

***A priori* error estimation of finite element models from first principles**

GANGAN PRATHAP

National Aerospace Laboratories, Bangalore 560 017, India, and
Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064,
India

e-mail: gp@css.cmmacs.ernet.in

MS received 30 May 1998

Abstract. The quality of finite element computational results can be assessed only by providing rational criteria for evaluating errors. Most exercises in this direction are based on *a posteriori* error estimates, based primarily on experience and intuition. If finite element analysis has to be considered a rational science, it is imperative that procedures to compute *a priori* error estimates from first principles are made available. This paper captures some efforts in this direction.

Keywords. Finite elements; error analysis; finite element discretizations; analytical quantification; stress correspondence paradigm.

1. Introduction

The finite element method (FEM) provides numerical results to physical problems which are already described by mathematical models. The answer may be wrong because the physical model was wrongly described by the mathematical model. We are not interested in this issue here. Further, as our computational models are now created and manipulated using digital computers, there are errors which occur due to the fact that information in the form of numbers can be stored only to a finite precision (“word length” as it is called) at every stage of the computation. These are called round-off errors. We shall assume here that in most problems we deal with, word length is sufficient so that round-off error is not a major headache.

The real issue for us to grapple with now is that the computational model prepared to simulate the mathematical model may be faulty and can lead to errors. In the process of replacing the continuum region by finite elements, errors originate in many ways. From physical intuition, we can argue that this will depend on the type, shape and number of elements that we use, the grading or density of the mesh used, the way distributed loads are assigned to nodes, the manner in which boundary conditions are modelled by

specification of nodal degrees of freedom etc. These are the discretization errors that can occur.

Most of such errors are difficult to quantify analytically or determine in a logically coherent way. We can only rely on heuristic judgement (based on intuition and experience) to understand how best to minimize errors. However, we shall now look only at that category of discretization error that appears because the computational or discretized model uses trial functions which are an approximation of the true solution to the mathematical model. It seems possible that to some extent, analytical quantification of these errors is possible.

The finite element method is based on variational procedures, like the minimum total potential principle and Hamilton's principle, which offer a powerful methodology to replace continuum structures with approximate systems from which one can compute quantities of interest such as displacements, strains and stresses. However, the minimum total potential principle by itself is unable to assess the quality of the solution; how the approximations made during the derivation process lead to errors or ensure a certain degree of accuracy. It is therefore necessary to introduce first principles axiomatised from a generalised principle called the Hu–Washizu principle (Hu 1955), and this provided considerable explanatory power in assessing the quality of solution using finite element models. The stress correspondence paradigm (Prathap 1996) shows that the FEM procedure intrinsically samples stresses first and produces displacement fields only in a secondary manner. Similarly, a momentum correspondence paradigm can be introduced to evaluate elastodynamic problems.

2. The stress correspondence principle and error estimates

The central issue in error analysis of the finite element method is to determine how exactly the finite element discretisation process responds to the actual continuum situation. We show (Prathap 1993, 1996) that in a sub-domain region covered by a finite element in an elastostatics problem, one can predict that the finite element solution (\bar{u} , $\bar{\epsilon}$, $\bar{\sigma}$) will correspond to the actual (u etc.) according to the orthogonality condition expressed by,

$$\int \delta \bar{\sigma}^T (\bar{\epsilon} - \epsilon) dV = 0. \quad (1)$$

This can serve as the basis for engineering error estimates for static structural problems. We shall briefly see how this is done here.

If displacement fields are chosen so that strains $\bar{\epsilon}$ (or stresses $\bar{\sigma}$) are complete to order x^{n-1} (for simplicity, a one-dimensional problem with x , or non-dimensional co-ordinate ζ is chosen), the finite element can represent a more accurate continuum field of order x^n in the best-fit manner shown in figure 1.

Let us assume that,

$$\sigma = a_0 + a_1 h L_1(\zeta) + a_2 h^2 L_2(\zeta) + \dots + a_n h^n L_n(\zeta), \quad (2)$$

where $L_n(\zeta)$ are suitably normalised Legendre polynomials. This form is specially chosen to exploit the orthogonality condition given by (1). The best-fit argument represented by (1) now leads to a finite element estimate

$$\sigma = a_0 + a_1 h L_1(\zeta) + a_2 h^2 L_2(\zeta) + \dots + a_{n-1} h^{n-1} L_{n-1}(\zeta). \quad (3)$$

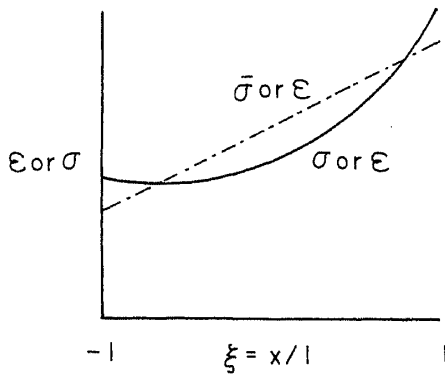


Figure 1. Best-approximation nature of stress and momentum correspondence principles.

Thus, a finite element computation produces approximate stresses and strains which are accurate to $O(h^n)$. It is possible to proceed from this to show that the error in the potential energy due to finite element discretisation is of the form: (using ϵ and σ interchangeably for simplicity)

$$\begin{aligned}
 \frac{1}{2} \int (\sigma^2 - \bar{\sigma}^2) dx &= \frac{1}{2} \int (\sigma - \bar{\sigma})(\sigma + \bar{\sigma}) dx \\
 &= \frac{1}{2} \int \bar{\sigma}(\sigma - \bar{\sigma}) dx + \frac{1}{2} \int \sigma(\sigma - \bar{\sigma}) dx \\
 &= 0 + \frac{1}{2} \int \sigma(\sigma - \bar{\sigma}) dx \\
 &= \frac{1}{2} a_n^2 h^{2n} \int L_n^2(\zeta) dx, \tag{4}
 \end{aligned}$$

where the condition that $\bar{\sigma}$ is a best-fit of σ as well as the orthogonality of the Legendre polynomials has been used. Thus if approximate strain or stress fields are accurate to $O(h^n)$, the potential energies are accurate to $O(h^{2n})$, after the process of finite element discretisation. The quality of the stiffness matrix is therefore governed by this order of accuracy and the overall gross errors (predictions of displacements, say) are removed at the optimal rate of $O(h^{2n})$. This is of course the optimal rate of convergence expected in an elastostatics problem; convergence rate may worsen due to errors in representing loads approximately, and may even improve due to fortuitous cancellation of errors! These predictions have been made possible because of the orthogonality conditions that emerged by using the generalised Hu–Washizu theorem instead of the minimum total potential energy principle.

3. The momentum correspondence principle

Mass matrices are derived rationally from the kinetic energy of an elastodynamic system by using Hamilton's principle. Just as the Hu–Washizu principle allowed the minimum total potential energy principle to be generalised to a three-field statement based on simultaneously varying displacement, strain and stress fields, a similar generalisation can be attempted of the kinetic energy.

If u and \bar{u} are the displacement fields, the velocity fields are v and \bar{v} (where v indicates velocity obtained by differentiating u with respect to time), the kinetic energies before and

after discretisation are $\frac{1}{2} \int v^T \rho v \, dx$ and $\frac{1}{2} \int \bar{v}^T \rho \bar{v} \, dx$ respectively, where ρ is the density. Let the generalised functional representing the discretisation process be written as

$$\Pi = \frac{1}{2} \int \bar{v}^T \rho \bar{v} \, dx + \int \bar{\sigma}^T (v - \bar{v}) \, dx, \quad (5)$$

where one can interpret $\bar{\sigma}$ as the Lagrangian multiplier removing the dislocation of the velocity potentials introduced by the discretisation process. There are three fields now, v , \bar{v} and $\bar{\sigma}$, and the variation leads to an orthogonality condition of particular significance for us here, namely

$$\int \delta \bar{v}^T \rho (v - \bar{v}) \, dx = 0. \quad (6)$$

This condition implies that in the finite element discretisation process, momentum or velocity fields are preserved in a best-fit manner. It follows from this that if v is replaced in a discretisation process by \bar{v} by choosing trial functions which are complete to order x^n , the procedure will have an accuracy of $O(h^{n+1})$ in representing displacements and velocities. Then, the kinetic energies, arguing by analogy from (4), are accurate after discretisation to $O(h^{2(n+1)})$.

4. Error estimates from the Rayleigh quotient

Clearly, in any consistent mass formulation involving the use of trial functions for displacements u (and hence velocities v) complete to x^n , the potential energy (and hence the quality of the stiffness matrix) is accurate to $O(h^{2n})$ whereas the kinetic energy (and hence the quality of the consistent mass matrix) is of $O(h^{2(n+1)})$. If the mass matrix is derived in an extra-variational way (non-consistent, as in a lumped matrix), this order of accuracy cannot be assured.

We can now use the Rayleigh's quotient approach to show how the quality of a finite element solution to a dynamic problem can be interpreted in terms of the qualities of the stiffness and mass matrices.

The Rayleigh quotient approximation (of which the finite element method is a piecewise version) is based on the description of the fundamental frequency as a minimum of the Rayleigh quotient. The piecewise element trial functions can be substituted and assembled directly into the quotient to obtain stiffness and mass matrices leading to the discretised eigenvalue problem, $Kd = \bar{\lambda}Md$, or in Rayleigh quotient form,

$$\bar{\lambda} = \frac{d^T K d}{d^T M d}, \quad (7)$$

where $\bar{\lambda}$ is the approximate eigenvalue and d , the eigenvector of displacements. This quotient can now be used to investigate the quality of the eigenvalues in terms of the order of error or rate of convergence, by independently assessing the errors in the strain (potential) energy of $d^T K d$ and kinetic energy $d^T M d$, using results from the stress correspondence and momentum correspondence principles we have described earlier.

We have seen above that in a finite element formulation using trial functions for displacements (velocities) complete to x^n , the potential energies (where, for simplicity, strains

are assumed to be first derivatives of the displacements) after discretisation are accurate to $O(h^{2n})$, whereas the kinetic energies are accurate to $O(h^{2(n+1)})$ if a consistent mass approach is strictly observed in formulating the mass matrix. From this, it can be argued that if λ , K and M are the quantities pertaining to an exact analytical solution, then the discretised solution $\bar{\lambda}$ is given by,

$$\begin{aligned}\bar{\lambda} &= \{d^T K d + O(h^{2n})\} / \{d^T M d + O(h^{2(n+1)})\} \\ &= \lambda + O(h^{2n}) - O(h^{2(n+1)}),\end{aligned}\tag{8}$$

if $\bar{\lambda}$ is a good approximation of λ .

It is apparent that the error that dominates is due to the formulation of the stiffness matrix. This accounts for the well-known upper bound nature of the eigenvalues obtained when consistent mass matrices are used. In any non-consistent (such as a lumped mass) formulation, the order of accuracy associated with the kinetic energy would be poorer than $O(h^{2(n+1)})$. By careful manipulation, it is possible to obtain a lumped mass matrix that has an accuracy $O(h^{2n})$, which is the accuracy of the potential energy term in the numerator of the Rayleigh quotient. It is further possible that the rate of convergence of the eigenvalue is as good as that obtained with a consistent mass formulation. It is also possible that lumping procedure can be so chosen that the errors from the numerator and the denominator of $O(h^{2n})$ cancel out giving a higher rate of convergence. The so-called higher-order masses actually achieve such accuracy by this process of cancellation of errors.

5. Types of discretisation errors and a methodology for error estimation

We can recognize two kinds of discretization errors due to trial function approximation. The first arises because a problem with an infinitely large number of degrees of freedom is replaced by a model with a finite number of degrees of freedom. Therefore, except in very rare cases, the governing differential equations and boundary conditions are satisfied only approximately. The second appears due to the fact that by overlooking certain essential requirements beyond that specified by continuity and completeness, the mathematical model can alter the physics of the problem. In both cases, we should be able to satisfy ourselves that the discretization process which led to the computational model has introduced a certain predictable degree of error and that this converges at a predictable rate, i.e. the error is removed in a predictable manner as the discretization is improved in terms of mesh refinement. Error analysis is the attempt to make such predictions *a priori*, or rationalize the errors in a logical way, *a posteriori*, after the errors are found.

To carry out error analysis, new procedures have to be invented. These must be set apart from the first-order tradition procedures that carry out the discretization (creating the computational model from the mathematical model) and solution (computing the approximate results). Thus, we must design auxiliary procedures that can trace errors in an *a priori* fashion from basic paradigms (conjectures or guesses). These error estimates or predictions can be seen as consequences computed from our guesses about how the FEM works. These errors must now be verified by constructing simple digital computation exercises. This is what we seek to do now. If this cycle can be completed, we can assure ourselves that we

have carved out a scientific basis for error analysis. This is a very crucial element of our study. The FEM, or for that matter, any body of engineering knowledge, or engineering methodology, can be said to have acquired a scientific basis only when it has incorporated within itself, these auxiliary procedures that permit its own self-criticism. Therefore, error analysis, instead of being only *a posteriori* or *post mortem* studies, as it is usually practised, must ideally be founded on *a priori* projections computed from intelligent paradigms which can be verified (falsified) by digital computation.

We next take note of the newly established paradigm that the Ritz-type and FEM approximations seek strains/stresses in a 'best-fit' manner. From such an interpretation, we examine if it is possible to argue that errors, whether in displacements, stresses or energies, due to finite element discretization must diminish rapidly, at least in an $(l/L)^2$ manner or better, where a large structure (domain) of dimension L is sub-divided into elements (sub-domains) of dimension l . Thus, with ten elements in a one-dimensional structure, errors must not be more than a few percent. This is the usual range of problems where the continuity and completeness paradigms explain completely the performance of finite elements. We shall discover that a class of problems exists where errors are much larger – the discretizations fail in a dramatic fashion. Convergence and error analysis must now be founded on a more complex conceptual framework – new paradigms need to be introduced and falsified.

6. Cantilever beam idealized with linear Timoshenko beam elements

The example that is ideally suited for demonstrating how error analysis can be founded on first principles is a uniform beam with a tip load as shown in figure 2. We shall model it with linear (two-node) Timoshenko beam elements which represent the bending moment within each element by a constant. Since the bending moment varies linearly over the beam for this problem, the finite element will replace this with a staircase approximation. Thus, with increase in number of elements, the stress pattern will approach the true solution more closely and therefore the computed strain energy due to bending will also converge. Since the applied load is at the tip, it is very easy to associate the load with the deflection under it using Castigliano's theorem. It will then be possible to discern the convergence trend for the tip deflection. Our challenge is therefore to see if the best-fit paradigm can be used to predict the convergence rate for this example from first principles.

The dimensions of a cantilever under tip load (figure 2) are chosen such that the tip deflection under the load will be $w = 4.0$. The example chosen represents a thin beam so that the influence of shear deformation and shear strain energy is negligible.

We shall now discretize the beam using equal length linear elements based on Timoshenko theory. We use this element instead of the classical beam element for several reasons. This element serves to demonstrate the features of shear locking which arise from

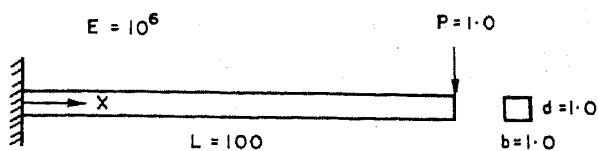


Figure 2. Cantilever beam under tip load.

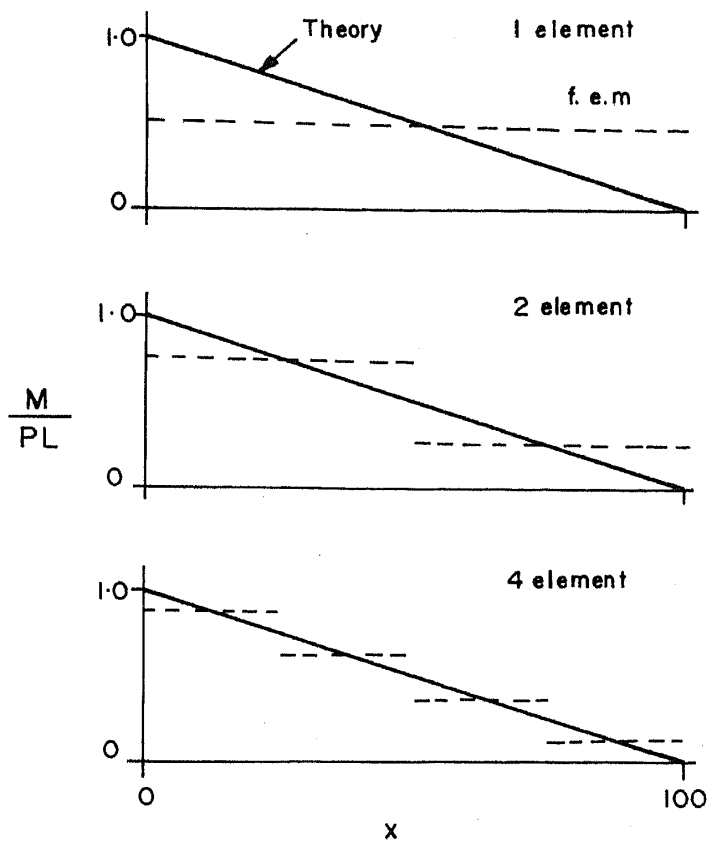


Figure 3. Bending moment diagrams for a one-, two- and four-element idealisations of a cantilever beam under tip load.

an inconsistent definition of the shear strains (which we examine in § 7). After correcting for the inconsistent shear strain, this element permits constant bending and shear strain accuracy within each element – the simplest representation possible under the circumstances and therefore an advantage in seeing how it works in this problem.

We shall restrict attention to the bending moment variation as we assume that the potential energy stored is mainly due to bending strain and that we can neglect the transverse shear strain energy for the dimensions chosen.

Figure 3 shows the bending moment diagrams for a 1, 2 and 4 element idealizations of the present problem using the linear element. The true bending moment (shown by the solid line) varies linearly. The computed (i.e. discretized) bending moments are distributed in a piecewise constant manner as shown by the broken lines. In each case, the elements pick up the bending moment at the centroid correctly – i.e. it is doing so in a ‘best-fit’ manner. What we shall now attempt to do for this problem is to relate this to the accuracy of results. We shall now interpret accuracy in the conventional sense, as that of the deflection at the tip under the load. Table 1 shows the normalized tip deflection with increasing idealizations (neglecting a very small amount due to shear deformation). An interesting pattern emerges. If error is measured by the norm $\{w - w(\text{FEM})\}/w$, it turns out that this is given exactly by the formula $1/4N^2$ where N is the number of elements. It can now be seen that this relationship can be established by arguing that this feature emerges from the fact that strains are sought in the ‘best-fit’ manner shown in figure 3.

Consider a beam element of length $2l$ (figure 4). Let the moment and shear force at the centroid be M and V . Thus the true bending moment over the element region for our

Table 1. Normalized tip deflections of a thin cantilever beam, $L/t = 100$.

N	Predicted	Computed
1	0.750	0.750
2	0.938	0.938
4	0.984	0.984

problem can be taken to vary as $M + Vx$. The discretized bending moment sensed by our linear element would therefore be M . We shall now compute the actual bending energy in the element region (i.e. from a continuum analysis) and that given by the finite element (discretized) model. We can show that

$$\text{Energy in continuum model} = (l/EI)(M^2 + V^2l^2/3), \quad (9)$$

$$\text{Energy in discretized model} = (l/EI)(M^2). \quad (10)$$

Thus, as a result of the discretization process involved in replacing each continuum segment of length $2l$ by a linear Timoshenko beam element which can give only a constant value M for the bending moment, there is a reduction (error) in energy in each element equal to $(l/EI)(V^2l^2/3)$. It is simple now to show from this that for the cantilever beam of length L with a tip load P , the total reduction in strain energy of the discretized model for the beam is $U/4N^2$ where $U = P^2L^3/6EI$ is the energy of the beam under tip load.

We are interested now in discovering how this error in strain energy translates into an error in the deflection under load P . This can be very easily deduced using Castigliano's second theorem. It is left to the reader to show that the tip deflections of the continuum and discretized model will differ as $\{w - w(\text{FEM})\}/w = 1/4N^2$.

Table 1 shows this predicted rate of convergence. Our foregoing analysis shows that this follows from the fact that if any linear variation is approximated in a piecewise manner by constant values as seen in figure 3, this is the manner in which the square of the error

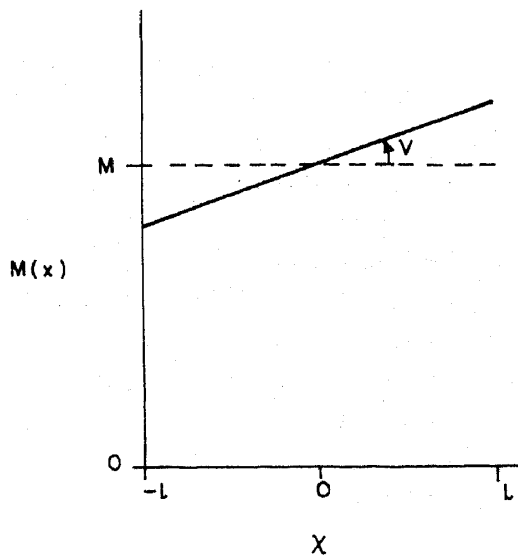


Figure 4. Bending moment variation in a linear beam element.

in the stresses/strains (or, equivalently, the difference in work or energy) will converge with idealization. Of course, in a problem where the bending moment is constant, the rate of convergence will be better than this (in fact, exact) and in the case where the bending moment is varying quadratically or at a higher rate, the rate of convergence will be decidedly less.

We also notice that convergence in this instance is from below. This can be deduced from the fact that the discretized potential energy \bar{U} is less than the actual potential energy U for this problem. It is frequently believed that the finite element displacement approach always underestimates the potential energy and a displacement solution is consequently described as a *lower bound solution*. However, this is not a universally valid generalization. We can see briefly below (the reader is in fact encouraged to work out the case in detail), where the cantilever beam has a uniformly distributed load acting on it using the same linear Timoshenko beam element for discretization, that this is not the case. It turns out that tip rotations converge from above (in a $1/2N^2$ rate) while the tip deflections are fortuitously exact. The *lower bound solution* nature has been disturbed because of the necessity of altering the load system at the nodes of the finite element mesh under the 'consistent load' lumping procedure.

As promised above, we now extend the concept of "best-fit" and variationally correct rate of convergence to the case of uniformly distributed load of intensity q on the cantilever with a little more effort. Now, when a finite element model is made, two levels of discretization error are introduced. First, the uniformly distributed load is replaced by consistent loads which are concentrated at element nodes. Thus, the first level of discretization error is due to the replacement of the quadratically varying bending moment in the actual beam with a linear bending moment within each beam element. Over the entire beam model, this variation is piecewise linear. The next level of error is due to the approximation implied in developing the stiffness matrix which we had considered above – this effectively senses a "best-approximated" constant value of the bending moment within each element of the linear bending moment appearing to act after load discretization.

With these assumptions, it is a simple exercise using Castigliano's theorem and fictitious tip force and moment P and M respectively to demonstrate that the finite element model of such a problem using two-noded beam elements will yield a fortuitously correct tip deflection ($w = qL^4/8EI$) for all idealizations (i.e. even with one element!) and a tip rotation that converges at the rate $1/2N^2$ from above to the exact value ($\theta = qL^3/6EI$). Thus, as far as tip deflections are concerned, the two levels of discretization errors have nicely cancelled each other to give correct answers. This can deceive an unwary analyst into believing that an exact solution has been reached. Inspection of the tip rotation confirms that the solution is approximate and converging.

We see from the foregoing analysis that by using linear Timoshenko beam elements for the tip-loaded cantilever, the energy for this problem converges as $O(h^2)$, where $h = 2l = L/N$ is the element size. We also see that this order of convergence carries over to the estimate of the tip deflections for this problem. Many text-books are confused over such relationships, especially those that proceed on the order of error analysis. These approaches arrive at conclusions, such as that the strain error is proportional to element size, i.e. $O(h)$, and displacement error is proportional to the square of the element size,

i.e. $O(h^2)$, for this problem. We can see that for this problem (see figure 3) this estimate is meaningful if we consider the maximum error in strain as occurring at element nodes (at centroids the errors are zero as these are optimal strain points). We also see that with element discretization, these errors in strain vanish as $O(h)$. We can also see that the strain energies are now converging at the rate of $O(h^2)$ and this emerges directly from the consideration that the discretized strains are 'best-fits' of the actual strain. This conclusion is not so readily arrived at in the order of error analysis methods, which often argue that if strains are accurate to $O(h)$, then strain energies are accurate to $O(h^2)$ because strain energy expressions contain squares of the strain. This conclusion is valid only for cases where the discretized strains are 'best-fit' approximations of the actual strains, as observed in the present example. If the 'best-fit' paradigm did not apply, the only valid conclusion that could be drawn is that the strains which have $O(h)$ error will produce errors in strain energy which are $O(2h)$.

It is also possible to argue that errors, whether in displacements, stresses or energies, due to finite element discretization must converge rapidly, at least in a $O(h^2)$ manner or better. If a large structure (domain) of dimension L is sub-divided into elements (sub-domains) of dimension l , one expects errors of the order of $(l/L)^2$. Thus, with ten elements in a one-dimensional structure, errors must not be more than a few percent. We discover however that a class of problems exist where errors are much larger – the discretizations fail in a dramatic fashion, and this cannot be resolved by the classical (pre-1977) understanding of the finite element method. A preliminary study of the issues involved will be taken up next; the linear Timoshenko beam element serves to expose the factors clearly.

7. The locking phenomena

By and large, finite elements work without difficulty. However, there are spectacular failures as well. These are what are now called the 'locking' problems. By locking, we mean that finite element solutions vanish quickly to zero (errors saturating quickly to nearly 100%!) as certain parameters (the penalty multipliers) become very large. It was not clear why the displacement type method, as it was understood around 1977, should produce for such problems, answers that are only a fraction of a percent of the correct answer with a practical level of discretization. Studies in recent years have established that an aspect known as consistency must be taken into account (Prathap 1993).

The consistency paradigm requires that the interpolation functions chosen to initiate the discretization process must also ensure that any special constraints that are anticipated must be allowed for in a consistent way. Failure to do so causes solutions to lock to erroneous answers. The paradigm showed how elements can be designed to be free of these errors. It also enabled error-analysis procedures that allowed errors to be traced to the inconsistencies in the representation to be developed.

7.1 Locking and the linear Timoshenko beam element

The two-noded beam element based on the shear flexible Timoshenko beam theory will need only C^0 continuity and can be based on simple linear interpolations. It was therefore

very attractive for general purpose applications. However, the element was beset with problems, as we shall presently see.

The strain energy of a Timoshenko beam element of length $2l$ can be written as the sum of its bending and shear components as

$$\int \left(\frac{1}{2} EI \kappa^T \kappa + \frac{1}{2} kGA \gamma^T \gamma \right) dx, \tag{11}$$

where

$$\kappa = \theta_{,x} \tag{12a}$$

$$\gamma = \theta - w_{,x}. \tag{12b}$$

In (12a) and (12b), w is the transverse displacement and θ the section rotation. E and G are the Young's and the shear moduli and k the shear correction factor used in Timoshenko's theory. I and A are the moment of inertia and the area of cross-section respectively.

In the conventional procedure, linear interpolations are chosen for the displacement field variables as,

$$N_1 = (1 - \zeta)/2, \tag{13a}$$

$$N_2 = (1 + \zeta)/2, \tag{13b}$$

where the dimensionless coordinate $\zeta = x/l$ varies from -1 to $+1$ for an element of length $2l$. This ensures that the element is capable of strain-free rigid body motion and can recover a constant state of strain (completeness requirement), and that the displacements are continuous within the element and across the element boundaries (continuity requirement). We can compute the bending and shear strains directly from these interpolations using the strain gradient operators given in (12a) and (12b). These are then introduced into the strain energy computation in (11), and the element stiffness matrix is calculated in an analytically or numerically exact (a 2-point Gauss-Legendre integration rule) way.

We shall now model a cantilever beam under a tip load using this element, considering the case of a "thin" beam with $E = 1000$, $G = 37,500,000$, $t = 1$, $L = 4$, using a fictitiously large value of G to simulate the "thin" beam condition. Table 2 shows that the normalized tip displacements are dramatically in error. We can carefully examine table 2 to see the trend as the number of elements are increased. The tip deflections obtained, which are several orders of magnitude lower than the correct answer, are directly related to the square of the number of elements used for the idealization. In other words, the discretization process has introduced an error so large that the resulting answer has a stiffness related to

Table 2. Normalised tip deflections.

No. of elements	"Thin" beam
1	0.200×10^{-5}
2	0.800×10^{-5}
4	0.320×10^{-4}
8	0.128×10^{-3}
16	0.512×10^{-3}

the inverse of N^2 . This is clearly unrelated to the physics of the Timoshenko beam and is also not the usual sort of discretization error encountered in the finite element method. It is this very phenomenon that is known as *shear locking*.

The error in each element must be related to the element length, and therefore when a beam of overall length L is divided into N elements of equal length h , the additional stiffening introduced in each element due to shear locking is seen to be proportional to h^2 . In fact, numerical experiments showed that the locking stiffness progresses without limit as the element depth t decreases. Thus, we now have to look for a mechanism that can explain how this spurious stiffness of $(h/t)^2$ can be accounted for by considering the mathematics of the discretization process.

It is clear from the formulation of the linear Timoshenko beam element using exact integration (we shall call it the *field-inconsistent* (FI) element) that ensuring the *completeness* and *continuity* conditions is not enough in some problems. We shall propose a requirement for a *consistent* interpolation of the constrained strain fields as the necessary paradigm to make our understanding of the phenomena complete.

If we start with linear trial functions for w and θ , as we had done in (13) above, we can associate two generalized displacement constants with each of the interpolations in the following manner.

$$w = a_0 + a_1(x/l), \quad (14a)$$

$$\theta = b_0 + b_1(x/l). \quad (14b)$$

We can relate such constants to the field-variables in this element in a discretized sense; thus, $a_1/l = w_{,x}$ at $x = 0$, $b_0 = \theta$ and $b_1/l = \theta_{,x}$, at $x = 0$. This denotation would become useful when we try to explain how the discretization process can alter the infinitesimal description of the problem if the strain fields are not consistently defined.

If the strain-fields are now derived from the displacement fields given in (14), we get

$$\kappa = (b_1/l), \quad (15a)$$

$$\gamma = (b_0 - a_1/l) + b_1(x/l). \quad (15b)$$

An exact evaluation of the strain energies for an element of length $h = 2l$ will now yield the bending and shear strain energy as,

$$U_B = \frac{1}{2}(EI)(2l)\{(b_1/l)\}^2, \quad (16a)$$

$$U_S = \frac{1}{2}(kGA)(2l)\{(b_0 - a_1/l)^2 + \frac{1}{3}b_1^2\}. \quad (16b)$$

It is possible to see from this that in the constraining physical limit of a very thin beam modelled by elements of length $2l$ and depth t , the shear strain energy in (16b) must vanish. An examination of the conditions produced by this requirement shows that the following constraints would emerge in such a limit,

$$b_0 - a_1/l \rightarrow 0, \quad (17a)$$

$$b_1 \rightarrow 0. \quad (17b)$$

Constraint (17a) is field-consistent as it contains constants from both the contributing displacement interpolations relevant to the description of the shear strain field. These

constraints can then accommodate the true Kirchhoff constraints in a physically meaningful way, i.e. in an infinitesimal sense this is equal to the condition $(\theta - w_{,x}) \rightarrow 0$ at the element centroid. In direct contrast, constraint (17b) contains only a term from the section rotation θ . A constraint imposed on this will lead to an undesired restriction on θ . In an infinitesimal sense, this is equal to the condition $\theta_{,x} \rightarrow 0$ at the element centroid (i.e. no bending is allowed to develop in the element region). This is the 'spurious constraint' that leads to shear locking and violent disturbances in the shear force prediction over the element, as we shall see presently.

7.1a *An error model for the field-consistency paradigm:* We must now determine that this field-consistency paradigm leads us to an accurate error prediction. We know that the discretized finite element model contains an error which can be recognized when digital computations made with these elements are compared with analytical solutions, where available. The consistency requirement has been offered as the missing paradigm for the error-free formulation of the constrained media problems. We must now devise an operational procedure that will trace the errors due to an inconsistent representation of the constrained strain field and obtain precise *a priori* measures for these. We must then show by actual numerical experiments with the original elements that the errors are as projected by these *a priori* error models. Only such an exercise will complete the scientific validation of the consistency paradigm. Fortunately, a procedure we shall call the *functional re-constitution* technique makes it possible to do this verification.

7.1b *Functional re-constitution:* We have postulated that the error of shear locking originates from the spurious shear constraint in (17b). We must now devise an error model for the case where the inconsistent element is used to model a beam of length L and depth t . The strain energy for such a beam can be set up as,

$$\Pi = \int_0^L \left\{ \frac{1}{2} EI \theta_{,x}^2 + \frac{1}{2} kGA (\theta - w_{,x})^2 \right\} dx. \quad (18)$$

If an element of length $2l$ is isolated, the discretization process produces an energy for the element of the form given in (16). In this equation, the constants which were introduced due to the discretization process can be replaced by the continuum (i.e. the infinitesimal) description. Thus, we note that in each element, the constants in (16a) and (16b) can be traced to the constants in (14a) and (14b) and can be replaced by the values of the field variations θ , $\theta_{,x}$ and $w_{,x}$ at the centroid of the element. Thus, the strain energy of deformation in an element is,

$$\pi_e = \frac{1}{2} (EI)(2l)(\theta_{,x})^2 + \frac{1}{2} (kGA)(2l)(\theta - w_{,x})^2 + \frac{1}{6} (kGAL^2)(\theta_{,x})^2. \quad (19)$$

Thus the constants in the discretized strain energy functional have been re-constituted into an equivalent continuum or infinitesimal form. From this re-constituted functional, we can argue that an idealization of a beam region of length $2l$ into a linear displacement type finite element would produce a modified strain energy density within that region of

$$\bar{\pi}_e = \frac{1}{2} (EI + (kGAL^2/3)(\theta_{,x})^2) + \frac{1}{2} (kGA)(\theta - w_{,x})^2. \quad (20)$$

This strain energy density indicates that the original physical system has been altered due to the presence of the inconsistent term in the shear strain field. Thus, we can postulate that a beam of length L modelled by equal elements of length $2l$ will have a re-constituted functional

$$\bar{\Pi} = \int_0^L \left\{ \frac{1}{2}(EI + (kGA l^2/3)(\theta_{,x})^2 + \frac{1}{2}(kGA)(\theta - w_{,x})^2 \right\} dx. \quad (21)$$

We now understand that the discretized beam is stiffer in bending (i.e. greater flexural rigidity) by the factor $kGA l^2/3EI$. For a thin beam, this can be very large, and produces the additional stiffening effect described as *shear locking*.

7.1c *Numerical experiments to verify error prediction:* Our functional re-constitution procedure (note that this is an auxiliary procedure, distinct from the direct finite element procedure that yields the stiffness matrix) allows us to critically examine the consistency paradigm. It indicates that an exactly-integrated or field-inconsistent finite element model tends to behave as a shear flexible beam with a much stiffened flexural rigidity I' . This can be related to the original rigidity I of the system by comparing the expressions in (18) and (21) as,

$$I'/I = 1 + kGA l^2/3EI. \quad (22)$$

We must now show through a numerical experiment that this estimate for the error, which has been established entirely *a priori*, starting from the consistency paradigm and introducing the functional re-constitution technique, anticipates very accurately, the behaviour of a field-inconsistent linearly interpolated shear flexible element in an actual digital computation. Exact solutions are available for the static deflection W of a Timoshenko cantilever beam of length L and depth t under a vertical tip load. If W_{FEM} is the result from a numerical experiment involving a finite element digital computation using elements of length $2l$, the additional stiffening can be described by a parameter as

$$e_{\text{FEM}} = W/W_{\text{FEM}} - 1. \quad (23)$$

From (22), we already have an *a priori* prediction for this factor as

$$e = I'/I - 1 = kGA l^2/3EI. \quad (24)$$

We can now re-interpret the results shown in table 2 for the thin beam case. Using (23) and (24), we can argue *a priori* that the inconsistent element will produce normalized tip deflections ($W_{\text{FEM}}/W = 1/(1 + e)$). Since $e \gg 1$, we have

$$W_{\text{FEM}}/W = (N^2/5) \times 10^{-5}, \quad (25)$$

for the thin beam. Table 3 shows how the predictions made, thus compare with the results obtained from an actual finite element computation using the field-inconsistent (FI) element.

This has shown us that the consistency paradigm can be scientifically verified. Traditional procedures such as counting constraint indices, or computing the rank or condition

Table 3. Normalised tip deflections for the thin beam computed from FEM model and predicted from error model, (25).

No. of elements	Computed (FEM)	Predicted
1	0.200×10^{-5}	0.200×10^{-5}
2	0.800×10^{-5}	0.800×10^{-5}
4	0.320×10^{-4}	0.320×10^{-4}
8	0.128×10^{-3}	0.128×10^{-3}
16	0.512×10^{-3}	0.512×10^{-3}

number of the stiffness matrices could offer only a heuristic picture of how and why locking sets in.

It will be instructive to note here that conventional error analysis norms in the finite element method are based on the percentage error or equivalent in some computed value as compared to the theoretically predicted value. We have seen now that the error of shear locking can be exaggerated without limit as the structural parameter that acts as a penalty multiplier becomes indefinitely large. The percentage error norms therefore saturate quickly to a value approaching 100% and do not sensibly reflect the relationship between error and the structural parameter even on a logarithmic plot. A new error norm called the additional stiffening parameter, e can be introduced to recognize the manner in which the errors of the locking kind can be blown out of proportion by a large variation in the structural parameter. Essentially, this takes into account the fact that the spurious constraints give rise to a spurious energy term and consequently alters the rigidity of the system being modelled. In many other examples (e.g. Mindlin plates, curved beams etc.), it was seen that the rigidity, I , of the field consistent system and the rigidity, I' , of the inconsistent system, were related to the structural parameters in the form, $I'/I = \alpha(l/t)^2$ where l is an element dimension and t is the element thickness. Thus, if w is the deflection of a reference point as predicted by an analytical solution to the theoretical description of the problem and w_{FEM} is the deflection predicted by a field inconsistent finite element model, we would expect the relationship described by (24). A logarithmic plot of the new error norm against the parameter (l/t) will show a quadratic relationship that will continue indefinitely as (l/t) is increased. This was found to be true of the many constrained media problems. By way of illustration of the distinction made by this definition, we shall anticipate again the results above. If we represent the conventional error norm in the form $E = (W - W_{FEM})/W$, and plot both E and the new error norm e from the results for the same problem using 4 field inconsistent elements against the penalty multiplier (l/t) on a logarithmic scale, the dependence is as shown in figure 5. It can be seen that E saturates quickly to a value approaching 100% and cannot meaningfully show how the error propagates as the penalty multiplier increases indefinitely. On the other hand, e captures this relationship very accurately.

8. Concluding remarks

We have attempted to show here that if the first principles are carefully identified, it is possible to set up error estimates on an *a priori* basis. This would enhance the validation

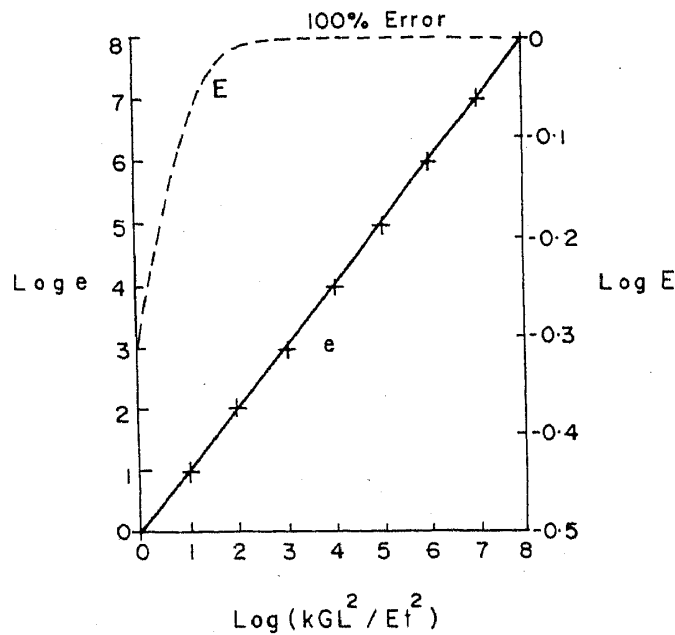


Figure 5. Variation of error norms e , E with structural parameter kGL^2/Et^2 for a cantilever beam under tip shear force.

of new design of finite elements so that not only are computed results compared with analytical results, it is also possible to give a measure to the quality of the solution by assigning rates of convergence and order of error estimates which have been arrived at from first principles using simple engineering analysis.

The author is very grateful to many colleagues, too numerous to mention, who have at some time or the other worked on the computational aspect of many of these problems. He thanks Dr T S Prahlad for his support. The concepts reviewed here were developed partly under the support of grants-in-aid schemes of the Aeronautics Research and Development Board of the Ministry of Defence and this financial support is gratefully acknowledged.

References

- Hu H C 1955 On some variational principles in the theory of elasticity and the theory of plasticity. *Sci. Sin.* 4: 33-54
- Prathap G 1993 *The finite element method in structural mechanics* (Dordrecht: Kluwer Academic Press)
- Prathap G 1996 Finite element analysis and the stress correspondence paradigm. *Sādhanā* 21: 525-546