Mathematics

FINITE ELEMENT ANALYSIS AS COMPUTATION

What the textbooks don't teach you about finite element analysis

Chapter 3

Completeness and continuity: How to choose shape functions?

Gangan Prathap
Director
NISCAIR, S.V. Marg
New Delhi - 110016

Contents

Chapter 3

Completeness and continuity: How to choose shape functions?

3.1 Introduction

We can see from our study so far that the quality of approximation we can achieve by the RR or fem approach depends on the admissible assumed trial, field or shape functions that we use. These functions can be chosen in many ways. An obvious choice, and the one most universally preferred is the use of simple polynomials. It is possible to use other functions like trigonometric functions but we can see that the least squares or best-fit nature of stress prediction that the finite element process seeks in an instinctive way motivates us to prefer the use of polynomials. When this is done, interpretations using Legendre polynomials become very convenient. We shall see this happy state of affairs appearing frequently in our study here.

An important question that will immediately come to mind is - How does the accuracy or efficiency of the approximate solution depends on our choice of shape functions. Accuracy is represented by a quality called convergence. By convergence, we mean that as we add more terms to the RR series, or as we add more nodes and elements into the mesh that replaces the original structure, the sequence of trial solutions must approach the exact solution. We want quantities such as displacements, stresses and strain energies to be exactly recovered, surely, and if possible, quickly. This leads to the questions: What kind of assumed pattern is best for our trial functions? What are the minimum qualities, or essences or requirements that the finite element shape functions must show or meet so that convergence is assured. Two have been accepted for a long time: *continuity (or compatibility)* and *completeness*.

3.2 Continuity

This is very easy to visualize and therefore very easy to understand. A structure is sub-divided into sub-regions or elements. The overall deformation of the structure is built-up from the values of the displacements at the nodes that form the net or grid and the shape functions within elements. Within each element, compatibility of deformation is assured by the fact that the choice of simple polynomial functions for interpolation allows continuous representation of the displacement field. However, this does not ensure that the displacements are compatible between element edges. So special care must be taken otherwise, in the process of representation, gaps or overlaps will develop.

The specification of continuity also depends on how strains are defined in terms of derivatives of the displacement fields. We know that a physical problem can be described by the stationary condition $\delta\Pi=0$, where $\Pi=\Pi(\phi)$ is the functional. If Π contains derivatives of the field variable ϕ to the order m, then we obviously require that within each element, $\bar{\phi}$, which is the approximate field chosen as trial function, must contain a complete polynomial of degree m. So that $\bar{\phi}$ is continuous within elements and the completeness requirements (see below) are also met. However, the more important requirement now is that continuity of field variable ϕ must be maintained across element boundaries - this requires that the trial function $\bar{\phi}$ and its derivatives through order m-1 must be continuous across element edges. In most natural formulations in solid mechanics, strains are defined by first derivatives of the displacement fields. In such cases, a simple continuity of the displacement fields across element edges suffices - this is called C^0 continuity. Compatibility between adjacent elements for problems which require only C^0 -continuity can be easily assured if the displacements along the side of an element depend only on the displacements specified at all that nodes placed on that edge.

There are problems, as in the classical Kirchhoff-Love theories of plates and shells, where strains are based on second derivatives of displacement fields. In this case, continuity of first derivatives of displacement fields across element edges is demanded; this is known as C^1-continuity. Elements, which satisfy the continuity conditions, are called *conforming* or *compatible* elements.

We shall however find a large class of problems (Timoshenko beams, Mindlin plates and shells, plane stress/strain flexure, incompressible elasticity) where this simplistic view of continuity does not assure reasonable (i.e. practical) rates of convergence. It has been noticed that formulations that take liberties, e.g. the non-conforming or incompatible approaches, significantly improve convergence. This phenomenon will be taken up for close examination in a later chapter.

3.3 Completeness

We have understood so far that in the finite element method, or for that matter, in any approximate method, we are trying to replace an unknown function $\phi(x)$, which is the exact solution to a boundary value problem over a domain enclosed by a boundary by an approximate function $\bar{\phi}(x)$ which is constituted from a set of trial, shape or basis functions. We desire a trial function set that will ensure that the approximation approaches the exact solution as the number of trial functions is increased. It can be argued that the convergence of the trial function set to the exact solution will take place if $\bar{\phi}(x)$ will be sufficient to represent any well behaved function such as $\phi(x)$ as closely as possible as the number of functions used becomes indefinitely large. This is called the completeness requirement.

In the finite element context, where the total domain is sub-divided into smaller sub-regions, completeness must be assured for the shape functions used within each domain. The continuity requirements then provide the compatibility of the functions across element edges.

We have also seen that what we seek is a best-fit arrangement in some sense between $\bar{\phi}(x)$ and $\phi(x)$. From Chapter 2, we also have the insight that this best-fit can be gainfully interpreted as taking place between strain or stress quantities. This has important implications in further narrowing the focus of the completeness requirement for finite element applications in particular.

By bringing in the new perspective of a best-fit strain or stress paradigm, we are able to look at the completeness requirements entirely from what we desire the strains or stresses to be like. Of paramount importance now is the idea that the approximate strains derived from the set of trial functions chosen must be capable in the limit of approximation (i.e. as the number of terms becomes very large) to describe the true strain or stress fields exactly. This becomes very clear from the orthogonality condition derived in Equation 2.25.

From the foregoing discussion, we are now in a position to make a more meaningful statement about what we mean by completeness. We would like displacement functions to be so chosen that no straining within an element takes place when nodal displacements equivalent to a rigid body motion of the whole element are applied. This is called the *strain* *free rigid body motion condition.* In addition, it will be necessary that each element must be able to reproduce a state of constant strain; i.e. if nodal displacements applied to the element are compatible with a constant strain state, this must be reflected in the strains computed within the element. There are simple rules that allow these conditions to be met and these are called the completeness requirements. If polynomial trial functions are used, then a simple assurance that the polynomial functions contain the constant and linear terms, etc. (e.g. *1, x* in a one-dimensional C^0 problem;) 1, x, y in a two-dimensional C^0 problem so that each element is certain to be able to recover a constant state of strain, will meet this requirement.

3.3.1 Some properties of shape functions for C^0 elements

The displacement field is approximated by using shape functions N_i within each element and these are linked to nodal degrees of freedom u_i. It is assumed that these shape functions allow elements to be connected together without generating gaps or overlaps, i.e. the shape functions satisfy the continuity requirement described above. Most finite element formulations in practical use in general purpose application require only what is called C^0 continuity - i.e. a field quantity ϕ must have interelement continuity but does not require the continuity of all of its derivatives across element boundaries.

C^0 shape functions have some interesting properties that derive from the completeness requirements imposed by rigid body motion considerations. Consider a field ϕ (usually a displacement) which is interpolated for an n-noded element according to

$$\phi = \sum_{i=1}^{n} N_i \phi_i \qquad (3.1)$$

where the N_i are assumed to be interpolated using natural co-ordinates (e.g. ξ, η for a 2-dimensional problem). It is simple to argue that one property the shape functions must have is that N_i must define the distribution of ϕ within the element domain when the degree of freedom ϕ_i at node *i* has unit value and all other nodal ϕ*'s* are zero. Thus

i) $N_i = 1$ when $x = x_i$ (or $\xi = \xi_i$) and $N_i = 0$ when $x = x_j$ (or $\xi = \xi_j$) for $i \neq j$.

Consider now, the case of the element being subjected to a rigid body motion so that at each node a displacement $u_i=1$ is applied. It is obvious from physical considerations that every point in the element must have displacements $u=1$. Thus from Eq 3.1 we have

ii) $\Sigma N_i = 1$.

This offers a very simple check for shape functions. Another useful check comes again from considerations of rigid body motion - that a condition of zero strain must be produced when rigid body displacements are applied to the nodes. The reader is encouraged to show that this entails conditions such as

iii) $\Sigma N_{i,\xi} = 0$, $\Sigma N_{i,\eta} = 0$

for a 2-dimensional problem.

Note that for C^1 elements, where derivatives of ϕ are also used as nodal degrees of freedom, these rules apply only to those N_i which are associated with the translational degrees of freedom.

3.4 A numerical experiment regarding completeness

Let us now return to the uniform bar carrying an axial load P as shown in Fig. 2.1. The exact solution for this problem is given by a state of constant strain. Thus, the polynomial series $\bar{u}(x) = a_0 + a_1 x$ (see Section 2.3) will contain enough terms to ensure completeness. We could at the outset arrange for a_0 to be zero to satisfy the essential boundary condition. With $\bar{u}_1(x) = a_1 x$ we were able to obtain the exact solution to the problem. We can now investigate what will happen if we omit the $a_1 x$ term and start out with the trial function set $\bar{u}_2(x) = a_2 x^2$. It is left to the reader to show that this results in a computed stress variation $\bar{\sigma}_2 = 3Px/2EAL$; i.e. a linear variation is seen instead of the actual constant state of stress. The reader is now encouraged to experiment with RR solutions based on other incomplete sets like $\bar{u}_3(x) = a_3 x^3$ and $\bar{u}_{2+3}(x) = a_2 x^2 + a_3 x^3$. Figure 3.1 shows how erroneous the computed stresses σ_2, σ_3 and σ_{2+3} are.

The reason for this is now apparent. The term that we have omitted is the one needed to cover the state of constant strain. In this case, this very term also provides the exact answer. The omission of this term has led to the loss of completeness. In fact, the reader can verify that as long as this term is left out of any trial set containing polynomials of higher order, there will not be any sort of clear convergence to the true state of strain. In the present example, because of the absence of the linear term in $\bar{u}$ (the constant strain term therefore vanishes), the computed strain vanishes at the origin!

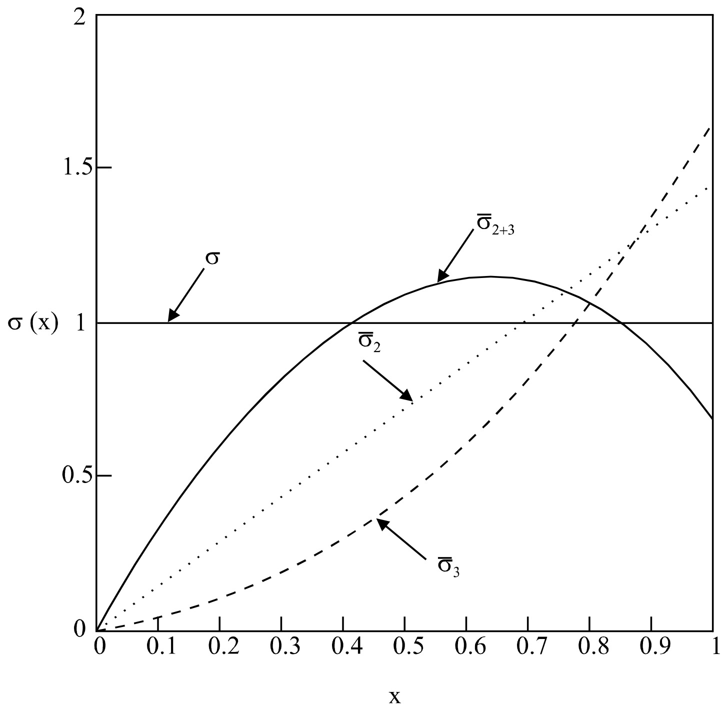

Fig. 3.1 Stresses from RR solutions based on incomplete polynomials

3.5 Concluding remarks

We have now laid down some specific qualities (continuity and completeness) that trial functions must satisfy to ensure that approximate solutions converge to correct results. It is also hoped that if these requirements are met, solutions from coarse meshes will be reasonably accurate and that convergence will be acceptably rapid with mesh refinement. This was the received wisdom or reigning paradigm around 1977. In the next chapter, we shall look at the twin issues of errors and convergence from this point of view.

Unfortunately, elements derived rigorously from only these basic paradigms can behave in unreasonably erratic ways in many important situations. These difficulties are most pronounced in the lowest order finite elements. The problems encountered were called *locking*, *parasitic shear*, etc. Some of the problems may have gone unrecorded with no adequate framework or terminology to classify them. As a very good example, for a very long time, it was believed that curved beam and shell elements performed poorly because they could not meet the *strain-free rigid body motion* condition. However, more recently, the correct error inducing mechanism has been discovered and these problems have come to be called *membrane locking*. An enlargement of the paradigm is now inevitable to take these strange features into account. This shall be taken up from Chapter 5.