1 The Finite Element Method: Introductory Remarks

The finite element method is a numerical technique for obtaining approximate solutions to a wide spectrum of engineering problems. Although originally developed to study the stresses in complex airframe structures, it has since been extended and applied to a broad field in continuum mechanics. Because of its diversity and flexibility as an analysis tool, this particular technique is receiving much attention in both academia and industry fields.

Although this brief comment on the finite element method answers the question posed by the section heading, it does not give us the operational definition we need to apply the method to a particular problem. Such an operational definition, along with a description of the method fundamentals, requires considerably more than just one paragraph to develop. Hence, the first segment of this book is devoted to basic concepts and fundamental theory. Before discussing more aspects of the finite element method, we should first consider some of the circumstances leading to its inception, and we should briefly contrast it with other numerical techniques.

In more and more engineering situations today, we find that it is necessary to obtain approximate numerical solutions to problems rather than exact closed-form solutions. For example, we may want to find the load capacity of a plate that has several stiffeners and odd-shaped holes, the concentration of pollutants during nonuniform atmospheric conditions, or the rate of fluid flow through a passage of arbitrary shape. Without too much effort, we can write down the governing equations and boundary conditions for these problems, but we immediately see that no simple analytical solution can be found. The difficulty in these examples lies in the fact that either the geometry or some other features of the problem are irregular. Analytical closed-form solutions to problems of this type seldom exist, yet those are the kinds of problems that engineers are called on to solve.

The resourcefulness of the analyst usually comes to the rescue, providing several alternatives to overcome the dilemma. One possibility is to make meaningful assumptions, to ignore the difficulties, and to reduce the problem to one that can be handled. Sometimes this procedure works. However, more often than not, it leads to serious inaccuracies or totally wrong answers. Now that large-scale digital computers are widely available, a more viable alternative is to retain the complexities of the problem and try to find an approximate numerical solution.

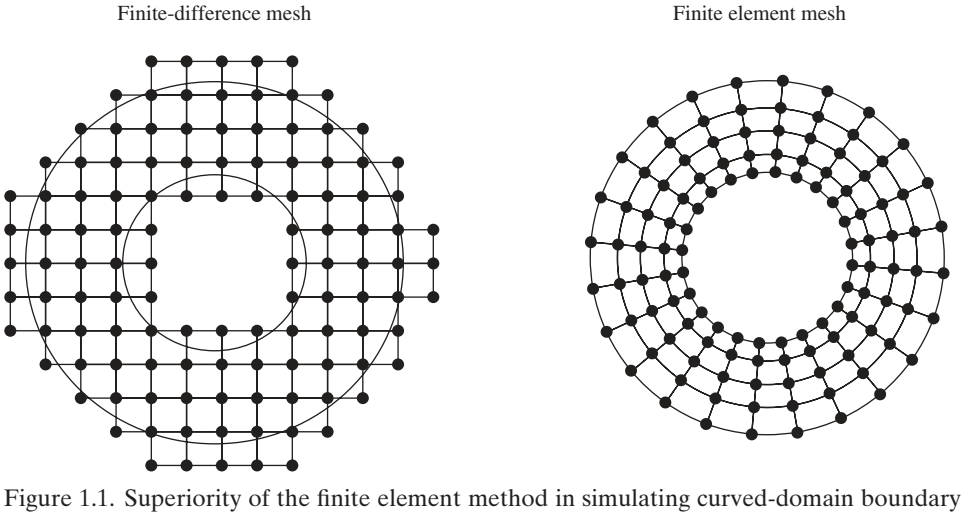

Figure 1.1. Superiority of the finite element method in simulating curved-domain boundary segments.

Several approximate numerical analysis methods have evolved over the years, with the most commonly used method being the finite-difference technique. The familiar finite-difference model of a problem gives pointwise approximation to the governing equations. The model, formed by writing difference equations for an array of grid points, is improved as more points are used. With finite-difference techniques, we can treat some fairly difficult problems, but, for example, when we encounter irregular geometries or unusual specification of boundary conditions, we find that finite-difference techniques become too hard to use.

In addition to the finite-difference method, another, more recent numerical method, known as the *finite element method*, has emerged. Unlike the finite-difference method, which envisions the solution domain as an array of grid points, the finite element method envisions the solution region as a buildup of small, nonoverlapping and interconnected subregions termed *elements*. A finite element model of a problem gives a *piecewise* approximation to the governing equations. The basic premise of the finite element method is that a solution domain can be modeled or approximated analytically by replacing it with an assemblage of discrete elements. Since these elements can be put together in a variety of ways, they can be used to represent exceedingly complex shapes.

As an example of how finite-difference and finite element models might be used to represent a complex geometric shape, consider the annular passage in Figure 1.1. Note the superiority of the finite element method in handling such problems because the elements can be shaped in such a way that matches the solution-domain curved boundary segments. Figure 1.1 is meant to simply illustrate the finite element model in contrast to a typical finite-difference mesh of the same annulus.

The Mathematical Approach: A Variational Interpretation

This approach is helpful in gaining an understanding of the finite element method, but instrumental difficulties arise when we try to apply it to complex problems. In this section, we take a broader view and interpret the finite element method as

an approximate means of solving variational problems. At this point, however, we cannot faithfully discuss the many specific techniques that are useful for particular types of problems. These specialized aspects of the finite element method will be introduced later in this and succeeding chapters.

To set the stage for the introduction of the mathematical concepts and to give them a place in the overall collection of solution techniques, we begin with a general discussion of the continuum problems of mathematical physics.

After briefly mentioning some of the more popular solution techniques for different classes of problems, we establish the necessary technology and definitions to show how variational problems and the finite element method are related. The variational basis of the finite element method dictates the criteria to be satisfied by the so-called element interpolation functions and enables us to make definitive statements about the convergence of results as we use an ever-increasing number of smaller and smaller elements.

After a discussion of the variational approach to the formulation of element equations, we consider a detailed example. The last segment of this chapter contains the problem of how to find variational principles for use in the finite element method.

Continuum Problems

Problems in engineering and science fall into two fundamentally different categories depending on which point of view we adopt. One point of view is that all matter consists of single particles that retain their identity and nature as they move through space. Their position in space at any instant is given by the coordinates in some reference frame, and these coordinates are functions of time – the only independent variable for any process. This viewpoint, known as the *Lagrangian* viewpoint, is the basis for *Newtonian* particle mechanics.

The other viewpoint, the one we will use, stems from the *continuum* rather than the molecular or particle approach to nature. In the continuum (sometimes termed *Eulerian*) viewpoint, we say that all processes are characterized by field quantities that are defined at every point in space. The independent variables in continuum problems are the coordinates of space and time. The Eulerian viewpoint allows us to focus our attention on one point in space and then observe the phenomena occurring there.

Continuum problems are concerned with fields of temperature, stress, mass concentration, displacement, electromagnetic, and acoustic potentials, to name just a few examples. These problems arise from phenomena in nature that are approximately characterized by partial differential equations and their boundary conditions.

We will briefly discuss the nature of continuum problems typically encountered and some of the possible means of solution. Then we will return to the topic of solving these problems using the finite element method. Continuum problems of mathematical physics are often referred to as *boundary-value problems* because their solution is sought in some domain defined by a given boundary, on which certain constraints termed *boundary conditions* are specified. Except for free-boundary

problems, the boundary shape and its location are always known. Sometimes, however, it is the analyst's responsiblity to extract it from a bigger system. The boundary may be defined by a curve or a surface of n-dimensional space, and the domain it defines may be finite or infinite depending on the extremities of the boundary. The boundary is said to be *closed* if conditions affecting the solution of the problem are specified everywhere on the boundary (even though part of the boundary may extend to infinity) and *open* if part of the boundary extends to infinity and no boundary conditions are specified on the part at infinity. It is important to note that our definition of a boundary value problem departs from the usual one. The usual definition distinguishes between boundary value problems and initial value problems, where time is an independent variable. Because of our definition of the domain boundary, we may describe all partial differential equations and their boundary conditions as boundary value problems.

Terminology and Preliminary Considerations: Types of Nodes

In Figures 1.2 through 1.6, the number of basic element shapes and typical locations of nodes assigned to these elements are illustrated. We allude to the exterior and interior nodes, but now we can be more specific. Nodes are classified as either exterior or interior depending on their location relative to the geometry of an element. Exterior nodes lie on the boundary of an element, and they represent the points of connection between bordering elements. Nodes positioned at the corners of elements, along the edges, or on the surfaces are all exterior nodes. For one-dimensional elements such as those in Figure 1.2, there are only two exterior nodes because only the ends of the element connect to other one-dimensional elements. In contrast to exterior nodes, interior nodes are those that do not connect with neighboring elements.

Degrees of Freedom

Two other features, in addition to shape, characterize a particular element type: (1) the number of nodes assigned to the element and (2) the number and type of nodal variables chosen for it. Often the nodal variables associated with the element are referred to as the *degrees of freedom* of the element. This terminology, which we will adopt, is a spin-off from the solid mechanics field, where the nodal variables are usually nodal displacements and sometimes derivatives of displacements. Nodal

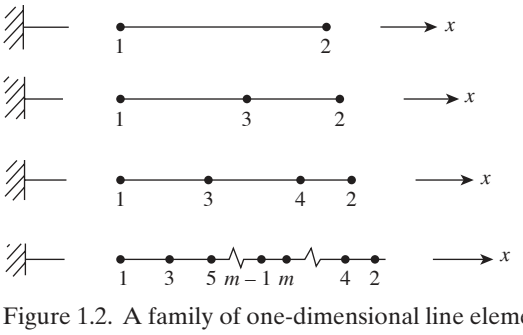

Figure 1.2. A family of one-dimensional line elements.

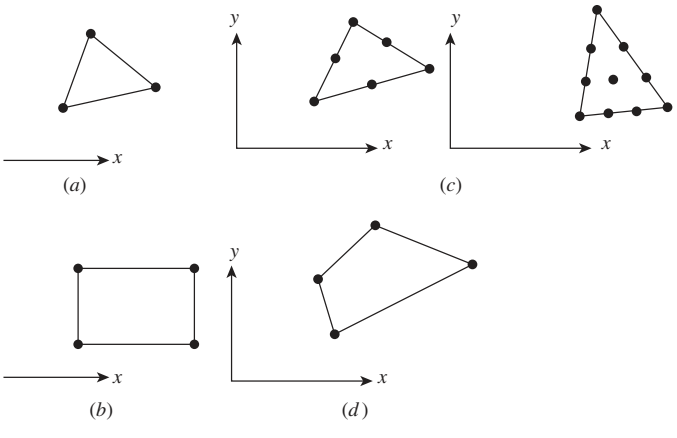

Figure 1.3. Examples of two-dimensional elements.

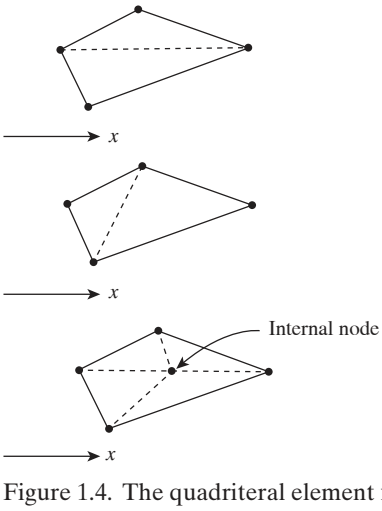

Figure 1.4. The quadriteral element formed by combining triangles.

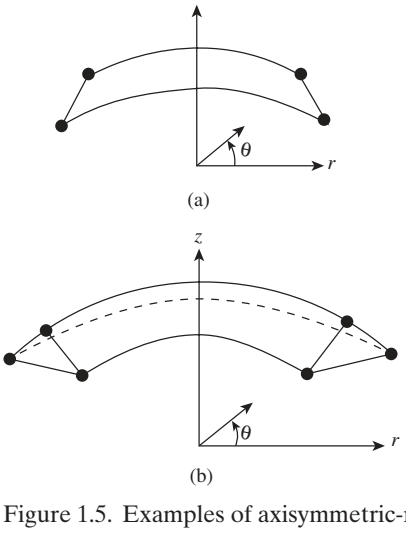

Figure 1.5. Examples of axisymmetric-ring elements.

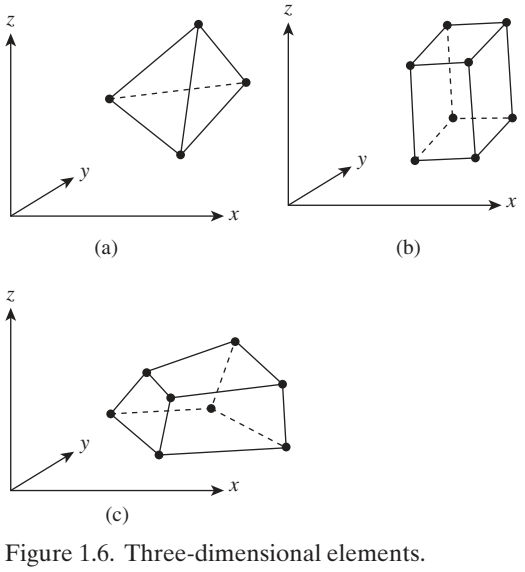

Figure 1.6. Three-dimensional elements.

degrees of freedom can be interior or exterior in relation to the element boundaries, depending on whether they are assigned to interior or exterior nodes.

In the fluid mechanics section of this book, one of the first few problems (Chapter 11) has to do with the fluid flow in multiply-connected solution domains. In this case, the field variable is the velocity potential ϕ. In addition, however, the circulation around the lifting body is added to a specific group of computational points as a *nodeless* degree of freedom. With the aid of this chapter's contents, therefore, the student is encouraged to generalize the term *degree of freedom* in a manner that is consistent with the problem physics.

Interpolation Functions: Polynomials

In the finite element literature, the functions used to represent the behavior of a field variable within an element are called *interpolation functions*, *shape functions*, or *approximation functions*. We have used and will continue to use only the first term in this text. Although it is conceivable that many types of functions could serve as interpolation functions, only polynomials have received widespread use. The reason is that polynomials are relatively easy to manipulate mathematically. In other words, they can be integrated or differentiated without difficulty. Trigonometric functions also possess this property, but they are seldom used. Here we will employ only polynomials of various types and orders to generate suitable interpolation functions. The polynomials we will consider follow.

One Independent Variable

In one dimension, a general complete nth-order polynomial may be written as follows:

$$P_n(x) = \sum_{i=0}^{T_n^{(1)}} \alpha_i x^{(i)}$$

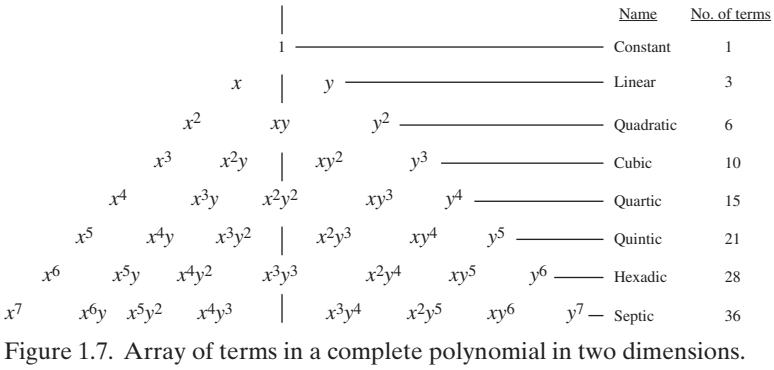

Figure 1.7. Array of terms in a complete polynomial in two dimensions.

where the number of terms in the polynomial is $T_n^{(1)} = n+1$. For $n=1$, $T_1^{(1)} = 2$, and $P_1(x) = \alpha_) + \alpha_1 x$. For $n=2$, $T_2^{(1)} = 3$, and $p_2(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2, \ldots$, and so on.

Two Independent Variables

In two dimensions, a complete nth-order polynomial may be written as follows:

$$P_n(x,y) = \sum_{k=1}^{T_n^{(2)}} \alpha_k x^i y^j, \qquad i+j \le n$$

where the number of terms in the polynomial is $T_n^{(2)} = [(n+1)(n+2)]/2$. For $n=1$, $T_1^{(2)} = 3$, and $P_1(x,y) = \alpha_1 + \alpha_2 x + \alpha_3 y$. For $n=2$, $T_2^{(2)} = 6$, and $P_2(x,y) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 xy + \alpha_5 x^2 + \alpha_6 y^2, \ldots$, and so on.

Gallagher [1] suggested a convenient way to illustrate the terms in a complete two-dimensional polynomial. If the terms are placed in a triangular array of ascending order, we obtain an arrangement similar to the Pascal triangle (Figure 1.7). We note that the sum of exponents of any term in this triangular array is the corresponding number in the well-known Pascal triangle of binomial coefficients.

Three Independent Variables

In three dimensions, a complete nth-order polynomial may be written as follows:

$$P_n(x,y,z) = \sum_{l=1}^{T_n^{(3)}} \alpha_l x^i y^j z^k \qquad i+j+k \le n$$

where the number of terms in the polynomial is

$$T_n^{(3)} = \frac{(n+1)(n+2)(n+3)}{6}$$

For $n=1$, $T_1^{(3)} = 4$, and $P_1(x,y,z) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 z$. For $n=2$, $T_2^{(3)} = 10$, and $P_2(x,y,z) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 z + \alpha_5 xy + \alpha_6 xz + \alpha_7 yz + \alpha_8 x^2 + \alpha_9 y^2 + \alpha_{10} z^2, \ldots$, and so on.

The terms in a complete three-dimensional polynomial may also be arrayed in a manner that is analogous to the triangular array in two dimensions. The array

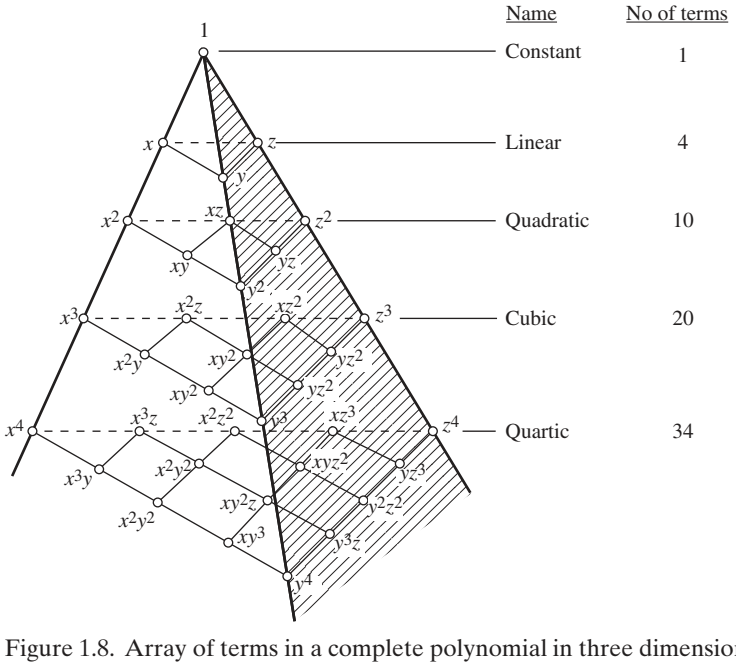

Figure 1.8. Array of terms in a complete polynomial in three dimensions.

becomes a tetrahedron, with the various terms placed at different planar levels, as shown in Figure 1.8.

Deriving Interpolation Functions

Thus far we have seen how a field variable can be represented within an element as a polynomial series whose coefficients are the generalized coordinates. In this section we will see how the interpolation functions for the physical degree of freedom are derived. These interpolation functions emerge from the basic procedure for expressing the generalized coordinates in terms of the nodal degrees of freedom.

The basic ideas can be illustrated through a simple example in two dimensions. Suppose that we wish to construct a rectangular element with nodes positioned at the element corners (Figure 1.9). If we assign one value of ϕ to each node, the element, then, will have four degrees of freedom, and we may select, as an interpolation model, a four-term polynomial such as

$$\phi(x,y) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 xy$$

The generalized coordinates may now be found by evaluating this interpolation function at each of the four nodes and then inverting the resulting set of simultaneous equations. Thus we may write

$$\phi_1 = \alpha_1 + \alpha_2 x_1 + \alpha_3 y_1 + \alpha_4 x_1 y_1$$

$$\phi_2 = \alpha_1 + \alpha_2 x_2 + \alpha_3 y_2 + \alpha_4 x_2 y_2$$

$$\phi_3 = \alpha_1 + \alpha_2 x_3 + \alpha_3 y_3 + \alpha_4 x_3 y_3$$

$$\phi_4 = \alpha_1 + \alpha_2 x_4 + \alpha_3 y_4 + \alpha_4 x_4 y_4$$

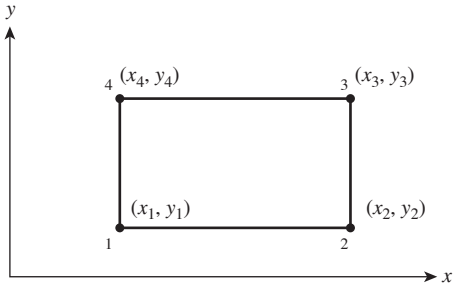

Figure 1.9. A rectangular element with sides parallel to the axes of the global coordinate system.

or, in matrix notation,

$$\{\phi\} = [G]\{\alpha\}$$

where the preceding column vectors and matrix are defined as follows:

$$\{\phi\} = \begin{Bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{Bmatrix}$$

$$[G] = \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 \\ 1 & x_2 & y_2 & x_2y_2 \\ 1 & x_3 & y_3 & x_3y_3 \\ 1 & x_4 & y_4 & x_4y_4 \end{bmatrix}$$

$$\{\alpha\} = \begin{Bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{Bmatrix}$$

In principle, then, we can express the generalized coordinates as the solution of the preceding matrix equation for $\{\alpha\}$, that is,

$$\{\alpha\} = [G]^{-1}\{\phi\}$$

Expressing the terms of the interpolation polynomial in the original expression for $\phi(x, y)$ as a product of a row vector and a column vector, we can write

$$\phi = [P]\{\alpha\}$$

where $$[P] = [1 \text{ x y xy}]$$

Thus, through simple substitution, we get

$$\phi = [P][G]^{-1}\{\phi\} = [N]\{\phi\}$$

with $$[N] = [P][G]^{-1}$$

This expression for the field variable ϕ, though obtained for one case, is generally applicable to all straight-sided elements. The original interpolation polynomial $[P]\{\alpha\}$ should not be confused with the interpolants N_i associated with the nodal degree of freedom. The distinction to note here is that $[P]\{\alpha\}$ is an interpolation function that applies to the whole element and expresses the field-variable behavior in terms of the generalized coordinates, whereas the interpolants N_i refer to individual nodes and individual degrees of freedom, and they represent the field-variable behavior. It is easy to see from the last expression of ϕ that the function N_i referring to node i takes on unit value at node i and zero value at all other nodes of the element.

The procedure for expressing the generalized coordinates in terms of the nodal degrees of freedom is actually the method commonly used to derive the nodal interpolation functions N_i. The procedure is straightforward and may be carried out easily, but sometimes difficulties are encountered. For some types of elements models, the inverse of the matrix $[G]$ may not exist for all orientations of the element in the *global* coordinate system. If an explicit expression for $[G]^{-1}$ is obtained algebraically, it may be possible to see under what conditions $[G]^{-1}$ does not exist and then try to avoid those circumstances when constructing the element mesh. Such an approach, however, is seldom recommended. Another disadvantage stems from the computational effort required to obtain $[G]^{-1}$ when it exists. For a large number of elements with many degrees of freedom, the computational cost can be prohibitive.

These reasons have motivated many researchers to try to obtain the nodal interpolation functions N_i by inspection, often relying on the use of special coordinate systems called *natural coordinates*. This particular topic is discussed separately in Appendix A for the general three-dimensional heat-conduction problem.

Throughout the remainder of this book, many elements with different shapes in two- and three-dimensional applications, including curve-sided elements, will be used. In each case, the shape functions will be defined and used in solving a variety of heat transfer and fluid mechanics problems.

Natural Coordinates

A local coordinate system that relies on the element geometry for its definition and whose coordinates range between zero and unity within the element is known as a *natural coordinate system*. Such systems have the property that one particular coordinate has a unity value at one node of the element and zero value at the other node(s); its variation between nodes is linear. We may construct natural coordinate systems for two-node line elements, three-node triangular elements, four-node quadrilateral elements, four-node tetrahedral elements, and so on.

The use of natural coordinates in deriving interpolation functions is particularly advantageous because special closed-form integration formulas can often be used to evaluate the integrals in the element equations. Natural coordinates also play a crucial role in the development of curve-sided elements.

The basic purpose of a natural coordinate system is to describe the location of a point inside an element in terms of the coordinates associated with the nodes of the element. We denote the natural coordinates as L_i, $(i = 1,2,\ldots,n)$, where n is the number of external nodes of the element. One coordinate is associated with