A Least-Squares Finite-Element Method for the Stokes Equations with Improved Mass Balances

X. YE
Department of Mathematics and Statistics
University of Arkansas at Little Rock
Little Rock, AR 72204, U.S.A.

(Received and accepted July 1998)

Abstract—We prove the convergence of a least-squares mixed finite-element method for the Stokes equations with zero residual of mass conservation. For the standard least squares mixed finite-element method, the equations for continuity of mass, and momentum are minimized in a global sense. Therefore, the mass may not be conserved at every point of the discretization. Recently, in [1], a modified least squares finite-element method is developed to enforce near zero residual of mass conservation. This is achieved by attaching a discrete divergence free constrain to the standard least squares finite-element method, and as a consequence, the number of equations is increased. In this paper, we take a different approach to improve the conservation of mass and reduce the number of the equations. This method does not require LBB condition on the finite-dimensional subspaces and the resulting bilinear form is symmetric and positive definite. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords—Least squares, Finite element, Stokes equations.

1. INTRODUCTION

In this paper, we consider a least-squares mixed finite-element formulation for the Stokes equations with zero residual of mass conservation. In the classical mixed formulation of Stokes equations [2], the Ladyzhenskaya-Babuska-Brezzi (LBB) [3,4] consistency condition needs to be satisfied, which translates to a compatibility condition for the velocity space and the pressure space. This restricts the choice of finite-element spaces, some of the best known and widely used finite-element spaces are excluded. The least squares mixed finite-element formulations have the advantage of not being restrained to the LBB condition. This advantage makes it more flexible to choose the finite-element spaces for the unknowns. The least squares methods have been intensively studied recently for elliptic boundary value problems of order 2m [5–7], the Stokes equations [8–10], and the Navier-Stokes equations [11,12]. The standard least squares mixed finite-element methods minimize the equations for continuity of mass and momentum in a global sense. Therefore, the mass may not be conserved at every point of the discretization. In [1], a numerical experiment has been conducted using the standard least squares finite-element method to simulate a cylinder moving along the centerline of a narrow channel. It is observed that mass...
is not conserved in each element, especially in the areas with large gradients of the variables. These areas are the places of most interest. To improve the mass conservation of the solutions, the standard least squares finite-element method is modified [1] for the Stokes equations such that the method nearly conserves the mass at every point. This has been done by adding an extra restriction \( \int_{\Omega} \nabla \cdot u = 0 \) to the standard least squares finite-element method to ensure that the equation for conservation of mass is satisfied in every element in an average sense. For this modified method, the cost to improve the conservation of mass is increasing the number of unknowns and equations. For the numerical example in [1], the modified method adds 2262 more equations to the original system of 11125 equations. In this paper, we develop a method in which the solutions are discrete divergence free including satisfying the basic requirement \( \int_{\Omega} \nabla \cdot u = 0 \) and, at the same time, the number of unknowns is reduced compared with the resulting system from the standard least squares finite-element method. This can be achieved by approximating the velocity from a subspace of the standard finite-element space, in which all the functions are discrete divergence free.

2. THE LEAST SQUARES MIXED FORMULATION

Let \( \Omega \) be a bounded domain in \( \mathbb{R}^2 \) with smooth boundary \( \Gamma \). We consider the Stokes equations

\[
\begin{align*}
-\Delta u + \nabla p &= f, \quad \text{in } \Omega, \\
-\nabla \cdot u &= 0, \quad \text{in } \Omega, \\
u &= 0,
\end{align*}
\]

where \( u = (u_1, u_2) \) denotes the velocity, \( p \) the pressure, and \( f \) the body force. In order to formulate the least squares approximation for (2.1)-(2.3), it is necessary to cast the Stokes equations into a first-order system. Here we use the approach in [10] to rewrite the equations (2.1)-(2.3) to a first-order system. For completeness, we restate some useful abstract results in [10].

**Problem Q.** Let \( X, P, \) and \( Q \) be Hilbert spaces such that \( P = P^* \) and \( Q = Q^* \). Let \( D \in L(X, Q) \) and \( B \in L(X, P) \). For a given \( f \in X^* \), find \((u, p, q) \in X \times P \times Q\) such that

\[
\begin{align*}
q &= Du, \\
D^*q + B^*p &= f, \\
Bu &= 0.
\end{align*}
\]

In order to numerically implement the least squares principle for Problem Q, it is necessary to introduce another Hilbert space \( N \) such that

\[
N \hookrightarrow X^*,
\]

and the norm of \( N \) is easy to compute. In applications, the norms of \( P \) and \( Q \) are usually simple to compute while the norm of \( X^* \) is very complicated. Let

\[
W = \{(u, p, q); (u, p, q) \in X \times P \times Q, D^*q + B^*p \in N\}
\]

equipped with the norm

\[
\| (u, p, q) \|_W = \left( \| u \|_X^2 + \| p \|_P^2 + \| q \|_Q^2 + \| D^*q + B^*p \|_N^2 \right)^{1/2}.
\]

We introduce the bounded bilinear form \( a \) on \( W \times W \) via

\[
a(u, p, q; \tilde{u}, \tilde{p}, \tilde{q}) \equiv (q - Du, \tilde{q} - D\tilde{u})_Q + (D^*q + B^*p, D^*\tilde{q} + B^*\tilde{p})_N + (Bu, B\tilde{u})_P.
\]

Then, the following result holds.
THEOREM 2.1. Suppose there exists a constant $c > 0$ such that
\begin{align}
\|D u\|_Q &\geq c\|u\|_X, \quad \forall u \in X, \\
\|B^* p\|_X &\geq c\|p\|_P, \quad \forall p \in P.
\end{align}

Then there exists a constant $\gamma > 0$ such that
\begin{align}
a(u, p, q; u, p, q) &\geq \gamma \| (u, p, q) \|^2_W, \quad \forall (u, p, q) \in W.
\end{align}

PROOF. See Theorem 2.1 and Corollary 2.1 in [10].

As an application of the abstract Problem Q, the Stokes equations (2.1) and (2.2) can be cast into the first order system
\begin{align}
q &= \nabla u, \\
-\nabla \cdot q + \nabla p &= f, \\
-\nabla \cdot u &= 0,
\end{align}
where
\begin{align}
D u &= \nabla u = \left( \frac{\partial u_j}{\partial x_1} \right)_{2 \times 2}, \\
D^* q &= -\nabla \cdot q = -\left( \sum_{j=1}^{2} \frac{\partial q_{x_j}}{\partial x_j} \right)_{2 \times 1},
\end{align}
and $B u = -\nabla \cdot u, B^* p = \nabla p.$

Let
\begin{align}
X &= [H^1_0(\Omega)]^2, \quad Q = [L^2(\Omega)]^4, \quad N = [L^2(\Omega)]^2,
\end{align}
and chose the space $P$ with special care such that
\begin{align}
P &= \left\{ p - \frac{1}{|\Omega|} \int_\Omega p \, dx ; p \in L^2(\Omega) \right\}.
\end{align}

Let
\begin{align}
W &= \{ (u, p, q) ; (u, p, q) \in X \times P \times Q, -\nabla \cdot q + \nabla p \in N \}
\end{align}
equipped with the norm
\begin{align}
\|(u, p, q)\|_W &= \left( \|u\|^2_1 + \|p\|^2_0 + \|q\|^2_0 + \|-\nabla \cdot q + \nabla p\|^2 \right)^{1/2},
\end{align}
where $\| \cdot \|_i$ is the standard Sobolev norm for $H^1(\Omega)$ with $H^0(\Omega) = L^2(\Omega)$.

Then the least squares variational problem for (2.1)–(2.3) is as follows.

Find $U = (u, p, q) \in W$ such that
\begin{align}
a(U, V) &= F(V), \quad \forall V = (\tilde{u}, \tilde{p}, \tilde{q}) \in W,
\end{align}
where
\begin{align}
a(U, V) &= (q - \nabla u, \tilde{q} - \nabla \tilde{u}) + (-\nabla \cdot q + \nabla p, -\nabla \cdot \tilde{q} + \nabla \tilde{p}) + (\nabla \cdot u, \nabla \cdot \tilde{u}), \\
F(V) &= (f, -\nabla \cdot \tilde{q} + \nabla \tilde{p}),
\end{align}
and $(f, q) = \int_\Omega f g.$
COROLLARY 2.2. For the least squares formulation of the Stokes equations defined in (2.21), the inequalities (2.11) and (2.12) are satisfied. Therefore, there exists a constant \( \gamma \) such that
\[
a(u, p, q; u, p, q) \geq \gamma \| (u, p, q) \|_{W}^{2}, \quad \forall (u, p, q) \in W.
\]
(2.22)

PROOF. See [10].

To discretize (2.21), let \( T_h \) be a partition of \( \Omega \) into finite elements, i.e., \( \Omega = \bigcup_{i \in T_h} \Omega_i \) with \( h = \max \{ \text{diam}(\Omega_i) : \Omega_i \in T_h \} \). Here we use a quadrilateral element. Associated with \( T_h \), let \( X_h \) be a finite-element subspace of \( X \) consisting of continuous piecewise biquadratic polynomials and, \( P_h \) and \( Q_h \) be finite-element subspaces of \( P \) and \( Q \), respectively, consisting of, at least, continuous piecewise linear polynomials. Let \( W_h \) be a subspace of \( W \):
\[
W_h = \{ (u_h, p_h, q_h) : (u_h, p_h, q_h) \in X_h \times P_h \times Q_h, \ -\nabla \cdot q + \nabla p \in N \}.
\]
(2.23)

Then the standard least squares finite-element approximation to (2.21) is: find \( U_h = (u_h, p_h, q_h) \in W_h \) such that
\[
a(U_h, V) = F(V), \quad \text{for all } V = (\bar{u}, \bar{p}, \bar{q}) \in W_h.
\]
(2.24)

3. ZERO RESIDUAL OF MASS CONSERVATION

For the least squares finite-element formulation (2.24), the solutions may not conserve mass well as observed in [1]. To improve the mass conservation, an extra condition
\[
\int_{\Omega_i} \nabla \cdot u = 0, \quad i = 1, \ldots M,
\]
(3.1)
where \( M \) is the total number of elements, is attached to the system resulting from the standard least squares finite-element mixed formulation [7]. The disadvantage of doing this is increasing the number of equations.

Let \( \Lambda_h \) be a space containing continuous piecewise bilinear polynomials associated with triangulation \( T^h \). Define
\[
D^h = \{ v \in X^h \mid (\nabla \cdot v, \lambda) = 0, \lambda \in \Lambda_h \}.
\]
All the functions in \( D^h \) are discrete divergence free and, of course, satisfy (3.1). To ensure the solutions (2.1)-(2.3) are discrete divergence free, we redefine the solutions space \( W_h \subset W \):
\[
W_h = \{ (u_h, p_h, q_h) : (u_h, p_h, q_h) \in D^h \times P^h \times Q^h, \ -\nabla \cdot q + \nabla p \in N \}.
\]
(3.2)

Then, the standard least squares finite-element formulation (2.24) can be changed to: find \( U_h = (u_h, p_h, q_h) \in W_h \) such that
\[
a(U_h, V) = F(V), \quad \text{for all } V = (\bar{u}, \bar{p}, \bar{q}) \in W_h.
\]
(3.3)

Since \( D^h \) is a subspace of \( X^h \), system (3.3) is smaller and mass is conserved in an average sense. The price for these advantages is that the condition number of the smaller system (3.3) is worse. To be more specific, the condition number of the original least square system (2.24) is equivalent to second-order problem and the condition number of (3.3) is equivalent to bihomonic problem. However, with the development of multigrid and domain decomposition techniques, one can solve second-order or bihomonic problems with the convergence rate independent of the condition number. To implement (3.3), an explicit basis for \( D^h \) is needed. In the following, we
will illustrate how to obtain a basis of $D^h$ for a uniform mesh on a rectangular domain $\Omega$. This result can be extended to the nonuniform mesh (see [13]).

For the mesh shown in Figure 1, a piecewise biquadratic shape function $\phi_i$ is associated with an interior node $i$. The number of $\phi_i$ is the same as the number of interior nodes which is $l = (2m - 1)(2n - 1)$.

We choose the following basis for the finite-element velocity space $X^h$:

$$\Phi_1 = (\phi_1, 0)^T, \Phi_2 = (0, \phi_1)^T, \ldots, \Phi_{L-1} = (\phi_1, 0)^T, \Phi_L = (0, \phi_1)^T. \quad (3.4)$$

The simple fact is that any element in the vector space $X^h$ (in particular, those in $D^h$) can be expressed as a linear combination of the basis functions $\Phi_i$, $i = 1, \ldots, L$. Therefore, the basis functions of the discrete divergence free subspace $D^h$ can be found if the appropriate coefficients of this combination can be found.

To derive a basis of $D^h$ for any $m \times n$ mesh, the $2 \times 2$ mesh is considered first. A $2 \times 2$ mesh and the order of the nodes for velocity are shown in Figure 2. For the $2 \times 2$ mesh, there are nine interior nodes and $X^h = \text{span}\{\Phi_1, \Phi_2, \ldots, \Phi_{18}\}$, where $\Phi_1, \Phi_2, \ldots, \Phi_{18}$ are defined in (3.4).

The following ten functions form a basis of $D^h$ for the $2 \times 2$ mesh (see [14]),

$$\begin{align*}
\Psi_1 &= \Phi_{13} - \Phi_{15} + \Phi_{17}, \\
\Psi_2 &= \Phi_1 - \Phi_3 + \Phi_5, \\
\Psi_3 &= \Phi_2 - \Phi_4 + \Phi_{14}, \\
\Psi_4 &= \Phi_6 - \Phi_{12} + \Phi_{18}, \\
\Psi_5 &= -\Phi_1 + \Phi_2 - 4\Phi_9 + 4\Phi_{10} - \Phi_{17} + \Phi_{18}, \\
\Psi_6 &= \Phi_5 + \Phi_6 + 4\Phi_9 + 4\Phi_{10} + \Phi_{13} + \Phi_{14}, \\
\Psi_7 &= \Phi_{11} - \Phi_{16} - 2\Phi_{17} + 2\Phi_{18}, \\
\Psi_8 &= 2\Phi_1 - 2\Phi_2 + \Phi_4 - \Phi_7, \\
\Psi_9 &= -\Phi_4 + 2\Phi_5 + 2\Phi_6 - \Phi_{11}, \\
\Psi_{10} &= -\Phi_7 + 2\Phi_{13} + 2\Phi_{14} - \Phi_{16}. 
\end{align*} \quad (3.5)$$
The basis vector function $\Psi_i$ has nonzero linear combination coefficients only for certain $\Phi_j$s. In Figure 3, for each $\Psi_i$, the nodes are marked if the coefficients are nonzero.

Since any $m \times n$ mesh contains many $2 \times 2$ submeshes, the divergence free functions $\Psi_1, \Psi_2, \ldots, \Psi_{10}$ in (3.5) for a $2 \times 2$ mesh or macro element can be used as blocks to build a divergence free basis of $D^h$ for an $m \times n$ mesh. If we generate ten functions from (3.5) for each $2 \times 2$ macro element in an $m \times n$ mesh, they will be linearly dependent. In general, to construct a basis for $D^h$ is not an easy task. We will illustrate how to derive a basis of $D^h$ for a $4 \times 4$ mesh. For an $m \times n$ mesh, more details can be found in [13].

For the $4 \times 4$ mesh in Figure 4, there are 49 interior nodes for velocity marked in Figure 3. The dimension of $D^h$ is 74. These 74 basis functions of $D^h$ can be obtained as shown in Figure 5-7.
4. ERROR ESTIMATE

Since $X^h$, $P^h$, and $Q^h$ are subspaces consisting of, at least, continuous piecewise bilinear polynomials, then, from the approximation theorem [14],

\[
\inf_{\tilde{u} \in X^h} \|u - \tilde{u}\|_1 \leq \|u - u_I\|_1 \leq c h^k \|u\|_{k+1, \Omega},
\]
\[
\inf_{\tilde{p} \in P^h} \|p - \tilde{p}\|_1 \leq \|p - p_I\|_1 \leq c h^k \|p\|_{k+1, \Omega},
\]

and

\[
\inf_{\tilde{q} \in Q^h} \|q - \tilde{q}\|_1 \leq \|q - q_I\|_1 \leq c h^k \|q\|_{k+1, \Omega},
\]

where $u_I \in X^h$, $p_I \in P^h$, and $q_I \in Q^h$ are the standard finite-element interpolants of $u$, $p$, and $q$, respectively, and $u$, $p$, and $q$ are in $H^2(\Omega)$.

Subtracting (3.3) from (2.21), the error has the orthogonality property

\[
a(U - U^h, V) = 0, \quad \forall V = (\tilde{u}, \tilde{p}, \tilde{q}) \in W^h.
\]

Then, we have the following theorem.

**Theorem 4.1.** Let $U = (u, p, q) \in W$ and $U^h = (u^h, p^h, q^h) \in W^h$ be the solutions of (2.21) and (3.3), respectively. Then

\[
\|u - u^h\|_1 + \|p - p^h\|_0 + \|q - q^h\|_0 \leq c h (\|u\|_2 + \|p\|_2 + \|q\|_2).
\]
PROOF. For $V = (\tilde{u}, \tilde{p}, \tilde{q}) \in W^h$, using the inequality (2.22) and orthogonality relation (4.4),

$$\| U - U^h \|^2 \leq C a (U - U^h, U - U^h)$$

$$= C a (U - U^h, U - V) + C a (U - U^h, V - U^h)$$

$$= C \{ (q - q^h, q - q) + \nabla \cdot (p - p^h), -\nabla \cdot (q - q) + \nabla \cdot (p - p) \}$$

$$\leq C (\| q - q^h \|_0 \| q - q \|_0 + \| q - q^h \|_0 \| u - \tilde{u} \|_1$$

$$+ \| u - u^h \|_1 \| q - q \|_0 + \| u - u^h \|_1 \| u - \tilde{u} \|_1$$

$$+ \| -\nabla \cdot (q - q^h) + \nabla \cdot (p - p^h) \|_0 \| -\nabla \cdot (q - q) + \nabla \cdot (p - p) \|_0$$

$$+ \| u - u^h \|_1 \| u - \tilde{u} \|_1)$$

$$\leq C (\| q - q^h \|^2 + \| u - u^h \|^2 + \| -\nabla \cdot (q - q^h) + \nabla \cdot (p - p^h) \|^2)^{1/2}$$

$$\leq C \| U - U^h \|_W \| U - V \|_W.$$  (4.5)

Then,

$$\| u - u^h \|_1 + \| p - p^h \|_0 + \| q - q^h \|_0 \leq C \{ \inf_{\tilde{u} \in D^h} \| u - \tilde{u} \|_1 + \inf_{\tilde{p} \in P^h} \| p - \tilde{p} \|_1$$

$$\| q - q \|_0 + \| q - q^h \|_0$$

$$\leq C \left\{ \inf_{\tilde{u} \in D^h} \| u - \tilde{u} \|_1 + \inf_{\tilde{p} \in P^h} \| p - \tilde{p} \|_1$$

$$+ \inf_{\tilde{q} \in Q^h} \| q - \tilde{q} \|_1 \right\}.$$  (4.6)

It is well known that for the quadrilateral element, piecewise biquadratic polynomial space $X^h$ and piecewise bilinear polynomial space $\Lambda^h$ satisfy the Ladyzhenskaya-Babuska-Brezzi (LBB) consistency condition, i.e.,

$$\sup_{v \in X^h} \frac{\langle \nabla \cdot v, \mu \rangle}{\| v \|_1} \leq \beta^* \| \mu \|_0, \quad \forall \mu \in \Lambda^h.$$  (4.7)

By Theorem 1.1 in [2],

$$\inf_{w \in D^h} \| u - w \|_1 \leq c \inf_{v \in X^h} \| u - v \|.$$  (4.8)

From (4.1)-(4.3) and (4.7), (4.6) becomes

$$\| u - u^h \|_1 + \| p - p^h \|_0 + \| q - q^h \|_0 \leq C h (\| u \|_2 + \| p \|_2 + \| q \|_2) \leq C h.$$  (4.9)

This completes the proof.

REFERENCES