ADVANCES IN COMPUTER METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS
R. VICHNEVETSKY (EDITOR)
PUBL. A I C A - 1975

AN ANALYSIS OF THE PENALTY METHOD AND EXTRAPOLATION FOR THE STATIONARY STOKES EQUATIONS

Richard S. Falk
Department of Mathematics
Rutgers University
New Brunswick, N. J. 08903

Summary

A major difficulty in the finite element method for approximating Stokes equations is the treatment of the incompressibility condition $\operatorname{div} \vec{v} = 0$. In this paper we use a penalty method approach to eliminate this problem and then show how extrapolation can be used to compute approximate solutions with higher order of accuracy using matrices with lower condition numbers than arise in the simple penalty method.

Introduction

In using finite element methods to approximate the solution of Stokes equations, a major difficulty is the construction of trial functions satisfying $\operatorname{div} \vec{v} = 0$, or some other condition which approximates it.

In this paper we analyze a "penalty method" for avoiding this problem and show how extrapolation can be used to improve the order of accuracy of the approximate solution.

The idea is based on a paper of J.T. King[4], where extrapolation procedures are used to achieve optimal accuracy in the Aubin-Babuška penalty method for the approximation of elliptic boundary value problems with Dirichlet type boundary conditions.

In this paper we consider the approximation of the stationary Stokes equations, i.e.

Problem (P): Find $\vec{u} = (u_1,\ldots,u_N)$ and p defined on Ω such that

$$-\nu\, \Delta \vec{u} + \overrightarrow{\text{grad}}\, p = \vec{f} \quad \text{in } \Omega$$
$$\operatorname{div} \vec{u} = 0 \quad \text{in } \Omega$$
$$\vec{u} = \vec{0} \quad \text{on } \partial\Omega$$

where $\vec{u}$ is the fluid velocity, p is the pressure, $\vec{f}$ are the body forces per unit mass and $\nu>0$ is the dynamic viscosity.

The literature on the theory and numerical analysis of the Navier-Stokes equations is immense. We mention only the recent work of Crouzeix and Raviart[3] on a finite element method for the problem we consider here and also the work of Temam[6] which contains results that we use in this paper and also an excellent bibliography.

An outline of the paper is as follows. In Section 2 we describe the notation to be used in the remainder of the paper. Section 3 contains the description of the approximate problem and the derivation of the error estimates. Finally, in Section 4, we make some brief comments about the method.

Notation

Let Ω be a bounded domain in $\mathbb{R}^N$. Denote by (u,v) the $L^2(\Omega)$ inner product.

$$\int_\Omega u(x)v(x)dx, \quad \text{and by}$$

$||v||_0$ the norm $(v,v)^{1/2}$. Let m be a non-negative integer and let $C^\infty(\bar{\Omega})$ denote the set of infinitely differentiable functions on $\bar{\Omega}$. Then $H^m(\Omega)$ will denote the completion of $C^\infty(\bar{\Omega})$ in the norm

$$||v||_m = \Big(\sum_{|\alpha|\le m} ||D^\alpha v||_0^2\Big)^{1/2} .$$

Now let $C_0^\infty(\Omega)$ be the set of infinitely differentiable functions with compact support in Ω and denote the completion of $C_0^\infty(\Omega)$ in the above norm by $H_0^m(\Omega)$.

For m a negative integer we define $H^m(\Omega)$ as the completion of $C^\infty(\Omega)$ with respect to the norm

$$||v||_m = \sup_{w\in C^\infty(\bar{\Omega})} \frac{(v,w)}{||w||_{-m}} .$$

We now define corresponding spaces for vector valued functions $\vec{v} = (v_1,\ldots,v_N)$. Let $[L^2(\Omega)]^N$ be the space of $\vec{v}$ with components $v_i \in L^2(\Omega)$. The scalar product in $[L^2(\Omega)]^N$ is given by

$$(\vec{u},\vec{v}) = \int_\Omega \vec{u}(x)\cdot\vec{v}(x)dx = \int_\Omega \sum_{i=1}^{N} u_i(x)v_i(x)dx.$$

Let $[H^m(\Omega)]^N$ be the space of $\vec{v}$ with components $v_i \in H^m(\Omega)$ and let $||\vec{v}||_m = (\sum_{i=1} ||v_i||_m^2)^{1/2}$.

Finally, for convenience we introduce the bilinear form

$$a(\vec{u},\vec{v}) = \nu \sum_{k=1}^{N} \sum_{\ell=1}^{N} \int_\Omega \frac{\partial u_k}{\partial x_\ell} \frac{\partial v_k}{\partial x_\ell}\, dx$$

defined on $[H_0^1(\Omega)]^N \times [H_0^1(\Omega)]^N$, and the corresponding norm $||\vec{u}||_E^2 = a(\vec{u},\vec{u})$.

Approximate Problem and Error Estimates

We begin our discussion with a statement of a regularity result for the solution of Problem (P) and an approximability assumption on the subspaces we will use in its approximation.

Lemma 1 (see Temam[4]). Let Ω be an open set of class C^s, $s\ge 2$, and let $\vec{f} \in [H^{s-2}(\Omega)]^N$ and $g \in H^{s-1}(\Omega)$ be given with $\int_\Omega g dx = 0$. Then there exist unique functions $\vec{u}$ and p (p is unique up to a constant) which are solutions of the generalized Stokes problem

$$-\nu\, \Delta \vec{u} + \overrightarrow{\text{grad}}\, p = \vec{f} \quad \text{in } \Omega$$
$$\operatorname{div} \vec{u} = g \quad \text{in } \Omega$$
$$\vec{u} = \vec{0} \quad \text{on } \partial\Omega$$

and satisfy $\vec{u} \in [H^s(\Omega)]^N$, $p \in H^{s-1}(\Omega)$ and the estimates $||\vec{u}||_s + ||p||_{s-1/\mathbb{R}} \leq C_o \{||\vec{f}||_{s-2} + ||g||_{s-1}\}$ $s \geq 1$ where C_o is a constant depending only on ν, s, and Ω.

$$(||p||_{s-1/\mathbb{R}} = \inf_{c\in\mathbb{R}} ||p+c||_{s-1}) .$$

The approximability assumption on the approximating subspaces we allow can be described as follows. Let h, $0<h<1$ be a parameter and $S_h^{\circ\, r}$ any one parameter family of finite dimensional subspaces of $H_0^1(\Omega)$ satisfying

(*) For any $u \in H^s(\Omega) \cap H_0^1(\Omega)$, $2 \leq s \leq r$ there exists $\bar{u} \in S_h^{\circ\, r}$ such that $||u-\bar{u}||_1 \leq C_1 h^{s-1} ||u||_s$.

References for the construction of such subspaces can be found in (1).

The approximate problem we will consider for the approximation of Problem (P) is given by:

Problem (P_h): Find $\vec{u}_h \in [S_h^{\circ\, r}]^N$ such that $a(\vec{u}_h,\vec{v}_h) + \gamma h^{-\sigma}$ $(\text{div } \vec{u}_h, \text{div } \vec{v}_h) = (\vec{f},\vec{v}_h)$ $\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$ where $\gamma>0$, and $\sigma \geq 0$ are constants.

We now turn our attention to the derivation of estimates which relate the quantities $\vec{u}_h$ and $\vec{u}$. These will immediately give error estimates for the penalty method (P_h) and also will serve as the crucial preliminary result for the derivation of the error estimates for extrapolation applied to Problem (P_h).

<u>Theorem 1</u>. Let $(\vec{u},p)$ be the solution of Problem (P) and $\vec{u}_h$ the solution of Problem (P_h). Define $(\vec{u}^m,p^m)$ $m \geq 1$ as solutions of the generalized Stokes problem (see Lemma 1)

$$-\nu \, \Delta \, \vec{u}^m + \overrightarrow{\text{grad}} \; p^m = \vec{0} \quad \text{in} \quad \Omega$$
$$\text{div } \vec{u}^m = -p^{m-1} \quad \text{in} \quad \Omega$$
$$\int_\Omega p^m(x)dx = 0$$
$$\vec{u}^m = \vec{0} \quad \text{on } \partial\Omega, \quad \text{where}$$

$(\vec{u}^0,p^0) = (\vec{u},p)$.

Then

$$||\vec{u}_h-\vec{u}-\vec{w}_k(\gamma)||_E \leq T_k \, (\vec{u},\vec{w}_k(\gamma);\vec{\phi}) \;\forall\; \vec{\phi} \in [S_h^{\circ\, r}]^N$$

where $\vec{w}_k(\gamma) = \sum_{m=1}^{k} [\gamma^{-1}h^\sigma]^m \vec{u}^m$ and

$T_k \, (\vec{u},\vec{w};\vec{\phi}) = \{a(\vec{\phi}-\vec{u}-\vec{w}, \vec{\phi}-\vec{u}-\vec{w})$
$+ \gamma h^{-\sigma} \; ||\text{div } (\vec{\phi}-\vec{u}-\vec{w} - [\gamma^{-1}h^\sigma]^{k+1}\vec{u}^{k+1})||_0^2\}^{1/2}$.

<u>Proof</u>. The exact solution $(\vec{u},p)$ satisfies $a(\vec{u},\vec{v}) + (\overrightarrow{\text{grad}} \; p,\vec{v}) = (\vec{f},\vec{v}) \;\forall\; \vec{v} \in [H_0^1(\Omega)]^N$ and $\text{div } \vec{u} = 0$. The approximate solution $\vec{u}_h$ satisfies $a(\vec{u}_h,\vec{v}_h) + \gamma h^{-\sigma}$ $(\text{div } \vec{u}_h, \text{div } \vec{v}_h) = (f,\vec{v}_h)$ $\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$. Hence:

$a(\vec{u}_h-\vec{u},\vec{v}_h) + \gamma h^{-\sigma}(\text{div } (\vec{u}_h-\vec{u}), \text{div } \vec{v}_h) - (\overrightarrow{\text{grad}} \; p,\vec{v}_h) = 0$ $\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$. Since $-(\overrightarrow{\text{grad}} \; p, \vec{v}_h) = (p, \text{div } \vec{v}_h)$ $\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$, we have
$a(\vec{e},\vec{v}_h) + \gamma h^{-\sigma}$ $(\text{div } \vec{e} + \gamma^{-1}h^\sigma p, \text{div } \vec{v}_h) = 0$
$\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$, where $\vec{e} = \vec{u}_h - \vec{u}$.

By the definition of $\vec{u}^m$, p^m, $a(\vec{u}^m,\vec{v}) + (\overrightarrow{\text{grad}} \; p^m,\vec{v}) = 0 \;\; \forall \vec{v} \in [H_0^1(\Omega)]^N$ so that
$a(\vec{u}^m,\vec{v}) - (p^m, \text{div } \vec{v}) = 0$ or
$a(\vec{u}^m,\vec{v}) + (\text{div } \vec{u}^{m+1}, \text{div } \vec{v}) = 0 \;\; \forall \vec{v} \in [H_0^1(\Omega)]^N$.
Multiplying the above equation by $[\gamma^{-1}h^\sigma]^m$ and summing from $m=1,\ldots,k-1$, we obtain:

$$a(\sum_{m=1}^{k-1} [\gamma^{-1}h^\sigma]^m\vec{u}^m, \vec{v})$$
$$+ \gamma h^{-\sigma} \; (\text{div} \sum_{m=1}^{k-1} [\gamma^{-1}h^\sigma]^{m+1}\vec{u}^{m+1}, \text{div } \vec{v}) = 0$$

so that $a(\sum_{m=1}^{k} [\gamma^{-1}h^\sigma]^m\vec{u}^m, \vec{v})$

$$+ \gamma h^{-\sigma}(\text{div} \sum_{m=1}^{k} [\gamma^{-1}h^\sigma]^m\vec{u}^m, \text{div } \vec{v})$$
$$-a([\gamma^{-1}h^\sigma]^k\vec{u}^k, \vec{v}) - \;\gamma h^{-\sigma}(\gamma^{-1}h^\sigma \text{ div } \vec{u}^1, \text{div } \vec{v})=0$$
$$\forall \vec{v} \in [H_0^1(\Omega)]^N.$$

Now $a(\vec{u}^k,\vec{v}) = (p^k, \text{div } \vec{v}) = (-\text{div } \vec{u}^{k+1}, \text{div } \vec{v})$ and $\text{div } u^1 = -p^0 = -p$. Abbreviating $\vec{w}_k(\gamma) = \sum_{m=1}^{k} [\gamma^{-1}h^\sigma]^m\vec{u}^m$ by $\vec{w}$, we may rewrite the above as
$a(\vec{w},\vec{v}) + \gamma h^{-\sigma}$ $(\text{div } \vec{w}, \text{div } \vec{v}) + ([\gamma^{-1}h^\sigma]^k \text{ div } \vec{u}^{k+1},\text{div } \vec{v})$
$+ \gamma h^{-\sigma}$ $(\gamma^{-1}h^\sigma p, \text{div } \vec{v}) = 0 \;\; \forall \vec{v} \in [H_0^1(\Omega)]^N$.

Subtracting this from the equation for $\vec{e}$, we obtain:
$a(\vec{e}-\vec{w},\vec{v}_h) + \gamma h^{-\sigma}$ $(\text{div } [\vec{e}-\vec{w}-[\gamma^{-1}h^\sigma]^{k+1}\vec{u}^{k+1}], \text{div } \vec{v}_h) = 0$
$\forall \vec{v}_h \in [S_h^{\circ\, r}]^N$.

This implies that $\vec{u}_h$ minimizes over $[S_h^{\circ\, r}]^N$ the functional (in $\vec{\phi}$) $T_k \, (\vec{u},\vec{w};\vec{\phi}) = \{a(\vec{\phi}-\vec{u},\vec{w}, \vec{\phi}-\vec{u}-\vec{w})$
$+ \gamma h^{-\sigma}||\text{div } (\vec{\phi}-\vec{u}-\vec{w} - [\gamma^{-1}h^\sigma]^{k+1}\vec{u}^{k+1})||_0^2\}^{1/2}$.
Hence we have that $||\vec{e}-\vec{w}||_E = [a(\vec{e}-\vec{w}, \vec{e}-\vec{w})]^{1/2} \leq$
$T_k \, (\vec{u},\vec{w};\vec{u}_h) \leq T_k \, (\vec{u},\vec{w};\vec{\phi}) \;\; \forall \vec{\phi} \in [S_h^{\circ\, r}]^N$.

<u>Theorem 2</u>. Let $(\vec{u}^m,p^m)$, $m=0,\ldots,k+1$ and $\vec{w}_k(\gamma)$ be as defined in Theorem 1. Suppose $\vec{f} \in [H^{s-2}(\Omega)]^N$, $2 \leq s \leq r$, and h_o and γ_o satisfy $C_o\gamma_o^{-1}h_o^{\sigma}<1$. Then for all $h \leq h_o$ and $\gamma \geq \gamma_o$,

$$||\vec{u}_h-\vec{u}-\vec{w}_k(\gamma)||_E \leq C_2\{(1+[\gamma h^{-\sigma}]^{1/2}h^{s-1}+[C_o\gamma^{-1}h^\sigma]^{k+1}\}||\vec{f}||_{s-2}$$

where C_2 is independent of σ, h, $\vec{u}$, and k.

<u>Proof</u>. Again using the abbreviation $\vec{w}_k(\gamma) = \vec{w}$, we have from Theorem 1,

$$||\vec{u}_h-\vec{u}-\vec{w}||_E \leq T_k \, (\vec{u},\vec{w};\vec{\phi}) \quad \forall \vec{\phi} \in [S_h^{\circ\, r}]^N.$$

Now $T_k \, (\vec{u},\vec{w};\vec{\phi})$
$= \{a(\vec{\phi}-\vec{u}-\vec{w}, \vec{\phi}-\vec{u}-\vec{w})$
$+ \gamma h^{-\sigma} \; ||\text{div } (\vec{\phi}-\vec{u}-\vec{w} + [\gamma^{-1}h^\sigma]^{k+1}\vec{u}^{k+1})||_0^2\}^{1/2}$
$\leq ||\vec{\phi}-\vec{u}-\vec{w}||_E+[\gamma h^{-\sigma}]^{1/2}||\text{div } (\vec{\phi}-\vec{u}-\vec{w} + [\gamma^{-1}h^\sigma]^{k+1}\vec{u}^{k+1})||_0$.

Choose $\vec{\phi} = \vec{\phi}^0 + \sum_{j=1}^{k+1} [\gamma^{-1}h^{\sigma}]^j\vec{\phi}^j$

where $\vec{\phi}^0,\ldots,\vec{\phi}^{k+1} \in [\overset{o}{S_h}{}^r]^N$ are to be determined.

Then $||\vec{u}_h-\vec{u}-\vec{w}||_E$

$$\leq \sum_{j=0}^{k} [\gamma^{-1}h^{\sigma}]^j\ ||\vec{u}^j-\vec{\phi}^j||_E + [\gamma^{-1}h^{\sigma}]^{k+1}\ ||\vec{\phi}^{k+1}||_E$$
$$+\ [\gamma h^{-\sigma}]^{\frac{1}{2}} \sum_{j=0}^{k+1} [\gamma^{-1}h^{\sigma}]^j\ ||\text{div}\ (\vec{u}^j-\vec{\phi}^j)||_0.$$

Using the easily established inequality

$$||\text{div}\ \vec{z}||_0 \leq \sqrt{N/\nu}\ \ ||\vec{z}||_E$$

and the triangle inequality

$$||\vec{\phi}^{k+1}||_E \leq ||\vec{u}^{k+1}-\vec{\phi}^{k+1}||_E + ||\vec{u}^{k+1}||_E$$

we obtain

$$||\vec{u}_h-\vec{u}-\vec{w}||_E \leq$$
$$(1 + [\tfrac{N}{\nu}\ \gamma h^{-\sigma}]^{\frac{1}{2}} \sum_{j=0}^{k+1} [\gamma^{-1}h^{\sigma}]^j\ ||\vec{u}^j-\vec{\phi}^j||_E$$
$$+ [\gamma^{-1}h^{\sigma}]^{k+1}\ ||\vec{u}^{k+1}||_E.$$

By Lemma 1, $(\vec{u}^m,p^m) \in [H^s(\Omega)]^N \times H^{s-1}(\Omega)\ \ \forall\ m \geq 1$ and $||\vec{u}^m||_s + ||p^m||_{s-1} \leq C_o\ ||p^{m-1}||_{s-1}$ so that $||\vec{u}^m||_s + ||p^m||_{s-1} \leq C_o{}^m\ ||p||_{s-1} \leq C_o{}^{m+1}\ ||\vec{f}||_{s-2}$ (again using Lemma 1).

Similarly, $||u^{k+1}||_E \leq \nu\ C_o{}^{k+1}\ ||p||_0$

$$\leq \nu\ \ C_o{}^{k+2}\ ||\vec{f}||_{-1} \leq \nu\ C_o{}^{k+2}\ ||\vec{f}||_{s-2} \qquad (s\geq 2).$$

Hence, for appropriate choices of the $\vec{\phi}^j$, we have by our approximability assumption (*), that

$$||\vec{u}^j-\vec{\phi}^j||_E \leq \nu\ C_1h^{s-1}\ ||\vec{u}^j||_s$$
$$\leq \nu\ C_1[C_o]^{j+1}\ h^{s-1}\ ||\vec{f}||_{s-2} \qquad j=0,\ldots,k+1.$$

We therefore obtain the estimate

$$||\vec{u}_h-\vec{u}-\vec{w}||_E\leq C_o\ \ \nu\{C_1(1+\tfrac{N}{\nu}\ \gamma h^{-\sigma}]^{\frac{1}{2}})(\sum_{j=0}^{k} [C_o\gamma^{-1}h^{\sigma}]^j)h^{s-1} +$$
$$[C_o\gamma^{-1}h^{\sigma}]^{k+1}\}||\vec{f}||_{s-2}.$$

Since $C_o\gamma_o^{-1}h_o{}^{\sigma}<1$ we have $\forall$ $h\leq h_o$ and $\gamma\geq\gamma_o$
$$||\vec{u}_h-\vec{u}-\vec{w}||_E\leq C_2\{(1+[\gamma h^{-\sigma}]^{\frac{1}{2}})h^{s-1}+[C_o\gamma^{-1}h^{\sigma}]^{k+1}\}||\vec{f}||_{s-2}.$$

Corollary 2.1. $||\vec{u}_h-\vec{u}-\vec{w}_k(\gamma)||_E \leq C_3h^{\lambda}\ ||\vec{f}||_{s-2}$ where $\lambda=\min\ (s-1-\sigma/2,\ \sigma(k+1))$ and C_3 is a constant independent of h and $\vec{u}$, but dependent on γ and k. (When useful we will write $C_3=C_3(\gamma,k)$).

Remark 1. One easily sees that Theorems 1 and 2 hold also when $k=0$ with the interpretation that $\vec{w}_0(\gamma)=0$. Hence we have $||\vec{u}_h-\vec{u}||_E \leq C_3\ h^{\beta}\ ||\vec{f}||_{s-2}$ where $\beta=\min\ (s-1-\sigma/2,\ \sigma)$. We achieve optimality when $\sigma=2(s-1)/3$, for which we obtain $||\vec{u}_h-\vec{u}||_E \leq C_3\ h^{2(s-1)/3}\ ||\vec{f}||_{s-2}$. Thus this special case of our theorem gives an order of convergence estimate for the "penalty method" without extrapolation.

Using Theorem 2 and Corollary 2.1, we can now show how extrapolation can be used to obtain approximate solutions with higher order of accuracy than given above.

Let $\gamma_0,\ldots,\gamma_k$ be distinct and choose $a_0,\ldots,a_k$ so that

$$\sum_{i=0}^{k} a_i = 1$$
$$\sum_{i=0}^{k} a_i\ \gamma_i^{-j} = 0 \qquad\qquad 1\leq j\leq k$$

We define $\vec{u}_h^{(k)}(\gamma) = \sum_{i=0}^{k} a_i\ \vec{u}_h^{(0)}(\gamma_i)$ where $\vec{u}_h^0(\gamma)$ is the solution of Problem (P_h) with weight $\gamma h^{-\sigma}$. We remark that the coefficients a_i exists and are unique as the above system is a Vandermonde.

In the case that $\gamma_i=2^i\gamma$ for some $\gamma>0$ we can give an explicit definition of the k^th extrapolate, $\vec{u}_h^{(k)}(\gamma)$. Define

$$\vec{u}_h^{(k)}(\gamma) = \frac{2^k\ \vec{u}_h^{(k-1)}(2\gamma) - \vec{u}_h^{(k-1)}(\gamma)}{2^k-1} \qquad k\geq 1.$$

We can then prove the following:

Theorem 3. Assume the hypotheses of Theorem 2 hold. Then $||\vec{u}-\vec{u}_h^{(k)}(\gamma)||_E \leq C_4(\gamma,k)h^{\lambda}\ ||\vec{f}||_{s-2}$ where $\lambda=\min\ (s-1-\sigma/2,\ \sigma(k+1))$ and C_4 is a constant independent of h and u, but dependent on γ and k.

Proof. From the definition of the a_i, we have that

$$||\vec{u}-\vec{u}_h^k(\gamma)||_E$$
$$= ||\ \sum_{i=0}^{k} a_i\ (\vec{u}-\vec{u}_h^{(0)}(2^i\gamma) + \sum_{j=1}^{k}\ [2^{-i}\gamma^{-1}h^{\sigma}]^j\vec{u}^j)||_E$$
$$\leq \sum_{i=0}^{k} |a_i|\ \ ||\vec{u}-\vec{u}_h^{(0)}(2^i\gamma) + \vec{w}_k(2^i\gamma)||_E$$
$$\leq \sum_{i=0}^{k} |a_i|\ C_3h^{\lambda}\ ||\vec{f}||_{s-2} \text{ (using Corollary 2.1)}$$
$$\leq C_4(\gamma,k)\ h^{\lambda}\ ||\vec{f}||_{s-2}$$

where $\lambda=\min\ (s-1-\sigma/2,\ \sigma(k+1))$.

Remark 2. For fixed k, the optimal choice of σ occurs when

$$s-1-\sigma/2 = \sigma(k+1), \text{ i.e}$$
$$\sigma = (s-1)/(k+3/2)$$

for which we obtain

$$||\vec{u}_h-\vec{u}_h^{(k)}(\gamma)||_E \leq C_4\ (\gamma,k)\ h^{\mu}\ ||\vec{f}||_{s-2}$$

where $\mu=(s-1)(k+1)/(k+3/2)$.

Remark 3. From the previous remark, one observes that the optimal choice of σ depends on the regularity of the solution and the number of extrapolations to be performed. A natural question then is whether the convergence of the method is affected if one overestimates the regularity of the solution. From Theorem 3 one has that the method will converge if $s-1-\sigma/2>0$. If one mistakenly assumes additional regularity for the solution, then using the criterion of Remark 2 and the approximability assumption (*), one may choose σ as large as $(r-1)/(k+3/2)$. If s is known to be at least 2, then for convergence we require $(r-1)/(2k+3)<1$, i.e. $k>(r-4)/2$. Using piecewise cubics, for example, one extrapolation will guarantee convergence.

For $k=0,1,2,$ and 3 and the corresponding optimal choices of σ, we get the following error bounds for $||\vec{u}-\vec{u}_h^{(k)}(\gamma)||_E$ of the form Ch^{μ}.

k	σ	λ
0	$\frac{2}{3}(s-1)$	$\frac{2}{3}(s-1)$
1	$\frac{2}{5}(s-1)$	$\frac{4}{5}(s-1)$
2	$\frac{2}{7}(s-1)$	$\frac{6}{7}(s-1)$
3	$\frac{2}{9}(s-1)$	$\frac{8}{9}(s-1)$

Thus extrapolation gives an improvement in the order of accuracy and also, since σ is decreasing, in the condition number of the matrix used to compute the approximate solution. (Using standard techniques it is easy to show the condition number is $0(h^{-2-\sigma})$. For example, when $s=4$ (i.e. using piecewise cubics) one obtains $0(h^2)$ for $k=0$ and $0(h^{8/3})$ for $k=3$. (The best possible is $0(h^3)$). The condition number is improved from $0(h^{-4})$ to $0(h^{-8/3})$.

Remark 4. From the analysis presented, it may at first appear that at least theoretically, a very small σ and large k may be desirable to achieve a low condition number and high accuracy. Putting aside the practical difficulties of using a large value for k, it must be remembered that the constant C_4 depends on k and in fact can be shown to approach infinity at a rate $\geq 0[\sqrt{2})^k]$.

Hence some numerical experiments will be needed to find what choices of k are reasonable to use in practice.

Using a variant of the Nitsche duality argument, we can derive the following error estimate in lower norms for the extrapolated penalty method.

Theorem 4. Suppose $\vec{f} \epsilon [H^{s-2}(\Omega)]^N$ $2\leq s\leq r$ and h_o and γ_o satisfy $C_o\,\gamma_o^{-1}\,h_o^{\sigma} < 1$. Then for all $h\leq h_o$, $\gamma\geq\gamma_o$ and $\alpha\geq 2$,

$$||\vec{u}-\vec{u}_h^{(t)}(\gamma)||_{2-\alpha} \leq C_6(\gamma,t,\sigma)h^{\beta}\ ||\vec{f}||_{s-2}$$

where $\beta=\min\ (\alpha-1-\sigma/2 + \lambda_t,\ \sigma(t+1))$ and $\lambda_t=\min\ (s-1-\sigma/2,\ \sigma(t+1))$.

Remark 5. If $\sigma = \frac{s-1}{k+3/2}$, (the optimal choice for the $||\cdot||_E$ norm estimate), then there will be no improvement in accuracy in lower norms unless $t>k$. For $t\geq k$, $\lambda_t = s-1-\sigma/2$ which implies that

$$||\vec{u}-\vec{u}_h^{(t)}(\gamma)||_{2-\alpha} \leq C_6 h^{\delta}\ ||\vec{f}||_{s-2}$$

where $\delta=\min\ (s+\alpha-2-\sigma,\ \sigma(t+1))$.

For example, if $\sigma = \frac{s-1}{k+3/2} < 2$ and $t \geq \frac{sk+2-s/2}{s-1}$ then $||\vec{u}-\vec{u}_h^{(t)}(\gamma)||_0 \leq C_6\ h^{s-\sigma}\ ||\vec{f}||_{s-2}$, while $||\vec{u}-\vec{u}_h^{(k)}(\gamma)||_E \leq C_4\ h^{s-1-\sigma/2}\ ||\vec{f}||_{s-2}$.

Comments

As remarked earlier, the advantage of using the extrapolated penalty method is that one can achieve higher accuracy while using matrices with lower condition numbers.

Although several linear systems must be solved to obtain the extrapolated approximate solution, no other inner products need be computed than the ones already required in the usual penalty method.

Finally, we remark again that since what we have presented here is an asymptotic error analysis, numerical studies would be useful to determine what choices of γ and k are reasonable to use in practice.

References

(1) I. Babuška, "The Finite Element Method with Penalty", Math. Comp., 27 (1973), pp. 221-228.

(2) M. Crouseix and P.A. Raviart, "Conforming and Nonconforming Finite Element Methods for Solving the Stationary Stokes Equations I, Revue Francaise d' Automatique, Informatique et Recherche Operationelle, 7 annee, decembre 1973, R-3, pp. 33-76.

(3) J.T. King, "New Error Bounds for the Penalty Method and Extrapolation", preprint.

(4) O.A. Ladyzhenskaya, The Mathematical Theory of Viscous Incompressible Flow, Gordon and Breach, New York, 1962.

(5) R. Temam, On the Theory and Numerical Analysis of the Naviar-Stokes Equations, Lecture Note #9, University of Maryland, June, 1973.