STRESS SAMPLING POINTS FOR LINEAR TRIANGLES IN THE FINITE ELEMENT METHOD

N. LEVINE

NUMERICAL ANALYSIS REPORT 10/82

1. INTRODUCTION

In 1971 Veryard [21] noted an improvement in the accuracy of gradients of biquadratic Galerkin approximations which took place when the gradients were sampled at the second order Gauss points in each element. Since then such points of exceptional accuracy of derivatives of finite element approximations have come to be known as "stress points" and their existence as an example of the phenomenon of "superconvergence".

Strang and Fix [19] prophetically had this to say: "We believe that the stress points can be located in the following way. The leading term in the errors is governed by the problem of approximating polynomials P_k of degree k (sic), in energy, by the trial functions in S^h --- Then stress points are identified by the property that the true stresses (derivatives of P_k) coincide with their approximations (derivatives of a lower-degree polynomial)." Also, in two dimensions, "exceptional points for one stress component need not be exceptional for the others. The midpoints of an edge seem to be likely to be exceptional for derivatives along the edge but not for stresses in the direction of the normal".

In a paper submitted in 1974 (published in 1976), Barlow [2] went further in the case of rectangular elements. He noted that the gradient of the interpolant, from a "serendipity" polynomial space incomplete in P_4, of any polynomial in P_4 is matched exactly at the 2x2 Gauss points. It was to be hoped that this would also apply to the gradient of the Galerkin approximation and that this exact matching would significantly improve the accuracy of the gradient (or in other words, that the error was indeed dominated by approximations to elements of P_4).

The first precise statement and complete proof of superconvergence at Gauss points was given for the case of linear and quadratic serendipity elements in up to three dimensions by Zlámal at the Rome conference in 1975 [23]. His work was later extended to cover curved isoparametric elements of any degree, with various numerical quadrature schemes ([24], [11]). In all three papers, as in this report, the term "superconvergence" is associated with the sampling of gradients to an average accuracy of $O(h^k)$ in cases where the global accuracy of first derivatives is $O(h^{k-1})$. We note that two papers published in 1977 ([5], [20]) which discussed local averaging processes for boosting an $O(h^r)$ error in displacements and their derivatives to $O(h^{2r-1})$ would not result in an improvement for the common linear and bilinear elements.

Very little has been published on triangular elements: whether they have stress points and if so where has been open to question. Moan [13] claimed without proof that, as with quadrilateral elements, stress points and Gauss points were one and the same. His intuitive argument was based on the assumption that the Galerkin least-squares approximation to gradients is "almost local" and can therefore be analysed in one element in complete isolation from all others. Zienkiewicz [22] similarly presented as "physically obvious" the equivalence of Gauss points and optimal stress sampling points in both quadrilaterals and triangles. It is now evident that the reasoning of these last authors, though fortuitously successful for quadrilaterals, assumes too much; the global nature of the Galerkin approximation cannot be neglected. However, Barlow's more cautious approach has survived; indeed it can be viewed as an introduction to Zlámal's work.

We recall that Strang and Fix predicted tangential-derivative stress points at the midpoints of element edges. (This result, applied to quadrilaterals, holds in addition to those of Zlámal.) In this report we show that the prediction is good for linear elements on triangles. The question of whether in some sense these midpoints are "optimal" sampling points is discussed.

In Section 3 we give a crucial bound on the normed difference of the Galerkin and interpolating approximations to the solution of a model problem. The derivation of this bound requires the combination of error terms between neighbouring elements in a manner not necessary when considering the corresponding problem on serendipity elements (i.e. rectangles). Otherwise Sections 3 and 4 (this latter completes the superconvergence proof) closely follow Zlámal's work [23]. In particular the bounds obtained are non-constructive (and clearly non-optimal).

Two key results from Sections 3 and 4 are reworked from an optimal standpoint in Section 5. The method used does not appear in the literature. It is based on Taylor expansions which directly exploit more smoothness than those employed by Meinguet [12]; however the full smoothness available is only used indirectly. Although our method is very successful here its applicability to other problems may be limited.

We discuss the possibilities of pointwise (as opposed to "on average") superconvergence in Section 6; this is followed by a note on Moan's suggestion for sampling at centroids. In Section 8 we consider extensions to the model problem introduced in Section 3. Finally, we present some numerical evidence and pose the question of where, however good or bad in some average sense the midpoints may be, it is "best" to sample the gradient of a finite element solution on linear triangles.

Before we embark on the preliminary notation etc., we have two points to note. The first is that separate components of a vector at different places are usually not what is required, so that some averaging process to recover both components at a single point must be devised (e.g. linear interpolation). The other is that the error bounds that follow invoke high order seminorms which will not in general be easy to calculate or even approximate. Therefore it may well turn out that the most practical results in this report are the numerical indications that, whatever the magnitude of the error, the midpoint is "a priori" the best place to sample.

2. PRELIMINARIES

The results in this report are presented in the context of Sobolev spaces. In this section we introduce the relevant notation and go on to some lemmas which will be used later.

For any open region in $\mathbb{R}^n$ (n = 1 or 2), parametised here by R, we denote by $H^m(R)$ the Sobolev space of functions which together with their generalised derivatives up to order m inclusive are in $L_2(R)$. The norm in $\mathbb{R}^2$ is given by

$$\| u \|^2_{m,R} = \int\int_R \sum_{i+j\le m} \left[\left(\frac{\partial}{\partial x}\right)^i \left(\frac{\partial}{\partial y}\right)^j u\right]^2. \tag{2.1}$$

We will also use the seminorm

$$|u|^2_{m,R} = \int\int_R \sum_{i+j=m} \left[\left(\frac{\partial}{\partial x}\right)^i \left(\frac{\partial}{\partial y}\right)^j u\right]^2. \tag{2.2}$$

In either case, if the subscript $_{,R}$ is absent, integration is implicitly over the domain Ω in which the differential equation is posed. By $(\cdot,\cdot)$ we denote the inner product in $L_2(\Omega)$. (Ω is introduced in Section 3).

The subspace $H^m_0(R)$ of $H^m(R)$ is formed by completing in the H^m - norm the set of infinitely differentiable functions with compact support in R. It can loosely be thought of as the subspace of H^m whose members vanish on ∂R.

Finally, the space of polynomials of degree less than k in R is denoted by $P_k(R)$ (for example P_3 is the space of quadratics).

In all that follows the letter C stands for a generic, positive number, different at each appearance unless identified by a subscript, "constant" in that it is independent of any functions in H^m (such as the unknown u) and of the discretisation parameter h which will be introduced below.

We require some results from the literature, in a simplified form in that they need only apply to the triangle Δ with vertices (0,0), (1,0), (1,1) and the quadrilateral Q with vertices (0,0), (0,-1), (1,0), (1,1) rather than to "unit diameter regions satisfying the ordinary cone condition", and in that they only refer to spaces based on the L_2 norm in $\mathbb{R}^2$. In the following, R denotes either of the regions Q or Δ, its boundary is ∂R.

Lemma 1 (Taylor's theorem). Let x and y be $\mathbb{R}^2$ vectors representing points on a line Γ and let α denote a 2-index. (See, for example, [8] or [17] for multi-index notation). If $w \in H^m(\Gamma)$ then

$$w(x) = \sum_{|\alpha|<m} w^{(\alpha)}(y) \frac{(x-y)^\alpha}{\alpha!} + m \sum_{|\alpha|=m} \frac{(x-y)^\alpha}{\alpha!} \int_0^1 w^{(\alpha)}(x+s(y-x)) s^{m-1} \, ds.$$

Lemma 2 (Trace theorem) If $w \in H^{m+1}(R)$ then $w \in H^m(\partial R)$. (For a proof of the more general statement see [1]). As a corollary, if the line Γ forms part of ∂R and if $w \in H^{m+1}(R)$ then the result of Lemma 1 holds.

Lemma 3 (Sobolev's representation). Let x and y denote $\mathbb{R}^2$ vectors and α denote a 2-index. Let B be an open ball in R and $\phi \in C_0^\infty(B)$ such that $\int\int_B \phi(y)dy = 1$. Define

$$k_\alpha(x,y) = \frac{|\alpha|}{\alpha!} (x-y)^\alpha \int_0^1 \phi(x+s^{-1}(y-x)) \, s^{-3} ds.$$

If $w \in H^m(R)$ then

$$w(x) = \sum_{|\alpha|<m} \int\int_B \phi(y) \, w^{(\alpha)}(y) \frac{(x-y)}{\alpha!} \, dy + \sum_{|\alpha|=m} \int\int_R k_\alpha(x,y) w^{(\alpha)}(y) dy.$$

(This is a "smoothed" version of Taylor's theorem; for proof see [8] or [18]).

Lemma 4 ("Sobolev lemma"). If $w \in H^2(R)$ then $\max_R |w| \le C_1 \| w \|_{2,R}$. (For a proof of the more general statement from Sobolev's representation see [17] or [18]).

If $w \in H^2(\Delta)$ then since w is defined at each point in Δ we can define the interpolant to w (denoted w_I) as the linear function on Δ which takes the same values as w at the vertices of Δ. Also if $w \in H^2(\Omega)$ where Ω is any region which has been exactly subdivided into a union of triangles we shall define the interpolant w_I to w to be the function which interpolates w in each triangle (so that w_I is continuous and piecewise linear). We clearly have

Lemma 5 Let $w \in H^2(\Delta)$ and let w_I be its interpolant. Then

$$\max_{\Delta} |w_I| = \max_{\substack{\text{vertices} \\ \text{of } \Delta}} |w| \leq \max_{\Delta} |w| \leq C_1 \| w \|_{2,\Delta}$$

and
$$|\underline{\nabla} w_I| \leq 2 \max_{\Delta} |w_I| \leq 2C_1 \| w \|_{2,\Delta} .$$

Lemma 6 (Bramble-Hilbert). Let F be a linear functional on $H^m(R)$ such that (i) $|F(w)| \leq C_2 \| w \|_{m,R}$ $\forall w \in H^m(R)$ and

(ii) $F(q) = 0$ $\forall q \in P_m(R)$.

Then
$$|F(w)| \leq C_2 C_3 \, |w|_{m,R} \qquad \forall w \in H^m(R)$$

for some constant C_3 which depends only on R.

(For a complete proof of the more general statement see [7], [17] or both [4] and [14]).

3. AN ERROR BOUND

In this section we derive an error bound central to the superconvergence proofs, restricting ourselves to the case of a model boundary-value problem. The extension to a more practical class of problems will be discussed in section 8.

We take Ω to be a square of side K in $\mathbb{R}^2$, place (x,y) co-ordinate axes parallel to the boundary $\partial\Omega$ and consider approximations to the solution u of the problem

$$\left.\begin{array}{rlll} u = 0 & \text{on} & \partial\Omega & \\ -\nabla^2 u = f & \text{in} & \Omega & . \end{array}\right\} \qquad (3.1)$$

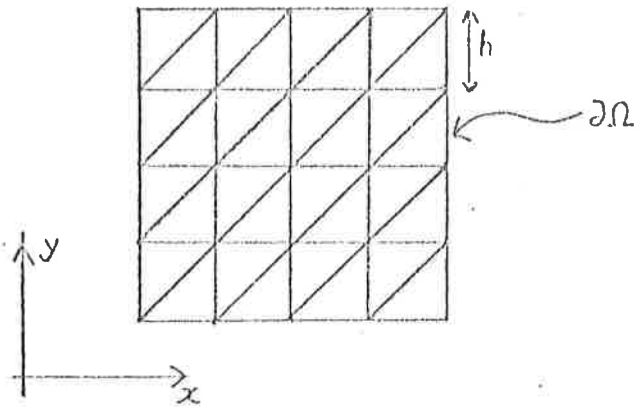
We will work with the weak form associated with this PDE, namely

$$\left.\begin{array}{l} u \in H_0^1(\Omega) \\ a(u,v) = (f,v) \quad \forall v \in H_0^1(\Omega), \end{array}\right\} \qquad (3.2)$$

where we define

$$a(u,v) = \int\int_{\Omega} \underline{\nabla}u \cdot \underline{\nabla}v \quad . \tag{3.3}$$

For each $h \in \{K/2, K/3, K/4, \ldots\}$ we can partition Ω into uniform squares of side h and of the same orientation as Ω and thence into triangles by means of a diagonal of slope $+1$ in each square (see Fig. 1).



(Figure 1)

We then define the finite element subspace $S_0^h \subset H_0^1(\Omega)$ as the space of continuous functions on Ω which vanish on $\partial\Omega$ and vary linearly in each triangular element. We get a sequence of "finite element" (or "Galerkin") approximations to u with diminishing parameter h by

$$\left.\begin{array}{l} u_h \in S_0^h \\ a(u_h,v) = (f,v) \qquad \forall v \in S_0^h \quad . \end{array}\right\} \tag{3.4}$$

Also, for each h we define the interpolant u_I to u as the member of S_0^h which takes the same value as u at each node of the triangulization, thus corresponding to the definition in Section 2. Note that this definition implies a smoothness restriction on u additional to that in (3.2). We shall in fact require

$$u \in H_0^3(\Omega) \quad . \tag{3.5}$$

It is a standard result (see for example [6]) that

$$|u-u_h|_1 \leq |u-u_I|_1 \leq Ch|u|_2 . \tag{3.6}$$

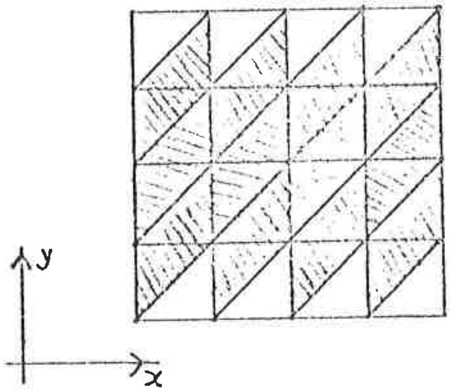
This rate of convergence is optimal for global L_2 sampling of the quantities $u-u_h$ and $u-u_I$ (see Section 9). In Section 4 we will improve this order by sampling at stress points; the following theorem will be central to the result.

Theorem 1 Let u_I and u_h be the interpolant and the Galerkin approximation to u as defined above, where u satisfies (3.2) and (3.5). Then

$$|u_h-u_I|_1 \leq C_4h^2|u|_3. \tag{3.7}$$

Proof Consider the quantity $a(u-u_I,v)$ for some $v\epsilon S_0^h$. We write this as the sum $\iint_\Omega \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x} + \iint_\Omega \frac{\partial}{\partial y}(u-u_I)\frac{\partial v}{\partial y}$ and bound each integral separately.

Write $\Omega = (\underset{j}{\cup}\{A_j\}) \cup (\underset{j}{\cup}\{B_j\})$, where the A_j are pairs of adjacent triangles with a common edge parallel to the x-axis (there are $O(1/h^2)$ such A_j) and the B_j are the $O(1/h)$ single triangles left over on the boundary $\partial\Omega$; none of the A_j or B_j intersect. (See Fig. 2).



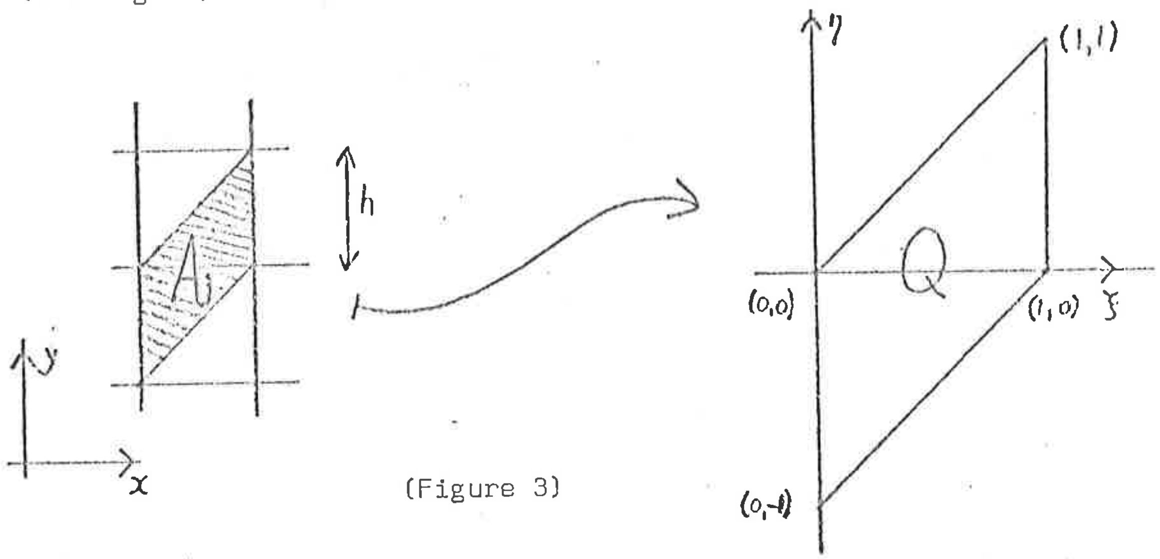
(Figure 2) In this example the 12 pairs A_j are shaded; there are 8 triangles B_j.

Now $\iint_\Omega \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x} = \sum_j \iint_{A_j} \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x} + \sum_j \iint_{B_j} \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x}$.

But $v\epsilon S_0^h \Rightarrow v = 0$ on $\partial\Omega \Rightarrow \frac{\partial v}{\partial x} = 0$ in each B_j.

So $\sum_j \iint_{B_j} \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x} = 0.$

For each A_j we transform $(x,y) \to (\xi,\eta)$ by means of a translation and enlargement so that $A_j \to Q$, the quadrilateral introduced in Section 2 (see Fig. 3).



(Figure 3)

Here and in later sections we adopt the notation that the image of a function $w(x,y)$ transformed in this manner is written as $\tilde{w}(\xi,\eta) = w(x(\xi,\eta),\, y(\xi,\eta))$. Clearly the Jacobian $\frac{\partial(x,y)}{\partial(\xi,\eta)}$ of this transformation is h^2 and $\partial w/\partial x = h^{-1}\ \partial\tilde{w}/\partial\xi$. Further

$$|\tilde{w}|_{k,Q} = h^{k-1}\ |w|_{k,A_j} \qquad \forall w\in H^k(A_j)\ ; \qquad (3.8)$$

it is instructive to note the rôle of (3.8) in what follows and compare it with the corresponding result on norms : $\|\tilde{w}\|_{k,Q} \le Ch^{-1}\,\|w\|_{k,A_j}$.

We have

$$\int\int_{A_j} \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x}\,dx\,dy = \int\int_Q h^{-1}\ \frac{\partial}{\partial\xi}(\tilde{u}-\tilde{u}_I)\cdot h^{-1}\frac{\partial\tilde{v}}{\partial\xi}\cdot h^2 d\xi d\eta =$$

$$= \int\int_Q \frac{\partial}{\partial\xi}(\tilde{u}-\tilde{u}_I)\frac{\partial\tilde{v}}{\partial\xi}\,d\xi d\eta = F(\tilde{u}) \qquad \text{say.}$$

Now
$$|\tilde{u}-\tilde{u}_I|_{1,Q} \le |\tilde{u}|_{1,Q} + |\tilde{u}_I|_{1,Q} \qquad (*)$$

$$\le |\tilde{u}|_{1,Q} + C\,\|\tilde{u}\|_{2,Q} \qquad (*)$$

$$\le C\ \|\tilde{u}\|_{3,Q} \qquad (*)$$

(where lemma 5 has been applied separately to the 2 triangles on either side of $\eta = 0$ which make up Q), and so $|F(\tilde{u})| \le \|\frac{\partial}{\partial\xi}\,(\tilde{u}-\tilde{u}_I)\,\|_{0,Q}\,\|\frac{\partial\tilde{v}}{\partial\xi}\|_{0,Q}$

$$\le C\,\|\tilde{u}\|_{3,Q}\,\|\tilde{v}_\xi\|_{0,Q}\ . \qquad (3.9)(*)$$

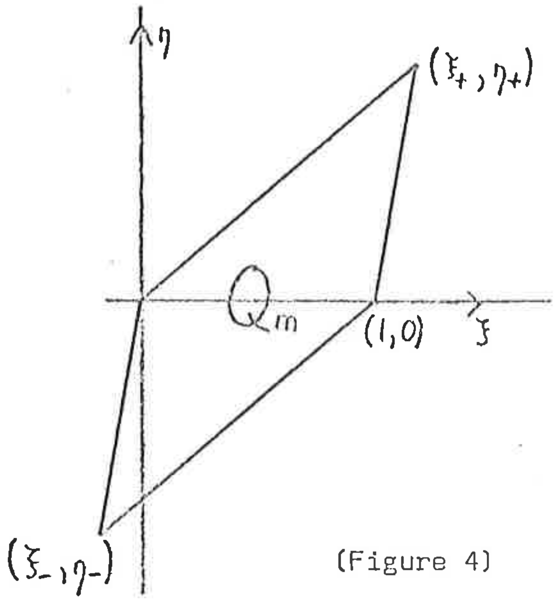
Further $F(q) = 0 \quad \forall q \epsilon P_3(Q)$. (3.10)

For $\partial\tilde{v}/\partial\xi = \tilde{v}(1,0) - \tilde{v}(0,0)$ is a constant over Q, so that $F(q) = \partial\tilde{v}/\partial\xi \int\int_Q \frac{\partial}{\partial\xi}(q-q_I)d\xi d\eta$. We now use:

Lemma 7 Let Q_m be the quadrilateral with vertices $(0,0)$, (ξ_-,η_-), $(1,0)$, (ξ_+,η_+), as in Fig. 4, such that $\xi_+ + \xi_- = 1$, $\eta_+ + \eta_- = 0$. The functional

$$G(q) = \int\int_{Q_m} (\underline{\nabla} q - \underline{\nabla} q_I)d\xi d\eta = \underline{0} \qquad \forall q \epsilon P_3(Q_m).$$

Note: The conditions on $\xi_\pm, \eta_\pm$ are precisely that Q_m is a parallelogram.



(Figure 4)

Proof If $q \epsilon P_2$ then $q_I = q$ (i.e. linears are interpolated exactly). So if G can be shown to annihilate $\xi^2, \xi\eta$ and η^2 then it annihilates a basis of P_3 and hence the whole of P_3.

But if $q = \xi^2$ then $G(q) = \int\int_{Q_m} \begin{pmatrix} 2\xi-1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{6}[(2\xi_+-1)\eta_+ - (2\xi_--1)\eta_-] \\ 0 \end{pmatrix} = \underline{0}$;

if $q = \xi\eta$ then $G(q) = \underset{(\eta>0)}{\int\int} \begin{pmatrix} \eta \\ \xi-\xi_+ \end{pmatrix} + \underset{(\eta<0)}{\int\int} \begin{pmatrix} \eta \\ \xi-\xi_- \end{pmatrix} = \begin{pmatrix} \frac{1}{6}[\eta_+^2-\eta_-^2] \\ \frac{1}{6}[(1-2\xi_+)\eta_+-(1-2\xi_-)\eta_-] \end{pmatrix} = \underline{0}$;

if $q = \eta^2$ then $G(q) = \underset{(\eta>0)}{\int\int} \begin{pmatrix} 0 \\ 2\eta-\eta_+ \end{pmatrix} + \underset{(\eta<0)}{\int\int} \begin{pmatrix} 0 \\ 2\eta-\eta_- \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{6}[-\eta_+^2 + \eta_-^2] \end{pmatrix} = \underline{0}$;

this completes the proof of the lemma.

By (3.9), (3.10) and lemma 6

$$|F(\underset{-}{\tilde{u}})| \leq C \ |\tilde{u}|_{3,Q} \|\tilde{v}_\xi\|_{0,Q} . \qquad (3.11)$$

Transforming back to (x,y) co-ordinates, (3.8) gives

$$\left| \iint_{A_j} \frac{\partial}{\partial x}(u-u_I) \frac{\partial v}{\partial x}\, dx\, dy \right| \leq Ch^2 \, |u|_{3,A_j} \, \| v_x \|_{0,A_j}$$

and so

$$\left| \iint_{\Omega} \frac{\partial}{\partial x}(u-u_I) \frac{\partial v}{\partial x} \right| \leq Ch^2 \sum_j |u|_{3,A_j} \cdot \| v_x \|_{0,A_j}$$

$$\leq Ch^2 \Big(\sum_j |u|^2_{3,A_j}\Big)^{\frac{1}{2}} \Big(\sum_j \| v_x \|^2_{0,A_j}\Big)^{\frac{1}{2}} \qquad (*)$$

(Cauchy-Schwarz)

$$\leq Ch^2 \, |u|_3 \, \| v_x \|_0 \; . \qquad (*)$$

An identical estimate holds for $\iint_{\Omega} \frac{\partial}{\partial y}(u-u_I) \frac{\partial v}{\partial y}$, so that (using Cauchy-Schwarz again)

$$|a(u-u_I, v)| \leq Ch^2 |u|_3 \, |v|_1 \quad ; \qquad (3.12)(*$$

recall that this holds for any $v \epsilon S^h_0$.

Finally $|u_I - u_h|^2_1 = a(u_I - u_h,\, u_I - u_h)$ by (3.3)

$= a(u_I - u_h, v)$ where $v = u_I - u_h \epsilon S^h_0$

$= a(u_I - u, v)$ by (3.2) and (3.4)

$\leq Ch^2 |u|_3 \; |v|_1$ by (3.12)

$= Ch^2 |u|_3 \, |u_I - u_h|_1$

and so $|u_I - u_h|_1 \leq Ch^2 |u|_3$ as required.

4. THE SUPERCONVERGENCE RESULT

In Theorem 3 below, Theorem 1 is used to prove superconvergence of the gradient of $u-u_h$. We also make use of the corresponding superconvergence result for $u-u_I$; this is given first. In this section Ω, h, S_0^h, u, u_I and u_h are as defined in Section 3.

If T is any triangle we write S(T) for the set of midpoints of the sides of T. If $P\epsilon S(T)$ we write D_Pu for the component at P of $\underline{\nabla}u$ parallel to that side of T.

Theorem 2 Let T be any triangle in the covering of Ω and let $P\epsilon S(T)$. Then

$$|D_P(u-u_I)| \le C_5 h\ |u|_{3,T}\ . \tag{4.1}$$

Proof We transform $(x,y) \to (\xi,\eta)$ by means of an affine transformation so that $T \to \Delta$, the triangle introduced in Section 2, and $P \to (\frac{1}{2},0)$. The Jacobian $J = \frac{\partial(x,y)}{\partial(\xi,\eta)}$ of this transformation is h^2,

also $$|D_P w| \le h^{-1}\ |\partial\tilde{w}/\partial\xi|_{(\frac{1}{2},0)} \qquad (4.2)(*)$$

and $$|\tilde{w}|_{k,\Delta} \le Ch^{k-1}\ |w|_{k,T}. \qquad (4.3)(*)$$

(This statement is weaker than the corresponding (3.8) because of the possibility of shearing the (x,y) system to map $T \to \Delta$ appropriately).

Let $$F(\tilde{u}) = \frac{\partial}{\partial\xi}(\tilde{u}-\tilde{u}_I)\Big|_{(\frac{1}{2},0)}\ .$$

Then $$|F(\tilde{u})| \le \left|\frac{\partial\tilde{u}}{\partial\xi}\right|_{(\frac{1}{2},0)} + \left|\frac{\partial u_I}{\partial\xi}\right| \qquad (*)$$

$$\le \max_{\Delta}\ |\underline{\nabla}\tilde{u}| + |\underline{\nabla}\tilde{u}_I| \qquad (*)$$

$$\le C\,\|\underline{\nabla}\tilde{u}\|_{2,\Delta} + C\,\|\tilde{u}\|_{2,\Delta} \qquad \text{(by Lemmas 4 and 5)}$$

$$\le C\,\|\tilde{u}\|_{3,\Delta} \qquad (4.4)(*)$$

Further $$F(q) = 0\quad \forall q\epsilon P_3(\Delta). \tag{4.5}$$

For if $q\epsilon P_2$ then $q_I = q$ (i.e. linears are interpolated exactly).

Also $F(\xi^2) = (2\xi-1)\big|_{(\frac{1}{2},0)} = 0$, $F(\xi\eta) = (\eta-0)\big|_{(\frac{1}{2},0)} = 0$ and $F(\eta^2) = (0-0) = 0$.

Hence F annihilates a basis of P_3 and therefore the whole of P_3.

By (4.4), (4.5) and lemma 6

$$|F(\tilde{u})| \le C \; |\tilde{u}|_{3,\Delta} \, . \qquad (4.6)$$

Transforming back to (x,y) co-ordinates, (4.2) and (4.3) give

$$\begin{aligned} |D_P(u-u_I)| &\le h^{-1} \; |F(\tilde{u})| \\ &\le Ch^{-1} \; |\tilde{u}|_{3,\Delta} \\ &\le Ch \; |u|_{3,T} \qquad \text{as required.} \end{aligned}$$

Theorem 3 (Superconvergence). Under the above conditions

$$h\left(\sum_{\substack{P\in \cup S(T)\\ T\subset\Omega}} |D_P(u-u_h)|^2\right)^{\frac{1}{2}} \le C_6 h^2 \; |u|_3 \quad . \qquad (4.7)$$

Proof The functionals D_P, $P\in \cup S(T)$ are locally bounded in the 1-seminorm on the space of piecewise linears. For if $v\in S_0^h$ and $P\in S(T)$ for some $T\subset\Omega$ then

$$\begin{aligned} |D_P(v)| &\le |\underline{\nabla} v|_{(T)} \qquad\qquad (*) \\ &= |v|_{1,T}/(\text{meas. } T)^{\frac{1}{2}} \\ &\le Ch^{-1} \; |v|_{1,T} \quad . \end{aligned}$$

Substituting $v = u_I-u_h$, squaring and summing we get

$$\begin{aligned} \sum_P |D_P(u_I-u_h)|^2 &\le Ch^{-2} \; |u_I-u_h|_1^2 \\ &\le Ch^2 \; |u|_3^2 \quad \text{(by Theorem 1).} \qquad (4.8) \end{aligned}$$

Also, squaring and summing the result of Theorem 2 over all P,

$$\sum_P |D_P(u-u_I)|^2 \le Ch^2 \; |u|_3^2 \quad . \qquad (4.9)$$

Finally, the linearity of the functionals D_P gives

$$D_P(u-u_h) = D_P(u-u_I) + D_P(u_I-u_h)$$

and by the triangle inequality

$$\sum_P |D_P(u-u_h)|^2 \le 2\left(\sum_P |D_P(u-u_I)|^2 + \sum_P |D_P(u_I-u_h)|^2\right) \qquad (4.10)(*$$

whence by (4.8) and (4.9)

$$\sum_p |D_p(u-u_h)|^2 \leq Ch^2 \; |u|_3^2$$

and so $\quad h\left(\sum_p |D_p(u-u_h)|^2\right)^{\frac{1}{2}} \leq Ch^2 \; |u|_3$ as required.

5. OPTIMALITY OF ERROR BOUNDS

Some stages in the proofs of Theorems 1-3, indicated by (*), are weak in that they lead to unnecessarily large values for constants such as C_4. This is because they follow the non-constructive approach which is usual for superconvergence proofs (such as Zlámal's) and which tends to hide major inefficiencies of proof. Two prime examples are the use of the triangle inequality $(|u-u_I| \leq |u| + |u_I|)$ and the weakening of norm bounds (e.g. $|\tilde{u}|_{1,Q} \leq \|\tilde{u}\|_{3,Q}$) in the derivations of (3.9) and (4.4). In consequence it is possible that, given values for the constants C_1 and C_3 in Lemmas 4 and 6, the final superconvergence bound could be non-optimal by as much as an order of magnitude for a problem of practical size.

We derive here optimal bounds corresponding to those parts of the proofs of Theorems 1 and 2 referred to above. This alternative approach is based directly on Taylor expansions and has the additional advantage of avoiding the non-constructive Bramble-Hilbert lemma. We start with an alternative derivation of (3.11).

We take Q to be the quadrilateral introduced in Section 2 and Q_+, Q_- to be the intersections of Q with the half-spaces $\{\eta > 0\}$, $\{\eta < 0\}$ respectively.

We define $$K(\xi,\eta) = \begin{cases} \xi-\eta & \text{in } Q_+ \\ \xi-\eta-1 & \text{in } Q_- \end{cases} \qquad (5.1)$$

and take ϕ_* to be a solution to the boundary-value problem

$$\left.\begin{aligned} -\nabla^2\phi_* + K/2 &= 0 \qquad \text{in } Q \\ \partial\phi_*/\partial n &= 0 \qquad \text{on } \partial Q \end{aligned}\right\} \qquad (5.2)$$

Theorem 4 Let $\tilde{u} \epsilon H^3(Q)$ and consider the functional

$F(\tilde{u}) = \int\int_Q \frac{\partial}{\partial\xi}(\tilde{u}-\tilde{u}_I)d\xi d\eta$, where $\tilde{u}_I$ interpolates $\tilde{u}$ in each of Q_+, Q_-.

Then $$F(\tilde{u}) = \int\int_Q K(\xi,\eta) \quad \phi(\xi,\eta)d\xi d\eta, \tag{5.3}$$

where $\phi = \tilde{u}_{\xi\eta}$,

and $$|F(\tilde{u})| \le C_7|\phi|_{1,Q}, \tag{5.4}$$

where $C_7 = 2|\phi_*|_{1,Q}$; this bound is attained when $\phi = \phi_*$.

Notes: (i) Although ϕ_* is only defined up to an additive constant, $|\phi_*|_{1,Q}$ is precisely defined.

(ii) Numerical solutions of (5.2) indicate a value of C_7 between 0.137 and 0.138.

Proof (For ease of notation we drop the ~).

$$F(u) = \int\int_Q \frac{\partial}{\partial\xi}(u-u_I)d\xi d\eta$$

$$= \int_{\partial Q} u(\xi,\eta)d\eta - u(1,0) + u(0,0)$$

$$= \int_0^1 f(\eta)d\eta$$

where $f(\eta) = u(1,\eta) - u(\eta,\eta) - u(0,-\eta) + u(1-\eta,-\eta) - u(1,0) + u(0,0)$.

We use Taylor's theorem (Lemma 1) with a change of integration variable:

$$\left.\begin{aligned} u(1,\eta) &= u(1,0) + \int_0^\eta u_\eta(1,t)dt \\ u(\eta,\eta) &= u(\eta,0) + \int_0^\eta u_\eta(\eta,t)dt \\ u(0,-\eta) &= u(0,0) - \int_{-\eta}^0 u_\eta(0,t)dt \\ u(1-\eta,-\eta) &= u(1-\eta,0) - \int_{-\eta}^0 u_\eta(1-\eta,t)dt \end{aligned}\right\} \text{(see Fig. 5)}$$

so that
$$f(\eta) = \int_0^\eta (u_\eta(1,t) - u_\eta(\eta,t)dt$$
$$+ \int_{-\eta}^0 (u_\eta(0,t) - u_\eta(1-\eta,t)dt$$
$$+ u(1-\eta,0) - u(\eta,0).$$

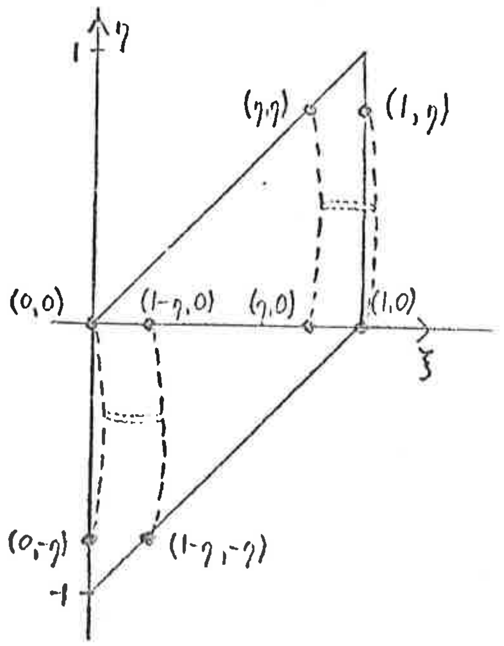
We employ Taylor again:

$$\left.\begin{aligned} u_\eta(\eta,t) &= u_\eta(1,t) - \int_\eta^1 u_{\xi\eta}(\xi,t)d\xi \\ u\ (1-\eta,t) &= u_\eta(0,t) + \int_0^{1-\eta} u_{\xi\eta}(\xi,t)d\xi \end{aligned}\right\} \quad \text{(see Fig. 5)}$$

(Note that $u \in H^3(Q)$ and so line integrals of second derivatives do converge, by the trace theorem (Lemma 2)).

Therefore
$$f(\eta) = \int_{t=0}^{\eta} \int_{\xi=\eta}^{1} u_{\xi\eta}(\xi,t)d\xi dt - \int_{t=-\eta}^{0} \int_{\xi=0}^{1-\eta} u_{\xi\eta}(\xi,t)d\xi dt$$
$$+ u(1-\eta,0) - u(\eta,0)$$

and since
$$\int_0^1 (u(1-\eta,0) - u(\eta,0)d\eta = 0$$
we have
$$F(u) = \int_{\eta=0}^{1} \int_{t=0}^{\eta} \int_{\xi=\eta}^{1} \phi(\xi,t)d\xi dt d\eta$$
$$- \int_{\eta=0}^{1} \int_{t=-\eta}^{0} \int_{\xi=0}^{1-\eta} \phi(\xi,t)d\xi dt d\eta.$$



(Figure 5)

Cancellation via Taylor's theorem is indicated by ------ for the first stage

====== for the second

Now the first integral is equal to

$$\int_{t=0}^{1}\int_{\eta=t}^{1}\int_{\xi=\eta}^{1} \phi(\xi,t)d\xi d\eta dt$$ (re-ordering the first two integrations)

$$= \int_{t=0}^{1}\int_{\xi=t}^{1}\int_{\eta=t}^{\xi} \phi(\xi,t)d\eta d\xi dt$$ (re-ordering the second two)

$$= \int_{t=0}^{1}\int_{\xi=t}^{1} (\xi-t)\ \phi(\xi,t)d\xi dt$$

$$= \int\int_{Q_+} (\xi-\eta)\ \phi(\xi,\eta)d\xi d\eta$$ (writing η for t);

similarly the second is equal to

$$\int_{t=-1}^{0}\int_{\eta=-t}^{1}\int_{\xi=0}^{1-\eta} \phi(\xi,t)d\xi d\eta dt$$

$$= \int_{t=-1}^{0}\int_{\xi=0}^{1+t}\int_{\eta=-t}^{1-\xi} \phi(\xi,t)d\eta d\xi dt$$

$$= \int_{t=-1}^{0}\int_{\xi=0}^{1+t} (1-\xi+t)\ \phi(\xi,t)d\xi dt$$

$$= \int\int_{Q_-} (1-\xi+\eta)\ \phi(\xi,\eta)d\xi d\eta\ .$$

Hence $F(u) = \int\int_Q K\phi$ as desired.

We want to find the constant C_7 such that (5.4) holds and is as sharp as possible. In other words,

$$C_7 = \sup \frac{\left|\int\int_Q K\phi\right|}{|\phi|_{1,Q}} \ ;$$

the supremum is over $u\epsilon H^3(Q)$, i.e. $\phi\epsilon H^1(Q)$.

This is equivalent to $$C_7^{-1} = \inf_{\substack{\phi\epsilon H^1(Q)\\ \left|\int\int_Q K\phi\right|=1}} |\phi|_{1,Q}\ . \qquad (5.5)$$

Squaring (5.5), if we find a pair $(\phi_\lambda, \lambda) \epsilon H^1(Q) \times \mathbb{R}$ which minimises (w.r.t. ϕ_λ)

$$\left.\begin{aligned} &\int\!\!\int_Q |\underline{\nabla}\phi_\lambda|^2 + \lambda K\phi_\lambda \\ \text{subject to} \quad &\int\!\!\int_Q K\phi_\lambda = 1 \end{aligned}\right\} \tag{5.6}$$

then

$$C_7^{-2} = \int\!\!\int_Q |\underline{\nabla}\phi_\lambda|^2 \quad .$$

Now the function ϕ_* introduced above (5.2) satisfies

$$\int\!\!\int_Q (\underline{\nabla}\phi_* \cdot \underline{\nabla}\psi + K\psi/2) = 0 \qquad \forall\psi\epsilon H^1(Q) \tag{5.7}$$

and is thus an extremal of

$$\int\!\!\int_Q |\underline{\nabla}\phi\,|^2 + K\phi \quad . \tag{5.8}$$

Indeed, the second variation of this functional is $\int\!\!\int_Q |\underline{\nabla}\psi|^2$, which is always non-negative, so that ϕ_* minimises that functional over $H^1(Q)$. Therefore, comparing (5.6) with (5.8),

$$\phi_\lambda = \lambda\phi_* \; .$$

Pausing only to take note that $\int\!\!\int_Q K(\lambda\phi_*) = 1$ and that $\int\!\!\int_Q |\underline{\nabla}\phi_*|^2 = -\int\!\!\int_Q K\phi_*/2$ (substitute ϕ_* for ψ in (5.7)), we have

$$\begin{aligned} C_7^{-2} &= \int\!\!\int_Q |\underline{\nabla}\phi_\lambda|^2 \\ &= \lambda^2 \int\!\!\int_Q |\underline{\nabla}\phi_*|^2 \\ &= \frac{-\lambda^2}{2} \int\!\!\int_Q K\phi_* \\ &= \frac{-1}{2\int\!\!\int_Q K\phi_*} \end{aligned}$$

and finally
$$C_7 = (-2 \int\int_Q K\phi_*)^{\frac{1}{2}}$$
$$= (4 \int\int_Q |\underline{\nabla}\phi_*|^2)^{\frac{1}{2}}$$
$$= 2|\phi_*|_{1,Q} \quad .$$

Hence $|F(u)| \leq C_7|\phi|_{1,Q}$ and the bound is attained by any $u\epsilon H^3(Q)$ satisfying $u_{\xi\eta} = \phi_*$.

We move on to the alternative derivation of (4.6). We take Δ to be the triangle introduced in Section 2 and Γ to be that part of its boundary $\partial\Delta$ for which $\eta = 0$. We define

$$K(\xi,0) = \begin{cases} \xi & (\xi < \frac{1}{2}) \\ \xi-1 & (\xi > \frac{1}{2}) \end{cases} \tag{5.9}$$

on Γ and take ϕ_* to be a solution of the boundary-value problem

$$\left.\begin{aligned} -\nabla^2\phi_* &= 0 \quad \text{in} \quad \Delta \\ \partial\phi_*/\partial n &= \begin{cases} -K/2 & \text{on} \quad \Gamma \\ 0 & \text{on} \quad \partial\Delta\backslash\Gamma \end{cases} \end{aligned}\right\} \tag{5.10}$$

where $\partial/\partial n$ represents differentiation along the outward normal.

<u>Theorem 5</u> Let $\tilde{u}\epsilon H^3(\Delta)$ and consider the functional

$$F(\tilde{u}) = \frac{\partial}{\partial\xi}(\tilde{u}-\tilde{u}_I)\Big|_{(\frac{1}{2},0)} \quad , \text{ where } \tilde{u}_I \text{ interpolates } \tilde{u}.$$

Then $F(\tilde{u}) = \int_\Gamma K(\xi,0)\,\phi(\xi,0)d\xi,$ (5.11)

where $\phi = \tilde{u}_{\xi\xi}$,

and $|F(\tilde{u})| \leq C_8|\phi|_{1,\Delta}$, (5.12)

where $C_8 = 2|\phi_*|_{1,\Delta}$; this bound is attained when $\phi = \phi_*$.

Notes: (i) As above, $|\phi_*|_{1,\Delta}$ is precisely defined.

(ii) Numerical solutions of (5.10) indicate a value of C_8 in the region of 0.12.

Proof (Again we omit the ~).

$$F(u) = \frac{\partial}{\partial\xi}(u-u_I)\Big|_{(\frac{1}{2},0)} = \frac{\partial u}{\partial\xi}(\tfrac{1}{2},0) - u(1,0) + u(0,0)$$

By Taylor's theorem (Lemma 1) with a change of integration variable,

$$u(1,0) = u(\tfrac{1}{2},0) + \tfrac{1}{2}u_\xi(\tfrac{1}{2},0) + \int_{\frac{1}{2}}^{1} u_{\xi\xi}(\xi,0).(1-\xi)d\xi$$

$$u(0,0) = u(\tfrac{1}{2},0) - \tfrac{1}{2}u_\xi(\tfrac{1}{2},0) + \int_{0}^{\frac{1}{2}} u_{\xi\xi}(\xi,0).\xi\ d\xi \quad .$$

Therefore $F(u) = \int_\Gamma k\ \phi d\xi$ as required. We now proceed precisely as in Theorem 4 above. C_8 is such that (5.12) holds and the bound is attained, whence

$$C_8^{-1} = \inf_{\substack{\phi\in H^1(\Delta) \\ |\int_\Gamma K\phi| = 1}} |\phi|_{1,\Delta} \quad ; \tag{5.13}$$

if $(\phi_\lambda,\lambda) \in H^1(\Delta)\times\mathbb{R}$ minimises $\iint_\Delta |\underline{\nabla}\phi_\lambda|^2 + \lambda\int_\Gamma K\phi_\lambda$ subject to $\int_\Gamma K\phi_\lambda = 1$ (5.14)

then
$$C_8^{-2} = \iint_\Delta |\underline{\nabla}\phi_\lambda|^2 \quad .$$

The function ϕ_* defined by (5.10) satisfies

$$\iint_\Delta \underline{\nabla}\phi_*\cdot\underline{\nabla}\psi + \tfrac{1}{2}\int_\Gamma K\psi = 0 \qquad \forall\psi\in H^1(\Delta) \tag{5.15}$$

and is thus an extremal of

$$\iint_\Delta |\underline{\nabla}\phi\ |^2 + \int_\Gamma K\phi \quad . \tag{5.16}$$

As before the second variation is non-negative, so that ϕ_* minimises that functional over $H^1(\Delta)$. Hence

$$\phi_\lambda = \lambda\phi_*,$$

also $\int_\Gamma K\ (\lambda\phi_*) = 1$ and (from (5.15)) $\iint_\Delta |\underline{\nabla}\phi_*|^2 = -\tfrac{1}{2}\int_\Gamma K\phi_*$.

Exactly as above, $C_8^{-2} = \iint_\Delta |\underline{\nabla}\phi_\lambda|^2 = \dfrac{-1}{2\int_\Gamma K\phi_*}$

whence finally $C_8 = 2|\phi_*|_{1,\Delta}$.

So $|F(u)| \leq C_8|\phi|_{1,\Delta}$ and the bound is attained by any $u \in H^3(\Delta)$ satisfying $u_{\xi\xi} = \phi_*$.

Remarks

(1) We note that these two results depend on the error functional being written in the form $F(\tilde{u}) = \int K\phi$, where ϕ is one of the second derivatives of $\tilde{u}$ and $\int K = 0$. Clearly it would be possible to write $F = \sum_{|\alpha|=1} \int K_\alpha \phi^{(\alpha)}$, where α represents a 2-index, by means of Lemma 3. However, the weights K_α are neither cheap nor simple to evaluate. Also, given $K_{\binom{1}{0}}$ and $K_{\binom{0}{1}}$ the best possible direct result would be $|F| \leq \sum_{|\alpha|=1} \| K_\alpha \|_0 \| \phi^{(\alpha)} \|_0$; in examples of this type it is unusual for $\frac{\partial}{\partial\eta}(K_{\binom{1}{0}}) = \frac{\partial}{\partial\xi}(K_{\binom{0}{1}})$ and we would not expect the resulting bound to be attained. So although this alternative form implies $|F(\tilde{u})| \leq C|\tilde{u}|_{3,R}$ (where R is either Δ or Q) it is of little help in evaluating C optimally. Our approach exploits the full smoothness of $\tilde{u}$ only indirectly, taking Taylor expansions with only the second derivatives of $\tilde{u}$ in the remainder term. The success of the scheme is undoubtedly due to the special nature of F, in particular only one component of the gradient is being estimated and there is sufficient cancellation of error terms for F to be bounded by just two of the four third derivatives of $\tilde{u}$. This last property leads to second order equations for ϕ_* instead of the fourth order problems which occur in the examples considered by Meinguet. In [12] he presents a method based on direct manipulation of the remainder in the smoothed Taylor expansion of $\tilde{u}$; this is more generally applicable than the above but is definitely non-optimal and results in bounds which are considerably weaker. For example, we can write (5.12) in the form $|F(\tilde{u})| \leq (0.12) \times |\tilde{u}_{\xi\xi}|_{1,\Delta}$; an analogue derived by direct manipulation is $|F(\tilde{u})| \leq (1.05) \times |\tilde{u}_\xi|_{2,\Delta}$. Incidentally, both approaches are significant advances on the Sard kernel/Schwarz inequality method of Barnhill et al [3].

(2) The results of Theorems 4 and 5 are optimal in the sense of Golumb and Weinberger [10], in that they give values of constants for which the bounds (5.4) and (5.12) are attained. The continuation of this optimality to the results of Theorems 1 to 3 is discussed below; we note here that, as they stand, these theorems are not optimal in the above sense. This should not be taken as a serious flaw, since the calculation of such optimal bounds would be highly impractical. We would rather strike a balance of "quasi-optimality" between the optimal efficiency of Golumb and Weinberger and the less efficient simplicity of Zlámal's approach.

It appears that the steps marked (*) in the proofs of Theorems 1 to 3 fall into three categories, one of which (stemming from the non-constructive route) has been dealt with above. The second is that of pure inefficiency in the names of brevity of proof and ease of notation. Examples are $|\cdot|_{k,\cup A_j} \leq |\cdot|_{k,\Omega}$ in the derivation of (3.12), $|\tilde{w}|_{k,\Delta} \leq Ch^{k-1}|w|_{k,T}$ in (4.3) and $|D_p(v)| \leq |\underline{\nabla}v|_{(T)}$ in the derivation of (4.8). These can all be corrected: by the introduction of a more cumbersome notation and long-winded proof, a rotation-invariant definition of norms and seminorms, the use of more than one model triangle Δ, etc. We also note that, in Theorem 5, there may be something to be gained by bounding $F(\tilde{u})$ in terms of the values of derivatives of $\tilde{u}$ in both the triangles of which the sampling point is on the boundary. However, there is little point in burdening the proofs with such details while more intransigent shortcomings remain.

The last type has just two instances, the (double) use of the Cauchy-Schwarz inequality in the derivation of (3.12), e.g. $\sum_j |u|_{3,A_j} |v|_{1,A_j} \leq (\sum|u|_{3,A_j}^2)^{\frac{1}{2}} (\sum|v|_{1,A_j}^2)^{\frac{1}{2}}$ and the use of the triangle inequality to obtain (4.10). In both cases it is likely though not certain that the bound is sharp, but unlikely that it is attained by that function $u \in H_0^3(\Omega)$ which attains the optimal superconvergence bound (4.7). Unfortunately both these stages are essential to the structure of Zlámal's approach and cannot be avoided. So it is at this point that we are forced into some degree or non-optimality; we can only rely on numerical evidence to measure the seriousness of this retreat.

6. POINTWISE SUPERCONVERGENCE

We move on to consider one of the central features of the superconvergence result. We have shown that in a mean square sense the points $P \in \underset{T}{\cup} S(T)$ are tangential derivative stress points. This result can be weakened by the Cauchy-Schwarz inequality to give superconvergence of arithmetic-mean error (corresponding to Zlámal's earlier results):

$$\begin{aligned}\sum_P |D_P(u-u_h)| &\le \Big(\sum_P |D_P(u-u_h)|^2\Big)^{\frac{1}{2}} \Big(\sum_P 1^2\Big)^{\frac{1}{2}}\\ &\le Ch|u|_3 \cdot Ch^{-1}\end{aligned}$$

and so
$$h^2\sum_P |D_P(u-u_h)| \le Ch^2|u|_3 \; . \qquad (6.1)$$

Indeed the more general Hölder inequality gives superconvergence of averages based on ℓ_q norms, $1 \le q \le 2$ ($q = 2$ being the result of Theorem 3); on the other hand it is not clear whether there exist corresponding results for higher norms (ℓ_q, $q > 2$) and in particular for the supremum norm, $\max\limits_{P \in \cup S(T)} |D_P(u-u_h)|$. Even if it is possible to bound these individual quantities it is not clear what form this bound would take. It is therefore convenient to relax the notation and work with orders of magnitude rather than inequalities.

We must impose a further smoothness restriction on u. For the statement of Theorem 2, namely $|D_P(u-u_I)| \le Ch|u|_{3,T}$, does not imply $O(h^2)$ gradient convergence at P unless $u \in H^s(\Omega)$, $s > 4$, in which case

$$\begin{aligned}|D_P(u-u_I)|^2 &\le Ch \cdot \sum_{i+j=3} \int_T \left[\left(\frac{\partial}{\partial x}\right)^i \left(\frac{\partial}{\partial y}\right)^j u\right]^2\\ &\le Ch^2 \cdot \sup_{\Omega} \sum_{i+j=3} \left[\left(\frac{\partial}{\partial x}\right)^i \left(\frac{\partial}{\partial y}\right)^j u\right]^2 . \; (\text{meas } T)\\ &\le Ch^2 \cdot \|u\|_s^2 \cdot h^2/2,\end{aligned}$$

i.e.
$$|D_P(u-u_I)| \le Ch^2\|u\|_s \quad .$$

Under these circumstances pointwise superconvergence occurs equivalently in the quantities $(u-u_h)$ and (u_I-u_h). Furthermore, there is numerical evidence that if u does not satisfy this smoothness constraint then the error in pointwise sampling for $(u-u_h)$ is worse than $O(h^2)$. (That this holds for $(u-u_I)$ can be shown by considering $u = (x^2+y^2)^{5/2}$). We will therefore discuss the convergence of $D_P(u_I-u_h)$ under the assumption that $u\epsilon H^s(\Omega)$, $s > 4$.

The relaxed-notation version of (4.8) is $h(\sum|D_P(u_I-u_h)|^2)^{\frac{1}{2}} = O(h^2)$. Though this does not imply $|D_P(u_I-u_h)| = O(h^2)$ $\forall P$ it does give $|D_P(u_I-u_h)| = O(h)$ $\forall P$. Further the number of points P at which the convergence rate is no better than $O(h^k)$ cannot be greater than $O(h^{2(1-k)})$.

So the proportion of points $P\epsilon\underset{T}{\cup}S(T)$ at which second order convergence of $|D_P(u_I-u_h)|$ does not occur is $o(1)$, i.e. the number of such points is $o(h^{-2})$. The only natural ways of selecting such a limited number of points would be in a layer close to the boundary $\partial\Omega$ or in narrow regions in the interior of Ω where the smoothness of some aspect of the problem or its method of solution was open to question. If all such aspects are sufficiently regular in the interior we can claim that superconvergence will be pointwise (away from the boundary) as opposed to only in mean-square. Further if the boundary and the conditions imposed on it are sufficiently smooth and well represented then superconvergence will occur at every point P in $\underset{T}{\cup}S(T)$. It should be stressed here that such terms as "narrow" and "sufficiently regular" have not been defined. In particular there might be severe restrictions on generalisations of the model problem.

We consider possible methods for proving pointwise convergence, recalling that it is not clear what form such bounds would take. We seek modifications to the ℓ_2 proof in Theorem 3 and note that the summation which leads to an averaged error bound is necessary because of the form of (3.7). Unfortunately

(3.7) is a sharp result as regards order of convergence (this can be shown simply either analytically or numerically). Also there is no bound of the form $\| u_I - u_h \|_{1,T} \leq Ch^2 |u|_{s,R}$, $T \subset R$, diam $(R) = O(h)$, for some seminorm $|\cdot|_s$. For although $\| u_I - u_h \|_{1,T}$ vanishes when $u \epsilon P_3(T)$ and the mesh is sufficiently regular near T (indeed under this set of conditions $u_I \equiv u_h$), the most local bound that can be obtained on $u_I - u_h$ is of the form given by Nitsche and Schatz [16] : $\| u_I - u_h \|_{1,T} \leq C(h \| u \|_{2,R} + \| u_I - u_h \|_{-s,R})$ for some positive integer s. They note that although $\| u_I - u_h \|_{-s,R}$ may have a high order of convergence it would usually be bounded via $\| \cdot \|_{-s,R} \leq \| \cdot \|_{-s,\Omega}$; at any rate it cannot be bounded locally and so the Bramble-Hilbert lemma applied directly to $\| u_I - u_h \|_{1,T}$ will not yield even an $O(h)$ convergence.

It appears therefore that an intermediate stage stronger than Theorem 1 is necessary. Methods based on the approximability of the Green's function (see [15] for references) look hopeful but are not yet strong enough to handle the cancellation effects between elements necessary for obtaining a sharp bound on $|u_I - u_h|_{1,T}$.

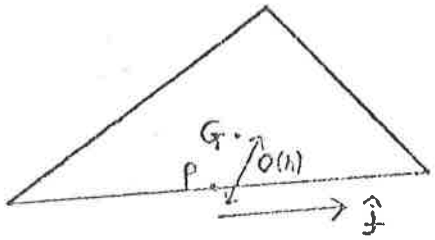
7. SAMPLING AT CENTROIDS

We discuss here Moan's claims about the location of stress points; since we have second order convergence of $|D_p(u-u_h)|$ for almost all $P \epsilon \underset{T}{\cup} S(T)$, we can show that there cannot be such a rate of convergence at the centroids of the elements.

Note first that it is not reasonable to expect particularly good convergence of just one component of the gradient at a centroid, because there is no single direction for such a component which can be chosen in a natural manner. So it is sufficient to show that the component of $\underline{\nabla}(u-u_h)$ parallel to one triangle side is no better than $O(h)$; the same will apply similarly for a second side and component and hence by the above for all naturally chosen components. For simplicity we will work under the assumption of pointwise superconvergence, with its implied requirement that $u \epsilon H^s(\Omega)$, $s > 4$.

We write e_h for the quantity $u-u_h$, and G for the centroid of the triangle, T, in question (See Fig. 6). The unit vector $\hat{\underline{j}}$ is parallel to one of the sides (midpoint P) and it is assumed that

$$D_P(e_h) \equiv (\hat{\underline{j}}\cdot\underline{\nabla})e_h|_P = O(h^2).$$



(Figure 6)

Since $u \epsilon H^3$ the Sobolev representation of Taylor's theorem (Lemma 3) gives

$$(\hat{\underline{j}}\cdot\underline{\nabla})\;(u|_G - u|_P) = O(h)\cdot\int\!\!\int_B \phi(y)\cdot(\underline{\nabla}^T A\underline{\nabla})u(y)dy + R, \qquad (7.1)$$

where A is some matrix,

it is simply verified that $|R| \le Ch|u|_{3,T} \le Ch^2\|u\|_s$ and

as before B is an open ball in T with $\phi \epsilon C_0^\infty(B)$, $\int\!\!\int_B \phi(y)dy = 1$,

y denoting an $\mathbb{R}^2$ vector.

But we also have

$$\begin{aligned}(\hat{\underline{j}}\cdot\underline{\nabla})u|_G &= \hat{\underline{j}}\cdot(\underline{\nabla}u_h|_G + \underline{\nabla}e_h|_G)\\ &= \hat{\underline{j}}\cdot(\underline{\nabla}u_h|_P + \underline{\nabla}e_h|_G) \quad \text{(since } \underline{\nabla}u_h \text{ is constant in } T\text{)}\\ &= \hat{\underline{j}}\cdot(\underline{\nabla}u|_P - \underline{\nabla}e_h|_P + \underline{\nabla}e_h|_G),\end{aligned}$$

so that
$$(\hat{\underline{j}}\cdot\underline{\nabla})(u|_G - u|_P) = (\hat{\underline{j}}\cdot\underline{\nabla})e_h|_G + O(h^2). \qquad (7.2)$$

Subtracting (7.1) from (7.2)

$$(\hat{\underline{j}}\cdot\underline{\nabla})e_h|_G = O(h)\cdot\int\!\!\int_B \phi(y)(\underline{\nabla}^T A\underline{\nabla})u(y)dy + O(h^2)$$

$$\neq O(h^2) \quad \text{for general } u.$$

In fact, as is easily shown, continuity restrictions imply that for any method of approximation by piecewise linears in $\mathbb{R}^2$, the gradient of a quadratic $q \in P_3(\Omega)$ will in general fail to be matched exactly, at the centroids of $O(h^{-2})$ of the triangles. From this it can be (non-rigorously) inferred that centroids are not stress points, even in a mean sense.

8. EXTENDING THE MODEL PROBLEM

We present here a summary of those aspects of extending the model problem which are particular to approximations on triangular elements. We are concerned mainly with means of writing the integral $a(u-u_I,v)$ as a sum of integrals whose terms both vanish for quadratic u and have integrands with local support, in cases where Ω is not a square.

For example, let us suppose that Ω is a parallelogram; we position the co-ordinate axes so that the x-axis is parallel to two segments of $\partial\Omega$ and the other two segments have slope m. We partition Ω into congruent parallelograms and thence into similar triangles by means of diagonals with (least) positive slope (see Fig. 7). (It is assumed throughout that these triangles are non-degenerate; implications of degeneracy are noted below). Finally, we construct S_0^h etc. on this mesh in the usual way.

Theorem 6 Let $u \in H_0^3(\Omega)$ and let $u_I \in S_0^h$ interpolate u, where Ω and S_0^h are as defined above. Let $v \in S_0^h$. Then $|a(u-u_I,v)| \le Ch^2|u|_3|v|_1$ holds (this is (3.12)) and, as a corollary, so does Theorem 1.

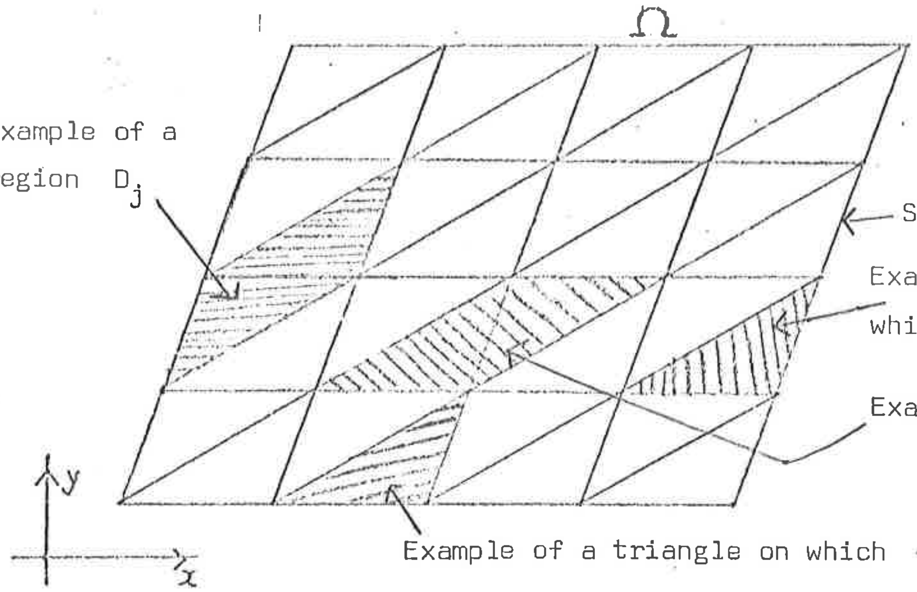
Proof Let $e = u - u_I$.

We recall that $\underline{\nabla} v$ is only piecewise constant but that it has a component which is constant over any pair of adjoining triangles, namely that component parallel to the common edge. If that edge has slope m then this component of $\underline{\nabla}$ is $\frac{\partial}{\partial y} + \frac{1}{m}\frac{\partial}{\partial x}$, whence the decomposition

$$a(e,v) = \iint_\Omega \underline{\nabla} e \cdot \underline{\nabla} v$$

$$= \sum_j \left(\frac{\partial v}{\partial y} + \frac{1}{m}\frac{\partial v}{\partial x}\right)\Bigg|_{C_j} \cdot \int_{C_j} \frac{\partial e}{\partial y} + \sum_j \frac{\partial v}{\partial x}\Bigg|_{D_j} \cdot \int_{D_j} \left(\frac{\partial e}{\partial x} - \frac{1}{m}\frac{\partial e}{\partial y}\right),$$

where we take the C_j and D_j to be adjoining pairs of triangles with common edges of slope m and 0 respectively. As in the proof of Theorem 1 there remain a number of integrals over single triangles on the boundary which disappear because $v = 0$ on $\partial\Omega$. (See Fig. 7).



(Figure 7)

Example of a region D_j

Slope m

Example of a triangle on which $\left(\frac{\partial v}{\partial y} + \frac{1}{m}\frac{\partial v}{\partial x}\right) = 0 \quad \forall v \in S_0^h$

Example of a region C_j

Example of a triangle on which $\frac{\partial v}{\partial x} = 0 \; \forall v \in S_0^h$

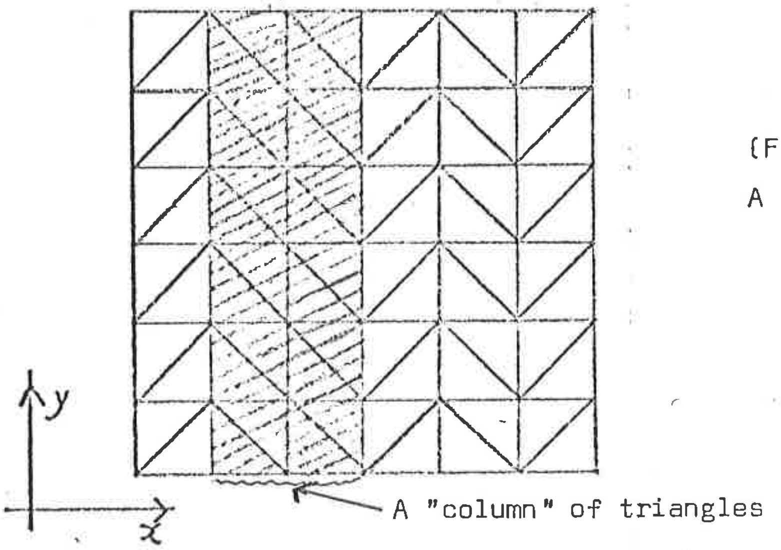
We need to show that if u is quadratic then the integrals $\int_{C_j} \frac{\partial e}{\partial y}$, $\int_{D_j} \frac{\partial e}{\partial x}$, $\int_{D_j} \frac{\partial e}{\partial y}$ all vanish for each j. Once this is established the bounds on $a(u-u_I,v)$ and $|u_I-u_h|_1$ can be derived exactly as in Theorem 1 and are not reproduced here.

Now if $\int_{D_j} \underline{\nabla e}$ vanishes for all quadratic u then so does $\int_{C_j} \frac{\partial e}{\partial y}$ (by a reflection which maps a D_j onto a C_j). But we can map each D_j onto the parallelogram Q_m introduced in Lemma 7 (Section 3) by a change of origin and scale. Hence, by Lemma 7, $\int_{D_j} \underline{\nabla e}$ does vanish for quadratic u and the proof is complete.

Remark

We define a mesh for which the result of Theorem 1 holds for all $u \in H^3$ to be "superconvergent". With the non-degeneracy conditions mentioned earlier, the superconvergence results of Theorems 2 and 3 also hold on such meshes for all $u \in H^3$. In particular, the above theorem with $m = -\sqrt{3}$ implies that a mesh of equilateral triangles is superconvergent.

We move on to consider "chevron" meshes, such as that in Fig. 8. Though we examine meshes based on square-grids, the extension to parallelogram-grids follows the same lines as above, both here and in all that follows.



(Figure 8)
A Chevron Mesh

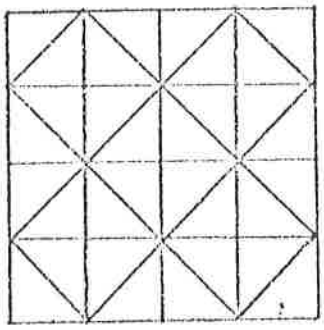
By "chevron" mesh we mean any mesh with exactly six elements meeting at each internal node. This condition is sufficient to ensure that Ω can be exactly partitioned into "columns" of triangles; a "column" consists of one or more adjacent, entire columns (or rows) of the squares which make up the triangulisation of Ω, such that each square in a given column is divided into its two triangles with the same orientation (i.e. hypotenuses in a single column all have the same slope, +1 or -1).

Theorem 7 Chevron meshes are superconvergent.

Sketch proof We bound $a(e,v)$ as a sum of integrals over each of these columns. For example, consider a column, such as the one shaded in Fig. 8, for which each hypotenuse has slope -1; call this region R. (Note that the "width" of R could be anything from 1 square upwards). Then $a_R(e,v)$ is bounded by being viewed as a complete parallelogram-grid-based mesh with $m = -1$; a crucial point is that the C_j can exactly cover R and so there are no left-over triangles on the "long" edges of ∂R (internal to Ω) on which $v = 0$ is not guaranteed. By symmetry the integrals over columns where each hypotenuse has slope $+1$ are also bounded and so the superconvergence proof proceeds in the usual way.

Remarks

(1) We note here that the criss-cross mesh (See Fig. 9) necessary in [9] for high-order derivative convergence in the mixed method of Fix et al. does not have six elements surrounding each node, cannot be arranged into columns and is not superconvergent (see numerical evidence in Section 9).



(Figure 9)
A criss-cross mesh similar to that used in [9].

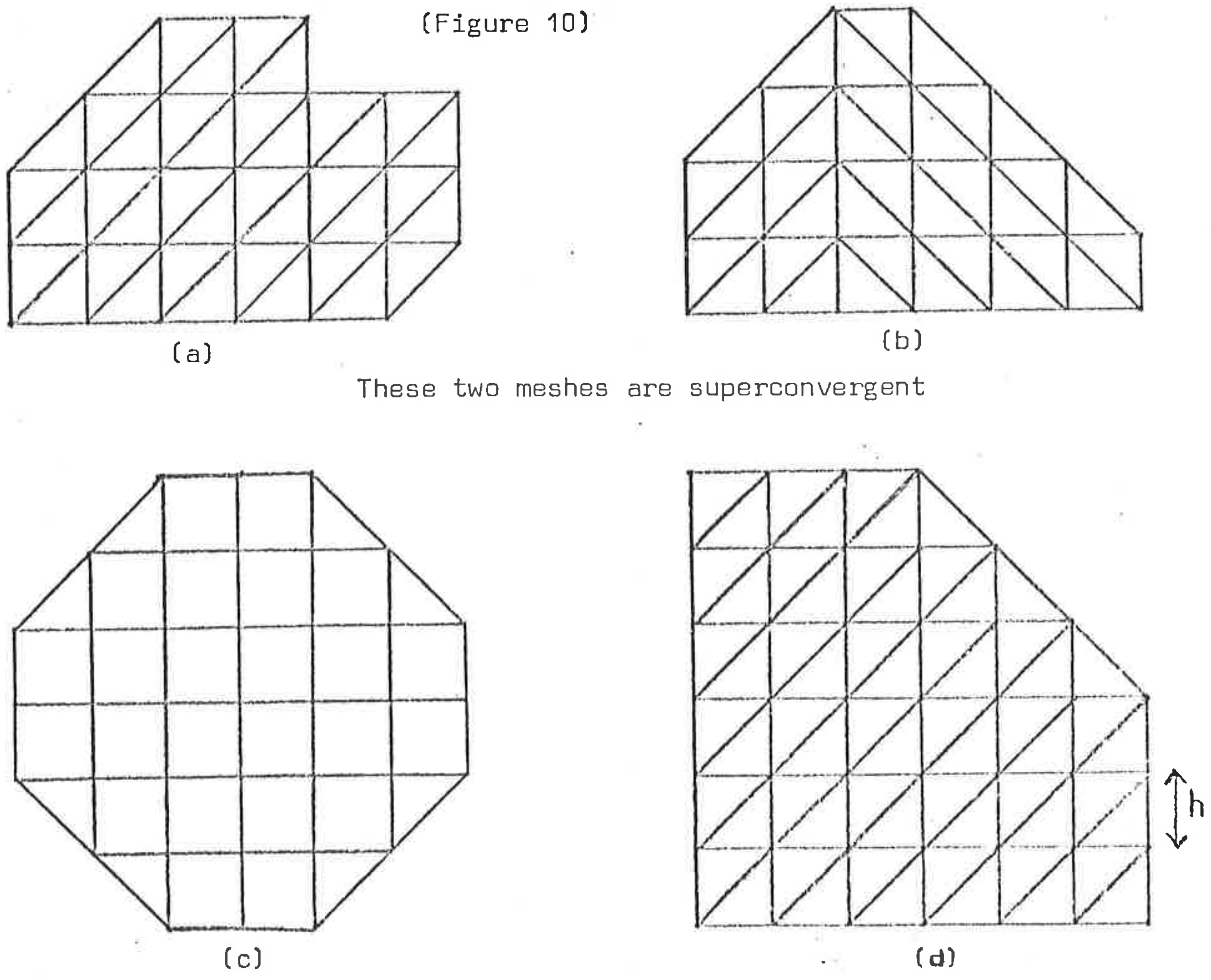
(2) The superconvergent property of chevron meshes becomes useful when we consider regions, Ω, which are not parallelograms. Suppose that the boundary $\partial\Omega$ consists of segments parallel to the x- and y- axes and the line $y = x$ (and relative lengths such that a triangulation, based on a square grid, can exactly cover Ω). (See Fig. 10(a). Then the superconvergent property can be obtained exactly as in Theorem 1, with the omission of some of the "left-over"

boundary triangles (such as the B_j). With chevron meshes, certain regions Ω are superconvergent whose boundaries consist of segments parallel to the axes and both lines $y = \pm x$. The condition on such regions is that a partition of Ω into columns (as above) must exist, such that in columns where there are no element edges of slope -1 there are no boundary segments of that slope, likewise for the slope $+1$ (see Fig. 10(b)).

Because of this necessary restriction it appears that no mesh on an octagonal region is superconvergent (Fig.10(c)). This case is qualitatively equivalent to the mesh shown in Fig.10(d). Here, making the minor smoothness assumption on u that $u \in H^s$, $s > 3$, we obtain at best

$$|a(u-u_I,v)| \leq Ch^{3/2} \, |u|_s \, |v|_1 \; .$$

This $O(h^{\frac{1}{2}})$ drop in accuracy is numerically confirmed in the next section.



(Figure 10)

These two meshes are superconvergent

The triangulation cannot be completed so as to give a superconvergent mesh.

This mesh too is not superconvergent. Note that there are internal nodes which are not surrounded by six elements.

(3) It is to be expected that the above triangulation considerations are irrelevant in the case of Laplace's equation on the square. (The boundary data can no longer be homogeneous - we will assume that they can be represented exactly in a piecewise linear subspace S_E^h of $H^1(\Omega)$). As this is rather a special case it is analysed in the appendix; again there is supporting numerical evidence in Section 9.

We turn now to a brief mention of those extensions of the model problem which are not particular to approximations on triangles. The case of general self-adjoint problems with mixed (homogeneous) boundary data follows [23] exactly and is not reproduced here. The effects of numerical quadrature and domains Ω which are such that the solution mesh must be placed on a curved grid (for example, regions with curved boundaries) can be dealt with as in [24], however the details have not been resolved and it is not yet clear what restrictions would be necessary on the order of integration and on the regularity of elements. The representation of non-homogeneous boundary data is not considered in the literature with regard to superconvergence although it is important and will require investigation. In the case of Poisson's equation, the above extensions are demonstrated numerically in Section 9.

We conclude this section with a remark on triangle degeneracy. If we apply the method of Theorem 2 to any triangle T with least angle α and maximum and minimum sides h_{max} and h_{min}, the result (4.1) becomes (after [6])

$$|D_P(u-u_I)| \leq C \frac{h_{max}^2}{h_{min}(\sin\alpha)^{\frac{1}{2}}} |u|_{3,T} \,. \tag{8.1}$$

Under the restriction $\sin\alpha \geq C_g > 0$ (an absolute constant over all triangles), h_{min}/h_{max} is also bounded away from zero and (8.1) reduces to (4.1); similarly the results of Theorem 1 and hence Theorem 3 still hold.

9. NUMERICAL EXAMPLES

In the examples that follow we investigate finite element approximations to Poisson's equation with non-homogeneous Dirichlet boundary conditions. In each case, we consider a region Ω, a series of triangulations with decreasing parameter h and a given function u on Ω. For each triangulation we define S_E^h to be the piecewise linear space which interpolates values of u at the boundary nodes (i.e. along $\partial\Omega$). We take $u_I \epsilon S_E^h$ to be the interpolant of u and $u_h \epsilon S_E^h$ to solve

$$a(u_h, v) = (f, v)_h \qquad \forall v \epsilon S_0^h, \tag{9.1}$$

where $a(\cdot,\cdot)$ and S_0^h are as defined in earlier sections, $(\cdot,\cdot)_h$ is a numerical approximation to the L_2 inner product $(\cdot,\cdot)$ and f is the (known) function equal to $-\nabla^2 u$.

(1) We take Ω to be the unit square $(0,1) \times (0,1)$, triangulated as in Fig. 1 (Section 3).

Let $s = 1.3\ (x-1) + y$. We will consider exact solutions u given by

$$u = \text{sign}(s)\ s^\alpha \quad \text{for} \quad \alpha = 2,3,4$$

(so that $u \epsilon H^\alpha(\Omega)$, $u \notin H^{\alpha+1}(\Omega)$). For each value of α we solve (9.1) for the range of meshes given by $h = \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots, \frac{1}{15}$ and display in Table 1 the observed convergence rates for various measures of the error in gradient sampling. (Unless otherwise stated, the numerical quadrature scheme $(\cdot,\cdot)_h$ has 7 points per triangle and is of order 5, in all these examples).

	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
$(\sum_P \|D_P(u-u_h)\|^2/n)^{\frac{1}{2}}$	$1.15h^{1\cdot 5}$	$2.2h^2$	$3.2h^2$
$\max_P \|D_P(u-u_h)\|$	$2.3h$	$2.7h^{1\cdot 8}$	$11.4h^2$
$(\sum_P \|D_P(u-u_I)\|^2/n)^{\frac{1}{2}}$	$0.69h^{1\cdot 5}$	$1.3h^2$	$2.55h^2$
$\max_P \|D_P(u-u_I)\|$	$1.71h$	$2.15h^2$	$10.4h^2$
$\|u-u_h\|_{1,\Omega}$	$2.35h$	$3.60h$	$5.60h$
$\|u_I-u_h\|_{1,\Omega}$	$0.87h^{1\cdot 5}$	$1.52h^2$	$1.65h^2$

Table 1: Gradient errors when true solution is
$u = \text{sign}(s)\, s^{\alpha}$, $s = 1.3(x-1) + y$.
($n(= O(h^{-2})$ is the number of mid-points $P \in \cup S(T)$, $T \subset \Omega$.)

Our first observation is that $u \in H^3$ is indeed a necessary condition for superconvergence; $u \in H^4$ is necessary to guarantee the pointwise result. However, even when $u \notin H^3$ the midpoint sampling procedure is considerably better than global sampling. (At the pessimistic end - $\alpha = 2$, $h = 1/3$ - we have $\max_p |D(u-u_h)| \approx 0.47$ and $|u-u_h|_{1,\Omega} \approx 0.71$; for reasonable values of h the improvement is more marked). Two general points are that the error when sampling the gradient of the interpolant u_I tends to be less than that of u_h and that when both the L_2 and L_∞ errors are $O(h^2)$ the mean-square error is usually smaller than the pointwise error by a factor of between two and five. All the above features have been confirmed in a variety of other examples.

We note from this example and others that, as mentioned in Section 6, $O(h^2)$ is the greatest rate of convergence which we can expect for $|u_I-u_h|_1$ for general u (so that in this sense (3.7) is sharp). It is also of interest to compare these observations of $|u_I-u_h|_1$ with optimal predictions based on Theorem 4, which when followed through in the manner of Theorem 1 give

$$|u_I-u_h|_1 \leq 2^{\frac{1}{2}}\cdot C_7\cdot h^2\cdot |u_{xy}|_{1,\Omega} \qquad .$$

For $\alpha = 3,4$ we have $|u_{xy}|_1 = 12.79$ and 25.41 respectively,

predicting $\qquad |u_I-u_h|_1 \leq 2.50h^2 \quad$ for $\alpha = 3$

and $\qquad |u_I-u_h|_1 \leq 4.96h^2 \quad$ for $\alpha = 4$.

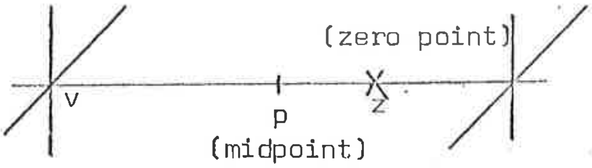
The observed improvements on these figures are due to the loss of optimality in the derivation of (3.12) noted in Section 5 and to the smoothness of u (especially for $\alpha = 4$, $u\in H^4$) when compared with that function whose restriction to each triangle pair A_j maps onto the quadrilateral Q to give the function ϕ_* from Theorem 4.

(2) We have hitherto considered the success of the midpoint gradient sampling scheme only in terms which are open to analysis: the optimal sampling location is specified and the error there measured. A quite different situation arises if we specify an "ideal" error (i.e. zero) and then measure the locations of sampling points which yield this error. To be specific, for each element edge we locate that point (assumed unique) where the component of $\underline{\nabla}(u-u_h)$ parallel to that edge is zero ("the zero point") and denote the ratio of the point's distance from the midpoint to the half-length of the edge by d (see Fig. 11). We divide the range

$0 \le d \le 1$ into smaller intervals and tabulate against d the number N of zero points in each interval for the set of element edges. We expect the distribution N/d to be clustered around $d = 0$ (zero points should be close to the midpoints) with a weaker grouping in non-superconvergent cases.

In this example we take Ω, u, u_h to be as before, with $\alpha = 2,3,4$ and $h = \frac{1}{5}, \frac{1}{7}$. We display the distributions N/d in Table 2. For example, when $\alpha = 2$, $h = \frac{1}{5}$ we find 20 zero points in $0 \le d \le 0.05, \ldots,$ none in $0.75 \le d \le 1$ and three element edges where there is no zero point. (Element edges connecting two points on the boundary $\partial\Omega$ have errors which are pure interpolation; zero points from these edges are not included in the distribution tables).

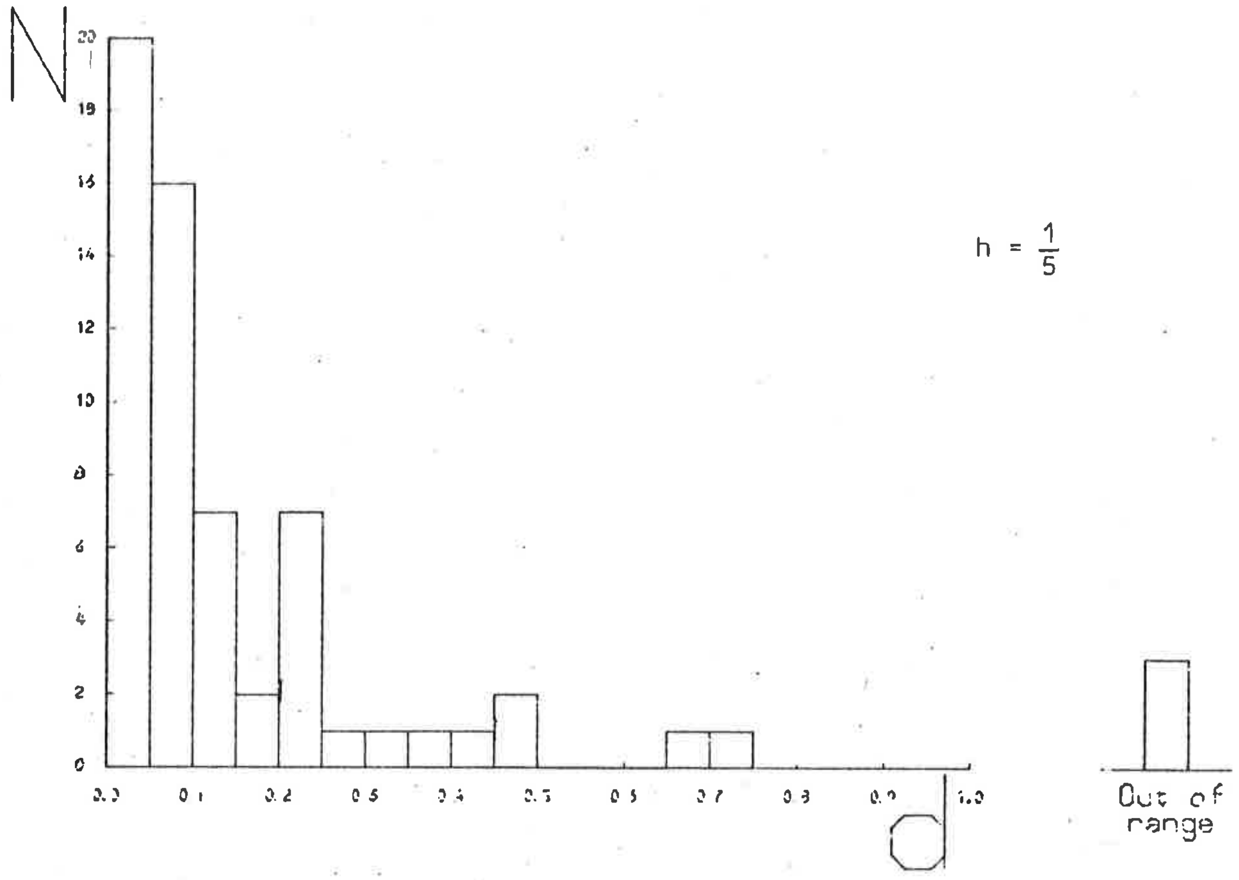
We get a picture which does not completely match expectations. Zero points appear to be clustered around midpoints even when there is no superconvergence; this may be because this measure of error is not greatly influenced by large errors in a small region of Ω. The grouping deteriorates as α increases, indeed it appears from this and other examples to be very sensitive to properties of u and Ω other than those directly connected with superconvergence. So although this example does lend further weight to the policy of sampling components of the gradient at midpoints of triangle sides, it goes a fair way towards deflating the philosophy of "superconvergence if and only if...".



(Figure 11)

$$d = \frac{|pz|}{|pv|}$$

Table 2 : Distribution of zero points, u = sign(s) $\cdot s^{\alpha}, \alpha = \cdot 2$



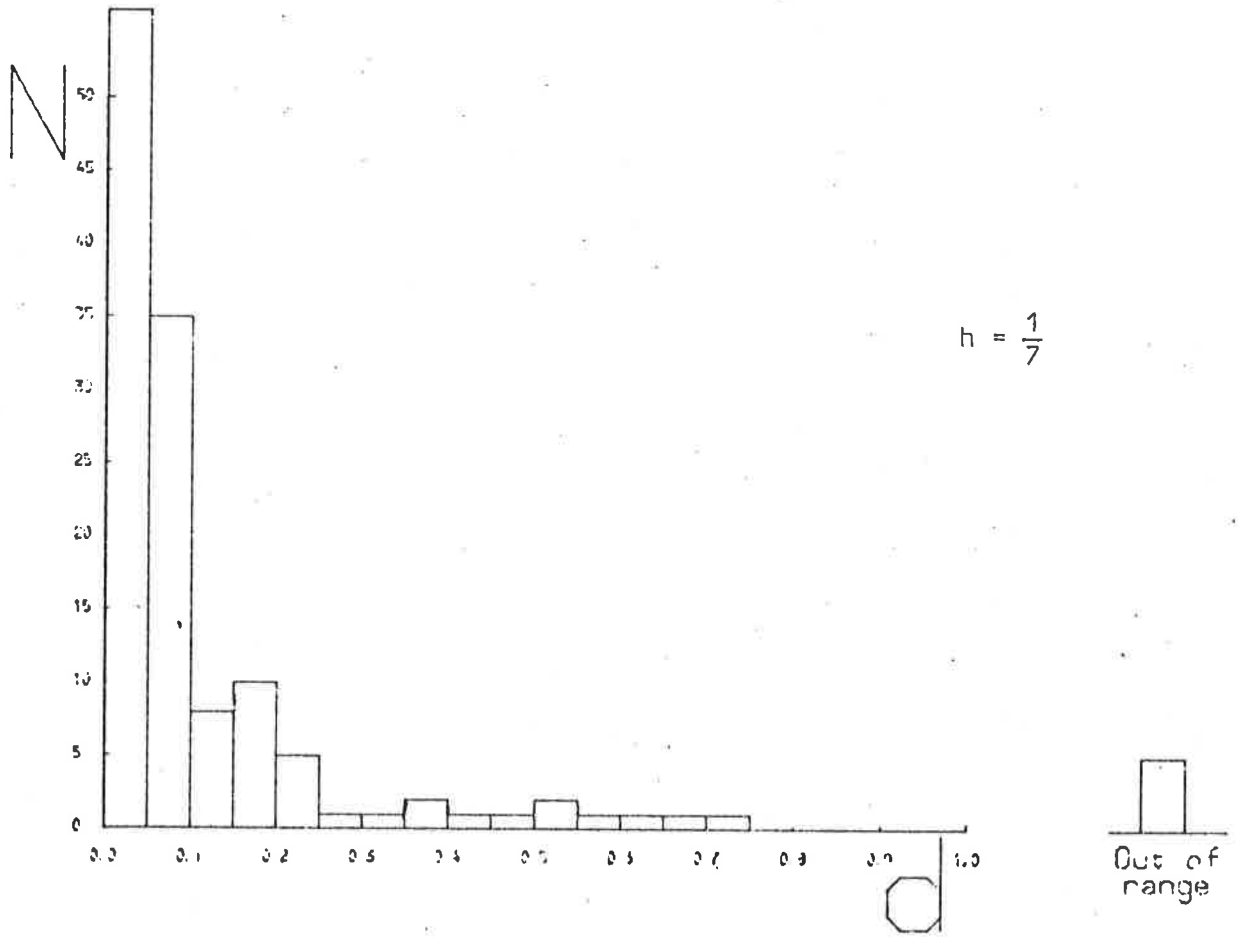
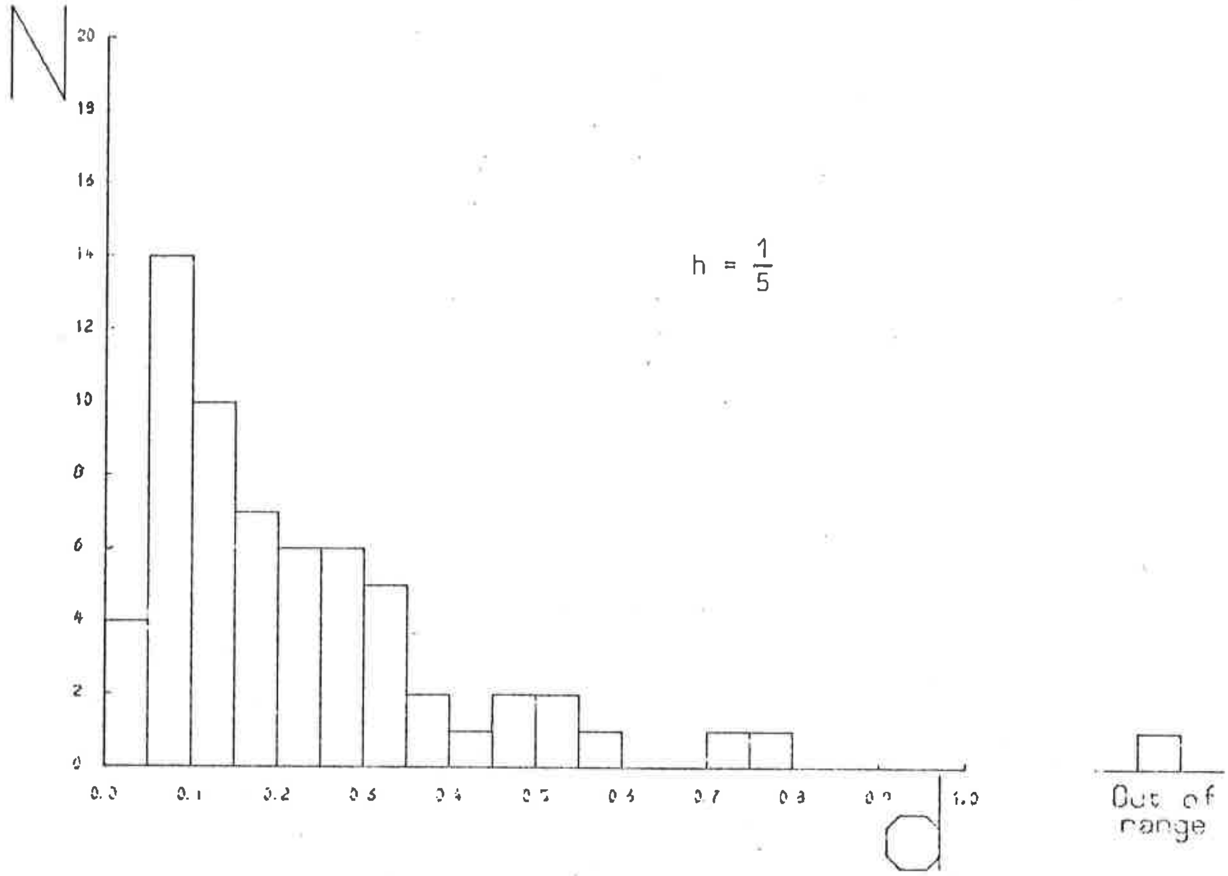


Table 2 (contd.): $\alpha = 3$



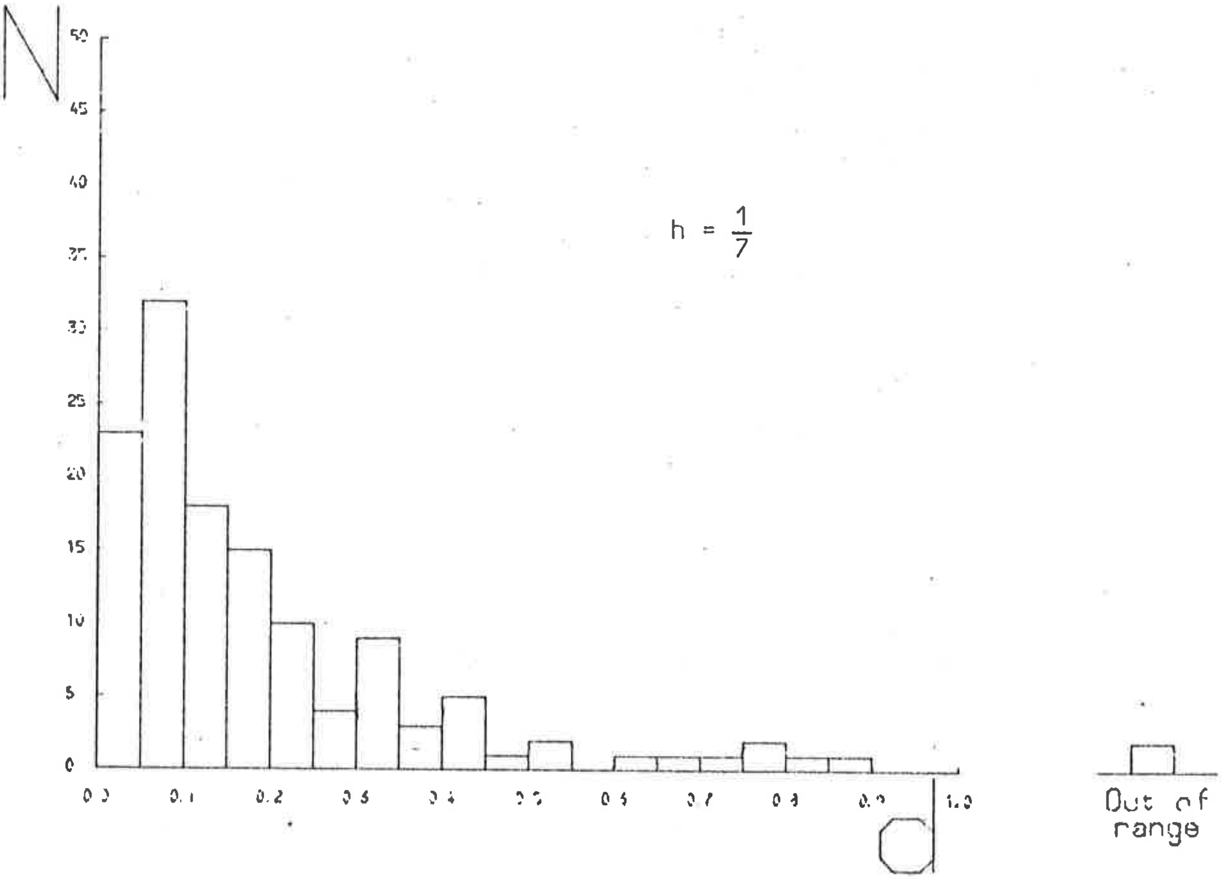
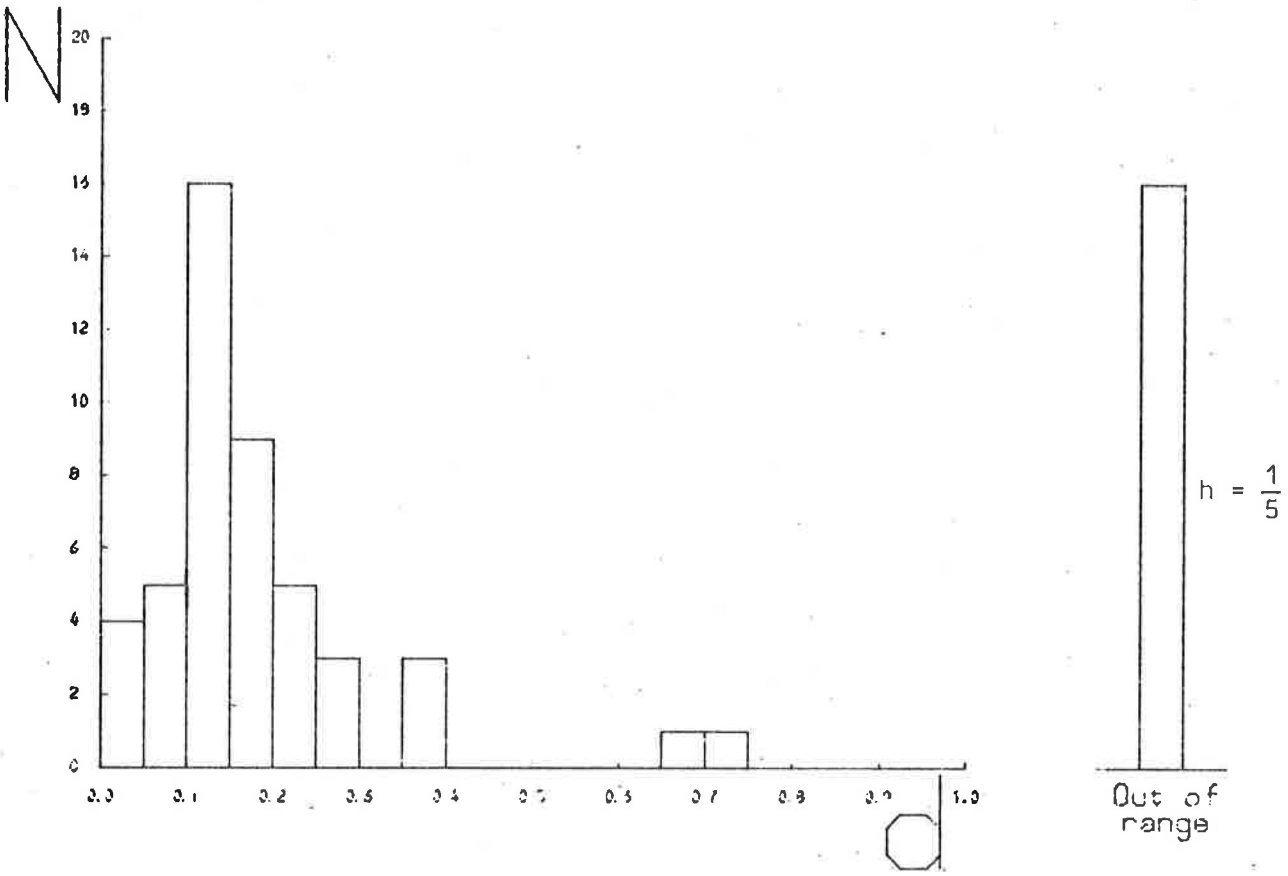
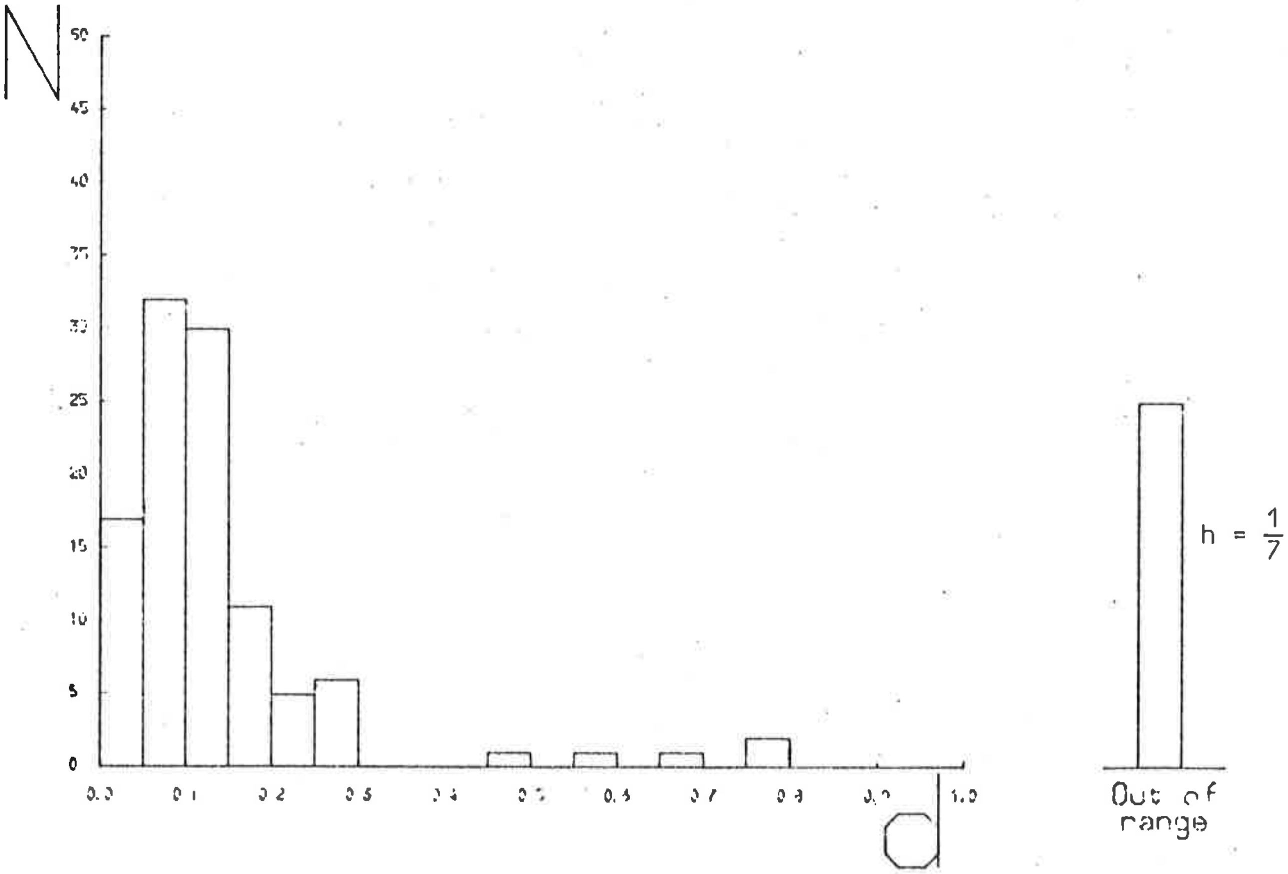


Table 2 (contd.): α = 4





(3) We given further evidence in favour of midpoint sampling even in the absence of superconvergence.

(a) We take Ω to be the unit square, triangulated as in Fig. 9 (see remark 1 to Theorem 7, Section 8), we let $u = (x-0.75)^2 + (y+1)^2$ and solve (9.1) for $h = \frac{1}{3}, \dots, \frac{1}{9}$. As predicted, there is no second order gradient convergence: we observe

$$(\sum_P |D_P(u-u_h)^2/n|)^{\frac{1}{2}} \simeq 0.26h$$

(Note that $D_P(u-u_I) = 0 \; \forall P$ because u is quadratic; i.e. there is no interpolation contribution to the error).

(b) We take Ω to be the truncated unit square triangulated as shown in Fig. 10(d)(see remark 2 to Theorem 7), and solve (9.1) for $h = \frac{1}{2}, \frac{1}{4}, \dots, \frac{1}{12}$ with u as above. Predictably, again, we observe

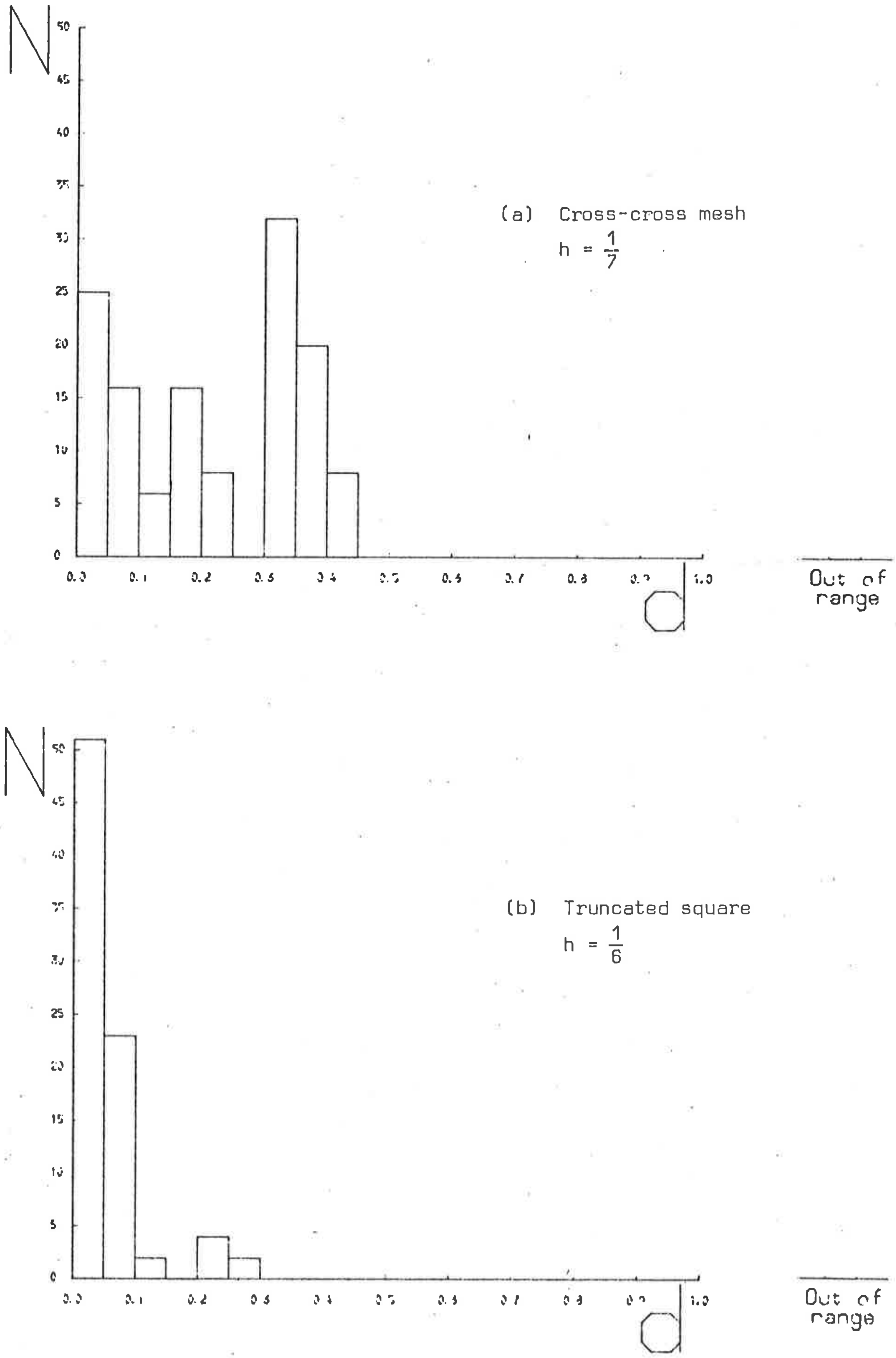
$$(\sum_P |D_P(u-u_h)^2/n|)^{\frac{1}{2}} \simeq 0.20h^{3/2}.$$

Zero point distributions from these two examples are shown in Table 3. We see that even though neither mesh is superconvergent there is a strong grouping of zero points around the mid-points. (In case (b) where the required conditions on the mesh are only violated locally - on the boundary - the grouping is particularly good).

(4) To confirm that superconvergence is not lost for Laplace's equation when mesh conditions are relaxed (see remark 3 to Theorem 7), we take Ω to be the truncated square again, triangulated as above, and solve (9.1) for $u = \log((x-0.75)^2 + (y+1)^2)^{\frac{1}{2}}$, $h = \frac{1}{2}, \dots, \frac{1}{10}$. Superconvergence is observed, with $(\sum_P |D_P(u-u_h)^2/n)^{\frac{1}{2}} \simeq 0.043h^2$.

(5) We investigate the effect of relaxing the order of the numerical quadrature $(\cdot,\cdot)_h$ in (9.1). We take Ω to be the unit square, triangulated as in Fig. 1 and let $u = ((x-0.75)^2 + (y+0.1)^2)^{3/2}$, solving (9.1) for $h = \frac{1}{3}, \dots, \frac{1}{11}$. With the usual fifth-order scheme we observe

Table 3 : Zero points for criss-cross mesh and truncated square (both non-superconvergent)



(a) Cross-cross mesh $h = \frac{1}{7}$

(b) Truncated square $h = \frac{1}{6}$

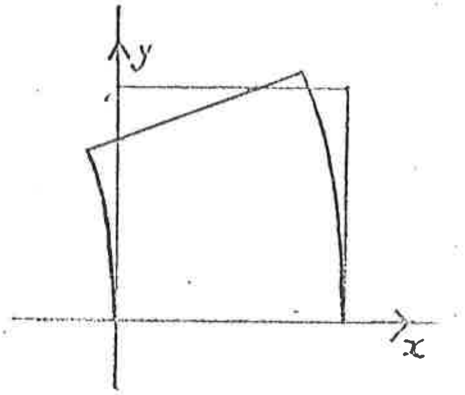
$(\sum_P |D_P(u-u_h)|^2/n)^{\frac{1}{2}} \approx 0.30h^2$; with a first-order scheme we observe $(\sum_P |D_P(u-u_h)|^2/n)^{\frac{1}{2}} \approx 0.35h^2$ (a slight loss of accuracy but no loss of order).

(6) Finally, we take a look at the other common "variational crime", that of approximate representation of the boundary $\partial\Omega$. (Note that approximate representation of boundary conditions has been implicitly covered in all the above examples). We map the unit square into the sector shown in Fig. 12 by the transformation:

$$x \rightarrow x_1 = (x+2)/(1+y^2/4)^{\frac{1}{2}} - 2$$

$$y \rightarrow y_1 = y\ (1+x_1/2)$$

The grid points are transformed in this way and then joined by straight lines to give a triangulation topologically equivalent to that in Fig. 1. Note that the two curved boundary segments are not represented exactly.



(Figure 12)

Distortion of Ω

We again take $u = (x_1-0.75)^2 + (y_1+1)^2$ and solve (9.1) for h (= mesh spacing on x_1-axis) $= \frac{1}{3},\ldots,\frac{1}{11}$. Once again we obtain superconvergence with

$$(\sum_P |D_P(u-u_h)|^2/n)^{\frac{1}{2}} \approx 0.039h^2.$$

We consider the zero point distribution for $h = \frac{1}{7}$; though visually dull it is mathematically encouraging since $d \leq 0.05$ for all 131 zero points.

Conclusion

Superconvergence is resistant to a surprising degree of variational abuse, however it is very touchy about the mesh topology and the smoothness of the function under approximation. On the other hand it is clear from the experiments with zero point distributions that even when superconvergence fails there is no a priori reason for sampling gradients on linear triangles by any other method than that of single components picked up at triangle mid-points.

Acknowledgements

I would like to thank Professor K.W. Morton for invaluable discussions, Paul Samuels for the use of his finite element code and the Science & Engineering Research Council for its financial support.

APPENDIX : RELAXING TRIANGULATION CONDITIONS FOR LAPLACE'S EQUATION

Theorem 8 . In the case of Laplace's equation on a region Ω with exactly represented boundary data, all meshes based on square grids are superconvergent.

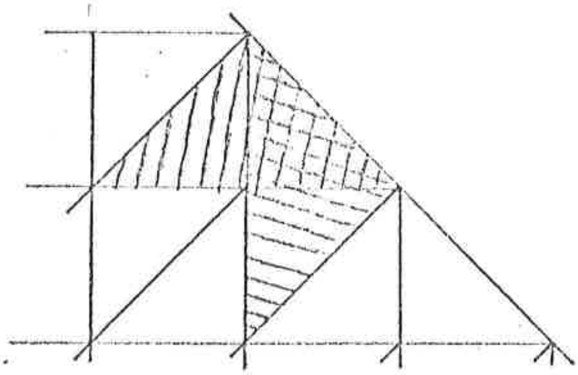
Proof In the interior of Ω the finite element equations reduce to the standard homogeneous five-point finite difference approximation and so u_h is not affected by the signs of the slopes of diagonals within the triangulation. We may therefore take all these slopes to be $+1$ and are left with consideration of boundary contributions to the error for the mesh shown in Figure 10(d).

We consider the x- and y- derivative contributions to $a(u-u_I, v)$ from the boundary region together (where $v\epsilon S_0^h$). For we have

$$|a(u-u_I,v)| \leq Ch^2 |u|_3 \ |v|_1 + \text{boundary term;}$$

$$\text{the boundary term} = \sum_j \int_{E_j} \frac{\partial}{\partial x}(u-u_I)\frac{\partial v}{\partial x} + \int_{F_j} \frac{\partial}{\partial y}(u-u_I)\frac{\partial v}{\partial y} \quad ,$$

where the E_j and F_j are, as shown in Fig. 13, partly overlapping pairs of triangles near the boundary.



(Figure 13)

▤ - The corresponding E_j

▨ - One of the F_j

Note that $\left.\frac{\partial v}{\partial x}\right|_{E_j} = \left.\frac{\partial v}{\partial y}\right|_{F_j}$ when $v\epsilon S_0^h$

As usual, each integral vanishes in the cases $u\epsilon P_2$ and $u = xy$. In addition, since $\partial v/\partial x|_{E_j} = \partial v/\partial y|_{F_j}$, the sum of the E_j- and corresponding F_j - integrals vanishes when $u = x^2 - y^2$. So this sum vanishes for all harmonic quadratics (but not for $u = x^2 + y^2$).

Some care must be taken in using the Bramble-Hilbert lemma here. We write the above sum of integrals, over a single pair E_j and F_j as the functional Au. By means of the Sobolev representation (lemma 3) we can write any $u \in H^3(\Omega)$ in the form

$$u = Q_1 + Q_2 + R,$$

where Q_1 is a harmonic quadratic,

Q_2 is a multiple of $\xi^2+\eta^2$,

R is dependent on the third derivatives of u in $E_j \cup F_j$

and the (ξ,η) origin and open ball B used in the expansion are in $E_j \cap F_j$.

We write Π for the projection

$$\Pi u = Q_1 + HR,$$

where H is the projection which takes any function on $E_j \cup F_j$ to one with the same values on the boundary of that region but which is harmonic in the interior. By trace and regularity theorems, H is bounded and therefore

$$\| \Pi u \|_{3,E_j \cup F_j} \leq C \| u \|_{3,E_j \cup F_j},$$

so that
$$|A\Pi u| \leq C \| u \|_{3,E_j \cup F_j} \, |v|_{1,E_j \cup F_j} \quad \forall u \in H^3. \tag{A1}$$

Also, for all $u \in P_3$,

$$\begin{aligned} A\Pi u &= A\Pi \, (Q_1+Q_2) \qquad \text{(since } R=0) \\ &= AQ_1 \\ &= 0. \end{aligned} \tag{A2}$$

So by (A1), (A2) and the Bramble-Hilbert lemma applied to the functional $A\Pi$ we have

$$|Au| = |A\Pi u| \leq Ch^2 \, |u|_{3,E_j \cup F_j} \cdot |v|_{1,E_j \cup F_j} \qquad \text{for all } u \in H^3$$

for which $u = \Pi u$, i.e. for all harmonic u in H^3. Summing over j, we get back to $|a(u-u_I,v)| \leq Ch^2|u|_3\ |v|_1$ for all harmonic $u\epsilon H^3$, as required for superconvergence.

REFERENCES

[1] Aubin, J.P. "Approximation of elliptic boundary-value problems", Wiley-Interscience, 1972.

[2] Barlow, J. "Optimal stress locations in finite-element models", Int. J. Num. Meth. Eng., 10, 243-51, 1976.

[3] Barnhill, R.E., Gregory, J.A. & Whiteman, J.R. "The expansion and application of Sard kernel theorems to compute finite element error bounds", in "The mathematical foundations of the finite element method with applications to partial differential equations", (ed. Aziz), Academic Press, 749-55, 1972.

[4] Bramble, J.H. & Hilbert, S.R. "Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation", SIAM J. Numer. Anal., 7, 113-124, 1970.

[5] Bramble, J.H. & Schatz, A.H. "Higher order local accuracy by averaging in the finite element method", Math. Comp., 31, 94-111, 1977.

[6] Bramble, J.H. & Zlámal, M. "Triangular elements in the finite element method", Math. Comp., 24, 809-20, 1970.

[7] Ciarlet, P.G. "The finite element method for elliptic problems", North-Holland, 1978.

[8] Dupont, T. & Scott, R. "Polynomial approximation of functions in Sobolev spaces", Math. Comp., 34, 441-63, 1980.

[9] Fix, G.J., Gunzburger, M.D. & Nicolaides, R.A. "On mixed finite element methods for first order elliptic systems", Numer. Math., 37, 29-48, 1981.

[10] Golumb, M. & Weinberger, H.F. "Optimal approximation and error bounds", in "On numerical approximation", (ed. Langer), University of Wisconsin Press, 117-90, 1959.

[11] Le Saint, P. & Zlámal, M. "Superconvergence of the gradient of finite element solutions", Rev. Française Automat. Informat. Recherche Opérationelle Sér. Range Anal. Numér., 13, 139-66, 1979.

[12] Meinguet, J. "Sharp 'a priori' error bounds for polynomial approximation in Sobolev spaces", to appear in "Multivariate Approximation Theory", Birkhäuser Verlag, 1982.

[13] Moan, T. "Experiences with orthogonal polynomials and 'best' numerical integration formulas on a triangle", Z.A.M.M., 54, 501-8, 1974.

[14] Morrey, C. "Multiple integrals in the calculus of variations", Springer-Verlag, 1966.

[15] Nitsche, J.A. "L_∞-Error analysis for finite elements" in "The mathematics of finite elements and applications, III" (ed. Whiteman), 174-86, 1978.

[16] Nitsche, J.A., & Schatz, A.H. "Interior estimates for Ritz-Galerkin methods", Math. Comp., 28, 937-58, 1974.

[17] Oden, J.T. & Reddy, J.N. "An introduction to the mathematical theory of finite elements", Wiley-Interscience, 1976.

[18] Sobolev, S.L. "Application of functional analysis in mathematical physics", Leningrad, 1950 (translation: American Mathematical Society, 1963).

[19] Strang, G. & Fix, G. "An analysis of the finite element method". Prentice-Hall, 1973.

[20] Thomée, V. "High order approximations to derivatives in the finite element method", Math. Comp., 31, 652-60, 1977.

[21] Veryard, D.A. "Problems associated with the convergence of isoparametric and mixoparametric finite elements", M.Sc. thesis, University of Wales, 1971.

[22] Zienkiewicz, O.C. "The finite element method" (3rd edition), Mc.Graw-Hill, 1977.

[23] Zlámal, M. "Some superconvergence results in the finite element method", in "Mathematical aspects of finite element methods", (eds. Galligani, Magenes), Springer-Verlag, 351-62, 1977.

[24] Zlámal, M. "Superconvergence and reduced integration in the finite element method", Math. Comp., 32, 663-85, 1978.