P. Lesaint
P. A. Raviart

On a Finite Element Method for Solving the Neutron Transport Equation

<http://www.numdam.org/item?id=PSMIR_1974___S4_A8_0>

ON A FINITE ELEMENT METHOD FOR SOLVING

THE NEUTRON TRANSPORT EQUATION

by P. LESAINT [*] *and* P.A. RAVIART [**]

[*] Service de Mathématiques Appliquées, Centre d'Etudes de Limeil, B.P. 27, 94190 VILLENEUVE-SAINT-GEORGES FRANCE.

[**] Analyse Numérique, Tour 55-65, Université PARIS VI , 4 Place Jussieu, 75230 PARIS CEDEX 05 , FRANCE.

1 - INTRODUCTION.

Let Ω be a _convex_ open set in the (x,y)-plane with boundary Γ
Denote by n_x and n_y the components of the outward unit vector normal
to Γ .

Let Q be the unit disk in the (μ,ν)-plane. We consider the
following problem : Find a function $u = u(x,y,\mu,\nu)$ such that

$$(1.1) \qquad \mu \frac{\partial u}{\partial x} + \nu \frac{\partial u}{\partial y} + \sigma u = f \qquad \text{in} \qquad \Omega \times Q ,$$

$$(1.2) \qquad u(x,y,\mu,\nu) = 0 \qquad \text{if} \qquad (x,y) \in \Gamma, \ (\mu n_x + \nu n_y)(x,y) < 0 .$$

Equation (1.1) is the neutron transport equation.: the function
$u(x,y,\mu,\nu)$ represents the flux of neutrons at the point (x,y) in th
angular direction (μ,ν) , σ is the nuclear cross section and f
stands for the scattering, the fission and the inhomogeneais source terms.
The boundary condition (1.2) simply means that no neutrons are entering
the system from outside.

In this paper, we shall be only concerned with the spatial discretization
of problem (1.1), (1.2). Thus, we shall assume that the angular direction
(μ,ν) is fixed and we shall consider the reduced problem : Given a function
f defined over Ω , find a function u defined over Ω such that

$$(1.3) \qquad \begin{cases} \mu \dfrac{\partial u}{\partial x} + \dfrac{\partial u}{\partial y} + \sigma u = f \text{ in } \Omega , \\[2mm] u = 0 \text{ on } \Gamma_- , \end{cases}$$

where

$$(1.4) \qquad \Gamma_- = \{ (x,y) \mid (x,y) \in \Gamma, \ (\mu n_x + \nu n_y)(x,y) < 0 \} .$$

This paper will be devoted to the numerical approximation of problem (1.3) by a finite element method using triangular or quadrilateral elements which has been recently introduced by Reed and Hill [17] and which appears to be very effective in practice. Other finite element methods for solving the neutron transport equation have been introduced by several authors (cf. for instance [10] , [14] , [15] ,[16]). We refer to [12] for a mathematical discussion of some of them.

An outline of the paper is as follows. In § 2 , we study a discontinuous Galerkin method for ordinary differential equations using polynomials od degree k . This Galerkin method is shown to be strongly A-stable and of order $2k + 1$. In § 3 , we introduce the finite element method as a generalization of the discontinuous Galerkin method of § 2 . We prove the existence and uniqueness of the approximate solution and we give an algorithm for computing this approximate solution. In § 4 , we derive general error bounds in L^2-norm. Finally, we give in § 5 a super convergence result.

Note that problem (1.3) is a simple but important example of a first-order hyperbolic problem. In fact, the finite element method studied in this paper provides an effective way for numerically solving such problems. For other finite element methods for solving first order systems of partial differential equations, we refer to [11] , [13] .

For the sake of simplicity, we have confined ourselves to polygonal domains Ω . It is probably an easy matter to handle general curved domains by using curved isoparametric elements and the analysis given in [5], [6].

2 - A DISCONTINUOUS GALERKIN METHOD FOR ORDINARY DIFFERENTIAL EQUATIONS

We begin by studying the numerical solution of the ordinary differential equation

$$(2.1) \quad \begin{cases} u'(x) = f(x,u(x)) \quad , \quad x \geqslant x_o \quad , \\ u(x_o) = u_o \quad , \end{cases}$$

on a finite interval $[x_o , x_o + a]$ by a _discontinuous Galerkin method_. For _continuous_ Galerkin methods and related collocation methods, we refer for instance to Axelsson [1] , de Boor and Swartz [2] , Hulme [9] .

Let $x_n = x_o + nh$, $0 \leqslant n \leqslant N$ (Nh = a) be a uniform mesh for the sake of simplicity. Then, we may approximate u on each subinterval $[x_n , x_{n+1}]$ by a _kth degree polynomial_ u_h . We require that u_h satisfies on each subinterval $[x_n , x_{n+1}]$, $0 \leqslant n \leqslant N-1$:

$$(2.2) \quad \begin{cases} (u_h(x_{n+}) - u_h(x_{n-}))v(x_n) + \\ \quad + \int_{x_n}^{x_{n+1}} \{u_h'(x) - f(x,u_h(x))\} \, v(x) \, dx = 0 \quad \text{for all} \quad v \in P_k \end{cases}$$

with the initial condition

$$(2.3) \quad u_h(x_{o-}) = u_o \quad ,$$

where P_k denotes the space of all polynomials of degree $\leqslant k$. Notice that the function u_h _is discontinuous in general at the mesh points_ x_n .

To obtain a computational form of (2.2)-(2.3), we replace the integral in (2.2) by an interpolatory quadrature formula

$$(2.4) \qquad \int_{x_n}^{x_{n+1}} \varphi(x) \, dx = h \sum_{i=1}^{k+1} b_i \, \varphi(x_{n,i}) + O(h^{p+1}) \, ,$$

$$(2.5) \qquad x_{n,i} = x_n + \xi_i h \, , \quad 1 \leqslant i \leqslant k+1 \, , \quad \xi_1 = 0 \, ,$$

where b_i and ξ_i are the weights and abscissae for $[0,1]$.

Notice that $k+1 \leqslant p \leqslant 2k+1$. Then (2.2) becomes :

$$(2.6) \qquad \begin{aligned} &(u_h(x_{n+}) - u_h(x_{n-}))v(x_n) + \\ &\quad + h \sum_{i=1}^{k+1} b_i \, \{u_h'(x_{n,i}) - f(x_{n,i} \, , \, u_h(x_{n,i}))\} \, v(x_{n,i}) = 0 \end{aligned}$$

$$\text{for all } v \in P_k$$

Let us now show that the discrete Galerkin method (2.3), (2.6) is equivalent to some implicit Runge-Kutta method. We define

$$(2.7) \qquad \begin{cases} u_n = u_h(x_{n-}) \, , \\ u_{n,1} = u_h(x_{n+}) = u_h(x_{n,1}) \, , \\ u_{n,i} = u_h(x_{n,i}) \, , \quad 2 \leqslant i \leqslant k+1 \end{cases}$$

We introduce the Lagrange interpolation coefficients

$$(2.8) \qquad \ell_i(x) = \prod_{\substack{j=2 \\ j \neq i}}^{k+1} \frac{x - \xi_j}{\xi_i - \xi_j} \, , \quad 2 \leqslant i \leqslant k+1 \, .$$

Lemma 1   The discrete Galerkin method (2.3),(2.6) is equivalent to the following implicit Runge-Kutta method

$$(2.9) \qquad \begin{cases} u_{n,i} = u_n + h \sum_{j=1}^{k+1} a_{ij} \, f(x_{n,j} \, , \, u_{n,j}) \, , \quad 1 \leqslant i \leqslant k+1 \, , \\ u_{n+1} = u_n + h \sum_{j=1}^{k+1} b_j \, f(x_{n,j} \, , \, u_{n,j}) \, , \end{cases}$$

where

$$(2.10) \quad \begin{cases} a_{i1} = b_1 \quad , \quad 1 \leqslant i \leqslant k+1 \ , \\ \\ a_{ij} = \displaystyle\int_0^{\xi_i} \ell_j(x) \, dx - b_1 \ell_j(\xi_1) \ , \quad 1 \leqslant i \leqslant k+1 \ , \quad 2 \leqslant j \leqslant k+1 \end{cases}$$

Proof   Let us introduce the basics $\{ v_i \}_{1 \leqslant i \leqslant k+1}$ of the space P_k defined by

$$v_i(x_{n,j}) = \delta_{ij} \cdot \ , \quad 1 \leqslant i,j \leqslant k+1$$

By replacing successively in (2.6) v by v_i , we find that an equivalent form of (2.6) is given by

$$(2.11) \quad \begin{cases} u_h(x_{n+}) - u_h(x_{n-}) + hb_1(u_h'(x_{n,1}) - f(x_{n,1} , u_h(x_{n,1})) = 0 \\ \\ u_h'(x_{n,i}) - f(x_{n,i} , u_h(x_{n,i})) = 0 \ , \quad 2 \leqslant i \leqslant k+1 \end{cases}$$

In the subinterval $[x_n , x_{n+1}]$, we have $u_h' \in P_{k-1}$ so that

$$u_h'(x) = \sum_{j=2}^{k+1} \ell_j \left(\frac{x-x_n}{h} \right) u_h'(x_{n,j})$$

and by (2.11)

$$(2.12) \quad u_h'(x) = \sum_{j=2}^{k+1} \ell_j \left(\frac{x-x_n}{h} \right) f(x_{n,j} , u_h(x_{n,j})) \ .$$

Taking $x = x_n = x_{n,1}$ in (2.12) , substituting this expression in the 1st equation (2.11) and using (2.7) , we obtain

$$(2.13) \quad u_{n,1} = u_n + hb_1 \{ f(x_{n,1} , u_{n,1}) - \sum_{j=2}^{k+1} \ell_j(\xi_1) f(x_{n,j} , u_{n,j}) \} \ .$$

On the other hand, we may write for $2 \leqslant i \leqslant k+1$

$$u_h(x_{n,i}) = u_h(x_{n,1}) + \int_{x_{n,1}}^{x_{n,i}} u_h'(x) \, dx$$

and by (2.7), (2.12), (2.13)

$$(2.14) \qquad u_{n,i} = u_n + h \{ b_1 \, f(x_{n,1} , u_{n,1}) + \sum_{j=2}^{k+1} (\int_0^{\xi_i} \ell_j(x) \, dx - b_1 \ell_j(\xi_1)) $$
$$f(x_{n,j} , u_{n,j}) \}$$

Similarly, we have

$$u_h(x_{n+1}-) = u_h(x_{n,1}) + \int_{x_n}^{x_{n+1}} u'(x) \, dx$$

and then

$$u_{n+1} = u_n + h \{ b_1 \, f(x_{n,1} , u_{n,1}) + \sum_{j=2}^{k+1} (\int_0^1 \ell_j(x) \, dx - b_1 \ell_j(\xi_1)) $$
$$f(x_{n,j} , u_{n,j}) \}$$

By noticing that

$$\int_0^1 \ell_j(x) \, dx = \sum_{i=1}^{k+1} b_i \ell_j(\xi_i) = b_1 \ell_j(\xi_1) + b_j \quad ,$$

we get

$$(2.15) \qquad u_{n+1} = u_n + h \sum_{j=1}^{k+1} b_j (x_{n,j} , u_{n,j}) \; .$$

The equations (2.13) - (2.15) are identical to the equations (2.9),(2.10). We then have proved that the discrete Galerkin method leads to the one-step method (2.9), (2.10). Conversely, the Runge-Kutta method (2.9), (2.10) can be clearly viewed as a discrete Galerkin method.

Theorem 1   **The discrete Galerkin method (2.3), (2.6) is a one-step method of order p .**

Proof   Following Butcher [3] , Crouzeix [7] , we know that the conditions

$$(2.16) \qquad \sum_{j=1}^{k+1} b_j \, \xi_j^{\ell} = \frac{1}{\ell+1} \quad , \quad 0 \leqslant \ell \leqslant p-1 \quad ,$$

$$(2.17) \qquad \sum_{j=1}^{k+1} a_{ij} \, \xi_j^{\ell} = \frac{\xi_i^{\ell+1}}{\ell+1} \quad , \quad 0 \leqslant \ell \leqslant k-1 \quad , \quad 1 \leqslant i \leqslant k+1$$

$$(2.18) \qquad \sum_{i=1}^{k+1} b_i a_{ij} \, \xi_i^{\ell} = \frac{1}{\ell+1} b_j (1- \xi_j^{\ell+1}) \quad , \quad k+\ell \leqslant p-1 \quad , \quad 1 \leqslant j \leqslant k+1 \quad ,$$

are sufficient for the Runge-Kutta method (2.9) to be of order p .

Let us show that these conditions hold in the present case.

First, conditions (2.16) simply mean that the interpolatory quadrature formula (2.4) is exact for all polynomials of degree $\leqslant$ p-1.

Next, consider conditions (2.17). Using (2.8) , we may write

$$x^{\ell} = \sum_{j=2}^{k+1} \ell_j(x) \, \xi_j^{\ell} \quad , \quad 0 \leqslant \ell \leqslant k-1 \quad ,$$

so that

$$\xi_1^{\ell} = \sum_{j=2}^{k+1} \ell_j(\xi_1) \, \xi_j^{\ell} \quad , \quad 0 \leqslant \ell \leqslant k-1 \quad ,$$

$$\frac{\xi_i^{\ell+1}}{\ell+1} = \sum_{j=2}^{k+1} \left(\int_0^{\xi_i} \ell_j(x) \, dx \right) \xi_j^{\ell} \quad , \quad 0 \leqslant \ell \leqslant k-1 \quad , \quad 1 \leqslant i \leqslant k+1$$

Using (2.10) , we have

$$\sum_{j=1}^{k+1} a_{ij} \, \xi_j^\ell = b_1 (\xi_1^\ell - \sum_{j=2}^{k+1} \ell_j (\xi_1) \, \xi_j^\ell) + \sum_{j=2}^{k+1} (\int_0^{\xi_i} \ell_j(x) \, dx) \, \xi_j^\ell$$

and by the previous relations

$$\sum_{j=1}^{k+1} a_{ij} \, \xi_j^\ell = \frac{\xi_i^{\ell+1}}{\ell+1} \, , \quad 0 \leqslant \ell \leqslant k-1 \, , \quad 1 \leqslant i \leqslant k+1$$

Finally, let us show that conditions (2.18) hold . We begin by noticing that

$$(2.19) \qquad \sum_{i=1}^{k+1} b_i a_{i1} \xi_i^\ell = b_1 \sum_{i=1}^{k+1} b_i \, \xi_i^\ell = \frac{b_1}{\ell+1} \, , \quad 0 \leqslant \ell \leqslant p-1 \, .$$

On the other hand, following Crouzeix [7], we may write for all continuous function φ

$$(2.20) \qquad \int_0^1 x^\ell (\int_0^x \varphi(y) \, dy) \, dx = \frac{1}{\ell+1} \int_0^1 (1-x^{\ell+1}) \, \varphi(x) \, dx \, .$$

Taking $\varphi \in P_{k-1}$, we obtain for $k+\ell \leqslant p-1$

$$\int_0^1 x^\ell (\int_0^x \varphi(y) \, dy) \, dx = \sum_{i=1}^{k+1} b_i \xi_i^\ell \int_0^{\xi_i} \varphi(y) \, dy =$$

$$= \sum_{i=1}^{k+1} b_i \xi_i^\ell \sum_{j=2}^{k+1} (\int_0^{\xi_i} \ell_j(y) \, dy) \, \varphi(\xi_j)$$

and by (2.10)

$$(2.21) \qquad \int_0^1 x^\ell (\int_0^x \varphi(y) \, dy) \, dx = \sum_{i,j=1}^{k+1} b_i \, a_{ij} \, \xi_i^\ell \, \varphi(\xi_j) \, , \, \varphi \in P_{k-1} \, , \, k+\ell \leqslant p-1$$

Similarly, we get

$$(2.22) \qquad \int_0^1 (1-x^{\ell+1})\varphi(x) \, dx = \sum_{j=1}^{k+1} b_j(1-\xi_j^{\ell+1})\varphi(\xi_j), \quad \varphi \in P_{k-1}, \quad k+\ell \leqslant p-1.$$

Hence, combining (2.19) - (2.22) , we have for all $\varphi \in P_{k-1}$ and for $k+\ell \leqslant p-1$

$$\sum_{j=1}^{k+1} \left(\sum_{i=1}^{k+1} b_i a_{ij} \xi_i^\ell - \frac{1}{\ell+1} b_j(1-\xi_j^{\ell+1}) \right)\varphi(\xi_j) = 0$$

This implies

$$\sum_{i=1}^{k+1} b_i a_{ij} \xi_i^\ell = \frac{1}{\ell+1} b_j(1-\xi_j^{\ell+1}), \quad k+\ell \leqslant p-1, \quad 2 \leqslant j \leqslant k+1 \qquad \square$$

For investigating the stability properties of the one-step method (2.9) , we consider the differential equation

$$(2.23) \qquad u' = \lambda u$$

where λ is a complex constant with $\mathrm{Re}(\lambda) < 0$.

Lemma 2.   Applied to the differential equation (2.23) , the one-step method (2.9) , (2.10) gives

$$(2.24) \qquad u_{n+1} = R(\lambda h)u_n$$

where $R(z) = \dfrac{P(z)}{Q(z)}$ is the quotient of two polynomials $P(z)$ and $Q(z)$ of degree $\leqslant k$ and $k+1$ respectively.

Proof   Applied to (2.23), the one-step method (2.9) becomes (2.25)

(2.25) $u_{n,i} = u_n + \lambda h \sum_{j=1}^{k+1} a_{ij} u_{n,j}$, $1 \leqslant i \leqslant k+1$,

(2.26) $u_{n+1} = u_n + \lambda h \sum_{j=1}^{k+1} b_j u_{n,j}$.

Using obvious notations, we may write equations (2.25) in the form

$$(I - \lambda h [a_{ij}]) [u_{n,i}] = u_n [1]$$

where the identity matrix I and $[a_{ij}]$ are $(k+1) \times (k+1)$-matrices.

Since $a_{i1} = b_1$, $1 \leqslant i \leqslant k+1$, we get from Cramer's rule

$$u_{n,i} = \frac{P_i(\lambda h)}{Q(\lambda h)} u_n$$, $1 \leqslant i \leqslant k+1$

where $P_1(z)$ is a polynomial of degree k whose leading coefficient

is $b_1^{-1} \det(a_{ij})$, $P_i(z)$, $2 \leqslant i \leqslant k+1$, are polynomials of degree $\leqslant k-1$

and where $Q(z)$ is a polynomial of degree k+1 whose leading

coefficient is $\det(a_{ij})$.

Using (2.26) , we obtain

$$u_{n+1} = \frac{P(\lambda h)}{Q(\lambda h)} u_n$$

where

$$P(z) = Q(z) - z \sum_{j=1}^{k+1} P_j(z) .$$

Clearly, in P(z), the coefficient of z^{k+1} vanishes . The lemma is

then proved.   □

Let us now recall the following definition. A one-step method

is <u>strongly A-stable</u> if

$$(2.27) \quad \begin{cases} |R(z)| < 1 & \text{for} \quad Re(z) < 0 \\ |R(z)| \to 0 & \text{as} \quad Re(z) \to -\infty \end{cases}$$

Theomem 2   The Galerkin method (2.2) , (2.3) is a strongly A-stable one-step method of order 2k+1 .

Proof   Consider first the discrete Galerkin method (2.3) , (2.6) associated with the Gauss-Radau abcissae ξ_i , $1 \leqslant i \leqslant k+1$ ($\xi_1 = 0$). Then, we have $p = 2k+1$ in (2.4). By Theorem 1 , this discrete Galerkin method is a one-step method of order 2k+1 so that

$$R(z) = \exp(z) + O(z^{2k+2}) .$$

Moreover, by lemma 2 , $R(z)$ is the quotient of two polynomials $P(z)$ and $Q(z)$ of degree $\leqslant k$ and $\leqslant k+1$ respectively. Then, necessarily , $R(z)$ is the subdiagonal (k+1,k) Padé rational approximation of $\exp(z)$. Using a result of Axelsson [1] , we know that such a Padé approximation satisfies conditions (2.27). Hence, the discrete Galerkin method (2.3) , (2.6) associated with the Gauss-Radau abscissae is a strongly A-stable one-step method of order 2k+1 .

Now, it is a simple but lengthy matter to prove that the Galerkin method (2.2) , (2.3) and the Gauss-Radau discrete Galerkin method (2.3) (2.6) are one-step methods of the same order 2k+1 . Moreover, these two methods coincide when applied to the differential equation (2.23). This completes the proof of the theorem.   �‫

3 - A FINITE ELEMENT METHOD FOR THE NEUTRON TRANSPORT EQUATION

Consider now our neutron transport problem (1.3). First, we need some notations. Let us denote by $L^2(\Omega)$ the space of real-valued functions v which are square integrable over Ω. We provide $L^2(\Omega)$ with the usual norm

$$(3.1) \qquad \| v \|_{0,\Omega} = \left(\int_\Omega |v(x)|^2 \, dx \right)^{1/2} .$$

Given any integer $m > 0$, let

$$(3.2) \qquad H^m(\Omega) = \{ v \,|\, v \in L^2(\Omega) \,,\, \partial^\alpha v \in L^2(\Omega) \,,\, |\alpha| \leqslant m \}$$

be the usual Sobolev space provided with the norm

$$(3.3) \qquad \| v \|_{m,\Omega} = \left(\sum_{|\alpha| \leqslant m} \| \partial^\alpha v \|_{0,\Omega}^2 \right)^{1/2}$$

In (3.2), (3.3), $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$ is a multiindex, $|\alpha| = \alpha_1 + \alpha_2$

and

$$\partial^\alpha = \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \left(\frac{\partial}{\partial x_2} \right)^{\alpha_2}$$

We shall also use the following semi-norm

$$(3.4) \qquad |v|_{m,\Omega} = \left(\sum_{|\alpha| = m} \| \partial^\alpha v \|_{0,\Omega}^2 \right)^{1/2}$$

Let us introduce the operator

$$(3.5) \qquad A = \mu \frac{\partial}{\partial x} + \nu \frac{\partial}{\partial y} + \sigma$$

and the space

$$(3.6) \qquad D(A) = \{ v \,|\, v \in L^2(\Omega) \,,\, \mu \frac{\partial v}{\partial x} + \nu \frac{\partial v}{\partial y} \in L^2(\Omega) \}$$

Then, as a consequence of [8], we have the following result.

Theorem 3   Assume that $\sigma \in L^{\infty}(\Omega)$ and $f \in L^{2}(\Omega)$. Then, problem (1.3) has a unique strong solution $u \in D(A)$.

Using the change of unknown function

$$u = \exp(\lambda(\frac{x}{\mu} + \frac{y}{\nu}))w ,$$

equation (1.3) becomes

$$\mu \frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} + (\sigma+\lambda)u = \exp(-\lambda(\frac{x}{\mu} + \frac{y}{\nu}))f .$$

Thus, by eventually changing $\sigma(x,y)$ into $\sigma(x,y) + \lambda$, we can confine ourselves to the case where σ is positive. More precisely, we shall assume in all the sequel that

(3.7) $M \geqslant \sigma(x,y) \geqslant \alpha > 0$ a.e. in Ω .

Let us now generalize the one-dimensional discontinuous Galerkin method of § 2 to our two-dimensional neutron transport problem. For the sake of simplicity, we shall assume in the following that Ω is a polygon. In order to approximate problem (1.3) , we first construct a triangulation $\mathcal{C}_h$ of $\overline{\Omega}$ with triangles and convex quadrilaterals K with diameters $\leqslant h$. With any $K \in \mathcal{C}_h$, we associate a finite-dimensional space P_K of real-valued functions defined over K such that

(3.8) $P_K \subset H^1(K)$.

We then consider the finite-dimensional space

(3.9) $V_h = \{v | v \in L^2(\Omega) , v_{|K} \in P_K \text{ for all } K \in \mathcal{C}_h \}$.

It is worthwhile to notice that in general <u>a function $v \in V_h$ does</u>

<u>not satisfy any continuity requirement at the interelement</u>

<u>boundaries</u>.

Let $K \in \mathscr{C}_h$ and let ∂K be the boundary of K . We set :

$$(3.10) \begin{cases} \partial_- K = \{(x,y) \mid (x,y) \in \partial K , (\mu n_x + \nu n_y)(x,y) < 0\} , \\ \partial_+ K = \{(x,y) \mid (x,y) \in \partial K , (\mu n_x + \nu n_y)(x,y) > 0\} , \end{cases}$$

where n_x and n_y are the components of the outward unit vector

normal to the boundary ∂K .

Then, the finite element approximation of problem (1.3) that

we shall consider here can be stated as follows. <u>Find a function</u>

 $u_h \in V_h$ such that for all $K \in \mathscr{C}_h$

$$(3.11) \quad \int_{\partial_- K} (\mu n_x + \nu n_y)(u_h - \xi_h) v \, ds + \int_K (A u_h - f) v \, dxdy = 0 \quad \forall v \in P_K$$

where

$$(3.12) \quad \xi_h = \begin{cases} 0 & \underline{on} \ \partial_- K \cap \Gamma_- \\ \\ \underline{outward \ trace \ of} \ u_h \ \underline{on} \ \partial_- K - (\partial_- K \cap \Gamma_-) \end{cases}$$

Clearly, this method appears to be the direct generalization of

the discontinuous Galerkin method (2.2) , (2.3).

Before proving existence and uniqueness of the solution $u_h \in V_h$

we shall show that there exists an ordering of the elements of $\mathscr{C}_h$

well suited for numerically solving equations (3.11) , (3.12).

<u>Lemma 3</u>   <u>There exists an ordering</u> $K_1 , K_2 ,\dots, K_I$ <u>of the</u>

<u>elements of</u> $\mathscr{C}_h$ <u>such that, for all</u> $i = 1 ,\dots, I$, <u>each side of</u>

 $\partial_- K_i$ <u>is either a subset of</u> Γ_- <u>or a substet of</u> $\partial_+ K_j$ <u>for some</u> $j < i$.

Proof   Let us introduce first some notations. We shall say

that K is a boundary element if at least one side of ∂K is

a subset of Γ and that K is a semi-boundary element if one

and only one vertex of K belongs to Γ . Let us consider Γ_-
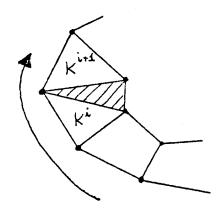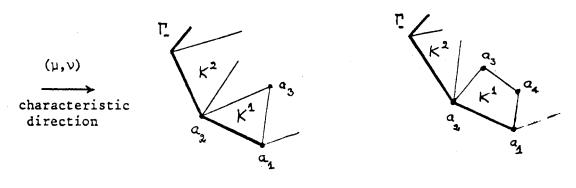
and let us number clockwise the corresponding

Fig. 1

boundary elements $K^1, K^2, \ldots, K^s$. Two consecutive boundary elements

K^i and K^{i+1} can have a common side or not. In the latter case

(cf. Fig. 1) , there exists at least one semi boundary element located

between K^i and K^{i+1} . Then, we shall say that a side of

$K^i (\text{resp.} K^{i+1})$ is semi-common with $K^{i+1} (\text{resp.} K^i)$ if it is a subset

of the union of the semi-boundary elements located betwen K^i and K^{i+1} .

Next, we show that there exists at least one boundary element   K

such that $\partial_- K \subset \Gamma_-$. To prove this, let us assume on the contrary

that $\partial_- K^i \not\subset \Gamma_-$ for all i = 1 , $\ldots$, s and let us show that this

assumption leads to a contradiction.

Consider the first boundary element K^1 and use the notations

of Fig. 2.

(μ,ν)

$\longrightarrow$

characteristic
direction
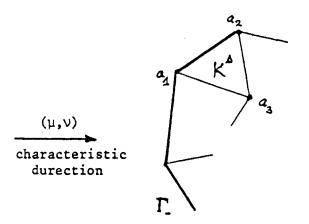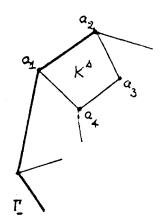
triangular case   quadrilateral case

Fig. 2

In the triangular case (resp. in the quadrilateral case), the side

$[a_1 , a_3]$ (resp. $[a_1 , a_4]$) of K^1 is a subset of $\partial_+ K^1$.

Otherwise, K^1 would not be the first boundary element of Γ_- . Then,

the side $[a_2 , a_3]$ of K^1 which is common or semi-common with

K^2 belongs to $\partial_- K^1$. Otherwise, we should get $\partial_- K^1 = [a_1 , a_2] \subset \Gamma_-$

which is excluded. Therefore, the side of K^2 which is common or

semi-common with K^1 belongs to $\partial_+ K^2$. More generally, we get for

every $i = 1 ,\ldots,$ s-1 the following property : the side of K^i which

is common or semi common with K^{i+1} is a subset of $\partial_- K^i$ and therefore

the side of K^{i+1} which is common or semi-common with K^i is a subset

of $\partial_+ K^{i+1}$. Now consider the last boundary element K^s and use

the notations of Fig. 3

triangular case   quadrilateral case
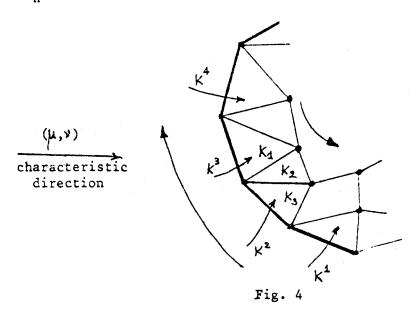
Fig. 3

In the triangular case (resp. in the quadrilateral case) , the side

$[a_1 , a_3]$ (resp. $[a_1 , a_4]$) of K^S is a subset of $\partial_+ K^S$.

Moreover, the side $[a_2 , a_3]$ is a subset of $\partial_+ K^S$. Otherwise,

K^S would not be the last boundary element of Γ_- . Thus, we get

$\partial_- K^S = [a_1 , a_2] \subset \Gamma_-$ which has been excluded. The existence of a

boundary element K such that $\partial_- K \subset \Gamma_-$ is then proved.

Now, choose for K_1 a boundary element of Γ_- such that

$\partial_- K_1 \subset \Gamma_-$ and define $\Omega_1 = \Omega - \Omega \cap K_1$, $\Gamma_{1-} = \partial_- \Omega_1$. Note that

each side of Γ_{1-} is either a subset of Γ_- or a subset of $\partial_+ K_1$.

By the previous argument, there exist a boundary element K_2 of

Γ_{1-} such that $\partial_- K_2 \subset \Gamma_{1-}$, etc ... . Repeating this process, we

take into account all the elements of $\mathcal{C}_h$ and we obtain an ordering

K_1 , K_2 ..., K_I of the elements of $\mathcal{C}_h$ such that the desired property

holds. □

This proof suggests <u>an ordering algorithm</u> for the elements of

$\mathscr{C}_h$ which is effectively used in practise.

(μ,ν)

characteristic
direction

Fig. 4

Consider the sequence K^1, K^2 ,..., K^s of boundary elements of Γ_- .

For K_1 we choose the first element K of this sequence which

satisfies $\partial_- K \subset \Gamma_-$. Let K^p be this element (p=3 in Fig.4).

From $K_1 = K^p$, we then number counterclockwise K_1 , K_2 ,..., K_r

the boundary and semi boundary elements located between K^p and K^1

which satisfy the following condition : for all i = 1 ,..., r , each

side of $\partial_- K_i$ is either a subset of Γ_- or a subset of $\partial_+ K_j$

for some j < i (r=3 in Fig.4). Next, we replace the set Ω by

$\Omega \setminus \overline{\Omega} \cap (\bigcup_{i=1}^{r} K_i)$ and we repeat the process ...

We are now able to prove

Theorem 4   Assume that $f \in L^2(\Omega)$ and that condition (3.7) holds.

Then, there exists a unique function $u_h \in V_h$ which satisfies

equations (3.11) and (3.12) for all $K \in \mathscr{C}_h$.

<u>Proof</u>   Clearly, the finite element method (3.11),(3.12) is

equivalent to a $N \times N$ linear system of equations with $N = \dim V_h$.

Then, it is sufficient to prove the uniqueness of the solution u_h .

Thus, let us assume that $f = 0$ and let us show that necessarily

$u_h = 0$. Let K_1 , K_2 ,..., K_I be an ordering of the elements

$K \in \mathcal{C}_h$ such that the condition of Lemma 3 holds. If $u_h = 0$ in

$K_1 \cup K_2 \cup \ldots \cup K_{i-1}$, then $\xi_h = 0$ on $\partial_- K_i$ and equation (3.11)

becomes in $K = K_i$

$$- \int_{\partial_- K_i} (\mu n_x + \nu n_y) u_h v \, ds + \int_{K_i} (A u_h) v \, dxdy = 0 \qquad \forall v \in P_{K_i}$$

Taking $v = u_h$ and using Green's formula

$$\int_{K_i} (\mu \frac{\partial u_h}{\partial x} + \nu \frac{\partial u_h}{\partial y}) u_h \, dxdy = \frac{1}{2} \int_{\partial K_i} (\mu n_x + \nu n_y) u_h^2 \, ds \quad ,$$

we get

$$\int_{\partial_+ K_i} (\mu n_x + \nu n_y) u_h^2 \, ds - \int_{\partial_- K_i} (\mu n_x + \nu n_y) u_h^2 \, ds + \int_{K_i} \sigma u_h^2 \, dxdy = 0$$

Using (3.7) and (3.10) , we obtain $u_h = 0$ in K_i . Therefore, using

an inductive argument, we get $u_h = 0$ in Ω . □

In practice, the computation of the approximate solution $u_h \in V_h$

goes along the following lines :

(i)   Find an ordering K_1 , K_2 ,..., K_I of the elements $K \in \mathcal{C}_h$

which satisfies the condition of Lemma 3 , for instance by using the

previous algorithm ;

(ii)   Compute successively u_h in K_1 , K_2 ,..., K_I . The computation of u_h in each K_i has a local character and involves the numerical solution of a $d_i \times d_i$ linear system where $d_i = \dim P_{K_i}$.

In other words, by using an ordering of $\mathcal{C}_h$ such that the condition of Lemma 3 holds, the $N \times N$ matrix of the approximate problem becomes block triangular and the ith diagonal block is a $d_i \times d_i$ matrix associated with the ith element K_i .

Note that, in many practical problems, the geometry of Ω and the triangulation $\mathcal{C}_h$ are so simple that step (i) becomes obvious.

4 - GENERAL ERROR BOUNDS

Let us now derive some estimates for the error $u_h - u$ when the solution u of problem (1.3) is smooth enough. We begin with

Lemma 4   For any $K \in \mathscr{C}_h$, any $v \in P_K$ and any function $\eta \in L^2(\partial\_K)$, we have the estimate :

$$
(4.1) \quad \left|
\begin{aligned}
& \tfrac{1}{2}\int_{\partial_+K}(\mu n_x + \nu n_y)(u_h - v)^2 ds + \tfrac{1}{2}\int_{\partial_-K}(\mu n_x + \nu n_y)(\xi_h - \eta)^2 ds - \\
& -\tfrac{1}{2}\int_{\partial_-K}(\mu n_x + \nu n_y)((u_h - v) - (\xi_h - \eta))^2 ds + \int_K \sigma(u_h - v)^2 dxdy = \\
& = \int_{\partial_+K}(\mu n_x + \nu n_y)(u-v)(u_h - v) ds + \int_{\partial_-K}(\mu n_x + \nu n_y)(u-\eta)(u_h - v) ds \\
& + \int_K (u-v)A^*(u_h - v)\, dxdy
\end{aligned}
\right.
$$

where A^* is the formal adjoint of the operator A , i.e.

$$(4.2) \quad A^* = -\mu\frac{\partial}{\partial x} - \nu\frac{\partial}{\partial y} + \sigma .$$

Proof   Given $p \in P_K$ and $\eta \in L^2(\partial\_K)$, we set :

$$(4.3) \quad w = u_h - v \in P_K \quad , \quad \zeta = \xi_h - \eta .$$

Consider the expression

$$(4.4) \quad X_h = -\int_{\partial_-K}(\mu n_x + \nu n_y)(w-\zeta)w\, ds + \int_K (Aw)w\, dxdy$$

First, using Green's formula, we obtain

$$X_h = \frac{1}{2}\int_{\partial K}(\mu n_x + \nu n_y)w^2 ds - \int_{\partial_- K}(\mu n_x + \nu n_y)(w-\zeta)w\, ds +$$

$$+ \int \sigma w^2\, dxdy \ .$$

Since

$$(w-\zeta)w = \frac{1}{2}(w^2-\zeta^2+(w-\zeta)^2) \qquad ,$$

we get

(4.5)
$$\left\{ \begin{aligned} X_h &= \frac{1}{2}\int_{\partial_+ K}(\mu n_x + \nu n_y)w^2 ds + \frac{1}{2}\int_{\partial_- K}(\mu n_x + \nu n_y)\zeta^2 ds - \\ \\ &\quad - \frac{1}{2}\int_{\partial_- K}(\mu n_x + \nu n_y)(w-\zeta)^2 ds + \int_K \sigma w^2 dxdy \end{aligned} \right.$$

On the other hand, using (3.11) , we obtain

$$X_h = \int_{\partial_- K}(\mu n_x + \nu n_y)(v-\eta)w\, ds + \int_K (f-Av)w\, dxdy$$

and therefore

$$X_h = \int_{\partial_- K}(\mu n_x + \nu n_y)(v-\eta)w\, ds + \int_K A(u-v)w\, dxdy$$

Since $u \in D(A)$, we may write

$$\int_K A(u-v)w\, dxdy = \int_K (u-v)A^*w\, dxdy + \int_{\partial K}(\mu n_x + \nu n_y)(u-v)w\, ds$$

so that

(4.6)
$$\left\{ \begin{aligned} X_h &= \int_{\partial_+ K}(\mu n_x + \nu n_y)(u-v)w\, ds + \int_{\partial_- K}(\mu n_x + \nu n_y)(u-\eta)w\, ds + \\ \\ &\quad + \int_K (u-v)A^*w\, dxdy \ . \end{aligned} \right.$$

By combining (4.3) , (4.5) and (4.6) , we get the desired estimate. □

In order to get explicit error bounds, we need to define more precisely the finite-dimensional spaces P_K . Let K be an element of $\mathcal{C}_h$. If K is a triangle, there exists an affine invertible mapping F_K which maps a reference triangle $\hat{K}$ onto K ($\hat{K}$ is usually chosen as a unit isosceles rectangular triangle). If K is a non-degenerate convex quadrilateral, there exists a biaffine invertible mapping F_K which maps the reference element $\hat{K} = [-1, +1]^2$ onto K . Note that this mapping F_K becomes affine when K is a parallelogram.

In both cases, let $\hat{P} \in H^1(\hat{K})$ be a finite-dimensional space of real-valued functions defined over the reference element $\hat{K}$. We shall always assume in the following that

(4.7) $P_K = \{p \mid p = \hat{p} \circ F_K^{-1} , \hat{p} \in \hat{P} \}$.

We shall make a constant use of the one-to-one correspondence

$$\hat{v} \longrightarrow v = \hat{v} \circ F_K^{-1} \quad , \quad v \longrightarrow \hat{v} = v \circ F_K$$

between the functions $\hat{v}$ defined over $\hat{K}$ and the functions v defined over K .

For any integer $m \geqslant 0$, let P_m denote the space of all polynomials of degree $\leqslant m$ in the two variables x,y and let Q_m denote the space of all polynomials of the form

$$p(x,y) = \sum_{i,j=0}^{m} c_{ij} x^i y^j \quad , \quad c_{ij} \in \mathbb{R} .$$

We shall need

Hypothesis H.1. <u>There exists an integer</u> $k \geq 0$ <u>such that :</u>

(4.8) $\qquad P_k \subset \hat{P}$ $\qquad$ if $\hat{K}$ <u>is the reference triangle</u> ,

(4.9) $\qquad Q_k \subset \hat{P}$ $\qquad$ if $\hat{K}$ <u>is the reference quadrilateral</u> $[-1, +1]^2$.

Let us now introduce the following geometrical parameters :

$$(4.10) \quad \begin{cases} h(K) = \text{diameter of } K \\[4pt] \rho(K) = \sup \{ \text{ diameter of the circles contained in } K \} \, , \\[4pt] \theta_i(K) \, , \; 1 \leq i \leq 4 = \text{angles of } K \text{ if } K \text{ is a quadrilateral.} \end{cases}$$

Hypothesis H.2. <u>There exists a constant</u> $\sigma > 1$ <u>independent of</u> h <u>such that</u>

(4.11) $\quad h(K) \leq \sigma\rho(K)$ $\quad$ for all $K \in \mathscr{C}_h$.

Moreover, <u>there exists a constant</u> γ <u>independent of</u> h <u>with</u> $0 < \gamma < 1$ <u>such that</u>

(4.12) $\quad \displaystyle\max_{1 \leq i \leq 4} |\cos \theta_i(K)| \leq \gamma$ $\quad$ for all quadrilateral $K \in \mathscr{C}_h$.

Given a reference element $\hat{K}$, we define $\hat{\Pi}$ to be the orthogonal projection operator in $L^2(\hat{K})$ upon $\hat{P}$. For any $K \in \mathscr{C}_h$, we define $\Pi_K \in \mathscr{L}(L^2(K) \, ; \, P_K)$ by

(4.13) $\quad \widehat{\Pi_K v} = \hat{\Pi} \hat{v}$ $\quad$ for all $v \in L^2(K)$.

Then, for any $v \in L^2(\Omega)$, we define $\Pi_h v$ to be the function in V_h such that

(4.14) $\quad \Pi_h v|_K = \Pi_K v$ $\quad$ for all $K \in \mathscr{C}_h$.

Let us now state some standard results which can be easily proved by using the techniques of Ciarlet & Raviart [4] , [5] .

Lemma 5   Assume that hypothesis H.2 holds. Then, there exists a constant $C > 0$ independent of $K \in \mathscr{C}_h$ such that for all $p \in P_K$

$$(4.15) \qquad |p|_{1,K} \leqslant C(h(K))^{-1} \| p \|_{0,K}$$

$$(4.16) \qquad \| p \|_{0,K'} \leqslant C(h(K))^{-1/2} \| p \|_{0,K} \quad ,$$

where K' is any side of K and $\| p \|_{0,K'} = (\int_{K'} |p|^2 \, ds)^{1/2}$.

Lemma 6 .   Assume that hypotheses H.1 , H.2 and (4.13) hold. Then, there exists a constant $C > 0$ independent of $K \in \mathscr{C}_h$ such that for all $v \in H^{k+1}(K)$

$$(4.17) \qquad |v - \Pi_K v|_{m,K} \leqslant C(h(K))^{k+1-m} \| v \|_{k+1,K} \quad , \quad m = 0,1 \ ,$$

$$(4.18) \qquad \| v - \Pi_K v \|_{0,K'} \leqslant C(h(K))^{k+1/2} \| v \|_{k+1,K}$$

where K' is any side of K .

Let K_1 , K_2 ,..., K_I be a fixed ordering of the elements of $\mathscr{C}_h$ which satisfies the condition of Lemma 3 . For all $i = 1$,..., I , we set

$$(4.19) \qquad \overset{\circ}{\Omega}_i = \bigcup_{j=1}^{i} K_j$$

and we define $\partial_+ \Omega_i$ and $\partial_- \Omega_i$ in the usual way . Note that $\partial_- \Omega_i \subset \Gamma_-$.

Theorem 5   Assume that Hypotheses H.1 and H.2 hold. Assume

in addition that the solution u of problem (1.3) belongs to

$H^{k+1}(\Omega)$. Then, there exists a constant C > 0 independent of h

such that for all i = 1 ,..., I

(4.20) $\| u_h - u \|_{0,\Omega_i} \leq Ch^k \| u \|_{k+1,\Omega_i}$

(4.21) $(\int_{\partial_+\Omega_i} (\mu n_x + \nu n_y)(u_h - u)^2 ds)^{1/2} \leq Ch^k \| u \|_{k+1,\Omega_i}$

(4.22) $(- \sum_{j=1}^{i} \int_{\partial_-K_j} (\mu n_x + \nu n_y)(u_h - \xi_h)^2 ds)^{1/2} \leq Ch^k \| u \|_{k+1,\Omega_i}$!

Proof   For any $K \in \mathscr{C}_h$, we define

(4.23) $\eta_h = \begin{cases} 0 & \text{on} \quad \partial_-K \cap \Gamma_- \quad , \\ \text{outward trace of} \quad \Pi_h u \quad \text{on} \quad \partial_-K - (\partial_-K \cap \Gamma_-) \end{cases}$

We start from equation (4.1) with $v = \Pi_h u$, $\eta = \eta_h$ and we estimate

the corresponding right hand side member. First, we have

$$\left| \int_K (u-\Pi_h u)A^*(u_h - \Pi_h u) \, dxdy \right| \leq c_1 \| u-\Pi_h u \|_{0,K} \| u_h -\Pi_h u \|_{1,K} \qquad (1)$$

and by (4.15) , (4.17)

(4.24) $\left| \int_K (u-\Pi_h u)A^*(u_h - \Pi_h u) \, dxdy \right| \leq c_2 h^k \| u \|_{k+1,K} \| u_h - \Pi_h u \|_{0,K}$

Next, using (4.16) and (4.18) , we obtain

(4.25) $\left| \int_{\partial_+K} (\mu n_x + \nu n_y)(u-\Pi_h u)(u_h -\Pi_h u) \, ds \right| \leq c_3 h^k \| u \|_{k+1,K} \| u_h -\Pi_h u \|_{0,K}$

---

(1) In all of the sequel, we shall denote by c_i various constants

independent of h .

Similarly, we get

$$(4.26) \qquad \left| \int_{\partial_- K} (\mu n_x + \nu n_y)(u - \eta_h)(u_h - \Pi_h u) \, ds \right| \leq c_4 h^k \| u \|_{k+1, \mathcal{D}_K} \| u_h - \Pi_h u \|_{0,K}$$

where $\mathcal{D}_K$ is the union of the elements of $\mathcal{C}_h$ which have a side contained in $\partial_- K$.

Thus, combining (4.1) with $v = \Pi_h u$, $\eta = \eta_h$, (4.24), (4.25), (4.26) and using (3.7) , we obtain

$$\left\{ \begin{array}{l} \dfrac{1}{2} \displaystyle\int_{\partial_+ K} (\mu n_x + \nu n_y)(u_h - \Pi_h u)^2 \, ds - \\[3mm] -\dfrac{1}{2} \displaystyle\int_{\partial_- K} (\mu n_x \cdot \nu n_y)((u_h - \Pi_h u) - (\xi_h - \eta_h))^2 ds + \alpha \| u_h - \Pi_h u \|_{0,K}^2 \leq \\[3mm] \leq -\dfrac{1}{2} \displaystyle\int_{\partial_- K} (\mu n_x + \nu n_y)(\xi_h - \eta_h)^2 ds + c_5 h^k \| u \|_{k+1, K \cup \mathcal{D}_K} \| u_h - \Pi_h u \|_{0,K} \end{array} \right.$$

Summing over all the elements K_j , $1 \leq j \leq i$, and using (3.12) , (4.23) we get

$$(4.27) \left\{ \begin{array}{l} \dfrac{1}{2} \displaystyle\int_{\partial_+ \Omega_i} (\mu n_x + \nu n_y)(u_h - \Pi_h u)^2 ds - \\[3mm] - \dfrac{1}{2} \displaystyle\sum_{j=1}^{i} \int_{\partial_- K_j} (\mu n_x + \nu n_y)((u_h - \Pi_h u)-(\xi_h - \eta_h))^2 ds + \alpha \| u_h - \Pi_h u \|_{0,\Omega_i}^2 \leq \\[3mm] \leq c_6 h^k \| u \|_{k+1, \Omega_i} \| u_h - \Pi_h u \|_{0,\Omega_i} \end{array} \right.$$

From (4.17) and (4.27), we deduce :

$$\| u_h - u \|_{0,\Omega_i} \leq \| u_h - \Pi_h u \|_{0,\Omega_i} + \| \Pi_h u - u \|_{0,\Omega_i} \leq$$

$$\leq \frac{c_6}{\alpha} h^k \| u \|_{k+1,\Omega_i} + c_7 h^{k+1} \| u \|_{k+1,\Omega_i}$$

so that (4.20) holds.

Next, we have by (4.27)

$$\left(\int_{\partial_+\Omega_i} (\mu n_x + \nu n_y)(u_h - \Pi_h u)^2 ds \right)^{1/2} \leq c_8 h^k \| u \|_{k+1,\Omega_i}$$

and by (4.18)

$$\left(\int_{\partial_+\Omega_i} (\mu n_x + \nu n_y)(\Pi_h u - u)^2 ds \right)^{1/2} \leq c_9 h^{k+1/2} \| u \|_{k+1,\Omega_i}$$

This proves inequality (4.21).

Similarly, we have by (4.27)

$$\left(- \sum_{j=1}^{i} \int_{\partial_- K_j} (\mu n_x + \nu n_y)((u_h - \Pi_h u) - (\xi_h - \eta_h))^2 ds \right)^{1/2} \leq c_{10} h^k \| u \|_{k+1,\Omega_i}$$

and by (4.18)

$$\left(- \sum_{j=1}^{i} \int_{\partial_- K_j} (\mu n_x + \nu n_y)(\Pi_h u - u)^2 ds \right)^{1/2} \leq c_{11} h^{k+1/2} \| u \|_{k+1,\Omega_i} \quad ,$$

$$\left(- \sum_{j=1}^{i} \int_{\partial_- K_j} (\mu n_x + \nu n_y)(\eta_h - u)^2 ds \right)^{1/2} \leq c_{12} h^{k+1/2} \| u \|_{k+1,\Omega_i}$$

This implies inequality (4.22). □

5 - A SUPERCONVERGENCE RESULT

Let us notice that the error estimates of theorem 5 are not optimal in the exponent of the parameter h . In fact, numerical calculations have shown that these error bounds could not be improved in general. However, the one-dimensional results of § 2 clearly indicate that better estimates must hold in some special cases. Indeed, we shall prove in this § that the rate of convergence of our finite element method is $O(h^{k+1})$ when all the elements $K \in \mathcal{C}_h$ are rectangles and when $\hat{P} = Q_k$. In all the sequel, we shall confine ourselves to this particular case.

On the interval $[-1,+1]$, let $-1 < \theta_1 < \theta_2 < ... < \theta_{k+1} = 1$ denote the (k+1) Gauss-Radau quadrature abcissae. In the reference square $\hat{K} = [-1 , +1]^2$, we consider the points $\hat{a}_{ij}$ with coordinates (θ_i , θ_j) , $1 \leqslant i,j \leqslant k+1$.
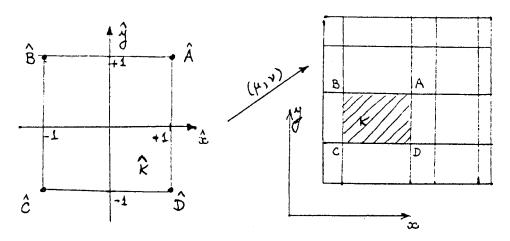

Fig. 5.

Just for convenience, we shall assume that the sides of the rectangles $K \in \mathscr{C}_h$ are parallel to the (x,y) axes and that the coefficients μ, ν are > 0. Given a rectangle K with vertices A, B, C, D as in Fig. 5, we denote by F_K the affine invertible mapping such that $A = F_K(\hat{A}), \ldots, D = F_K(\hat{D})$. Given a function $v \in C^\circ(K)$, we define $r_K v$ as the unique polynomial of Q_k which interpolates v at the points $a_{ij} = F_K(\hat{a}_{ij})$, $1 \leqslant i, j \leqslant k+1$. Then, for any $v \in C^\circ(\overline{\Omega})$, we define $r_h v$ to be the function in V_h such that

$$(5.1) \qquad r_h v \big|_K = r_K v \quad \text{for al } K \in \mathscr{C}_h .$$

We provide $L^\infty(\Omega)$ with the following norm

$$\| v \|_{0,\infty,\Omega} = \sup \ \{ |v(x)| \ ; \ x \in \Omega \}$$

Given any integer $m \geqslant 0$, let

$$W^{m,\infty}(\Omega) = \{ \ v \, | \, v \in L^\infty(\Omega), \ \partial^\alpha v \in L^\infty(\Omega) \ , |\alpha| \leqslant m \}$$

be the Sobolev space provided with the norm

$$\| v \|_{m,\infty,\Omega} = \max \ \{ \ \| \partial^\alpha v \|_{0,\infty,\Omega} \ ; \ |\alpha| \leqslant m \ \} .$$

Using [4] for instance, one can easily prove

Lemma 7   Assume that Hypothesis H.2 holds. Then, there exists a constant $c > 0$ independent of $K \in \mathscr{C}_h$ such that

$$(5.2) \ \| v - r_K v \|_{0,K} \leqslant c(h(K))^{k+1} \| v \|_{k+1,K} \quad \text{for all } v \in H^{k+1}(K) ,$$

$$(5.3) \ \| v - r_K v \|_{0,K'} \leqslant c(h(K))^{k+3/2} \| v \|_{k+1,\infty,K} \quad \text{for all}$$

$v \in W^{k+1,\infty}(K)$ where K' is any side of $\partial_+ K$

We are now able to prove

Theorem 6   Assume that all the elements $K \in \mathcal{C}_h$ are rectangles, that $\hat{P} = Q_k$ and that Hypothesis H.2 holds. Assume in addition, that the solution u of problem (1.3) belongs to $H^{k+2}(\Omega)$. Then, there exists a constant $c > 0$ independent of h such that for all $i = 1, \ldots, I$

$$(5.4) \quad \| u_h - u \|_{0,\Omega_i} \leqslant ch^{k+1} \| u \|_{k+2,\Omega_i} \quad,$$

$$(5.5) \quad (\int_{\partial_+\Omega_i} (\mu n_x + \nu n_y)(u_h - u)^2 ds)^{1/2} \leqslant ch^{k+1} (\| u \|_{k+2,\Omega_i} + \| u \|_{k+1,\infty,\Omega_i})$$

Proof   For any $K \in \mathcal{C}_h$, we now define

$$(5.6) \quad \eta_h = \begin{cases} 0 \quad \text{on} \quad \partial_- K \cap \Gamma_- \quad, \\ \text{outward trace of } r_h u \text{ on } \partial_- K - (\partial_- K \cap \Gamma_-). \end{cases}$$

We start from equation (4.1) with $v = r_h u$, $\eta = \eta_h$. The corresponding right hand side may be written in the form

$$(5.7) \quad X_K(u, u_h - r_h u) = Z_K(u, u_h - r_h u) + \int_K \sigma(u - r_h u)(u_h - r_h u) \, dxdy$$

where

$$(5.8) \quad \left\{ \begin{array}{l} Z_K(u,w) = \int_{\partial_+ K} (\mu n_x + \nu n_y)(u - r_h u)w \, ds + \int_{\partial_- K} (\mu n_x + \nu n_y)(u - \eta_h) \, wds - \\[2mm] \qquad\qquad - \int_K (u - r_h u)(\mu \frac{\partial w}{\partial x} + \nu \frac{\partial w}{\partial y}) \, dxdy \; . \end{array} \right.$$

We now use the following essential lemma which will be proved later.

Lemma 8   With the same assumptions as in Theorem 6 , there exists a constant $c > 0$ independent of $K \in \mathcal{C}_h$ such that for all $w \in Q_k$

$$(5.9) \qquad |z_K(u,w)| \leqslant C(h(K))^{k+1} \|u\|_{k+2,K} \|w\|_{0,K}$$

Using (5.2),(5.7) and (5.9), we obtain for all $K \in \mathcal{C}_h$

$$(5.10) \qquad |X_K(u,u_h - r_h u)| \leqslant c_1 h^{k+1} \|u\|_{k+2,K} \|u_h - r_h u\|_{0,K}$$

Thus, combining (4.1) with $v = r_h u, \eta = \eta_h$, (5.10) and using (3.7), we get

$$
\left\{
\begin{aligned}
&\frac{1}{2}\int_{\partial_+ K} (\mu n_x + \nu n_y)(u_h - r_h u)^2 ds + \alpha \|u_h - r_h u\|_{0,K}^2 \leqslant \\
&\leqslant -\frac{1}{2}\int_{\partial_- K} (\mu n_x + \nu n_y)(\xi_h - \eta_h)^2 ds + c_1 h^{k+1} \|u\|_{k+2,K} \|u_h - r_h u\|_{0,K}
\end{aligned}
\right.
$$

Summing over all the elements K_j , $1 \leqslant j \leqslant i$, we obtain

$$(5.11) \qquad
\begin{aligned}
&\frac{1}{2}\int_{\partial_+ \Omega_i} (\mu n_x + \nu n_y)(u_h - r_h u)^2 ds + \alpha \|u_h - r_h u\|_{0,\Omega_i}^2 \leqslant \\
&\leqslant c_1 h^{k+1} \|u\|_{k+2,\Omega_i} \|u_h - r_h u\|_{0,\Omega_i}
\end{aligned}
$$

Thus, the estimate (5.4) and (5.5) are simple consequences of inequality (5.11) and Lemma 7.

Proof of Lemma 8   Consider a rectangle $K \in \mathcal{C}_h$ with vertices A,B,C,D (cf.Fig.5). Let us denote by Δx (resp. Δy) the length of the side AB (resp.BC). We may write

$$(5.12) \qquad z_K(u,w) = \mu z_{K,x}(u,w) + \nu z_{K,y}(u,w)$$

with

$$z_{K,x}(u,w) = \int_D^A (u-r_h u)w \, dy - \int_C^B (u-\eta_h)w \, dy - \int_K (u-r_h u) \frac{\partial w}{\partial x} \, dxdy \ ,$$

$$z_{K,y}(u,w) = \int_B^A (u-r_h u)w \, dx - \int_C^D (u-\eta_h)w \, dx - \int_K (u-r_h u) \frac{\partial w}{\partial y} \, dxdy.$$

By using the one-to-one correspondence $v \to \hat{v} = v \circ F_K$, we get :

$$(5.13) \qquad z_{K,x}(u,w) = \frac{\Delta y}{2} \hat{z}_{\hat{x}}(\hat{u},\hat{w})$$

with

$$\hat{z}_{\hat{x}}(\hat{u},\hat{w}) = \int_{-1}^{+1} (\hat{u}(1,\hat{y}) - \hat{r}\hat{u}(1,\hat{y}))\hat{w}(1,\hat{y}) \, d\hat{y} -$$

$$- \int_{-1}^{+1} (\hat{u}(1,\hat{y}) - \hat{\eta}(\hat{y}))\hat{w}(1,\hat{y}) \, d\hat{y} - \int_{-1}^{+1} \int_{-1}^{+1} (\hat{u}-\hat{r}\hat{u}) \frac{\partial \hat{w}}{\partial \hat{y}} \, d\hat{x}d\hat{y} \ ,$$

where $\hat{r}\hat{u}$ is the polynomial of Q_k which interpolates $\hat{u}$ at the

points $\hat{a}_{ij}$, $1 \leqslant i,j \leqslant k+1$, and where $\hat{\eta}$ is the polynomial of degree

$\leqslant k$ which interpolates the function $\hat{y} \to \hat{u}(-1,\hat{y})$ at the points

θ_i , $1 \leqslant i \leqslant k+1$.

Clearly

$$\hat{z}_{\hat{x}}(\hat{u},\hat{w}) = 0 \quad \text{for all} \quad \hat{u},\hat{w} \in Q_k \ .$$

Now, when $\hat{u} = \hat{x}^{k+1}$, we have

$$\hat{u}(1,\hat{y}) = \hat{r}\hat{u}(1,\hat{y}) = 1 \ , \ \hat{u}(-1,\hat{y}) = \hat{\eta}(\hat{y}) = (-1)^{k+1} \ .$$

Moreover, $\hat{r}\hat{u}$ does not depend on $\hat{y}$ and then, for all $\hat{w} \in Q_k$, the

function $\hat{x} \to (\hat{u}-\hat{r}\,\hat{u})(\hat{x}) \frac{\partial \hat{w}}{\partial \hat{x}} (\hat{x},\hat{y})$ is a polynomial of degree $\leqslant 2k$

which vanishes at the $(k+1)$ Gauss-Radau points θ_i. Therefore,

$$\int_{-1}^{+1} (\hat{u}-\hat{r}\hat{u})\hat{w}\ d\hat{x} = 0 \quad \text{for all } \hat{w} \in Q_k .$$

Thus, when $\hat{u} = \hat{x}^{k+1}$, we get

$$\hat{z}_{\hat{x}}(\hat{u},\hat{w}) = 0 \quad \text{for all } \hat{w} \in Q_k .$$

On the other hand, when $\hat{u} = y^{k+1}$, $\hat{r}\hat{u}$ is independent of $\hat{x}$ so that we obtain by integration by parts

$$\int_{-1}^{+1}\int_{-1}^{+1} (\hat{u}-\hat{r}\hat{u}) \frac{\partial\hat{w}}{\partial\hat{x}}\ d\hat{x}d\hat{y} = \int_{-1}^{+1} (\hat{u}(1,\hat{y})-\hat{r}\hat{u}(1,\hat{y}))\hat{w}(1,\hat{y})\ d\hat{y} -$$

$$- \int_{-1}^{+1} (\hat{u}(1,\hat{y})-\hat{n}(\hat{y}))\hat{w}(1,\hat{y})\ d\hat{y} .$$

This give again

$$\hat{z}_{\hat{x}}(\hat{u},\hat{w}) = 0 \quad \text{for all } \hat{w} \in Q_k .$$

Therefore, we have proved that

$$\hat{z}_{\hat{x}}(\hat{u},\hat{w}) = 0 \quad \text{for all } \hat{u} \in P_{k+1} \text{ and all } \hat{w} \in Q_k .$$

Then, for fixed $\hat{w} \in Q_k$, the linear functional $\hat{u} \to \hat{z}_{\hat{x}}(\hat{u},\hat{w})$ is continuous over $H^{k+2}(\hat{K})$ with norm $\leqslant c_1 \|\hat{w}\|_{0,\hat{K}}$ and vanishes over P_{k+1}. By the Bramble-Hilbert lemma in the form given in [4 ,Lemma 6], we get for all $\hat{u} \in H^{k+2}(\hat{K})$ and all $\hat{w} \in Q_k$

$$|\hat{z}_{\hat{x}}(\hat{u},\hat{w})| \leqslant c_2|\hat{u}|_{k+2,\hat{K}} \|\hat{w}\|_{0,\hat{K}} .$$

Going back to the element K by using the correspondence

$\hat{v} \rightarrow v = \hat{v} \circ F_K^{-1}$ and (5.13) , we obtain for all $u \in H^{k+2}(K)$ and all $w \in Q_k$

(5.14) $\left| z_{K,x}(u,w) \right| \leqslant c_3 (h(K))^{k+1} \| u \|_{k+2,K} \| w \|_{o,K}$

Likewise, we get

(5.15) $\left| z_{K,y}(u,w) \right| \leqslant c_4 (h(K))^{k+1} \| u \|_{k+2,K} \| w \|_{o,K}$

Then, combining (5.12), (5.14) and (5.15) , we obtain the desired inequality (5.9). □

Note that the error estimates of Theorem 6 are now optimal in the exponent of the parameter h . However, as the one-dimensional results of § 1 suggest, we conjecture that, for any rectangle $K \in \mathcal{C}_h$, there exist some points of $\partial_+ K$ where even more precise error bounds hold. Unfortunately, we have not been able to prove the existence of such points.

REFERENCES

[1] Axelsson, O.,   " A class of A-stable methods " , B.I.T. 9 (1969),185-199.

[2] de Boor, C., and   " Collocation at Gaussian points ", SIAM J. Numer. Anal.
     B. Swartz   10 (1973) 582-606.

[3] Butcher, J.C.,   " Implicit Runge-Kutta processes ", Math. Comp. 18
                            (1964) , 50-64.

[4] Ciarlet, P.G., and   " General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with
     P.A. Raviart   applications to finite element methods " , Arch. Rat.
                            Mech. Anal. 46 (1972), 177-199.

[5] Ciarlet, P.G., and   " Interpolation theory over curved elements, with applica-
     P.A. Raviart   tions to finite element methods " , Comp. Meth. Appl.
                            Mech. Eng. 1 (1972), 217-249.

[6] Ciarlet, P.G., and   " The combined effect of curved boundaries and numerical
     P.A. Raviart   integration in isoparametric finite element methods " ,
                            The Mathematical Foundations of the Finite Element
                            Method with Applications to Partial Differential Equations
                            (A.K. Aziz, Editor), 409-474, Academic Press, New-York,1972.

[7] Crouzeix, M.,   Thesis , to appear.

[8] Friedrichs, K.O.,   " Symmetric positive differential equations " , Comm. Pure
                            Appl. Math. 11 (1958), 333-418.

[9] Hulme, B.L.,   " Discrete Galerkin and related one-step methods for ordinary
                            differential equations " , Math. Comp. 26 (1972), 881-891.

[10] Kaper, H.G., G.K. Leaf " Application of finite element techniques for the numerical
     and A.J. Lindeman   solution of the neutron transport and diffusion equations "
                            Proc. Conf. on Transport Theory , 2nd Conf.710107 , Los
                            Alamos (1971).

[11] Lesaint, P.,   " Finite element methods for symmetric hyperbolic equations " ,
                            Numer. Math. 21 (1973) , 244-255.

[12] Lesaint, P.,   " Finite element methods for the transport equation " , to
                            appear in RAIRO , série Mathématiques.

[13] Lesaint, P.,   Thesis , to appear.

[14] Lesaint, P., and   " Isoparametric finite element methods for the neutron
     J. Gérin-Roze   transport equation " , to appear.

[15] Miller, W.F. Jr,   " The application of phase-space finite elements to
     E.E Lewis, and   the two-dimensional transport equation in x-y
     E.C. Rossow   geometry " , to appear in Nucl. Sci. Eng.

[16] Ohnishi, T.,   " Application of finite element solution technique to
                            neutron diffusion and transport equations " , Proc.
                            conf. on New Developments in Reactor Mathematics
                            and Applications , CONF - 710302, Idaho Falls (1971).

[17] Reed, W.H., and   " Triangular mesh methods for the neutron transport
     T.R. Hill   equation ",to appear in Proc. Amer. Nucl. Soc.

TRAVAUX DU LABORATOIRE ASSOCIE 189

Année Universitaire 1973-1974.

M. CROUZEIX, P.A. RAVIART. Conforming and Nonconforming Finite Element Methods

for Solving the Stationary Stokes Equations . I

(Ref. 73001).

M. CROUZEIX, J.M. THOMAS. Eléments finis et problèmes elliptiques dégénérés.

(Ref. 73002).

H. BREZIS, G. DUVAUT . Ecoulements avec sillages autour d'un profil symétrique

sans incidence.

(Ref. 73003).

V. GIRAULT. Theory of a Finite Difference Method on irregular Networks for

Solving General Elliptic Dirichlet Boundary Value

Problems of Order Two.

(Ref. 73004) .

P.A. RAVIART. Méthode des éléments finis. Cours de D.E.A analyse numérique

rédigé par J.M. THOMAS.

(Ref. 73005) .

P.A. RAVIART. The Use of Numerical Integration in Finite Element Methods for

Solving Parabolic Equations.

(Ref. 73006) .

A. PERRONNET. Décompositions du problème de Stokes, stationnaire et évolutif

par dualisation de certaines contraintes. Rédigé en

collaboration avec M. BONNET.

(Ref. 73007) .

J.P. BOURGUIGNON, H. BREZIS. Remarks on the Euler Equation.

(Ref. 73008) .

J.L. LIONS, P.A. RAVIART. Séminaires d'analyse numérique.(1972-1973)

(Ref. 73009) .

J.L. LIONS, H. BREZIS. Séminaires sur les équations aux dérivées partielles.

(1972-1973) (Ref. 73010) .

H. BREZIS. Solutions with Compact Support of Variational Inequalities.

(Ref. 73010) .

P.A. RAVIART. Finite Element Methods for Solving the Stationary Stokes and

Navier-Stokes Equations.

(Ref. 73011) .

P. FAURRE. Identification et filtrage statistique. D.E.A. analyse numérique

(1973-1974) (Ref. 73012) .

P.G. CIARLET. Sur l'élément de Clough et Tocher.

(Ref. 73013) .

P.G. CIARLET. Quelques méthodes d'éléments finis pour le problème d'une plaque

encastrée.

(Ref. 73014) .

TRAVAUX DU LABORATOIRE ASSOCIE 189

Année Universitaire 1974-1975.

C. GERHARDT. Existence and Regularity of Capillary Surfaces.

(Ref. 74001).

P.G. CIARLET, P.A. RAVIART. A Mixed Finite Element Method for the Biharmonic

Equation.

(Ref. 74002).

F. MURAT, SIMON. Quelques résultats sur le contrôle par un domaine géométrique.

(Ref. 74003).

J.L. LIONS, P.A. RAVIART. Séminaires d'analyse numérique. (1973-1974).

(Ref. 74004).

L. TARTAR. Interpolation. Cours de III cycle rédigé par F. MURAT.

(Ref. 74005).

P.G. CIARLET, P.A. RAVIART. La méthode des éléments finis pour les problèmes

aux limites elliptiques.Chap. I.

(Ref. 74006).

M.F. BIDAUT. Théorèmes d'existence et d'existence "en général" pour des

problèmes de contrôle optimal.

(Ref. 74007).