SIAM J. NUMER. ANAL.
Vol. 13, No. 5, October 1976

A PENALTY AND EXTRAPOLATION METHOD FOR THE STATIONARY STOKES EQUATIONS*

RICHARD S. FALK† AND J. THOMAS KING‡

Abstract. A penalty-extrapolation procedure is analyzed for avoiding the problem of the construction of trial functions satisfying $\operatorname{div} \mathbf{v}=0$ and $\mathbf{v}=\mathbf{0}$ on $\partial\Omega$ in the approximation of the stationary Stokes equations.

1. Introduction. In using finite element methods to approximate the solution of Stokes equations, a major difficulty is the construction of trial functions satisfying $\operatorname{div} \mathbf{v}=0$, or some other condition which approximates it. This problem is compounded by the fact that the trial functions must also vanish on the boundary.

In this paper we analyze a "penalty method" for avoiding both these problems and show how extrapolation can be used to improve the order of accuracy of the approximate solution while using matrices with lower condition numbers than arise in the simple penalty method.

The idea is based on a paper of J.T. King [7], 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 condition.

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

Problem (P). Find $\mathbf{u}=(u_1,\cdots,u_N)$ and p defined on Ω such that

$$-\nu\Delta\mathbf{u}+\mathbf{grad}\, p=\mathbf{f} \quad \text{in } \Omega,$$

$$\operatorname{div}\mathbf{u}=0 \quad \text{in } \Omega,$$

$$\mathbf{u}=\mathbf{0} \quad \text{on } \partial\Omega,$$

where $\mathbf{u}$ is the fluid velocity, p is the pressure, $\mathbf{f}$ is the body force 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 the recent work of Crouzeix and Raviart [4] on a finite element method for the problem we consider here and also the work of Temam [10] which contains results we use in this paper and also an excellent bibliography. In Falk [6] another type of "penalty method" is used to obtain optimal error estimates for this problem, and in Falk [5] the techniques we will use in this paper were exploited to obtain results for this problem in the case where the trial functions do satisfy the boundary conditions.

An outline of the paper is as follows. In § 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 error estimates. Finally, in § 4, we make some comments about the method.

---

* Received by the editors February 20, 1975, and in revised form November 16, 1975.

† Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903. This work was supported by the National Science Foundation under Grant MPS74-05795.

‡ Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221.

2. Notation. Let Ω be a bounded domain in $\mathbb{R}^N$ with "smooth" boundary, $\partial\Omega$. For $\beta \geqq 0$ let $H^\beta(\Omega)$ and $H^\beta(\partial\Omega)$ denote the Sobolev spaces of order β on Ω and $\partial\Omega$, respectively, with associated norms $\|\cdot\|_\beta$ and $|\cdot|_\beta$, respectively. For definitions and characterizations of these spaces, the conventions of [9] are adopted.

For $L_2(\Omega)$ and $L_2(\partial\Omega)$ we will denote the inner products by

$$(u, v) = \int_\Omega uv \, dx$$

and

$$\langle u, v\rangle = \int_{\partial\Omega} uv \, d\Gamma,$$

respectively.

We will also need the spaces $H^\beta(\Omega)$ for $\beta < 0$. For $\psi \in C^\infty(\bar{\Omega})$ and $\beta < 0$, we define

$$\|\psi\|_\beta = \sup_{\phi \in C^\infty(\bar{\Omega})} \frac{|(\psi, \phi)|}{\|\phi\|_{-\beta}}.$$

Then $H^\beta(\Omega)$, for $\beta < 0$ is defined as the completion of $C^\infty(\bar{\Omega})$ with respect to the above norm.

We now define corresponding spaces for vector-valued functions $\mathbf{v} = (v_1, \cdots, v_N)$. Let $[H^\beta(\Omega)]^N$ be the space of $\mathbf{v}$ with components v_i in $H^\beta(\Omega)$ and let

$$\|\mathbf{v}\|_\beta = \left(\sum_{i=1}^N \|v_i\|_\beta^2 \right)^{1/2}.$$

The scalar product in $[L^2(\Omega)]^N$ is given by

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

The spaces $[H^\beta(\partial\Omega)]^N$ are defined in a similar fashion.

Finally for convenience we introduce the bilinear form

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

defined on $[H^1(\Omega)]^N \times [H^1(\Omega)]^N$ and the corresponding seminorm

$$\|\mathbf{u}\|_E^2 = a(\mathbf{u}, \mathbf{u}).$$

3. Approximate problem and error estimates. We begin our discussion with a description of the class of finite-dimensional spaces we will use in the approximation of Problem (P):

Let h, $0 < h < 1$, be a parameter and $r \geqq 2$ an integer. Let S_h^r be any one parameter family of finite-dimensional subspaces of $H^1(\Omega)$ having the following approximation property:

(I) For any $u \in H^\beta(\Omega)$, $\beta \geqq 1$, there exists $\bar{u} \in S_h^r$ such that

$$\|u-\bar{u}\|_0 + h\|u-\bar{u}\|_1 \leqq Ch^s \|u\|_\beta,$$

where $s = \min(\beta, r)$ and C is a constant independent of h and u. Many examples of subspaces satisfying (I), for various choices of r, may be found in the literature. We refer the reader to the list cited in [1, Chap. 4].

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

Problem ($\mathrm{P_h}$). Find $\mathbf{u}_h \in [S_h^r]^N$ such that for every $\mathbf{v}_h \in [S_h^r]^N$,

$$a(\mathbf{u}_h, \mathbf{v}_h) + \gamma h^{-\sigma_1}(\operatorname{div} \mathbf{u}_h, \operatorname{div} \mathbf{v}_h) + \gamma h^{-\sigma_2}\langle \mathbf{u}_h, \mathbf{v}_h\rangle = (\mathbf{f}, \mathbf{v}_h),$$

where $\gamma > 0$, $\sigma_1 \geqq 0$, $\sigma_2 \geqq 0$ are constants.

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

To obtain these results we need to make use of a regularity result for the solution of the generalized Stokes problem and also some well-known results from partial differential equations. We state these results without proof for the convenience of the reader. Following that, we prove two lemmas which we will need to establish our main results.

LEMMA 1 (see Temam [10]). *Let Ω be an open bounded set of class C^β, $\beta \geqq 2$ an integer, and let $\mathbf{f} \in [H^{\beta-2}(\Omega)]^N$, $g \in H^{\beta-1}(\Omega)$, and $\boldsymbol{\phi} \in [H^{\beta-(1/2)}(\partial\Omega)]^N$, where*

$$\int_\Omega g \, dx = \int_{\partial\Omega} \boldsymbol{\phi} \cdot \boldsymbol{\eta} \, d\Gamma,$$

$\boldsymbol{\eta}$ = *unit outward normal.*

Then there exists unique functions $\mathbf{u}$ and p (p is unique up to a constant) which are solutions of the generalized Stokes problem:

$$-\nu\Delta\mathbf{u} + \mathbf{grad}\, p = \mathbf{f} \quad \text{in } \Omega,$$

$$\operatorname{div} \mathbf{u} = g \quad \text{in } \Omega,$$

$$\mathbf{u} = \boldsymbol{\phi} \quad \text{on } \partial\Omega,$$

with $\mathbf{u} \in [H^\beta(\Omega)]^N$, $p \in H^{\beta-1}(\Omega)$ and satisfy the estimate

$$\|\mathbf{u}\|_\beta + \|p\|_{(\beta-1)/\mathbb{R}} \leqq C_0\{\|\mathbf{f}\|_{\beta-2} + \|g\|_{\beta-1} + |\boldsymbol{\phi}|_{\beta-(1/2)}\},$$

where C_0 is a constant depending only on ν, β and Ω and

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

LEMMA 2 (cf. [9]). *Let $u \in H^\beta(\Omega)$, $\beta > \frac{1}{2}$. Then there exists a trace of the function u on $\partial\Omega$ and*

$$|u|_{\beta-(1/2)} \leqq C\|u\|_\beta,$$

where C is a constant independent of u.

LEMMA 3 (cf. [9]). *Let $u \in H^\beta(\Omega)$, $\beta > \frac{3}{2}$. Then there exists a trace $\partial u/\partial \eta$ on $\partial\Omega$ and*

$$\left|\frac{\partial u}{\partial \eta}\right|_{\beta-(3/2)} \leqq C\|u\|_\beta,$$

where C is a constant independent of u, and $\partial u/\partial \eta$ denotes the normal derivative of u.

LEMMA 4 (cf. [7]). *For $\mathbf{u} \in [H^1(\Omega)]^N$, there exists a constant C independent of u, such that*

$$\|\mathbf{u}\|_1^2 \leqq C\{a(\mathbf{u}, \mathbf{u}) + |\mathbf{u}|_0^2\}.$$

We remark that C will be used to denote a positive constant not necessarily the same in any two places.

LEMMA 5 (cf. [3]). *For $u \in H^1(\Omega)$ and $\varepsilon > 0$, there is a constant C independent of u and ε such that*

$$|u|_0^2 \leqq \varepsilon\|u\|_0^2 + C\varepsilon^{-1}\|u\|_1^2.$$

LEMMA 6. *Let $u \in [H^\beta(\Omega)]^N$, $\beta \geqq 1$. Then there exists a $\mathbf{v} \in [S_h^r]^N$ such that*

$$\|\mathbf{u}-\mathbf{v}\|_E + [\gamma h^{-\sigma_1}]^{(1/2)}\|\operatorname{div}(\mathbf{u}-\mathbf{v})\|_0 + [\gamma h^{-\sigma_2}]^{(1/2)}|\mathbf{u}-\mathbf{v}|_0$$
$$\leqq C[h^{s-1} + \gamma^{1/2}h^{s-1-\sigma_1/2} + \gamma^{1/2}h^{s-1/2-\sigma_2/2}]\|\mathbf{u}\|_\beta,$$

where $s = \min(\beta, r)$ and C is a constant independent of h and $\mathbf{u}$.

Proof. The lemma follows from the approximability assumption (I), Lemma 5 with $\varepsilon = h^{-1}$, and the fact that

$$\|\operatorname{div}\mathbf{u}\|_0 \leqq C\|\mathbf{u}\|_1.$$

LEMMA 7. *Let $\mathbf{f} \in [H^{\beta-2}(\Omega)]^N$, $\beta \geqq 2$ and let $(\mathbf{u}^m, p^m) = (\mathbf{z}^m + h^{\sigma_2-\sigma_1}\mathbf{w}^m, Q^m + h^{\sigma_2-\sigma_1}q^m)$, $1 \leqq m \leqq \beta - 1$, where $(\mathbf{z}^m, Q^m)$ and $(\mathbf{w}^m, q^m)$ are the respective solutions of the following generalized Stokes problems:*

(A)
$$-\nu\Delta\mathbf{z}^m + \mathbf{grad}\, Q^m = \mathbf{0} \quad \text{in } \Omega,$$
$$\operatorname{div}\mathbf{z}^m = -p^{m-1} \quad \text{in } \Omega,$$
$$\mathbf{z}^m = \left[\frac{1}{\mu(\partial\Omega)}\int_\Omega -p^{m-1}\,dx\right]\boldsymbol{\eta} \quad \text{on } \partial\Omega.$$

(B)
$$-\nu\Delta\mathbf{w}^m + \mathbf{grad}\, q^m = \mathbf{0} \quad \text{in } \Omega,$$
$$\operatorname{div}\mathbf{w}^m = \frac{1}{\mu(\Omega)}\int_{\partial\Omega}\left(p^{m-1} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right)d\Gamma \quad \text{in } \Omega,$$
$$\mathbf{w}^m = p^{m-1}\boldsymbol{\eta} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta} \quad \text{on } \partial\Omega,$$

and

(D)
$$\frac{1}{\mu(\partial\Omega)}\int_\Omega -p^{m-1}\,dx = h^{2(\sigma_2-\sigma_1)}\frac{1}{\mu(\Omega)}\int_{\partial\Omega}\left(p^{m-1} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right)d\Gamma,$$

where $\mu(\Omega)$ and $\mu(\partial\Omega)$ denote the N- and $(N-1)$-dimensional measures of Ω and $\partial\Omega$, respectively, and $(\mathbf{u}^0, p^0)=(\mathbf{u}, p)$. Then

$$\mathbf{z}^m \in [H^{\beta-m+1}(\Omega)]^N,$$
$$\mathbf{w}^m \in [H^{\beta-m}(\Omega)]^N, \qquad m=1,\cdots,\beta-1,$$

and

$$\|\mathbf{z}^m\|_{\beta-m+1} \leqq C\|\mathbf{f}\|_{\beta-2},$$
$$\|\mathbf{w}^m\|_{\beta-m} \leqq C\|\mathbf{f}\|_{\beta-2},$$

for some constant C independent of h and $\mathbf{f}$ (C will depend on m).

Proof. We first note that by Lemma 1, Q^m and q^m and hence p^m are determined uniquely up to a constant. This constant is then specified by condition (D). It will also be necessary to make use of solutions of problems (A) and (B) with different normalizations. We denote by Q^m_* and q^m_* the respective solutions of problems (A) and (B) with the normalizations

$$\int_\Omega Q^m_*\,dx=0, \qquad \int_\Omega q^m_*\,dx=0, \qquad m \geqq 1.$$

Let $p^m_* = Q^m_* + h^{\sigma_2-\sigma_1} q^m_*$, $m \geqq 1$, and define constants D^m by

$$p^m = p^m_* + D^m.$$

For $m=0$, let p^* denote the solution of Problem (P) such that $\int_\Omega p^*\,dx=0$, and p the solution of Problem (P) satisfying condition (D).

Then using condition (D), it is easily seen that for $m \geqq 1$,

$$D^{m-1} = \frac{\dfrac{-h^{2(\sigma_2-\sigma_1)}}{\mu(\Omega)}\displaystyle\int_{\partial\Omega}\left(p^{m-1}_* - \nu\frac{\partial \mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right)d\Gamma}{\dfrac{\mu(\Omega)}{\mu(\partial\Omega)} + h^{2(\sigma_2-\sigma_1)}\dfrac{\mu(\partial\Omega)}{\mu(\Omega)}}.$$

Hence

$$\begin{aligned}|D^{m-1}| &\leqq C(\Omega)h^{2(\sigma_2-\sigma_1)}\left|\int_{\partial\Omega}\left(p^{m-1}_* - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right)d\Gamma\right| \\ &\leqq C(\Omega)h^{2(\sigma_2-\sigma_1)}\left|p^{m-1}_* - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right|_0 \\ &\leqq C(\Omega)h^{2(\sigma_2-\sigma_1)}[\|p^{m-1}_*\|_1 + \|\mathbf{u}^{m-1}\|_2]\end{aligned} \tag{3.1}$$

(using Lemmas 2 and 3).

By Lemma 1, $(\mathbf{z}^m, Q^m) \in [H^{\beta-m+1}(\Omega)]^N \times H^{\beta-m}(\Omega)$ and

$$\begin{aligned}
&\|\mathbf{z}^m\|_{\beta-m+1} + \|Q_*^m\|_{\beta-m} \\
&\quad \leqq C_0\left[\|p^{m-1}\|_{\beta-m} + \left|\frac{1}{\mu(\partial\Omega)}\int_\Omega -p^{m-1}\,dx\right| |\boldsymbol{\eta}|_{\beta-m+(1/2)}\right] \\
&\quad \leqq C_0\left[\|p_*^{m-1}\|_{\beta-m} + \|D^{m-1}\|_0 + \frac{\mu(\Omega)}{\mu(\partial\Omega)}|D^{m-1}|\ |\boldsymbol{\eta}|_{\beta-m+(1/2)}\right] \qquad (3.2)\\
&\quad \leqq C_0 C(\Omega)[\|p_*^{m-1}\|_{\beta-m} + |D^{m-1}|] \\
&\quad \leqq C_0 C(\Omega)[\|p_*^{m-1}\|_{\beta-m} + h^{2(\sigma_2-\sigma_1)}(\|p_*^{m-1}\|_1 + \|\mathbf{u}^{m-1}\|_2)],
\end{aligned}$$

where we have used (3.1).

Also by Lemma 1, $(\mathbf{w}^m, q^m) \in [H^{\beta-m}(\Omega)]^N \times H^{\beta-m-1}(\Omega)$ and we have the estimate

$$\begin{aligned}
&\|\mathbf{w}^m\|_{\beta-m} + \|q_*^m\|_{\beta-m-1} \\
&\quad \leqq C_0\left[\frac{(\mu(\Omega))^{1/2}}{\mu(\Omega)}\left|\int_{\partial\Omega}\left(p^{m-1} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\cdot\boldsymbol{\eta}\right)d\Gamma\right| + \left|p^{m-1}\boldsymbol{\eta} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\right|_{\beta-m-(1/2)}\right] \\
&(3.3) \\
&\quad = C_0\left[\frac{(\mu(\Omega))^{1/2}}{\mu(\partial\Omega)}\left|\int_\Omega -p^{m-1}\,dx\right| h^{2(\sigma_1-\sigma_2)} + \left|p^{m-1}\boldsymbol{\eta} - \nu\frac{\partial\mathbf{u}^{m-1}}{\partial\eta}\right|_{\beta-m-(1/2)}\right] \\
&\quad \leqq C_0 C(\Omega)[|D^{m-1}|h^{2(\sigma_1-\sigma_2)} + \|p_*^{m-1}\|_{\beta-m} + |D^{m-1}| + \|\mathbf{u}^{m-1}\|_{\beta-m+1}]
\end{aligned}$$

(using Lemmas 2 and 3)

$$\leqq C_0 C(\Omega)[\|p_*^{m-1}\|_1 + \|\mathbf{u}^{m-1}\|_2 + \|p_*^{m-1}\|_{\beta-m} + \|\mathbf{u}^{m-1}\|_{\beta-m+1}].$$

Hence for $\beta \geqq m+1$,

$$\begin{aligned}
&\|\mathbf{u}^m\|_{\beta-m} + \|p_*^m\|_{\beta-m+1} \\
&\quad \leqq C_0 C(\Omega)[\|p_*^{m-1}\|_{\beta-m} + \|\mathbf{u}^{m-1}\|_{\beta-m+1} + h^{2(\sigma_2-\sigma_1)}(\|p_*^{m-1}\|_1 + \|\mathbf{u}^{m-1}\|_2)] \qquad (3.4)\\
&\quad \leqq C_0 C(\Omega)[\|p_*^{m-1}\|_{\beta-m} + \|\mathbf{u}^{m-1}\|_{\beta-m+1}].
\end{aligned}$$

It easily follows from (3.2) and (3.4) that

$$\begin{aligned}
\|\mathbf{z}^m\|_{\beta-m+1} + \|Q_*^m\|_{\beta-m} &\leqq [C_0 C(\Omega)]^m [\|p_*\|_{\beta-1} + \|\mathbf{u}\|_\beta] \\
&\leqq C\|\mathbf{f}\|_{\beta-2} \qquad \text{(by Lemma 1)}
\end{aligned}$$

and from (3.3) and (3.4) that

$$\begin{aligned}
\|\mathbf{w}^m\|_{\beta-m} + \|q_*^m\|_{\beta-m-1} &\leqq [C_0 C(\Omega)]^m [\|p_*\|_{\beta-1} + \|\mathbf{u}\|_\beta] \\
&\leqq C\|\mathbf{f}\|_{\beta-2}
\end{aligned}$$

again by Lemma 1.

THEOREM 1. *Let* $(\mathbf{u}, p)$ *be the solution of Problem* (P) *with* $\mathbf{f}\in[H^{\beta}(\Omega)]^N$ *and* $\mathbf{u}_h$ *the solution of Problem* (P_h). *Let* $(\mathbf{u}^m, p^m)$, $m\geqq 1$, *be as defined in Lemma* 7. *Then* $\mathbf{u}_h$ *minimizes over* $[S_h^r]^N$ *the functional* (*in* $\boldsymbol{\phi}$)

$$\begin{aligned}T_k(\mathbf{u},\boldsymbol{\psi}_k,\boldsymbol{\phi})=\{a(\boldsymbol{\phi}-\mathbf{u}-\boldsymbol{\psi}_k,\boldsymbol{\phi}-\mathbf{u}-\boldsymbol{\psi}_k)\\ +\gamma h^{-\sigma_1}\|\mathrm{div}\,(\boldsymbol{\phi}-\mathbf{u}-\boldsymbol{\psi}_k-[\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1})\|_0^2\\ +\gamma h^{-\sigma_2}|\boldsymbol{\phi}-\mathbf{u}-\boldsymbol{\psi}_k-[\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}|_0^2\}^{1/2},\end{aligned}$$

where $k\leqq\beta-2$.

Proof. Integrating by parts, we easily see that the exact solution $(\mathbf{u}, p)$ satisfies

$$a(\mathbf{u},\mathbf{v})-(p,\mathrm{div}\;\mathbf{v})+\left\langle p-\nu\frac{\partial\mathbf{u}}{\partial\eta},\mathbf{v}\right\rangle=(\mathbf{f},\mathbf{v})\quad\forall\,\mathbf{v}\in[H^1(\Omega)]^N$$

along with $\mathrm{div}\,\mathbf{u}=0$ in Ω and $\mathbf{u}=0$ on $\partial\Omega$. The approximate solution $\mathbf{u}_h$ satisfies

$$a(\mathbf{u}_h,\mathbf{v}_h)+\gamma h^{-\sigma_1}(\mathrm{div}\,\mathbf{u}_h,\mathrm{div}\,\mathbf{v}_h)+\gamma h^{-\sigma_2}\langle\mathbf{u}_h,\mathbf{v}_h\rangle=(\mathbf{f},\mathbf{v}_h)\quad\forall\,\mathbf{v}_h\in[S_h^r]^N.$$

Hence

$$a(\mathbf{u}_h-\mathbf{u},\mathbf{v}_h)+\gamma h^{-\sigma_1}(\mathrm{div}(\mathbf{u}_h-\mathbf{u}),\mathrm{div}\,\mathbf{v}_h)+\gamma h^{-\sigma_2}\langle\mathbf{u}_h-\mathbf{u},\mathbf{v}_h\rangle+(p,\mathrm{div}\,\mathbf{v}_h)-\left\langle p\boldsymbol{\eta}-\nu\frac{\partial\mathbf{u}}{\partial\eta},\mathbf{v}_h\right\rangle=0\quad\forall\,\mathbf{v}_h\in[S_h^r]^N.$$

Letting $\mathbf{e}=\mathbf{u}_h-\mathbf{u}$, we may rewrite the above as

$$a(\mathbf{e},\mathbf{v}_h)+\gamma h^{-\sigma_1}(\mathrm{div}\,\mathbf{e}+\gamma^{-1}h^{\sigma_1}p,\mathrm{div}\,\mathbf{v}_h)+\gamma h^{-\sigma_2}\left\langle\mathbf{e}-\gamma^{-1}h^{\sigma_2}\left(p\boldsymbol{\eta}-\nu\frac{\partial\mathbf{u}}{\partial\eta}\right),\mathbf{v}_h\right\rangle=0. \tag{3.5}$$

It follows easily from the definitions of $(\mathbf{z}^m, Q^m)$, $(\mathbf{u}^m, q^m)$ that

$$a(\mathbf{z}^m,\mathbf{v})-(Q^m,\mathrm{div}\,\mathbf{v})+\left\langle Q^m\boldsymbol{\eta}-\nu\frac{\partial\mathbf{z}^m}{\partial\eta},\mathbf{v}\right\rangle=0$$

and

$$a(\mathbf{w}^m,\mathbf{v})-(q^m,\mathrm{div}\,\mathbf{v})+\left\langle q^m\boldsymbol{\eta}-\nu\frac{\partial\mathbf{u}^m}{\partial\eta},v\right\rangle=0$$

for all $\mathbf{v}\in[H^1(\Omega)]^N$. Hence from the definition of $(\mathbf{u}^m, p^m)$ we have

$$a(\mathbf{u}^m,\mathbf{v})-(p^m,\mathrm{div}\,\mathbf{v})+\left\langle p^m\boldsymbol{\eta}-\nu\frac{\partial\mathbf{u}^m}{\partial\eta},\mathbf{v}\right\rangle=0.$$

Using the definitions of $\mathbf{z}^m$ and $\mathbf{w}^m$, we get

$$a(\mathbf{u}^m,\mathbf{v})+(\mathrm{div}\,\mathbf{z}^{m+1},\mathrm{div}\,\mathbf{v})+\langle\mathbf{w}^{m+1},\mathbf{v}\rangle=0$$

and hence that

$$a(\mathbf{u}^m, \mathbf{v}) + (\operatorname{div} \mathbf{u}^{m+1}, \operatorname{div} \mathbf{v}) + h^{\sigma_1-\sigma_2}\langle \mathbf{u}^{m+1}, \mathbf{v}\rangle - h^{\sigma_2-\sigma_1}(\operatorname{div} \mathbf{w}^{m+1}, \operatorname{div} \mathbf{v}) - h^{\sigma_1-\sigma_2}\langle \mathbf{z}^{m+1}, \mathbf{v}\rangle = 0.$$

If we use the divergence theorem and condition (D) on p^m, the last two terms cancel, and we are left with

$$a(\mathbf{u}^m, \mathbf{v}) + (\operatorname{div} \mathbf{u}^{m+1}, \operatorname{div} \mathbf{v}) + h^{\sigma_1-\sigma_2}\langle \mathbf{u}^{m+1}, \mathbf{v}\rangle = 0$$

for all $\mathbf{v} \in [H^1(\Omega)]^N$.

Multiplying the above equation by $[\gamma^{-1}h^{\sigma_1}]^m$ and summing from $m=1$, $k-1$, we obtain

$$a\left(\sum_{m=1}^{k-1} [\gamma^{-1}h^{\sigma_1}]^m \mathbf{u}^m, \mathbf{v}\right) + \gamma h^{-\sigma_1}\left(\operatorname{div} \sum_{m=1}^{k-1} [\gamma^{-1}h^{\sigma_1}]^{m+1}\mathbf{u}^{m+1}, \operatorname{div} \mathbf{v}\right) + \gamma h^{-\sigma_2}\left\langle \sum_{m=1}^{k-1} [\gamma^{-1}h^{\sigma_1}]^{m+1}\mathbf{u}^{m+1}, \mathbf{v}\right\rangle = 0$$

or

$$\begin{aligned} &a\left(\sum_{m=1}^{k} [\gamma^{-1}h^{\sigma_1}]^m \mathbf{u}^m, \mathbf{v}\right) + \gamma h^{-\sigma_1}\left(\operatorname{div} \sum_{m=1}^{k} [\gamma^{-1}h^{\sigma_1}]^{m}\mathbf{u}^{m}, \operatorname{div} \mathbf{v}\right) \\ &+ \gamma h^{-\sigma_2}\left\langle \sum_{m=1}^{k} [\gamma^{-1}h^{\sigma_1}]^{m}\mathbf{u}^{m}, \mathbf{v}\right\rangle - a([\gamma^{-1}h^{\sigma_1}]^k\mathbf{u}^k, \mathbf{v}) - (\operatorname{div} \mathbf{u}^1, \operatorname{div} \mathbf{v}) \\ &- h^{\sigma_1-\sigma_2}\langle \mathbf{u}^1, \mathbf{v}\rangle = 0. \end{aligned} \tag{3.6}$$

Now

$$\begin{aligned} &-a([\gamma^{-1}h^{\sigma_1}]^k\mathbf{u}, \mathbf{v}) - (\operatorname{div} \mathbf{u}^1, \operatorname{div} \mathbf{v}) - h^{\sigma_1-\sigma_2}\langle \mathbf{u}^1, \mathbf{v}\rangle \\ &= \gamma h^{-\sigma_1}(\operatorname{div}[\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}, \operatorname{div} \mathbf{v}) \\ &+ \gamma h^{-\sigma_2}\langle [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}, \mathbf{v}\rangle + (p, \operatorname{div} \mathbf{v}) - \left\langle p\boldsymbol{\eta} - \nu\frac{\partial \mathbf{u}}{\partial \eta}, \mathbf{v}\right\rangle \end{aligned} \tag{3.7}$$

(using the definitions of $(\mathbf{u}^m, p^m)$ and condition (D)).

Set $\boldsymbol{\psi}_k = \sum_{m=1}^k [\gamma^{-1}h^{\sigma_1}]^m\mathbf{u}^m$. Then it follows easily from (3.5), (3.6) and (3.7) that

$$a(\mathbf{e} - \boldsymbol{\psi}_k, \mathbf{v}_h) + \gamma h^{-\sigma_1}(\operatorname{div}(\mathbf{e} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}), \operatorname{div} \mathbf{v}_h) + \gamma h^{-\sigma_2}\langle \mathbf{e} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}, \mathbf{v}_h\rangle = 0$$

for all $v_h \in [S_h^r]^N$. This implies that $\mathbf{u}_h$ minimizes over $[S_h^r]^N$ the functional (in $\boldsymbol{\phi}$)

$$\begin{aligned} T_k(\mathbf{u}, \boldsymbol{\psi}_k, \boldsymbol{\phi}) = \{&a(\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k, \boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k) \\ &+ \gamma h^{-\sigma_1}\|\operatorname{div}(\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1})\|_0^2 \\ &+ \gamma h^{-\sigma_2}|\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}|_0^2\}^{1/2}. \end{aligned}$$

THEOREM 2. *Let* $(\mathbf{u}^m, p^m)$, $m = 0, 1, \cdots, k+1$, *and* $\boldsymbol{\psi}_k$ *be as defined in the proof of Theorem* 1. *Suppose* $\mathbf{f} \in [H^{\beta-2}(\Omega)]^N$, $2 < \beta$, $2 \leq r$ *and* $1 \leq k \leq \beta - 2$. *Then* *for* $\sigma_2 = 1+\sigma_1$ *(the "optimal" choice),*

$$\|\mathbf{u}_h - \mathbf{u} - \boldsymbol{\psi}_k\|_1 \leqq Ch^{\mu_k(\sigma_1)}\|\mathbf{f}\|_{\beta-2}, \tag{3.8}$$

where

$$\mu_k(\sigma_1) = \min [s-1-\sigma_1/2,\ \sigma_1(k+1),\ s-\sigma_1/2+(\sigma_1-1)(k+1)],$$

$s = \min(\beta, r)$ *and* C *is a constant independent of* h *and* $\mathbf{f}$ *(but dependent on* γ *and* k*).*

Proof. By the triangle inequality and Lemma 4, we have

$$\|\mathbf{u}_h - \mathbf{u} - \boldsymbol{\psi}_k\|_1 \leqq C\{T_k(\mathbf{u}, \boldsymbol{\psi}_k, \mathbf{u}_h) + [\gamma^{-1}h^{\sigma_1}]^{k+1}|\mathbf{u}^{k+1}|_0\}.$$

From Theorem 1, we have that

$$\|\mathbf{u}_h - \mathbf{u} - \boldsymbol{\psi}_k\|_1 \leqq C\{T_k(\mathbf{u}, \boldsymbol{\psi}_k, \boldsymbol{\phi}) + [\gamma^{-1}h^{\sigma_1}]^{k+1}|\mathbf{u}^{k+1}|_0\}$$

for all $\boldsymbol{\phi} \in [S_h^r]^N$.

Now

$$\begin{aligned} T_k(\mathbf{u}, \boldsymbol{\psi}_k, \boldsymbol{\phi}) \leqq \|\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k\|_E &\\ + [\gamma h^{-\sigma_1}]^{1/2}\|\operatorname{div}(\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1})\|_0 &\\ + [\gamma h^{-\sigma_2}]^{1/2}|\boldsymbol{\phi} - \mathbf{u} - \boldsymbol{\psi}_k - [\gamma^{-1}h^{\sigma_1}]^{k+1}\mathbf{u}^{k+1}|_0. & \end{aligned}$$

Set $\boldsymbol{\phi} = \boldsymbol{\phi}^0 + \sum_{j=1}^{k+1}\{\boldsymbol{\chi}^j + h^{\sigma_2-\sigma_1}\boldsymbol{\xi}^j\}$ where $\boldsymbol{\phi}^0, \boldsymbol{\chi}^j, \boldsymbol{\xi}^j \in [S_h^r]^N$. Using the norm

$$|||\mathbf{v}|||_\gamma = \|\mathbf{v}\|_E + [\gamma h^{-\sigma_1}]^{1/2}\|\operatorname{div}\mathbf{v}\|_0 + [\gamma h^{-\sigma_2}]^{1/2}|\mathbf{v}|_0,$$

the triangle inequality and Theorem 1, we obtain

$$\begin{aligned} \|\mathbf{u}_h - \mathbf{u} - \boldsymbol{\psi}_k\|_1 \leqq C\{&|||\boldsymbol{\phi}^0 - \mathbf{w}^0|||_\gamma \\ &+ [\gamma^{-1}h^{\sigma_1}]^{k+1}(\|\mathbf{u}^{k+1}\|_E + |\mathbf{u}^{k+1}|_0) \\ &+ \sum_{j=1}^{k+1}[\gamma^{-1}h^{\sigma_1}]^j[|||\boldsymbol{\chi}^j - \mathbf{z}^j|||_\gamma + h^{\sigma_2-\sigma_1}|||\boldsymbol{\xi}^j - \mathbf{w}^j|||_\gamma\}. \end{aligned} \tag{3.9}$$

Using Lemmas 1, 2 and 6, we obtain from (3.9)

$$\begin{aligned} \|\mathbf{u}_h - \mathbf{u} - \boldsymbol{\psi}_k\|_1 \leqq \{&[Ch^{s-1} + C\gamma^{1/2}h^{s-1-\sigma_1/2} + C\gamma^{1/2}h^{s-1/2-\sigma_2/2}] \\ &+ C[\gamma^{-1}h^{\sigma_1}]^{k+1} \\ &+ \sum_{j=1}^{k+1}[\gamma^{-1}h^{\sigma_1}]^j([Ch^{s-j} + C\gamma^{1/2}h^{s-j-\sigma_1/2} \\ &\qquad + C\gamma^{1/2}h^{s-j+1/2-\sigma_2/2}] \\ &+ h^{\sigma_2-\sigma_1}[Ch^{s-j-1} + C\gamma^{1/2}h^{s-j-1-\sigma_1/2} \\ &\qquad + C\gamma^{1/2}h^{s-j-1/2-\sigma_2/2}])\}\|\mathbf{f}\|_{\beta-2} \\ \leqq \{&C\gamma^{1/2}h^{s-1-\sigma_1/2}[1 + h^{-(1/2)[\sigma_2-\sigma_1-1]}] + C[\gamma^{-1}h^{\sigma_1}]^{k+1} \\ &+ C\gamma^{1/2}h^{s-\sigma_1/2}[1 + h^{-(1/2)[\sigma_2-\sigma_1-1]}][1 + h^{[\sigma_2-\sigma_1-1]}] \\ &\cdot \sum_{j=1}^{k+1}[\gamma^{-1}h^{(\sigma_1-1)}]^j\}\ \|\mathbf{f}\|_{\beta-2}. \end{aligned} \tag{3.10}$$

Now for $\sigma_2-\sigma_1-1\geqq 0$, the right-hand side of (3.10) is bounded by

$$C\{\gamma^{1/2}h^{s-1-\sigma_1/2-(1/2)(\sigma_2-\sigma_1-1)}+[\gamma^{-1}h^{\sigma_1}]^{k+1}+\gamma^{1/2}h^{s-\sigma_1/2-1/2[\sigma_2-\sigma_1-1]}\sum_{j=1}^{k+1}[\gamma^{-1}h^{\sigma_1-1}]^j\}\|\mathbf{f}\|_{\beta-2}.$$

Hence the best choice of σ_2 is as small as possible, i.e., $\sigma_2=1+\sigma_1$. Now for $\sigma_2-\sigma_1-1\leqq 0$ we have

$$\|\mathbf{u}_h-\mathbf{u}-\boldsymbol{\psi}_k\|_1\leqq C\{\gamma^{1/2}h^{s-1-\sigma_1/2}+[\gamma^{-1}h^{\sigma_1}]^{k+1}+\gamma^{1/2}h^{s-\sigma_1/2+(\sigma_2-\sigma_1-1)}\sum_{j=1}^{k+1}[\gamma^{-1}h^{(\sigma_1-1)}]^j\}\|\mathbf{f}\|_{\beta-2}.$$

Hence the best choice of σ_2 is as large as possible, i.e., σ_1+1.

In either case then, the best choice is $\sigma_2=\sigma_1+1$, for which we obtain:

$$\begin{aligned}\|\mathbf{u}_h-\mathbf{u}-\boldsymbol{\psi}_k\|_1&\leqq C\{\gamma^{1/2}h^{s-1-\sigma_1/2}+[\gamma^{-1}h^{\sigma_1}]^{k+1}\\&\quad+\gamma^{1/2}h^{s-\sigma_1/2}\sum_{j=1}^{k+1}[\gamma^{-1}h^{(\sigma_1-1)}]^j\}\|\mathbf{f}\|_{\beta-2}\\&\leqq C\{\gamma^{1/2}h^{s-1-\sigma_1/2}+[\gamma^{-1}h^{\sigma_1}]^{k+1}\\&\quad+\gamma^{1/2}h^{s-\sigma_1/2}(\gamma^{-1}h^{(\sigma_1-1)}+[\gamma^{-1}h^{(\sigma_1-1)}]^{k+1})\}\|\mathbf{f}\|_{\beta-2}\\&\leqq C\{\gamma^{1/2}h^{s-1-\sigma_1/2}+[\gamma^{-1}h^{\sigma_1}]^{k+1}\\&\quad+\gamma^{1/2}h^{s-\sigma_1/2}[\gamma^{-1}h^{(\sigma_1-1)}]^{k+1}\}\|\mathbf{f}\|_{\beta-2}.\end{aligned}$$

Theorem 2 follows immediately.

Remark. One easily sees that Theorems 1 and 2 hold also when $k=0$ with the interpretation that $\boldsymbol{\psi}_k=0$. Hence with $\sigma_2=1+\sigma_1$, we obtain

$$\|\mathbf{u}_h-\mathbf{u}\|_1\leqq Ch^{\tau}\|\mathbf{f}\|_{\beta-2},$$

where $\tau=\min[s-1-\sigma_1/2,\sigma_1]$ $(s=\min(\beta,r))$. The best choice for σ_1 is $\sigma_1=\frac{2}{3}(s-1)$ for which we obtain

$$\|\mathbf{u}_h-\mathbf{u}\|_1\leqq Ch^{2(s-1)/3}\|\mathbf{f}\|_{\beta-2}.$$

Thus this special case of our theorem gives an order of convergence estimate for the "penalty method" without extrapolation.

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

Let $\gamma_0,\cdots,\gamma_m$ be distinct and choose $a_0,\cdots,a_m$ so that

$$\begin{aligned}&\sum_{i=0}^m a_i=1,\\ &\sum_{i=0}^m a_i\gamma_i^{-j}=0,\qquad 1\leqq j\leqq m.\end{aligned}\tag{3.11}$$

We define $\mathbf{u}_h^{(m)} = \sum_{i=0}^m a_i \mathbf{u}_h^{(0)}(\gamma_i)$ where $\mathbf{u}_h^{(0)}(\gamma)$ is the solution of Problem (P_h) with boundary weights $\gamma h^{-\sigma_1}$, $\gamma h^{-\sigma_2}$. We remark that the coefficients a_i exist and are unique as the system (3.11) is a Vandermonde.

In the case that $\gamma_i = 2^i\gamma$ for some $\gamma > 0$, we can give an explicit definition of the mth extrapolate, $u_h^{(m)}$.

Define

$$\mathbf{u}_h^{(j)}(\gamma) = \frac{2^j \mathbf{u}_h^{(j-1)}(2\gamma) - \mathbf{u}_h^{(j-1)}(\gamma)}{2^j - 1} \qquad j \geqq 1.$$

We can then prove the following:

THEOREM 3. *Assume the hypotheses of Theorem 2 hold. Then*

$$\|\mathbf{u} - u_h^{(m)}(\gamma)\|_1 \leqq Ch^{\mu(\sigma_1)}\|\mathbf{f}\|_{\beta-2},$$

where

$$\mu(\sigma_1) = \max_{0 \leqq k \leqq \min(m,\, \beta-2)} \min[s - 1 - \sigma_1/2,\, \sigma_1(k+1),\, s - \sigma_1/2 + (\sigma_1 - 1)(k+1)]$$

and C is a constant independent of σ_1, h and $\mathbf{f}$ (but dependent on m and γ). We have assumed $\sigma_2 = 1 + \sigma_1$.

Proof. For $0 \leq k \leq \min(m, \beta - 2)$, we have from (3.11) that

$$\begin{aligned}\|\mathbf{u} - \mathbf{u}_h^{(m)}(\gamma)\|_1 &= \left\| \sum_{i=0}^m a_i\Big(\mathbf{u} - \mathbf{u}_h^{(0)}(2^i\gamma) + \sum_{j=1}^k [2^{-i}\gamma^{-1}h^{\sigma_1}]^j \mathbf{u}^j\Big)\right\|_1 \\ &\leqq \sum_{i=0}^m |a_i|\, \|\mathbf{u} - \mathbf{u}_h^0(2^i\gamma) + \boldsymbol{\psi}_k(2^i\gamma)\|_1 \leqq Ch^{\mu_k(\sigma_1)}\|\mathbf{f}\|_{\beta-2}\end{aligned}$$

using (3.8). Hence

$$\|\mathbf{u} - \mathbf{u}_h^{(m)}(\gamma)\|_1 \leqq Ch^{\mu(\sigma_1)}\|\mathbf{f}\|_{\beta-2},$$

where

$$\mu(\sigma_1) = \max_{0 \leqq k \leqq \min(m,\, \beta-2)} \mu_k(\sigma_1).$$

COROLLARY 3.1. *For $0 \leqq \min(m, \beta - 2) \leqq s - 2$ the optimal choice of σ_1 in Theorem 3 is given by*

$$\sigma_1 = \frac{s-1}{k + 3/2}$$

for which we obtain

$$\|\mathbf{u} - u_h^{(m)}(\gamma)\|_1 \leqq Ch^{\mu}\|\mathbf{f}\|_{\beta-2},$$

where

$$\mu = (s-1)(k+1)/(k + \tfrac{3}{2}),$$

$k=\min(m, \beta-2)$, $s=\min(\beta, r)$ and C is a constant independent of h and $\mathbf{f}$ (but dependent on m and γ).

For $s-1\leqq\min(m, \beta-2)$, the optimal choice of σ_1 in Theorem 3 is given by $\sigma_1=(s-1)/s$ for which we obtain

$$\|\mathbf{u}-u_h^{(m)}(\gamma)\|_1\leqq Ch^{\mu}\|\mathbf{f}\|_{\beta-2},$$

where

$$\mu=\frac{(s-1)(2s-1)}{2s}.$$

(Note that this second case can only occur when $\beta\geqq r+1$.)

Proof. From Theorem 3, we have that

$$\|\mathbf{u}-\mathbf{u}_h^{(m)}(\gamma)\|_1\leqq Ch^{\mu(\sigma_1)}\|\mathbf{f}\|_{\beta-2}.$$

Let

$$\mu_k=\max_{\sigma_1\geqq 0}\mu_k(\sigma_1).$$

We first observe that

$$\mu_k\leqq\max_{\sigma_1\geqq 0}\min\,[s-1-\sigma_1/2,\ \sigma_1(k+1)].$$

Since $s-1-\sigma_1/2$ decreases as σ_1 increases and $\sigma_1(k+1)$ increases when σ_1 increases, we have that

$$\max_{\sigma_1\geqq 0}\min\,[s-1-\sigma_1/2,\ \sigma_1(k+1)]=(s-1)(k+1)/(k+\tfrac{3}{2});$$

that is, the maximum occurs when $s-1-(\sigma_1/2)=\sigma_1(k+1)$, i.e.,

$$\sigma_1=\frac{s-1}{k+\frac{3}{2}}.$$

Hence

$$\mu_k\leqq\frac{s-1}{k+\frac{3}{2}}(k+1).$$

We now show μ_k achieves this value for $\sigma_1=(s-1)/(k+\frac{3}{2})$ for all $0\leqq k\leqq s-2$. To do so we need only establish that

$$s-\sigma_1/2+(\sigma_1-1)(k+1)\geqq\frac{(s-1)(k+1)}{k+\frac{3}{2}}$$

for $\sigma_1=(s-1)/(k+\frac{3}{2})$. This will imply that

$$\mu_k\left(\frac{s-1}{k+\frac{3}{2}}\right)=\frac{s-1}{k+\frac{3}{2}}(k+1).$$

Now for $0 \leqq k \leqq s-\frac{3}{2}$, we have $s-1 \geqq k+\frac{1}{2}$ and hence

$$(s-1)(k+1) \geqq (k+\tfrac{1}{2})(k+1) \geqq k(k+\tfrac{3}{2})$$

so that

$$\frac{(s-1)(k+1)}{k+\frac{3}{2}} \geqq k.$$

Now

$$\begin{aligned} s-\sigma_1/2+(\sigma_1-1)(k+1) &= (s-1-\sigma_1/2)+\sigma_1(k+1)-k \\ &= 2\left[\frac{(s-1)(k+1)}{k+\frac{3}{2}}\right]-k \qquad \left(\text{for } \sigma_1 = \frac{s-1}{k+\frac{3}{2}}\right) \\ &\geqq \frac{(s-1)(k+1)}{k+\frac{3}{2}} \qquad \text{by the above.} \end{aligned}$$

Let $\mu = \max_{0 \leqq k \leqq \min(m,\, \beta-2)} \mu_k$. Clearly μ_k increases with increasing k. Hence for $0 \leqq \min(m, \beta-2) \leqq s-2$, the optimal choice is

$$\sigma_1 = \frac{s-1}{k+\frac{3}{2}} \quad \text{and} \quad \mu = \frac{s-1}{k+\frac{3}{2}}(k+1),$$

where $k = \min(m, \beta-2)$.

Now for $s-1 \leqq k \leqq \beta-2$ we have $s-1-k \leqq 0$ so that $\frac{1}{2}\sigma_1 \geqq s-1-k$ and hence $\sigma_1(k+1) \geqq s-1-k+\sigma_1(k+\frac{1}{2})$. Hence $\mu_k(\sigma_1) = \min[s-1-\sigma_1/2,\ s-\sigma_1/2+(\sigma_1-1)(k+1)]$. This minimum occurs when $s-1-\sigma_1/2 = s-\sigma_1/2+(\sigma_1-1)(k+1)$, i.e., $\sigma_1 = k/(k+1)$ for which we obtain

$$\mu_k\left(\frac{k}{k+1}\right) = s-1-\frac{k}{2(k+1)}.$$

Since μ_k decreases with increasing k in this range, to determine μ and the optimal σ_1, when $s-1 \leqq \min(m, \beta-2)$, we need only compare the two choices $k = s-2$, $\sigma_1 = (s-1)/(k+\frac{3}{2})$, $\mu_k = [(s-1)/(k+\frac{3}{2})](k+1)$ and $k = s-1$, $\sigma_1 = k/(k+1)$, $\mu_k = s-1-k/(2(k+1))$. The first choice of k gives $\sigma_1 = (s-1)/(s-\frac{1}{2})$, $\mu_k = [(s-1)/(s-\frac{1}{2})](s-1)$. The second choice gives $\sigma_1 = (s-1)/s$, $\mu_k = s-1-(s-1)/2s$. Now $2s^2-2s \leqq 2s^2-2s+\frac{1}{2}$ so that $2s(s-1) \leqq (s-\frac{1}{2})(2s-1)$ and hence $(s-1)(s-1)/(s-\frac{1}{2}) \leqq (s-1)(2s-1)/2s = (s-1)-(s-1)/2s$. Hence for $s-1 \leqq \min(m, \beta-2)$, the optimal choice of σ_1 is $(s-1)/s$ with the corresponding $\mu = (s-1)(2s-1)/2s$.

From Corollary 3.1, one observes that the optimal choice of σ_1 depends on the regularity of the solution. We now examine the question of whether the convergence of the method is affected if one incorrectly assumes too high a degree of regularity for the solution and bases the choice of σ_1 on that assumption.

Suppose then, that $\mathbf{u} \in [H^\beta(\Omega)]^N$, $2 \leqq \beta$, but we assume incorrectly that $\mathbf{u} \in [H^\rho(\Omega)]^N$, $\rho > \beta$. Further suppose we do m extrapolations, choosing the optimal σ_1 for our mistaken regularity assumption; i.e., for $0 \leqq \min(m, \rho-2) \leqq$

$\min(\rho, r)-2$, we choose

$$\sigma_1 = \frac{\min(\rho, r)-1}{\min(m, \rho-2)+\frac{3}{2}}, \tag{3.12}$$

and for $r-1 \leqq \min(m, \rho-2)$, we choose

$$\sigma_1 = \frac{r-1}{r}. \tag{3.13}$$

By Theorem 3,

$$\|\mathbf{u}-\mathbf{u}_h^{(m)}(\gamma)\|_1 \leqq Ch^{\mu(\sigma_1)}\|\mathbf{f}\|_{\beta-2},$$

where $\mu(\sigma_1) = \max_{0\leqq k\leqq \beta-2} \mu_k(\sigma_1)$ and

$$\mu_k(\sigma_1) = \min[s-1-\sigma_1/2, \sigma_1(k+1), s-\sigma_1/2+(\sigma_1-1)(k+1)], \qquad s = \min(\beta, r).$$

We first note that for $\beta \geqq r \geqq 2$, it follows easily that using the σ_1 described above, $\mu_0(\sigma_1)$ and hence $\mu(\sigma_1)>0$. The troublesome case occurs for $2\leqq \beta < r$. In this case, $\mu_k(\sigma_1) = \min[\beta-1-\sigma_1/2, \sigma_1(k+1), \beta-\sigma_1/2+(\sigma_1-1)(k+1)]$. Since $\beta-\sigma_1/2+(\sigma_1-1)(k+1) = \beta-k-1+\sigma_1(k+\frac{1}{2})$ and $0\leqq k\leqq \beta-2$, $\mu_k(\sigma_1)$ will be >0 if $\beta-1-(\sigma_1/2)>0$, i.e., if $\sigma_1<2(\beta-1)$.

If σ_1 is chosen equal to $(r-1)/r$ (the case where $r-1\leqq \min(m, \rho-2)$), then clearly $\sigma_1 < 1 \leqq 2(\beta-1)$.

For $0\leqq \min(m, \rho-2) \leqq \min(\rho, r)-2$, we had

$$\sigma_1 = \frac{\min(\rho, r)-1}{\min(m, \rho-2)+\frac{3}{2}}.$$

Hence we require

$$\frac{\min(\rho, r)-1}{\min(m, \rho-2)+\frac{3}{2}} < 2(\beta-1).$$

Here the worst case occurs when $m\leqq r-2 \leqq \rho-2$. We then require $(r-1)/(m+\frac{3}{2}) \leqq 2(\beta-1)$. For $\beta-2$ this will be assured if $m+\frac{3}{2} > (r-1)/2$, i.e., $m>(r-4)/2$.

An example of the problem discussed above occurs when $\beta=2$, but $\rho=r=4$; i.e., we are using piecewise cubics and mistakenly assuming $u\in[H^4(\Omega)]^N$ instead of the correct regularity $[H^2(\Omega)]^N$. If we plan to do no extrapolations, then choosing σ_1 by (3.12), we would have $\sigma_1=2$, and hence would not be guaranteed convergence of the method since $s-1-\sigma_1/2=0$. However, if we had decided to do one extrapolation and again chose σ_1 by (3.12), then we would have $\sigma_1=\frac{6}{5}$ and convergence would be guaranteed.

Summing up, we are guaranteed convergence as long as

$$\sigma_1 < 2[\min(\beta, r)-1].$$

Since $\beta \geqq 2$ and $r \geqq 2$, choosing $\sigma_1 < 2$ will be sufficient. If σ_1 is chosen according to the optimal formulas (3.12) and (3.13) based on the regularity and number of extrapolations performed, then no matter what regularity we incorrectly assume for the solution, we are guaranteed convergence provided we base the choice of σ_1 on doing more than $(r-4)/2$ extrapolations.

Examples for Corollary 3.1 are given by the following.

For $\beta = 4$ and using a subspace of type $[S_h^4]^N$ (e.g., piecewise cubics) we obtain for $k = 0, 1$ and 2 and the corresponding optimal choices of σ_1, the following error bounds for $\|u - u_h^{(k)}(\gamma)\|_1$ of the form $Ch^{\mu}\|\mathbf{f}\|_{\beta-2}$.

k	σ_1	μ
0	2	2
1	$\frac{6}{5}$	$\frac{12}{5}$
2	$\frac{6}{7}$	$\frac{18}{7}$

For $\beta \geqq 5$ one additional extrapolate gives improvement

$$k = 3, \quad \sigma_1 = \tfrac{3}{4}, \quad \mu = \tfrac{21}{8}.$$

Thus extrapolation gives an improvement in the order of accuracy and also, since σ_1 is decreasing, in the condition number of the matrix used to compute the approximate solution. (Using standard techniques it is easy to show that the condition number is $O(h^{-2-\sigma_1})$ for $\sigma_2 = 1 + \sigma_1$.)

We also remark that it is shown in Falk [5] that if the trial functions satisfy the boundary conditions, then additional extrapolations can be used to further increase the order of accuracy and further lower the condition number.

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

THEOREM 4. *Suppose* $\mathbf{f} \in [H^{\beta-2}(\Omega)]^N$, $2 \leq \beta$, $2 \leqq r$, *and* $0 \leqq t \leqq \beta - 2$. *Then for* $\sigma_2 = 1 + \sigma_1$, $\alpha \geqq 2$,

$$\|\mathbf{u} - \mathbf{u}_h^t(\gamma)\|_{2-\alpha} \leqq C[h^{\lambda_t + \lambda_l} + h^{\sigma_1(t+1)}]\|\mathbf{f}\|_{\beta-2},$$

where $\lambda_t = \min[s - 1 - (\sigma_1/2), \sigma_1(t+1), s - (\sigma_1/2) + (\sigma_1 - 1)(t+1)]$, $s = \min(r, \beta)$ *and* $\lambda_l = \min[\min(\alpha, r) - 1 - (\sigma_1/2), \quad \sigma_1(l+1), \quad \min(\alpha, r) - (\sigma_1/2) + (\sigma_1 - 1)\cdot(l+1)]$, $0 \leqq l \leqq \alpha - 2$.

Proof. The proof follows in a manner analogous to that of Theorem 2 by introducing a sequence of auxiliary generalized Stokes problems and using the orthogonality conditions derived in Theorem 1. We omit the technical details.

Remark. For $0 \leqq k \leqq s - 2$, if $\sigma_1 = (s-1)/(k + \frac{3}{2})$, (the optimal choice for the $\|\cdot\|_1$ norm estimate) then there will be no improvement in accuracy in lower norms unless $t > k$.

As an example, let us consider the error in the L^2-norm ($\alpha = 2$) using a subspace of type $[S_h^4]^N$ (e.g., cubic splines). Choosing σ_1 as the optimal choice for 1-norm error estimates, we get the following error bounds for $\|\mathbf{u} - \mathbf{u}_h^t(\gamma)\|_0$ of the form $Ch^{\delta}\|\mathbf{f}\|_{\beta-2}$.

k	β	t	σ_1	δ
0	$\geqq 4$	0	2	2
1	$\geqq 4$	1	$\frac{6}{5}$	$\frac{12}{5}$
2	$\geqq 4$	2	$\frac{6}{7}$	$\frac{18}{7}$
3	$\geqq 5$	3	$\frac{3}{4}$	3
3	$\geqq 6$	4	$\frac{3}{4}$	3
2	$\geqq 5$	3	$\frac{6}{7}$	$\frac{22}{7}$

4. Comments. The extrapolated penalty method presented here for obtaining approximations to the stationary Stokes equations has the following features:

(i) The trial functions are not required to satisfy any auxiliary conditions such as div $\mathbf{v}=0$ or $\mathbf{v}=\mathbf{0}$ on $\partial\Omega$.

(ii) By using extrapolation, one can achieve higher accuracy using matrices with lower condition numbers than arise in the "simple" penalty method.

(iii) 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 "simple" penalty method.

REFERENCES

[1] A. K. AZIZ, ed., *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, Academic Press, New York, 1972.

[2] I. BABUŠKA, *The finite element method with penalty*, Math. Comp., 27 (1973), pp. 221–228.

[3] J. H. BRAMBLE AND V. THOMÉE, *Semidiscrete least-squares methods for a parabolic boundary value problem*, Ibid., 26 (1972), pp. 633–648.

[4] M. CROUZEIX AND P. A. RAVIART, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations I*, Rev. Française Automat. Informat. Recherche Opérationelle, 7 (1973), pp. 33–76.

[5] R. S. FALK, *An analysis of the penalty method and extrapolation for the stationary Stokes equations*, Advances in Computer Methods for Partial Differential Equations, R. Vichnevetsky, ed., L'Association Internationale pour le Calcul Analogique, 1975, pp. 66–69.

[6] ——— *A finite element method for the stationary Stokes equations using trial functions which do* not *have to satisfy div* $\boldsymbol{v}=0$, Math. Comp., (1976), to appear.

[7] J. T. KING, *New error bounds for the penalty method and extrapolation*, Numer. Math, 23 (1974), pp. 153–166.

[8] O. A. LADYZHENSKAYA, *The Mathematical Theory of Viscous Incompressible Flow*, Gordon and Breach, New York, 1962.

[9] J. L. LIONS AND E. MAGENES, *Problèmes aux limites non homogènes et applications*, vol. 1, Dunod, Paris, 1968.

[10] R. TEMAM, *On the theory and numerical analysis of the Navier–Stokes equations*, Lecture Notes #9, University of Maryland, College Park, 1973.