in Proceedings, Conference Reliability of Methods for Engineering Analysis
Univ. College Swansea, U.K., July, 1986

STUDIES OF FINITE ELEMENT PROCEDURES -- THE INF-SUP CONDITION, EQUIVALENT FORMS AND APPLICATIONS

Franco Brezzi
Dipartimento di Meccanica Strutturale and Istituto di Analisi Numerica del C.N.R., 27100 Pavia, Italy

Klaus-Jürgen Bathe
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

This is a tutorial paper on the mathematical analysis of finite element discretizations. We are concerned with the convergence behavior of elements when internal constraints need to be satisfied, in particular, the incompressibility constraints in elasticity or fluid flow. As a model problem, we consider the solution of incompressible, or slightly compressible elasticity. First we discuss the general conditions to be satisfied by the finite element spaces in order to obtain optimal convergence behavior. This leads to the inf-sup condition and various equivalent forms that are more easily used. We then apply the inf-sup condition to the 9-node displacement/linear pressure two-dimensional element and make some remarks on how other elements and problems would be considered.

1. Introduction

In fluid and structural mechanics we often have to minimize functionals of the form

$$J(v) = J_1(v) + \varepsilon^{-1} J_2(v) \quad . \tag{1}$$

In some applications, the form (1) is naturally obtained in the physical formulation, and ε is some positive physical quantity that is very small. In other applications, the true physical functional to be minimized is actually $J_1(v)$ and the second term is introduced only as a penalty term, to take account of a physical constraint. In other words, we have to minimize $J_1(v)$ under the constraint $J_2(v) = 0$ and we choose to minimize, instead, (1) for some $\varepsilon > 0$ very small. Here, therefore, ε is a mathematical parameter that is somehow at our choice: however the use of an ε too large will produce unsatisfactory results, because the constraint $J_2(v) = 0$ will be poorly satisfied. In what follows, we shall not distinguish between these two (philosophically different) cases. We will just focus our attention on the question: "What happens to the minimizers of (1) (before and, mainly, after discretization) when ε is very small?"

In many practical cases, a very small ε is known to be a considerable source of difficulties, and the most popular remedy is "reduced" integration for the functional $J_2(v)$. However, it is not clear when the reduced integration is necessary and, most of all, how much do we have to "reduce" the integration for saving a satisfactory accuracy. The solution of problems of type (1) using reduced integration has been in recent years connected to the use of mixed finite element methods. From the mathematical point of view, such a connection really works only in special cases and is just false in others. One analysis technique that is in some sense related to reduced integration is mixed interpolation, and it is this approach that we will consider in this paper.

Mixed finite element methods have received careful attention by mathematicians: sufficient conditions have been introduced to guarantee their stability, convergence and error estimates. Among them the inf-sup condition became popular in the framework of incompressible fluids, probably because the other conditions were trivially satisfied by any reasonable discretization while the satisfaction of the inf-sup condition was (and still is) the source of difficulties.

The objective of this paper is to discuss the inf-sup condition -- and its various forms -- as it applies in finite element analysis. As a model problem we consider the solution of linear elastic (almost) incompressible solids. The discussion does not present new information, but is directed towards providing insight into some mathematical conditions for finite elements and into techniques to analyze whether these conditions are satisfied.

In the next section we study the use of a general sequence of finite element spaces to solve (1) and discuss conditions on these spaces to obtain optimal convergence properties. These requirements lead to the inf-sup condition.

In Section 3 we then analyze the use of mixed interpolation with projection to solve (almost) incompressible elasticity problems. In this analysis the inf-sup condition is used, and we give the details of analysis to show that the (well-known) 9-node displacement/3 parameter pressure element is a reliable element to use in general two-dimensional solutions. This analysis also gives insight into how other elements could be analyzed.

We conclude the paper with some short remarks regarding the role of the inf-sup condition in the analysis of other types of problems.

2. Setting of the Problem Considered

In the following analysis we shall consider problems of the form

$$\inf_{v \in V} \; J_1(v) + \varepsilon^{-1} J_2(v) - C(v) \qquad (2)$$

where V is a Hilbert space of admissible functions, the J_i $(i=1,2)$ are quadratic functionals in V and $C(v)$ is a linear functional on V; ε is a very small real positive number. Problems of type (2) arise frequently in engineering analysis. Their origin can be two-fold: sometimes $J_1 + \varepsilon^{-1} J_2$ represents the potential energy of a system and ε^{-1} is some large physical parameter used to define the energy. This is the case for instance in the linear elasticity of nearly incompressible materials and in the analysis of Mindlin-Reissner plates. In other applications, the form (2) appears after introduction of a penalty term. This is the case in the analysis of incompressible fluids, when a penalty term $\varepsilon^{-1} \int (\operatorname{div} \underline{u})^2 \, dx$ is added to the potential energy to account for the incompressibility condition.

The common feature in all these cases is that after discretization an ε which is too small (employed with the mesh size h) is a potential source of major difficulties.

In order to start the discussion of such difficulties, let us consider two very simple, purely academic examples.

Example 1: In this example we will not, actually, discuss a problem of the type (2). We shall just make some remarks on the approximation of functions by finite element spaces. Let V_h be the space of piecewise linear continuous functions on $(-1, 1)$ vanishing at the endpoints. Consider $v(x) = 1-x^2$; it is a nice and smooth function. We know that

$$v \in C^2(-1,1) \quad \text{and} \quad \inf_{v_h \in V_h} \|v-v_h\|_1 \le c\,h$$

(here $\| \ \|_1$ is the energy norm: $\|\varphi\|_1^2 = \int_{-1}^1 (\varphi')^2 dx$). However,

$$\inf_{v_h \in V_h \cap C^1} \|v-v_h\|_1 = \inf_{v_h \in \{0\}} \|v-v_h\|_1 = \|v\|_1 = \sqrt{8/3} \quad \forall\ h.$$

Of course, we know that we would not proceed this way; but we want to point out the following philosophy. We are given a sequence of finite element spaces (here V_h) that allows us to approximate in a satisfactory manner every function in some space V (here functions in $C^2(-1,1)$, vanishing at the end-points). We choose a function v in V and recognize a nice property (here, to have a continuous derivative). Now (big mistake) we try to approximate the function v by means of functions that are in V_h <u>and</u> share the same nice property (here to have a continuous derivative). We find that the only function in V_h with continuous derivative is the zero function. As we shall see, this simple example displays the essence of the "locking" phenomenon.

<u>Example 2</u>: This again will be a very academic case, but will teach us something. Assume that we want to minimize

$$J_1(v) - C(v) = \int_0^1 (v')^2 \, dx - 4 \int_0^1 v \, dx$$

over the space of functions v(x) defined in (-1,1), vanishing at $x = \pm 1$, smooth and <u>symmetric</u>: $v(x) = v(-x)$. The solution is $v(x) = 1 - x^2$. In order to enforce the symmetry condition, we add a penalty term

$$\varepsilon^{-1} J_2(v) := \varepsilon^{-1} \int_0^1 (v(x)-v(-x))^2 \, dx.$$

The new problem is now to minimize

$$\int_0^1 (v')^2 \, dx + \varepsilon^{-1} \int_0^1 (v(x)-v(-x))^2 \, dx - 4 \int_0^1 v(x) \, dx$$

over the smooth functions $v(x)$ vanishing at $x = \pm 1$. Again the unique solution is $v(x) = 1 - x^2$ for every positive ε. Now we discretize with piecewise linear functions vanishing at $x = \pm 1$, but for each integer n, we split the subinterval [0,1] into n equal parts and the subinterval [-1,0] into n+1 equal parts (we are really looking for trouble!). Now the only symmetric function in our finite element subspace is the function identically zero. If we try to minimize the functional on the subspace V_h for very small ε (compared to $h = 1/n$) the term $\varepsilon^{-1} \int (v(x)-v(-x))^2\, dx$ will be an unbearable burden for our trial minimizer that will force the solution to remain at zero.

In both examples we made the same mistake. We had a nice function $v(x)$ and a nice sequence of subspaces V_h that was able to approximate the function in a satisfactory manner. However, by imposing a certain condition on the solution in V_h we could not obtain a reasonable solution. The important point is that when we minimize a functional of type (2) on a finite element subspace V_h with $\varepsilon > 0$ very small, we are actually minimizing

$$J_1(v) - C(v)$$

on the <u>intersection</u> of V_h with the set of functions v_h such that $J_2(v_h)=0$. If the space of intersection is too small, we loose in accuracy.

The two small examples above suggest our strategy of analysis. Assume that we are given, together with the problem (2), a sequence of finite element subspaces V_h that can approximate every smooth function of V with a "satisfactory accuracy". What are the conditions on V_h and $J_2(v)$ such that, no matter how small is $\varepsilon>0$, the discrete solution of

$$\inf_{v_h \in V_h} J_1(v_h) + \varepsilon^{-1} J_2(v_h) - C(v_h) \tag{3}$$

approximates the solution of (2) with the same "satisfactory accuracy"?

The above question has two weak points: it is not formulated in a mathematically precise manner and it is too abstract to be properly understood. Hence it will be more convenient to discuss a particular case.

3. <u>The Linear Elasticity Problem Solved Using Exact Integration</u>

We consider as a model problem the linear elasticity problem [1]

$$\inf_{\underline{v}\in V} \; 1/2\; a(\underline{v},\underline{v}) + \lambda/2 \int_\Omega (\text{div } \underline{v})^2 \, dx - \int_\Omega \underline{f}\cdot\underline{v}\, dx \qquad (4)$$

where

$$a(\underline{u},\underline{v}) := 2\,\mu \int_\Omega \sum_{i,j}^{3} \varepsilon_{ij}^{D}(\underline{u})\; \varepsilon_{ij}^{D}(\underline{v})\, dx$$

$$\varepsilon_{ij}^{D}(\underline{u}) := \varepsilon_{ij}(\underline{u}) - 1/3 \text{ div } \underline{u}\; \delta_{ij}$$

$$\varepsilon_{ij}(\underline{u}) := 1/2\, (\partial u_i/\partial x_j + \partial u_j/\partial x_i)$$

$$\lambda = \frac{E}{3(1-2\nu)} \text{(bulk modulus)}, \quad \mu = \frac{E}{2(1+\nu)} \text{(shear modulus)}$$

E = Young's modulus, ν = Poisson's ratio

$V = \{ \underline{v} \mid \partial v_i/\partial x_j \in L^2(\Omega)\ (i,j=1,2,3),$

$v_i|_{\partial\Omega} = 0\ (i=1,2,3) \}$

$$\|\underline{v}\|_V^2 = \sum_{i,j} \|\partial v_i/\partial x_j\|^2_{L^2(\Omega)} \;;\; \|q\|_0^2 = \|q\|^2_{L^2(\Omega)} \quad (*).$$

The use of homogeneous Dirichlet boundary conditions is mainly to simplify the notation. Let $\underline{u}$ be the minimizer of (4) and let V_h be a sequence of (finite element) subspaces to be specified later on. It is clear that the discrete problem

$$\inf_{\underline{v}_h\in V_h} \; 1/2\; a(\underline{v}_h,\underline{v}_h) + \lambda/2 \int_\Omega (\text{div } \underline{v}_h)^2 \, dx - \int_\Omega \underline{f}\cdot\underline{v}_h dx \qquad (5)$$

has a unique solution $\underline{u}_h$. If the sequence V_h has good approximation properties the quantity

$$d(\underline{u},V_h) := \inf_{\underline{v}_h\in V_h} \|\underline{u}-\underline{v}_h\|$$

---

*In the following discussion we will not explicitly give the subscripts on the norms but always imply that a vector $\underline{w}$ has norm $\|\underline{w}\|_V$ and a scalar γ has norm $\|\gamma\|_0$.

will be small for small h. Moreover, it must be pointed out that the solution $\underline{u}$ will be uniformly smooth no matter how large λ, $\lambda > 0$. Our purpose is to find <u>conditions on</u> V_h <u>such that</u>

$$\|\underline{u}-\underline{u}_h\| \le c\ d(\underline{u},V_h) \tag{6}$$

<u>with a constant</u> c <u>independent of</u> h <u>and</u> λ.

The inequality (6) means that the distance between the continuous solution $\underline{u}$ and the discrete solution $\underline{u}_h$ must be asymptotically of the same order of magnitude as $d(\underline{u},V_h)$. For instance if $d(\underline{u},V_h) = O(h^2)$ then (6) requires $\|\underline{u}-\underline{u}_h\|$ to be also $O(h^2)$.

On studying (5) we remark that for large λ the quantity $\|\text{div } \underline{u}_h\|$ will be small. The larger is λ, the smaller is $\|\text{div } \underline{u}_h\|$. As we have seen in Example 2, for very large λ we are practically trying to approximate $\underline{u}$ in the <u>intersection</u> of V_h with the space of vectors with vanishing divergence. It is now convenient, for the following discussion, to introduce some notation. For any scalar function $q(\underline{x})$ defined on Ω we set

$$K(q) = \{\underline{v} \mid \underline{v} \in V,\ \text{div } \underline{v} = q\}$$

$$K_h(q) = \{\underline{v}_h \mid \underline{v}_h \in V_h,\ \text{div } \underline{v}_h = q\} = V_h \cap K(q).$$

It is clear that $K_h(q) \subseteq K(q)$ for all q and for all h. When λ grows to $+\infty$ the solution $\underline{u}$ will be closer and closer to $K(0)$ while the solution $\underline{u}_h$ will be closer and closer to $K_h(0)$.

Hence it is reasonable to consider directly the two limit problems

$$\inf_{\underline{v} \in K(0)} \quad 1/2\ a(\underline{v},\underline{v}) - \int \underline{f}\cdot\underline{v} \quad (*) \tag{7}$$

and

$$\inf_{\underline{v}_h \in K_h(0)} \quad 1/2\ a(\underline{v}_h,\underline{v}_h) - \int \underline{f}\cdot\underline{v}_h\,. \tag{8}$$

For the sake of simplicity, let us call again $\underline{u}$ and $\underline{u}_h$ the minimizers of (7) and (8), respectively. We still want (6) to hold true. It is now easy to note that, since $\underline{u}_h \in K_h(0)$,

---

* We imply integration over the domain Ω.

$$d(\underline{u},K_h(0)) := \inf_{\underline{v}_h \in K_h(0)} \|\underline{u}-\underline{v}_h\| \le \|\underline{u}-\underline{u}_h\|$$

so that, *if* (6) holds, *then*

$$d(\underline{u},K_h(0)) \le c\, d(\underline{u},V_h). \tag{9}$$

The following Proposition 1 will show that actually

$$\|\underline{u}-\underline{u}_h\| \le \gamma\, d(\underline{u},K_h(0)) \tag{10}$$

so that $\|\underline{u}-\underline{u}_h\| \simeq d(\underline{u},K_h(0))$ and hence (6) and (9) are equivalent.

Proposition 1. Let $\underline{u}$ and $\underline{u}_h$ be the minimizers of (7) and (8) respectively. Then (10) holds with γ independent of h.

Proof. It is enough to note that:

(i) $\quad \exists\, \alpha > 0\ \forall\, \underline{v} \in K(0) \quad a(\underline{v},\underline{v}) \ge \alpha\, \|\underline{v}\|^2$

(ii) $\quad \exists\, M > 0 \quad \forall\, \underline{v}_1,\underline{v}_2 \in V \quad a(\underline{v}_1,\underline{v}_2) \le M\|\underline{v}_1\|\ \|\underline{v}_2\|$

and that the Euler equations of (7) and (8) are

(iii) $\quad a(\underline{u},\underline{v}) = (\underline{f},\underline{v})\ \forall\, \underline{v} \in K(0)$

(iv) $\quad a(\underline{u}_h,\underline{v}_h) = (\underline{f},\underline{v}_h)\ \forall\, \underline{v}_h \in K_h(0).$

Then the (well-known) argument goes like this: for any $\tilde{\underline{u}}_h \in K_h(0)$ we have

$\alpha\|\underline{u}-\underline{u}_h\|^2 \le$ (use(i))

$\le a(\underline{u}-\underline{u}_h,\underline{u}-\underline{u}_h) =$ (add and subtract $\tilde{\underline{u}}_h$)

$= a(\underline{u}-\underline{u}_h,\underline{u}-\tilde{\underline{u}}_h) + a(\underline{u}-\underline{u}_h,\ \tilde{\underline{u}}_h-\underline{u}_h) =$ (use (iii), (iv))

$= a(\underline{u}-\underline{u}_h,\underline{u}-\tilde{\underline{u}}_h) \le$ (use (ii))

$\le M\ \|\underline{u}-\underline{u}_h\|\ \|\underline{u}-\tilde{\underline{u}}_h\|$

so that, for any $\tilde{\underline{u}}_h$ in $K_h(0)$ we have

$$\|\underline{u}-\underline{u}_h\| \le M/\alpha\ \ \|\underline{u}-\tilde{\underline{u}}_h\|$$

which implies (10) with $\gamma = M/\alpha$. ∎

We have assumed so far that the forcing vector $\underline{f}$ is a fixed quantity. Hence, we essentially considered one single, specific continuous solution $\underline{u}$. However, if we want our theory to hold for all possible right-hand sides $\underline{f}$, we must require that (9) (or equivalently (6)) holds for all possible solutions $\underline{u}$. That is, we must require that (9) holds for any $\underline{u}$ in V. We need now a further notation:

$$D_h = \{q \mid q = \text{div } \underline{v}_h \text{ for some } \underline{v}_h \text{ in } V_h\}.$$

In other words, D_h is the space of the divergences of vectors belonging to V_h.

Proposition 2. Condition (9) holds for every $\underline{u} \in K(0)$ if and only if the following condition holds.

$$\left.\begin{array}{l} \text{For all } q \in D_h \text{ and for all } \underline{u} \in K(q) \\ \displaystyle\inf_{\underline{v}_h \in K_h(q)} \|\underline{u}-\underline{v}_h\| \le c \inf_{\underline{v}_h \in V_h} \|\underline{u}-\underline{v}_h\| \end{array}\right\} \qquad (11)$$

with c independent of h.

Proof. (11) implies (9) by simply taking $q = 0$. Assume now (9) to hold; take a $q \neq 0$ and a $\underline{u}$ in $K(q)$. Since $q \in D_h$ we have $q = \text{div } \underline{w}_h$ for some $\underline{w}_h$ in V_h. Set $\underline{u}_0 = \underline{u} - \underline{w}_h$ and write $\underline{v}_h \in K_h(q)$ as $\underline{v}_h = \underline{v}_h^0 + \underline{w}_h$ with $\underline{v}_h^0 \in K_h(0)$

$$\inf_{\underline{v}_h \in K_h(q)} \|\underline{u}-\underline{v}_h\| = \inf_{\underline{v}_h^0 \in K_h(0)} \|\underline{u} - \underline{v}_h^0 - \underline{w}_h\| =$$

$$= \inf_{\underline{v}_h^0 \in K_h(0)} \|\underline{u}_0-\underline{v}_h^0\| \le \text{ (use (9)) } \le$$

$$\le c \inf_{\underline{v}_h \in V_h} \|\underline{u}_0-\underline{v}_h\| = c \inf_{\underline{v}_h \in V_h} \|\underline{u}_0 + \underline{w}_h - \underline{v}_h\|$$

$$= c \inf_{\underline{v}_h \in V_h} \|\underline{u}-\underline{v}_h\| \qquad \blacksquare$$

The above result simply expresses the fact that $K(q)$ and $K_h(q)$ are simple translations of $K(0)$ and $K_h(0)$ by a common vector (here: $\underline{w}_h$). This clearly does not change the quality of the approximation. Now comes a slightly more sophisticated result.

Theorem 1. Condition (11) is equivalent to

$$\forall\ q \in D_h\ \exists\ \underline{w}_h \in K_h(q) \qquad \|\underline{w}_h\| \le c\|q\| \tag{12}$$

c independent of q, $\underline{w}_h$ and h.

Before starting the proof, we note that a basic result in partial differential equations ensures that the "continuous" version of (12) always holds true, namely

$$\forall\ q \in D\ \exists\ \underline{w} \in K(q) \qquad \|\underline{w}\| \le \gamma\ \|q\| \tag{13}$$

γ independent of q and $\underline{w}$

(with $D = \{q | q \in L^2(\Omega), \int q = 0\}$; note that $D_h \subset D$). We refer for instance to Ladyzhenskaya [2] or Temam [3] for a proof of (13).

Proof of Theorem 1. We assume first that (11) holds, and we prove (12). For this, let $q \in D_h$ and let $\underline{w} \in K(q)$ be such that $\|\underline{w}\| \le \gamma\ \|q\|$. Such a $\underline{w}$ surely exists due to (13). Take now $\underline{w}_h \in K_h(q)$ such that

$$\|\underline{w}-\underline{w}_h\| = \inf_{\underline{v}_h \in K_h(q)} \|\underline{w}-\underline{v}_h\| .$$

From (11) we have

$$\|\underline{w}-\underline{w}_h\| \le c \inf_{\underline{v}_h \in V_h} \|\underline{w}-\underline{v}_h\| \le c\|\underline{w}\| \quad (\text{use } \underline{v}_h = \underline{0})$$

and therefore

$$\|\underline{w}_h\| \le \|\underline{w}-\underline{w}_h\| + \|\underline{w}\| \le (1+c)\ \|\underline{w}\| \le (1+c)\ \gamma\ \|q\| .$$

We have proved that (11) implies (12). Assume now that (12) holds true. For any $q \in D_h$ and for any $\underline{u} \in K(q)$ let $\overline{\underline{w}}_h \in V_h$ be such that

$$\|\underline{u} - \overline{\underline{w}}_h\| = \inf_{\underline{v}_h \in V_h} \|\underline{u} - \underline{v}_h\| . \tag{14}$$

Unfortunately $\underline{\bar{w}}_h \notin K_h(q)$ in general. Set now $\tilde{q} = q - \operatorname{div} \underline{\bar{w}}_h = \operatorname{div} \underline{u} - \operatorname{div} \underline{\bar{w}}_h \in D_h$. Since the div operator is continuous from V into $L^2(\Omega)$ we have

$$\|\tilde{q}\| \leq c_1 \|\underline{u}-\underline{\bar{w}}_h\|. \quad (c_1 \text{ is actually} = \sqrt{3}) \tag{15}$$

Take now $\underline{w}_h \in K_h(\tilde{q})$ such that

$$\|\underline{w}_h\| \leq c_2 \|\tilde{q}\|. \tag{16}$$

This is possible due to (12). Note that $\operatorname{div}(\underline{w}_h+\underline{\bar{w}}_h) = \tilde{q} + \operatorname{div} \underline{\bar{w}}_h = q$.

Hence

$$\inf_{\underline{v}_h \in K_h(q)} \|\underline{u}-\underline{v}_h\| \leq \|\underline{u} - (\underline{w}_h+\underline{\bar{w}}_h)\| \leq \text{(triangle ineq)}.$$

$$\leq \|\underline{u} - \underline{\bar{w}}_h\| + \|\underline{w}_h\| \leq \quad \text{(use (16))}$$

$$\leq \|\underline{u} - \underline{\bar{w}}_h\| + c_2\|\tilde{q}\| \leq \quad \text{(use (15))}$$

$$\leq \|\underline{u} - \underline{\bar{w}}_h\| + c_2 c_1 \|\underline{u} - \underline{\bar{w}}_h\| =$$

$$= (1+c_2c_1) \|\underline{u} - \underline{\bar{w}}_h\| = \text{(use (14))}$$

$$= (1+c_2c_1) \inf_{\underline{v}_h \in V_h} \|\underline{u} - \underline{v}_h\| \qquad \blacksquare$$

An equivalent formulation of (12) is the following inf-sup condition

$$\inf_{q \in D_h} \sup_{\underline{v}_h \in V_h} \frac{\int q \operatorname{div} \underline{v}_h}{\|\underline{v}_h\| \, \|q\|} \geq \beta > 0 \tag{17}$$

with β independent of h.

Let us see for instance that (12) implies (17). Let q be an element of D_h. From (12) we have that there exists a $\underline{w}_h$ with $\operatorname{div} \underline{w}_h = q$ and $\|\underline{w}_h\| \leq c\,\|q\|$. Multiplying this last inequality by $\|q\|$ we obtain $\|\underline{w}_h\| \, \|q\| \leq c\|q\|^2$. Now $\|q\|^2 = \int q^2 = \int q \operatorname{div} \underline{w}_h$. Therefore, $\|\underline{w}_h\| \, \|q\| \leq c \int q \operatorname{div} \underline{w}_h$. This can be written as

$$\frac{\int q \operatorname{div} \underline{w}_h}{\|q\| \, \|\underline{w}_h\|} \geq \frac{1}{c} \quad . \tag{18}$$

Summarizing we have: for any $q \in D_h$ there exists a $\underline{w}_h \in V_h$ such that (18) holds. This clearly implies (17) with $\beta \geq 1/c$. The opposite implication (17) $\Rightarrow$ (12) is also true but the proof is a little more technical. Since we are not going to use the form (17) here, we omit the proof, see for instance Brezzi [4].

A more interesting equivalent formulation of (12) is given by the following proposition.

Proposition 3. Condition (12) is equivalent to

$$\left.\begin{array}{ll} \forall\, \underline{u} \in V\ \exists\, \underline{u}_I \in V_h \text{ such that} & \\ \text{(i)} \quad \int \operatorname{div}(\underline{u}-\underline{u}_I)\, q = 0 \quad \forall\, q \in D_h & \\ \text{(ii)} \quad \|\underline{u}_I\| \leq c\, \|\underline{u}\| & \\ \text{with } c \text{ independent of } \underline{u},\ \underline{u}_I \text{ and } h. & \end{array}\right\} \tag{19}$$

Proof. Assume first that (12) holds and prove (19). For this, let $\underline{u} \in V$ be given and let $\bar{q}$ be the L^2-projection of $\operatorname{div} \underline{u}$ onto D_h; hence $\bar{q} \in D_h$ and $\int(\operatorname{div} \underline{u} - \bar{q})q = 0\ \forall\, q \in D_h$, and

$$\|\bar{q}\| \leq \|\operatorname{div} \underline{u}\| \leq \sqrt{3}\,\|\underline{u}\| \tag{20}$$

Using (12) we can find a $\underline{u}_I \in K_h(\bar{q})$ with

$$\|\underline{u}_I\| \leq c\|\bar{q}\| . \tag{21}$$

Now since $\underline{u}_I \in K_h(\bar{q})$ we have that (19;i) is satisfied, and using (20), (21) we have that (19;ii) is also satisfied.

Assume now that (19) is satisfied and prove (12). For every $q \in D_h$ we have (using (13)) that there exists a $\underline{w} \in K(q)$ with

$$\|\underline{w}\| \leq \gamma\, \|q\| . \tag{22}$$

Using now (19) with $\underline{u} = \underline{w}$ we obtain a $\underline{w}_I$ such that

$$\int (q - \operatorname{div} \underline{w}_I)\delta q = 0 \quad \forall\, \delta q \in D_h$$

which implies $\operatorname{div} \underline{w}_I = q$ and hence $\underline{w}_I \in K_h(q)$. Moreover by (19;ii) we have $\|\underline{w}_I\| \leq c\, \|\underline{w}\|$ which together with (22) gives $\|\underline{w}_I\| \leq c\, \gamma\, \|q\|$. Therefore (12) is proved by taking $\underline{w}_h = \underline{w}_I$. ■

Let us now comment on our formulations of the inf-sup condition. Condition (6) is really what we want: this condition implies an optimal error bound. If we want to satisfy it for every possible right hand side $\underline{f}$ we must satisfy it for every $\underline{u}$ in V. If we want that our method is uniformly effective for $\lambda \to +\infty$ we are naturally led to require (9) for every $\underline{u} \in K(0)$. The formulation (11) is just an intermediate step to reach the powerful formulation (12). Actually, if we know that the inf-sup condition is satisfied and if we want to use this fact for proving other results, the form (12) is highly recommended: it is simple, it is easy to manipulate and it contains all the information. On the other hand, if we want to prove that our discretization satisfies the inf-sup condition then the form (19) (also known as "Fortin's trick") is the more easy way in almost all cases. We shall use (19) in the next section.

Since we observed that (12) is a convenient form, let us express it in words and consider the interesting case q very small.

The form (12) says: it is not forbidden to have "large" vectors (i.e., $\|\underline{w}\|$ is large) with small divergence in the vector space V_h as long as there are enough "small" vectors with small divergence.

Namely, consider that we have an approximation $\underline{u}_I$ to the solution $\underline{u}$. Convergence is measured by the norm $\|\underline{u}-\underline{u}_I\|$. Assume that $q = \operatorname{div} u$ is very small and $\operatorname{div} \underline{u}_I$ is too large (we recall that $\operatorname{div} \underline{u}_I$ is multiplied by λ and any error will have a very large effect on the solution). Hence we want to add a vector, say $\underline{w}_h$, to $\underline{u}_I$ in order to more closely satisfy the required (small) value on the divergence. The addition of $\underline{w}_h$ to $\underline{u}_I$ should give a vector with smaller divergence and closer to (or at least not farther away from) $\underline{u}$.

In other words, if the only $\underline{w}_h$ vector with small divergence we can add to $\underline{u}_I$ is a "large" vector then the error $\|\underline{u}-\underline{u}_I\|$ will increase and we encounter convergence difficulties. On the other hand, if (12) does hold (that is if the kernel $K_h(0)$ is large enough) then we can find a "small" $\underline{w}_h$ (with also a small divergence) to correct the vector $\underline{u}_I$ to obtain a new approximation $\underline{u}_I-\underline{w}_h$ which is still close to $\underline{u}$ but has a much smaller divergence.

If we consider our condition (12) as "$K_h(0)$ must be large enough" we can better understand the so-called "locking" phenomenon. We have seen that, basically, locking corresponds to $K_h(0) = \{0\}$ or, more generally $K_h(0)$ very small. Hence, locking corresponds in general to a strong violation of (12). We should also point out that the absence of locking (i.e., $K_h(0) \neq \{0\}$) does not imply that we obtain an accurate solution (that is, it does not imply that $K_h(0)$ is large enough).

Let us end this section with a brief summary. We have observed (in (6)) that we want the error $\|\underline{u}-\underline{u}_h\|$ to be of the same order as $d(\underline{u},V_h)$, uniformly in $\lambda \to +\infty$. This took us to consider the two limit problems (7) and (8) and require again (6) with $\underline{u}$ and $\underline{u}_h$ now the solutions of the limit problems. We found (6) to be equivalent to (9). Since we want our argument to hold for every righthand side $\underline{f}$, we require (9) to hold for every $\underline{u}$ in K(0). Then we proved that this was equivalent to (11), that (11) was equivalent to (12) and that (12) was equivalent to the inf-sup condition (17). Finally, we gave as a last equivalent formulation (19), the so-called "Fortin's trick". Now we can (reasonably) ask: what are the finite element spaces such that (19) (hence (17), (12), $\cdots$, (6)) is satisfied? Unfortunately, not many and , in general, not the space that we would like to use. Under minor assumptions on the triangulation, Scott and Vogelius [5] proved that (17) is satisfied when V_h is the space of piecewise polynomials of degree $k \geq 4$. If we want to use lower-order methods, it seems that we must do something special when dealing with the term $\lambda \int (\operatorname{div} v)^2 dx$ in the functional. This is done in the next section.

4. <u>Mixed Interpolation with Projection</u>

There are various approaches for reducing the bad influence of a term like $\lambda\int(\operatorname{div} \underline{v})^2$ in our functional. For the sake of simplicity, we shall just analyze one family of such methods. The general idea for the whole family consists of

(1) choosing a new finite element space Q_h (usually made of functions that are discontinuous across the elements) and

(2) substitute in place of $\lambda\int(\operatorname{div} \underline{v})^2$ the "weaker" term

$$\lambda \int (P(\operatorname{div} \underline{v}))^2$$

where P is, by definition, the L^2-projection operator on Q_h. This means that for a given $\varphi \in L^2(\Omega)$ the projection $P\varphi$ is defined as the unique element $p = P\varphi$ in Q_h such that

$$\int (p-\varphi)q = 0 \quad \forall\, q \in Q_h . \tag{23}$$

Note that, if Q_h consists of discontinuous basis functions the projection (23) can be computed element by element. Let us consider the continuous and the discrete problem and try to deduce an error estimate.

Continuous problem: find $\underline{u} \in \underline{V}$ such that

$$a(\underline{u},\underline{v}) + \lambda \int \operatorname{div} \underline{u} \operatorname{div} \underline{v} = \int \underline{f}\cdot\underline{v} \qquad \forall\, \underline{v} \in V. \tag{24}$$

Discrete problem: find $\underline{u}_h \in V_h$ such that

$$a(\underline{u}_h,\underline{v}_h) + \lambda \int (P \operatorname{div} \underline{u}_h)(P \operatorname{div} \underline{v}_h) = \int \underline{f}\cdot\underline{v}_h \qquad \forall\, \underline{v}_h \in V_h. \tag{25}$$

Recall that the bilinear form $a(\underline{u},\underline{v})$, defined in (4) satisfies (i) of Proposition 1. Also,

$$\text{(i')} \qquad \alpha\|\underline{v}_h\|^2 \le a(\underline{v}_h, \underline{v}_h) + \lambda\, \|PD\underline{v}_h\|^2 .$$

Let now $\underline{u}_I$ be an element of V_h to be chosen later on "close to $\underline{u}$". In order to simplify the formulae, we shall also write here D instead of div, we shall set $\underline{w}_h := \underline{u}_I - \underline{u}_h$ and recall that P being a projection operator,

$$\int (P\chi)(P\phi)dx = \int \chi\, P\phi\, dx \qquad \forall\, \chi,\ \phi \in L^2(\Omega).$$

Now we have:

$$\begin{aligned} a(\underline{w}_h,\underline{w}_h) &= a(\underline{u}_I-\underline{u}_h,\underline{w}_h) = \\ &= a(\underline{u}_I-\underline{u},\underline{w}_h) + a(\underline{u}-\underline{u}_h,\underline{w}_h) = \\ &= I + a(\underline{u},\underline{w}_h) - a(\underline{u}_h,\underline{w}_h) = \text{(use (24)(25))} \\ &= I + \int\underline{f}\cdot\underline{w}_h - \lambda\int D\underline{u}\, D\underline{w}_h - \left(\int\underline{f}\cdot\underline{w}_h - \lambda\int PD\underline{u}_h\, PD\underline{w}_h\right) = \\ &= I + \lambda\left[\int(PD\underline{u}-D\underline{u})D\underline{w}_h - \int PD\underline{u}\, D\underline{w}_h + \int PD\underline{u}_h\, PD\underline{w}_h\right] = \\ &= I + II + \lambda\left[-\int PD\underline{u}\, D\underline{w}_h + \int PD\underline{u}_h\, PD\underline{w}_h\right] = \\ &= I + II + \lambda\left[\int(PD\underline{u}_I-PD\underline{u})D\underline{w}_h + \int(PD\underline{u}_h-PD\underline{u}_I)PD\underline{w}_h\right] = \\ &= I + II + \lambda\int(PD\underline{u}_I-PD\underline{u})D\underline{w}_h - \lambda\, \|PD\underline{w}_h\|^2 \end{aligned} \tag{26}$$

with obvious notations for I and II. From (i') and (26) we have (moving the last term to the left-hand side)

$$\begin{aligned} \alpha\|\underline{w}_h\|^2 &\le a(\underline{w}_h, \underline{w}_h) + \lambda\|PD\underline{w}_h\|^2 \le I + II \\ &\quad + \lambda \int (PD\underline{u}_I - PD\underline{u})D\underline{w}_h = I + II + III. \end{aligned} \tag{27}$$

Let us analyze the three terms on the right-hand side: we can easily bound I and using (ii) of Proposition 1 we have

$$I = a(\underline{u}_I-\underline{u},\underline{w}_h) \le M\|\underline{u}-\underline{u}_I\| \; \|\underline{w}_h\|. \tag{28}$$

In order to make I small we must choose $\underline{u}_I$ close to $\underline{u}$: no surprise! The term II measures the error in the projection

$$II = \lambda \int(PD\underline{u}-D\underline{u})D\underline{w}_h = \int(P-I)p\; D\underline{w}_h \le$$
$$\le \|(P-I)p\| \; \|\underline{w}_h\| \qquad (p:=\lambda D\underline{u}) \; . \tag{29}$$

In order to make II small we have to choose P close to the identity, or, in other words, Q_h large enough. Somehow here we pay for substituting for the original term $\int(\text{div } \underline{u}_h)^2$ the term $\int(P \text{ div } \underline{u}_h)^2$. This is also very reasonable. The third term, however, can introduce major difficulties,

$$III = \lambda \int(PD\underline{u}_I - PD\underline{u})D\underline{w}_h. \tag{30}$$

If λ is very large, then $PD\underline{u}$ will be very small (recall that $\lambda D\underline{u}$ is bounded). Hence we need $PD\underline{u}_I$ to be also very small, with $\underline{u}_I$ close to $\underline{u}$ as needed in the estimate of I. In the limit for $\lambda\to\infty$ we need an approximation $\underline{u}_I$ of $\underline{u}$ with P div $\underline{u}_I$ = 0 (now div $\underline{u}$ = 0). More generally, if we want III to be small uniformly in λ, we cannot accept that $\int(PD\underline{u}_I-PD\underline{u})D\underline{w}_h$ is some $O(h^k)$: we need that it is <u>exactly zero</u>. So this is the condition: we must find $\underline{u}_I$ such that

$$\int(PD\underline{u}_I-PD\underline{u})D\underline{v}_h = 0 \quad \forall\; \underline{v}_h \in V_h.$$

This can be written as

$$\int(D\underline{u}_I-D\underline{u})PD\underline{v}_h = 0 \qquad \forall\; \underline{v}_h \in V_h. \tag{31}$$

For every $\underline{u}_I$ satisfying (31) we will obtain using (27) to (31)

$$\|\underline{u}_h-\underline{u}\| \le c(\|\underline{u}-\underline{u}_I\| + \|(I-P)p\|) \tag{32}$$

where as usual $p = \lambda D\underline{u} = \lambda$ div $\underline{u}$.

We propose to keep the inf-sup condition in the form (31) (32): that is, for every $\underline{u}_I$ satisfying (31) we obtain the bound (32). It is clear that in practice this means: we must be able to build (or to show that it exists) an "interpolation" of $\underline{u}$ that satisfies (31) and is close to $\underline{u}$. Note that this, in particular, strongly excludes "locking" behavior. Consider for simplicity the limiting case in which div $\underline{u}$ = 0. We need that the set

$$\overline{K}_h(0) = \{\underline{v}_h | \underline{v}_h \in V_h, \quad P \text{ div } \underline{v}_h = 0\} \tag{33}$$

is large enough, so that in $\overline{K}_h(0)$ we can find a $\underline{u}_I$ close to $\underline{u}$. If $\overline{K}_h(0) = \{0\}$ (or if it contains only a few elements) this is clearly not possible (unless the true solution $\underline{u}$ is itself zero).

Let us comment further on this formulation. First of all, note that we can always take P = identity. This in general is not convenient, but it is clearly allowed. The choice P = identity brings us back to the original formulation of Section 1. Therefore, we have proved in particular that if $\underline{u}$ and $\underline{u}_h$ are the solutions of (4) and (5) respectively, then for any $\underline{u}_I \in V_h$ satisfying

$$\int \text{div}(\underline{u}-\underline{u}_I)\text{div } \underline{v}_h = 0 \qquad \forall \ \underline{v}_h \in V_h \tag{34}$$

we have

$$\|\underline{u}-\underline{u}_h\| \le c\|\underline{u}-\underline{u}_I\| \tag{35}$$

with c independent of $\underline{u}$, $\underline{u}_I$ and h.

It can also be interesting to compare this result with (19) of Section 1. Basically there we had that if (19) holds, then we have optimal error bounds. Condition (19) essentially required that for any $\underline{u}$ in $\underline{V}$ we can find a $\underline{u}_I$ in V_h satisfying (34) and $\|\underline{u}_I\| \le c \|\underline{u}\|$. Here we just say that $\underline{u}$ being the true solution, for any $\underline{u}_I$ satisfying (34) we have (35) which in itself is not an optimal error bound but can become such in applications. Finally we remark that the implication (34) $\Rightarrow$ (35) is not a condition: it is a theorem that holds for every choice of V_h.

Let us now comment on the more useful cases P $\neq$ identity, to see what benefit is reached. Consider for the sake of simplicity the limiting case $\lambda = +\infty$, so that div $\underline{u} = 0$. We have seen that the aim is to find a $\underline{u}_I$ satisfying P div $\underline{u}_I = 0$. If P is the identity operator, such $\underline{u}_I$ can be difficult to find. This will be the case if Q_h is too large. However, as we reduce the space Q_h, the condition P div $\underline{u}_I = 0$ becomes less and less difficult to enforce: for instance if Q_h is made of piecewise constants then we have only to require that div $\underline{u}_I$ has zero mean value on each element, which gives a much wider choice for $\underline{u}_I$. Hence we have more possibilities of finding a $\underline{u}_I$ that is close to $\underline{u}$. Using the notation (33) it is easy to observe that $\overline{K}_h(0)$ becomes larger as the space Q_h becomes smaller.

Let us consider, as an example, how to deal with (31) in some practical 2-D cases. For instance we can consider V_h to consist of piecewise biquadratic polynomials (usually called Q_2) and P to be the projection on the space of linear functions (usually called P_1): we are then using the Q_2 - P_1 scheme. Given $\underline{u}$ smooth we must find an interpolation $\underline{u}_I$ such that, for each element e,

$$\int_e (\operatorname{div} \underline{u}_I - \operatorname{div} \underline{u}) q = 0 \tag{36}$$

for all q polynomials of degree ≤ 1 in e. To define $\underline{u}_I$ we must prescribe the values of each component at the nine element nodes (vertices, midpoints and center). We start with the vertices V_1, V_2, V_3, V_4 and require

$$\underline{u}_I(V_j) = \underline{u}(V_j) \qquad \text{(8 conditions).} \tag{37}$$

Then we adjust the values at the midpoints $M_1, \ldots, M_4$ in such a way that

$$\int_{L_j} (\underline{u}-\underline{u}_I)\cdot\underline{n}\, d\ell = \int_{L_j} (\underline{u}-\underline{u}_I)\cdot\underline{\tau}\, d\ell = 0 \qquad \text{(8 conditions)} \tag{38}$$

for every edge $L_1, \ldots, L_4$, with $\underline{n}$ = normal vector and $\underline{\tau}$ = tangential vector. Note that (38) in particular implies, for every constant q

$$\int_e \operatorname{div}(\underline{u}-\underline{u}_I)\, q = q \int_{\partial e} (\underline{u}-\underline{u}_I)\cdot\underline{n}\, d\ell = 0. \tag{39}$$

We are left to use the 2 degrees of freedom at the element center node. We choose these in such a way that

$$\int_e \operatorname{div}(\underline{u}-\underline{u}_I) x_1 = \int_e \operatorname{div}(\underline{u}-\underline{u}_I) x_2 = 0 . \tag{40}$$

Note now that easily (39) and (40) imply (36). Note also that $\underline{u}_I$, constructed element by element through (37), (38), (40) will be continuous from element to element (we need this continuity). Finally, note that clearly if $\underline{u}$ is a (vector) polynomial of degree ≤ 2 on e we obtain $\underline{u}_I \equiv \underline{u}$ and this ensures optimal bounds for $\|\underline{u}-\underline{u}_I\|$. We also observe that it is not so easy to require in addition to (40) the condition $\int \operatorname{div}(\underline{u}-\underline{u}_I)x_1x_2 = 0$, that is, to use a bilinear variation instead of a linear variation on pressure. We cannot give up easily conditions (37), (38) because of the continuity requirements (that is, the values prescribed for $\underline{u}_I$ on one edge must depend only on the values of $\underline{u}$ on that edge). Once we used (37), (38) we employed already 16 degrees of freedom and the 2 degrees freedom left are just sufficient to obtain (40), while (39) is automatically satisfied (because of (38)). However, a possibility might be to define $\underline{u}_I$ by blocks of elements instead of using the element by element construction (see e.g., Johnson-Pitkäranta [6]).

The same arguments used for the Q_2 - P_1 element can be employed for the P_2^* - P_1 element on triangles, where P_2^* is the space of piecewise quadratics augmented with the cubic bubble functions. We fix the three vertices first, then the midpoints in order to preserve the mean value of the normal components: this takes care of q = constant as in (39). Then we are left with the two bubble functions that we can use as in (40).

Incidentally we have also proved the convergence of the P_2 -P_0 element for triangles and of the Q_2^{red} - P_0 element for quadrilaterals(where Q_2^{red} is the 8-node element): we do not use the two bubble functions and the two conditions (40).

There are also less easy cases; for instance, the Taylor-Hood element, for which it is possible to have (34) (hence optimal error bounds) but the construction is complicated (see Bercovier-Pironneau [7] and Verfürth [8]). The Q_1 - P_0 element is also a major source of difficulties and cannot be treated completely by our simplified analysis, see for example [6], [9].

Remark. We only considered two-dimensional examples. However the same approach is used for three-dimensional cases. Roughly: using one degree of freedom per face we can deal with a constant q. Then we use the bubble functions to deal with a linear q (if necessary). We still must use vertices and edges to ensure the continuity of $\underline{u}_I$, unless we work with nonconforming trial functions.

Let us finally discuss how we can obtain estimates on the error in the pressure, that is on $\|\lambda D\underline{u} - \lambda PD\underline{u}_h\|$. Using the formulation of this section we obtain only

$$\lambda\|PD\underline{w}_h\|^2 \le c(\|\underline{u}-\underline{u}_I\| + \|(P-I)p\|)^2 \qquad (41)$$

where $\underline{w}_h$ is always $\underline{u}_I$ - $\underline{u}_h$. This error bound is not good enough because the left-hand side is multiplied by λ and not λ^2.

In order to obtain a better error bound on the pressures we can proceed as follows. Assume always that $\underline{u}_I$ satisfies (31). Then we have from (24) and (25)

$$\left.\begin{aligned} \lambda \int PD\underline{u}_h\, PD\underline{v}_h &= -\, a(\underline{u}_h,\underline{v}_h) + (\underline{f},\underline{v}_h) \\ \lambda \int PD\underline{u}_I\, PD\underline{v}_h &= -a(\underline{u},\underline{v}_h) + (\underline{f},\underline{v}_h) \\ &\quad + \lambda \int (PD\underline{u}-D\underline{u})D\underline{v}_h \end{aligned}\right\} \forall\, \underline{v}_h \in V_h$$

so that

$$\begin{aligned} \lambda \int PD\underline{w}_h\, PD\underline{v}_h &= a(\underline{u}_h-\underline{u},\underline{v}_h) \\ &\quad + \lambda \int (PD\underline{u}-D\underline{u})\, D\underline{v}_h \qquad \forall\, \underline{v}_h \in V_h. \end{aligned} \tag{42}$$

Assume now that our choice of spaces V_h and Q_h is such that

$$\forall\, q \in P(D_h)\ \exists\, \underline{z}_h \in \overline{K}_h(q), \qquad \|\underline{z}_h\| \le c\|q\| \tag{43}$$

where

$$P(D_h) = \{q \mid q \in Q_h,\quad q = PD\underline{v}_h \ \text{ for some } \underline{v}_h \in V_h\}$$

and (according to (33))

$$\overline{K}_h(q) = \{\underline{v}_h \mid \underline{v}_h \in V_h,\ PD\underline{v}_h = q\}.$$

Then we can apply (43) to $q = PD\underline{w}_h$ and have, using $\underline{v}_h = \underline{z}_h$ in (42),

$$\begin{aligned} \lambda\|PD\underline{w}_h\|^2 &= a(\underline{u}_h-\underline{u},\underline{z}_h) + \lambda \int (PD\underline{u}-D\underline{u})D\underline{z}_h \le M\|\underline{u}-\underline{u}_h\|\,\|\underline{z}_h\| \\ &\quad + \|(I-P)p\|\ \|D\underline{z}_h\| \le \\ &\le c_1(\|\underline{u}-\underline{u}_h\|\ \|PD\underline{w}_h\| + \|(I-P)p\|\ \|PD\underline{w}_h\|) \end{aligned}$$

so that actually

$$\lambda\ \|PD\underline{w}_h\| \le c_1(\|\underline{u}-\underline{u}_h\| + \|(I-P)p\|). \tag{44}$$

Then we can use the triangle inequality and obtain

$$\lambda\|D\underline{u}-PD\underline{u}_h\| \leq \lambda\ \|D\underline{u}\ -\ PD\underline{u}\| + \lambda\|PD\underline{u}\ -\ PD\underline{u}_I\| +$$
$$+ \lambda\ \|PD\underline{w}_h\| = \lambda\|D\underline{u}\ -\ PD\underline{u}\| + \lambda\ \|PD\underline{w}_h\| \tag{45}$$

(because $PD\underline{u} = PD\underline{u}_I$ from (31)).

Using now (45), (44) and (32) we obtain

$$\|\lambda D\underline{u}\ -\ \lambda PD\underline{u}_h\| \leq \|(I-P)p\| + c_2(\|(I-P)p\| + \|\underline{u}-\underline{u}_I\|)$$
$$\leq (1+c_2)(\|(I-P)p\| + \|\underline{u}-\underline{u}_I\|) \tag{46}$$

which is an error bound as good as (32).

Let us now see what kind of sufficient condition can be used in order to obtain (43). Note that (43) has a form similar to (12), and it is therefore <u>easy to use</u>. We shall give now an equivalent formulation that is similar to (19) and therefore <u>easy to prove</u>. Namely,

$$\left.\begin{array}{ll} \multicolumn{2}{l}{\forall\ \underline{u} \in V\ \exists\ \underline{u}_I \in V_h \text{ such that}} \\ \text{(i)} & \int(D\underline{u}-D\underline{u}_I)q = 0\ \forall\ q \in\ P(D_h) \\ \text{(ii)} & \|\underline{u}_I\| \leq c\ \|\underline{u}\| \\ \multicolumn{2}{l}{\text{with } c \text{ independent of } \underline{u},\ \underline{u}_I \text{ and } h.} \end{array}\right\} \tag{47}$$

The proof of the equivalence between (43) and (47) is practically identical to the proof of the equivalence between (12) and (19). In its turn, the construction of $\underline{u}_I$ to show that (47) holds in practical cases follows the same ideas that we have used for the construction of a $\underline{u}_I$ satisfying (31) (up to some minor technical difficulties connected with the fact that a general $\underline{u}$ in V can be less regular than the solution of a smooth problem).

5. <u>Concluding Remarks</u>

The objective in this paper was to discuss the inf-sup condition in its various forms and show some applications. This condition plays an important role in the development of reliable finite elements that need to satisfy internal constraints, such as the incompressibility condition.

We considered here the special case of linear elasticity problems. We would arrive at the same results when studying the case of incompressible fluids. The only difference would be a rather philosophical issue: for incompressible fluids we really would like to enforce the condition div $\underline{u} = 0$ and in a formulation of type (2), ϵ is a penalty parameter. On the other hand, for linear elasticity problems λ is a physical parameter and one would like to have estimates that hold uniformly for λ very large and for λ "not so large" (as we did in this last section).

The situation can be quite different in other applications, like the analysis of Mindlin-Reissner plates. For Mindlin-Reissner plates we have a more complicated operator instead of the divergence operator (see Bathe-Brezzi for an example of analysis of such problems [10]). However, we believe that the ideas discussed here, as pertaining to some reference problem -- the linear elasticity problem treated here -- are also helpful in approaching more complex situations.

References

[1] K. J. Bathe, Finite Element Procedures in Engineering Analysis, Prentice-Hall, 1982.

[2] O.A. Ladyzhenskaya, The Mathematical Theory of Viscous Incompressible Flows, Gordon and Breach, 1969.

[3] R. Temam, Navier-Stokes Equations, North Holland, 1977.

[4] F. Brezzi, "On the Existence, Uniqueness, and Approximation of Saddle-Point Problems Arising from Lagrange Multipliers ", R.A.I.R.O., B-R2, pp. 129-151, 1974.

[5] L. R. Scott and M. Vogelius, "Norm Estimates for a Maximal Right Inverse of the Divergence Operator in Spaces of Piecewise Polynomials", M^2.A.N., Vol. 19, pp. 111-143, 1985.

[6] C. Johnson and J. Pitkäranta, "Analysis of Some Mixed Finite Element Methods Related to Reduced Integration", Math. Comp., Vol. 38, pp. 375-400, 1982.

[7] M. Bercovier and O. Pironneau, "Error Estimates for Finite Element Solution of the Stokes Problem in the Primitive Variables", Numer. Math. Vol. 33, pp. 211-224, 1979.

[8] R. Verfürth, "Error Estimates for a Mixed Finite Element Approximation of the Stokes Equations", R.A.I.R.O., Vol. 18, pp. 175-182, 1984.

[9] J. Pitkäranta and R. Stenberg, "Error Bounds for the Approximation of the Stokes Problem Using Bilinear/Constant Elements on Irregular Quadrilateral Meshes", The Mathematics of Finite Elements and Applications V, Academic Press, pp. 325-334, 1985.

[10] K. J. Bathe and F. Brezzi, "On the Convergence of a Four-Node Plate Bending Element Based on Mindlin/Reissner Plate Theory and a Mixed Interpolation", The Mathematics of Finite Elements and Applications V, Academic Press, pp. 491-503, 1985.