Tadeusz J. Janik

THE EFFECT OF QUADRATURE ERRORS
ON THE FINITE ELEMENT SOLUTION
OF SOME NONLINEAR ELLIPTIC PROBLEMS

1. Introduction

The main purpose of this paper is to investigate the errors introduced in the approximate solutions of the nonlinear elliptic problems by quadrature formulas.

Our aim is to give sufficient conditions on this quadrature scheme which insure that the effect of numerical integration is of the same order as the approximation errors of the Galerkin method.

Numerical quadratures are used more and more frequently to compute the integrals over separate elements but this step of the finite element method has received not too much attention from numerical analysis.

The effect of quadrature errors in the finite element methods was analyzed by Ciarlet and Raviart (e.g. [2]) for linear elliptic problems. The alternative ways were presented by Strang and Fix [9] also for linear problems. The analogous result for mild nonlinear problems on rectangular regions was obtained by Herbold and Varga [7].

The method we shall employ is the straightforward transfer of the method presented by Ciarlet [1] on this nonlinear case.

2. The nonlinear elliptic problem

Let Ω be a bounded domain in Euclidian R^n with a sufficiently smooth boundary Γ and consider the following nonlinear elliptic problem,

$$(2.1a) \quad - \sum_{i=1}^n \partial_i (a_i(x, u) \partial_i u) + a_0(x, u) = 0 \quad \text{in } \Omega$$

$$(2.1b) \quad u(x) = 0 \quad \text{on } \Gamma.$$

We shall assume regularity and ellipticity conditions, namely for any $C > 0$

(R1) $a_i(x, u)$, $i = 0, 1, \dots, n$ are k -times continuously differentiable functions mapping $\Omega \times [-C, C]$ into bounded subset of R^1 ,

(R2) there exist constants C_0, C_1 such that for any $i = 1, \dots, n$ and any $(x, u) \in \Omega \times [-C, C]$, $0 < C_0 \leq a_i(x, u) \leq C_1$ and

$$(\partial/\partial u)(a_0(x, u)) - \sum_{i=1}^n [(\partial/\partial u)(a_i(x, u)) \partial_i u]^2 / a_i(x, u) \geq 0.$$

Let $W^{m,q}(\Omega)$, for any integer $m \geq 0$ and any number q , $1 \leq q \leq \infty$, be the linear space consisting of those functions in $L^q(\Omega)$ whose distributional derivatives $\partial^\alpha v$ up to the order m belong to $L^q(\Omega)$ with the norm,

$$\|v\|_{m,q,\Omega} = \left(\sum_{|\alpha| \leq m} \int_{\Omega} |\partial^\alpha v|^q dx \right)^{\frac{1}{q}}, \quad \text{if } 1 \leq q < \infty,$$

$$\|v\|_{m,\infty,\Omega} = \max_{|\alpha| \leq m} \left\{ \text{ess. sup} |\partial^\alpha v(x)| \right\}, \quad \text{if } q = \infty.$$

We shall also use the semi-norms,

$$|v|_{m,q,\Omega} = \left(\sum_{|\alpha|=m} \int_{\Omega} |\partial^\alpha v|^q dx \right)^{\frac{1}{q}}, \quad \text{if } 1 \leq q < \infty,$$

$$|v|_{m,\infty,\Omega} = \max_{|\alpha|=m} \left\{ \text{ess. sup} |\partial^\alpha v(x)| \right\}, \quad \text{if } q = \infty.$$

$W_0^{m,q}(\Omega)$ denotes the closure of the space $\mathcal{D}$ in the space $W^{m,q}(\Omega)$. Given a subset A of $\mathbb{R}^n$ and given a function $v \in C^m(A)$, the notation $\|v\|_{m,\infty,A}$ and $|v|_{m,\infty,A}$ will denote the norm $\max_{|\alpha| \leq m} \sup_{x \in A} |\partial^\alpha v(x)|$ and the semi-norm $\max_{|\alpha|=m} \sup_{x \in A} |\partial^\alpha v(x)|$, respectively.

We adopt the following notations

$$W_0^{m,2}(\Omega) = H^m(\Omega), \quad W_0^{m,2}(\Omega) = H_0^m(\Omega) \quad \text{and} \quad \|\cdot\|_{m,2,\Omega} = \|\cdot\|_{m,\Omega}.$$

Furthermore, let $\partial_i v = (\partial/\partial x_i)v$, $i = 1, \dots, n$, and the scalar product in $L^2(\Omega)$ is $(u, v) = \int_{\Omega} u v dx$. The weak form of (1.1a,b) is to find $u \in H_0^1(\Omega)$ such that

$$(2.2) \quad B(a(u), v) + Q(a_0(u), v) = 0 \quad \text{for all } v \in H_0^1(\Omega),$$

where

$$B(a(u), v) = \sum_{i=1}^n (a_i(\cdot, u) \partial_i u, \partial_i v),$$

$$Q(a_0(u), v) = (a_0(\cdot, u), v),$$

are nonlinear forms.

We remark that assumptions (R1), (R2) are sufficient to insure the existence of the unique solution $u \in W^{k+1,\infty}(\Omega) \cap W_0^{1,\infty}(\Omega)$; for this result see, Ladyzenskaja, Uralceva [8] and Douglas, Dupont, Serrin [5].

3. The Galerkin solution

Here and in the following we shall mostly use the notations introduced by Ciarlet [1].

We approximate the problem (2.2) by the finite element method using piecewise polynomial spline functions. For this

purpose, let $\bar{\Omega}$ be polyhedral domain which is a union of a finite number of closed subsets $e \in \mathcal{T}_h$, where $\mathcal{T}_h$ are triangulations of $\bar{\Omega}$. If $\bar{\Omega}$ is not a polyhedral domain, the isoparametric elements can be used as discussed by Zlamal [10] and Ciarlet, Raviart [2].

We shall assume that the finite element space V_h made up of finite elements (e, P_e, Σ_e) , $e \in \mathcal{T}_h$, satisfies the following conditions:

(F1) let $h_e = \text{diam}(e)$, $e \in \mathcal{T}_h$ and $h = \max_{e \in \mathcal{T}_h} (h_e)$, then the

associated family of triangulations $\mathcal{T}_h$ is κ -regular, i.e. each $e \in \mathcal{T}_h$ contains a ball with radius $\kappa^{-1}h$ and is contained in a ball with radius κh . The constant κ does not depend on h . The parameter h of the family belongs to a set whose infimum equals zero.

(F2) all the finite elements (e, P_e, Σ_e) , $e \in \mathcal{T}_h$ are affine equivalent to a single reference finite element (E, P_E, Σ_E) . (In the isoparametric case all the finite elements will be isoparametrically equivalent to the reference finite element).

(F3) all the finite elements (e, P_e, Σ_e) , $e \in \mathcal{T}_h$, are of class C^0 .

Restricting ourselves for simplicity to the case where $P_e = P_k(e)$ -space of all polynomials in $x_1, \dots, x_n$ of degree $\leq k$ for all $e \in \mathcal{T}_h$, we define finite dimensional subspaces S_h by

$$S_h = \{v_h \in V_h : v_h|_{\Gamma} = 0\}.$$

It follows that the inclusion

$$S_h \subset H_0^1(\Omega)$$

holds and if we assume in addition to (F1-3) that the following inclusion is satisfied

$$(3.1) \quad H^{k+1}(E) \subset C^S(E),$$

where s is the maximal order of partial derivatives occurring in the definition of the set $\sum_E$, then the following approximation property holds:

(P1) for any integer $m = 2, \dots, k+1$ and any number $q \in [2, +\infty]$ there exists a linear operator $\Pi_h: W^{m,q}(\Omega) \cap W_0^{1,q}(\Omega) \rightarrow S_h$ such that for all $v \in W^{m,q}(\Omega) \cap W_0^{1,q}(\Omega)$

$$\left(\sum_{e \in \mathcal{T}_h} \|v - \Pi_h v\|_{j,q,e}^q \right)^{\frac{1}{q}} \leq C h^{m-j} \|v\|_{m,q,\Omega}, \quad j=0,1,\dots,m, \text{ if } q < +\infty,$$

$$\max_{e \in \mathcal{T}_h} \|v - \Pi_h v\|_{j,\infty,e} \leq C h^{m-j} \|v\|_{m,\infty,\Omega}, \quad j=0,1,\dots,m, \text{ if } q = +\infty,$$

where the constant C is independent of h .

The inclusion (3.1) (Sobolev's imbedding theorem) is satisfied if $k > (n/2) - 1 + s$.

Under the assumptions (F1-3) the following inverse inequality is valid on the subspace V_h (and of course on S_h):

(P2) for any integers $0 \leq j \leq m$ and any number $q \in [1, +\infty]$ there exists a constant C independent of h such that for all $v_h \in V_h$

$$\left(\sum_{e \in \mathcal{T}_h} |v_h|_{m,q,e}^q \right)^{\frac{1}{q}} \leq C h^{j-m} \left(\sum_{e \in \mathcal{T}_h} |v_h|_{j,q,e}^q \right)^{\frac{1}{q}} \quad \text{if } q < +\infty,$$

$$\max_{e \in \mathcal{T}_h} |v_h|_{m,\infty,e} \leq C h^{j-m} \max_{e \in \mathcal{T}_h} |v_h|_{j,\infty,e} \quad \text{if } q = +\infty.$$

For the proof of the results (P1) and (P2) see Ciarlet [1].

The Galerkin approximation of the solution of (2.2) is now defined as the solution $u_h \in S_h$ of

$$(3.2) \quad B(a(u_h), v_h) + (Q(a_0(u_h)), v_h) = 0 \quad \text{for all } v_h \in S_h.$$

Douglas, Dupont [4] have shown that the difference between the solution of (2.2) and its Galerkin approximation (3.2) is

$$(3.3) \quad \|u - u_h\|_{j, \Omega} \leq Ch^{k+1-j} \|u\|_{k+1, \Omega}, \quad j = 0 \text{ or } 1.$$

Prehse [6] has obtained the pointwise estimate of the error

$$(3.4) \quad \|u - u_h\|_{0, \infty, \Omega} \leq Ch^{k+1} (\ln h)^{c(n)} \|u\|_{k+1, \infty, \Omega},$$

where $c(n)$ depends on n only. It gives the regularity of Galerkin solution, namely from (P1), (P2) and (3.3)

$$(3.5) \quad \left(\sum_{e \in \mathcal{T}_h} \|u_h\|_{k, e}^2 \right)^{\frac{1}{2}} \leq \|u\|_{k, \Omega} + \left(\sum_{e \in \mathcal{T}_h} \|u - u_h\|_{k, e}^2 \right)^{\frac{1}{2}} \leq \\ \leq \|u\|_{k, \Omega} + Ch \|u\|_{k+1, \Omega} \leq C \|u\|_{k+1, \Omega}$$

and likewise using (P1), (P2) and (3.4)

$$(3.6) \quad \max_{e \in \mathcal{T}_h} \|u_h\|_{k, \infty, e} \leq \|u\|_{k, \infty, \Omega} + \max_{e \in \mathcal{T}_h} \|u - u_h\|_{k, \infty, e} \leq \\ \leq \|u\|_{k, \infty, \Omega} + Ch (\ln h)^{c(n)} \|u\|_{k+1, \infty, \Omega} \leq C \|u\|_{k+1, \infty, \Omega}.$$

4. The effect of quadrature errors

In practice, even if the functions a_i , $i = 0, 1, \dots, n$, are linear and have simple analytical expression, the integrals appearing in (3.2) are seldom computed exactly. Instead, they are approximated through the process of numerical integration. According to (P2), computing the integral over e amounts to computing the integral over E . So let us consider one of the integrals appearing in (3.2) e.g. $\int_e f(x) dx$ and let

$$F_e : E \in \hat{x} \longrightarrow x = F_e(\hat{x}) = G_e \hat{x} + g_e \in e$$

be the invertible affine mapping which maps E onto e . Then,

$$(4.1) \quad \int_e f(x) dx = |\det G_e| \int_E \hat{f}(\hat{x}) d\hat{x},$$

where $f(x) = \hat{f}(\hat{x})$ for all $x \in F_e(\hat{x})$, $\hat{x} \in E$. The quadrature scheme over E consists in replacing the integral by a finite sum which can be symbolically written as

$$(4.2) \quad \int_E \hat{f}(\hat{x}) d\hat{x} \sim \sum_{l=1}^L \hat{w}_l \hat{f}(\hat{b}_l).$$

For simplicity we shall consider only formulas for which the nodes $\hat{b}_l \in E$ and the weights $\hat{w}_l$ are strictly positive. In view of (4.1) and (4.2), we see that the quadrature scheme over the set E automatically induces the quadrature scheme over the set e , namely

$$(4.3) \quad \int_e f(x) dx \sim \sum_{l=1}^L w_{l,e} f(b_{l,e})$$

with weights $w_{l,e}$ and nodes $b_{l,e}$ defined by

$$(4.4) \quad w_{l,e} = |\det G_e| \hat{w}_l \quad \text{and} \quad b_{l,e} = F_e(\hat{b}_l).$$

Let us introduce the quadrature error functionals

$$(4.5) \quad I_e(f) = \int_e f(x) dx - \sum_{l=1}^L w_{l,e} f(b_{l,e}),$$

$$(4.6) \quad \hat{I}(\hat{f}) = \int_E \hat{f}(\hat{x}) d\hat{x} - \sum_{l=1}^L \hat{w}_l \hat{f}(\hat{b}_l)$$

which are related

$$(4.7) \quad I_e(f) = |\det G_e| \hat{I}(\hat{f}).$$

We shall assume that the quadrature scheme (4.2) is exact for the space $P_{2k-2}(E)$, i.e.

$$(4.8) \quad \int_E \hat{p}(\hat{x}) d\hat{x} = \sum_{l=1}^L \hat{w}_l \hat{p}(\hat{b}_l) \quad \text{for all} \quad \hat{p} \in P_{2k-2}(E).$$

Such a rule is said to be of precision or of degree $2k-2$. Davis and Rabinowitz [3] have shown that for any bounded and closed subset A of R^n there exists a number of nodes in A and positive weights such that the quadrature scheme is of required precision. Also, see [3] for general introduction to the subject of numerical integration.

Finally, instead of solving the Galerkin equation (3.2) we are looking for a discrete solution $U_h \in S_h$ which satisfies

$$(4.9) \quad B_h(a(U_h), v_h) + Q_h(a_0(U_h), v_h) = 0 \quad \text{for all } v_h \in S_h,$$

where modified nonlinear forms are given by

$$\begin{aligned} B_h(a(U_h), v_h) &= \\ &= \sum_{e \in \mathcal{T}_h} \sum_{l=1}^L w_{l,e} \sum_{i=1}^n a_i(b_{l,e}, U_h(b_{l,e})) \partial_i U_h(b_{l,e}) \partial_i v_h(b_{l,e}), \\ Q_h(a_0(U_h), v_h) &= \sum_{e \in \mathcal{T}_h} \sum_{l=1}^L w_{l,e} a_0(b_{l,e}, U_h(b_{l,e})) v_h(b_{l,e}). \end{aligned}$$

Let

$$\begin{aligned} (4.10) \quad A_h(v_h, z_h; y_h) &= B_h(a(v_h), y_h) - B_h(a(z_h), y_h) + \\ &+ Q_h(a_0(v_h), y_h) - Q_h(a_0(z_h), y_h) \quad \text{for } v_h, z_h, y_h \in S_h, \end{aligned}$$

where apparently the functional $A_h(v_h, z_h; y_h)$ is linear in y_h . We now prove that (4.8) is the sufficient condition on a quadrature scheme which insures the forms (4.10) are uniformly S_h -coercive i.e. there exists a constant $\alpha > 0$ independent of the subspace S_h such that for all $v_h, z_h \in S_h$

$$(4.11) \quad A_h(v_h, z_h; v_h - z_h) \geq \alpha |v_h - z_h|_{1,\Omega}^2.$$

Notice that such an assumption implies the existence of the unique discrete solution U_h of (4.9).

T h e o r e m 4.1. Let there be given a quadrature scheme (4.2) satisfying (4.8). Then there exists a constant $\alpha > 0$ independent of h such that for all forms (4.10), all spaces S_h , and for all bounded $v_h, z_h \in S_h$

$$A_h(v_h, z_h; v_h - z_h) \geq \alpha \|v_h - z_h\|_{1, \Omega}^2.$$

P r o o f . First, we notice that since the function $\sum_{i=1}^n (\partial_i \hat{p})^2$ belongs to the space $P_{2k-2}(\mathbb{E})$ for all $\hat{p} \in \hat{P} = P_k(\mathbb{E})$ we have

$$(4.12) \quad \|\hat{p}\|_{1, \mathbb{E}}^2 = \int_{\mathbb{E}} \sum_{i=1}^n (\partial_i \hat{p})^2 d\hat{x} = \sum_{l=1}^L \hat{w}_l \sum_{i=1}^N (\partial_i \hat{p}(\hat{b}_l))^2.$$

From the Taylor formula we have

$$\begin{aligned} & \sum_{i=1}^n \left\{ a_i(b_{1,e}, v_h(b_{1,e})) \partial_i v_h(b_{1,e}) \partial_i (v_h - z_h)(b_{1,e}) - \right. \\ & \quad \left. - a_i(b_{1,e}, z_h(b_{1,e})) \partial_i z_h(b_{1,e}) \partial_i (v_h - z_h)(b_{1,e}) \right\} + \\ & + a_0(b_{1,e}, v_h(b_{1,e}))(v_h - z_h)(b_{1,e}) - a_0(b_{1,e}, z_h(b_{1,e}))(v_h - z_h)(b_{1,e}) = \\ & = \int_0^1 W^T M_h W ds, \end{aligned}$$

where

$$M_h = \begin{bmatrix} a_1(b_{1,e}, y_h) & 0 & \dots & 0 & (\partial/\partial u)(a_1(b_{1,e}, y_h)) \partial_1 y_h \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & \dots & 0 & a_n(b_{1,e}, y_h) & (\partial/\partial u)(a_n(b_{1,e}, y_h)) \partial_n y_h \\ 0 & \dots & \dots & 0 & (\partial/\partial u)(a_0(b_{1,e}, y_h)) \end{bmatrix},$$

$$W^T = [\partial_1 p_e(b_{1,e}), \dots, \partial_n p_e(b_{1,e}), p_e(b_{1,e})]$$

with $y_h = (z_h + s(v_h - z_h))(b_{1,e})$ and $p_e = (v_h - z_h)|_e$.

The matrix M_h is uniformly positive definite for bounded $v_h, z_h \in S_h$, since the assumptions (R1-2) are valid.

Thus, with the positivity of the weights,

$$(4.13) \quad A_h(v_h, z_h; v_h - z_h) \Big|_e \geq C \sum_{l=1}^L w_{l,e} \sum_{i=0}^n (\partial_i p_e(b_{l,e}))^2 \geq \\ \geq C \sum_{l=1}^L w_{l,e} \sum_{i=1}^n (\partial_i p_e(b_{l,e}))^2.$$

Ciarlet [1] has shown that

$$\sum_{l=1}^L w_{l,e} \sum_{i=1}^n (\partial_i p_e(b_{l,e}))^2 \geq \\ \geq |\det G_e| \|G_e\|^{-2} \sum_{l=1}^L \hat{w}_l \sum_{i=1}^n (\partial_i \hat{p}_e(\hat{b}_l))^2,$$

where $\hat{p}_e \in \hat{P}$ is the polynomial associated with p_e through the affine mapping.

From (4.12) and (F1) we can estimate the right-hand side of the last inequality,

$$(4.14) \quad A_h(v_h, z_h; v_h - z_h) \Big|_e \geq C |\det G_e| \|G_e\|^{-2} |\hat{p}_e|_{1,E}^2 \geq \\ \geq C (\|G_e\| \|G_e\|^{-1})^{-2} |p_e|_{1,e}^2 \geq C |p_e|_{1,e}^2,$$

since from (F1-2) $|p_e|_{1,e} \leq C \|G_e^{-1}\| |\det G_e|^{\frac{1}{2}} |\hat{p}_e|_{1,E}$ and $\|G_e\| \|G_e\|^{-1} \leq C$, for these results, see Ciarlet [1].

Thus, we have obtained that there exists a constant $\alpha > 0$ independent of $e \in \mathcal{T}_h$ and h such that

$$A_h(v_h, z_h; v_h - z_h) \Big|_e \geq \alpha |v_h - z_h|_{1,e}^2.$$

By adding the above inequalities for all $e \in \mathcal{T}_h$ we get the Theorem 4.2.

We can now prove the following result about an abstract error estimate.

T h e o r e m 4.2. Consider a family of discrete problems for which the associated approximate nonlinear forms (4.10) are uniformly S_h -coercive. Let u_h be the solution of the Galerkin equation (3.2) and U_h be the solution of the modified by numerical integration equation (4.9). Then there exists a constant C independent of the space S_h such that

$$(4.15) \quad |u_h - U_h|_{1,\Omega}^2 \leq C \left\{ |B(a(u_h), u_h - U_h) - B_h(a(u_h), u_h - U_h)| + |Q(a_0(u_h), u_h - U_h) - Q_h(a_0(u_h), u_h - U_h)| \right\}.$$

P r o o f . With the assumption of uniform S_h -coercivity we may write,

$$\begin{aligned} \alpha |u_h - U_h|_{1,\Omega}^2 &\leq A_h(u_h, U_h; u_h - U_h) = \\ &= B_h(a(u_h), u_h - U_h) - B_h(a(U_h), u_h - U_h) + \\ &+ Q_h(a_0(u_h), u_h - U_h) - Q_h(a_0(U_h), u_h - U_h) \leq \\ &\leq |B_h(a(u_h), u_h - U_h) - B(a(u_h), u_h - U_h)| + \\ &+ |Q_h(a_0(u_h), u_h - U_h) - Q(a_0(u_h), u_h - U_h)|, \end{aligned}$$

where we have used (3.2) and (4.9) with $v_h = u_h - U_h \in S_h$, from this the theorem follows immediately.

We can now turn to the problem of estimating the both terms appearing in the right-hand side of the inequality (4.15). We shall determine when these quadrature errors are consistent with (i.e. of the same order as) the approximation errors of the Galerkin method (3.2).

By definition of the quadrature error functionals $I_\bullet(\cdot)$ (cf. (4.5)) we have, for all $v_h \in S_h$,

$$(4.16) \quad B(a(u_h), v_h) - B_h(a(u_h), v_h) = \sum_{e \in \mathcal{T}_h} I_\bullet \left(\sum_{i=1}^n a_i(u_h) \partial_i u_h \partial_i v_h \right),$$

$$(4.17) \quad Q(a_0(u_h), v_h) - Q_h(a_0(u_h), v_h) = \sum_{e \in \mathcal{T}_h} I_e(a_0(u_h), v_h).$$

In the next theorem we shall obtain "local" quadrature error estimates.

T h e o r e m 4.3. Let us assume that the quadrature scheme is exact for the space $P_{2k-2}(e)$. Moreover let for $i = 0, 1, \dots, n$, $a_i(\cdot, u(\cdot)) \in W^{k, \infty}(e)$. Then there exists a constant C independent of $e \in \mathcal{T}_h$ such that for all $u_h, v_h \in P_k(e)$,

$$(4.18) \quad I_e(a_i(u_h), \partial_i u_h \partial_i v_h) \leq \\ \leq Ch_e^k \|a_i(u_h)\|_{k, \infty, e} \|u_h\|_{k, e} \|v_h\|_{1, e},$$

$$(4.19) \quad I_e(a_0(u_h), v_h) \leq Ch_e^k (\text{meas}(e))^{\frac{1}{2}} \|a_0(u_h)\|_{k, \infty, e} \|v_h\|_{1, e}.$$

P r o o f . For a given $\hat{v} \in P_{k-1}(E)$ and any $\hat{\varphi} \in W^{k, \infty}(E)$, using the Bramble-Hilbert lemma (see [1]), there exists a constant C such that

$$|\hat{I}(\hat{\varphi} \hat{v})| \leq C |\hat{\varphi}|_{k, \infty, E} |\hat{v}|_{0, E}.$$

Now we shall get an error estimate for the expression

$$I_e(a_i(u_h), \partial_i u_h \partial_i v_h) \text{ for } a_i \in W^{k, \infty}(e) \text{ and } \partial_i u_h, \partial_i v_h \in P_{k-1}(e).$$

From (4.7) we have that

$$I_e(a_i(u_h), \partial_i u_h \partial_i v_h) = |\det G_e| \hat{I}(\hat{a}_i(\hat{u}) \partial_i \hat{u} \partial_i \hat{v})$$

with $\hat{a}_i \in W^{k, \infty}(E)$ and $\partial_i \hat{u}, \partial_i \hat{v} \in P_{k-1}(E)$, $i = 1, \dots, n$.

Using the following inequality

$$|\varphi v|_{m, q, A} \leq C \sum_{j=0}^m |\varphi|_{m-j, q, A} \|v\|_{j, \infty, A}$$

for $\varphi \in W^{m,q}(A)$ and $v \in W^{m,\infty}(A)$ with C independent of the set A and taking into account that $|\partial \hat{u}|_{k,\infty,E} = 0$ we get that

$$\begin{aligned} |\hat{\varphi}|_{k,\infty,E} &= |\hat{a}_1(\hat{u}) \partial_1 \hat{u}|_{k,\infty,E} \leq \\ &\leq C \sum_{j=0}^{k-1} |\hat{a}_1(\hat{u})|_{k-j,\infty,E} |\partial_1 \hat{u}|_{j,\infty,E} \leq \\ &\leq C \sum_{j=0}^{k-1} |\hat{a}_1(\hat{u})|_{k-j,\infty,E} |\partial_1 \hat{u}|_{j,E}, \end{aligned}$$

where we have used the equivalence of norms over the finite-dimensional space $P_{k-1}(E)$.

Thus, we obtain for $a_1 \in W^{k,\infty}(E)$ and $\partial_1 \hat{u}, \partial_1 \hat{v} \in P_{k-1}(E)$

$$\begin{aligned} (4.20) \quad &|\hat{I}(\hat{a}_1(\hat{u}) \partial_1 \hat{u} \partial_1 \hat{v})| \leq \\ &\leq C \left(\sum_{j=0}^{k-1} |\hat{a}_1(\hat{u})|_{k-j,\infty,E} |\partial_1 \hat{u}|_{j,E} \right) |\partial_1 \hat{v}|_{0,E}. \end{aligned}$$

Ciarlet [1] has shown that

$$(4.21) \quad |\hat{p}|_{j,q,E} \leq Ch_e^j |\det G_e|^{-\frac{1}{q}} |p|_{j,q,e}, \quad j=0,\dots,k; \quad 1 \leq q \leq \infty.$$

It follows that

$$(4.22) \quad |\partial_1 \hat{u}|_{j,E} \leq Ch_e^j |\det G_e|^{-\frac{1}{2}} |\partial_1 u_h|_{j,e}, \quad j=0,\dots,k-1,$$

$$(4.23) \quad |\partial_1 \hat{v}|_{0,E} \leq C |\det G_e|^{-\frac{1}{2}} |\partial_1 v_h|_{0,e}.$$

Since $u_h|_e$ is bounded (polynomial) and $a_i(x,u)$, $i=1,\dots,n$, are k -times continuously differentiable functions of $\bar{\Omega} \times [-C,C]$ (cf. (R1)) we have that the functions $\hat{a}_1(\hat{u})$ belongs to the space $W^{k,\infty}(E)$.

Thus, from (4.21)

$$(4.24) \quad |\hat{a}_i(\hat{u})|_{k-j, \infty, E} \leq Ch_e^{k-j} |a_i(u_h)|_{k-j, \infty, e}, \quad j=0, \dots, k-1.$$

Combining inequalities (4.20), (4.22), (4.23) and (4.24) we get for $a_i(u_h) \in W^{k, \infty}(e)$; $\partial_i u_h, \partial_i v_h \in P_{k-1}(e)$, $i=1, \dots, n$

$$\begin{aligned} & |I_e(a_i(u_h) \partial_i u_h \partial_i v_h)| \leq \\ & \leq Ch_e^k \left(\sum_{j=0}^{k-1} \|a_i(u_h)\|_{k-j, \infty, e} |\partial_i u_h|_{j, e} \right) |\partial_i v_h|_{0, e} \leq \\ & \leq Ch_e^k \|a_i(u_h)\|_{k, \infty, e} \|\partial_i u_h\|_{k-1, e} |\partial_i v_h|_{0, e}. \end{aligned}$$

Analogously, we can prove the other fact from the theorem, namely it suffices to notice that

$$\hat{I}(\hat{a}_0(\hat{u})\hat{v}) = \hat{I}(\hat{a}_0(\hat{u})\hat{\Pi}_1\hat{v}) + \hat{I}(\hat{a}_0(\hat{u})(\hat{v} - \hat{\Pi}_1\hat{v})),$$

where $\hat{\Pi}_1$ is the orthogonal projection in the space $L^2(E)$ onto the subspace $P_1(E)$ and the central idea is to apply the result that

$$|\hat{a}_0(\hat{u})|_{k-j, \infty, E} \leq Ch_e^{k-j} |a_0(u_h)|_{k-j, \infty, e}$$

which follows from (4.21) and from the condition (R2).

Now, we can come to introduce the "global" consistency error estimates with u_h - the Galerkin solution of (3.2).

Theorem 4.4. With the all previous assumptions the following inequalities hold for any $v_h \in S_h$

$$(4.25) \quad |B(a(u_h), v_h) - B_h(a(u_h), v_h)| \leq Ch^k \|u\|_{k+1, \Omega} |v_h|_{1, \Omega},$$

$$(4.26) \quad |Q(a_0(u_h), v_h) - Q_h(a_0(u_h), v_h)| \leq Ch^k (\text{meas}(\Omega))^{\frac{1}{2}} |v_h|_{1, \Omega},$$

where u_h is the Galerkin solution of (3.2) and C does not depend on h .

P r o o f . Using (4.16), Theorem 4.3 and the Cauchy-Schwarz inequality we obtain for any $v_h \in S_h$

$$\begin{aligned}
 & |B(a(u_h), v_h) - B_h(a(u_h), v_h)| \leq \\
 & \leq \sum_{e \in \mathcal{T}_h} \sum_{i=1}^h |I_e(a_i(u_h)|_e \partial_i u_h|_e \partial_i v_h|_e| \leq \\
 & \leq C \sum_{e \in \mathcal{T}_h} h_e^k \sum_{i=1}^n \|a_i(u_h)\|_{k, \infty, e} \|u_h\|_{k, e} |v_h|_{1, e} \leq \\
 & \leq Ch^k \left(\sum_{i=1}^n \max_{e \in \mathcal{T}_h} \|a_i(u_h)\|_{k, \infty, e} \right) \left(\sum_{e \in \mathcal{T}_h} \|u_h\|_{k, e}^2 \right)^{\frac{1}{2}} |v_h|_{1, \Omega}.
 \end{aligned}$$

To complete the proof of (4.25) it suffices to use the assumption (R1) and the inequality (3.6) to obtain

$$\sum_{i=1}^n (\max_{e \in \mathcal{T}_h} \|a_i(u_h)\|_{k, \infty, e}) \leq C.$$

Likewise, using (4.17), Theorem 4.3 and (R2), (3.6), we get (4.26). In order to complete the quadrature error estimate, we have to use Theorem 4.2 and Theorem 4.4 with $v_h = u_h - U_h$. Thus,

$$|u_h - U_h|_{1, \Omega} \leq Ch^k \|u\|_{k+1, \Omega}.$$

Therefore we have given sufficient conditions on the quadrature scheme (it should be of degree $2k-2$) which insure that the effect of numerical integration does not decrease the order of convergence of the Galerkin solution. This follows from the triangle inequality that

$$\|u - U_h\|_{1, \Omega} \leq Ch^k \|u\|_{k+1, \Omega},$$

where u and U_h are the solution of (2.2) and (4.9), respectively.

Acknowledgment. The author appreciates several helpful suggestions made by Dr. A.Wakulicz.

REFERENCES

- [1] P.G. C i a r l e t : The finite element method for elliptic problems, North-Holland, Amsterdam, 1978.
- [2] P.G. C i a r l e t , P.-A. R a v i a r t : The combined effect of curved boundaries and numerical integration in isoparametric finite element methods, in the mathematical foundations of the finite element method with applications to partial differential equations, (ed. A.K. Aziz), Ac. Press, New York (1972) 409-474.
- [3] P.J. D a v i s , P. R a b i n o w i t z : Methods of numerical integration, Ac. Press, New York, 1975.
- [4] J. D o u g l a s , T. D u p o n t : A Galerkin method for a nonlinear Dirichlet problem, Math. Comp. 29 (1975), 689-696.
- [5] J. D o u g l a s , T. D u p o n t , J. S e r - r i n : Uniqueness and comparison theorems for elliptic equations in divergence form, Arch. for Rational Mech. and Anal. 42 (1971), 157-168.
- [6] J. F r e h s e : Optimale gleichmassige Konvergenz der Methode der finiten elemente bei quasilinearen N-dimensionalen Randwertproblemen, Z. Ang. Math. Mech. 57 (1977), 229-231.
- [7] R.J. H e r b o l d , R.S. V a r g a : The effect of quadrature errors in the numerical solution of two-dimensional boundary value problems by variational techniques, Aequationes Math. 7 (1971), 36-58.
- [8] O.A. L a d y z e n s k a j a , N.N. U r a l c e v a : Linear and quasilinear elliptic equations, Ac. Press, New York, 1968.

- [9] G. S t r a n g , G.J. F i x : An analysis of the finite element method, Prentice-Hall, New Jersey, 1973.
- [10] M. Z l a m a l : Curved elements in the finite element method, I, II, SIAM J. Numer. Anal., 10 (1973), 229-240; 11 (1974), 347-362.

INSTITUTE OF MATHEMATICS, TECHNICAL UNIVERSITY OF WARSAW,
00-661 WARSZAWA

Received October 16, 1985.