frac{\mathrm{d}w}{\mathrm{d}x} \frac{\mathrm{d}u}{\mathrm{d}x} \mathrm{d}x \,. \tag{1.13}$$

Noteworthy, the boundary terms on the right-hand side are replaced by considering the actual boundary conditions for the problem at hand. For instance, $EA \frac{\mathrm{d}u}{\mathrm{d}x}$ would be replaced by $\bar{N}$ if the BC at the corresponding end is of Neumann type, or by $h\left(\bar{u} - u|_{\Gamma_h}\right)$ if it is of Robin type. If the BC is of Dirichlet type, $EA \frac{\mathrm{d}u}{\mathrm{d}x}$ is replaced by the reaction force R. Since at this stage we have no concrete boundary conditions for our model boundary value problem, we will use the "boundary terms" notation.

By replacing (1.13) into (1.10) and collecting terms, the weak formulation is then stated: Find $u \in \mathcal{E}(\Omega)$ such that

$$B(u, w) = L(w) \qquad \forall\, w \in \mathcal{E}(\Omega), \tag{1.14}$$

3. More formally, for solving partial differential equations both trial and weight functions are chosen from Hilbertian Sobolev spaces $\mathcal{H}^k(\Omega)$, which comprises the set of functions that are square integrable in Ω, and which contains derivatives up to order k which are also square integrable in Ω. For our 1-D bar our space is $\mathcal{H}^1(\Omega)$, and it is mathematically described as

$$\mathcal{H}^1(\Omega) \equiv \left\{ w \in L^2(\Omega) : \frac{\mathrm{d}w}{\mathrm{d}x} \in L^2(\Omega) \right\}.$$

A function w is *square integrable* on Ω, *i.e.*, it belongs to $L^2(\Omega)$ if

$$\int_\Omega w^2 \, \mathrm{d}x < \infty.$$

where the *bilinear form*[4] is given by

$$B(u,w) = \int_0^l \left(EA \frac{\mathrm{d}w}{\mathrm{d}x} \frac{\mathrm{d}u}{\mathrm{d}x} + kwu \right) \mathrm{d}x\,, \tag{1.15}$$

and the *linear form* by

$$L(w) = \int_0^l wb \ \mathrm{d}x + \text{boundary terms}. \tag{1.16}$$

Noteworthy, by comparing Eqs. (1.11) and (1.12) with (1.15) it can be seen that

$$U(u) \equiv \frac{1}{2} B(u,u) \qquad \text{and} \qquad \|u\|_{\mathcal{E}(\Omega)} = \sqrt{\tfrac{1}{2} B(u,u)}. \tag{1.17}$$

Consider the case when dealing with homogeneous Dirichlet boundary conditions, *i.e.*, $\Gamma_\mathrm{u} \neq \emptyset$, $\bar{u}|_{\Gamma_\mathrm{u}} = 0$. Since Eq. (1.14) has to be valid for any function w, certain choices of weighting function would simplify the equation to be solved. For instance, by defining

$$\mathcal{E}_0(\Omega) := \left\{ w |\ w \in \mathcal{E}(\Omega)\,,\ w|_{\Gamma_\mathrm{u}} = 0 \right\},$$

choosing a weight function $w \in \mathcal{E}_0$ causes the term containing the reaction force to vanish from the formulation.

Formally, when dealing with non-homogenous Dirichlet boundary conditions we need to take a special consideration regarding the choice of function spaces. Defining the following spaces

$$\mathcal{V}(\Omega) = \left\{ v : v(x) \in \mathbb{R},\ \ v \in \mathcal{H}^1(\Omega) \right\}, \tag{1.18}$$

$$\mathcal{V}_0(\Omega) = \left\{ v \in \mathcal{V}(\Omega) :\ v|_{\Gamma_\mathrm{u}} = 0 \right\}, \tag{1.19}$$

where $\mathcal{H}^1(\Omega)$ denotes the first-order Hilbertian Sobolev function space on Ω (as defined in Footnote 3). We define a *linear variety* as $\mathcal{V}_\star \equiv \tilde{u} + \mathcal{V}_0$, where $\tilde{u} \in \mathcal{V} :\ \tilde{u}|_{\Gamma_\mathrm{u}} = \bar{u}$. $\mathcal{V}_\star$ is a translation of the space $\mathcal{V}_0$ by $\tilde{u}$ so that every element in $\mathcal{V}_\star$ satisfies the non-homogeneous Dirichlet boundary condition. The weak form of the boundary value problem can then be expressed as: Find the displacement field $u \in \mathcal{V}_\star$ such that

$$B(u,w) = L(w) \qquad \forall w \in \mathcal{V}_0, \tag{1.20}$$

4. For $u, v, w \in \mathcal{V}$ and $c \in \mathbb{R}$, the bilinear form has the following properties:

$$\begin{aligned} B(u,u) &\geq 0, \\ B(u,v) &= B(v,u)\,, \\ B(u+v,w) &= B(u,w) + B(v,w)\,, \\ B(cu,v) &= cB(u,v)\,. \end{aligned}$$

which is equivalent to finding the "unknown" part of solution from the space $\mathcal{V}_0$ and add it to a known function $\tilde{u}$ that satisfies the essential boundary conditions.

Equivalently, we could also write: Given $u = v + \tilde{u}$, find $v \in \mathcal{V}_0$ such that

$$B(v, w) = L(w) - B(\tilde{u}, w) \qquad \forall\, w \in \mathcal{V}_0(\Omega). \tag{1.21}$$

This last expression explicitly states that the known non-zero essential boundary condition modifies the right hand side (which will eventually become the force vector once (1.21) is discretized, as explained later in § 1.1.4). We will use this last expression henceforth for defining the weak problem statement when non-homogeneous essential boundary conditions are present.

If at any end the BC is of Robin type, two terms are added to the formulation from which $h\, u|_{\Gamma_u}$ is added to the bilinear form and $h\bar{u}$ to the linear form.

Example 4 (One-dimensional cracked bar with spring (revisited)). In Exercise 2 we looked at obtaining the exact solution for a 1-D cracked bar that was connected with a linear spring (refer to Fig. 1.4). Derive the weak formulation for this problem.

Solution: The weak or variational formulation is equivalent to the principle of virtual work. Therefore, we obtain the virtual work done by the bar and by the spring. The virtual work done by the external force is simply $W_e = v(l)\, P$, where we denote $v = \delta u$ our virtual displacement field. The virtual work of internal forces in the bar is

$$W_i = \int_0^l \frac{\mathrm{d}v}{\mathrm{d}x} k \frac{\mathrm{d}u}{\mathrm{d}x}\, \mathrm{d}x,$$

where we note that $\frac{\mathrm{d}v}{\mathrm{d}x}$ is the virtual strain. To derive the virtual work of internal forces on the spring, Figure 1.4 shows the free body diagram of the spring, and the jump in displacement and virtual displacement at $x = x_\Gamma = l/2$. The force on the linear spring is given by

$$f_s^+ = f_s^- = k_\Gamma [\![u(x_\Gamma)]\!],$$

where $[\![u(x_\Gamma)]\!] = u(x_\Gamma^+) - u(x_\Gamma^-)$ denotes the displacement jump across discontinuity at $x = x_\Gamma$. The virtual work of f_s^+ and f_s^- are

$$\begin{aligned} {}^I W_s^+ &= f_s^+ v(x_\Gamma^+), \\ {}^I W_s^- &= -f_s^- v(x_\Gamma^-), \end{aligned}$$

so the virtual work of internal forces on the spring is given by

$${}^I W_s^+ + {}^I W_s^- = f_s^+ v(x_\Gamma^+) - f_s^+ v(x_\Gamma^-) = f_s^+ [\![v(x_\Gamma)]\!] = k_\Gamma [\![u(x_\Gamma)]\!] [\![v(x_\Gamma)]\!].$$

The weak formulation of the problem can then be stated as: Find $u(x) \in \mathcal{U}_0$ such that

$$\int_0^l \frac{\mathrm{d}v}{\mathrm{d}x} k \frac{\mathrm{d}u}{\mathrm{d}x}\, \mathrm{d}x + [\![v(x_\Gamma)]\!] k_\Gamma [\![u(x_\Gamma)]\!] = v(l) P \qquad \forall v(x) \in \mathcal{U}_0,$$

with

$$\mathcal{U}_0 = \left\{ u : \int_0^l \frac{1}{2} k \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 \mathrm{d}x + \frac{1}{2} k_\Gamma [\![u(x_\Gamma)]\!]^2 < \infty,\ u(0) = 0 \right\}.$$

1.1.3 The Galerkin formulation

The weak formulation given by Eq. (1.14) implies that the solution be found in the infinite-dimensional function space $\mathcal{E}(\Omega)$. In other words, we seek a function of the form

$$u(x) = \sum_{i=1}^{\infty} \varphi_i(x) U_i, \tag{1.22}$$

where the functions $\varphi_i : \Omega \to \mathbb{R}$ form a basis that spans $\mathcal{E}(\Omega)$, and $U_i \in \mathbb{R}$ are coefficients to those basis functions usually referred to as *degrees of freedom (DOFs)*. One can immediately foresee that such approach is not practical. Instead of searching for the solution in such space, we would rather relax the requirement on $u(x)$ by searching for it in a finite-dimensional space $\mathcal{E}^h(\Omega) \subset \mathcal{E}(\Omega)$. That is to say, by truncating (1.22) to a finite number of terms, we accept the fact that an error may be associated with our trial solution $u^h(x) \in \mathcal{E}^h(\Omega)$. Nevertheless, this poses no problem as long as we have information about the error and we know how to reduce it, which we will investigate later in this chapter.

Our finite-dimensional problem statement now reads: Find $u^h \in \mathcal{E}^h(\Omega)$ such that

$$B(u^h, w^h) = L(w^h) \qquad \forall\, w^h \in \mathcal{E}^h(\Omega), \tag{1.23}$$

where the linear and bilinear forms now use trial and weight functions taken from the finite-dimensional space $\mathcal{E}^h(\Omega)$. Eq. (1.23) is usually referred to as the Galerkin formulation. If u^h and w^h are taken from the same space the procedure is called Bubnov-Galerkin, and otherwise it is called Petrov-Galerkin.

Eq. (1.23) could be solved already for the entire bar *à la* Rayleigh-Ritz by choosing trial and weight functions that act *globally* on the entire bar and which are kinematically admissible (*i.e.*, satisfy *a priori* the boundary conditions). For instance, one could take the trial solution to be a truncated Fourier series $u^h(x) = \sum_{j=1}^{n} a_j \cos(j\pi x) + \sum_{j=1}^{n} b_j \sin(j\pi x)$, $a_j, b_j \in \mathbb{R}$.

1.1.3.0.1 Orthogonality of Galerkin error

The weak formulation (1.14) and its discrete counterpart (1.23) can be used to give some insight into the solution obtained by the Galerkin form (1.23). In Eq. (1.14) u denotes the exact solution, and since such equation is valid for all $w \in \mathcal{E}(\Omega)$,

$$B(u, w^h) = L(w^h) \qquad \forall\, w^h \in \mathcal{E}^h(\Omega), \tag{1.24}$$

since $\mathcal{E}^h(\Omega) \subset \mathcal{E}(\Omega)$. Then subtracting Eq. (1.23) from (1.24) we get

$$B(u - u^h, w^h) = 0 \qquad \forall\, w^h \in \mathcal{E}^h(\Omega). \tag{1.25}$$

Eq. (1.25) states that the error in the approximation $u - u^h$ is orthogonal (with respect to the bilinear form) to the subspace $\mathcal{E}^h(\Omega)$. In this context, two functions f and g are said to be orthogonal with respect to the bilinear form B if $B(f, g) = 0$. In other words,

u^h is the projection (in the sense of the bilinear operator) of the exact solution onto the subspace $\mathcal{E}^h(\Omega)$. This means that from all $w^h \in \mathcal{E}^h(\Omega)$, the approximated solution u^h obtained by the Galerkin finite-dimensional formulation is the one that minimizes the error with respect to the exact solution. This is demonstrated also in the following theorem.

Theorem 1.1.1. *$u^h \in \mathcal{E}^h$ minimizes the error with respect to the exact solution, i.e.,*

$$u^h = \underset{v^h \in \mathcal{E}^h}{\arg\min} \left\| u - v^h \right\|_{\mathcal{E}^h} \tag{1.26}$$

We do a proof by contradiction: Consider that $v^h \in \mathcal{E}^h$ has a lower error, then

$$\begin{aligned}
\|e\|^2_{\mathcal{E}^h} &> \left\| u - v^h \right\|^2_{\mathcal{E}^h} \\
B(e,e) &> B(u - u^h + u^h - v^h, u - u^h + u^h - v^h) \\
\cancel{B(e,e)} &> \cancel{B(e,e)} + B(e, u^h - v^h) + B(u^h - v^h, e) + B(u^h - v^h, u^h - v^h) \\
0 &> \cancel{2B(e, u^h - v^h)} + B(u^h - v^h, u^h - v^h)
\end{aligned}$$

We first add zero and then use three properties of the bilinear form[4]. Then we use the symmetry property and Eq. (1.25). And thus, since $B(u,u) \geq 0$, then v^h cannot have a smaller error than u^h.

A generalization of Eq. (1.26) for *non-symmetric* bilinear forms is provided by Céa's approximation theorem [23]:

Theorem 1.1.2 (Céa's approximation theorem)**.** *Let $\mathcal{V}$ be a closed Hilbert space with norm $\|\cdot\|_{\mathcal{V}}$. Let $B : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ be a bilinear form, not necessarily symmetric with the following properties:*

- *B is continuous, and therefore bounded:*

$$|B(u,v)| \leq C \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}} \qquad C > 0, \forall u, v \in \mathcal{V}$$

- *B is coercive (bounded below) on $\mathcal{V}$, i.e., there exists a constant $\alpha > 0$ such that*

$$B(u,u) \geq \alpha \|u\|^2_{\mathcal{V}} \qquad \forall u \in \mathcal{V}$$

then

$$\left\| u - u^h \right\|_{\mathcal{V}} \leq \frac{C}{\alpha} \left\| u - v^h \right\|_{\mathcal{V}} \qquad v^h \in \mathcal{V}$$

Proof.

$$\begin{aligned}
\alpha \left\| u - u^h \right\|^2_{\mathcal{V}} \leq B\left(u - u^h, u - u^h\right) &= B\left(u - u^h, u - u^h + v^h - v^h\right) \\
&= B\left(u - u^h, u - v^h\right) + \cancel{B\left(u - u^h, v^h - u^h\right)} \\
&\leq C \left\| u - u^h \right\|_{\mathcal{V}} \left\| u - v^h \right\|_{\mathcal{V}} \qquad \forall v^h \in \mathcal{V}.
\end{aligned} \tag{1.27}$$

The cancelled term results from Eq. (1.25).

Theorem 1.1.3. *Given the energy space $\mathcal{E}^h(\Omega)$, the error $e \equiv u - u^h$ in the energy norm can be determined as*

$$\left\|u - u^h\right\|_{\mathcal{E}(\Omega)} = \sqrt{\frac{1}{2}\left[B(u,u) - B\left(u^h, u^h\right)\right]} \tag{1.28}$$

Proof.

$$\begin{aligned} 2\left\|u - u^h\right\|^2_{\mathcal{E}(\Omega)} &= B(e, u - u^h) \\ &= B(e,u) - \cancel{B(u - u^h, u^h)} = B(u - u^h, u) + B(u - u^h, u^h) \\ &= B(u,u) - \cancel{B(u^h, u)} + \cancel{B(u, u^h)} - B(u^h, u^h) = B(u,u) - B(u^h, u^h) \end{aligned}$$

In this proof the first cancellation is due to (1.25) and the second one due to the symmetry of the bilinear norm[4]. This is an important result that tell us we can compute the error in the approximation by just computing the strain energy of our approximate solution and compare that to the strain energy of the exact solution.

1.1.4 The finite element discrete equations

Instead of having functions that behave globally, one could divide the bar in a number of smaller segments and use simpler functions that act *locally* in those segments. Even if these functions are chosen to be low-order polynomials, they may still be able to represent complicated solutions because the discretization segments can be created as small as needed. This is the rationale behind the finite element method.

Let us subdivide our bar in Figure **??** at n node locations into $n-1$ finite elements such that $\overline{\Omega} = \cup_{i=1}^{n-1} \overline{e}_i$, $e_i \cap e_j = \emptyset$ for $i \neq j$, as shown in Figure 1.7a.

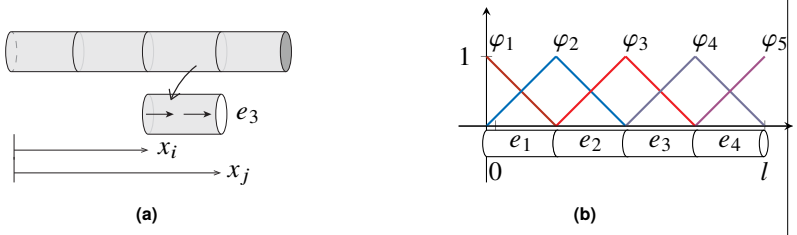

FIGURE 1.7 (a) 1-D bar discretized using finite elements; (b) Linear shape functions.

With a Bubnov-Galerkin projection, consider both the weight function and the trial solution of the form

$$w^h = \sum_{i=1}^{n} \varphi_i w_i = \boldsymbol{\varphi W} \quad \text{and} \quad u^h = \sum_{i=1}^{n} \varphi_i u_i = \boldsymbol{\varphi U} \tag{1.29}$$

where $\boldsymbol{U} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}^\intercal$ and $\boldsymbol{W} = \begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix}^\intercal$ are coefficient vectors (the former vector collects all degrees of freedom in the system) and

$\boldsymbol{\varphi} = \begin{bmatrix}\varphi_1 & \varphi_2 & \dots & \varphi_n\end{bmatrix}$ is a vector that collects all shape functions, whose ith component given by

$$\varphi_i = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{for } x \in [x_{i-1}, x_i], \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & \text{for } x \in [x_i, x_{i+1}], \\ 0 & \text{otherwise.} \end{cases} \tag{1.30}$$

The shape function φ_i is a piecewise linear polynomial that has a maximum value of one at node i, it ramps down to zero at nodes $i-1$ and $i+1$, and it is equally zero elsewhere. These functions satisfy the so-called Kronecker-δ property since $\varphi_i(x_j) = \delta_{ij}$. Because of this property, the degree of freedom u_i physically represents the displacement of the bar at x_i, *i.e.*, $u_i = u(x_i)$. Notice also that $\sum_{i=1}^{n} \varphi_i = 1, \forall x \in \overline{\Omega}$ and therefore the chosen basis forms *a partition of unity*. This property will be widely exploited in this book when describing enriched formulations.

By inserting (1.29) into (1.23), we see that

$$B\Big(\sum_{i=1}^{n} \varphi_i u_i, \sum_{j=1}^{n} \varphi_j u_j\Big) = L\Big(\sum_{i=1}^{n} \varphi_i w_i\Big) \quad \forall\, w_i \in \mathbb{R},\ i = 1, \dots, n \tag{1.31}$$

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} B(\varphi_i, \varphi_j) u_j = \sum_{i=1}^{n} w_i L(\varphi_i) \quad \forall\, w_i \in \mathbb{R},\ i = 1, \dots, n \tag{1.32}$$

$$\sum_{i=1}^{n} w_i \left[\sum_{j=1}^{n} B(\varphi_i, \varphi_j) u_j - L(\varphi_i)\right] = 0 \quad \forall\, w_i \in \mathbb{R},\ i = 1, \dots, n \tag{1.33}$$

which can be written in matrix form as

$$\boldsymbol{W}^{\intercal}\left[\boldsymbol{K}\boldsymbol{U} - \boldsymbol{F}\right] = 0, \quad \forall\, \boldsymbol{W} \in \mathbb{R}^n. \tag{1.34}$$

where $K_{ij} = B(\varphi_i, \varphi_j)$ and $F_i = L(\varphi_i)$. And because this equation needs to be satisfied for all $\boldsymbol{W} \in \mathbb{R}^n$, the expression in brackets must be equal to the zero vector, and thus our discrete system of equations is

$$\boldsymbol{K}\boldsymbol{U} = \boldsymbol{F} \tag{1.35}$$

We now consider an alternative approach to obtaining the discrete equations. Now consider an element of the discretization $e_i = [x_i, x_j]$, where equilibrium can also be described by Eq. (1.23):

$$\int_{x_i}^{x_j} \left(EA \frac{\mathrm{d}w^h}{\mathrm{d}x} \frac{\mathrm{d}u^h}{\mathrm{d}x} + k w^h u^h\right) \mathrm{d}x = \int_{x_i}^{x_j} w^h b \ \mathrm{d}x + w^h(x_j) N(x_j) - w^h(x_i) N(x_i), \quad \forall\, w^h \in \mathcal{E}^h, \tag{1.36}$$

where $N(x_i) \equiv N_i$ and $N(x_j) \equiv N_j$ represent the forces at both ends of the element. In this element, the only nonzero shape functions are φ_i, φ_j, and thus the displacement field can thus be written as

$$u^h = \varphi_i u_i + \varphi_j u_j = \boldsymbol{\varphi}\boldsymbol{u}, \tag{1.37}$$

where $\boldsymbol{\varphi} = \begin{bmatrix}\varphi_i & \varphi_j\end{bmatrix}$ is the element shape function vector, and $\boldsymbol{u} = \begin{bmatrix}u_i & u_j\end{bmatrix}^\intercal$ is the element degree of freedom vector.

Within the element this function can be written as in Eq. (1.37) as $w^h = \boldsymbol{\varphi}\boldsymbol{w} = \boldsymbol{w}^\intercal\boldsymbol{\varphi}^\intercal$, where $\boldsymbol{w} = \begin{bmatrix}w_i & w_j\end{bmatrix}^\intercal$ is now an arbitrary vector, *i.e.*, the statement $\forall\, w^h$ in Eq. (1.23) implies $\forall\, \boldsymbol{w}$ since $\boldsymbol{\varphi}$ is known. Then Eq. (1.36) is written as

$$\boldsymbol{w}^\intercal\left[\int_{x_i}^{x_j}(EA\boldsymbol{B}^\intercal\boldsymbol{B} + k\boldsymbol{\varphi}^\intercal\boldsymbol{\varphi})\ \mathrm{d}x\ \boldsymbol{u}\right] = \boldsymbol{w}^\intercal\left[\int_{x_i}^{x_j}\boldsymbol{\varphi}^\intercal b\ \mathrm{d}x + \begin{bmatrix}-N_i\\ N_j\end{bmatrix}\right], \quad \forall\, \boldsymbol{w}, \tag{1.38}$$

where $\boldsymbol{B} = \frac{\mathrm{d}\boldsymbol{\varphi}}{\mathrm{d}x}$ denotes the strain-displacement matrix and coefficient vectors have been taken out of the integrals. Since Eq. (1.38) must be valid for any $\boldsymbol{w}$, then

$$\underbrace{\int_{x_i}^{x_j}(EA\boldsymbol{B}^\intercal\boldsymbol{B} + k\boldsymbol{\varphi}^\intercal\boldsymbol{\varphi})\ \mathrm{d}x}_{\mathbf{k}_e}\ \boldsymbol{u} = \underbrace{\int_{x_i}^{x_j}\boldsymbol{\varphi}^\intercal b\ \mathrm{d}x + \begin{bmatrix}-N_i\\ N_j\end{bmatrix}}_{\mathbf{f}_e}, \tag{1.39}$$

where $\mathbf{k}_e$ denotes the local element stiffness and $\mathbf{f}_e$ the local element force vector. Thus the system

$$\mathbf{k}_e\mathbf{U}_e = \mathbf{f}_e \tag{1.40}$$

represents the discrete static equilibrium for the eth element. Clearly, the solution over the entire bar can only be obtained after considering the contributions of all finite elements in the discretization. This process, which is called *assembly*, results in the global stiffness matrix and global force vector. To wit,

$$\mathbf{K} = \mathop{\mathbb{A}}_{e}\mathbf{k}_e, \qquad \mathbf{F} = \mathop{\mathbb{A}}_{e}\mathbf{f}_e, \tag{1.41}$$

where $\mathbb{A}$ is the assembly operator. Note that the assembly process will produce the cancellation of many of the forces in Eq. (1.39). In other words, for a given node the force N_j from the preceding element cancels N_i from the subsequent one, leaving only those that represent the BCs in the final discrete equation. This means that at the actual boundaries of the bar, N_i and N_j in the first and last elements, respectively, need to be changed according to the actual boundary condition (see the discussion following Eq. (1.13)). Notice that the finite element discrete formulation allows us to also consider concentrated nodal forces at other nodal locations. Finally, if non-homogeneous Dirichlet boundary conditions are present, then w^h does not include the corresponding shape function at the prescribed node, *i.e.*, the weight function is zero at prescribed node locations as

to eliminate the reaction force. But since the displacement is known at prescribed nodes, the unknown part of the solution can be represented in the same way as the weight function and thus this situation poses no further difficulties.

The resulting discrete equation that describes the static equilibrium for the entire bar is

$$\mathbf{KU} = \mathbf{F}, \tag{1.42}$$

where $\mathbf{U}$ is a vector that collects all DOFs in the bar (global displacement vector). This equation can now be solved by a direct or an iterative solver. Since the trial and weight functions are of the same form (Bubnov-Galerkin), the resulting stiffness matrix is symmetric positive semi-definite. The only zero eigenvalue physically corresponds to the rigid body translation of the bar, a situation that produces no deformation–the strain energy is zero. If there are Dirichlet boundary conditions then the system is positive definite and there exists a unique solution to Eq. (1.42).

1.1.5 The iso-parametric mapping

Let us consider a single finite element of the bar $e = [x_i, x_j]$. Consider a bijective map Q, which maps a reference or master element $\hat{e} = [-1, +1]$ into e through the coordinate ξ, *i.e.*, $Q(\xi) : \hat{e} \rightarrow e$. This mapping is schematically shown in Figure 1.8a.

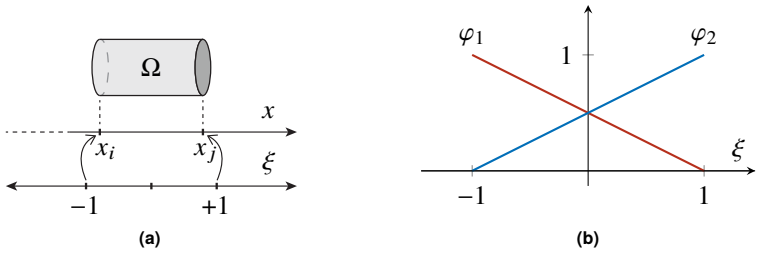

FIGURE 1.8 (a) Mapping from ξ to x coordinates for a finite element; (b) Shape functions in the master element.

From all possible mappings, consider the linear mapping given by

$$x(\xi) = \underbrace{\frac{1-\xi}{2}}_{\varphi_i} x_i + \underbrace{\frac{1+\xi}{2}}_{\varphi_j} x_j = \sum_i \varphi_i(\xi)\, x_i. \tag{1.43}$$

This choice is by no means arbitrary, and it is called an *iso-parametric mapping* because the trial solution is interpolated in the same way (see Eq. (1.37)). We could also opt for a *sub-* or *super-parametric* mapping, where the geometry interpolation is of lower or higher order, respectively, than that of the trial solution.

In the iso-parametric case, the inverse mapping $Q^{-1}(x) : e \to \hat{e}$ is given by

$$\xi(x) = \frac{2x - x_i - x_j}{x_j - x_i}. \tag{1.44}$$

The mapping is mostly used for integration, since

$$\int_{x_i}^{x_j} f(x)\, \mathrm{d}x = \int_{-1}^{+1} f\left(\xi^{-1}\right) j\, \mathrm{d}\xi = \sum_{i=1}^{n_{\mathrm{GP}}} f\left(\xi_i^{-1}\right) w_i j, \tag{1.45}$$

where $j \equiv \mathrm{d}x/\mathrm{d}\xi$ is called the Jacobian of the transformation, and the last term evaluates numerically the integral by sampling the function f at n_{GP} Gauss quadrature points with coordinate ξ_i and corresponding weight w_i. For this mapping, j represents the ratio relation between the lengths of physical and master elements, *i.e.*, $j = l_e/2$. The benefits of this transformation, which are not obvious for the1-D case, will become evident in higher dimensions.

Example 5 (*h*–**FEM solution to the bi-material interface problem**)**.** For the bi-material interface problem in Example 1, the exact solution was found to be the linear field given by Eq. (1.3). Given the bilinear and linear forms for this problem, given respectively by

$$B(u, v) = \int_0^l k(x) \frac{\mathrm{d}v}{\mathrm{d}x} \frac{\mathrm{d}u}{\mathrm{d}x}\, \mathrm{d}x \quad \text{and} \quad L(v) = v(l)P, \tag{1.46}$$

where $k(x) = E(x)A(x)$, obtain the finite element solution.

Solution: Finite element approximations studied so far are piece-wise linear. In other words, they are C^0–continuous at element boundaries. Therefore, as long as a node is placed at x_Γ, any finite element solution should capture the jump in the gradient for this problem. Moreover, because the exact solution is linear, a linear finite element approximation should recover (1.46) *exactly*. We therefore discretize the bar by only two elements, placing the middle node exactly at the location of the material interface (see Figure 1.9).

Taking our trial and weight functions as

$$u^h(\xi) = \sum_{i \in \iota_h} \varphi_i(\xi) U_i, \quad \text{and} \quad v^h(\xi) = \sum_{i \in \iota_h} \varphi_i(\xi) W_i, \tag{1.47}$$

respectively, where $\varphi_i(\xi)$ are the linear shape functions introduced in subsection 1.1.5, the discrete forms of (1.46) are given by

$$B\left(u^h, v^h\right) = \boldsymbol{v}^{\mathsf{T}} \underbrace{\int_0^l \boldsymbol{B}^{\mathsf{T}} E A \boldsymbol{B}\, \mathrm{d}x}_{\boldsymbol{K}} \boldsymbol{u} \quad \text{and} \quad L\left(v^h\right) = \boldsymbol{v}^{\mathsf{T}} \underbrace{\boldsymbol{\varphi}^{\mathsf{T}}(l) P}_{\boldsymbol{F}}, \tag{1.48}$$

which leads to the system equation

$$\boldsymbol{v}^{\mathsf{T}}[\boldsymbol{K}\boldsymbol{U} - \boldsymbol{F}] = 0 \quad \forall \boldsymbol{v} \quad \Rightarrow \quad \boldsymbol{K}\boldsymbol{U} = \boldsymbol{F}.$$

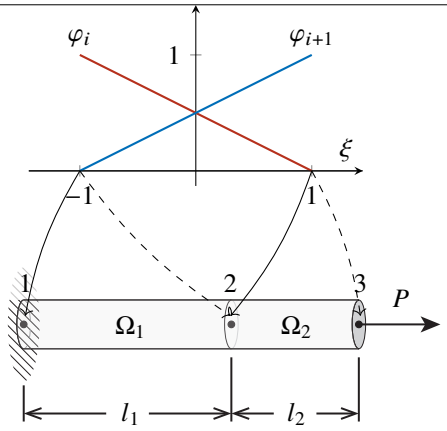

FIGURE 1.9 FEM discretization for the bi-material interface problem showing the iso-parametric mapping.

With this configuration, the load vector is simply $\boldsymbol{F} = \boldsymbol{\varphi}(l)\,P = \begin{bmatrix}0 & 0 & P\end{bmatrix}^\intercal$. The stiffness matrix is obtained by assembling the contribution of each element

$$\boldsymbol{K} = \underbrace{\int_0^{x_\Gamma} k_1 \boldsymbol{B}^\intercal \boldsymbol{B}\ \mathrm{d}x\,[x]}_{\boldsymbol{k}_1} + \underbrace{\int_{x_\Gamma}^{l} k_2 \boldsymbol{B}^\intercal \boldsymbol{B}\ \mathrm{d}x\,[x]}_{\boldsymbol{k}_2}.$$

Using the iso-parametric mapping, and recalling that Young moduli and cross sectional areas are constant at either side of the interface, the stiffness matrix of either element is obtained as

$$\boldsymbol{k}_i = \int_{-1}^{1} k_i \begin{bmatrix}-\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\begin{bmatrix}-\frac{1}{2} & \frac{1}{2}\end{bmatrix}\frac{2}{l_i}\,\mathrm{d}\xi = \bar{k}_i \begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix},\ \text{with}\ \bar{k}_i = \frac{k_i}{l_i}\ \text{and}\ i = 1, 2. \tag{1.49}$$

The assembly of the local stiffness matrices is done considering the connectivity (element freedom table), which for Ω_1 and Ω_2 are, respectively, $\begin{bmatrix}1 & 2\end{bmatrix}$ and $\begin{bmatrix}2 & 3\end{bmatrix}$. The system of linear equations $\boldsymbol{KU} = \boldsymbol{F}$ is

$$\begin{matrix}\\1\\2\\3\end{matrix}\!\!\begin{matrix}\begin{matrix}1 & 2 & 3\end{matrix}\\ \begin{bmatrix}\bar{k}_1 & -\bar{k}_1 & 0\\ -\bar{k}_1 & \bar{k}_1+\bar{k}_2 & -\bar{k}_2\\ 0 & -\bar{k}_2 & \bar{k}_2\end{bmatrix}\end{matrix}\begin{bmatrix}U_1\\U_2\\U_3\end{bmatrix} = \begin{matrix}\\1\\2\\3\end{matrix}\!\!\begin{bmatrix}0\\0\\P\end{bmatrix}\quad \text{whose solution is } \boldsymbol{U} = \begin{bmatrix}0\\ P/\bar{k}_1\\ P/\bar{k}_1 + P/\bar{k}_2\end{bmatrix}, \tag{1.50}$$

and therefore the DOFs exactly correspond to the displacement field at the nodes. By replacing $\boldsymbol{U}$ in the trial solution u^h, it can be verified that (1.3) is recovered.

Example 6 (*h*–**FEM solution to constrained bar pullout problem 3).** Obtain the finite element solution for the boundary value problem given in Example 3 for $P = -1\,\mathrm{N}$, $\alpha = 1\,\mathrm{m}$, and 4, 8, 16, and 32 elements of uniform size. Use a domain $x \in [0, 10\,\mathrm{m}]$ fix the displacement at $x = 10\,\mathrm{m}$ obtained by applying Eq. (1.8). In addition, create a mesh convergence plot of the relative error in the

energy norm, defined as

$$\epsilon = \frac{\left\|u^h - u\right\|_{\mathcal{E}(\Omega)}}{\|u\|_{\mathcal{E}(\Omega)}} = \sqrt{\frac{U^h - U}{U}}, \tag{1.51}$$

where U^h is the energy obtained by the finite element analysis and U is the exact energy obtained by applying Eq. (1.9).

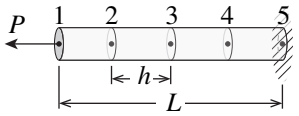

FIGURE 1.10 Discretization for the bar pullout problem.

Solution: To solve this problem, the bar is discretized using finite elements, as shown in Figure 1.10 for a discretization into 4 elements and 5 nodes. Noting that for this problem $\alpha^2 = k/EA$, the 2×2 stiffness matrix for a given finite element e can be written as

$$\mathbf{k}_e = \int_e \left(\boldsymbol{b}^\intercal \boldsymbol{b} + \alpha^2 \boldsymbol{\varphi}^\intercal \boldsymbol{\varphi}\right) \mathrm{d}x = \sum_{i=1}^{2} \left(\boldsymbol{b}^\intercal \boldsymbol{b} + \alpha^2 \boldsymbol{\varphi}^\intercal \boldsymbol{\varphi}\right) w_i j, \tag{1.52}$$

where the continuous integral is transformed into a 2-Gauss rule numeric quadrature given by (1.45). After assembly, the 5×5 system $\mathbf{KU} = \mathbf{F}$ is obtained. The rigid body translation is eliminated by prescribing the exact value of the displacement at node 5, *i.e.*, $U_5 = -0.000\,045\,399\,9$ m. Also, the unit load is also prescribed in the right hand side load as $F_1 = -1$ N.

The problem is solved numerically with 4, 8, 16, and 32 finite elements of uniform size. The displacement field as a function of position is given in Figure 1.11a. As the figure shows, the more elements are used, the closer the displacement field is to the exact solution. In order to

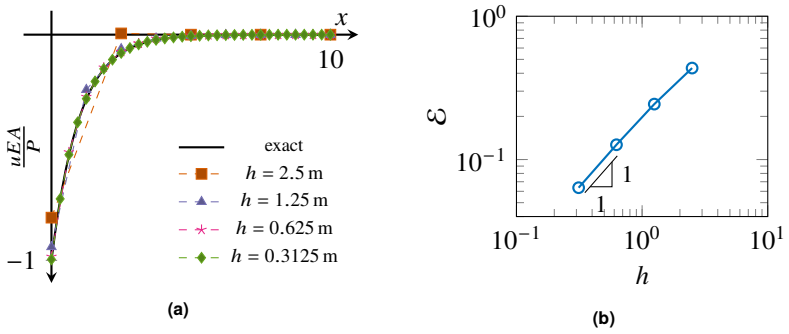

FIGURE 1.11 (a) uEA/P for different mesh sizes; (b) error as a function of mesh size h.

quantify the error with respect to the exact solution, we need a reference value for the *exact energy* in the system. The result of the energy given by Eq. (1.9) was obtained for an infinite bar length. Thus, by chancing the upper integration limit to $x = 10$ m, the reference energy is computed as $U = 0.499\,999\,998\,969\,423\,2$ J. The energy obtained from the finite element solution is computed as $U^h = \frac{1}{2}\mathbf{U}^\intercal\mathbf{K}\mathbf{U}$, where $\mathbf{K}$ is the global stiffness matrix and $\mathbf{U}$ the resulting displacement vector. The error as a function of mesh size is illustrated in Figure 1.11b, with a constant convergence rate $\beta_h = 1$ (rounded from $\beta_h = 0.9956110380498996$). Thus, for this problem h–FEM gives an *algebraic* rate of convergence, *i.e.*, the convergence rate remains constant as the mesh size h decreases.

1.2 THE ELASTOSTATICS PROBLEM IN HIGHER DIMENSIONS

The 1-D bar discussed in the previous section gives considerable insight into how the finite element formulation is obtained. The process in 3-D is remarkably similar as we will see below, though we give a slightly more formal presentation. As before, we start with the strong form of the boundary value problem.

1.2.1 Strong form

Consider the d-dimensional Euclidean vector space $\mathbb{R}^d$, $d = 2, 3$, spanned by a chosen orthonormal basis $\{\boldsymbol{e}_i\}$. A point in $\mathbb{R}^d$ is represented by its coordinates in such basis as $\boldsymbol{x} = x_i \boldsymbol{e}_i$ This space is equipped with the inner product structure that induces the norm $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$, $\forall \boldsymbol{v} \in \mathbb{R}^d$. Now assume a body $\Omega \subset \mathbb{R}^d$ (*cf.* Figure 1.12 for 3-D) which represents a solid composed of an infinite number of particles. The closure of the body is $\overline{\Omega}$ and the boundary $\overline{\Omega} \setminus \Omega \equiv \partial\Omega \equiv \Gamma$ has outward unit normal vector $\boldsymbol{n}$. The boundary is divided into regions $\Gamma \equiv \overline{\Gamma_\mathrm{u}} \cup \overline{\Gamma_\mathrm{t}} \cup \overline{\Gamma_\mathrm{h}}$, where Dirichlet, Neumann, and Robin boundary conditions are prescribed, respectively. These regions are also disjoint, *i.e.*, $\Gamma_\mathrm{u} \cap \Gamma_\mathrm{t} \cap \Gamma_\mathrm{h} = \emptyset$.

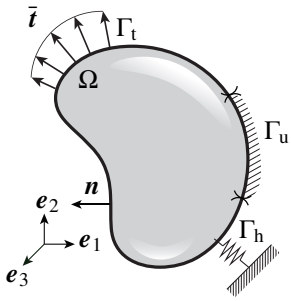

FIGURE 1.12 Schematic of a 3-D domain Ω.

The equilibrium equation in d-dimensional space is described by

$$\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} + \boldsymbol{b} = \boldsymbol{0} \qquad \forall \boldsymbol{x} \in \Omega, \tag{1.53}$$

where $\boldsymbol{\nabla}\cdot$ denotes the divergence operator, $\boldsymbol{\sigma} : \overline{\Omega} \to \mathbb{R}^d \times \mathbb{R}^d$ is the stress tensor, and $\boldsymbol{b} : \Omega \to \mathbb{R}^d$ is the body force vector.

As in the 1-D bar case thoroughly explained in Section 1.1, the strong form of the elastostatics boundary value problem in d–dimensional space is obtained after including boundary conditions

$$\boldsymbol{u} = \bar{\boldsymbol{u}} \qquad \text{on } \Gamma_\mathrm{u} \tag{1.54}$$

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \bar{\boldsymbol{t}} \qquad \text{on } \Gamma_\mathrm{t} \tag{1.55}$$

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \boldsymbol{h} \cdot \left(\bar{\boldsymbol{u}} - \boldsymbol{u}|_{\Gamma_\mathrm{h}}\right) \qquad \text{on } \Gamma_\mathrm{h} \tag{1.56}$$

where $\boldsymbol{h} : \Gamma_{\mathrm{h}} \to \mathbb{R}^d \times \mathbb{R}^d$ is a tensor that defines the elastic support. Contrary to what we have seen in the bar of Section 1.1, for this problem we can prescribe the three types of boundary conditions simultaneously. Clearly, the actual boundary value problem can include any combination of boundary conditions (1.54)-(1.56).

So far no constitutive relationship has been adopted to describe the relationship between stress $\boldsymbol{\sigma}$ and strain $\boldsymbol{\varepsilon}$. Here we adopt a linear relationship (Hooke's law), *i.e.*, $\boldsymbol{\sigma} = \boldsymbol{C}\boldsymbol{\varepsilon}$, where $\boldsymbol{C}$ is a fourth-order tensor. We also assume small deformation so that the strain tensor can be written as $\boldsymbol{\varepsilon} = \frac{1}{2}\left(\boldsymbol{\nabla u} + \boldsymbol{\nabla u}^{\intercal}\right)$. Notice that $\boldsymbol{\sigma}$ is a function of the displacement field through the constitutive relation.

1.2.2 Weak form

As before we start by setting up a weighted residual

$$\int_{\Omega} \boldsymbol{w} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} + \boldsymbol{b}) \, \mathrm{d}\Omega = 0, \qquad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{V}}, \tag{1.57}$$

where $\boldsymbol{\mathcal{V}}$ is a conveniently chosen function space. In d–dimensional space we turn to vector-valued function spaces

$$\boldsymbol{\mathcal{V}}(\Omega) \equiv \left[\mathcal{H}^1(\Omega)\right]^d = \left\{\boldsymbol{v} : \boldsymbol{v}(\boldsymbol{x}) \in \mathbb{R}^d \; \forall \boldsymbol{x} \in \Omega; \; v_i \in \mathcal{H}^1(\Omega), i = 1, \ldots, d\right\}, \tag{1.58}$$

$$\boldsymbol{\mathcal{V}}_0(\Omega) \equiv \left[\mathcal{H}^1_0(\Omega)\right]^d = \left\{\boldsymbol{v} \in \boldsymbol{\mathcal{V}}(\Omega) : \; \boldsymbol{v}|_{\Gamma_{\mathrm{u}}} = \boldsymbol{0}\right\}, . \tag{1.59}$$

In (1.58) and (1.59) $\mathcal{H}^1(\Omega)$ is the first-order Sobolev function space on Ω, defined earlier in footnote 3. Eq. (1.58) states that every component of the vector-valued function $\boldsymbol{v} \in \boldsymbol{\mathcal{V}}$ belongs to $\mathcal{H}^1(\Omega)$. The subscript 0 in Eq. (1.59) emphasizes the fact that vector-valued functions from the space vanish on the region with prescribed essential boundary conditions.

As with the 1-D bar studied earlier, we can lower the smoothness requirement on $\boldsymbol{u}$ by balancing derivatives. To that effect we use the identity $\boldsymbol{w} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{\sigma}) = \boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}^{\intercal} \cdot \boldsymbol{w}) - \boldsymbol{\nabla w} : \boldsymbol{\sigma}$, where : denotes the contraction between two tensors, *i.e.*, $\mathbf{S} : \mathbf{T} = S_{ij} T_{ij}$ (summation convention implied). Eq. (1.57) can be rewritten as[5]

$$\int_{\Omega} (\boldsymbol{\nabla w} : \boldsymbol{\sigma}) \, \mathrm{d}\Omega = \int_{\Omega} \boldsymbol{w} \cdot \boldsymbol{b} \, \mathrm{d}\Omega + \int_{\Omega} \boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}^{\intercal} \cdot \boldsymbol{w}) \, \mathrm{d}\Omega, \qquad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{V}}. \tag{1.60}$$

By using the divergence theorem, the last term can be converted to an integral

5. Note that we have chosen to keep $\boldsymbol{\sigma}^{\intercal}$ though the stress tensor is symmetric, *i.e.*, $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\intercal}$, which is the result of the balance of angular momentum.

over the boundary:

$$\begin{aligned}\int_\Omega \boldsymbol{\nabla}\cdot(\boldsymbol{\sigma}^\intercal\cdot\boldsymbol{w})\ \mathrm{d}\Omega &= \int_\Gamma \boldsymbol{n}\cdot(\boldsymbol{\sigma}^\intercal\cdot\boldsymbol{w})\ \mathrm{d}\Gamma\,,\\ &= \int_\Gamma \boldsymbol{w}\cdot(\boldsymbol{\sigma}\cdot\boldsymbol{n})\ \mathrm{d}\Gamma\,,\\ &= \int_{\Gamma_\mathrm{t}} \boldsymbol{w}\cdot\bar{\boldsymbol{t}}\ \mathrm{d}\Gamma + \int_{\Gamma_\mathrm{u}} \boldsymbol{w}\cdot\boldsymbol{r}\ \mathrm{d}\Gamma + \int_{\Gamma_\mathrm{h}} \boldsymbol{w}\cdot\boldsymbol{h}\cdot(\bar{\boldsymbol{u}}-\boldsymbol{u})\ \mathrm{d}\Gamma\,.\end{aligned} \tag{1.61}$$

In case the elastostatics boundary value problem has non-homogeneous boundary conditions, we define the linear variety $\mathcal{V}_\star \equiv \tilde{\boldsymbol{u}} + \mathcal{V}_0$ as a translation of $\mathcal{V}_0$ by a function $\tilde{\boldsymbol{u}} \in \mathcal{V}$ that satisfies the Dirichlet boundary condition, *i.e.*, $\tilde{\boldsymbol{u}}|_{\Gamma_\mathrm{u}} = \bar{\boldsymbol{u}}$. Given $\boldsymbol{u} = \boldsymbol{v} + \tilde{\boldsymbol{u}}$, find $\boldsymbol{v} \in \mathcal{V}_0$ such that

$$B(\boldsymbol{v},\boldsymbol{w}) = L(\boldsymbol{w}) - B(\tilde{\boldsymbol{u}},\boldsymbol{w}) \qquad \forall\,\boldsymbol{w} \in \mathcal{V}_0(\Omega)\,. \tag{1.62}$$

The bilinear and linear forms in Eq. (1.62) are given by

$$B(\boldsymbol{v},\boldsymbol{w}) = \int_\Omega (\boldsymbol{\nabla}\boldsymbol{w} : \boldsymbol{\sigma})\ \mathrm{d}\Omega + \int_{\Gamma_\mathrm{h}} \boldsymbol{w}\cdot\boldsymbol{h}\cdot\boldsymbol{v}\ \mathrm{d}\Gamma\,, \tag{1.63}$$

$$L(\boldsymbol{w}) = \int_\Omega \boldsymbol{w}\cdot\boldsymbol{b}\ \mathrm{d}\Omega + \int_{\Gamma_\mathrm{t}} \boldsymbol{w}\cdot\bar{\boldsymbol{t}}\ \mathrm{d}\Gamma + \int_{\Gamma_\mathrm{h}} \boldsymbol{w}\cdot\boldsymbol{h}\cdot\bar{\boldsymbol{u}}\ \mathrm{d}\Gamma\,. \tag{1.64}$$

Notice that as in the bar formulation, the boundary conditions arise naturally after balancing the derivatives between the trial and the weight functions. In addition, by conveniently choosing $\boldsymbol{w} \in \mathcal{V}_0$ the integral over the part of the domain with prescribed Dirichlet BC – containing the reaction forces – drops out.

1.2.3 Discrete formulation

The Galerkin formulation is obtained by choosing finite dimensional vector-valued function spaces $\mathcal{V}^h \subset \mathcal{V}$ and $\mathcal{V}_0^h \subset \mathcal{V}_0$ so that the weak formulation reads: Find $\boldsymbol{u}^h \in \mathcal{V}_\star^h$ such that

$$B\left(\boldsymbol{u}^h,\boldsymbol{w}^h\right) = L\left(\boldsymbol{w}^h\right) \qquad \forall\,\boldsymbol{w}^h \in \mathcal{V}_0^h(\Omega)\,. \tag{1.65}$$

Selecting trial and weight functions from the same space $\mathcal{V}_0^h$ gives us the Bubnov-Galerkin approach, which results in a symmetric stiffness matrix.

We now focus on the finite element discretized equations. To that purpose the domain Ω is discretized in finite elements so that $\Omega^h = \mathrm{int}\,(\cup_i \overline{e}_i)$ and $e_i \cap e_j = \emptyset, \forall i \neq j$. Contrary to the 1-D case and unless the geometry is relatively simple, in higher dimensions the discretized domain $\Omega^h \approx \Omega$ due to discretization error. Consequently, the boundary of the domain $\Gamma^h \approx \Gamma$. After

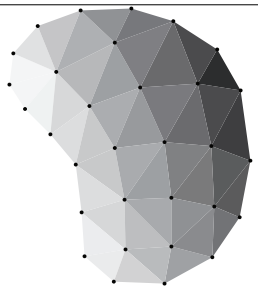

FIGURE 1.13 Domain discretization Ω^h.

choosing the finite-dimensional function spaces, typically as the vector-valued function spaces of low-order polynomials, the trial and weight functions can be written as

$$\boldsymbol{u}^h = \boldsymbol{\Phi}\mathbf{U}_e, \qquad \boldsymbol{w}^h = \boldsymbol{\Phi}\mathbf{W}_e. \tag{1.66}$$

Contrary to the 1-D case studied earlier, in the latter equation $\boldsymbol{\Phi}$ is understood to have the correct dim ensions for dealing with d–space. In other words, $\boldsymbol{\Phi}$ can be obtained as $\boldsymbol{\Phi} \otimes \mathbf{I}$, *i.e.*, the Kronecker product between the 1-D vector containing shape functions and the $d \times d$ identity matrix.

For a particular element e the static equilibrium is written as

$$\int_e \left(\boldsymbol{\nabla} \boldsymbol{w}^h : \boldsymbol{\sigma}\right) \, \mathrm{d}\Omega = \int_e \boldsymbol{w}^h \cdot \boldsymbol{b} \, \mathrm{d}\Omega + \int_{\Gamma_e} \boldsymbol{w}^h \cdot \underbrace{(\boldsymbol{\sigma} \cdot \boldsymbol{n}_e)}_{\boldsymbol{t}_e} \, \mathrm{d}\Gamma, \qquad \forall \, \boldsymbol{w}^h \in \boldsymbol{\mathcal{V}}_0^h, \tag{1.67}$$

where $\boldsymbol{n}_e$ is the unit normal to the element and thus $\boldsymbol{t}_e$ the traction field on the element sides. Since the stress tensor is symmetric, $\boldsymbol{\nabla w} \cdot \boldsymbol{\sigma} = \mathrm{sym}\,(\boldsymbol{\nabla w}) \cdot \boldsymbol{\sigma}$, where $\mathrm{sym}\,(\boldsymbol{\nabla w}) \equiv \frac{1}{2}\,(\boldsymbol{\nabla w} + \boldsymbol{\nabla w}^{\intercal})$. Thus, by considering $\boldsymbol{\varepsilon} = \mathbf{B}\boldsymbol{u}$, where $\mathbf{B}$ is the strain-displacement matrix, Eq. (1.67) can be rewritten as

$$\mathbf{W_e}^{\intercal} \left[\underbrace{\int_e \mathbf{B}^{\intercal} \boldsymbol{C} \mathbf{B} \, \mathrm{d}\Omega}_{\mathbf{k}_e} \, \mathbf{U}_e \right] = \mathbf{W}_e^{\intercal} \left[\underbrace{\int_e \boldsymbol{\Phi}^{\intercal} \boldsymbol{b} \, \mathrm{d}\Omega + \int_{\Gamma_e} \boldsymbol{\Phi}^{\intercal} \boldsymbol{t}_e \, \mathrm{d}\Gamma}_{\mathbf{f}_e} \right], \qquad \forall \, \mathbf{W}_e. \tag{1.68}$$

And since this equation has to hold for any $\mathbf{W}_e$, local equilibrium at the element level is $\mathbf{k}_e\mathbf{U}_e = \mathbf{f}_e$. Following the assembly procedure discussed at the end of Section 1.1.4, the final discrete equation that describes the static equilibrium of the body is

$$\mathbf{KU} = \mathbf{F}, \tag{1.69}$$

where $\mathbf{K}$, and $\mathbf{F}$ are obtained by Eq. (1.41).

1.3 HEAT CONDUCTION

The steady-state heat conduction problem in 1-D is governed by Poisson's equation

$$k\frac{\mathrm{d}u}{\mathrm{d}x}[2] + f = 0 \qquad \forall\, x \in \Omega, \tag{1.70}$$

where $k : \overline{\Omega} \to \mathbb{R}$ denotes the thermal conductivity, $f : \Omega \to \mathbb{R}$ the heat source, $u : \overline{\Omega} \to \mathbb{R}$ is the temperature[6] In the case of no source term, Eq. 1.70 reduces to Laplace's equation, *i.e.*, $\nabla^2 u = 0$.

The boundary value problem is fully determined after prescribing two boundary conditions out of the following three conditions:

- Given the prescribed temperature $\bar{u} : \Gamma_\mathrm{u} \to \mathbb{R}$, the *essential* or *Dirichlet* BC is
$$u|_{\Gamma_\mathrm{u}} = \bar{u}$$
- With prescribed heat flux $\bar{q} : \Gamma_\mathrm{q} \to \mathbb{R}$, the *Natural* or *Neumann* is
$$-k\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_{\Gamma_\mathrm{q}} n = \bar{q},$$
where n is again either -1 or 1 for the left and right end, respectively. The heat flux is therefore positive when heat leaves the bar and negative otherwise, a sign convention analogous to that used earlier.
- Finally, the *mixed*, *Robin* or in this case *convective* boundary condition is
$$-k\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_{\Gamma_\mathrm{h}} n = h\,(u_\infty - u)$$
where for the case of the bar h is the heat transfer coefficient and u_∞ is the ambient temperature.

Example 7 (**Weak formulation in 1-D**). Use the method of weighted residuals to obtain the weak formulation for the 1-D bar: Find the temperature field $u \in \mathcal{V}_\star$ such that

$$B\,(u, w) = L\,(w) \qquad \forall\, w \in \mathcal{V}_0. \tag{1.71}$$

Notice that a non-zero temperature boundary condition is needed for the use of the linear variety $\mathcal{V}_\star$. Consider a convective BC for the other end of the bar and write the bilinear and linear forms above.

In $\mathbb{R}^d$ Euclidean space, the strong form of the problem is presented similarly. Consider a closed domain $\Omega \subset \mathbb{R}^d$, with closure $\overline{\Omega}$ and boundary $\partial\Omega \equiv \Gamma$,

6. We have opted to use the same symbol used earlier for displacements. Thus, u must be understood as the primal variable of the boundary value problem.

the latter composed of disjoint regions Γ_u, Γ_q, and Γ_h. Given the thermal conductivity tensor $\boldsymbol{\kappa} : \overline{\Omega} \to \mathbb{R}^d \times \mathbb{R}^d$, the heat source $f : \Omega \to \mathbb{R}$, prescribed temperature $\bar{u} : \Gamma_u \to \mathbb{R}$, prescribed heat flux $\bar{q} : \Gamma_q \to \mathbb{R}$, the temperature u_∞, the heat transfer coefficient $h : \Gamma_h \to \mathbb{R}$, find the temperature field $u \in C^2$ such that

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\kappa} \boldsymbol{\nabla} u) + f = 0 \qquad \text{in } \Omega, \tag{1.72}$$

with boundary conditions

$$u = \bar{u} \qquad \text{on } \Gamma_u, \tag{1.73}$$

$$\boldsymbol{\kappa} \boldsymbol{\nabla} u \cdot \boldsymbol{n} = \bar{q} \qquad \text{on } \Gamma_q, \tag{1.74}$$

$$\boldsymbol{\kappa} \boldsymbol{\nabla} u \cdot \boldsymbol{n} = h\,(u_\infty - u) \qquad \text{on } \Gamma_h. \tag{1.75}$$

As before, the weak form of (1.72) is: Given $u = v + \tilde{u}$, $u \in \mathcal{V}_\star$, find $v \in \mathcal{V}_0$ such that

$$B\,(v, w) = L\,(w) - B\,(\tilde{u}, w) \qquad \forall\, w \in \mathcal{V}_0, \tag{1.76}$$

where the bilinear and linear forms are given by

$$B\,(v, w) = \int_\Omega \boldsymbol{\nabla} w \cdot (\boldsymbol{\kappa} \boldsymbol{\nabla} v)\ \mathrm{d}\Omega + \int_{\Gamma_\mathrm{h}} h w v\ \mathrm{d}\Gamma\,, \tag{1.77}$$

and

$$L\,(w) = \int_\Omega w f\ \mathrm{d}\Omega + \int_{\Gamma_\mathrm{h}} h w u_\infty\ \mathrm{d}\Gamma + \int_{\Gamma_\mathrm{q}} w \bar{q}\ \mathrm{d}\Gamma\,, \tag{1.78}$$

respectively.

The discrete formulation can then be obtained by choosing finite dimensional function spaces $\mathcal{V}^h \subset \mathcal{V}$ and $\mathcal{V}_0^h \subset \mathcal{V}_0$. Then the finite dimensional weak form reads: Find $v^h = u^h - \tilde{u}^h \in \mathcal{V}_0^h$ such that

$$B\left(v^h, w^h\right) = L\left(w^h\right) - B\left(\tilde{u}^h, w^h\right) \qquad \forall\, w^h \in \mathcal{V}_0^h\,(\Omega)\,, \tag{1.79}$$

where $\mathcal{V}_\star^h \equiv \tilde{u} + \mathcal{V}_0^h$ is a linear variety that allows us to handle non-homogeneous essential boundary conditions since $\tilde{u}|_{\Gamma_\mathrm{u}} = \bar{u}$. The Bubnov-Galerkin formulation is obtained by choosing trial and weight functions from the same $\mathcal{V}_0^h$. To obtain the finite element discrete equations, we apply the discrete formulation to elements of the domain discretization $\Omega^h \approx \Omega$. Let the trial and weight functions within the eth element be given by $u^h = \boldsymbol{\Phi} \mathbf{U}_e$ and $w^h = \boldsymbol{\Phi} \mathbf{W}_e$, respectively. Inserting these into (1.79) leads to the local conductivity matrix and local source vector:

$$\mathbf{k}_e = \int_e \mathbf{B}^\mathsf{T} \boldsymbol{\kappa} \mathbf{B}\ \mathrm{d}\Omega + \int_{\Gamma_e} h \boldsymbol{\Phi}^\mathsf{T} \boldsymbol{\Phi}\ \mathrm{d}\Gamma \tag{1.80}$$

$$\mathbf{f}_e = \int_e \boldsymbol{\Phi}^\mathsf{T} f\ \mathrm{d}\Omega + \int_{\Gamma_h} h u_\infty \boldsymbol{\Phi}^\mathsf{T}\ \mathrm{d}\Gamma + \int_{\Gamma_q} \bar{q} \boldsymbol{\Phi}^\mathsf{T}\ \mathrm{d}\Gamma\,, \tag{1.81}$$

These equations also contain the contributions of the boundary conditions, though it is understood such integrals will be non-zero only if the element has a side that belongs to the respective domain boundary. After assembling the contribution of all elements, $\mathbf{K} = \mathbb{A}_e \, \mathbf{k}_e$ and $\mathbf{F} = \mathbb{A}_e \, \mathbf{f}_e$, and the final discrete system of linear equations is $\mathbf{KU} = \mathbf{F}$.

A priori error estimates for h–FEM

Depending on the *smoothness* or *regularity* of the exact solution, problems can be classified in three categories [2]:

- **a** $\boldsymbol{u}$ is an analytic function in Ω, *i.e.*, it can be expanded in Taylor series;
- **b** $\boldsymbol{u}$ is analytic in Ω, except for a finite number of sets of zero measure, *i.e.*, points and curves in 2-D or points, curves, and surfaces in 3-D;
- **c** $\boldsymbol{u}$ is not in Category A nor in B.

It can be shown that the h–version of the finite element method exhibits *algebraic convergence*. This means that for uniform meshes–where the mesh size h remains roughly the same throughout the discretization–the error of the finite element solution is given by

$$\left\| \boldsymbol{u} - \boldsymbol{u}^h \right\|_{\mathcal{E}(\Omega)} \leq C_1 h^{\beta_h} \, \|\boldsymbol{u}\|_{\mathcal{E}(\Omega)} \,, \tag{1.82}$$

where C_1 is a constant, h the mesh size, and β_h is the rate of convergence. Table 1.1 provides the convergence rates for problems in each category. With the relative error in energy $e_{\mathcal{E}} \equiv \left\| \boldsymbol{u} - \boldsymbol{u}^h \right\|_{\mathcal{E}(\Omega)} / \|\boldsymbol{u}\|_{\mathcal{E}(\Omega)}$, and taking the natural logarithm we get

$$\ln e_{\mathcal{E}} = \ln C_1 + \beta_h \ln h$$

which is the equation of a line with slope β_h in $\ln h \times \ln e_{\mathcal{E}}$ space. Back in Example 9 we obtained the convergence rate $\beta_h = 1$ for a smooth problem discretized by a linear partition of unity. The rates for h–FEM are summarized in Table 1.1.

	Category		
	A	B	C
Convergence rate	$\beta_h = p^{\dagger}$	$\beta_h = \begin{cases} \min(p, k-1)^{\ddagger} & \text{for uniform } h \\ p & \text{for optimal mesh} \end{cases}$	$\beta_h > 0$

† $p \equiv$ polynomial order

‡ $k \equiv$ regularity of the solution. For $\boldsymbol{u} \in \mathcal{H}^k$, k is very large or $k > 1$ for categories A and B, respectively.

TABLE 1.1 Convergence rates based on problem category.

1.4 PROBLEMS

PROBLEM 1.0.— BOUNDARY VALUE PROBLEM WITH $u \in C^\infty$

Consider the differential equation

$$-\frac{\mathrm{d}u}{\mathrm{d}x}[2] = \frac{2\left(a + a^3 b(b - x + 1)\right)}{\left(a^2 b^2 + 1\right)^2}, \tag{1.83}$$

with boundary conditions (BCs) $u(0) = u(1) = 0$, where a is constant and $b = x - x_b$. The solution to this boundary value problem (BVP) is given by

$$u(x) = (1 - x)\left(\arctan(ab) + \arctan(ax_b)\right).$$

This solution has a peak centered near $x = x_b$ whose sharpness depends on the value of a. For large a the solution has a "near discontinuity" while for small a the solution is very smooth. Figure 1.14 shows the exact solution and its derivative for $a = \{0.5, 50\}$.

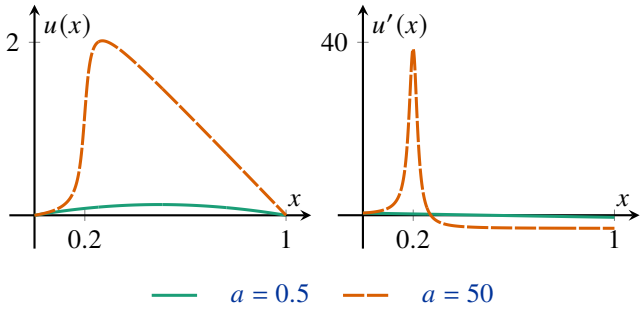

FIGURE 1.14 Exact solution and its derivative for problem (1.83) considering $a = \{0.5, 50\}$.

1. Solve problem (1.83) for $x_b = 0.2$ and $a = 0.5$. Use 2, 4, 8, 16 and 32 evenly spaced linear and quadratic *Lagrangian* elements (10 runs total). Make a log-log plot of the relative error in the energy norm in ordinates, defined as

$$e_r = \sqrt{\frac{U - U_h}{U}},$$

versus the mesh size h and another one versus the number of degrees of freedom (DOFs) (2 plots and 4 curves total). The exact strain energy for this problem is $U = 0.0408777548$. Calculate the rates of convergence in the energy norm for both element types and indicate them in the convergence plots. How do the computed values compare with the theoretical values (knowing that the solution is very smooth)?
2. Using the data points (e_r and N) from the h–version and the *a posteriori* error estimate, evaluate the value of the exact strain energy. Compare this with the actual value.
3. Solve (1.83) for $x_b = 0.2$ and $a = 50$ and repeat the above h–version convergence study for *only* the quadratic element case. This time use 5, 10, 20 and 40 elements evenly spaced. Make a log-log plot and compute the rate of convergence in the energy norm. If the plot is not linear, use the last two data points to compute the (asymptotic) rate of convergence. The exact strain energy, U, for this problem is $U = 25.138142063$.

PROBLEM 1.0.— MATERIAL DISCONTINUITY

Consider the boundary value problem

$$-\mathrm{d}x\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right) + Cu = T(x) \qquad 0 < x < L, \tag{1.84}$$

with BCs $u(0) = 0$ and $u(L) = 1$. The bar consists of two materials with elastic moduli $E_1 = 10\,000$ and $E_2 = 1000$. The length of the bar is $L = 10$ and the cross section is $A = 1$. The material interface is located at $x_\Gamma = L/2$. The bar is subjected to a distributed force per unit length given by

$$T(x) = 25x - \frac{15}{2}x^2 + \frac{1}{2}x^3. \tag{1.85}$$

The solution to this problem is given by [24]

$$u(x) = \begin{cases} \frac{1}{E_1}\left(E_2 B x + g(x)\right) & x \le x_\Gamma \\ B(x - L) + 1 + \frac{1}{E_2}\left(g(x) - g(L)\right) & x > x_\Gamma \end{cases}, \tag{1.86}$$

where the constant B and the function $g(x)$ are given by

$$B = \frac{E_1 E_2 - g(x_\Gamma)(E_2 - E_1) - g(L) E_1}{E_2\left((E_2 - E_1) x_\Gamma + L E_1\right)}, \qquad g(x) = -\frac{25}{6}x^3 + \frac{5}{8}x^4 - \frac{1}{40}x^5,$$

respectively. Figure 1.15 show the solution (1.86) and its derivative, respectively.

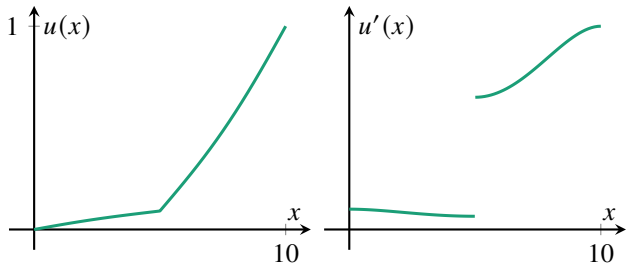

FIGURE 1.15 Exact solution and its derivative for problem (1.84) subjected to load (1.85).

1. Solve the boundary value problem using the h–version of the FEM and uniform meshes. Use *i)* a sequence of meshes with an even number of elements, and *ii)* a sequence of elements with an odd number of elements. In case *i)* the material discontinuity is at a node of the mesh, while in case *ii)* it is at the center of an element. Note that in this case, the stiffness matrix of the element containing the material interface must be integrated using integration elements with the right material properties. Use linear and quadratic elements. For each element type used, create a convergence plot of the relative error in the energy norm versus the number of DOFs. The exact energy can be computed from Eq. (1.86):

$$U(u(x)) = \int_0^{x_\Gamma} \frac{E_1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 \mathrm{d}x + \int_{x_\Gamma}^{L} \frac{E_2}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 \mathrm{d}x\,,$$

or alternatively, you may use *a posteriori* error estimation. Calculate the rate of convergence for both element types and sequence of meshes in *i)* and *ii)*.