UNIVERSITE DE LIEGE
FACULTE DES SCIENCES APPLIQUEES
LABORATOIRE DE METHODES DE FABRICATION

SOME ASPECTS OF THE FINITE ELEMENT ERRORS

J.F. DEBONGNIE

1

A PRIORI ERROR ESTIMATIONS

J.F. DEBONGNIE
Laboratoire de Méthodes de Fabrication
U.Lg.

1. INTRODUCTION

A priori error estimations are obtained from a general analysis of boundary value problems and of their discretization. As a tool are used some functional analysis concepts. The main result is the *order of convergence* for a mesh refinement. However, some constants are implied, whose evaluation is difficult. Consequently, the a priori estimations are essentially of *qualitative* nature.

In what follows, the simple potential problem (Poisson's equation) will be used as a model, in order to introduce the fundamental concepts in their simplest form. However, the results are of a general nature, and will finally be applied to the more complex problem of elasticity.

2. POTENTIAL PROBLEMS

2.1.- Local and variational formulations
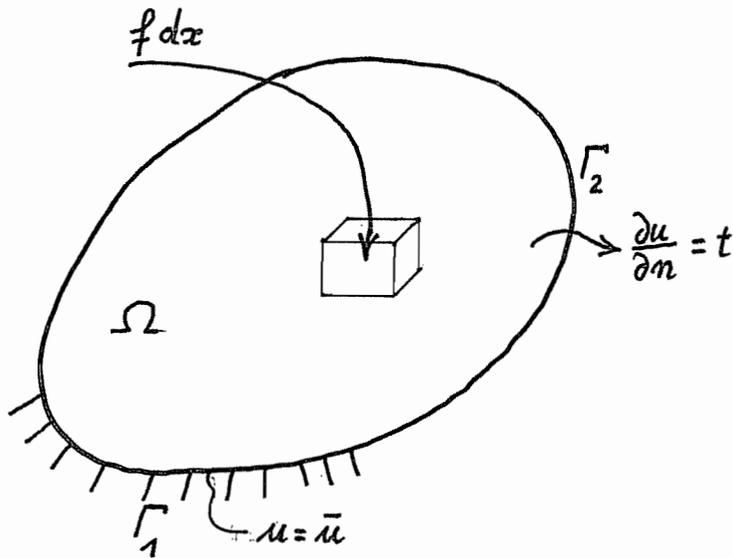
Let us consider the classical Poisson equation in a bounded open set Ω,

$$-\nabla u = f \text{ in } \Omega \tag{1}$$

The boundary Γ of Ω is splitted in two parts, namely,

- Γ_1 where the potential u has to conform with a prescribed value,

$$u = \bar{u} \text{ on } \Gamma_1 \tag{2}$$

- Γ_2 where Neumann conditions are applied,

$$\frac{\partial u}{\partial n} = t \text{ on } \Gamma_2 \tag{3}$$

As is well known from Dirichlet, this problem admits a variational formulation, which is as follows

Among all functions u that verify $u = \bar{u}$ on Γ_1, find the particular one which minimizes the functional

$$\mathscr{E}(u) = \frac{1}{2}\int_\Omega D_i u D_i u \, dx - \int_\Omega fu \, dx - \int_{\Gamma_2} tu \, d\Gamma \qquad (4)$$

2.2.- More precisely

The abovementioned formulation is somewhat naïve, because we did not specify how regular the functions have to be.

At a first glance, one should be temptated to impose the condition $u \in C^2(\Omega)$ with C^1 extension on $\overline{\Omega}$, so that the strong equations (1) to (3) will be pointwise verified. But such a point of view is a too narrow one, because it is not true that the strong equations admit a solution whatever be the geometry and the data f and t.

In contrast, the variational form of the problem constitutes its natural frame, since it leads to a proof of existency. Moreover, the quantities which are involved in the functional $\mathscr{E}$ possess a more sound physical sense (energy) than local second derivatives.

So, the only condition which will be used is the fact that

$$\int_\Omega D_i u D_i u \, dx < \infty \qquad (5)$$

Admissible functions will then be defined as functions which are of finite energy (condition 5) and which verify the essential boundary condition (2) on Γ_1. The set of admissible functions will be noted U.

2.3.- The first variation

In order to find the minimum, the first variation $\delta\mathscr{E}$ of $\mathscr{E}$ has to be equated to zero. This writes

$$0 = \delta\mathscr{E} = \int_\Omega D_i u D_i \delta u \, dx - \int_\Omega f\delta u \, dx - \int_{\Gamma_2} t \, \delta u \, d\Gamma \qquad (6)$$

Here appear field variations δu, which have to be of the form

$$\delta u = u_1 - u_2, u_1 \in U, u_2 \in U \qquad (7)$$

On the part S_1 of the boundary, u_1 and u_2 necessarily have the same value $\bar{u}$, so that

$$\delta u\big|_{\Gamma_1} = \bar{u} - \bar{u} = 0 \qquad (8)$$

3

So, the *admissible variations* verify (5) and (8). They clearly form a linear (vector) space, which will be noted δU.

2.4.- Abstract form of the problem

Let us introduce the following notations

- $a(u,v) = \int_\Omega D_i u D_i v \, dx$ (9)

It is a symmetric bilinear functional

- $f(u) = \int_\Omega fu \, dx + \int_{S_2} tu \, dS$ (10)

It is a linear functional.

Our problem is now reduced to the following abstract form, *find $u \in U$ such that for each $\delta u \in \delta U$ one has*

$$a(u, \delta u) = f(\delta u) \quad (11)$$

Does this problem admit a solution ? And if the answer is positive, is this solution a unique one ? To answer these questions, we have to introduce a norm on a space V that contains U and δU, say $\|u\|_V$. Then, the bilinear functional a(u,v) has to be *bounded* by respect of the norm, a condition which writes

$$|a(u,v)| \leq A\|u\|_V \|v\|_V \quad (12)$$

Similarly, the linear functional f has to verify the boundedness condition

$$|f(u)| \leq C\|u\|_V \quad (13)$$

In this case, the quantity

$$\|f\|_{V'} = \sup_{u \in V} \frac{|f(u)|}{\|u\|_V} \quad (14)$$

is known as the norm of f (in the dual V').

A supplementary condition, which is called the *δU - ellipticity* condition, is the existency of a $\alpha > 0$ such that for each $v \in \delta U$

$$\alpha\|v\|^2 \leq a(v,v) \quad (15)$$

4

A problem whose functionals a and f verify conditions (12), (13) and (15) will be called a *well-posed elliptic problem*.

2.5.- Existency in the homogeneous case

In the homogeneous case $U = \delta U$, that is, for the potential problem, $\bar{u} = 0$ on S_1. In this case, it is natural to take $V = \delta U$, and the fact that a is elliptic and bounded leads to the relation

$$\alpha \|u\|_v^2 \leq a(u,u) \leq A \|u\|_v^2 \tag{16}$$

for each $u \in V$. This is to say that the energetic norm

$$\|u\|_a = \sqrt{a(u,u)} \tag{17}$$

is equivalent to the natural norm of the space. They define the same open and closed sets, and the Hilbert space V may be equipped with the scalar product a(u,v). The existency and unicity of the solution now follows from the celebrated *RIESZ representation theorem* [3,4,5].

Let V be a Hilbert space with a scalar product (u,v), and let f be a bounded linear functional. Then there exists an unique element $u \in V$ such that for each $v \in V$.

$$(u,v) = f(v) \tag{18}$$

and

$$\|u\|_v = \|f\|_{v'} \tag{19}$$

2.6.- The nonhomogeneous case

In the nonhomogeneous case, the first step is to find a particular element u_{part} in V such that on Γ_1,

$$u_{part} = \bar{u} \tag{20}$$

Without risk of confusion, this particular element u_{part} may be noted $\bar{u}$ in Ω also. Writing the solution u of problem (11) as

$$u = \bar{u} + v, \text{ with } v \in \delta U \tag{21}$$

this problem is now reduced to

$$a(\bar{u} + v, \delta u) = f(\delta u)$$

5

Developping the first member, this writes

$$a(\overline{u}, \delta\, u) + a(v, \delta\, u) = f(\delta\, u)$$

Here, the first term of the left member may be seen as a linear functional of δu, leading to the following homogeneous problem : find $v \in \delta\, U$ such that for each $\delta\, u \in \delta\, U$,

$$a(v, \delta\, u) = f(\delta\, u) - a(\overline{u}, \delta\, u) \equiv g(\delta\, u) \qquad (22)$$

and we return to the preceeding case.

The only difficulty concerns the unicity. In fact, we arbitrarily choosed a *particular* $\overline{u}$. Does the solution $u = \overline{u} + v$ depend on this choice ? Suppose it is so, and let be

- $u_1 = \overline{u}_1 + v_1 = solution\ with\ \overline{u}_1$
- $u_2 = \overline{u}_2 + v_2 = solution\ with\ \overline{u}_2$

Then

$$\begin{cases} a(v_1, \delta\, u) = f(\delta\, u) - a(\overline{u}_1, \delta\, u) \\ a(v_2, \delta\, u) = f(\delta\, u) - a(\overline{u}_2, \delta\, u) \end{cases}$$

Substracting,

$$a(v_1 - v_2, \delta\, u) = -a(\overline{u}_1 - \overline{u}_2, \delta\, u)$$

or

$$0 = a(\overline{u}_1 + v_1) - (\overline{u}_2 + v_2), \delta\, u) = a(u_1 - u_2, \delta\, u)$$

But $u_1 - u_2 = o$ on Γ_1, so that one may take $\delta\, u = u_1 - u_2$, which leads

$$a(u_1 - u_2, u_1 - u_2) = o$$

and from the δU - ellipticity, this implies $u_1 = u_2$.

2.7.- Return to the actual problem

The natural Hilbert space for the potential problem is the *SOBOLEV space H^1 (Ω)* whose elements u verify

$$u \in L^2(\Omega), D_i u \in L^2(\Omega),$$

with as scalar product

$$(u,v)_{1,\Omega} = \int_{\Omega} uv \ dx + \int_{\Omega} D_i u D_i v \ dx \qquad (23)$$

The corresponding norm is

$$\|u\|_{1,\Omega}^2 = \int_{\Omega} u^2 dx + \int_{\Omega} D_i u D_i u \ dx \qquad (24)$$

As can be seen, the used bilinear form

$$a(u,v) = \int_{\Omega} D_i u \ D_i v \ dx$$

is only *a part* of the scalar product (23) and moreover,

$$a(u,u) = \int_{\Omega} D_i u D_i u \ dx = |u|_{1,\Omega}^2$$

is not a norm but a *seminorm*, it is to say that

$$|u|_{1,\Omega} = o \not\Rightarrow u = o.$$

In fact,

$$|u|_{1,\Omega} = o \Rightarrow D_i u = o \Rightarrow u = cte.$$

However, with proper fixations, the ellipticity condition will hold. This is a consequence of some classical results about Sobolev spaces, which will be summarized in what follows.

2.8. Sobolev spaces

2.8.1. More generally, the space $H^m(\Omega)$ is composed of those functions $u \in L^2(\Omega)$ whose derivatives up to order m are also square integrable. The scalar product on $H^m(\Omega)$ is

$$(u,v)_{m,\Omega} = \int_{\Omega} (uv + D_i u D_i v + ... + D_{i_1...i_m} u D_{i_1...i_m} v) \ dx \qquad (25)$$

The corresponding norm is $\|u\|_{m,\Omega}$. Also of interest is the m-seminorm

$$|u|_{m,\Omega}^2 = \int_{\Omega} (D_{i_1...i_m} u \ D_{i_1...i_m} v) dx \qquad (26)$$

Of course, $H^o(\Omega) = L^2(\Omega)$.

2.8.2. Continuity

Are elements of $H^m(\Omega)$ continuous ? This depends on m and on the dimension of the space. The following result holds (Sobolev) :

$$u \in H^m(\Omega) \Rightarrow u \in C^o \quad \text{if } m > n/2 \qquad (27)$$

So, if n = 1, $H^1(\Omega) \Rightarrow C^o$, but it is not true if n = 2 or 3.

* Example of a discontinuous H^1 function in R^2 [7]

Let Ω be a circle of radius $R = \dfrac{1}{2}$ centered at the origin, and

$$u = \left(\ln \frac{1}{r} \right)^\alpha = (-\ln r)^\alpha$$

One has

$$\frac{du}{dr} = \alpha(-\ln r)^{\alpha-1} \cdot \left(-\frac{1}{r} \right)$$

and

$$\int_\Omega \left(\frac{du}{dr} \right)^2 dx = 2\pi \int_o^R r \left(\frac{du}{dr} \right)^2 dr = 2\pi \, \alpha^2 \int_o^{1/2} \frac{(-\ln r)^{2\alpha-2}}{r} dr$$

Setting $r = e^{-t}$, one finds

$$dr = -e^{-t} dt$$
$$-\ln r = t$$

and

$$\int_o^1 \frac{(-\ln r)^{2\alpha-2}}{r} dr = \int_{\ln 2}^\infty t^{2\alpha-2} dt = \frac{1}{2\alpha-1} \left[t^{2\alpha-1} \right]_{\ln 2}^\infty ,$$

a quantity which is finite when $\alpha < 1/2$. So, $u \in H^1(\Omega)$, but for $r \to o$, $u \to \infty$.

* For n = 1, every H^1 function is continuous

In fact, if $u \in H^1(]a,b[)$ and $x \in]a,b[$,

$$\left| u(x+h) - u(x) \right| = \left| \int_x^{x+h} u'(\xi) d\xi \right|$$

$$\leq \left[\int_x^{x+h} u'^2 \, d\xi \right]^{1/2} \cdot \left[\int_x^{x+h} 1 d\xi \right]^{1/2}$$

$$\leq \sqrt{h} \left(\int_a^b u'^2 \, d\xi \right)^{1/2}$$

So, u is continuous and moreover, it verifies a Hölder condition of degree 1/2.

2.8.3. Trace theorem

Any element u of $H^1(\Omega)$ admits on Γ a trace $u|_\Gamma$ which verifies

$$\int_\Gamma \left(u|_\Gamma \right)^2 d\Gamma \leq C \int_\Omega \|u\|_{1,\Omega}^2 \tag{28}$$

Remarks

a) This is not an optimal result. In fact, traces of H^1 functions define a more special space on Γ, which is called $H^{1/2}(\Gamma)$.
b) The condition $u|_\Gamma \in L^2(\Gamma)$ is not sufficient for $u|_\Gamma$ to be the trace of a $H^1(\Omega)$ function. Even the condition $u|_\Gamma \in L^2(\Gamma) \cap C^o(\Gamma)$ is not sufficient, as proved by Hadamard.

Corollary. To $u \in H^m(\Omega)$ are associated traces of u and its derivatives up to order (m-1), all of which lying in $L^2(\Gamma)$.

2.8.4.- A compacity theorem

The following very useful theorem is attributed to Rellich and Kondrasov.

From each bounded sequence $\{u_k\}$ of elements of $H^m(\Omega)$, it is possible to extract a subsequence $\{u_{k'}\}$ which converges in $H^l(\Omega)$, $l < m$.

2.9.- A general theorem about norm equivalences

A direct consequence of Rellich's theorem is the following theorem which may be used to prove a lot of ellipticity conditions.

Let b(u,v) be a bounded symmetric bilinear functional on $H^m(\Omega)$, such that

- $b(u,u) \geq 0$ *for each* $u \in H^m(\Omega)$ (29)
- *If u is a polynomial of degree (m-1) verifying b(u,u)=0, then u = 0*   (30)

Then, the quadratic form

$$\|u\|^2 = |u|^2_{m,\Omega} + b(u,u) \tag{31}$$

constitutes an equivalent norm for $H^m(\Omega)$.

Proof : We have to prove the existency of two strictly positive constants α and A such that

$$\alpha \|u\|^2_{m,\Omega} \leq \|u\|^2 \leq A\|u\|^2_{m,\Omega}$$

The second inequality being trivial, let us turn to the first one.

Suppose it is not verified. This to say that there exists a sequence $\{u_k\}$ such that

$$\frac{\|u_k\|}{\|u_k\|_{m,\Omega}} \to 0$$

Clearly, one may suppose that $\|u_k\|_{m,\Omega} = 1$. The sequence being bounded in $H^m(\Omega)$, it contains a subsequence $\{u_{k'}$ which converges in $H^{m-1}(\Omega)$, to some element u_o. So, we have

$$\begin{cases} |u_{k'}|_{m,\Omega} \leq \|u_{k'}\| \to 0 \\ \|u_{k'} - u_o\|_{m-1,\Omega} \to 0 \end{cases}$$

This implies

$$\lim_{p,q \to \infty} \|u_p - u_q\|_{m,\Omega} = 0,$$

that is the convergence in $H^m(\Omega)$ to a u_o which verifies

$$|u_o|_{m,\Omega} = 0$$

Thus, all derivatives of order m of u_o vanish, that is to say, u_o is a polynomial of degree (m-1).

Now,

$$b(u_o, u_o) = \lim b(u_{k'}, u_{k'}) = 0$$

so that $u_o = o$. But this implies $\|u_o\|_{m,\Omega} = 0$, which contradicts the fact that

$$\left\| u_o \right\|_{m,\Omega} = \lim \left\| u_{k'} \right\|_{m,\Omega} = 1$$

QED

2.10.- Ellipticity of the potential problem

Returning to the potential problem (6), let us consider the symmetric bilinear functional

$$b(u,v) = \int_{\Gamma_1} uv d\Gamma$$

It is bounded (trace theorem), and for any constant field u = c, one has

$$b(c,c) = c^2 \int_{\Gamma_1} d\Gamma = c^2 meas\ (\Gamma_1)$$

so that

$$b(c,c) = 0 \Rightarrow c = 0$$

provided meas $(\Gamma_1) \neq 0$ (this condition is a sharp one).

So, from the preceeding theorem, there exists a strictly positive α such that

$$\alpha \left\| u \right\|_{1,\Omega}^2 \leq \left| u \right|_{1,\Omega}^2 + \int_{\Gamma_1} u^2 d\Gamma \tag{32}$$

and, for the subspace δU which is characterized by the condition

$$u\Big|_{\Gamma_1} = 0,$$

this reduces to

$$\alpha \left\| u \right\|_{1,\Omega}^2 \leq \left| u \right|_{1,\Omega}^2 \qquad (\forall u \in \delta\ U) \tag{33}$$

which is precisely the needed inequality.

Exercise. Prove that inequality (33) holds in

$$\left\{ u \in H^1(\Omega) \Big| \int_{\Omega} u dx = 0 \right\}$$

(Poincaré's inequality)

3. RAYLEIGH-RITZ APPROXIMATIONS

3.1.- Introduction

The following developments are of a general nature, and it is why they are presented in a separated chapter. However, the reader may always refer to the potential problem as a model.

3.2.- The homogeneous case

We will first turn to the homogeneous problem characterized by $V = U = \delta U$. Obtaining the exact solution of the variational problem is impossible in the general case. It is therefore necessary to develop *approximate* solutions. One way to obtain such approximations consists to define a *finite dimensional subspace* $V_h = U_h = \delta U_h \subset \delta U$. The discretized problem is then

Find $u_h \in V_h$ such that for any $\delta u_h \in V_h$,

$$a(u_h, \delta u_h) = f(\delta u_h) \tag{1}$$

Two questions now arise,

(i)   *Does problem* (1) *admit an unique solution ?*
(ii)   *How different is this approximate solution to the true one ?*

3.3.- Existency and uniqueness of u_h

This is a consequence of the fact that V_h, being a finite dimensional subspace of V, is a closed subspace. It is therefore a Hilbert space, and Riesz' theorem applies.

3.4.- Discretization error

The discretization error will be measured by the distance $\|u - u_h\|_V$ between the true solution and the discretized one.

3.4.1.- The orthogonality property

The true problem consists to find $u \in V$ such that $a(u, \delta u) = f(\delta u)$

for any $\delta u \in V$. Since $V_h \subset V$, one has for any $\delta u_h \in V_h$,

$$a(u, \delta u_h) = f(\delta u_h) \tag{2}$$

Substracting (1) from this equation, one obtains

$$a(u - u_h, \delta u_h) = 0 \tag{3}$$

for any $\delta u_h \in V_h$. This is to say that *the error (u - u_h) is orthogonal to V_h.*

3.4.2.- Céa's lemma

Let us now compute

$$a(u - u_h, u - u_h) = a(u - u_h, u) - a(u - u_h, u_h)$$

The last term vanishes by virtue of the orthogonality property. Moreover, this property implies that for *any $v_h \in V_h$,*

$$a(u - u_h, v_h) = 0,$$

so that

$$a(u - u_h, u - u_h) = a(u - u_h, u - v_h)$$

Now, from the boundedness and the ellipticity of a,

$$\alpha \|u - u_h\|_V^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq A \|u - u_h\|_V \|u - v_h\|_V$$

Excluding the trivial case $u_h = u$, this gives

$$\boxed{\|u - u_h\|_V \leq \frac{A}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|} \tag{4}$$

So, the norm of the error is proportional to the least distance from u to V_h.

3.5.- The nonhomogeneous case

A strict Rayleigh-Ritz procedure is possible in the nonhomogeneous case only if there exists a particular $\bar{u} \in V_h$ which complies *exactly* with conditions on Γ_1. Other cases will be referred as non strict Rayleigh-Ritz approximations, and not considered here.

The true solution is of the form

$$u = \bar{u} + r, \, r \in \delta \, U,$$

and verifies, for each $\delta \, u \in \delta \, U,$

$$a(u, \delta \, u) = f(\delta \, u) \tag{5}$$

The discretized solution is of the form

$$u_h = \bar{u} + r_h, \quad r_h \in \delta \, U_h \subset \delta \, U$$

and verifies, for each $\delta \, u_h \in \delta \, U_h$,

$$a(u_h, \delta \, u_h) = f(\delta \, u_h) \tag{6}$$

Substracting (5) with $\delta \, u = \delta \, u_h$ from (6) leads to

$$a(u - u_h, \delta \, u_h) = 0, \tag{7}$$

so that the error is orthogonal to the discretized space of variations.

Let us now compute

$$\begin{aligned} a(u - u_h, u - u_h) &= a(u - u_h, u - \bar{u} - r_h) \\ &= a(u - u_h, u - \bar{u}) - a(u - u_h, r_h) \end{aligned}$$

and the last term vanishes. Furthermore, for any $w_h \in \delta \, U_h$,

$$a(u - u_h, u - u_h) = a(u - u_h, u - \bar{u} - w_h).$$

As $(u - u_h) \in \delta \, U$, one may use the δU-ellipticity of a to obtain

$$\alpha \| u - u_h \|_V^2 \le A \| u - u_h \|_V \| u - \bar{u} - w_h \|_V$$

and, from the arbitrariness of w_h,

$$\boxed{\left\| u - u_h \right\|_V \leq \frac{A}{\alpha} \inf_{v_h \in \bar{u} + \delta\, U_h} \left\| u - v_h \right\|_V} \qquad (8)$$

The norm of the error is thus proportional to the distance from u to the linear manifold $(\bar{u} + \delta\, U_h)$.

3.6.- Error evaluations from the interpolation theory

The fundamental results (4) and (8) refer to the *best* approximation of u by a v_h, in the sense of the V-norm. This best v_h is not easy to determine and to characterize, but supposing that some w_h is known (which is as closed as possible to u), one has

$$\inf_{v_h} \left\| u - v_h \right\| \leq \left\| u - w_h \right\|$$

and an *upper bound* of the error is obtained.

In the case of finite element approximations, the basic idea is that if the solution u of the variational problem is *sufficiently smooth*, it admits a finite element *interpolation* $\Pi_h u$, generally differing from u_h, but verifying

$$\left\| u - u_h \right\|_V \leq \frac{A}{\alpha} \left\| u - \Pi_h u \right\|_V \qquad (9)$$

This reduces the problem to an interpolation theory.

3.7.- Existency of the finite element interpolation

Any interpolation makes use of *local* values of the field, and thus supposes that u is continuous. From the theory of Sobolev spaces, this supposes at least $u \in H^2(\Omega)$ if n > 1 and $u \in H^1(\Omega)$ if n = 1, n being the dimension of the space.

For interpolations making use of local derivatives (beams, plates), the conditions are $u \in H^3(\Omega)$ if n > 1 and $u \in H^2(\Omega)$ if n = 1.

3.8.- Degree of an element

In each element e, the interpolation operator Π_h is said to be of degree k if for any polynomial p_k *of degree k*, one has

$$\Pi_h p_k = p_k \quad \text{on e.} \qquad (10)$$

3.9.- A fundamental lemma

The interpolation theory may be based on the following lemma :

Let u be a function on an open set e, and let $R_k u$ be its best approximation in the set $P_k(e)$ of polynomials of degree k on e, for the L^2-norm, that is,

$$(\forall p_k \in P_k(e)) \qquad (u - R_k u, p_k)_{o,e} = o \tag{11}$$

Then, if $u \in P_k(e)$, one has $R_k u$ = u. It follows from the theorem on norm equivalences that there exists a constant $\alpha > 0$ such that

$$\alpha \|u\|_{k+1,e}^2 \leq |u|_{k+1,e}^2 + \|R_k u\|_{o,e}^2 \tag{12}$$

Now, as

$$R_k(u - R_k u) = 0,$$

this implies

$$\alpha \|u - R_k u\|_{k+1,e}^2 \leq |u|_{k+1,e}^2 \tag{13}$$

3.10.- Interpolation error

Supposing that $u \in H^{k+1}(e)$, k being such that the interpolation operator is bounded (see § 3.7), and that Π_h is of degree k. Then, if $p_k \in P_k(e)$, one has

$$\Pi_h p_k = p_k$$

so that

$$u - \Pi_h u = u - p_k - \Pi_h u + \Pi_h p_k = (I - \Pi_h)(u - p_k)$$

and

$$\|u - \Pi_h u\|_{k+1,e} \leq \|I - \Pi_h\| \|u - p_k\|_{k+1,e} = C_1 \|u - p_k\|_{k+1,e}$$

Setting p_k = $R_k u$, this leads to

$$\|u - \Pi_h u\|_{k+1,e} \leq C_1 \|u - R_k u\|_{k+1,e} \leq \frac{C_1}{\sqrt{\alpha}} |u|_{k+1,e} = C_2 |u|_{k+1,e}$$

Since for m < k+1,

$$\left| u - \Pi_h u \right|_{m,e} \le \left\| u - \Pi_h u \right\|_{m,e} \le \left\| u - \Pi_h u \right\|_{k+1,e} ,$$

one obtains

$$\left| u - \Pi_h u \right|_{m,e} \le C_2 \left| u \right|_{k+1,e} \tag{14}$$

3.11.- Scale effect, the affine theory

The constant appearing in (14) *depends on the open set e*. To obtain this dependency, let us suppose that each element of the finite element model is related to an *unique reference element* ε in the space (ξ₁, ..., ξₙ), by an affine transformation

$$\xi \in \varepsilon \leftrightarrow x_i = A_{ij}\xi_j + b_i \in e \tag{15}$$

with constant A_{ij} and b_i (this is always the case with triangular elements). With such an affine transformation, polynomials of degree k are transformed in polynomials of the same degree.

Now, if at u defined in e corresponds $\hat{u}(\xi) = u(x(\xi))$, one has $\left(\partial_i = \dfrac{\partial}{\partial \xi_i} \right)$

$$\partial_i \hat{u} = \partial_i x_j D_j u = A_{ji} D_j u \tag{16}$$

and

$$D_i u = D_i \xi_j \partial_j \hat{u} = A_{ji}^{-1} \partial_j \hat{u} \tag{17}$$

From this follows

$$D_i u D_i u = A_{ji}^{-1} A_{ki}^{-1} \partial_j \hat{u} \partial_k \hat{u} = (A^{-1} A^{-T})_{jk} \partial_j \hat{u} \partial_k \hat{u} = G_{jk}^{-1} \partial_j \hat{u} \partial_k \hat{u}$$

where the symmetric positive definite matrix

$$G = (A^{-1} A^{-T})^{-1} = A^T A$$

is the well known metric tensor, defined by

$$G_{ij} = \partial_i x_k \partial_j x_k \tag{18}$$

Its positive eigenvalues measure the principal values of the length in e for a unit length in ε. Let us note

$$h^2 = \lambda_{\max}(G) \quad \text{and} \quad \rho^2 = \lambda_{\min}(G) \tag{19}$$

Then,

$$\frac{1}{h^2}\partial_i\hat{u}\partial_i\hat{u} \leq G_{jk}^{-1}\partial_j\hat{u}\partial_k\hat{u} = D_iu\,D_iu \leq \frac{1}{\rho^2}\partial_i\hat{u}\partial_i\hat{u}$$

Applying this result to higher order derivatives leads to

$$\frac{1}{h^{2l}}\partial_{i_1\ldots i_l}\hat{u}\,\partial_{i_1\ldots i_l}\hat{u} \leq D_{i_1\ldots i_l}u\,D_{i_1\ldots i_l}u \leq \frac{1}{\rho^{2l}}\partial_{i_1\ldots i_l}\hat{u}\,\partial_{i_1\ldots i_l}\hat{u}$$

Integrating on e, one obtains

$$\frac{|\det A|}{h^{2l}}|\hat{u}|_{l,\varepsilon}^2 \leq |u|_{l,e}^2 \leq \frac{|\det A|}{\rho^{2l}}|\hat{u}|_{l,\varepsilon}^2 \tag{20}$$

Let us now apply this result to make (14) more precise.

One has first

$$|u - \Pi_h u|_{m,e} \leq \frac{|\det A|^{1/2}}{\rho^l}|\hat{u} - \hat{\Pi}_h u|_{m,\varepsilon} \tag{21}$$

and for any $\hat{p}_k \in P_k(\varepsilon)$, $p_h \in P_k(e)$ and $\Pi_h p_k = p_k$, which implies $\Pi_h p_k = \hat{p}_k$. So, result (14) may be used in ε and

$$|\hat{u} - \Pi_h u|_{m,\varepsilon} \leq C(\varepsilon)|\hat{u}|_{k+1,\varepsilon} \tag{22}$$

Now, from (20),

$$|\hat{u}|_{k+1,\varepsilon} \leq \frac{h^{k+1}}{|\det A|^{1/2}}|u|_{k+1,e} \tag{23}$$

Assembling results (21) to (23) leads to

$$|u - \Pi_h u|_{m,e} \leq C(\varepsilon)\cdot\frac{h^{k+1}}{\rho^m}|u|_{k+1,e} \tag{24}$$

3.12.- The error on the whole domain

The finite element model will be said *uniform* if there exists strictly positive constants K_1 and K_2 such that

$$\max_{e} h_e \leq K_1 h \qquad \inf_{e} \rho_e \geq K_2 h,$$

h being the diameter of a given element. For such a mesh,

$$|u - \Pi_h u|_{m,e} \leq C \, h^{k+1-m} |u|_{k+1,e} \, ,$$

with C independent from the element.

Consequently,

$$|u - \Pi_h u|^2_{m,\Omega} = \sum_e |u - \Pi_h u|^2_{m,e} \leq C^2 h^{2(k+1-m)} \sum_e |u|^2_{k+1,e} \leq C^2 h^{2(k+1-m)} |u|^2_{k+1,\Omega}$$

For the m-norm, one has

$$\|u - \Pi_h u\|^2_{m,\Omega} = \sum_{l=o}^{m} |u - \Pi_h u|^2_{l,\Omega} \leq \sum_{l=o}^{m} C_l^2 h^{2(k+1-l)} |u|^2_{k+1,\Omega}$$

from which

$$\|u - \Pi_h u\|_{m,\Omega} \leq \left[C + O(h^2)\right] h^{k+1-m} |u|_{k+1,\Omega} \tag{25}$$

and for sufficiently small values of h, the factor $\left[C + O(h^2)\right]$ is bounded by a constant C*.

3.13.- Conclusions

Let a(u,v) involve derivatives of order m. Then the suited space V is $H^m(\Omega)$. The form a is supposed to be bounded and elliptic, and the linear form f has to be bounded.

Then, if finite elements of degree k are used (see 3.8 for the definition of the degree), and if the solution $u \in H^{k+1}(\Omega)$, the finite element solution verifies

$$\|u - u_h\|_{m,\Omega} \leq C \cdot h^{k+1-m} |u|_{k+1,\Omega} \tag{26}$$

But if u is not so smooth, say, $u \in H^2(\Omega)$, $m < l < k+1$, a less favourable result is obtained,

$$\|u - u_h\|_{m,\Omega} \leq C \, h^{l+1-m} |u|_{l+1,\Omega} \tag{27}$$

In the case where $u \in H^m(\Omega)$, it is possible to prove that

$$\|u - u_h\|_{m,\Omega} \to o \quad \text{when } h \to o,$$

by a density argument, but no order of convergence can be guaranteed [8].

4. ELASTICITY PROBLEMS

4.1.- Local equations

The local equations are here

- Compatibility
$$\varepsilon_{ij} = \frac{1}{2}(D_i u_j + D_j u_i) \text{ in } \Omega$$
$$u_i = \overline{u}_i \text{ on } \Gamma_1$$

- Equilibrium
$$D_j \sigma_{ji} + f_i = o \text{ in } \Omega \qquad f_i = \text{body forces}$$
$$n_j \sigma_{ji} = t_i \text{ on } \Gamma_2$$

- Constitutive equations
$$\sigma_{ij} = C_{ijkl} \varepsilon_{kl}$$

with

$$\beta \, \varepsilon_{ij} \varepsilon_{ij} \le C_{ijkl} \varepsilon_{ij} \varepsilon_{kl} \le B \, \varepsilon_{ij} \varepsilon_{ij}$$

β and B being *strictly* positive constants independent from x.

4.2.- Variational formulation

The variational formulation is

$$\delta \, \mathscr{E} = o , \tag{1}$$

with

$$\mathscr{E} = \mathscr{U} + \mathscr{P} , \tag{2}$$

where

$$= \frac{1}{2} \int_\Omega C_{ijkl} \varepsilon_{ij}(u) \varepsilon_{kl}(u) \, dx \tag{3}$$

and

$$\mathscr{P} = -\int_\Omega f_i u_i dx - \int_{\Gamma_2} t_i u_i d\Gamma \tag{4}$$

4.3.- Applicability

Except the fact that we are here confronted to vectors, the ideas are the same as for the potential case.

a) Admissible displacements have to verify

$$u_i = \bar{u}_i \text{ on } \Gamma_1. \tag{5}$$

Their set will be called U.

b) Admissible displacement variations, being differences of admissible displacements, verify

$$\delta u_i = o \text{ on } \Gamma_1. \tag{6}$$

They form a linear space δU

c) The boundedness of the strain energy variation

$$a(u,v) = \int_\Omega C_{ijkl} \varepsilon_{ij}(u) \, \varepsilon_{kl}(v) \, dx$$

follows from

$$\left| a(u,v) \right| \le B \left\| \varepsilon_{ij}(u) \right\|_{o,\Omega} \left\| \varepsilon_{kl}(v) \right\|_{o,\Omega} \le B \left\| \underset{\sim}{u} \right\|_{1,\Omega} \left\| \underset{\sim}{v} \right\|_{1,\Omega} \tag{7}$$

d) The greatest difficulty lies in ellipticity. In fact, a(u,u) contains some *combinations* of derivatives. One has

$$U(u) = \int_\Omega C_{ijkl} \varepsilon_{ij}(u) \, \varepsilon_{kl}(u) \, dx \ge \beta \int_\Omega \varepsilon_{ij}(u) \, \varepsilon_{ij}(u) \, dx$$

The only elementary case is $\bar{u}_i = o$ on $\Gamma_1 = \Gamma$. Then

$$\int_\Omega \varepsilon_{ij}(u) \, \varepsilon_{ij}(u) \, dx = \frac{1}{4} \int_\Omega (D_i u_j + D_j u_i)(D_i u_j + D_j u_i) \, dx$$

$$= \frac{1}{2} \int_\Omega D_i u_j \, D_i u_j \, dx + \frac{1}{2} \int_\Omega D_i u_j \, D_j u_i \, dx$$

For the last term, note that when $u_i = o$ on Γ,

$$\int_\Omega D_i u_j \, D_j u_i \, dx = -\int_\Omega u_j D_{ij} u_i \, dx = \int_\Omega D_j u_j D_i u_i \, dx,$$

so that

$$\int_\Omega \varepsilon_{ij}(u)\varepsilon_{ij}(u)\ dx = \frac{1}{2}\int_\Omega D_i u_j D_i u_j\ dx + \frac{1}{2}\int_\Omega D_i u_i D_j u_j\ dx$$

$$\geq \frac{1}{2}\int_\Omega D_i u_j D_i u_j\ dx = \frac{1}{2}\left|u\right|^2_{1,\Omega} \geq \frac{\alpha}{2}\left\|u\right\|^2_{1,\Omega}$$

due to the fixations.[1]

The general case is *very difficult*, and the result is known as *Korn's inequality*,

$$\alpha\left\|u\right\|^2_{1,\Omega} \leq \int_\Omega \varepsilon_{ij}(u)\varepsilon_{ij}(u)\ dx + \sum_i \left(\int_\Omega u_i dx\right)^2 + \sum_i \left(\int_\Omega \omega_i dx\right)^2 \tag{7}$$

with

$$\omega_i = e_{ijk} D_j u_k\ .$$

This implies that $H^1(\Omega)$ - ellipticity is obtained when the fixations guarantee the absence of rigid body modions.

4.4.- Convergence results

By the same way as for potential problems, one obtains, with elements of degree k,

$$\left\|\underset{\sim}{u} - \underset{\sim}{u}_h\right\|_{1,\Omega} \leq Ch^{k+1-1}\left|\underset{\sim}{u}\right|_{k+1,\Omega}$$

provided $\underset{\sim}{u} \in H^{k+1}(\Omega)$.

----

[1] Another trivial case is $u_i = D_i\varphi,\ \varphi \in H^2(\Omega)$. Then, $\varepsilon_{ij} = D_{ij}\varphi$ and

$$\int_\Omega \varepsilon_{ij}\varepsilon_{ij} dx = \int_\Omega D_{ij}\varphi\ D_{ij}\varphi\ dx = \int_\Omega D_j u_i D_j u_i\ dx\ .$$

So, when fixations implie $\left(u_i = ct \Rightarrow u_i = o\right)$, it follows from the general theorem 2.9 that

$$\int_\Omega \varepsilon_{ij}\varepsilon_{ij} dx \geq \alpha\left\|u\right\|^2_{1,\Omega}$$

BIBLIOGRAPHY

[1] S. GODOUNOV - *Equations de la Physique mathématique*. Mir, Moscou, 1973.

[2] V. MIKHAILOV - *Equations aux Dérivées partielles*. Mir, Moscou, 1980.

[3] F. RIESZ, B. Sz. NAGY - *Leçons d'Analyse fonctionnelle*. 6ᵉ éd., Gauthier-Villars, Paris, 1975.

[4] H.G. GARNIR - *Les Problèmes aux Limites de la Physique mathématique*. Birkhäuser Verlag, Basel & Stuttgart, 1958.

[5] S.C. BRENNER, L.R. SCOTT - *The mathematical Theory of finite Element methods*. Springer Verlag, New York, 1994.

[6] G. STRANG, G. FIX - *An Analysis of the finite Element Method*. Prentice Hall, 1973.

[7] J.L. LIONS - *Cours d'Analyse numérique*. Ecole polytechnique, promotion 1972.

[8] P.G. CIARLET - *Numerical Analysis of the finite Element Method*. Les presses de l'Université de Montréal, 1976.

SUPERCONVERGENCE

1. Interpolation property of some one-dimensional finite elements

The most classical one-dimensional finite elements have the property of interpolation. As an example, consider the variational problem

$$\frac{1}{2}\int_o^l N\left(\frac{du}{dx}\right)^2 dx - \int_o^l fu\, dx\ stat \tag{1}$$

with u = o at x = o and x = 1. It represents the equation of a very tight spring submitted to a transverse distributed load f. When using first degree elements, one obtains an approximate solution which is *exact* at the nodes. Why such a result ?

It is a consequence of the following abstract result. Let us consider a variational problem

$$a(u, \delta u) = f(\delta u) \tag{2}$$

and a finite element decomposition of the domain Ω. Let B be the space of admissible displacements that have zero traces on the element boundaries ("bubbles"), and let S be the orthogonal complement of B in the space V of admissible displacements, that is

$$s \in S \Leftrightarrow \forall b \in B, a(s,b) = 0 \tag{3}$$

Then, the solution is of the form

$$u = s + b, s \in S, b \in B$$

and verifies

$$\forall \delta s \in S \quad f(\delta s) = a(s + b, \delta s) = a(s, \delta s) \tag{4}$$

and

$$\forall \delta b \in B \quad f(\delta b) = a(s + b, \delta b) = a(b, \delta b) \tag{5}$$

As can be seen, problems (4) and (5) are *uncoupled* and their solutions are independent. *As a consequence, when omitting the bubbles, the true solution is obtained at the interelement boundaries.*

Let us now turn to the meaning of condition (3). It is to say that in each element,

$$a_e(s,b) = 0$$

for any admissible field which vanishes on the boundary. But this is equivalent to say that *s is a solution to the homogeneous problem.*

A. One-dimensional problems

In the case of one-dimensional problems, the set of homogeneous solutions is generally of finite dimension. As an example, the abovementioned problem (1) corresponds to the differential equation

$$-N\frac{d^2u}{dx^2} - f = 0$$

The homogeneous problem is

$$\frac{d^2u}{dx^2} = 0$$

and admits the general solution

$$u = \alpha_1 + \alpha_2 x$$

Therefore, using first degree elements, the discretized solution is exact at the nodes.

As a second example, let us consider a beam, whose variational principle writes

$$\frac{1}{2}\int_o^l EIw''^2 \, dx - \int_o^l pw dx \quad stat.$$

Here, the general solution of the associated homogeneous problem

$$EIw^{IV} = 0$$

is a third degree polynomial. As a consequence, provided the inertia I is constant in each element, a third degree finite element solution leads to exact values of w and w' at the nodes.

As a third example, the axisymmetric solutions of Poisson's equation in a two-dimensional space minimize the functional

$$\frac{1}{2}\int_{R_1}^{R_2} r\left(\frac{du}{dr}\right)^2 dr - \int_{R_1}^{R_2} furdr$$

The associated homogeneous equation is

$$\frac{d}{dr}\left(r\frac{du}{dr}\right) = 0$$

from which

$$r\frac{du}{dr} = \alpha_1,$$

$$\frac{du}{dr} = \frac{\alpha_1}{r},$$

and

$$u = \alpha_1 lnr + \alpha_2.$$

Consequently, specialized finite elements containing the functions "ct" and "lnr" would interpolate the exact solution at the nodes.

B. Two and three-dimensional problems

Here, it is no more possible to adopt as a basis in each element the set of solutions of the homogeneous problems, because this space is of infinite dimension. Let us consider the Poisson equation

$$-\nabla u = f$$

Here, the associate homogeneous equation is

$$-\nabla u = 0$$

and its solutions are the harmonic functions which form a non finite dimensional subspace. *There is no boundary interpolation.*

2. A theorem on bubbles

There is a discretized form of the above theorem. *Suppose that a finite element space V_h contains only fields that verify the homogeneous equation in each element* (e.g., for harmonic problems, first degree elements). *Then, adding bubbles in each element will not modify the solution at the element boundaries.*

In fact, if B_h is the space of added bubbles, the finite element solution will be of the form

$$u_h = s_h + b_h, \quad s_h \in V_h, \quad b_h \in B_h$$

and verify

$$\forall \delta s_h \in V_h \quad f(\delta s_h) = a(s_h + b_h, \delta s_h) = a(s_h, \delta s_h)$$

3

and

$$\forall \delta b_h \in B_h \quad f(\delta b_h) = a(s_h + b_h, \delta b_h) = a(b_h, \delta b_h)$$

so that both problems are decoupled.

3. Nodal superconvergence

It is often accepted that in multidimensional problems, a special accuracy on the displacements is obtained *at the nodes*. We do not know any correct proof of such a phenomenon. Let us citate a classical argument due to Dupont and Douglas [6]. Following these authors, let $G_o(x)$ be the Riesz representator of the Dirac measure at some point x_o,

$$f_o(v) = v(x_o).$$

It is to say that

$$a(G_o, v) = f_o(v) = v(x_o)$$

Then,

$$u(x_o) - u_h(x_o) = f_o(u - u_h) = a(G_o, u - u_h)$$

and for any $v_h \in V_h$,

$$u(x_o) - u_h(x_o) = a(G_o - v_h, u - u_h) \le A\|G_o - v_h\|_v \|u - u_h\|_v$$

so that

$$|u(x_o) - u_h(x_o)| \le A \inf_{v_h \in V_h} \|G_o - v_h\|_v \|u - u_h\|_v.$$

"Normally", say STRANG & FIX [6], "the term $\|G_o - v_h\|_v$ will add some finite power of h to the h^{k+1-m} coming from $\|u - u_h\|_v$".

Unfortunately, the Dirac measure is *not bounded* in $H^1(\Omega)$ when the space dimension is greater than 1... and the argument does not hold. It is bounded for the plate problem, but the corresponding solution is in $H^2(\Omega)$, no more. Therefore,

$$|u(x_o) - u_h(x_o)| \le o(h^{k+1-m}) \text{ in place of } O(h^{k+1-m}), \text{ that is, for third degree plate}$$
elements,

$$|u(x_o) - u_h(x_o)| \le o(h^2)$$

4

4. Stress points

Starting with the nodal superconvergence, *stress points* of an element of degree k are defined as those points where the element, when interpolating a polynomial of degree (k+1), leads to the exact stresses.

a) First degree rectangle

For a second degree polynomial,

$$u = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2$$

the interpolate is of the form

$$\Pi u = \beta_1 + \beta_2 x + \beta_3 y + \beta_4 xy .$$

It is easy to see that the difference is of the form

$$\Delta u = \gamma_1 (x^2 - b^2) + \gamma_2 (y^2 - b^2)$$

So,

$$\frac{\partial \Delta u}{\partial x} = 2\gamma_1 x, \quad \frac{\partial \Delta u}{\partial y} = 2\gamma_2 y$$

so that the stress point is the center of the element.

b) Second degree rectangle

Here, the interpolation formula contains 8 parameters. A third degree polynomial being of the form

$$u = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2 + \alpha_7 x^3 + \alpha_8 x^2 y + \alpha_9 xy^2 + \alpha_{10} y^3$$

contains 10 parameters. Therefore, Δu must have two parameters. It is easy to see that

$$\Delta u = \gamma_1 x(x^2 - a^2) + \gamma_2 y(y^2 - b^2)$$

So,

$$\frac{\partial \Delta u}{\partial x} = \gamma_1 (3x^2 - a^2) = 0 \ at \ x = \pm \frac{a}{\sqrt{3}}$$

5

$$\frac{\partial \Delta u}{\partial y} = \gamma_2 (3y^2 - b^2) = 0 \; at \; y = \pm \frac{b}{\sqrt{3}},$$

which are the well known *Gauss points*.

c) Triangular element

Here, the preceeding approach leads to a stress point which depends on the solution ! It is generally accepted that the stress point of a linear triangle is its center of gravity.

5. Stress points : another approach

5.1. General idea

Finite elements lead to a best mean square approximation of stresses. For sufficiently refined meshes, it is rational to admit that the *elemental* mean square approximation is nearly the best one.

So, consider an isolated element of degree k. Consider any function of degree (k+1), and find its best approximation by the element in the L^2 sense for the derivatives. The points where the function and its approximation coincide are stress points, where superconvergence occurs.

5.2. The first degree triangle

Let us consider a first degree triangle. The coordinates will be chosed such that their origin is the gravity center of the triangle.

A general function of degree 2 is of the form

$$u = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2$$

and its derivatives are

$$\frac{\partial u}{\partial x} = \alpha_2 + 2\alpha_4 x + \alpha_5 y$$

$$\frac{\partial u}{\partial y} = \alpha_3 + \alpha_5 x + 2\alpha_6 y$$

Now, in the element, the considered function is

6

$$v = \beta_1 + \beta_2 x + \beta_3 y$$

and

$$\frac{\partial v}{\partial x} = \beta_2, \quad \frac{\partial v}{\partial y} = \beta_3$$

Substracting, one obtains the following form

$$u - v = \gamma_1 + \gamma_2 x + \gamma_3 y + \gamma_4 x^2 + \gamma_5 xy + \gamma_6 y^2$$

and

$$\frac{\partial (u-v)}{\partial x} = \gamma_2 + 2\gamma_4 x + \gamma_5 y$$

$$\frac{\partial (u-v)}{\partial y} = \gamma_3 + \gamma_5 x + 2\gamma_6 y$$

Now, the condition of best approximation of u by v is

$$\int_\Omega \left(\frac{\partial (u-v)}{\partial x} \frac{\partial \delta v}{\partial x} + \frac{\partial (u-v)}{\partial y} \frac{\partial \delta v}{\partial y} \right) dxdy = 0,$$

that is

$$\delta \beta_2 \to \delta \beta_2 \int_\Omega \left(\gamma_2 + 2\gamma_4 x + \gamma_5 y \right) = 0 \Rightarrow \gamma_2 = 0$$

$$\delta \beta_3 \to \delta \beta_3 \int_\Omega \left(\gamma_3 + \gamma_5 x + 2\gamma_6 y \right) = 0 \Rightarrow \gamma_3 = 0$$

It follows that

$$\frac{\partial (u-v)}{\partial x} = 2\gamma_4 x + \gamma_5 y$$

$$\frac{\partial (u-v)}{\partial x} = \gamma_5 x + 2\gamma_5 y$$

and both expressions vanish at the centroid x = 0, y = 0, which is the stress point.

5.3. The bilinear rectangle

The element approximation is

$$v = \beta_1 + \beta_2 x + \beta_3 y + \beta_4 xy$$

and

$$\frac{\partial v}{\partial x} = \beta_2 + \beta_4 y \quad \frac{\partial v}{\partial y} = \beta_3 + \beta_4 x$$

For the same expression of u as in 5.2, the approximation conditions are now

$$\delta\beta_2 \to \int_\square \frac{\partial(u - v)}{\partial x} dx\, dy = 0 \Rightarrow \gamma_2 = 0$$

$$\delta\beta_3 \to \int_\square \frac{\partial(u - v)}{\partial y} dx\, dy = 0 \Rightarrow \gamma_3 = 0$$

$$\delta\beta_4 \to 0 = \int_\square \left[y\frac{\partial(u - v)}{\partial x} + x\frac{\partial(u - v)}{\partial y} \right] dx\, dy = \gamma_5 \int_\square (x^2 + y^2) dx\, dy \Rightarrow \gamma_5 = 0$$

So,

$$\frac{\partial(u - v)}{\partial x} = 2\gamma_4 x$$

$$\frac{0(u - v)}{\partial y} = 2\gamma_6 y$$

and both vanish at the center of the element.

5.4. The biquadratic rectangle

The element field is here

$$v = \beta_1 + \beta_2 x + \beta_3 y + \beta_4 x^2 + \beta_5 xy + \beta_6 y^2 + \beta_8 x^2 y + \beta_9 xy^2 + \beta_{11} x^2 y^2$$

It will be compared to a full cubic,

$$u = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2 + \alpha_7 x^3 + \alpha_8 x^2 y + \alpha_9 xy^2 + \alpha_{10} y^3$$

The difference is of the form

8

$$u - v = \gamma_1 + \gamma_2 x + \gamma_3 y + \gamma_4 x^2 + \gamma_5 xy + \gamma_6 y^2 + \gamma_7 x^3 + \gamma_8 x^2 y + \gamma_9 xy^2$$

Let us first compute the derivatives,

$$\begin{cases} \dfrac{\partial v}{\partial x} = \beta_2 + 2\beta_4 x + \beta_5 y + 2\beta_8 xy + \beta_9 y^2 + 2\beta_{11} xy^2 \\[2mm] \dfrac{\partial v}{\partial y} = \beta_3 + \beta_5 x + 2\beta_6 y + \beta_8 x^2 + 2\beta_9 xy + 2\beta_{11} x^2 y \end{cases}$$

and

$$\begin{cases} \dfrac{\partial(u-v)}{\partial x} = \gamma_2 + 2\gamma_4 x + \gamma_5 y + 3\gamma_7 x^2 + 2\gamma_8 xy + \gamma_9 y^2 + 2\gamma_{11} xy^2 \\[2mm] \dfrac{\partial(u-v)}{\partial y} = \gamma_3 + \gamma_5 x + 2\gamma_6 y + \gamma_8 x^2 + 2\gamma_9 xy + 3\gamma_{10} y^2 + 2\gamma_{11} x^2 y \end{cases}$$

Introducing the notation

$$I_{kl} = \int\limits_{\Omega} x^k y^l \, dx dy \,,$$

the approximation conditions are as follows

$$\delta\beta_2 \rightarrow \gamma_2 I_{00} + 3\gamma_7 I_{20} + \gamma_9 I_{02} + = 0 \qquad\qquad \text{(a)}$$

$$\delta\beta_3 \rightarrow \gamma_3 I_{00} + \gamma_8 I_{20} + 3\gamma_{10} I_{02} + = 0 \qquad\qquad \text{(b)}$$

$$\delta\beta_4 \rightarrow 2\gamma_4 I_{20} + 2\gamma_{11} I_{22} = 0 \qquad\qquad \text{(c)}$$

$$\delta\beta_5 \rightarrow \gamma_5 (I_{02} + I_{20}) = 0 \qquad\qquad \text{(d)}$$

$$\delta\beta_6 \rightarrow 2\gamma_6 I_{02} + 2\gamma_{11} I_{22} = 0 \qquad\qquad \text{(e)}$$

$$\delta\beta_8 \rightarrow 4\gamma_8 I_{22} + \gamma_3 I_{20} + \gamma_8 I_{40} + 3\gamma_{10} I_{22} = 0 \qquad\qquad \text{(f)}$$

$$\delta\beta_9 \rightarrow \gamma_2 I_{02} + 3\gamma_7 I_{22} + \gamma_9 I_{04} + 2\gamma_9 I_{22} = 0 \qquad\qquad \text{(g)}$$

$$\delta\beta_{11} \rightarrow 2\gamma_4 I_{22} + 2\gamma_{11} I_{24} + 2\gamma_6 I_{22} + 2\gamma_{11} I_{42} = 0 \qquad\qquad \text{(h)}$$

Note that the terms which are not present in u are β_7 and β_{10}. Therefore, coefficients which are not related to these may be set to zero. Let us try a solution where

$$\gamma_{11} = \gamma_9 = \gamma_8 = 0$$

Then, from (c), $\gamma_4 = 0$,
  from (d), $\gamma_5 = 0$
  from (e), $\gamma_6 = 0$

and (h) is verified identically. It remains only $\gamma_2, \gamma_3, \gamma_7, \gamma_{10}$.

From (a),

$$0 = \gamma_2 I_{00} + 3\gamma_7 I_{20} = \gamma_2 \cdot 4ab + 3\gamma_7 \cdot \frac{4a^3 b}{3} \Rightarrow \gamma_2 = -\gamma_7 a^2$$

This is consistent with (g),

$$0 = \gamma_2 I_{02} + 3\gamma_7 I_{22} = \gamma_2 \cdot \frac{4ab^3}{3} + 3\gamma_7 \cdot \frac{4a^3 b^3}{9}$$

In a similar way, one obtains from (b)

$$\gamma_3 = -\gamma_{10} b^2$$

and this is consistent with (f).

So, the solution is

$$\frac{\partial(u-v)}{\partial x} = 3\gamma_7 \left(x^2 - \frac{a^2}{3} \right)$$

$$\frac{\partial(u-v)}{\partial y} = 3\gamma_{10} \left(y^2 - \frac{b^2}{3} \right)$$

and vanishes at the Gauss points $x = \pm \dfrac{a}{\sqrt{3}}, y = \pm \dfrac{b}{\sqrt{3}}$.

This conforms with the result obtained in section 4. But it has to be mentioned that

$$u - v = \gamma_1 + \gamma_7 (x^3 - a^2 x) + \gamma_{10} (y^3 - b^2 x)$$

is not zero at the boundary, but parallel to this solution. The interpolation condition is thus not necessary in order to obtain stress points.

REGULARITY OF THE SOLUTION

The question of the regularity of the solution is crucial from the point of view of convergence. In fact, a full $O(h^{k+1-m})$ order is obtained only in cases where $u \in H^{k+1}$.

a) *Interior regularity*. It may be proved that if the coefficients of the equation are analytic and if the second member f is analytic, then, the solution is analytic in any

$$\Omega_\varepsilon = \{x \in \Omega | d(x, C\Omega)\} > \varepsilon, \qquad \varepsilon > 0$$

This property is a characteristic of elliptic problems (cf. LIONS & MAGENES).
So, a loss of regularity may be found

- in the vicinity of the boundary
- when the load is not regular

b) *Regular problems*
An elliptic second order problem

$$a(u, v) = \int_\Omega fu \, dx$$

is said to be regular if

$$f \in L^2(\Omega) \Rightarrow u \in H^2(\Omega).$$

This is the case of the equation

$$\sum_{ij} D_i(a_{ij} D_j u) + a_o u = f$$

with sufficiently smooth a_{ij} and a_o, and homogeneous Dirichlet or Neumann conditions on a *convex* set with Lipschitz-continuous boundary (*Not for mixed problems* !).

c) *Necessity of the convexity*

In order to prove the necessity of the convexity, let us examine the problem

$$-\nabla^2 u = f, \quad f \in L^2(\Omega)$$

on a circular sector or radius R and angle α. Dirichlet conditions u = 0 are assumed on the boundary. In order to find the solution, let us write

$$u = \sum_{k=1}^{\infty} u_k(r) \sin \frac{k\pi\theta}{\alpha}$$

One has

$$\int_{\Omega} |grad\ u|^2\ dx = \int_o^\alpha d\theta \int_o^R \left[\left(\frac{\partial u}{\partial r} \right)^2 + \left(\frac{1}{r} \frac{\partial u}{\partial \theta} \right)^2 \right] rdr$$

and

$$\frac{\partial u}{\partial r} = \sum_k u'_k \sin \frac{k\pi\theta}{\alpha}$$

$$\frac{1}{r} \frac{\partial u}{\partial \theta} = \sum_k \frac{u_k}{r} \frac{k\pi}{\alpha} \cos \frac{k\pi\theta}{\alpha}$$

So,

$$\int_{\Omega} |grad\ u|^2\ dx = \frac{\alpha}{2} \sum_k \int_o^R \left(u'^2_k + \frac{k^2\pi^2}{d^2} \frac{u_k}{r^2} \right) r\ dr$$

Similarly,

$$\int_{\Omega} f\ u\ dx = \sum_k \int_o^R f_k(r) u_k r\ dr$$

with

$$f_k(r) = \int_o^\alpha f(r,\theta) \sin \frac{k\pi\vartheta}{\alpha} d\theta$$

The problem is thus splitted in the following ones :

$$\frac{1}{2} \frac{\alpha}{2} \int_o^R \left(u'^2_k + \frac{k^2\pi^2}{\alpha^2} \frac{u_k^2}{r^2} \right) r\ dr - \int_o^R f_k u_k r\ dr \quad \underset{u_k}{stat}$$

This leads to

$$\frac{\alpha}{2} \left(-\frac{d}{dr} (ru'_k) + \frac{k^2\pi^2}{\alpha^2} \frac{u_k}{r} \right) = f_k r$$

The corresponding homogeneous equation is

$$-\frac{d}{dr} (ru'_k) + \frac{k^2\pi^2}{\alpha^2} \frac{u_k^2}{r} = 0$$

Let us try a solution of the form

$$u_k = r^l$$

One has

$$u'_k = lr^{l-1}$$
$$ru'_k = lr^l$$

$$\frac{d}{dr}(ru'_k) = l^2 r^{l-1}$$

and the equation writes

$$l^2 r^{l-1} = \frac{k^2 \pi^2}{\alpha^2} r^{l-1}$$

so that

$$l = \pm \frac{k\pi}{\alpha}$$

The solution has to give a finite energy, so that one must have

$$\int_o^R ru'^2_k \, dr < \infty,$$

that is,

$$l^2 \int_o^R r^{2l-1} dr < \infty$$

So, values of l such that (2l-1)>-1 are admitted. This condition writes

2l>0, l>0.

So, only positive powers have to be taken in account.
But in order to obtain $u \in H^2(\Omega)$, it is necessary to ensure

$$\int_o^R r \, u''^2_k \, dr < \infty,$$

and, as

$$u''_k = l(l-1)r^{l-2},$$

this implies

$$l^2(l-1)^2 \int_o^R r^{2l-3} dr < \infty, \quad (*)$$

so that it is necessary that

$$2l - 3 > -1,$$

that is

$$2l > 2, \quad l > 1$$

This is to say

$$l = \frac{k\pi}{\alpha} > 1$$

In the case l = 1, the factor $(l-1)^2$ in (*) suppress the problem. So, the condition is

$$l = \frac{k\pi}{\alpha} \geq 1$$

As can be seen, this not the case with k = 1 when $\alpha > \pi$, that is with a re-entrant angle. So, with a non convex shape, singularities appear, which are not elements of H^2.

Conclusion : an angle α leads to a solution comprising a term of the form

$$r^{\pi/\alpha}$$

which $\in H^2(\Omega)$ only if $\alpha \leq \pi$.