INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS, VOL. 3, 93–98 (1983)

SHORT COMMUNICATION

TWO COMMENTS ON: CONSISTENT vs REDUCED INTEGRATION PENALTY METHODS FOR INCOMPRESSIBLE MEDIA USING SEVERAL OLD AND NEW ELEMENTS[1]

MICHEL FORTIN

Départment de Mathématiques, Université Laval, Québec, G1K 7P4, Canada

1. INTRODUCTION

This note would like to develop some ideas of Engelman *et al.*[1] about penalty methods. Two remarks will be made: the first deals with the role of the mass matrix M in consistent penalty methods and the second with the effect of numerical quadrature on the Babǔska–Brezzi (B–B) compatibility condition.

2. THE ROLE OF THE MASS MATRIX

To remain in the same framework as the authors,[1] we consider a discrete Stokes problem,

$$\left.\begin{aligned} a(u_h, v_h) - \int_\Omega p_h \nabla \,.\, v_h \,\mathrm{d}x = \int_\Omega f \,.\, v_h \,\mathrm{d}x \qquad & \forall\, v_h \in V_h, \\ \int_\Omega \nabla \,.\, u_h q_h \,\mathrm{d}x = 0 \qquad & \forall\, q_h \in Q_h, \end{aligned}\right\} \tag{1}$$

where V_h is an N-dimensional space of approximate velocities, Q_h an L-dimensional space of approximate pressures and $a(u, v) = \int_\Omega \varepsilon_{ij}(u)\varepsilon_{ij}(v)\,\mathrm{d}x$. Let U and P be the vectors of 'nodal values' of $u_h \in V_h$ and $p_h \in Q_h$. We now define $A_{N\times N}, C_{L\times N}, M_{L\times L}, F_{N\times 1}$ by

$$\left.\begin{aligned} (AU, V)_{\mathbb{R}^N} &= a(u_h, v_h) && \forall\, u_h, v_h \in V_h, \\ (CU, Q)_{\mathbb{R}^L} &= \int_\Omega \nabla \,.\, u_h q_h \,\mathrm{d}x && \forall\, q_h \in Q_h, \quad \forall\, u_h \in V_h, \\ (MP, Q)_{\mathbb{R}^L} &= \int_\Omega p_h q_h \,\mathrm{d}x && \forall\, p_n, q_h \in Q_h, \\ (F, V)_{\mathbb{R}^N} &= \int_\Omega f.\, v_h \,\mathrm{d}x && \forall\, v_h \in V_h. \end{aligned}\right\} \tag{2}$$

Problem (1) is now equivalent to,

$$\left.\begin{aligned} AU + C^{\mathrm{T}}P &= F, \\ CU &= 0, \end{aligned}\right\} \tag{3}$$

0271–2091/83/010093–06$01.00

Received 30 November 1981
Revised 13 May 1982

which is nothing but a linearly constrained quadratic problem in $\mathbb{R}^N$, precisely

$$\inf_{CU=0} \tfrac{1}{2}(AU, U)_{\mathbb{R}^N} - (F, U)_{\mathbb{R}^N}. \tag{4}$$

In (3) the Lagrange multiplier P has been introduced to deal with the linear constraint $CU=0$. In mathematical programming, a standard procedure to solve constrained problems is to use a penalty method. A *penalty method has nothing to do, in itself, with finite elements.* It can, however, be applied to solve problems arising from finite element approximations such as (4). Let then S be *any* positive definite $L \times L$ matrix. Then $(S^{-1}P, P)$ is the square of a norm on $\mathbb{R}^L$ and it is possible to approximate problem (4) by the following problem:

$$\inf_U \tfrac{1}{2}(AU, U)_{\mathbb{R}^N} - (F, U)_{\mathbb{R}^N} + \frac{\lambda}{2}(S^{-1}CU, CU)_{\mathbb{R}^L}. \tag{5}$$

The solution U_λ is found by solving the linear system,

$$AU_\lambda + \lambda C^T S^{-1} CU_\lambda = F, \tag{6}$$

that can also be written as,

$$\left.\begin{aligned} AU_\lambda &= C^T P = F, \\ CU_\lambda &= \frac{1}{\lambda} SP. \end{aligned}\right\} \tag{7}$$

A *possible* choice of a suitable matrix S is $S=M$. However, any choice (including $S=I$ and computing M by some quadrature rules) will lead to a penalty method in which U_λ converges to the solution of (3) when $\lambda \to \infty$ and $\lambda S^{-1}CU$ converges to P. It is only necessary in practice to use a not too ill-conditioned matrix in order to avoid a bad scaling in the penalty factor. This makes it possible to use a penalty method even with approximations where the functions of Q_h are continuous across element boundaries and where M^{-1} is a *full* matrix and cannot be used in (6).

3. THE EFFECT OF NUMERICAL QUADRATURE ON THE BABŬSKA–BREZZI CONDITION

We have just said that a precise evaluation of the mass matrix M is not a necessary part of a consistent penalty method. The same is not true of an approximate quadrature rule applied to the computation of matrix C. Any error on the evaluation of $\int_\Omega p_h \nabla \cdot v_h \, dx$ changes the matrix C in (3) to another matrix $\tilde{C}$ and the solution itself will be changed as the discrete divergence-free condition is no longer the same.

By how much it will be changed is a matter for stability analysis. Engelman *et al.*[1] use a result of Brezzi[2] to make this analysis. Although their discussion is mostly right, I think a few points can be clarified and a small mistake corrected. Let us first recall some facts about the equivalence of reduced integration penalty methods and mixed methods (cf. Malkus and Hughes[3] and Oden[4]).

A reduced integration penalty (RIP) method is one in which a penalty term

$$\lambda \int_\Omega |\nabla \cdot u|^2 \, dx$$

is introduced in the 'continuous' (infinite dimensional) problem. When this term is computed in a discrete problem using finite elements, an inexact quadrature is used. For instance using a Q_2, 9-node element for velocities and using a 4-point (2×2) Gaussian rule to evaluate

$$\int_K |\nabla \cdot v_h|^2 \,\mathrm{d}x$$

is a RIP method. It was shown by Malkus and Hughes[3] that using a RIP method is equivalent to some mixed method in which the nodal values of pressure are defined at the quadrature point and in which the matrix C corresponding to

$$\int_\Omega p_h \nabla \cdot v_h \,\mathrm{d}x$$

would be evaluated by the same quadrature rule.

Thus a RIP method will be equivalent to a consistent penalty method if and only if the quadrature rule is exact for the computation of

$$\int_\Omega p_h \nabla \cdot v_h \,\mathrm{d}x.$$

This clearly comes out from the numerical results of Engelman *et al.*[1] Consider for instance the Q_2-Q_1 consistent method and the Q_2, 4-point RIP method. They yield the *same* results as long as the 4-point rule is exact to compute C, that is for quadrilaterals with straight sides and natural centroids. Whenever the centroid is displaced or the sides are curved the 4-point rule is no longer exact and the RIP method becomes equivalent to a modified mixed method in which $\tilde{C}$ is an approximation of C.

When

$$\int_\Omega p_h \nabla \cdot v_h \,\mathrm{d}x$$

is computed by a quadrature rule, it is necessary that the approximation of the mass matrix by the same rule be positive definite or equivalently that the quadrature points must *contain* a unisolvent set for the pressures q_h of Q_h. If this were not the case the kernel of C^{T} would become very large, pressure would be undetermined at element level and very exotic pressure modes would appear. (Furthermore if this approximate mass matrix is to be used as a penalty matrix it should be positive definite.)

For instance a Q_2-P_1 approximation with a 4-point rule is correct but collapses with a 1-point rule. In this last case one should replace Q_h by a P_0 (piecewise constant) approximation.

Thus we have a first check for the B–B condition when numerical quadrature is involved: the quadrature points must contain an unisolvent set for the pressure on every element. In Engelman *et al.*[1] this was incorrectly applied to the velocities instead of the pressure to conclude that a Q_2-P_1 approximation with $\int p_h \nabla \cdot v_h \,\mathrm{d}x$ evaluated by a 4-point rule does not satisfy the B–B condition.

This condition is, however, not sufficient. To go further we shall suppose that the mixed method with *exact quadrature* does satisfy the B–B condition which is

$$\sup_{v_h \neq 0} \frac{|b(v_h, p_h)|}{\|v_h\|_1} \geq k\, |p_h|_0. \tag{8}$$

Let then $b^{\cdot}(v_h, p_h)$ be an approximation of $b(v_h, p_h)$ such that

$$|b^{\cdot}(v_h, p_h) - b(v_h, p_h)| \leq C(h)\, \|v_h\|_1\, \|p\|_0, \tag{9}$$

where $C(h) \rightarrow 0$ as $h \rightarrow 0$. Then one has,

$$\sup_{v_h \neq 0} \frac{|b^{\cdot}(v_h, p_h)|}{\|v\|_1} \geq (k - C(h))\, |p_h|_0 \geq k'\, \|p\|_0, \tag{10}$$

for h small enough. Indeed one has,

$$k\, |p_h|_0 \leq \sup_{v_h \neq 0} \frac{|b(v_h, p_h)|}{\|v_h\|_1} \leq \sup_{v_h \neq 0} \frac{|b^{\cdot}(v_h, p_h)|}{\|v_h\|_1} + \sup_{v_h \neq 0} \frac{|b(v_h, p_h) - b^{\cdot}(v_h, p_h)|}{\|v_h\|_1}, \tag{11}$$

which clearly implies (10) if (9) is satisfied.

As can be seen from the previous proof, it would be sufficient to prove that one has,

$$C(h) < k, \tag{12}$$

for h small enough. This is indeed a much more difficult task as the precise values of the constants are very difficult to obtain. The real problem is then to check (9) and this requires a very careful analysis. We shall consider one case for which we know (cf. Fortin[5]) that (8) holds, that is the $Q_2 - P_1$ approximation and we shall evaluate the effect of using a (2×2) Gaussian quadrature rule with this approximation.

In this case there is no quadrature error if the elements are straight-sided quadrilaterals with 'natural centroids'.

We shall outline the proof of the following result:

For h small enough, the $Q_2 - P_1$ approximation with a (2×2) Gaussian rule satisfies the B–B condition if the elements are 'not too curved'. Precisely, the deviation of midpoints from the straight line and that of the centroid from its natural place must be $O(h^2)$ when h converges to zero.

It must be remarked that the condition stated is a standard one (cf. Ciarlet[6] and Ciarlet and Raviart[7]) in the theory of isoparametric quadrilateral approximations. The reader should also note that the proof relies heavily on the special form of the inverse of a 2×2 matrix and cannot be extended to the 3-D case. This is also consistent with the results of Engelman *et al.*[1]

Let us now sketch the proof. We consider the change of variables enabling us to pass from the reference element $\hat{K}$ to the element K. We have

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = F(\hat{x}) \begin{cases} q_1(\hat{x}_1, \hat{x}_2) + \sum_{k=5}^{9} c_k \hat{p}_k(\hat{x}_1, \hat{x}_2) \\ q_2(\hat{x}_1, \hat{x}_2) + \sum_{k=5}^{9} d_k \hat{p}_k(\hat{x}_1, \hat{x}_2) \end{cases} \tag{13}$$

or

$$x = Q(\hat{x}) + E(\hat{x})$$

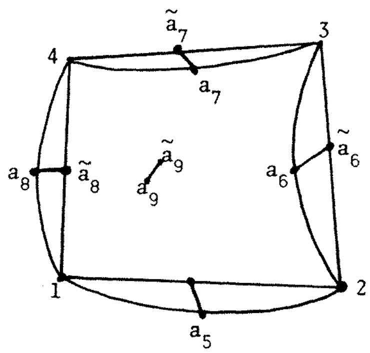

Figure 1

where c_k and d_k are the components of vector $\tilde{a}_k - a_k$, in Figure 1. Points $\tilde{a}_k$ refer to the midpoints of the straight-sided part of the element defined by the functions q_1 and q_2. By assumption we have $c_k = O(h^2)$ and $d_k = O(h^2)$. We have denoted $\hat{p}$ the basis functions associated with the nodes 5 to 9 in the usual definition of a Q_2 element.

The operator ∇_x on K is related to $\hat{\nabla}_{\hat{x}}$ on $\hat{K}$ by

$$\nabla = (DF)^{-1}\hat{\nabla}. \tag{14}$$

One has

$$\begin{aligned}(DF)^{-1} &= \frac{1}{J(\hat{x})}\begin{pmatrix} \dfrac{\partial q_2}{\partial \hat{y}} + \sum_k d_k \dfrac{\partial \hat{p}_k}{\partial \hat{y}}, & -\dfrac{\partial q_2}{\partial \hat{x}} - \sum_k d_k \dfrac{\partial \hat{p}}{\partial \hat{x}} \\ -\dfrac{\partial q_1}{\partial \hat{y}} - \sum_k c_k \dfrac{\partial \hat{p}_k}{\partial \hat{y}}, & \dfrac{\partial q_1}{\partial \hat{x}} + \sum_k c_k \dfrac{\partial \hat{p}_k}{\partial \hat{x}} \end{pmatrix} \\ &= \frac{1}{J}M_1 + \frac{1}{J}M_2. \end{aligned} \tag{15}$$

We have to estimate the error made when evaluating, by a (2×2) Gaussian rule, the expression

$$\begin{aligned}\int_K p_h \nabla \,.\, v_h \,\mathrm{d}x &= \int_K \hat{p}_h (DF)^{-1}\hat{\nabla} \,.\, \hat{v}_h J(x)\,\mathrm{d}\hat{x} \\ &= \int_K \hat{p}_h (M_1 \hat{\nabla}) \,.\, \hat{v}_h \,\mathrm{d}\hat{x} + \int_{\hat{K}} \hat{p}_h (M_2 \hat{\nabla}) \,.\, \hat{v}_h \,\mathrm{d}\hat{x}. \end{aligned} \tag{16}$$

The determinants J have cancelled and the expression has been split into two parts. There is no quadrature error for the first part that corresponds to the straight-sided part of the element (cf. also Leone *et al.*[8]) Only the terms of the second part induce errors and this error is bounded by

$$\Big(\sum_k c_k + d_k\Big)\hat{C}\,\|\hat{v}_h\|_{1,\hat{K}}\,|\hat{p}|_{0,\hat{K}}. \tag{17}$$

To prove (17) we use the continuity of the error on the space of polynomials containing $\hat{v}_h$ and $\hat{p}_h$ and the equivalence of norms on these finite dimensional spaces. The constant $\hat{C}$

depends only on $\hat{K}$. For details on the techniques used, the reader should refer to Ciarlet.[6] Now, using standard formulae for the change of variables, we bound (17) by,

$$\sum_k (c_k + d_k)\hat{C}\left(\inf_{\hat{x}} J(\hat{x})\right)^{-1} |p_h|_{0,K} \|DF\|_{1,\infty} \|v_h\|_{1,K} \leq 0(h)\, |p_h|_{0,K} \|v_h\|_{1,K}, \tag{18}$$

using again the methods of Ciarlet[6] and the assumption on c_k and d_k. This bound corresponds to condition (9) and we thus have the B–B condition.

Another less important although interesting case is the $Q_2 - P_0$ element with a 1-point quadrature rule. The above proof unfortunately cannot be extended to this case. The author conjectures that one should be able to prove that this approximation satisfies the B–B condition if the mesh is 'not too distorted', that is if one uses straight-sided quadrilaterals that tend to parallelograms when h tends to zero.

Following the same steps as in the proof one splits the integral into two parts. The first depends on the linear part of the change of variables and the second on the $\hat{x}\hat{y}$ terms. This last part converges to zero by the geometric condition. However, the first part induces a quadrature error, although only for basis functions associated with the midpoints.

I found no way of proving that this term converges to zero. It is therefore possible that this is a case where condition (12) should be proved to make theory coincide with experimental evidence.

REFERENCES

1. M. Engelman, R. L. Sani, P. M. Gresho and M. Bercovier, 'Consistent vs reduced integration penalty methods for incompressible media using several old and new elements', *Int. J. num. Meth. Fluids*, **2,** 25–43 (1982).
2. F. Brezzi, 'On the existence, uniqueness and approximation of saddle point problems arising from Lagrange multipliers', *RAIRO*, R.2 (1974) p. 129.
3. D. S. Malkus and T. J. R. Hughes, 'Mixed finite element methods—reduced and selective integration techniques: a unification of concepts', *Comp. Meth. App. Mech. Eng.*, **15,** 63 (1978).
4. J. T. Oden, 'RIP-methods for Stokesian flows', *TICOM Report 80–81*, Texas Institute for Computational Mechanics, University of Texas at Austin.
5. M. Fortin, 'Old and New elements for incompressible flows', *Int. J. num. Meth. Fluids* **1,** 347–367 (1981).
6. P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam–New York, 1980.
7. P. G. Ciarlet and P. A. Raviart, 'Interpolation theory over curved elements with applications to finite elements methods', *Comp. Meth. App. Mech. Eng.*, **1,** 217–274 (1972).
8. J. M. Leone, P. M. Gresho, T. C. Chan and R. L. Lee, 'A note on the accuracy of Gauss–Legendre Quadrature in the finite element method', *Int. J. num. Meth. Eng.*, **14,** 769–784 (1979)