

**Stabilized Finite Element Methods  
for the Stokes Problem**

Leopoldo P. Franca  
Laboratório Nacional de Computação Científica (LNCC/CNPq)  
Rua Lauro Müller 455  
22290 Rio de Janeiro, RJ - Brazil

Thomas J. R. Hughes  
Division of Applied Mechanics, Durand Building  
Stanford University  
Stanford, California 94305 - USA

Rolf Stenberg  
Faculty of Mechanical Engineering  
Helsinki University of Technology  
02150 Espoo - Finland

This is a preprint of Chapter 4 in  
*Incompressible Computational Fluid Dynamics*.  
M. Gunzburger and R.A. Nicolaides, Eds.  
Cambridge University Press 1993. pp. 87-107

## 1. Introduction

Physically the Stokes equations model "slow" flows of incompressible fluids or alternatively isotropic incompressible elastic materials. In Computational Fluid Dynamics, however, the Stokes equations have become an important model problem for designing and analyzing finite element algorithms. The reason being, that some of the problems encountered when solving the full Navier-Stokes equations are already present in the more simple Stokes equations. In particular, it gives the right setting for studying the stability problem connected with the choice of finite element spaces for the velocity and the pressure, respectively. It is well known that these spaces cannot be chosen independently when the discretization is based on the "Galerkin" variational form. This method belongs to the class of saddle-point problems for which an abstract theory has been developed by Brezzi [1974] and Babuška [1973]. The theory shows that the method is optimally convergent if the finite element spaces for velocity and pressure satisfy the "Babuška-Brezzi" or "inf-sup" condition. In computations the violation of this condition often leads to unphysical pressure oscillations and a "locking" of the velocity field, cf. Hughes [1987]. During the last decade this problem has been studied thoroughly and various velocity-pressure combinations have been shown to satisfy the Babuška-Brezzi condition. Unfortunately, however, it has turned out that many seemingly natural combinations do not satisfy it. (See Girault and Raviart [1986], Brezzi and Fortin [to appear], and references therein.)

In this chapter we will review a recent technique of "stabilizing" mixed methods. In this approach the standard Galerkin form is modified by the addition of mesh-dependent terms which are weighted residuals of the differential equations. By this technique it is possible to avoid the stability problem connected with the classical mixed methods and hence the convergence can be established for a wide family of simple interpolations.

This methodology was first used in connection with advective flows in the works of Brooks and Hughes [1982], Hughes *et al.* [1979,1989] and Johnson *et al.* [1981,1984]. Developments of these formulations to mixed methods, started in Hughes *et al.* [1986] motivated by the stabilization procedure proposed by Brezzi and Pitkäranta [1984] to the Stokes problem employing linear elements for velocity and pressure. Different from the formulation proposed by Brezzi and Pitkäranta, the work in Hughes *et al.* [1986] presented a consistent formulation which allowed the construction of higher order approximations with optimal accuracy. Later numerous variants and extensions have been proposed and analyzed, cf. Hughes and Franca [1987], Franca *et al.* [1988b], Brezzi and Douglas [1988], Karam Filho and Loula [1988], Pierre [1988,1989], Douglas and Wang [1989], Durán and

Notchitto [1989], Franca and Stenberg [1991], Silvester and Kechkar [1990], and Franca *et al.* [1990b]. The first papers Hughes *et al.* [1986] and Hughes and Franca [1987] contained an error analysis which was improved in Brezzi and Douglas [1988] and Pierre [1989]. In our paper Franca and Stenberg [1991] we presented a unified error analysis technique which can be used for all formulations.

Let us here also remark that these stabilization techniques are not restricted to problems in fluid mechanics. They can be applied to design stable finite element methods for a number of other problems in continuum mechanics, such as beams, plates and arches, see Franca and Hughes [1988a] and references therein.

The plan of the chapter is as follows. In the next section we introduce our notation and review the classical mixed formulation of the Stokes problem. In section 3 we consider stabilization techniques. First, we review our technique by giving a self-contained analysis of the method of Hughes and Franca [1987] as modified in Franca and Stenberg [1991]. This method appears particularly attractive since it preserves the symmetry of the original Stokes operator. Next, we consider the method by Douglas and Wang [1989] in which the symmetry of the discretization is abandoned in favour of better stability characteristics for higher order interpolations. Finally, we consider two methods (proposed in Franca and Stenberg [1991]) which give rise to symmetric discretizations and in addition have the same stability robustness as the method of Douglas and Wang [1989].

## 2. Preliminaries

The purpose of this chapter is to discuss the stabilization of finite element methods for the Stokes problem. Hence, we will without loss of generality consider the Stokes equations with viscosity equal to unity and with homogeneous Dirichlet boundary conditions

$$\begin{aligned} -\Delta \mathbf{u} + \nabla p &= \mathbf{f} && \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{0} && \text{on } \Gamma. \end{aligned} \tag{1}$$

Here  $\mathbf{u} = (u_1, u_2, \dots, u_N)$  is the velocity of the fluid,  $p$  is the pressure, and  $\mathbf{f}$  is the body force. The domain  $\Omega \subset \mathbb{R}^N$ ,  $N = 2$  or  $3$ , is assumed to be bounded with a polygonal or polyhedral boundary  $\Gamma$ .

Using the notation

$$B(\mathbf{w}, r; \mathbf{v}, q) = (\nabla \mathbf{w}, \nabla \mathbf{v}) - (\operatorname{div} \mathbf{v}, r) - (\operatorname{div} \mathbf{w}, q) \tag{2}$$

and

$$F(\mathbf{v}, q) = (\mathbf{f}, \mathbf{v}), \quad (3)$$

the variational formulation of the problem is the following. Given  $\mathbf{f} \in [H^{-1}(\Omega)]^N$ , find  $(\mathbf{u}, p) \in [H_0^1(\Omega)]^N \times L_0^2(\Omega)$  such that

$$B(\mathbf{u}, p; \mathbf{v}, q) = F(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in [H_0^1(\Omega)]^N \times L_0^2(\Omega). \quad (4)$$

Here and below we use standard notation (see Ciarlet [1978] for the notation not explicitly defined below). In particular, we denote by  $(\cdot, \cdot)_D$  the inner product in  $L^2(D)$ ,  $[L^2(D)]^N$  or  $[L^2(D)]^{N \times N}$  with the subscript  $D$  dropped for  $D = \Omega$ . Further, we denote the space of functions continuous on  $\Omega$  by  $C^0(\Omega)$  and

$$L_0^2(\Omega) = \left\{ q \in L^2(\Omega) \mid \int_{\Omega} q \, d\Omega = 0 \right\}.$$

$C$ ,  $C_j$ ,  $j \in \mathbb{N}$ , and  $C_I$  stand for various positive constants independent of the mesh parameter  $h$ .

REMARK. In the case of various "outflow" boundary conditions it is probably more correct to use  $(\boldsymbol{\varepsilon}(\mathbf{u}), \boldsymbol{\varepsilon}(\mathbf{v}))$  (here  $\boldsymbol{\varepsilon}(\mathbf{u})$  is the symmetric part of the velocity gradient) instead of  $(\nabla \mathbf{u}, \nabla \mathbf{v})$  in the variational and finite element formulations. For the present discussion this is an irrelevant matter since all results are trivially valid for this case as well. ■

The problem (4) is well posed, i.e. there is a positive constant  $C$  such that

$$\sup_{\substack{(\mathbf{v}, q) \in [H_0^1(\Omega)]^N \times L_0^2(\Omega) \\ (\mathbf{v}, q) \neq (\mathbf{0}, 0)}} \frac{B(\mathbf{u}, p; \mathbf{v}, q)}{\|\mathbf{v}\|_1 + \|q\|_0} \geq C(\|\mathbf{u}\|_1 + \|p\|_0) \quad \forall (\mathbf{u}, p) \in [H_0^1(\Omega)]^N \times L_0^2(\Omega). \quad (5)$$

This inequality is a simple consequence of the Poincaré inequality (or Korn's inequality when  $(\boldsymbol{\varepsilon}(\mathbf{u}), \boldsymbol{\varepsilon}(\mathbf{v}))$  is used in the variational formulation) and the condition

$$\sup_{\mathbf{0} \neq \mathbf{v} \in [H_0^1(\Omega)]^N} \frac{(\operatorname{div} \mathbf{v}, p)}{\|\mathbf{v}\|_1} \geq C\|p\|_0 \quad \forall p \in L_0^2(\Omega). \quad (6)$$

In the traditional mixed method the same variational problem is solved in some finite element subspaces. Let  $\mathbf{V}_h \subset [H_0^1(\Omega)]^N$  and  $P_h \subset L_0^2(\Omega)$  be defined as piecewise polynomials on a regular partitioning  $\mathcal{C}_h$  of  $\bar{\Omega}$  into elements consisting of triangles (tetrahedrons in  $\mathbb{R}^3$ ) or convex quadrilaterals (hexahedrons). We then get the so called GALERKIN METHOD. Find  $\mathbf{u}_h \in \mathbf{V}_h$  and  $p_h \in P_h$  such that

$$B(\mathbf{u}_h, p_h; \mathbf{v}, q) = F(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{V}_h \times P_h. \quad \blacksquare$$

In the discrete case the same steps used for proving that the Poincaré inequality and (6) implies (5) can be repeated. Hence, we have the following famous result.

PROPOSITION 2.1. (Babuška [1973], Brezzi [1974]) *If the finite element spaces  $\mathbf{V}_h$  and  $P_h$  satisfy the condition*

$$\sup_{\mathbf{0} \neq \mathbf{v} \in \mathbf{V}_h} \frac{(\operatorname{div} \mathbf{v}, p)}{\|\mathbf{v}\|_1} \geq C \|p\|_0 \quad \forall p \in P_h, \quad (7)$$

then the inequality

$$\sup_{\substack{(\mathbf{v}, q) \in \mathbf{V}_h \times P_h \\ (\mathbf{v}, q) \neq (\mathbf{0}, 0)}} \frac{B(\mathbf{u}, p; \mathbf{v}, q)}{\|\mathbf{v}\|_1 + \|q\|_0} \geq C(\|\mathbf{u}\|_1 + \|p\|_0) \quad \forall (\mathbf{u}, p) \in \mathbf{V}_h \times P_h$$

is valid. ■

It is evident that this excludes many finite element spaces and, as already noted in the introduction, many natural choices cannot be used. For a survey of methods which have been proven to satisfy this condition we refer to Girault and Raviart [1986] and Brezzi and Fortin [to appear].

In the case that Proposition 2.1 is valid, the method will converge optimally. If piecewise polynomials of degree  $k$  and  $l$  are used for the velocity and pressure, respectively, then we have the following error estimate

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 \leq C(h^k |\mathbf{u}|_{k+1} + h^{l+1} |p|_{l+1}),$$

provided that  $\mathbf{u} \in [H^{k+1}(\Omega)]^N$  and  $p \in H^{l+1}(\Omega)$ .

### 3. Stabilized methods

Let us consider approximation by Lagrange elements which are without doubt the most popular in practice. For convenience we will adopt the following notation

$$R_m(K) = \begin{cases} P_m(K) & \text{if } K \text{ is a triangle or tetrahedron,} \\ Q_m(K) & \text{if } K \text{ is a quadrilateral or hexahedron,} \end{cases}$$

where for each integer  $m \geq 0$ ,  $P_m(K)$  and  $Q_m(K)$  are the usual polynomial spaces on  $K$  (see Ciarlet [1978]).

The finite element spaces  $\mathbf{V}_h$  and  $P_h$  for approximating the velocity and the pressure, respectively, are then defined as

$$\mathbf{V}_h = \{\mathbf{v} \in [H_0^1(\Omega)]^N \mid \mathbf{v}|_K \in [R_k(K)]^N \quad \forall K \in \mathcal{C}_h\}, \quad (8)$$

$$P_h = \{p \in \mathcal{C}^0(\Omega) \cap L_0^2(\Omega) \mid p|_K \in R_l(K) \ \forall K \in \mathcal{C}_h\}, \quad (9a)$$

or

$$P_h = \{p \in L_0^2(\Omega) \mid p|_K \in R_l(K) \ \forall K \in \mathcal{C}_h\}. \quad (9b)$$

Here we have two alternatives for the pressure depending on if it is approximated continuously or not. Let us here also remark that for a two-dimensional domain one could mix triangles and quadrilaterals.

For stating our results it is convenient to use the following notation

$$S(K) = \begin{cases} P_N(K) & \text{if } K \text{ is a triangle or tetrahedron,} \\ Q_2(K) & \text{if } K \text{ is a quadrilateral or hexahedron,} \end{cases}$$

and

$$\mathbf{S}_h = \{ \mathbf{v} \in [H_0^1(\Omega)]^N \mid \mathbf{v}|_K \in [S(K)]^N \ \forall K \in \mathcal{C}_h \}. \quad (10)$$

Next, we recall that since  $\mathbf{V}_h$  is assumed to consist of piecewise polynomials, a simple scaling argument shows that there is a positive constant  $C_I$  such that

$$C_I \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{v}\|_{0,K}^2 \leq \|\nabla \mathbf{v}\|_0^2 \quad \forall \mathbf{v} \in \mathbf{V}_h. \quad (11)$$

Now let us define

METHOD I (Hughes and Franca [1987], Franca and Stenberg [1991]). Find  $\mathbf{u}_h \in \mathbf{V}_h$  and  $p_h \in P_h$  such that

$$B_h(\mathbf{u}_h, p_h; \mathbf{v}, q) = F_h(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{V}_h \times P_h,$$

with

$$B_h(\mathbf{w}, r; \mathbf{v}, q) = B(\mathbf{w}, r; \mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{w} + \nabla r, -\Delta \mathbf{v} + \nabla q)_K$$

and

$$F_h(\mathbf{v}, q) = F(\mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, -\Delta \mathbf{v} + \nabla q)_K. \quad \blacksquare$$

The first observation concerning this method (and all the methods to follow) is that it is consistent, i.e. if we (e.g.) assume that  $\mathbf{f} \in [L^2(\Omega)]^N$ , then the exact solution  $(\mathbf{u}, p)$  satisfies the discrete equation

$$B_h(\mathbf{u}, p; \mathbf{v}, q) = F_h(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{V}_h \times P_h. \quad (12)$$

The next result, when compared with Proposition 2.1, shows the superiority of this formulation compared with the Galerkin method.

**THEOREM 3.1.** *Assume that either  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset \mathcal{C}^0(\Omega)$  and that  $0 < \alpha < C_I$ . Then the bilinear form of Method I satisfies*

$$\sup_{\substack{(\mathbf{v}, q) \in \mathbf{V}_h \times P_h \\ (\mathbf{v}, q) \neq (\mathbf{0}, 0)}} \frac{B_h(\mathbf{u}, p; \mathbf{v}, q)}{\|\mathbf{v}\|_1 + \|q\|_0} \geq C(\|\mathbf{u}\|_1 + \|p\|_0) \quad \forall (\mathbf{u}, p) \in \mathbf{V}_h \times P_h. \quad \blacksquare$$

Before going into the details of proving this result, let us note that it implies the following optimal error estimate.

**THEOREM 3.2.** *Assume that either  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset \mathcal{C}^0(\Omega)$  and that  $0 < \alpha < C_I$ . For the approximate solution obtained with Method I we then have*

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 \leq C(h^k |\mathbf{u}|_{k+1} + h^{l+1} |p|_{l+1}),$$

provided the exact solution satisfies  $\mathbf{u} \in [H^{k+1}(\Omega)]^N$  and  $p \in H^{l+1}(\Omega)$ .

*Proof:* Let  $\tilde{\mathbf{u}} \in \mathbf{V}_h$  be the interpolant of  $\mathbf{u}$  and let  $\tilde{p} \in P_h$  be the interpolant of  $p$ . Theorem 3.1 now implies the existence of  $(\mathbf{v}, q) \in \mathbf{V}_h \times P_h$  such that

$$\|\mathbf{v}\|_1 + \|q\|_0 \leq C \tag{13}$$

and

$$\|\tilde{\mathbf{u}} - \mathbf{u}_h\|_1 + \|\tilde{p} - p_h\|_0 \leq B_h(\mathbf{u}_h - \tilde{\mathbf{u}}, p_h - \tilde{p}; \mathbf{v}, q). \tag{14}$$

The consistency (12) now gives

$$B_h(\mathbf{u}_h - \tilde{\mathbf{u}}, p_h - \tilde{p}; \mathbf{v}, q) = B_h(\mathbf{u} - \tilde{\mathbf{u}}, p - \tilde{p}; \mathbf{v}, q). \tag{15}$$

For the right hand side above Schwarz inequality gives

$$\begin{aligned} & B_h(\mathbf{u} - \tilde{\mathbf{u}}, p - \tilde{p}; \mathbf{v}, q) \\ & \leq C(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta(\mathbf{u} - \tilde{\mathbf{u}})\|_{0,K}^2 + \|p - \tilde{p}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla(p - \tilde{p})\|_{0,K}^2)^{1/2} \\ & \quad \cdot (\|\mathbf{v}\|_1^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{v}\|_{0,K}^2 + \|q\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla q\|_{0,K}^2)^{1/2}. \end{aligned} \tag{16}$$

Hence, the triangle inequality, the inverse inequalities (11) and

$$\left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla q\|_{0,K}^2 \right)^{1/2} \leq C \|q\|_0,$$

together with (13) to (16) give

$$\begin{aligned} \|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 &\leq C(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + (\sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta(\mathbf{u} - \tilde{\mathbf{u}})\|_{0,K}^2)^{1/2}) \\ &\quad + \|p - \tilde{p}\|_0 + (\sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla(p - \tilde{p})\|_{0,K}^2)^{1/2}. \end{aligned}$$

The following interpolation estimate is standard

$$\begin{aligned} \|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + (\sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta(\mathbf{u} - \tilde{\mathbf{u}})\|_{0,K}^2)^{1/2} + \|p - \tilde{p}\|_0 + (\sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla(p - \tilde{p})\|_{0,K}^2)^{1/2} \\ \leq C(h^k |\mathbf{u}|_{k+1} + h^{l+1} |p|_{l+1}). \end{aligned}$$

and the theorem is thus proved. ■

REMARKS. 1) With the regularity assumption (valid in the two-dimensional case for a convex polygon  $\Omega$ )

$$\|\mathbf{u}\|_2 + \|p\|_1 \leq C\|\mathbf{f}\|_0 \quad (17)$$

the usual duality argument gives the optimal  $L^2$ -estimate (cf. Franca and Stenberg [1991])

$$\|\mathbf{u} - \mathbf{u}_h\|_0 \leq C(h^{k+1} |\mathbf{u}|_{k+1} + h^{l+2} |p|_{l+1}).$$

2) It is important to note that for piecewise linear approximations to the velocity the bilinear and linear form reduce to

$$B_h(\mathbf{w}, r; \mathbf{v}, q) = B(\mathbf{w}, r; \mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\nabla r, \nabla q)_K$$

and

$$F_h(\mathbf{v}, q) = F(\mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \nabla q)_K,$$

and in this case no upper limit has to be imposed on  $\alpha$ . If we additionally make the simplification  $F_h(\mathbf{v}, q) = F(\mathbf{v}, q)$ , we get the method of Brezzi and Pitkäranta [1984]. Let us also remark that for isoparametric bilinear (in  $\mathbb{R}^2$ ) or trilinear (in  $\mathbb{R}^3$ )  $B_h$  and  $F_h$  can be defined as above without a loss of accuracy.

3) For continuous linear approximations (on triangles or tetrahedrons) for both the velocity and pressure it was noted in Pierre [1989] that there is a close connection between this method and a well known Galerkin method, the MINI element of Arnold, Brezzi and Fortin (cf. Girault and Raviart [1986] or Brezzi and Fortin [to appear]). In the MINI element the

stability is achieved by adding local "bubble functions" to the velocity space. When these bubble degrees of freedom are condensed, one gets the above stabilized method with a particular choice for the stabilization parameter  $\alpha$ , see Pierre [1989] for the details.

4) From the error estimate above it is clear that the optimal choice for the finite element spaces is obtained with  $l = k - 1$ . We have, however, preferred not to do this choice when presenting our result, since we want to emphasize how arbitrary one can choose the spaces and still obtain a stable method. Also, for the full Navier-Stokes equations it might be good to use other combinations than  $l = k - 1$  (such as equal order approximations). ■

Next, let us review the technique given in Franca and Stenberg [1991] for verifying the stability. The following lemma turns out to be very useful for analyzing both traditional Galerkin methods (cf. Stenberg [1990]) and the present stabilized methods. For Galerkin methods this idea was first used by Verfürth [1984] for the analysis of "Taylor-Hood" methods, i.e. methods with a continuous pressure approximation.

LEMMA 3.1. *Suppose that either one of the assumptions  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset \mathcal{C}^0(\Omega)$  is valid. Then there exist positive constants  $C_1$  and  $C_2$  such that*

$$\sup_{\mathbf{0} \neq \mathbf{v} \in \mathbf{V}_h} \frac{(\operatorname{div} \mathbf{v}, p)}{\|\mathbf{v}\|_1} \geq C_1 \|p\|_0 - C_2 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \quad \forall p \in P_h.$$

*Proof:* Consider first the case  $\mathbf{S}_h \subset \mathbf{V}_h$ . Denote by  $\Pi_h$  the  $L^2$ -projection onto the space of piecewise constants, i.e.

$$\Pi_h q|_K = \frac{1}{\operatorname{area}(K)} \int_K q \, d\Omega, \quad \forall K \in \mathcal{C}_h.$$

Since  $\mathbf{S}_h \subset \mathbf{V}_h$  the pair  $(\mathbf{V}_h, \Pi_h P_h)$  satisfies the stability inequality (7) (see e.g. Girault and Raviart [1986]). Hence there is a constant  $C_1$  such that for every  $p \in P_h$  there exists  $\mathbf{v} \in \mathbf{V}_h$ , with  $\|\mathbf{v}\|_1 = 1$ , such that

$$(\operatorname{div} \mathbf{v}, \Pi_h p) \geq C_1 \|\Pi_h p\|_0.$$

Using the interpolation estimate

$$\|(I - \Pi_h)p\|_{0,K} \leq C_3 h_K \|\nabla p\|_{0,K}, \quad \forall K \in \mathcal{C}_h,$$

we now get

$$\begin{aligned} (\operatorname{div} \mathbf{v}, p) &= (\operatorname{div} \mathbf{v}, \Pi_h p) + (\operatorname{div} \mathbf{v}, (I - \Pi_h)p) \\ &\geq C_1 \|\Pi_h p\|_0 - \|\mathbf{v}\|_1 \|(I - \Pi_h)p\|_0 \\ &= C_1 \|\Pi_h p\|_0 - \|(I - \Pi_h)p\|_0 \\ &\geq C_1 \|p\|_0 - (1 + C_1) \|(I - \Pi_h)p\|_0 \\ &\geq C_1 \|p\|_0 - (1 + C_1) C_3 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \end{aligned}$$

and the asserted estimate with the first assumption is proved.

Next, let us consider the case  $P_h \subset \mathcal{C}^0(\Omega)$ . Since  $p \in P_h \subset L_0^2(\Omega)$ , by the continuous version of the "inf-sup" condition, i.e (6), there is a non-vanishing  $\mathbf{w} \in [H_0^1(\Omega)]^N$  such that

$$(\operatorname{div} \mathbf{w}, p) \geq C_4 \|p\|_0 \|\mathbf{w}\|_1.$$

Further, one can show (e.g. Girault and Raviart [1986], pp. 109-111) that there is an interpolant  $\tilde{\mathbf{w}} \in \mathbf{V}_h$  to  $\mathbf{w}$  such that

$$\sum_{K \in \mathcal{C}_h} h_K^{-2} \|\mathbf{w} - \tilde{\mathbf{w}}\|_{0,K}^2 \leq C_5 \|\mathbf{w}\|_1$$

and

$$\|\tilde{\mathbf{w}}\|_1 \leq C_6 \|\mathbf{w}\|_1. \quad (18)$$

Integrating by parts on each  $K \in \mathcal{C}_h$ , and using the above estimates we get

$$\begin{aligned} (\operatorname{div} \tilde{\mathbf{w}}, p) &= (\operatorname{div} (\tilde{\mathbf{w}} - \mathbf{w}), p) + (\operatorname{div} \mathbf{w}, p) \\ &\geq (\operatorname{div} (\tilde{\mathbf{w}} - \mathbf{w}), p) + C_4 \|\mathbf{w}\|_1 \|p\|_0 \\ &= \sum_{K \in \mathcal{C}_h} (\mathbf{w} - \tilde{\mathbf{w}}, \nabla p)_K + C_4 \|\mathbf{w}\|_1 \|p\|_0 \\ &\geq - \left( \sum_{K \in \mathcal{C}_h} h_K^{-2} \|\mathbf{w} - \tilde{\mathbf{w}}\|_{0,K}^2 \right)^{1/2} \cdot \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} + C_4 \|\mathbf{w}\|_1 \|p\|_0 \\ &\geq \left\{ -C_5 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right) + C_4 \|p\|_0 \right\} \|\mathbf{w}\|_1. \end{aligned}$$

Dividing by  $\|\mathbf{w}\|_1$  we obtain

$$\frac{(\operatorname{div} \tilde{\mathbf{w}}, p)}{\|\mathbf{w}\|_1} \geq C_4 \|p\|_0 - C_5 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2}$$

and (18) then gives

$$\frac{(\operatorname{div} \tilde{\mathbf{w}}, p)}{\|\tilde{\mathbf{w}}\|_1} \geq \frac{(\operatorname{div} \tilde{\mathbf{w}}, p)}{C_6 \|\mathbf{w}\|_1} \geq C_6^{-1} \left\{ C_4 \|p\|_0 - C_5 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \right\}.$$

The assertion is thus proved. ■

Let us now use this lemma for the

*Proof of Theorem 3.1.* Let  $(\mathbf{u}, p) \in \mathbf{V}_h \times P_h$  be given and let  $\mathbf{w} \in \mathbf{V}_h$  be a function for which the supremum of Lemma 3.1 is obtained. We scale  $\mathbf{w}$  so that  $\|\mathbf{w}\|_1 = \|p\|_0$ . Using the stability estimate of Lemma 3.1, the inverse inequality (11) and the arithmetic-geometric mean inequality we get

$$\begin{aligned}
& B_h(\mathbf{u}, p; -\mathbf{w}, 0) \\
&= -(\nabla \mathbf{u}, \nabla \mathbf{w}) + (\operatorname{div} \mathbf{w}, p) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{u} + \nabla p, \Delta \mathbf{w})_K \\
&\geq -\|\nabla \mathbf{u}\|_0 \|\nabla \mathbf{w}\|_0 + C_1 \|p\|_0^2 - C_2 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \|p\|_0 \\
&\quad + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\Delta \mathbf{u}, \Delta \mathbf{w})_K - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\nabla p, \Delta \mathbf{w})_K \\
&\geq -\|\nabla \mathbf{u}\|_0 \|\nabla \mathbf{w}\|_0 + C_1 \|p\|_0^2 - C_2 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \|p\|_0 \\
&\quad - \alpha \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{u}\|_{0,K}^2 \right)^{1/2} \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{w}\|_{0,K}^2 \right)^{1/2} \\
&\quad - \alpha \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{w}\|_{0,K}^2 \right)^{1/2} \tag{19} \\
&\geq -\|\nabla \mathbf{u}\|_0 \|\nabla \mathbf{w}\|_0 + C_1 \|p\|_0^2 - C_2 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \|p\|_0 \\
&\quad - \alpha C_I^{-1} \|\nabla \mathbf{u}\|_0 \|\nabla \mathbf{w}\|_0 - \alpha C_I^{-1/2} \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K} \right)^{1/2} \|\nabla \mathbf{w}\|_0 \\
&\geq -C_3 \|\nabla \mathbf{u}\|_0 \|\nabla \mathbf{w}\|_0 + C_1 \|p\|_0^2 - C_4 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \|p\|_0 \\
&\geq -C_3 \|\nabla \mathbf{u}\|_0 \|p\|_0 + C_1 \|p\|_0^2 - C_4 \left( \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \right)^{1/2} \|p\|_0 \\
&\geq -\frac{C_3}{2\epsilon} \|\nabla \mathbf{u}\|_0^2 + \left( C_1 - \frac{\epsilon}{2} (C_3 + C_4) \right) \|p\|_0^2 - \frac{C_4}{2\epsilon} \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq -C_5 \|\nabla \mathbf{u}\|_0^2 + C_6 \|p\|_0^2 - C_7 \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2,
\end{aligned}$$

when choosing  $0 < \epsilon < 2C_1(C_3 + C_4)^{-1}$ . Next, we note that Poincaré's inequality and the

assumption  $0 < \alpha < C_I$  give

$$\begin{aligned}
B_h(\mathbf{u}, p; \mathbf{u}, -p) &= \|\nabla \mathbf{u}\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{u}\|_{0,K}^2 \\
&\geq (1 - \alpha C_I^{-1}) \|\nabla \mathbf{u}\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq C_8 (\|\mathbf{u}\|_1^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2),
\end{aligned} \tag{20}$$

Denote  $(\mathbf{v}, q) = (\mathbf{u} - \delta \mathbf{w}, -p)$ . Combining (19) and (20) gives

$$\begin{aligned}
B_h(\mathbf{u}, p; \mathbf{v}, q) &= B_h(\mathbf{u}, p; \mathbf{u} - \delta \mathbf{w}, -p) \\
&= B_h(\mathbf{u}, p; \mathbf{u}, -p) + \delta B_h(\mathbf{u}, p; -\mathbf{w}, 0) \\
&\geq (C_8 - \delta C_5) \|\nabla \mathbf{u}\|_0^2 + \delta C_6 \|p\|_0^2 + (C_8 - \delta C_7) \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq C (\|\nabla \mathbf{u}\|_0^2 + \|p\|_0^2)
\end{aligned} \tag{21}$$

when choosing  $0 < \delta < \min\{C_8 C_5^{-1}, C_8 C_7^{-1}\}$ . On the other hand we have

$$\begin{aligned}
\|\nabla \mathbf{v}\|_0 + \|q\|_0 &\leq \|\nabla \mathbf{u}\|_0 + \delta \|\nabla \mathbf{w}\|_0 + \|p\|_0 \\
&\leq C (\|\nabla \mathbf{u}\|_0 + \|p\|_0),
\end{aligned}$$

which combined with (21) (and the Poincaré inequality) proves the stability estimate. ■

For continuous approximations for the pressure the above method was introduced in Hughes and Franca [1987]. For discontinuous pressures, however, the original formulation contained "jump terms" on the pressure, i.e. the bilinear form was defined as

$$B_h(\mathbf{w}, r; \mathbf{v}, q) = B(\mathbf{w}, r; \mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{w} + \nabla r, -\Delta \mathbf{v} + \nabla q)_K - \beta \sum_{T \in \Gamma_h} h_T ([r], [q])_T,$$

where  $\Gamma_h$  denotes the collection of all element interfaces,  $[[s]]$  is the jump in  $s$  along the interface and  $h_T$  is the length or diameter of  $T$ .  $\beta$  is a positive parameter. With these jump terms there does not seem to be any reason to use a discontinuous pressures instead of continuous one. (Some arguments in favour of a modified jump formulation are, however, given in Silvester and Kechkar [1990].) Furthermore, it leads to a nonstandard assembly procedure. In Franca and Stenberg [1991] we, however, showed that with the assumption  $\mathbf{S}_h \subset \mathbf{V}_h$  these additional jump terms above are unnecessary. This has a considerable practical significance since it allows the pressure to be eliminated at the element level by

the penalty technique. Alternatively, one can use the augmented Lagrangian technique, see e.g. Brezzi and Fortin [to appear]. Here we will briefly review the penalty technique. We replace the bilinear form with

$$B_h^\epsilon(\mathbf{w}, r; \mathbf{v}, q) = B(\mathbf{w}, r; \mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{w} + \nabla r, -\Delta \mathbf{v} + \nabla q)_K - \epsilon(r, q),$$

and solve the problem

$$B_h^\epsilon(\mathbf{u}_h^\epsilon, p_h^\epsilon; \mathbf{v}, q) = F_h(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{V}_h \times P_h. \quad (22)$$

Now it is straightforward to show that if the original method is stable then the error estimates do not change if  $0 \leq \epsilon \leq Ch^s$ ,  $s = \min\{k+1, l+2\}$ , i.e. we have

$$\|\mathbf{u} - \mathbf{u}_h^\epsilon\|_1 + \|p - p_h^\epsilon\|_0 \leq C(h^k |\mathbf{u}|_{k+1} + h^{l+1} |p|_{l+1})$$

and also

$$\|\mathbf{u} - \mathbf{u}_h^\epsilon\|_0 \leq C(h^{k+1} |\mathbf{u}|_{k+1} + h^{l+2} |p|_{l+1})$$

when (17) holds.

Let now  $\{\phi_i\}_{i=1}^{N_u}$  and  $\{\psi_i\}_{i=1}^{N_p}$  be the basis for  $\mathbf{V}_h$  and  $P_h$ , respectively. The discretization (22) then leads to the following matrix equation

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ \mathbf{P} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{pmatrix}, \quad (23)$$

where  $\mathbf{U}$  and  $\mathbf{P}$  are the degrees of freedom for  $\mathbf{u}_h^\epsilon$  and  $p_h^\epsilon$ , respectively, and the matrices are given by

$$\begin{aligned} (\mathbf{A})_{ij} &= (\nabla \phi_i, \nabla \phi_j) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\Delta \phi_i, \Delta \phi_j)_K, \\ (\mathbf{B})_{ij} &= -(\operatorname{div} \phi_j, \psi_i) + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\Delta \phi_j, \nabla \psi_i)_K, \\ (\mathbf{C})_{ij} &= -\epsilon(\psi_i, \psi_j) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\nabla \psi_i, \nabla \psi_j)_K, \\ (\mathbf{F}_1)_i &= (\mathbf{f}, \phi_i) + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \Delta \phi_i)_K, \\ (\mathbf{F}_2)_i &= -\alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \nabla \psi_i)_K. \end{aligned}$$

Now, since  $\epsilon > 0$  the matrix  $\mathbf{C}$  is negatively definite and it can be inverted on each element separately (we have assumed that the pressure approximation is discontinuous). By eliminating  $\mathbf{P}$ , we get the following system for  $\mathbf{U}$  alone

$$(\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}^T) \mathbf{U} = \mathbf{F}_1 - \mathbf{BC}^{-1}\mathbf{F}_2.$$

Due to the assumption  $0 < \alpha < C_I$  the matrix  $\mathbf{A}$  is positively definite and since  $\mathbf{C}$  is negatively definite, the coefficient matrix above is positively definite.

We note here that for a continuous pressure or a discontinuous pressure with the "jump terms" of Hughes and Franca [1987] we of course obtain a discrete system of the form (23), but in those cases the matrix  $\mathbf{C}$  cannot be eliminated on each element separately.

A problem with the stabilized formulation presented above is the choice of the parameter  $\alpha$ . The most straight-forward way to get some insight into the dependency of  $\alpha$  is to perform a large number of calculations with different values of  $\alpha$  for various problems. In this respect we refer to Hughes *et al.* [1986], Franca *et al.* [1988b], Karam Filho and Loula [1988], Pierre [1988] and Franca *et al.* [1990a], where the results of some tests of this kind are reported for various methods. Alternatively, one could try to directly get some reliable estimates for the constant  $C_I$  in (11) which is the upper limit for  $\alpha$ . Some result in this direction can be found in Harari [to appear].

However, there are some stabilized formulations for which no upper bounds for the stability parameters are required. Of these we will first consider the following alternative. METHOD II (Douglas and Wang [1989]). Find  $\mathbf{u}_h \in \mathbf{V}_h$  and  $p_h \in P_h$  such that

$$B_h(\mathbf{u}_h, p_h; \mathbf{v}, q) = F_h(\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{V}_h \times P_h$$

with

$$B_h(\mathbf{w}, r; \mathbf{v}, q) = B(\mathbf{w}, r; \mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{w} + \nabla r, \Delta \mathbf{v} + \nabla q)_K$$

and

$$F_h(\mathbf{v}, q) = F(\mathbf{v}, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \Delta \mathbf{v} + \nabla q)_K. \quad \blacksquare$$

We note that the only difference to the previous formulation is that the sign in front of  $\Delta \mathbf{v}$  has been changed. This has, however, as consequence that the method is stable and optimally convergent for all positive values of  $\alpha$ .

**THEOREM 3.3.** Assume that either  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset \mathcal{C}^0(\Omega)$  and that  $\alpha > 0$ . For Method II we then have the following error estimate

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 \leq C(h^k |\mathbf{u}|_{k+1} + h^{l+1} |p|_{l+1}).$$

*Proof:* The analysis differs from that of the earlier method only with respect to the verification of the stability. To this end we let  $\gamma > 1$  be a parameter and estimate as follows

$$\begin{aligned}
B_h(\mathbf{u}, p; \mathbf{u}, -p) &= \|\nabla \mathbf{u}\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|-\Delta \mathbf{u} + \nabla p\|_{0,K}^2 \\
&= \|\nabla \mathbf{u}\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{u}\|_{0,K}^2 \\
&\quad + 2\alpha \sum_{K \in \mathcal{C}_h} h_K^2 (-\Delta \mathbf{u}, \nabla p)_K + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq \|\nabla \mathbf{u}\|_0^2 + \alpha(1 - \gamma) \sum_{K \in \mathcal{C}_h} h_K^2 \|\Delta \mathbf{u}\|_{0,K}^2 \\
&\quad + \left(1 - \frac{1}{\gamma}\right) \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq (1 + \alpha(1 - \gamma)C_I^{-1}) \|\nabla \mathbf{u}\|_0^2 + \left(1 - \frac{1}{\gamma}\right) \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq C_2 \|\nabla \mathbf{u}\|_0^2 + C_3 \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2 \\
&\geq C(\|\nabla \mathbf{u}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0,K}^2)
\end{aligned}$$

when we choose  $1 < \gamma < (1 + C_I \alpha^{-1})$ . With the use of Lemma 3.1 the stability is now proved as in the proof of Theorem 3.1. The error estimate then follows as in Theorem 3.2. ■

REMARK. In the case of a discontinuous pressure the original proposal of Douglas and Wang contained jump terms for the pressure. In Franca and Stenberg [1991] we showed that with the assumption  $\mathbf{S}_h \subset \mathbf{V}_h$  these are unnecessary. ■

The theoretical results proved for the two methods, suggest that the method of Douglas and Wang might be more "robust", by which we mean that the accuracy of the finite element solution is less sensitive to the choice of  $\alpha$  (except for linear velocities for which the methods coincide). The numerical results reported in Franca *et al.* [1990a] confirms this. On the other hand, the method of Douglas and Wang method does not give a symmetric system of equations.

However, it is possible to get methods with symmetric discrete system and without any upper bounds on the stability parameter. We rewrite the Stokes problem with  $\nabla \mathbf{u}$  as a new unknown, "the augmented stress"  $\boldsymbol{\sigma}$ :

$$\begin{aligned}
\boldsymbol{\sigma} - \nabla \mathbf{u} &= \mathbf{0} && \text{in } \Omega, \\
\operatorname{div} \mathbf{u} &= 0 && \text{in } \Omega, \\
-\mathbf{div} \boldsymbol{\sigma} + \nabla p &= \mathbf{f} && \text{in } \Omega, \\
\mathbf{u} &= \mathbf{0} && \text{on } \Gamma.
\end{aligned}$$

Here  $\mathbf{div}$  denotes the vector divergence applied to matrix functions.

The velocity and the pressure we approximate as before with the spaces (8) and (9), respectively, and for the augmented stress we introduce the space

$$\boldsymbol{\Sigma}_h = \{ \boldsymbol{\tau} \in [L^2(\Omega)]^{N \times N} \mid \boldsymbol{\tau}|_K \in [R_m(K)]^{N \times N} \ \forall K \in \mathcal{C}_h \}, \quad (24)$$

with  $m = k - 1$  for triangular and tetrahedral elements and  $m = k$  for quadrilaterals and hexahedrons.

REMARK. When  $\boldsymbol{\varepsilon}(\mathbf{u})$  is used in the formulation instead of  $\nabla \mathbf{u}$  then  $\boldsymbol{\sigma}$  is symmetric and  $\boldsymbol{\Sigma}_h$  can be reduced to consist of symmetric matrices. ■

Let us define the method directly in penalty form.

METHOD III (Franca and Stenberg [1991]). Find  $(\boldsymbol{\sigma}_h, p_h, \mathbf{u}_h) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h$  such that

$$B_h(\boldsymbol{\sigma}_h, p_h, \mathbf{u}_h; \boldsymbol{\tau}, q, \mathbf{v}) = F_h(\boldsymbol{\tau}, q, \mathbf{v}) \quad \forall (\boldsymbol{\tau}, q, \mathbf{v}) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h,$$

with

$$\begin{aligned}
B_h(\boldsymbol{\kappa}, r, \mathbf{w}; \boldsymbol{\tau}, q, \mathbf{v}) &= -(\boldsymbol{\kappa}, \boldsymbol{\tau}) + (\nabla \mathbf{w}, \boldsymbol{\tau}) + (\boldsymbol{\kappa}, \nabla \mathbf{v}) - (r, \operatorname{div} \mathbf{v}) - (\operatorname{div} \mathbf{w}, q) - \epsilon(r, q) \\
&\quad - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{div} \boldsymbol{\kappa} - \nabla r, \mathbf{div} \boldsymbol{\tau} - \nabla q)_K
\end{aligned}$$

and

$$F_h(\boldsymbol{\tau}, q, \mathbf{v}) = (\mathbf{f}, \mathbf{v}) + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \mathbf{div} \boldsymbol{\tau} - \nabla q)_K. \quad \blacksquare$$

For the method we have the optimal estimate.

THEOREM 3.4. Assume that either  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset C^0(\Omega)$  and that  $\alpha > 0$ . For the approximation with Method III we then have the following error estimate

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|p - p_h\|_0 + \|\mathbf{u} - \mathbf{u}_h\|_1 \leq C(h^{m+1} |\boldsymbol{\sigma}|_{m+1} + h^{l+1} |p|_{l+1} + h^k |\mathbf{u}|_{k+1}),$$

for all  $\epsilon$  in the range  $0 \leq \epsilon \leq Ch^s$ ,  $s = \min\{k + 1, l + 2\}$ . ■

*Proof:* We will not give the proof in full detail. We first note that the method is consistent. Hence, in analogy with the earlier methods we have to verify the stability condition which now is

$$\sup_{\substack{(\boldsymbol{\tau}, q, \mathbf{v}) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h \\ (\boldsymbol{\tau}, q, \mathbf{v}) \neq (\mathbf{0}, 0, \mathbf{0})}} \frac{B_h(\boldsymbol{\sigma}, p, \mathbf{u}; \boldsymbol{\tau}, q, \mathbf{v})}{\|\boldsymbol{\tau}\|_0 + \|q\|_0 + \|\mathbf{v}\|_1} \geq C(\|\boldsymbol{\sigma}\|_0 + \|p\|_0 + \|\mathbf{u}\|_1) \quad \forall (\boldsymbol{\sigma}, p, \mathbf{u}) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h.$$

To prove this we first use an inverse estimate similar to (11) and estimate as in the proof of Theorem 3.3 in order to get

$$\begin{aligned} B_h(\boldsymbol{\sigma}, p, \mathbf{u}; -\boldsymbol{\sigma}, -p, \mathbf{u}) &= \|\boldsymbol{\sigma}\|_0^2 + \epsilon \|p\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\mathbf{div} \boldsymbol{\sigma} - \nabla p\|_{0_K}^2 \\ &\geq C_1(\|\boldsymbol{\sigma}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0_K}^2) + \epsilon \|p\|_0^2. \end{aligned} \quad (25)$$

The second step is to use Lemma 3.1 in the same way as earlier and conclude that there is a velocity  $\mathbf{w} \in \mathbf{V}_h$ , with  $\|\mathbf{w}\|_1 \leq \|p\|_0$ , such that

$$B_h(\boldsymbol{\sigma}, p, \mathbf{u}; \mathbf{0}, 0, -\mathbf{w}) \geq C_2 \|p\|_0^2 - C_3(\|\boldsymbol{\sigma}\|_0^2 + \|\nabla \mathbf{u}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0_K}^2). \quad (26)$$

Next, the assumption that  $m = k - 1$  for triangles (tetrahedrons) and  $m = k$  for quadrilaterals (hexahedrons) implies that there is  $\boldsymbol{\kappa} \in \boldsymbol{\Sigma}_h$  such that

$$(\boldsymbol{\kappa}, \nabla \mathbf{u}) = \|\nabla \mathbf{u}\|_0^2 \quad (27a)$$

and

$$\|\boldsymbol{\kappa}\| \leq C \|\nabla \mathbf{u}\|_0. \quad (27b)$$

This gives

$$B_h(\boldsymbol{\sigma}, p, \mathbf{u}; \boldsymbol{\kappa}, 0, \mathbf{0}) \geq C_4 \|\nabla \mathbf{u}\|_0^2 - C_5(\|\boldsymbol{\sigma}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\nabla p\|_{0_K}^2). \quad (28)$$

The stability estimate is now obtained from (25), (26) and (28) by taking

$$(\boldsymbol{\tau}, q, \mathbf{v}) = (-\boldsymbol{\sigma} + \delta \boldsymbol{\kappa}, -p, \mathbf{u} - \delta^2 \mathbf{w})$$

with  $\delta$  positive and small enough. ■

Let us next discuss the implementation of the method. Let  $\{\chi_i\}_{i=1}^{N_\sigma}$ ,  $\{\phi_i\}_{i=1}^{N_u}$  and  $\{\psi_i\}_{i=1}^{N_p}$  be the basis for  $\Sigma_h$ ,  $\mathbf{V}_h$  and  $P_h$ , respectively. The discretization then leads to the following matrix equation

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \mathbf{B}^T & \mathbf{0} & \mathbf{D} \\ \mathbf{C}^T & \mathbf{D}^T & \mathbf{E} \end{pmatrix} \begin{pmatrix} \mathbf{S} \\ \mathbf{U} \\ \mathbf{P} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \mathbf{F}_3 \end{pmatrix}, \quad (29)$$

where  $\mathbf{S}$ ,  $\mathbf{U}$  and  $\mathbf{P}$  are the degrees of freedom for  $\sigma_h$ ,  $\mathbf{u}_h$  and  $p_h$ , respectively. The matrices are given by

$$\begin{aligned} (\mathbf{A})_{ij} &= -(\chi_i, \chi_j) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{div} \chi_i, \mathbf{div} \chi_j)_K, \\ (\mathbf{B})_{ij} &= (\nabla \phi_j, \chi_i), \\ (\mathbf{C})_{ij} &= \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\nabla \psi_j, \mathbf{div} \chi_i)_K, \\ (\mathbf{D})_{ij} &= -(\psi_j, \mathbf{div} \phi_i), \\ (\mathbf{E})_{ij} &= -\epsilon (\psi_i, \psi_j) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\nabla \psi_i, \nabla \psi_j)_K \end{aligned}$$

and

$$\begin{aligned} (\mathbf{F}_1)_i &= \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \mathbf{div} \chi_i)_K, \\ (\mathbf{F}_2)_i &= (\mathbf{f}, \phi_i), \\ (\mathbf{F}_3)_i &= -\alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\mathbf{f}, \nabla \psi_i)_K. \end{aligned}$$

Now, since the matrix  $\mathbf{A}$  is negatively definite and  $\Sigma_h$  consists of discontinuous functions,  $\mathbf{S}$  can be condensed in the assembling phase of the calculations. It is also important to note that in the condensation on one element each row of the matrix unknown  $\mathbf{S}$  can be eliminated separately and the matrices to invert are identical for all rows. Hence, this condensation is not as expensive as one might think at a first glance. This leads to the system

$$\begin{pmatrix} \widehat{\mathbf{A}} & \widehat{\mathbf{B}} \\ \widehat{\mathbf{B}}^T & \widehat{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ \mathbf{P} \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{F}}_1 \\ \widehat{\mathbf{F}}_2 \end{pmatrix},$$

with

$$\begin{aligned} \widehat{\mathbf{A}} &= -\mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}, \\ \widehat{\mathbf{B}} &= \mathbf{D} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{C}, \\ \widehat{\mathbf{C}} &= \mathbf{E} - \mathbf{C}^T \mathbf{A}^{-1} \mathbf{C} \end{aligned}$$

and

$$\begin{aligned}\widehat{\mathbf{F}}_1 &= \mathbf{F}_2 - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{F}_1, \\ \widehat{\mathbf{F}}_2 &= \mathbf{F}_3 - \mathbf{C}^T \mathbf{A}^{-1} \mathbf{F}_1.\end{aligned}$$

From (27) it follows that the matrix  $\mathbf{B}$  has full rank and hence the negative definiteness of  $\mathbf{A}$  implies that  $\widehat{\mathbf{A}}$  is positively definite. Next, we recall that we in (25) showed that the matrix

$$\begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{E} \end{pmatrix}$$

is negatively definite. Now, since  $\mathbf{A}$  and  $\mathbf{E}$  are both negatively definite, this implies that  $\widehat{\mathbf{C}}$  is negatively definite. Hence, we have a system of the same type as for Method II, cf. the equation (23). Again, if discontinuous pressures are used, we can eliminate  $\mathbf{P}$  locally and we obtain a positive system for  $\mathbf{U}$ .

For quadrilateral and hexahedral elements in the above formulations we have used equal order interpolation for the augmented stress and the velocity. The reason is that we need a pair of spaces satisfying the stability condition (27). To get the optimal approximation properties it is sufficient to take  $m = k - 1$ . Since the degrees of freedom for the augmented stress are condensed it would be preferable to choose the space  $\boldsymbol{\Sigma}_h$  as small as possible. Let us therefore close the paper by showing that it is possible to modify the above formulation with the same stabilization technique. Then any space for the augmented stress can be used, i.e. we let  $m$  be arbitrary and define

$$\boldsymbol{\Sigma}_h = \{ \boldsymbol{\tau} \in [L^2(\Omega)]^{N \times N} \mid \boldsymbol{\tau}|_K \in [R_m(K)]^{N \times N} \ \forall K \in \mathcal{C}_h \}, \quad (30a)$$

or

$$\boldsymbol{\Sigma}_h = \{ \boldsymbol{\tau} \in [C^0(\Omega)]^{N \times N} \mid \boldsymbol{\tau}|_K \in [R_m(K)]^{N \times N} \ \forall K \in \mathcal{C}_h \}. \quad (30b)$$

METHOD IV (Franca and Stenberg [1991]). Find  $(\boldsymbol{\sigma}_h, p_h, \mathbf{u}_h) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h$  such that

$$B_h(\boldsymbol{\sigma}_h, p_h, \mathbf{u}_h; \boldsymbol{\tau}, q, \mathbf{v}) = F_h(\boldsymbol{\tau}, q, \mathbf{v}) \quad \forall (\boldsymbol{\tau}, q, \mathbf{v}) \in \boldsymbol{\Sigma}_h \times P_h \times \mathbf{V}_h,$$

with

$$\begin{aligned}B_h(\boldsymbol{\kappa}, r, \mathbf{w}; \boldsymbol{\tau}, q, \mathbf{v}) &= -(\boldsymbol{\kappa}, \boldsymbol{\tau}) + (\nabla \mathbf{w}, \boldsymbol{\tau}) + (\boldsymbol{\kappa}, \nabla \mathbf{v}) + \beta(\boldsymbol{\kappa} - \nabla \mathbf{w}, \boldsymbol{\tau} - \nabla \mathbf{v}) \\ &\quad - (r, \operatorname{div} \mathbf{v}) - (\operatorname{div} \mathbf{w}, q) - \epsilon(r, q) - \alpha \sum_{K \in \mathcal{C}_h} h_K^2 (\operatorname{div} \boldsymbol{\kappa} - \nabla r, \operatorname{div} \boldsymbol{\tau} - \nabla q)_K. \quad \blacksquare\end{aligned}$$

**THEOREM 3.5.** *Assume that the finite element spaces satisfy either  $\mathbf{S}_h \subset \mathbf{V}_h$  or  $P_h \subset C^0(\Omega)$ , and further that  $0 < \beta < 1$  and  $\alpha > 0$ . For the approximation with Method IV we then have the following error estimate*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|p - p_h\|_0 + \|\mathbf{u} - \mathbf{u}_h\|_1 \leq C(h^{m+1} |\boldsymbol{\sigma}|_{m+1} + h^{l+1} |p|_{l+1} + h^k |\mathbf{u}|_{k+1}),$$

for all  $\epsilon$  in the range  $0 \leq \epsilon \leq Ch^s$ ,  $s = \min\{k + 1, l + 2, m + 2\}$ .

*Proof:* For the stability we note that

$$\begin{aligned} B_h(\boldsymbol{\sigma}, p, \mathbf{u}; -\boldsymbol{\sigma}, -p, \mathbf{u}) &= (1 - \beta)\|\boldsymbol{\sigma}\|_0^2 + \beta\|\nabla\mathbf{u}\|_0^2 + \epsilon\|p\|_0^2 + \alpha \sum_{K \in \mathcal{C}_h} h_K^2 \|\mathbf{div} \boldsymbol{\sigma} - \nabla p\|_{0_K}^2 \\ &\geq C(\|\boldsymbol{\sigma}\|_0^2 + \|\nabla\mathbf{u}\|_0^2 + \sum_{K \in \mathcal{C}_h} h_K^2 \|\mathbf{div} \boldsymbol{\sigma} - \nabla p\|_{0_K}^2), \end{aligned}$$

when  $0 < \beta < 1$ . The rest of the stability and error analysis is as before. ■

Of course this formulation leads to a second stability parameter for which the optimal choice is not known. However, for this parameter we have the explicit upper bound 1.

The matrix equations arising from this formulation is as (29) with a positive definite matrix instead of the zero matrix in the middle.

Let us finally remark that this method seems to be useful for some models of visco-elastic fluids, see Marchal and Crochet [1987]. These models contain spatial derivatives of the augmented stress and hence a continuous approximation for that variable is desirable. In the Galerkin formulation (obtained from above by choosing  $\alpha = 0$  and  $\beta = 0$ ) this leads to expensive elements as can be seen from Marchal and Crochet [1987] and Fortin and Pierre [1989].

## References

- Babuška, I. [1973]. The finite element method with Lagrangian multipliers. *Numer. Math.*, **20**, 179-192.
- Brezzi, F. [1974]. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers. *RAIRO Ser. Rouge* **8**, 129-151.
- Brezzi, F. and Pitkäranta, J. [1984]. On the stabilization of finite element approximations of the Stokes problem. *Efficient Solutions of Elliptic Systems, Notes on Numerical Fluid Mechanics*, **10** (Ed. by W. Hackbusch), Vieweg, Wiesbaden, 11-19.
- Brezzi, F. and Douglas, J. [1988]. Stabilized mixed methods for Stokes problem. *Numer. Math.* **53**, 225-236.
- Brezzi, F. and Fortin, M. [to appear]. *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, Heidelberg.
- Brooks, A. N. and Hughes, T.J.R. [1982]. Streamline upwind/Petrov-Galerkin formulations for convective dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg.* **32**, 199-259.

- Ciarlet, P. G. [1978]. *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam.
- Douglas, J. and Wang, J. [1989]. An absolutely stabilized finite element method for the Stokes problem. *Math. Comp.* **52**, 495-508.
- Durán, R. and Notchetto, R. [1989]. Pointwise accuracy of a stable Petrov-Galerkin approximation of the Stokes problem. *SIAM J. Num. Anal.* **26**, 1395-1406.
- Fortin, F. and Pierre, R. [1989]. On the convergence of the mixed method of Crochet and Marchal for viscoelastic flows. *Comput. Methods Appl. Mech. Engrg.* **73**, 341-350.
- Franca, L. P. and Hughes, T. J. R. [1988a]. Two classes of mixed finite element methods. *Comput. Methods Appl. Mech. Engrg.* **69**, 89-129.
- Franca, L.P., Hughes, T.J.R., Loula, A.F.D. and Miranda, I. [1988b]. A new family of stable element for nearly incompressible elasticity based on a mixed Petrov-Galerkin finite element formulation. *Numer. Math.* **53**, 123-141.
- Franca, L. P., Frey, S. L. and Hughes, T. J. R. [1990a]. Stabilized finite element methods: I. Application to the advective-diffusive model. *Report LNCC 32/90*, Rio de Janeiro, Brazil. Also, *Comput. Methods Appl. Mech. Engrg.*, to appear.
- Franca, L. P., Karam Filho, J., Loula, A. F. D. and Stenberg, R. [1990b] A convergence analysis of a stabilized method for the Stokes flow. *Report LNCC 19/90*, Rio de Janeiro, Brazil.
- Franca, L. P. and Stenberg, R. [1991]. Error analysis of some Galerkin-least-squares methods for the elasticity equations. *SIAM J. Num. Anal.* To appear
- Girault, V. and Raviart, P. A. [1986]. *Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms*, Springer-Verlag, Heidelberg.
- Harari, I. [to appear]. Ph.D. Thesis, Division of Applied Mechanics, Stanford University, Stanford, California.
- Hughes, T. J. R. [1987]. *The Finite Element Method. Linear Static and Dynamic Analysis*. Prentice-Hall.
- Hughes, T. J. R. and Brooks, A. N. [1979]. A multidimensional upwind scheme with no crosswind diffusion. *Finite Element Methods for Convection Dominated Flows* (Ed. by T.J.R. Hughes), ASME, New York, 19-35.
- Hughes, T. J. R., Franca, L. P. and Balestra, M. [1986]. A new finite element formulation for computational fluid dynamics: V. Circumventing the Babuška-Brezzi condition: A stable Petrov-Galerkin formulation Stokes problem accommodating equal-order inter-

- polations. *Comput. Methods Appl. Mech. Engrg.* **59**, 85-99.
- Hughes, T. J. R. and Franca, L. P. [1987]. A new finite element formulation for computational fluid dynamics: VII. The Stokes problem with various well-posed boundary conditions: symmetric formulations that converge for all velocity/pressure spaces. *Comput. Methods Appl. Mech. Engrg.* **65**, 85-96.
- Hughes, T. J. R., Franca, L. P. and Hulbert, G. M. [1989]. A new finite element formulation for computational fluid dynamics: VIII. The Galerkin-least-squares method for advective-diffusive equations. *Comput. Methods Appl. Mech. Engrg.* **73**, 173-189.
- Johnson, C. and Nävert, U. [1981]. An analysis of some finite element methods for advection-diffusion problems. *Analytical and Numerical Approaches to Asymptotic Problems in Analysis* (Ed. by O. Axelsson, L. S. Frank and A. Van Der Sluis), North-Holland, Amsterdam, 99-116.
- Johnson, C., Nävert, U. and Pitkäranta, J. [1984]. Finite element methods for linear hyperbolic problem. *Comput. Methods Appl. Mech. Engrg.*, **45**, 285-312.
- Karam Filho, J. and Loula, A.F.D. [1988]. New mixed Petrov-Galerkin finite element formulations for incompressible flow. *Proceedings of "II Encontro Nacional de Ciências Térmicas", ENCIT 88, Aguas de Lindóia, Brazil*, 172-175
- Marchal, J. M. and Crochet, M.J. [1987]. A new mixed finite element for calculating viscoelastic flow. *J. of Non-Newtonian Fluid Mech.* **26**, 77-114.
- Pierre, R. [1988]. Simple  $C^0$ -approximations for the computation of incompressible flows. *Comput. Methods Appl. Mech. Engrg.*, **68**, 205-228.
- Pierre, R. [1989]. Regularization procedures of mixed finite element approximations for the Stokes problem. *Num. Meths. Part. Diff. Eqs.*, **3**, 241-258
- Silvester, D. J. and Kechkar, N. [1990]. Stabilized bilinear-constant velocity-pressure finite elements for the conjugate gradient solution of the Stokes problem. *Comput. Methods Appl. Mech. Engrg.*, **79**, 71-86
- Stenberg, R. [1990]. A technique for analysing finite element methods for viscous incompressible flow. *Int. J. Num. Meths. Fluids.* **11**, 935-948.
- Verfürth, R. [1984]. Error estimates for a mixed finite element approximation of the Stokes problem. *RAIRO Anal. Numer.* **18**, 175-182.