AFFDL-TR-68-150

COMPLETENESS AND CONVERGENCE IN THE FINITE ELEMENT METHOD

Eduardo Romano de Arantes e Oliveira*

Laboratório Nacional De Engenharia Civil, Lisboa, Portugal

A study is made of the convergence to the exact solution of sequences of approximate solutions generated by elements with decreasing size. A general completeness criterion is justified which is shown to become also a convergence criterion if the second order derivative of the displacements corresponding to the successive approximate solutions remain bounded within each element, as the size decreases indefinitely. The conclusions are valid for all the different linear continuous structural theories, i.e., for two and three-dimensional Elasticity, as well as for beams, shells and plates. If the simplified theories which result from neglecting the transverse shear deformation are considered, the third derivatives of the transverse displacement must also remain bounded within each element, in order that convergence can be obtained.

* Head, Applied Mathematics Division

SECTION I

INTRODUCTION

The problem of convergence to the exact solution of a sequence of approximate solutions generated by patterns of finite elements decreasing in size is the main concern of this paper.

In the Ritz method (Reference 1), completeness is a sufficient condition for convergence. It is important to notice that the finite element method is a particularization of Ritz method only if compatiblity between elements is achieved (Reference 2). Convergence is however still possible if compatibility is violated. Completeness and convergence criteria are justified in the present paper which are valid even if compatibility between elements is not obtained. The different kinds of structures are simultaneously considered but a special Section will be dedicated to the simplified theories which result from neglecting the transverse shear deformation.

A general discussion of the subject was already presented in a former paper by the author (Reference 2). The finite element method, however, was considered as a general mathematical technique, while the present paper is only concerned with elastic problems. Such particularization makes the discussion less abstract and thus easier to follow.

Some other papers (References 3, 4, and 5) have been written which examine convergence in connection with particular kinds of elements. No general convergence criteria have however been presented or justified in those papers.

SECTION II

A SYNTHETIC FORMULATION OF ELASTICITY

Elastic theories involve three kinds of magnitudes, stresses, strains and displacements, whose vectors will be denoted by $\boldsymbol{\sigma}$, $\boldsymbol{\epsilon}$ and $\mathbf{u}$.

Such magnitudes are defined on a domain, D, corresponding to the body, and related by three kinds of field equations which can be symbolized as follows:

a) Equilibrium equations:

$$\mathbf{E}\boldsymbol{\sigma} = \mathbf{X} \tag{1}$$

b) Strain-displacement relations:

$$\boldsymbol{\epsilon} = \mathbf{D}\mathbf{u} \tag{2}$$

c) Stress-strain relations:

$$\boldsymbol{\sigma} = \mathbf{H}\boldsymbol{\epsilon} \tag{3}$$

$\mathbf{E}$ and $\mathbf{D}$ are first order differential operators, $\mathbf{X}$ is the vector of the body force density components, $\mathbf{H}$ is a symmetric positive definite matrix.

Equations 1, 2 and 3 are valid on D. On the boundary, B, the equilibrium equations become

$$\mathbf{N}\boldsymbol{\sigma} = \mathbf{p} \tag{4}$$

$\mathbf{N}$ is a matrix whose elements depend on the orientation of the normal vector at each boundary point. $\mathbf{p}$ is the vector of the tractions applied to the boundary.

The elastic analysis reduces to finding the solution of the system of field Equations 1, 2 and 3 which satisfies certain boundary conditions. The simplest and most important types of boundary conditions can be expressed directly in terms of displacements or tractions applied to the boundary. Let B_1 and B_2 denote the portions of the boundary where tractions or displacements are respectively prescribed. The boundary conditions are analytically expressed by

$$\mathbf{p} = \overline{\mathbf{p}} \quad \text{in } B_1 \tag{5}$$

$$\mathbf{u} = \overline{\mathbf{u}} \quad \text{in } B_2 \tag{6}$$

The stresses and strains can be eliminated and the problem can be formulated in terms of displacements only. Indeed, introducing Equations 2 and 3 into 1, there results

$$\mathbf{K}\ \mathbf{u} = \mathbf{X} \tag{7}$$

in which $\mathbf{K}$ is the second-order differential operator

$$\mathbf{K} = \mathbf{E}\ \mathbf{H}\ \mathbf{D} \tag{8}$$

The boundary conditions are also easily expressible in terms of displacements or displacement derivatives.

Any set of fields, $\boldsymbol{\sigma}$, $\boldsymbol{\epsilon}$, $\mathbf{u}$, defines an elastic field.

An elastic field is said to be compatible if Equations 2 and 3 and boundary conditions Equation 6 are obeyed. An elastic field is said to be equilibrated, with respect to a certain set of external forces, if it respects Equations 1, 3 and 4 and boundary conditions Equation 5. The exact solution is the elastic field which is simultaneously a compatible and an equilibrated one.

Assume now that $\mathbf{u}$ and $\boldsymbol{\sigma}$ are such that the derivatives involved in $\mathbf{D}$ and $\mathbf{E}$ exist everywhere in D. Then, the following relation holds

$$\int_D \boldsymbol{\sigma}^T (\mathbf{D}\,\mathbf{u})\ dD = \int_D (\mathbf{E}\boldsymbol{\sigma})^T \mathbf{u}\ dD + \int_B (\mathbf{N}\boldsymbol{\sigma})^T \mathbf{u}\ dB \tag{9}$$

In this relation, vectors $\boldsymbol{\sigma}$ and $\mathbf{u}$ are not necessarily related by the stress-strain relations, Equation 3.

The theorem of the minimum total potential energy, which is a consequence of Equation 9, states that the exact solution minimizes the functional

$$F = U - \int_D \overline{\mathbf{X}}^T \mathbf{u}\ dD - \int_B \overline{\mathbf{p}}^T \mathbf{u}\ dB \tag{10}$$

i.e., the total potential energy, in any class of compatible elastic fields which contains the exact solution.

The first term in F,

$$U = \int_D W\ dD = \frac{1}{2}\int_D \boldsymbol{\epsilon}^T \mathbf{H}\,\boldsymbol{\epsilon}\ dD = \frac{1}{2}\int_D (\mathbf{D}\,\mathbf{u})^T \mathbf{H}\,(\mathbf{D}\,\mathbf{u})\ dD \tag{11}$$

is the strain energy.

The upper dash on $\bar{X}$ and $\bar{p}$ indicates prescribed magnitudes.

The formulation which has been presented is quite general as it is valid not only for linear two- and three-dimensional Elasticity, but also for linear theories of plates, shells and beams.

In the case of a plate, for instance, vector $\mathbf{u}$ contains the transverse displacement and two rotations, vector $\boldsymbol{\epsilon}$ contains the curvatures and the transverse shear strains, vector $\boldsymbol{\sigma}$ contains the bending and twisting moments and the transverse shearing forces.

A very frequent simplification in the analysis of plates, shells and beams consists in neglecting the transverse shear deformation. This makes it possible to reduce the number of unknowns to one (the normal displacement) in the theory of plates, and three (the normal displacement and the tangential displacements) in the theory of thin shells.

Such simplified theories present the same scheme as the initial ones. Operators $\mathbf{E}$ and $\mathbf{D}$ are now however of the second order and operator $\mathbf{K}$ of the fourth order. The compatibility conditions involve the continuity of both the displacements and their first derivatives. As the rotations are expressed in terms of such magnitudes, the compatibility conditions are essentially the same, both in the simplified theories and in the corresponding initial ones where the transverse shear deformation is not neglected.

In the following discussion, the initial theories will first be considered. Operators $\mathbf{E}$ and $\mathbf{D}$ will thus be assumed of the first order, and continuity of displacements will be the only condition for compatibility. Then, at the end of the paper, the results established for the initial theories will be used in the discussion of the simplified theories.

SECTION III
THE FINITE ELEMENT METHOD

The finite element method is a general discretization technique. In this method, domain D is assumed to be decomposed into a finite number of subdomains D^e and families of fields are considered which have different analytical expressions within each subdomain.

A finite element is a closed subdomain, D^e , together with the family of elastic fields which are allowed to occur within it. Such family is a linear combination with coefficients q_i^e of a finite number of unit modes, so that each field of the family corresponds to ascribing particular values to the parameters q_i^e.

The values of the displacements at a certain number of points on the boundary of the element, called nodes or nodal points, are, as a rule, chosen as parameters. The allowed fields need not be introduced however by giving the expression of the displacements directly in terms of their own nodal values. They can indeed be given in terms of equal number of arbitrary parameters which in turn can be expressed in terms of those nodal values.

The type of an element refers to its general shape, nodal point specification and to the allowed fields, analytically defined by expressing the displacement vector $\mathbf{u}^e$ in terms of the parameters (generalized displacements) and the coordinates:

$$\mathbf{u}^e = \boldsymbol{\phi}^e (x_1, x_2, \cdots) \, \mathbf{q}^e \tag{12}$$

$\mathbf{q}^e$ is the vector of the nodal displacements q_i^e.

The elements ϕ_{ij}^e of $\boldsymbol{\phi}^e$ are supposed to be continuous in the closed subdomain occupied by the finite element. The unit modes are defined by the columns of $\boldsymbol{\phi}^e$

Each displacement component is assumed to depend only on its own nodal values. Thus, if q_j^e corresponds to the component u_i^e , all the magnitudes ϕ_{kj}^e, for which $k \neq i$, will vanish.

In order that Equation 12 can be homogeneous, q_j^e must be amenable to a nondimensional form

$$\phi_{ij}^e (x_1, x_2, \cdots) = \psi_{ij}^e \left(\frac{x_1}{\ell^e}, \frac{x_2}{\ell^e}, \cdots \right) \tag{13}$$

in which ℓ^e is a typical dimension of the element, for instance its maximum diameter. ψ^e_{ij} is a function which does not depend on the absolute dimensions of the element.

The different finite elements are made compatible through the specification of reduced compatibility conditions. These require that the values of the displacements be the same at coincident nodes of adjacent elements and equal the prescribed ones at the nodes located on B_2. For the sake of simplicity, such prescribed displacements are assumed to vanish.

A point of the domain is said to be a node of the system if it is a node for one or more elements.

Let $\mathbf{q}_n$ be the vector of the values of the displacements at every node of the system but those which are located on B_2. The reduced compatibility conditions can be expressed by writing for each element

$$\mathbf{q}^e = \mathbf{T}^e \mathbf{q}_n \tag{14}$$

where matrix $\mathbf{T}^e$ depends on the topology of the system.

Equations 12 and 14 show that the knowledge of $\mathbf{q}_n$ is enough for the definition of the field within every element of the system.

Introducing 14 in 12, there results

$$\mathbf{u}^e = \mathbf{\Phi}^e \mathbf{q}_n \tag{15}$$

in which

$$\mathbf{\Phi}^e = \boldsymbol{\phi}^e \mathbf{T}^e \tag{16}$$

Equation 15 is valid in D^e. The whole set of equations (15) (one for each element) piecewise defines a family C_n of elastic fields each of which corresponds to a certain vector $\mathbf{q}_n$. Index n refers to a certain degree of subdivision of D into subdomains.

The reduced compatibility conditions are generally not sufficient to make the displacement components continuous across the element boundaries. This depends on the type of the element.

If the type is such that the reduced compatibility conditions are sufficient to ensure continuity of the displacements across the element boundaries, the elements are said to be conforming. If continuity is violated, the elements are termed nonconforming.

The approximate solution which the finite element method provides for the elastic problem is determined by making the functional

$$F_n = \sum_e U^e - \int_D \bar{X}^T u \, dD - \int_{B_1} \bar{p}^T u \, dB \tag{17}$$

stationary in the class C_n of the elastic fields piecewise defined by Equation 15.

U^e is the strain energy of the element e

$$U^e = \frac{1}{2} \int_{D^e} (D u^e)^T H (D u^e) \, dD \tag{18}$$

If the continuity of the displacements across the element boundaries is achieved, the first term in F_n becomes the strain energy of the body and F_n becomes the total potential energy. The finite element method is then nothing else but a particularization of Ritz's method, characterized by the piecewise definition of the field.

The expression of F_n can be modified to take into account external forces distributed on the element interfaces. F_n becomes then

$$F_n = \sum_e U^e - \int_D \bar{X}^T u \, dD - \sum_e \int_{B_1^e} \bar{p}^T u \, dB$$

$$= \sum_e \left(U^e - \int_{D^e} \bar{X}^T u \, dD - \int_{B_1^e} \bar{p}^T u \, dB \right) \tag{19}$$

B_1^e denotes the portion of the boundary of element e where prescribed external forces are applied. No reactive forces are to be considered.

In order that the last term of Equation 19 can be calculated, the external forces distributed on the interfaces must be shared between the pair of elements which contact along each interface. We assume those forces to be equally distributed between the elements in contact.* The magnitude of $\bar{p}$ to be assigned to each element is thus half of the distribution density of the force applied to the interface.

* This assumption is necessary only if conformity is not achieved.

Introducing Equation 12 in Equation 19, we obtain

$$F_n = \frac{1}{2} \sum_e \left[q^{e^T} K^e \; q^e \; - 2 q^{e^T} \; \overline{Q}^e \right] \tag{20}$$

where

$$K^e = \int_{D^e} (D \, \phi^e)^T \; H \; (D \phi^e) \; dD \tag{21}$$

$$\overline{Q}^e = \int_{D^e} \phi^{e^T} \; \overline{X} \; dD + \int_{B_1^e} \phi^{e^T} \; \overline{p} \; dB \tag{22}$$

K^e is the stiffness matrix of the element. $\overline{Q}^e$ is the vector of the external generalized forces applied to the element.

Introducing Equation 14, we obtain

$$F_n = \frac{1}{2} \left[q_n^T (\sum_e T^{e^T} \; K^e \; T^e) \, q_n - 2 q_n^T \; (\sum_e T^{e^T} \overline{Q}^e) \right] \tag{23}$$

Making

$$K_n = \sum_e T^{e^T} \; K^e \; T^e = \sum_e \int_{D^e} (D \Phi^e)^T \; H \; (D \Phi^e) \, dD \tag{24}$$

$$\overline{Q}_n = \sum_e T^{e^T} \; \overline{Q}^e = \sum_e (\int_{D^e} \Phi^{e^T} \overline{X} \, dD + \int_{B_1^e} \Phi^{e^T} \overline{p} \, dB) \tag{25}$$

there results

$$F_n = \frac{1}{2} q_n^T \; K_n \; q_n - q_n^T \; \overline{Q}_n \tag{26}$$

The symmetric matrix K_n is called the stiffness matrix of the structure. $\overline{Q}_n$ is the vector of the external generalized forces applied to the structure.

The stationary conditions for F_n are obtained by equating to zero the derivatives of F_n with respect to the mutually independent parameters q_{ni}. It results in the system of linear equations

$$K_n \; q_n = \overline{Q}_n \tag{27}$$

Equation 27 permits the determination of the displacements whenever the structure is stable, i.e., whenever the prescribed generalized displacements are enough to prevent rigid body motion. Then, indeed, the columns of $\left[D\,\Phi^e \right]$ are linearly independent of D and, as H is definite positive, K_n is nonsingular and also definite positive.

Functional F_n can be expressed as

$$F_n = \frac{1}{2}\, q_n^T\, K_n\, q_n - q_n^T\, K_n\, q_{an} = \frac{1}{2}\left[(q_n - q_{an})^T K_n (q_n - q_{an}) - q_{an}^T\, K_n\, q_{an} \right] \tag{28}$$

in which q_{an} is the solution of Equation 27. q_{an} corresponds thus to the approximate solution.

As K_n is definite positive, the first term in Equation 28 is positive unless q_n equals q_{an}. This proves that the approximate solution minimizes F_n in C_n.

By virtue of Equation 27, $\sum_e U^e$ can be expressed in terms of the external generalized forces applied to the structure. Indeed,

$$\sum_e U^e = \frac{1}{2} \sum_e q^{e^T} K^e q^e = \frac{1}{2}\, q_n^T\, K_n\, q_n = \frac{1}{2}\, q_n^T\, \overline{Q}_n \tag{29}$$

in which $\overline{Q}_n$ is the vector of the external force corresponding to .

The same magnitude can also be expressed in terms of the external generalized forces applied to the elements. Indeed, introducing Equation 25 in Equation 29, there results

$$\sum_e U^e = \frac{1}{2}\, q_n^T \left(\sum_e T^{e^T} \overline{Q}^e \right) = \frac{1}{2} \sum_e \left(T^e q_n \right) \overline{Q}^e = \frac{1}{2} \sum_e q^e\, \overline{Q}^e \tag{30}$$

Equation 27 makes it possible to analyse the structure by the displacement method.

The force method may also be used. It starts from the reduced equilibrium conditions, which are expressed in terms of the vector

$$Q^e = K^e\, q^e \tag{31}$$

of the total generalized forces acting on each element. Multiplying both sides of Equation 31 by T^{e^T} and summing, there results

$$\sum_e T^{e^T} Q^e = \sum_e T^{e^T} K^e q^e \tag{32}$$

Introducing Equation 14, we obtain

$$\sum_e T^{e^T} Q^e = \sum_e T^{e^T} K^e T^e q_n = K_n q_n \tag{33}$$

and, finally, using Equation 27,

$$\sum_e T^{e^T} Q^e = \overline{Q}_n \tag{34}$$

Equation 34 expresses the reduced equilibrium conditions. It must not be confused with Equation 25. Q^e includes indeed not only the generalized external forces acting on the element (which are the only ones contained in $\overline{Q}^e$) but also the interaction forces between elements and the generalized reactions corresponding to the prescribed displacements.

SECTION IV
COMPLETENESS CRITERION

Completeness of a sequence of families C_n with respect to a given set C has a meaning provided we can compute the distance between any element u_n of each family and any element c of C. The sequence of families is said to be complete with respect to C if it is possible to find, for a specified $\epsilon>0$, an integer N, such that in each family with order $n>N$ there exists an element u_{cn} which satisfies the inequality

$$d(u_{cn}, c) < \epsilon \tag{35}$$

where c is any element belonging to C.

The following definition of distance is chosen as a basis for our discussion:

$$d(u_n, c) = \sqrt{\sum_e \int_{D^e} \left[D \; (u^e - u_c) \right]^T H \left[D(u^e - u_c) \right] dD} \tag{36}$$

in which u^e and u_c denote, within D^e, the displacement fields corresponding to u_n and c.

The expression under the square root is nothing else than the double of the sum of the strain energies corresponding to the difference of the two fields u_n and c. Such expression is never negative and cannot vanish unless the two fields coincide(*). It would be easy to show that other requirements are also satisfied which make Expression 36 a proper definition of distance.

A general criterion for completeness will be stated and justified in this section. Such criterion essentially does not differ from the one which was presented by Baseley et al. (Reference 5).

Let C be the set of the compatible elastic fields whose displacements have continuous and bounded second order derivatives within each element.

We wish to demonstrate that completeness with respect to C will be obtained if

a) the general analytical expression for u_i^e within element e is a polynomial with a number of arbitrary parameters equal to the number of unit modes corresponding to the element,

b) The terms of degree higher than the first can vanish regardless of the values taken by the constant term and the coefficients which affect the linear terms,

c) the constant term and the coefficients which affect the linear terms are completely arbitrary.

We remark that, once these conditions are respected, the displacement component u_i^e or its first derivatives can take up any arbitrary value throughout the element, if suitable values are ascribed to the parameters.

(*) The structure is assumed to be stable, i.e., such that the prescribed displacements are enough to prevent rigid body motion.

The displacements u_i of any field c belonging to C can be represented inside D^e by the following Taylor's expansion

$$u_i = u_i(O) + u_{i,j}(O)(x_j - x_j^o) + \frac{1}{2!} u_{i,jk}(O_i)(x_j - x_j^o)(x_k - x_k^o) \tag{37}$$

O and O_i are points in D^e. O_i depends on the coordinates of the point where u_i is to be determined.

Let us consider now the displacement field with components

$$u_{ti}^e = u_i(O) + u_{i,j}(O)(x_j - x_j^o) \tag{38}$$

within D^e, which we call tangent field to u at O.

As all the second-order derivatives are supposed to be bounded inside D^e, Equations 37 and 38 yield

$$\left| u_i - u_{ti}^e \right| < \frac{d}{2!} V_1 (\ell^e)^2 \tag{39}$$

in which V_1 is an upper bound for all the second derivatives and ℓ^e is the maximum diameter of element e. d is the total number of second order derivatives.

By considering similar expansions for the first derivatives of u_i, it is also possible to derive the inequality

$$\left| u_{i,j} - u_{ti,j}^e \right| < d V_1 \ell^e \tag{40}$$

As operator **D** involves derivatives of the first order, we have

$$\left[\mathbf{D}(u - u_t^e) \right]^T H \left[\mathbf{D}(u - u_t^e) \right] < V_2^e (\ell^e)^2 \tag{41}$$

for ℓ^e sufficiently small. V_2^e is a positive number.

The distance between the elastic field c, corresponding to **u**, and u_{t_n} corresponding to $\mathbf{u}_t^e$, is thus

$$d(u_t, c) = \sqrt{\sum_e \int_{D^e} \left[\mathbf{D}(u - u_t^e) \right]^T H \left[\mathbf{D}(u - u_t^e) \right] dD} < \sqrt{\sum_e V_2^e (\ell^e)^2 D^e}$$

$$\leq \sqrt{V_2 \ell_n^2 \sum_e D^e} = \sqrt{V_2 \ell_n^2 D} \tag{42}$$

in which V_2 and ℓ denote the maximum values of V_2^e and ℓ^e in the whole set of elements.

Equation 42 means that the distance between u_t and c tends to zero when n tends to infinity, that is, when the size of elements decreases indefinitely.

Consider now a type of finite element generating a sequence of families of fields, C_n, whose completeness is to be investigated.

Let u_{cn} be the field of C_n whose displacements take, at the nodes, the same values as the displacement of c. u_c^e corresponds to u_{cn} within D^e.

Suppose the general criterion to be satisfied. u_t^e is thus one of the displacement fields which can occur within the finite element. Let such field correspond to values q_{ti}^e of the parameters, i.e.,

$$u_t^e = \phi^e \; q_t^e \quad \text{within } D^e \tag{43}$$

On the other hand,

$$u_c^e = \phi^e \; q_c^e \tag{44}$$

From Equations 43 and 44, we obtain

$$\left| u_{ti}^e - u_{ci}^e \right| = \left| \phi_{ij}^e \left(q_{tj}^e - q_{cj}^e \right) \right| \quad , \text{ within } D^e \tag{45}$$

On the other hand,

$$\left| u_{ti,k}^e - u_{ci,k}^e \right| = \left| \phi_{ij,k}^e \left(q_{tj}^e - q_{cj}^e \right) \right| \quad \text{within } D^e \tag{46}$$

But, by virtue of Equation 13,

$$\phi_{ij,k}^e = \frac{1}{\ell^e} \frac{\partial}{\partial \left(\frac{x_k}{\ell^e} \right)} \left[\psi_{ij}^e \left(\frac{x_1}{\ell^e}, \cdots \right) \right] = \frac{1}{\ell^e} \chi_{ij}^e \left(\frac{x_1}{\ell^e}, \cdots \right) \tag{47}$$

As the absolute dimensions of the element do not appear explicitly in the functions ψ_{ij}^e and χ_{ij}^e, these functions remain bounded as the size of the element decreases. Assume the moduli of all those magnitudes remain below a positive number, V_3.

Then

$$\left| u^e_{ti} - u^e_{ci} \right| < V_3 \sum_j \left| q^e_{tj} - q^e_{cj} \right| \tag{48}$$

$$\left| u^e_{ti,j} - u^e_{ci,j} \right| < \frac{V_3}{\ell^e} \sum_j \left| q^e_{tj} - q^e_{cj} \right| \tag{49}$$

within D^e.

On the other hand, as u^e_i and u_i take the same values at the nodes, and the parameters q^e_i are the nodal values of the displacements, equation 39 permits to write

$$\left| q^e_{tj} - q^e_{cj} \right| < \frac{d}{2!} V_1 (\ell^e)^2 \tag{50}$$

Introducing Equation 50 into Equations 48 and 49, we obtain

$$\left| u^e_{ti} - u^e_{ci} \right| < \frac{V_1 V_3 N^e d}{2!} (\ell^e)^2 \tag{51}$$

$$\left| u^e_{ti,j} - u^e_{ci,j} \right| < \frac{V_1 V_3 N^e d}{2!} \ell^e \tag{52}$$

N^e being the total number of parameters corresponding to element e.

Equations 51 and 52 hold even if the first derivatives of u^e_{ci} are discontinuous in D^e. This is an important remark because sometimes (Reference 7) the element itself is considered subdivided into parts and the displacement field admits different analytical expressions within each part.

The similarity between Equations 51 and 52 and Equations 39 and 40 allows a jump straight to the inequality

$$d(u_{tn}, u_{cn}) < \sqrt{V_4 \ell_n^2 D} \tag{53}$$

in which u_{cn} represents the piecewise defined elastic field with displacements u^e_i within element e. V_4 is a positive number.

Combining Equations 42 and 53, we obtain finally

$$d(u_{cn}, c) \leq d(u_{cn}, u_{tn}) + d(u_{tn}, c) < \ell_n \sqrt{D} \left(\sqrt{V_4} + \sqrt{V_2} \right) \tag{54}$$

Equation 54 means that the distance between u_{cn} and c tends to zero with the size of the largest element, so that, as c is an arbitrary element of C, the completeness proof is finally achieved.

SECTION V
CONVERGENCE DISCUSSION

Consider any type of finite element which can generate a sequence $\{C_n\}$ of families of fields complete with respect to C. We wish to investigate if the sequence of approximate solutions, $\{u_{an}\}$, obtained by minimizing F_n in each family C_n converges to the exact solution, which is assumed to belong to C.

We know already that completeness implies convergence to the exact solution if it is associated with conformity. The finite element technique becomes then indeed a particularization of Ritz's method, in which conformity is a sufficient condition for convergence.

Let u_{en} be the field in C_n whose displacements take at the nodes the same values as the displacements of the exact solution.

As the exact solution, u_o, belongs to C, and completeness with respect to C is ensured, it is possible to determine N such that, for $n > N$,

$$d(u_o, u_{en}) < \epsilon \tag{55}$$

ϵ being a positive and arbitrarily small number.

As F_n is a continuous functional, we can find ϵ such that

$$F_n(u_{en}) = F_n(u_o) \pm \epsilon' \qquad (56)$$

ϵ' being also positive and arbitrarily small.

As u_{en} belongs to C_n, and u_{an} minimizes F_n in C_n,

$$F_n(u_{an}) \leq F_n(u_{en}) \qquad (57)$$

and

$$F_n(u_{an}) \leq F_n(u_o) + \epsilon' \qquad (58)$$

The approximate solution u_{an} equilibrates a system of external forces which generally differs from the system of external forces actually applied to the body. Let u_a denote the compatible field which equilibrates the same external forces as u_{an}, i.e. the exact solution corresponding to those forces.

We assume that u_a belongs to C, i.e. that the second derivatives of the corresponding displacements are continuous and bounded within each element. This assumption will later be discussed.

Let u_{bn} be the field in C_n whose displacements have the same values at the nodes as the displacement corresponding to u_a. As u_a belongs to C, and the sequence $\{C_n\}$ is complete with respect to C, it is possible to find N_1 such that, for $n > N_1$

$$d(u_a, u_{bn}) < \epsilon'' \qquad (59)$$

ϵ'' being a positive number, arbitrarily small.

We can also find N_1 such that the first derivatives of the displacements (see Equations 40 and 52), and thus the difference of the stresses, are smaller than ϵ'' within each subdomain, that is

$$\left| \boldsymbol{\sigma}_a - \boldsymbol{\sigma}_{bn} \right| < \boldsymbol{\epsilon}'' \qquad (60)$$

$\boldsymbol{\epsilon}''$ being a column vector with all its elements equal to ϵ''.

Let us consider now the generalized forces connected with the systems of external forces respectively corresponding to u_a and u_{bn}.

By virtue of Equation 22,

$$\bar{Q}_b^e - \bar{Q}_a^e = \int_{D^e} \phi^{e^T} (\bar{X}_b^e - \bar{X}_a^e)\, dD + \int_{B_I^e} \phi^{e^T} (\bar{p}_b^e - \bar{p}_a^e)\, dB \tag{61}$$

in which $\bar{X}_b^e$ and $\bar{p}_b^e$ correspond to u_{bn} and $\bar{X}_a^e$ and $\bar{p}_a^e$ correspond to u_a.

The body force term tends to be of higher order than the second term in Equation 59, thus tends to infinity. The difference of the generalized forces tends then to be expressed by

$$\bar{Q}_b^e - \bar{Q}_a^e = \int_{B_I^e} \phi^{e^T} (\bar{p}_b^e - \bar{p}_a^e)\, dB \tag{62}$$

By virtue of the continuity of the stresses within $\bar{D}^e$, Equation 60 is also valid on B^e. Thus,

$$\bar{p}_b^e - \bar{p}_a^e < \epsilon'' \tag{63}$$

and

$$\left| \bar{Q}_b^e - \bar{Q}_a^e \right| < \int_{B_I^e} \phi^{e^T} \epsilon''\, dB \tag{64}$$

On the other hand, by virtue of Equations 18, 30 and 36, the distance between u_{an} and u_{bn} is given by

$$2 \left[d(u_{an}, u_{bn}) \right]^2 = \sum_e (q_b^e - q_a^e)^T (\bar{Q}_b^e - \bar{Q}_a'^e) \tag{65}$$

As the external forces which equilibrate u_a and u_{an} are the same, $\bar{Q}_a'^e$ coincides with $\bar{Q}_a^e$. Thus,

$$d(u_{an}, u_{bn}) < \sqrt{\sum_e \left| q_b^{e^T} - q_a^{e^T} \right| \int_{B_I^e} \left| \phi^{e^T} \right| \epsilon''\, dB} \tag{66}$$

Equation 66 shows that the distance between u_{an} and u_{bn} can be so small as we wish.

Combining 66 with 59, we conclude that it is possible to find N_2 such, for $n > N_2$,

$$d(u_a, u_{an}) < \epsilon^{III} \tag{67}$$

ϵ''' being a positive arbitrarily small number.

As F_n is a continuous functional, it is then possible, given ϵ^{IV}, to determine ϵ''' such that

$$F_n(u_a) = F_n(u_{an}) \pm \epsilon^{IV} \tag{68}$$

As u_a belongs to C, the theorem of the total potential energy permits to write

$$F_n(u_a) \geq F_n(u_o) \tag{69}$$

and thus

$$F_n(u_o) \leq F_n(u_{an}) \pm \epsilon^{IV} \tag{70}$$

Combining Equations 56, 57 and 68, there results

$$F_n(u_{an}) \leq F_n(u_{en}) \leq F_n(u_{an}) \pm \epsilon' \pm \epsilon^{IV} \tag{71}$$

and thus

$$F_n(u_{en}) = F_n(u_{an}) + \epsilon^{V} \tag{72}$$

in which

$$0 < \epsilon^{V} < \epsilon' + \epsilon^{IV} \tag{73}$$

But, as u_{an} and u_{en} both belong to C_n, Equation 28 permits us to write

$$F_n(u_{en}) - F_n(u_{an}) = \frac{1}{2}(q_{en} - q_{an})^T K_n (q_{en} - q_{an}) = \frac{1}{2}\left[d(u_{en}, u_{an})\right]^2 = \epsilon^{V} \tag{74}$$

Combining Equation 74 and 55 we obtain finally

$$d(u_{an}, u_o) \leqslant d(u_{an}, u_{en}) + d(u_{en}, u_o) < \epsilon + \sqrt{2\epsilon^{V}} \tag{75}$$

which shows that $\{u_{an}\}$ converges to u_o.

It remains to investigate if u_a belongs to C.

Our reasoning will be based on a theorem which is known to be valid (Reference 8, p. 349) for linear elliptic equations of the form

$$\alpha_{ij}\, u_{,ij} + \beta_i\, u_{,i} + \gamma u = f \tag{76}$$

Such theorem states that the partial derivatives of order up to (m + 2) of u satisfy a Hölder condition with exponent α $(0<\alpha<1)$ in every bounded subdomain D^e with closure in a closed domain(*) $\overline{G}$, whenever all the derivatives of order up to m of f and the coefficients a_{ij}, b_i and c satisfy the same Hölder condition in $\overline{G}$.

The displacements will thus be Holder continuous, and therefore have continuous bounded second order derivatives in a closed subdomain , whenever the body force density is Hölder continuous in $\overline{D}^e$.

u_a is the compatible field which equilibrates the same external forces as u_{an}. Such forces are of two kinds: body forces distributed within each subdomain, and forces distributed on the subdomain interfaces and on B.

u_a will belong to C, that is, the second-order derivatives of the corresponding displacements will be continuous and bounded within each subdomain D^e, if the body force density corresponding to u_{an} is Hölder continuous within $\overline{D}^e$.

The problem now consists in knowing whether the body force density corresponding to u_{an} is continuous and bounded (and therefore Hölder continuous) within $\overline{D}^e$, no matter how large is n. This is namely the case if the type of the element is such that the body force density vanishes or is obliged to a prescribed bounded and continuous variation within each element, no matter the values of the corresponding generalized displacements.

If such a situation arises, there is no doubt that the completeness criterion is also a convergence criterion.

Our reasoning can be adapted to cases (Reference 7) in which the elements are subdivided into parts and the allowed fields have different analytical expressions within each part. The body force density is then generally not continuous within the element taken as a whole. Convergence will however be proved if each part is treated as a separate element.

(*) An upper dash denotes a domain and its boundary, i.e., a closed domain.

SECTION VI
SIMPLIFIED THEORIES

The results which have been established can be generalized to cover a much broader class of problems (Reference 2). The simplified theories which result from neglecting the transverse shear deformation (see Section 2) belong to such class.

The convergence criterion respecting the simplified theories can however be established in a more direct way, if the results are used which were obtained for the corresponding theories in which the transverse shear deformation is not neglected.

It was proved in the last Section that convergence will be obtained if completeness with respect to C is achieved and if the body force density components remain continuous and bounded within the elements, no matter of their size.

Completeness with respect to C is not however a necessary condition for convergence. Completeness with respect to a subset $C' \subset C$ will indeed be sufficient, if the exact solution belongs to C'.

Assume indeed that all the displacement components of every field in C' verify a certain set of linear relations

$$\mathbf{R}\mathbf{u} = 0 \tag{77}$$

in which $\mathbf{R}$ may be a differential operator.

Consider now polynomial expressions for the displacements, within a general finite element, and assume that the completeness criterion with respect to C is verified. Let C_n be the family of fields piecewise defined by such polynomial expressions made compatible by reduced compatibility conditions.

The introduction of Equation 77 implies a linear relationship between the coefficients of the polynomial expressions which generally involves the constant and linear terms. Such terms cease thus to be arbitrary and completeness with respect to C is therefore destroyed. Completeness with respect to C' will however still be maintained <u>if the relations between the coefficients are only those which result from Equation 77</u>. Let C'_n be the family of fields piecewise defined by polynomial expressions obeying such requirement and made compatible by the same reduced compatibility conditions as the fields of C_n.

The total number of independent coefficients corresponding to C_n is larger than the total number corresponding to C'_n. As the number of independent coefficients equals the number of nodal displacements, C_n corresponds to a larger number of nodes and nodal displacements than C'_n and, therefore, to a stiffness matrix $\mathbf{K}_n$ of larger order than the stiffness matrix $\mathbf{K}'_n$ corresponding to C'_n. Let $\mathbf{q}_n$ and $\mathbf{q}'_n$ be the vectors of nodal displacements corresponding to C_n and C'_n.

As the reduced compatibility conditions were supposed the same, both for C_n and C'_n, C'_n is a subset of C_n. The reduced compatibility conditions are assumed such that the rigid body motion of each single element is prevented, so that matrices $\mathbf{K}_n$ and $\mathbf{K}'_n$ are both definite positive.

To each field in C'_n corresponds a displacement vector $\mathbf{q}_n$, as well as a displacement vector $\mathbf{q}'_n$. The elements of vector $\mathbf{q}_n$ may be deduced from the corresponding vector $\mathbf{q}'_n$ by using the polynomial expressions of the displacements to determine the supplementary nodal displacements.

Let $\mathbf{q}_{an}$ be the displacement vector which minimizes F_n in C_n.

The expression of F_n for the fields of C_n may be given, according to Equation 28, by

$$F_n = \frac{1}{2} (\mathbf{q}_n - \mathbf{q}_{an})^T \mathbf{K}_n (\mathbf{q}_n - \mathbf{q}_{an}) - \frac{1}{2} \mathbf{q}_{an}^T \mathbf{K}_n \mathbf{q}_{an} \tag{78}$$

As C'_n is contained in C_n, the expression of F_n for the fields of C'_n can also be given by Equation 78.

F_n will be a minimum in C'_n if the first term in Equation 78 is a minimum. But such first term represents half of the square of the distance between the field corresponding to $\mathbf{q}_n$ and the field u_{an} corresponding to $\mathbf{q}_{an}$. The field which minimizes F_n in C'_n is thus the one in C'_n whose distance to u_{an} is a minimum.

Consider the sequence $\{u_{an}\}$ of the approximate solutions obtained by minimizing F_n in each family C_n. If the body force density components remain bounded within each element, no matter the size, the sequence $\{u_{an}\}$ converges to the exact solution.

But the exact solution was supposed to belong to C' and, since sequence $\{C'_n\}$ is complete with respect to C', there exists a field $u''_{an} \subset C'_n$ whose distance to the exact solution tends to zero when n tends to infinity.

The distance between u_{an} and the exact solution also tends to zero. This means that the distance between u_{an} and u''_{an} tends to zero.

As the distance between u_{an} and the field u'_{an} which minimizes F_n in C'_n is the minimum distance between u_{an} and any field in C'_n, it follows that the distance between u_{an} and u'_{an} cannot be larger than the distance between u_{an} and u''_{an}. Therefore, u'_{an} tends to u_{an}, and thus to the exact solution.

Completeness with respect to a set $C' \subset C$ which contains the exact solution, and the continuity and boundedness of the body force density components, were thus proved to be sufficient conditions for convergence.

It is now easy to establish convergence criteria for the simplified theories. The reasoning will be exemplified with the theory of plates (Reference 9), in which the displacements are the normal displacement, , and two rotations, and . We assume such magnitudes referred to a system of cartesian orthogonal coordinates, x_1, x_2, x_3. The middle plane of the plate corresponds to $x_3 = 0$.

The equilibrium equations are

$$N_{\alpha 3,\alpha} + p_3 = 0 \tag{79}$$

$$M_{\alpha\beta,\alpha} - N_{\beta 3} + m_\beta = 0 \tag{80}$$

and the strain-displacement relations are

$$e_{\alpha 3} = U_{3,\alpha} + I_\alpha \tag{81}$$

$$K_{\alpha\beta} = I_{\beta,\alpha} \tag{82}$$

In these equations, moments are denoted by $M_{\alpha\beta}$, transverse forces by $N_{\alpha 3}$, transverse shear deformations by $e_{\alpha 3}$ and curvatures by $K_{\alpha\beta}$. Greek indices can take up the numerical values 1 and 2.

Assume the stress-strain relations to be of the form

$$N_{\alpha 3} = H_{\alpha\beta}\, e_{\beta 3} \tag{83}$$

$$M_{\alpha\beta} = H_{\alpha\beta\gamma\delta}\, K_{\gamma\delta} \tag{84}$$

The equilibrium equations expressed in terms of displacement are then

$$H_{\alpha\beta}\,(U_{3,\beta\alpha} + I_{\beta,\alpha}) + p_3 = 0 \tag{85}$$

$$H_{\alpha\beta\gamma\delta}\, I_{\delta,\alpha\gamma} - H_{\beta\gamma}\,(U_{3,\gamma} + I_{\gamma}) + m_{\beta} = 0 \tag{86}$$

if the elastic coefficients $H_{\alpha\beta}$ and $H_{\alpha\beta\gamma\delta}$ are constant all over the domain.

Assume first that the transverse shear deformation is not neglected. The results obtained in Section V are then directly applicable. This means that convergence will be obtained if completeness with respect to C is achieved and if the body force density components remain continuous and bounded within the elements, whatever their size.

We remember that the body force density components, that is, the elements of vector **X** , are in this case both the applied force and moment distribution densities, p_3 , m_1 and m_2. Equations 85 and 86 show that such magnitudes will be continuous and bounded if the displacements (U_3, I_1 and I_2) together with their first and second derivatives, are continuous and bounded.

Completeness with respect to C will be obtained if the displacements are expressed within each element by

$$I_1 = a_{10} + a_{11}\,x_1 + a_{12}\,x_2 + P_1^{(2)} \tag{87}$$

$$I_2 = a_{20} + a_{21}\,x_1 + a_{22}\,x_2 + P_2^{(2)} \tag{88}$$

$$U_3 = a_{30} + a_{31}\,x_1 + a_{32}\,x_2 + P_3^{(2)} \tag{89}$$

in which $P_i^{(2)}$ denotes a polynomial whose terms are at least of the second degree. The coefficients a_{ij} must be arbitrary and the polynomials $P_i^{(2)}$ can vanish for any value of them.

Assume now that the transverse shear deformation is negligible, that is,

$$e_{\alpha 3} = 0 \tag{90}$$

This is exactly so if the elastic coefficients $H_{\alpha\beta}$ are supposed unbounded.

Introducing Equation 90 into Equation 81 there results

$$I_{\alpha} = -U_{3,\alpha} \tag{91}$$

which is a particularization of Equation 77.

Now let

$$P_3^{(2)} = a_{33} x_1^2 + a_{34} x_1 x_2 + a_{35} x_2^2 + P_3^{(3)} \tag{92}$$

Introducing Equations 89 and 92 into Equation 91, there results

$$I_1 = -a_{31} - 2a_{33} x_1 - a_{34} x_2 - P_{3,1}^{(3)} \tag{93}$$

$$I_2 = -a_{32} - a_{34} x_1 - 2a_{35} x_2 - P_{3,2}^{(3)} \tag{94}$$

Comparing Equations 87 and 88 with Equation 93, it becomes clear that a set of linear relations has been introduced between the coefficients a_{ij}.

Completeness is still achieved, however, with respect to the subset C' of the fields of C with vanishing transverse shear deformation(*), provided the coefficients a_{3i} $(i = 0, \cdots, 5)$ are all arbitrary. Such completeness, together with the continuity and boundedness of the body force density components, was shown to be sufficient for convergence if the exact solution belongs to C', as it is supposed to.

As $(U_{3,\gamma} + I_\gamma)$ vanishes and the elastic coefficients $H_{\beta\gamma}$ are unbounded, Equations 85 and 86 cannot be used for the discussion of which derivatives must be kept continuous and bounded in order that m_1, m_2 and p_3 be also continuous and bounded.

Such difficulty may be removed by remarking that the exact solution is in this case the limit of a sequence of exact solutions corresponding to increasing values of the elastic coefficients $H_{\beta\gamma}$. Each one of such exact solutions is the limit of a sequence of approximate solutions $\{u_{an}\}$ if the completeness criterion with respect to C and the condition of boundedness

(*) As C is the set of the fields whose displacements are continuous and bounded, together with their first and second derivatives, C' is, by virtue of Equation 91, the set of the fields in which displacement U_3 is continuous and bounded, together with its derivatives of order up to three.

and continuity of the displacements and their first and second derivatives are respected. Such conditions must still be valid for the limit case.

As by virtue of Equation 91, the continuity and boundedness of the rotations and their first and second derivatives are insured by the continuity and boundedness of the first, second and third derivatives of U_3, it becomes clear that the convergence conditions for the simplified theory of plates are simply the arbitrariness of the coefficient a_{3i} $(i = 0, 1, \cdots 5)$ and the boundedness and continuity within each element of U_3 and of its first, second and third derivatives.

SECTION VII
CONCLUSIONS

The results of the present paper may be summarized as follows:

a) A very simple completeness criterion has been given with respect to the set C (to which the exact solution is assumed to belong) of the compatible elastic fields whose displacement second derivatives remain continuous and bounded within each element, whatever the size.

b) The completeness criterion becomes a convergence criterion whenever compatibility is not violated across the element boundaries (Ritz method).

c) Completeness with respect to C also ensures convergence in any case, i.e., even if compatibility is violated, whenever the body force density components corresponding to the successive approximate solutions remain continuous and bounded within the elements as their size decreases indefinitely.

d) Continuity and limitation of the displacement derivatives of order up to two ensures continuity and limitation of the body force density components, so that convergence to the exact solution will be obtained whenever completeness is achieved and such derivatives remain continuous and bounded within the elements.

e) The condition of completeness with respect to C is not a necessary condition for convergence. Completeness with respect to a subset $C' \subset C$ which contains the exact solution is sufficient. This conclusion makes it possible to derive convergence criteria for the simplified theories which result from neglecting the transverse shear deformation.

f) In what concerns plates, an important subset of C is the set C' of the fields of C with vanishing transverse shear deformation. The completeness criterion with respect to C' is simply that the transverse displacement U_3 be given by

$$U_3 = a_{30} + a_{31} x_1 + a_{32} x_2 + a_{33} x_1^2 + a_{34} x_1 x_2 + a_{35} x_2^2 + P_3^{(3)} \tag{95}$$

in which the coefficients $a_{3i}(i = 0, 1, \cdots 5)$ are arbitrary. The polynomial $P_3^{(3)}$ can vanish for any values of the coefficients, and the rotations I_1 and I_2 must be deduced by using Equation 91.

g) In the case of the simplified theory of plates, sufficient conditions for convergence are completeness with respect to C', and the continuity and limitation, within each element, of the derivatives of U_3, of order up to three, corresponding to the successive approximate solutions.

h) Completeness with respect to C' is by itself, however, a sufficient condition for convergence whenever compatibility (continuity of U_3, I_1 and I_2) between elements is not violated.

It remains to apply these conclusions to actual cases.

In what concerns two- and three-dimensional elasticity, conformity is easy to achieve. Convergence will then be obtained if the completeness criterion is obeyed. Completeness and the condition of the body force density components remaining continuous and bounded within each element (this is mainly the case if the body forces vanishes, as happens frequently) would however ensure convergence even if conformity were not respected.

In what concerns plates (simplified theory), conformity has not been easy to achieve. If it is achieved, the completeness criterion (with respect to C') will be sufficient for convergence. If it is not achieved, convergence is ensured only when both the completeness criterion and the condition of the derivatives of the displacements of order up to three remaining continuous and bounded within each element are respected.

As the expressions for U_3 are polynomials at least of the third degree, the third derivatives of U_3 generally do not vanish and it may happen that they increase beyond all limit when the size of the elements decreases indefinitely.

This is why the mesh plays a role in convergence. Indeed, some types of mesh (like the union jack mesh in Reference 5,lead to unbounded third derivatives of U_3, whereas more regular types of mesh lead to bounded third derivatives of U_3, and thus to convergence.

It is possible to check the limitation of the third derivatives by inspecting the sequence of the approximate solutions, or by examining the behavior of groups of elements (Reference 5). Such procedures do not allow prediction of convergence before any computations have been made, but they make it possible to know that the sequence converges to the exact solution, even if such exact solution is not available.

SECTION VIII

REFERENCES

1. S. G. Mikhlin, Variational Methods of Mathematical Physics. New York, Pergamon Press, 1963.

2. E. R. Arantes Oliveira, "Theoretical Foundations of the Finite Element Method," Int. J. Solids Structures, Vol. 4, No. 10, October 1968, pp 929-952.

3. S. W. Key, A Convergence Investigation of the Direct Stiffness Method. Ph. D. Dissertation, University of Washington, 1966.

4. Pin Tong and T. H. H. Pian, "The Convergence of Finite Element Method in Solving Linear Elastic Problems," Int. J. Solids Structures, Vol. 3, No. 5, pp. 865-879, September 1967.

5. G. P. Bazeley, Y. K. Cheung, B. M. Irons, O. C. Zienkiewicz, "Triangular Elements in Plate Bending. Conforming and Non-Conforming Solutions." Proceedings of Conference on Matrix Methods in Structural Mechanics, Wright-Patterson AFB, Ohio, (AFFDL-TR-66-80) 1965.

6. O. C. Zienkiewicz and Y. K. Cheung, The Finite Element Method in Structural and Continuum Mechanics. London, Mc Graw-Hill, 1967.

7. R. W. Clough and J. L. Tocher, "Finite Element Stiffness Matrices for Analysis of Plate Bending." Proceedings of Conference on Matrix Methods in Structural Mechanics, Wright-Patterson AFB, Ohio, (AFFDL-TR-66-80) 1965.

8. R. Courant and D. Hilbert, "Methods of Mathematical Physics," Vol II. International Publishers (1962).

9. E. R. Arantes Oliveira, Introdução à Teoria das Estruturas. Lisboa Author's edition, 1966.