Approximation of generalized and steady Stokes problems using dual-mixed finite elements without enrichment

Jason S. Howell

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213-3890, USA. email: howell4@andrew.cmu.edu

SUMMARY

In this work a finite element method for a dual-mixed approximation of Stokes and generalized Stokes problems in two or three space dimensions is studied. A variational formulation of the generalized Stokes problems is accomplished through the introduction of the pseudostress and the trace-free velocity gradient as unknowns, yielding a twofold saddle point problem. The method avoids the explicit computation of the pressure, which can be recovered through a simple post-processing technique. Compared to an existing approach for the same problem, the method presented here reduces the global number of degrees of freedom by up to one third in two space dimensions. The method here also represents a connection between existing dual-mixed and pseudostress methods for Stokes problems. Existence, uniqueness, and error results for the generalized Stokes problem are given, and two new penalty methods for approximating the steady Stokes problem are derived. Numerical experiments that illustrate the theoretical results are presented.

KEY WORDS: Stokes problem, generalized Stokes problem, twofold saddle point problem, penalty method Raviart-Thomas, pseudostress, finite element method, dual-mixed method

1. INTRODUCTION

In this article a dual-mixed formulation and corresponding finite element approximation of generalized Stokes and steady Stokes problems is studied. The generalized Stokes problem with pure Dirichlet boundary conditions on the velocity is given by: Find \((\sigma, u, p)\) such that

\[
\sigma = \nu(\nabla u + (\nabla u)^T) - pI
\]

\(a u - \nabla \cdot \sigma = f \quad \text{in } \Omega,
\]

\(\nabla \cdot u = 0 \quad \text{in } \Omega,
\]

\(u = u_\Gamma \quad \text{on } \Gamma,
\]
where $\Omega$ is a bounded open subset of $\mathbb{R}^d$ with Lipschitz continuous boundary $\Gamma$. The fluid velocity is denoted by $\mathbf{u}$, and $\nabla \mathbf{u} := (\nabla \mathbf{u})_{ij} = \partial u_j / \partial x_i$ is the tensor gradient of $\mathbf{u}$. The pressure is denoted by $p$, and $f$ describes the external forces on the fluid. The stress $\mathbf{\sigma}$ is determined by the constitutive law (1.1). Here and throughout the paper the following notation is used: for tensors $\mathbf{\sigma} = (\sigma_{ij})$, $\mathbf{\tau} = (\tau_{ij})$, $\mathbf{\mathbf{\sigma}} : \mathbf{\tau} = \sum_{i,j} \sigma_{ij} \tau_{ij}$, $|\mathbf{\sigma}|^2 = \mathbf{\sigma} : \mathbf{\sigma}$. The constant $\nu > 0$ is the viscosity of the fluid, and $\alpha \geq 0$ is a parameter that may be proportional to the time step. When $\alpha = 0$, (1.1)–(1.4) represents the steady-state Stokes problem. It should be noted that the incompressibility condition (1.3) further implies that the boundary data $\mathbf{u}_\Gamma$ satisfies $\int_{\Gamma} \mathbf{u}_\Gamma \cdot \mathbf{n} \, d\Gamma = 0$ where $\mathbf{n}$ is the outer unit normal vector of $\Gamma$.

There is considerable interest in developing efficient, stable, and theoretically accurate numerical approximation schemes for viscous incompressible flows modeled by the Navier-Stokes equations [1, 2, 3, 4]. These methods often give rise to a system of nonlinear equations which require an iterative solution process, and time-dependent problems require the solution of these nonlinear systems at each time step. Many such approaches reduce to the repeated solution of the generalized Stokes problem [2], a second-order linear PDE. It should be noted that many approximation methods for the generalized Stokes problem are variants of methods for the steady Stokes problem.

For Newtonian flows, the constitutive law (1.1) can be incorporated directly into the momentum equation (1.2), thereby eliminating the total stress tensor $\mathbf{\sigma}$. Many approaches for solving the generalized and steady Stokes problems are therefore posed in a primal-mixed formulation of velocity and pressure unknowns only. A disadvantage of this approach is that when approximations of physically relevant variables $\nabla \mathbf{u}$ or $\mathbf{\sigma}$ are desired, the approximation for $\mathbf{u}$ must be differentiated, resulting in a loss of computational accuracy. Additionally, many non-Newtonian fluid models give rise to nonlinear constitutive laws, from which $\mathbf{\sigma}$ cannot be eliminated.

When $\mathbf{\sigma}$ or $\nabla \mathbf{u}$ is retained in the problem formulation, certain vector-tensor identities (such as Green’s formulas) can be applied and a dual-mixed formulation of (1.1)–(1.4) can be posed. Advantages of this formulation include a relaxed regularity requirement on the function space in which $\mathbf{u}$ resides, which leads to greater flexibility when choosing finite-dimensional subspaces to approximate the velocity. Another advantage is the reformulation of essential boundary conditions on velocity into natural boundary conditions which are easily incorporated into the variational formulations. This approach for Stokes problems shares many characteristics with stress-displacement formulations in linear elasticity, and much of the available research in those methods can aid in the development of dual-mixed methods for Stokes problems. Further motivation for studying dual-mixed approaches arises from the inherent coupling of fluid and elasticity systems in fluid-structure interaction (FSI) problems, as both systems can be approximated using dual-mixed methods with similar structure.

However, the inclusion of $\mathbf{\sigma}$ as an unknown quantity poses significant computational issues, namely (i) the increase in the number of unknowns to be solved for and (ii) the symmetry of the physical stress. The difficulty in using symmetric tensor finite elements in Stokes and elasticity problems is well-documented [5, 6]. Recently, new finite elements for symmetric tensors have been introduced [7, 8] and applied to the elasticity problem. An application of these elements to the Stokes problem is discussed in [9]. Another alternate approach is to impose the symmetry of $\mathbf{\sigma}$ weakly [10, 11]. In [12], the pseudostress was introduced as an additional variable and used in the modeling equations instead of the symmetric stress tensor. A similar approach that, in the context of fluid problems, introduces the pseudostress and velocity gradient as additional variables, has been employed in several recent works (see [13, 14, 15, 16, 17, 18]). Both approaches yield saddle-point problems which require certain inf-sup conditions to be satisfied by the chosen finite element approximation spaces.

The work in this paper is based on the approximation methods presented by Bustinza, Gatica, and González [19] for generalized Stokes problems and by Howell [20] for steady Stokes problems. Both
of these works are extensions of a dual-mixed approximation method for nonlinear and linear steady Stokes problems derived in [13] and [21]. This method introduces both the pseudostress and velocity gradient as additional unknowns, and the variational form yields a twofold saddle point problem. For \( k \geq 0 \) (\( k = 0 \) in [13], \( k \geq 0 \) in [21]), it has been shown that Raviart-Thomas elements of order \( k \) can be used for the pseudostress and discontinuous piecewise polynomial elements of degree \( k \) can be used for the velocity, its gradient, and the pressure. In [20], this approach is modified by restricting the space of velocity gradients to tensors with zero trace. By doing so, the pressure can be eliminated completely from the variational formulation and computed accurately via a postprocessing computation. In two space dimensions, this approach reduces the number of degrees of freedom required in [13] and [21] by 2 on each triangle. The method of [13] is extended to 2-D generalized Stokes problems in [19] (also described in [22]) for the case \( \alpha \gg \nu \). However, in [19] the theoretical results imply that the approximation space for the velocity gradient needs to be enriched to ensure it contains the deviatoric of the discrete pseudostress tensor.

The purpose of this work is to extend the method of [20] to the generalized problem and simplify the method of [19] by reducing the number of degrees of freedom necessary for solvability of the discrete problem. Employing a result from [9], it is shown here that when using trace-free velocity gradients, it is not necessary to enrich the space to include the deviatoric of the discrete pseudostress. When combined with the elimination of the pressure from the formulation, the method described here reduces the computational cost of the method in [19] for the lowest-order 2-D case by 4 degrees of freedom on each triangle (this is a true savings, as they are neither edge nor vertex degrees of freedom). The method is easily generalized to three space dimensions and is also shown to be suitable for all \( \alpha > 0 \), regardless of the relative size of \( \nu \). Also of interest is the role of the parameters \( \nu \) and \( \alpha \) in a priori and error estimates, which will be noted in the results. Due to the structure of the approximating linear system, the problem can be reduced to solving a single symmetric positive definite linear system for the pseudostress only, which will be of size \( dN_f \), where \( d \) is the space dimension and \( N_f \) is the number of edges (\( d = 2 \)) or faces (\( d = 3 \)) in the computational mesh. The method presented here also gives rise to new penalty methods for the steady Stokes problem.

This work also represents a bridge between the aforementioned works and other pseudostress methods for Stokes problems recently studied by Cai, Lee, and Wang [23]; Cai and Wang [24]; and Cai, Lee, Vassilevski, and Wang [25]. In these works the pseudostress is introduced as an auxiliary variable into the steady Stokes problem with pure Dirichlet boundary conditions and the constitutive equation (1.1) is inverted to express the pseudostress in terms of the velocity gradient. The pressure can then be eliminated and a saddle point variational formulation in velocity and pseudostress is derived, transforming essential boundary conditions for velocity into natural conditions in the same manner as described above. It is shown that Raviart-Thomas elements are suitable for the approximation of the pseudostress and discontinuous piecewise polynomials are used for velocity. A penalty term is then introduced in [24] and [25] to eliminate the velocity from the formulation (much in the same way that is traditionally done for the pressure in the velocity-pressure formulations, see [2]). This new term also represents the zeroth-order velocity term present in the generalized Stokes problem (1.1)–(1.4). The problem reduces to solving a symmetric positive definite linear system in pseudostress only, which as noted above will be of size \( dN_f \) in the lowest-order case. The method presented here differs from these approaches [23, 24, 25] by introducing the trace-free velocity gradient tensor as an additional variable and expressing the pseudostress as a function of the velocity gradient. By doing so, the method does not require the inversion of the constitutive law and is more appropriate for extension to problems with nonlinear constitutive laws that cannot be inverted analytically (Carreau Law fluids are an example of this type). A similarity of the two approaches is that both methods reduce to equivalent linear systems.
of the same size for the pseudostress, although they are derived in different ways.

A description of the notation used in this paper, the modification of the original problem, and the dual-mixed variational formulation is given in Section 2. Solvability of the continuous and discrete formulations and the description of appropriate finite element approximation spaces is described in 3. The new penalty methods for steady Stokes problems are described in Section 4. In Section 5, the reduction of the discretized linear problem to one in pseudostress only is described, and similarities to existing methods are discussed. Some numerical results that illustrate the theoretical results are given in Section 6.

2. MATHEMATICAL SETTING

Let Ω be a bounded open subset of \( \mathbb{R}^d \), \( d = 2, 3 \) with Lipschitz continuous boundary \( \Gamma \). The following conventions will be used throughout this paper: tensor functions will be represented by bold lowercase greek letters, vectors will be represented by bold lowercase latin letters, and function spaces will be represented by upper case letters. Norms will be denoted by either \( \| \cdot \|_X \) for a given function space (or product space) \( X \) or by \( \| \cdot \|_{m,r,\Omega} \) for the Sobolev space \( W^{m,r}(\Omega) \). When \( r = 2 \), the standard notation \( H^m(\Omega) = W^{m,2}(\Omega) \) will be used. The infinity norm will be denoted by \( \| \cdot \|_\infty \). For Banach spaces \( X \) and \( Y \), \( X^\ast \) denotes the dual space of \( X \) with associated norm \( \| \cdot \|_{X^\ast} \) and \( \mathcal{L}(X, Y) \) represents the set of all bounded linear operators from \( X \) to \( Y \). Let \( \langle \cdot, \cdot \rangle \) represent a duality pairing between Banach spaces, with a subscript when the particular spaces are not clear from the context.

2.1. Analysis of Saddle Point Problems

Here the relationship between twofold saddle point problems and single saddle point problems is discussed. The results from this section will then be applied to the problem of interest in Section 3.

Let \( U_1, U_2, P \) be reflexive Banach spaces, and let \( a_1 : U_1 \times U_1 \to \mathbb{R}, a_2 : U_2 \times U_2 \to \mathbb{R}, b_1 : P \times U_1 \to \mathbb{R}, b_2 : P \times U_2 \to \mathbb{R} \) all be continuous bilinear forms. The problem of interest is: given linear functionals \( f_1 \in U_1^\ast, f_2 \in U_2^\ast, g \in P^\ast \), find \((u_1, u_2, p) \in U_1 \times U_2 \times P \) satisfying

\[
\begin{align*}
    a_1(u_1, v_1) + b_1(p, v_1) &= f_1(v_1), \quad \forall v_1 \in U_1, \\
    a_2(u_2, v_2) + b_2(p, v_2) &= f_2(v_2), \quad \forall v_2 \in U_2, \\
    b(q, u_1) + b_2(q, u_2) &= g(q), \quad \forall q \in P.
\end{align*}
\]

With appropriate notation, the problem (2.1) can be transformed into a single saddle point problem. Letting \( a((u_1, u_2), (v_1, v_2)) = a_1(u_1, v_1) + a_2(u_2, v_2), b(q, (u_1, u_2)) = b_1(q, u_1) + b_2(q, u_2), \) and \( f((v_1, v_2)) = f_1(v_1) + f_2(v_2) \), we can write (2.1) as

\[
\begin{align*}
    a((u_1, u_2), (v_1, v_2)) + b(p, (v_1, v_2)) &= f((v_1, v_2)), \quad \forall (v_1, v_2) \in U_1 \times U_2, \\
    b(q, (u_1, u_2)) &= g(q), \quad \forall q \in P.
\end{align*}
\]

From the general theory of (single) saddle point problems (for example, see [26, 27, 1, 5, 28]), necessary and sufficient conditions for the well-posedness of problems of the form (2.2) are well-known. Let \( V \) denote the kernel of the linear operator \( B \in \mathcal{L}(U_1 \times U_2, P^\ast) \) defined by \( \langle B(u_1, u_2), q \rangle = g(q), \forall q \in P \) and \( \|B(u_1, u_2)\|_{\mathcal{L}(U_1 \times U_2, P^\ast)} \leq m \|u_1\|_U \|u_2\|_U \).
and there exists a constant $\beta$, i.e.,

$$V := \{ (u_1, u_2) \in U_1 \times U_2 \mid b(q, (u_1, u_2)) = 0 \quad \forall q \in P \}.$$  

Then the following result establishes the well-posedness of problem (2.34).

**Theorem 2.1** ([1] Theorem I.4.1, Corollary I.4.1) Assume that the bilinear form $a(\cdot, \cdot)$ is coercive over $V$ (V-elliptic), i.e., there exists a constant $m > 0$ such that $a((u_1, u_2), (u_1, u_2)) \geq m \| (u_1, u_2) \|_{U_1 \times U_2}$ for all $(u_1, u_2) \in V$. Then problem (2.1) is well-posed if and only if $b(\cdot, \cdot)$ satisfies the inf-sup condition, i.e., if there is a constant $\beta_B > 0$ such that

$$\inf_{q \in P} \sup_{(v_1, v_2) \in U_1 \times U_2} \frac{b(q, (v_1, v_2))}{\| (v_1, v_2) \|_{U_1 \times U_2} \| q \|_P} \geq \beta_B. \quad (2.3)$$

The condition (2.3), which can also be written as

$$\inf_{q \in P} \sup_{(v_1, v_2) \in U_1 \times U_2} \frac{b_1(q, v_1) + b_2(q, v_2)}{\| (v_1, v_2) \|_{U_1 \times U_2} \| q \|_P} \geq \beta_B. \quad (2.4)$$

is not easily verified in many situations, in continuous or discrete contexts. However, the equivalent problem (2.1) has a structure that lends itself to a different characterization of the inf-sup condition. Note that, for the appropriately defined linear operators $A_1$, $A_2$, $B_1$, and $B_2$, we have

$$\begin{bmatrix} A_1 & 0 & B_1^* \\ 0 & A_2 & B_2^* \\ B_1 & B_2 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ g \\ q \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ g \end{bmatrix} \iff \begin{bmatrix} A_1 & B_1^* & 0 \\ B_1 & B_2 & q \\ 0 & B_2^* & A_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ q \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ g \end{bmatrix}. \quad (2.5)$$

With the exception of the operator $A_2$, the problem on the right of (2.5) has the structure of a twofold saddle point problem. Problems with this structure have been previously analyzed [29, 30, 31] and it has been shown that, for $A_2 = 0$, the sufficient conditions for solvability include the surjectivity of the operators $B_1$ and $B_2$. (This is in fact the approach for dual-mixed steady Stokes problems [13, 20].) This will be guaranteed when the individual bilinear forms $b_1$ and $b_2$ satisfy appropriate inf-sup conditions, which may, in both the continuous and discrete contexts, be easier to prove than the condition (2.4).

However, the existing cache of results for single saddle point problems should not be forsaken by merely viewing the problem as a modified twofold saddle point problem. Thus the goal of the present analysis is to describe sufficient conditions for problems (2.5) to be well-posed by showing, with convenient choices of inf-sup conditions, (2.1) is really just a single saddle point problem (2.2) whose criteria for solvability has been met. This is motivated by recent results in [9] that give equivalent sets of inf-sup conditions for twofold saddle point problems. In the following analysis, define the $P$-kernel of $b_2$ by

$$Z_2 = \{ p \in P \mid b_2(p, u_2) = 0 \quad \forall u_2 \in U_2 \}.$$ 

**Theorem 2.2.** Assume that the bilinear form $a(\cdot, \cdot) = a_1(u_1, v_1) + a_2(u_2, v_2)$ is coercive over $V$, i.e., there exists a constant $m > 0$ such that $a_1(u_1, v_1) + a_2(u_2, v_2) \geq m \| (u_1, u_2) \|_{U_1 \times U_2}$ for all $(u_1, u_2) \in V$. Also assume that there exists a constant $\beta_1 > 0$ such that

$$\inf_{q \in Z_2} \sup_{v_1 \in U_1} \frac{b_1(q, v_1)}{\| v_1 \|_{U_1} \| q \|_P} \geq \beta_1, \quad (2.6)$$

and there exists a constant $\beta_2 > 0$ such that

$$\inf_{v_2 \in U_2} \sup_{q \in P} \frac{b_2(q, v_2)}{\| q \|_P \| v_2 \|_{U_2}} \geq \beta_2. \quad (2.7)$$

Then there exists a unique solution $(u_1, u_2, p) \in U_1 \times U_2 \times P$ to problem (2.2).
Proof
With the application of Theorem 2.1, the desired result will hold if it can be shown that the inf-sup conditions (2.6) and (2.7) imply (2.4). This is shown in the proof of Lemma 3.2 in [9], and will be presented here for clarity. Lemma 4.1 of §1 of [1] will be used ubiquitously below.

Let \( \tilde{Z} = \{ u_1 \in U_1 \mid b_1(q, u_1) = 0 \ \forall q \in P \} \) and let \( g \in P^* \). The inf-sup condition (2.6) implies that the operator \( B_1 \in \mathcal{L}(U_1, \tilde{Z}) \) is an isomorphism, thus there exists a \( u_1 \in U_1 \) such that

\[
\beta_1 ||u_1||_{U_1} \leq ||g||_{P^*} \text{ (the Hahn-Banach theorem is employed to extend a suitable } g \in Z^0_1 \text{ to } P^*). \]

Then the inf-sup condition (2.7) implies that \( B_2^* \in \mathcal{L}(U_2, Z^0_2) \) is an isomorphism (here \( Z^0_2 \) is the polar set of \( Z_2 \), i.e., \( Z^0_2 = \{ g \in P^* \mid g(p) = 0 \ \forall p \in Z_2 \} \)). Thus there exists a \( u_2 \in U_2 \) such that

\[
b_2(q, u_2) = g(q) - b_1(q, u_1) \quad \forall q \in P,
\]

as the functional \( \tilde{g} := g - B_1 u_1 \) is in \( Z^0_2 \), with

\[
\beta_2 ||u_2||_{U_2} \leq ||\tilde{g}||_{P^*} = ||g - B_1 u_1||_{P^*} \leq \left( 1 + \frac{1}{\beta_1} \right) ||g||_{P^*}.
\]

Thus \( (u_1, u_2) \) satisfies

\[
b_1(q, u_1) + b_2(q, u_2) = g(q) \quad \forall q \in P,
\]

and

\[
||u_1, u_2|| \leq \left( 1 + \frac{1}{\beta_1} \right) \left( 1 + \frac{1}{\beta_2} \right) ||g||_{P^*} = \frac{1 + \beta_1 + \beta_2}{\beta_1 \beta_2} ||g||_{P^*}.
\]

Let \( p \in P \) and let \( g \in P^* \) satisfy \( g(p) = ||g||_{P^*}^2 = ||p||_{P}^2 \). Then, with \( c = (1 + \beta_1 + \beta_2)/\beta_1 \beta_2 \),

\[
||u_1, u_2||_{U_1 \times U_2} \leq c ||p||_{P} \text{ and thus}
\]

\[
||p||_{P} = \frac{g(p)}{||p||_{P}} = \frac{b_1(p, u_1) + b_2(p, u_2)}{||p||_{P}} \leq \frac{b_1(p, u_1) + b_2(p, u_2)}{(1/c)(||u_1, u_2||_{U_1 \times U_2})}.
\]

Hence

\[
\frac{b_1(p, u_1) + b_2(p, u_2)}{(1/c)(||u_1, u_2||_{U_1 \times U_2})} \geq \frac{1 + \beta_1 + \beta_2}{\beta_1 \beta_2} ||p||_{P},
\]

which proves the inf-sup condition (2.4).

\[\square\]

Remark 2.1. In Theorem 2.2, the roles of \( u_1 \) and \( u_2 \) (and thus \( a_1, a_2, b_1, b_2 \)) can be interchanged. Thus (2.2) is well-posed if \( a(\cdot, \cdot) \) is coercive over \( V \) and there exist constants \( \beta_1, \beta_2 > 0 \) such that

\[
\inf_{q \in Z_1} \sup_{v_1 \in V_1} \frac{b_2(q, v_2)}{||v_2||_{U_2} ||q||_{P}} \geq \beta_2 \quad \text{and} \quad \inf_{v_1 \in V_1} \sup_{q \in P} \frac{b_1(q, v_1)}{||q||_{P} ||v_1||_{U_1}} \geq \beta_1,
\]

where \( Z_1 = \{ p \in P \mid b_1(p, u_1) = 0 \ \forall u_1 \in U_1 \} \).

2.2. Notation and Definitions

With the standard definition of the Hilbert space \( L^2(\Omega) \), define the following spaces:

\[
U := \left( L^2(\Omega) \right)^d, \quad T := \left( L^2(\Omega) \right)^{d \times d},
\]

\[
S := (H(\Omega; \text{div}))^d = \{ \tau \in T \mid \text{div} \tau \in U \}.
\]
Let \((\cdot, \cdot)\) represent the \(L^2(\Omega)\) inner product for scalar, vector, or tensor functions. Each of the above spaces has an associated norm induced from the inner product:

\[
\|u\|_U := (u, u)^{1/2} = \left( \int_{\Omega} u \cdot u \, d\Omega \right)^{1/2}, \quad \|\tau\|_T := (\tau, \tau)^{1/2} = \left( \int_{\Omega} \tau : \tau \, d\Omega \right)^{1/2},
\]

\[
\|\tau\|_S := \left( \|\tau\|_T^2 + \|\text{div} \, \tau\|_U^2 \right)^{1/2}.
\]

The space \(T^0\) of trace-free tensors in \(T\) will be utilized:

\[
T^0 := \left\{ \tau \in T \mid \sum_{i=1}^{d} \tau_{ii} = 0 \right\}.
\]

A representation of \(\tau \in T^0\) can be written as

\[
\tau = \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & -\tau_{11} \end{bmatrix} \quad \text{or} \quad \tau = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & -\tau_{11} - \tau_{22} \end{bmatrix}
\]

for \(d = 2\) or \(d = 3\), respectively.

### 2.3. Derivation of the Variational Formulation

The dual-mixed variational formulation of (1.1)–(1.4) is obtained by introducing the pseudostress and velocity gradient as auxiliary variables. Let \(\psi := \nu \nabla u - p I\). Then \(\sigma = \psi + \nu \nabla u^T\), and we have from (1.1),

\[
\text{div} \, \sigma = \text{div} (\psi + \nu (\nabla u)^T) = \text{div} \, \psi + \nu \text{div} (\nabla u)^T = \text{div} \, \psi
\]

as (1.3) implies \(\text{div} (\nabla u)^T = 0\). Then set \(\phi := \nabla u\) and observe that the incompressibility condition (1.3) is equivalent to \(\text{tr}(\phi) = 0\). Then problem (1.1)–(1.4) can be written as: Find \((\phi, u, \psi, p)\) such that

\[
\phi = \nabla u \quad (2.8)
\]

\[
\psi = \nu \phi - p I \quad (2.9)
\]

\[
\alpha u - \nabla \cdot \psi = f \quad \text{in } \Omega, \quad (2.10)
\]

\[
\text{tr}(\phi) = 0 \quad \text{in } \Omega, \quad (2.11)
\]

\[
u \phi - p I \quad \text{on } \Gamma, \quad (2.12)
\]

To derive a variational formulation for (2.8)–(2.12), it suffices to consider \(\phi \in T^0\). Then, for \(\tau \in S\), (2.8) and (2.12) give

\[
\int_{\Omega} \phi : \tau \, d\Omega = \int_{\Omega} \nabla u : \tau \, d\Omega
\]

\[
= -\int_{\Omega} u \cdot \text{div} \, \tau \, d\Omega + \int_{\Gamma} u_{\Gamma} \cdot (\tau \cdot n) \, d\Gamma, \quad \forall \tau \in S.
\] (2.13)

Testing against tensors in \(\zeta \in T^0\), (2.9) yields

\[
0 = \nu \int_{\Omega} \phi : \zeta \, d\Omega - \int_{\Omega} \psi : \zeta \, d\Omega - \int_{\Omega} p I : \zeta \, d\Omega
\]

\[
= \nu \int_{\Omega} \phi : \zeta \, d\Omega - \int_{\Omega} \psi : \zeta \, d\Omega, \quad \forall \zeta \in T^0,
\] (2.14)
as \( tr(\mathbf{c}) = 0 \) implies that \( p \mathbf{I} : \mathbf{c} = 0 \). Then, multiplying (2.10) by \( v \in U \) and integrating, the following variational form of (2.8)–(2.12) is established: Given \( f \in U \) and \( u_T \in H^{1/2}(\Gamma) \), find \( (\phi, u, \psi) \in T^0 \times U \times S \) such that

\[
\begin{align*}
\nu \int_{\Omega} \phi : \mathbf{c} \, d\Omega - \int_{\Omega} \psi : \mathbf{c} \, d\Omega &= 0, \quad \forall \mathbf{c} \in T^0, \\
\alpha \int_{\Omega} u \cdot v \, d\Omega - \int_{\Omega} v \cdot \text{div} \psi \, d\Omega &= \int_{\Omega} f : v \, d\Omega, \quad \forall v \in U, \\
- \int_{\Omega} \phi : \tau \, d\Omega - \int_{\Omega} u \cdot \text{div} \tau \, d\Omega &= - \int_{\Gamma} u_T \cdot (\tau \cdot n) \, d\Gamma, \quad \forall \tau \in S.
\end{align*}
\]

(2.15)

(2.16)

(2.17)

The solution to (2.15)–(2.17) is not unique, as \((\phi, u, \psi)\) and \((\phi, u, \psi + c \mathbf{I})\) for \( c \in \mathbb{R} \) will satisfy the equations. To force the variational problem to have a unique solution, there are two common approaches. One approach is to seek solutions where \( \psi \) and \( \tau \) are restricted to the space of tensors in \( S \) with mean trace zero, i.e.,

\[ \tau, \psi \in S^0 := \left\{ \tau \in S \mid \int_{\Omega} tr(\tau) \, d\Omega = 0 \right\}. \]

This is the approach employed in [25] and [18]. An alternate approach is to leave \( \psi, \tau \in S \) and include a Lagrange multiplier in the variational form that will force the mean of \( tr(\psi) \) to be zero. This is the approach employed in [13, 19, 32, 20] and will be used here. It should be noted that only nominal changes are required to apply the results in this work to the approach of seeking the solution in \( S^0 \), as it is a simplification of the method used here.

The mean trace zero constraint is thus enforced by adding the variational equation

\[ \eta \int_{\Omega} tr(\psi) \, d\Omega = 0, \quad \forall \eta \in \mathbb{R}, \]

to (2.16). To preserve the symmetry of the variational formulation, the equation \( \lambda \int_{\Omega} tr(\tau) \, d\Omega = 0 \) for all \( \tau \in S \) is added to (2.17) as the true value of \( \lambda \) is 0 (this can be seen by choosing \( \tau = \mathbf{I} \) in (2.17)). Additionally, the scaled inner product \( \alpha \lambda \eta \) (also 0) is added to the variational formulation in the momentum equation. Then the complete variational form of the dual-mixed generalized Stokes problem is: Given \( f \in U \) and \( u_T \in H^{1/2}(\Gamma) \), find \( (\phi, u, \lambda, \psi) \in T^0 \times U \times \mathbb{R} \times S \) such that

\[
\begin{align*}
\nu \int_{\Omega} \phi : \mathbf{c} \, d\Omega - \int_{\Omega} \psi : \mathbf{c} \, d\Omega &= 0, \quad \forall \mathbf{c} \in T^0, \\
\alpha \int_{\Omega} u \cdot v \, d\Omega + \alpha \lambda \eta - \int_{\Omega} v \cdot \text{div} \psi \, d\Omega + \eta \int_{\Omega} tr(\psi) \, d\Omega &= \int_{\Omega} f : v \, d\Omega, \quad \forall v \in U \times \mathbb{R}, \\
- \int_{\Omega} \phi : \tau \, d\Omega - \int_{\Omega} u \cdot \text{div} \tau \, d\Omega + \lambda \int_{\Omega} tr(\tau) \, d\Omega &= - \int_{\Gamma} u_T \cdot (\tau \cdot n) \, d\Gamma, \quad \forall \tau \in S.
\end{align*}
\]

(2.18)

(2.19)

(2.20)

Remark 2.2. If \( \phi \in T \) (not trace-free), then the variational formulation (2.18)–(2.20) will include pressure terms, i.e., the term \( \int_{\Omega} p \mathbf{I} : \mathbf{c} \, d\Omega \) will be added to the left hand side of (2.18) and the term \( \int_{\Omega} q \mathbf{I} : \phi \, d\Omega \) will be added to the left hand side of (2.20). This formulation with pressure is then equivalent that of [19] with the exception of the \( \alpha \lambda \eta \) term.

Copyright © 2009 John Wiley & Sons, Ltd.


Prepared using fidauth.cls
Define the variational forms

\[
a_1(\phi, \zeta) := \nu \int_{\Omega} \phi : \zeta \, d\Omega, \quad \phi, \zeta \in T^0, \tag{2.21}
\]

\[
a_2((u, \lambda), (v, \eta)) := \alpha \int_{\Omega} u \cdot v \, d\Omega + \alpha \lambda \eta, \quad u, v \in U, \eta, \lambda \in \mathbb{R}, \tag{2.22}
\]

\[
b_1(\psi, \zeta) := -\int_{\Omega} \psi : \zeta \, d\Omega, \quad \zeta \in T^0, \psi \in S, \tag{2.23}
\]

\[
b_2(\psi, (v, \eta)) := -\int_{\Omega} v \cdot \text{div} \psi \, d\Omega + \eta \int_{\Omega} \text{tr}(\psi) \, d\Omega, \psi \in S, v \in U, \eta \in \mathbb{R}. \tag{2.24}
\]

Using (2.21)–(2.24), the problem (2.18)–(2.20) is written as the twofold saddle point problem: Given \(f \in U\) and \(u_\Gamma \in H^{1/2}(\Gamma)\), find \((\phi, u, \lambda, \psi) \in T^0 \times U \times \mathbb{R} \times S\) such that

\[
a_1(\phi, \zeta) + b_1(\psi, \zeta) = 0, \quad \forall \zeta \in T^0, \tag{2.25}
\]

\[
a_2((u, \lambda), (v, \eta)) + b_2(\psi, (v, \eta)) = (f, v), \quad \forall (v, \eta) \in U \times \mathbb{R}, \tag{2.26}
\]

\[
b_1(\tau, \phi) + b_2(\tau, (u, \lambda)) = \langle u_\Gamma, \tau \cdot n \rangle, \quad \forall \tau \in S, \tag{2.27}
\]

where \((\cdot, \cdot)\) represents the standard \(L^2\) inner product and \((\cdot, \cdot)_\Gamma\) represents the duality pairing of \(H^{1/2}(\Gamma)\) and \(H^{-1/2}(\Gamma)\). It will also be convenient to use the operator notation

\[
A_1 : T^0 \to (T^0)^*, \quad (A_1^1(\phi, \zeta) := a_1(\phi, \zeta), \tag{2.28}
\]

\[
A_2 : (U \times \mathbb{R}) \to (U \times \mathbb{R})^*, \quad (A_2^2(u, \lambda), (v, \eta)) := a_2((u, \lambda), (v, \eta)), \tag{2.29}
\]

\[
B_1 : T^0 \to S^*, \quad (B_1^1(\phi, \tau) := b_1(\phi, \tau), \tag{2.30}
\]

\[
B_2 : (U \times \mathbb{R}) \to S^*, \quad (B_2^2(u, \lambda), \tau) := b_2(\tau, (u, \lambda)), \tag{2.31}
\]

where \((\cdot, \cdot)\) represents the appropriate duality pairing. Associate with \(B_1\) and \(B_2\) are the adjoint operators \(B_1^* : S \to (T^0)^*\) and \(B_2^* : S \to (U \times \mathbb{R})^*\), respectively. Also, define the combined variational forms

\[
a((\phi, u, \lambda), (\zeta; v, \eta)) := a_1(\phi, \zeta) + a_2((u, \lambda), (v, \eta)) \tag{2.32}
\]

\[
b(\psi, (\zeta; v, \eta)) := b_1(\psi, \zeta) + b_2(\psi, (v, \eta)) \tag{2.33}
\]

along with their corresponding operators

\[
A : (T^0 \times U \times \mathbb{R}) \to (T^0 \times U \times \mathbb{R})^*, \quad (A(\phi, u, \lambda), (\zeta; v, \eta)) := a((\phi, u, \lambda), (\zeta; v, \eta)), \tag{2.34}
\]

\[
B : (T^0 \times U \times \mathbb{R}) \to S^*, \quad (B(\phi, u, \lambda), \tau) := b((\phi, u, \lambda), \tau)). \tag{2.35}
\]

Two equivalent formulations of problem (2.18)–(2.20) under this notation are:

\[
\begin{bmatrix}
A_1 & 0 & B_1^* \\
0 & A_2 & B_2^* \\
B_1 & B_2 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\phi \\
(u, \lambda) \\
\psi \\
\end{bmatrix}
=
\begin{bmatrix}
0 \\
(f, v) \\
\langle u_\Gamma, \tau \cdot n \rangle \\
\end{bmatrix} \tag{2.36}
\]

and

\[
\begin{bmatrix}
A & B^* \\
B & 0 \\
\end{bmatrix}
\begin{bmatrix}
(\phi, u, \lambda) \\
\psi \\
\end{bmatrix}
=
\begin{bmatrix}
(f, v) \\
\langle u_\Gamma, \tau \cdot n \rangle \\
\end{bmatrix} \tag{2.37}
\]

These and other similar representations will be utilized in the subsequent analysis.
3. SOLVABILITY AND DISCRETE APPROXIMATION OF THE GENERALIZED STOKES PROBLEM

In this section the well-posedness of the variational problem (2.18)–(2.20) and its discrete analogue are discussed. This is accomplished by applying the results of Section 2.1 to the equivalent formulations (2.33) and (2.34). For the remainder of this section it is assumed that \( \alpha > 0 \) (corresponding to the generalized Stokes problem).

3.1. Continuous and Discrete Variational Problems

Let \( Z \) denote the kernel of the linear operator \( B \):

\[
Z := \{ (\phi, u, \lambda) \in T^0 \times U \times \mathbb{R} \mid b(\tau, (\phi, u, \lambda)) = 0 \quad \forall \tau \in S \} \subset T^0 \times U \times \mathbb{R}.
\]

It is clear that the linear operator \( A \) is coercive over \( Z \); we have

\[
a((\phi, u, \lambda), (\phi, u, \lambda)) = \nu \| \phi \|^2_T + \alpha \| u \|^2_U + \alpha | \lambda |^2 \geq \min(\alpha, \nu) \| (\phi, u, \lambda) \|_{T \times U \times \mathbb{R}},
\]

for all \( (\phi, u, \lambda) \in Z \). Thus the necessary hypothesis is that \( \inf \sup_{\tau \in S} \| \phi \|_T \right| \tau \right| \leq \| \phi \|_T \right| \tau \right|_S \), \( \forall \tau \in S \).

The continuity of the bilinear forms \( b_1 \) and \( b_2 \) is clear, we have

\[
b_1(\tau, (\phi, u, \lambda)) \leq \| \phi \|_T \right| \tau \right| \leq \| \phi \|_T \right| \tau \right|_S,
\]

and, provided \( |\Omega| \geq 1 \),

\[
b_2(\tau, (u, \lambda)) \leq \| u \|_U \right| \text{div} \tau \right|_U + | \lambda | |\Omega|^{1/2} | tr(\tau) |_{0,2,\Omega} \leq (|\Omega|)^{1/2} \| (u, \lambda) \|_{U \times \mathbb{R}} \right| \tau \right|_S.
\]

Let \( Z_2 \) be the kernel of \( B_2^* \):

\[
Z_2 := \{ \psi \in S \mid b_2(\psi, (v, \eta)) = 0 \quad \forall (v, \eta) \in U \times \mathbb{R} \}.
\]

From the definition of \( b_2 \), \( Z_2 \) can be written as

\[
Z_2 = \left\{ \psi \in S \mid \text{div} \psi = 0 \quad \text{and} \quad \int_{\Omega} \text{tr} \psi \right| d\Omega = 0 \right\}.
\]

Hence problem (2.34) is well-posed if the inf-sup conditions

\[
\inf_{\tau \in Z_2} \sup_{\phi \in T^0} \frac{b_1(\tau, (\phi, \phi))}{\| \phi \|_T \right| \tau \right|_S} \geq \beta_1,
\]

\[
\inf_{(u, \lambda) \in U \times \mathbb{R}} \sup_{\tau \in S} \frac{b_2(\tau, (u, \lambda))}{\| \tau \|_S \right| (u, \lambda) \|_{U \times \mathbb{R}}} \geq \beta_2,
\]

hold for some constants \( \beta_1, \beta_2 > 0 \) (here \( b_1 \) and \( b_2 \) are as defined in (2.23) and (2.24), respectively).

The inf-sup condition (3.5) has been established in Lemma 2.1 and Theorem 2.4 of [13]. The inf-sup condition (3.6) is proven in Lemma 3.2 of [20], and requires Lemma 3.1 of [33]. Thus the solvability of (2.34) is established via the application of Theorem 2.2. Then standard a priori estimates for the solution components are immediately established (Proposition II.1.3 of [5]). This is summarized in the result below.
Theorem 3.1. Let \( f \in U \) and \( u_T \in H^{1/2}(\Gamma) \). Then there exists a unique solution \((\phi, u, \lambda, \psi) \in T^0 \times U \times \mathbb{R} \times S \) to problem (2.18)–(2.20) satisfying

\[
\| (\phi, u, \lambda) \|_{T^0 \times U \times \mathbb{R} \times S} \leq \frac{1}{m} \| f \|_{U} + \frac{1}{\beta} \left( 1 + \frac{M}{m} \right) \| u_T \|_{H^{1/2}(\Gamma)},
\]

\[
\| \psi \|_{S} \leq \frac{1}{\beta} \left( 1 + \frac{M}{m} \right) \| f \|_{U} + \frac{M}{\beta^2} \left( 1 + \frac{M}{m} \right) \| u_T \|_{H^{1/2}(\Gamma)},
\]

where \( m = \min\{\nu, \alpha\} \), \( M = \max\{\nu, \alpha\} \), and \( \beta = (1 + \beta_1 + \beta_2)/\beta_1 \beta_2 \) where \( \beta_1, \beta_2 \) are given in (3.5)–(3.6).

Remark 3.1. There is no restriction on the relative size of \( \nu > 0 \) and \( \alpha > 0 \) in this formulation, in contrast to the assumption of \( \alpha \gg \nu > 0 \) in [19].

Remark 3.2. The estimates (3.7)–(3.8) above imply that

\[
\| (\phi, u, \lambda) \|_{T^0 \times U \times \mathbb{R} \times S} \leq C_1 (\| f \|_{U} + \| u_T \|_{H^{1/2}(\Gamma)})
\]

and

\[
\| \psi \|_{S} \leq C_2 (\| f \|_{U} + \| u_T \|_{H^{1/2}(\Gamma)})
\]

where \( C_1 = O(M/m) \) and \( C_2 = O(M^2/m) \).

The discrete variational formulation is derived by defining the finite-dimensional subspaces \( T^0_h \subseteq T^0, \)

\( S_h \subseteq S, \) and \( U_h \subseteq U. \) Then the discrete problem is: Given \( f \in U \) and \( u_T \in H^{1/2}(\Gamma), \) find \((\phi_h, u_h, \lambda_h, \psi_h) \in T^0_h \times U_h \times \mathbb{R} \times S_h \) such that

\[
\nu \int_{\Omega} \phi_h : \varsigma_h \, d\Omega - \int_{\Omega} \psi_h : \varsigma_h \, d\Omega = 0, \quad \forall \varsigma_h \in T^0_h,
\]

\[
\alpha \int_{\Omega} u_h \cdot v_h \, d\Omega + \nu h \int_{\Omega} v_h \cdot \text{div} \psi_h \, d\Omega + \eta_h \int_{\Omega} \text{tr}(\psi_h) \, d\Omega = \int_{\Gamma} f \cdot v_h \, d\Gamma, \quad \forall (v, \eta) \in U_h \times \mathbb{R}, \quad (3.10)
\]

\[
\int_{\Omega} \phi_h : \tau_h \, d\Omega - \int_{\Omega} u_h \cdot \text{div} \tau_h \, d\Omega + \int_{\Omega} \text{tr}(\tau_h) \, d\Omega = - \int_{\Gamma} u_T \cdot (\tau_h \cdot n) \, d\Gamma, \quad \forall \tau_h \in S_h. \quad (3.11)
\]

This can be written as

\[
a_1(\phi_h, \varsigma_h) + b_1(\psi_h, \varsigma_h) = 0,
\]

\[
a_2((u_h, \lambda_h), (v_h, \eta_h)) + b_2(\psi_h, (v_h, \eta_h)) = (f, v_h),
\]

\[
b_1(\tau_h, \phi_h) + b_2(\tau_h, (u_h, \lambda_h)) = \langle -u_T, \tau_h \cdot n \rangle_T,
\]

for all \((\varsigma, v_h, \eta, \tau_h) \in T^0_h \times U_h \times \mathbb{R} \times S_h. \) Define the finite-dimensional kernels

\[
Z_h := \{ (\phi_h, u_h, \lambda_h) \in T^0_h \times U_h \times \mathbb{R} \mid b(\tau_h, (\phi_h, u_h, \lambda_h)) = 0 \quad \forall \tau_h \in S_h \},
\]

and

\[
Z_{2h} := \{ \psi_h \in S_h \mid b_2(\psi_h, (v_h, \eta_h)) = 0 \quad \forall (v_h, \eta_h) \in U_h \times \mathbb{R} \}.
\]

The kernel \( Z_{2h} \) inherits the properties of \( Z_2 \), i.e.,

\[
Z_{2h} = \left\{ \psi_h \in S_h \mid \text{div} \psi_h = 0 \quad \text{and} \quad \int_{\Omega} \text{tr}(\psi_h) \, d\Omega = 0 \right\}.
\]
Theorem 3.2. Let \( f \in U \) and \( u^r \in H^{1/2}(\Gamma) \). Assume that \( a(\cdot, \cdot) = a_1(\cdot, \cdot) + a_2(\cdot, \cdot) \) is coercive over \( Z_h \), i.e., there is a constant \( m > 0 \) such that \( a((\phi_h, u_h, \lambda_h), (\phi_h, u_h, \lambda_h)) \geq m_h \| (\phi_h, u_h, \lambda_h) \|_{T \times U \times \mathbb{R}} \). Assume that there exist constants \( \beta_{1h}, \beta_{2h} > 0 \) such that
\[
\inf_{\tau_h \in Z_h} \sup_{\phi_h \in T_h} \frac{b_1(\tau_h, \phi_h)}{\| \phi_h \|_T \| \tau_h \|_S} \geq \beta_{1h},
\]
and
\[
\inf_{(u_h, \lambda_h) \in U \times \mathbb{R}} \sup_{\tau_h \in S_h} \frac{b_2(\tau_h, (u_h, \lambda_h))}{\| (u_h, \lambda_h) \|_{U \times \mathbb{R}}} \geq \beta_{2h}.
\]
Then problem (3.12) has a unique solution \((\phi_h, u_h, \lambda_h, \psi_h) \in T_h^0 \times U_h \times \mathbb{R} \times S_h \) satisfying
\[
\| (\phi_h, u_h, \lambda_h) \|_{T \times U \times \mathbb{R}} \leq \frac{1}{m_h} \| f \|_U + \frac{1}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \| u^r \|_{H^{1/2}(\Gamma)},
\]
\[
\| \psi_h \|_S \leq \frac{1}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \| f \|_U + \frac{M_h}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \| u^r \|_{H^{1/2}(\Gamma)},
\]
where \( M_h = \| a(\cdot, \cdot) \| \) and \( \beta_h = (1 + \beta_{1h} + \beta_{2h})/\beta_{1h}\beta_{2h} \).

Let \( \| b \| \) represent the operator norm of the bilinear form \( b(\cdot, \cdot) = b_1(\cdot, \cdot) + b_2(\cdot, \cdot) \). From (3.3) and (3.4), an upper bound for \( \| b \| \) is derived by
\[
\| b \| = \sup_{(\phi, u, \lambda) \in T \times U \times \mathbb{R}, \phi \in S} \frac{b(\psi, (\phi, u, \lambda))}{\| (\phi, u, \lambda) \|_{T \times U \times \mathbb{R}} \| \psi \|_S} \leq \frac{1}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \| f \|_U + \frac{M_h}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \| u^r \|_{H^{1/2}(\Gamma)},
\]
provided \( |\Omega| \geq 1 \). The following results follow directly from the approximation theory of single saddle point problems (see Proposition II.2.6 of [5] or Lemma 2.44 of [28], for example).

Theorem 3.3. Assume the hypotheses of Theorem 3.2 are satisfied. Then the approximation \((\phi_h, u_h, \lambda_h, \psi_h) \in T_h^0 \times U_h \times \mathbb{R} \times S_h \) satisfies
\[
\| (\phi, u, \lambda) - (\phi_h, u_h, \lambda_h) \|_{T \times U \times \mathbb{R}} \leq \left( 1 + \frac{M_h}{m_h} \right) \left( 1 + \frac{\| b \|}{\beta_h} \right) \inf_{(\phi_h, u_h, \lambda_h) \in T_h^0 \times U_h} \| (\phi, u, \lambda) - (\phi_h, u_h, \lambda_h) \|_{T \times U} + c \inf_{\tau_h \in S_h} \| \psi - \tau_h \|_S,
\]
and
\[
\| \psi - \psi_h \|_S \leq \frac{M_h}{\beta_h} \left( 1 + \frac{M_h}{m_h} \right) \left( 1 + \frac{\| b \|}{\beta_h} \right) \inf_{(\phi_h, u_h, \lambda_h) \in T_h^0 \times U_h} \| (\phi, u, \lambda) - (\phi_h, u_h, \lambda_h) \|_{T \times U} + \left( 1 + \frac{\| b \|}{\beta_h} + c \frac{M_h}{\beta_h} \right) \inf_{\tau_h \in S_h} \| \psi - \tau_h \|_S.
\]

If \( Z_h \subset Z \), then \( c = 0 \), otherwise \( c = \| b \|/m_h \).
The constitutive relation (2.9) indicates that \( \psi = \nu \phi - p I \). For trace-free \( \phi \), this implies that \( \text{tr}(\psi) = \text{tr}(p I) = np \). Thus the true pressure \( p \in L^2(\Omega) \) is given by \( p = \frac{1}{n} \text{tr}(\psi) \) when \( \phi \) is trace-free. The following result follows directly from Theorem 3.3.

**Corollary 3.1.** Let \( p \in P = L^2(\Omega) \) be given by \( p = \frac{1}{d} \text{tr}(\psi) \) where \( \psi \) is part of the solution to (2.18)-(2.20). Let \( P_h \) be a finite-dimensional subspace of \( P \) such that \( p_h \) given by \( p_h = \frac{1}{d} \text{tr}(\psi_h) \) where \( \psi_h \) is part of the solution to (3.12), is in \( P_h \). Then

\[
\| p - p_h \|_P \leq \sqrt{|\Omega|} \frac{\| S \|}{d} \| \psi - \psi_h \|_S.
\]

**Remark 3.3.** Theorem 3.3 implies that the error in approximating \( \psi \) will be \( O(M^2/m) \), while the other components will be approximated with an error of \( O(M/m) \).

### 3.2. Finite Element Approximation

In this section choices for the subspaces \( T_h^0, U_h, \) and \( S_h \) that satisfy the inf-sup conditions (3.14) and (3.15) are described.

Let \( \Omega \subset \mathbb{R}^d \) be a polygonal domain and \( T_h \) be a regular triangulation of \( \Omega \) into triangles \( (n = 2) \) or tetrahedrals \( (n = 3) \). Thus

\[
\Omega = \bigcup K, \quad K \in T_h,
\]

and assume that there exist constants \( \gamma_1, \gamma_2 \) such that

\[
\gamma_1 h \leq h_K \leq \gamma_2 \rho_K
\]

where \( h_K \) is the diameter of triangle (tetrahedral) \( K \), \( \rho_K \) is the diameter of the greatest ball (sphere) included in \( K \), and \( h = \max_{K \in T_h} h_K \).

The motivating factor in the choice of approximation spaces for \( U_h \) and \( S_h \) is the need for the spaces to satisfy (3.15). After considerations involving the Lagrange multiplier \( \eta_h \) in (3.15), this amounts to finding choices of \( U_h \) and \( S_h \) such that the divergence operator \( \text{div} : S_h \to U_h \) is surjective. When there is no symmetry requirement on tensors in \( S_h \), as is the case here, then any of the well-known vector-scalar pairs of spaces that satisfy a \( \text{div} \)-type inf-sup condition can be used by using the vectors as rows of the tensor \( \psi_h \) and the scalars as components of the vector \( u_h \). Some common choices of finite element spaces of this type that do not require vertex continuity are the Raviart–Thomas (RT) elements for \( n = 2 \) and \( n = 3 \) [34, 35] or the Brezzi–Douglas–Marini (BDM) elements for \( n = 2 \) and \( n = 3 \) [36]. In what follows the Raviart–Thomas elements will be employed, however the analysis can easily be adapted to accommodate the BDM elements.

Let \( k \geq 0 \) be an integer. Let \( K \in T_h \) and let \( P_k(K) \) be the set of all polynomials in the variables \( x_1, \ldots, x_n \) \((n = 2 \text{ or } 3)\) of degree less than or equal to \( k \) defined on the triangle \( K \). Let \( \mathbb{RT}_k(K) \) be the \( d \)-vector of Raviart-Thomas elements on \( K \) defined by

\[
\mathbb{RT}_k(K) = (P_k(K))^d + \left\{ \begin{array}{c} x_1 \\
... \\
x_n \end{array} \right\} : P_k(K).
\]
Then define the following discrete spaces:
\[
S_h := \left\{ \mathbf{\psi} \in S \mid \mathbf{\psi} = (\mathbf{\psi}_1, \ldots, \mathbf{\psi}_n)^T \big| K \in (\mathbb{RT}_k(K))^d, \right. \\
(\mathbf{\psi}_1, \ldots, \mathbf{\psi}_n)^T \big| K \in \mathbb{RT}_k(K), \quad \forall i \in \{1, \ldots, n\}, \quad \forall K \in \mathcal{T}_h \left. \right\}.
\]
\[
U_h := \left\{ \mathbf{u} \in U \mid \mathbf{u}|_K \in (\mathbb{P}_k(K))^d, \quad \forall K \in \mathcal{T}_h \right\}.
\]
It should be noted that the choice of space for the Lagrange multiplier \(\lambda\) is \(\mathbb{R}\) itself. These choices will satisfy the inf-sup condition (3.6), as shown in Theorem 3.1 of [13].

Lemma 3.1. There is a constant \(\beta_{2h} > 0\) such that (3.15) holds for the above choices of \(U_h\) and \(S_h\).

With the choice of finite element for \(S_h\) established, the choice for \(T^0_h\) need only satisfy (3.6), which amounts to the identity operator being surjective from \(T^0_h\) onto \(Z_{2h} \subset S_h\) where \(Z_{2h}\) is characterized by (3.13). The fact that only divergence free tensors in \(S_h\) with mean trace zero need be considered here illustrates the importance of Theorem 2.2. As the Raviart–Thomas vectors with zero divergence are merely vectors whose components are polynomials of degree \(k\) [5], an appropriate choice for \(T^0_h\) is simply the trace-free \(d \times d\) tensors with components in \(\mathbb{P}_k\), i.e.,
\[
T^0_h := \left\{ \mathbf{\phi} \in T^0 \mid \mathbf{\phi}|_K \in (\mathbb{P}_k(K))^d, \quad \forall K \in \mathcal{T}_h \right\}.
\]
This avoids the enrichment necessary for the corresponding discrete space of velocity gradients in [19], which requires that finite element basis for \(\phi_h\) be expanded to include the deviatoric of tensors with rows of Raviart–Thomas vectors. In two space dimensions, this requires an additional two degrees of freedom per triangle.

Remark 3.4. For \(\mathbf{\phi} \in T^0_h\), if \(n = 2\) then \(\mathbf{\phi}\) has three independent components, and if \(n = 3\), then \(\mathbf{\phi}\) has eight independent components.

Thus, as shown in Lemma 4.2 of [20], the above choices of \(T^0_h\) and \(S_h\) satisfy the required inf-sup condition.

Lemma 3.2. There is a constant \(\beta_{1h} > 0\) such that (3.14) holds for the above choices of \(T^0_h\) and \(S_h\).

Remark 3.5. As the Brezzi–Douglas–Marini vectors of order \(k \geq 1\) with zero divergence are also merely polynomials of degree \(k-1\), Lemma 3.2 holds for BDM elements in place of RT elements in the construction of \(S_h\).

It should be noted that, with the above choices of finite element spaces, it is clear that \(a(\cdot, \cdot)\) is coercive over \(Z_h\), and the constants \(\alpha_h, \beta_{1h}, \beta_{2h}, m_h\), and \(M_h\) in Theorems 3.2 and 3.3 are all independent of the mesh parameter \(h\). Standard approximation properties of the above finite element spaces are well-known [5, 28]. Thus, the error estimate below is a direct consequence of these properties, Lemmas 3.1 and 3.2, and Theorems 3.2 and 3.3.

Theorem 3.4. Let \(k \geq 0\) be an integer and let \(T^0_h\), \(U_h\), and \(S_h\) be as above. Then problem (3.12) has a unique solution \((\mathbf{\phi}_h, \mathbf{u}_h, \lambda_h, \mathbf{\psi}_h) \in T^0_h \times U_h \times \mathbb{R} \times S_h\) satisfying, for some constants \(C_1, C_2, C_3, C_4 > 0\) independent of \(h\),
\[
\|\mathbf{\phi} - \mathbf{\phi}_h, \mathbf{u} - \mathbf{u}_h, \lambda - \lambda_h\|_{T \times U \times \mathbb{R}} \leq h^{k+1} \left\{ C_1 \left(\|\mathbf{\phi}\|_{L^2, \Omega} + \|\mathbf{u}\|_{L^2, \Omega} + \|\text{div} \mathbf{\psi}\|_{L^2, \Omega}\right) + C_2 \left(\|\mathbf{\psi}\|_{L^2, \Omega} + \|\text{div} \mathbf{\psi}\|_{L^2, \Omega}\right) \right\},
\]
and
\[
\|\mathbf{\psi} - \mathbf{\psi}_h\|_{S} \leq h^{k+1} \left\{ C_3 \left(\|\mathbf{\phi}\|_{L^2, \Omega} + \|\mathbf{u}\|_{L^2, \Omega}\right) + C_4 \left(\|\mathbf{\psi}\|_{L^2, \Omega} + \|\text{div} \mathbf{\psi}\|_{L^2, \Omega}\right) \right\}.
\]
Remark 3.6. In the above error estimates, \( C_1 \) and \( C_4 \) are \( O(M/m) \), \( C_2 \) is \( O(1/m) \), while \( C_3 \) is \( O(M^2/m) \). Further dependence of the errors on \( M_h \) and \( m_h \) can be observed through the application of the a priori estimates (3.7) and (3.8).

Remark 3.7. When the above analysis is applied to the variational formulation (2.18)–(2.20) with non-trace-free \( \phi \) (i.e., the variational formulation includes the pressure \( p \)), it is not necessary to enrich the approximation space for \( \phi_h \) as is done in [19, 22] and the above choice for \( T_h^0 \) is sufficient for well-posedness of the discrete problem, with the choice of \( \mathbb{P}_k \) for the approximation space for \( p \).

3.3. A Posteriori Error Estimates and Adaptive Computation

It is briefly noted that the analysis of a posteriori error estimates given in Lemmas 4, 5, 7, and Theorem 6 of Section 4 of [19] can be directly applied to the problem (3.9)–(3.11) for \( n = 2 \) with nominal changes in the proofs and results (the pressure \( p \) can be dropped from the estimates derived there). Additionally, the adaptive computation algorithm described in Section 4 of [19] can be employed.

4. NEW PENALTY METHODS FOR THE STEADY STOKES PROBLEM

In this section new penalty methods for computing discrete approximations to the steady-state Stokes problem are presented. Traditional penalty methods for the Stokes problem [1, 2, 5] are sometimes known as “artificial compressibility” techniques and are motivated by the potential elimination of the pressure variable from the velocity-pressure mixed formulation. These are accomplished by adding a constant multiple of the pressure space inner product to the formulation. It is then shown that as the constant tends to zero, the penalized solution converges to the unpenalized solution.

The methods described here stand in contrast to the traditional penalty methods in that inner products on the velocity space and pseudostress space are introduced into the dual-mixed variational formulation of the steady Stokes problem. The first method introduces the velocity inner product as a way to eliminate the velocity and velocity gradient from the linear system of equations, leaving the pseudostress as the only unknown. This essentially transforms the steady Stokes problem to the generalized one that has been analyzed above, which has some properties that make it advantageous in terms of solution of linear systems, as will be described in Section 5. The second method introduces the pseudostress inner product as a way to eliminate that variable from the system of equations, leaving the velocity, its gradient, and the scalar Lagrange multiplier as the remaining unknowns.

The penalty methods are constructed from the discrete form of the dual-mixed variational formulation of the steady Stokes problem as given in [20]. This method is identical to the method (3.9)–(3.11) with \( \alpha = 0 \): Given \( f \) \( \in \mathcal{U} \) and \( \mathbf{u}_T \) \( \in H^{1/2}(\Gamma) \), find \( (\phi_h, \mathbf{u}_h, \lambda_h, \psi_h) \in T_h^0 \times \mathcal{U}_h \times \mathbb{R} \times S_h \) such that

\[
\nu \int_\Omega \phi_h : \mathbf{q}_h \, d\Omega - \int_\Omega \psi_h : \mathbf{q}_h \, d\Omega = 0, \tag{4.1}
\]

\[- \int_\Omega \mathbf{v}_h \cdot \text{div} \psi_h \, d\Omega + \eta_h \int_\Omega \text{tr}(\psi_h) \, d\Omega = \int_\Omega f \cdot \mathbf{v}_h \, d\Omega, \tag{4.2}\]

\[- \int_\Omega \phi_h : \mathbf{\tau}_h \, d\Omega - \int_\Omega \mathbf{u}_h \cdot \text{div} \mathbf{\tau}_h \, d\Omega + \lambda \int_\Omega \text{tr}(\mathbf{\tau}_h) \, d\Omega = - \int_\Gamma \mathbf{u}_T \cdot (\mathbf{\tau}_h \cdot \mathbf{n}) \, d\Gamma, \tag{4.3}\]

for all \( (\phi_h, \mathbf{v}_h, \eta_h, \mathbf{\tau}_h) \in T_h^0 \times \mathcal{U}_h \times \mathbb{R} \times S_h \). As shown in [20], the above problem is well-posed if the inf-sup conditions (3.14) and (3.15) hold with the chosen finite-dimensional subspaces \( T_h^0, \mathcal{U}_h, \) and \( S_h \).
Corresponding a priori and error estimates are shown there as well. Theorem 2.2 implies that
\[
\inf_{\tau_h \in S_h} \sup_{(\phi_h, u_h, \lambda_h) \in T_h^0 \times U_h \times R} \frac{b_1(\tau_h, \phi_h) + b_2(\tau_h, (u_h, \lambda_h))}{\|\phi_h, u_h, \lambda_h\|_{T_h^0 \times U_h \times R}} \geq \beta_h,
\]
holds as well, which will be used in the analysis below.

### 4.1. Penalty Method 1

In this section the constant \( \alpha \) in (3.9)--(3.11) is replaced with a small parameter, \( \varepsilon_1 \). When \( \varepsilon_1 > 0 \), the results from Section 3 can be applied to show well-posedness of the penalized problem. The result below shows that the difference between the penalized solution and the solution to (4.1)--(4.3) goes to zero as the size of \( \varepsilon_1 \) goes to zero.

**Theorem 4.1.** Let \((\phi_h, u_h, \lambda_h, \psi_h) \in T_h^0 \times U_h \times R \times S_h \) be the solution to (4.1)--(4.3) and let \((\phi_e, u_e, \lambda_e, \psi_e) \in T_h^0 \times U_h \times R \times S_h \) be the solution to (3.9)--(3.11) with \( \varepsilon_1 \) in place of \( \alpha \). Then
\[
\|\phi_h - \phi_e\|_T \leq \varepsilon_1 \left( \frac{\|b_1\|}{\sqrt{\beta_2 h}} \right) \|\phi_h, \lambda_h\|_{U \times R},
\]
\[
\|\phi_h, \lambda_h\| - |u_e, \lambda_e| \leq \varepsilon_1 \left( \frac{\|b_1\|^2}{\sqrt{\beta_2 h}} \right) \|\phi_h, \lambda_h\|_{U \times R},
\]
and
\[
\|\psi_h - \psi_e\|_S \leq \varepsilon_1 \left( \frac{1}{\beta_h} \left( 1 + \frac{\|b_1\|}{\beta_2 h} + \varepsilon_1 \frac{\|b_1\|^2}{\sqrt{\beta_2 h}} \right) \right) \|\phi_h, \lambda_h\|_{U \times R}.
\]

**Proof**

Using the notation established in Sections 2 and 3 above, the solution \((\phi_h, u_h, \lambda_h, \psi_h)\) satisfies
\[
a_1(\phi_h, \zeta_h) + b_1(\psi_h, \zeta_h) = 0,
\]
\[
a_1(\phi_e, \zeta_e) + b_1(\psi_e, \zeta_e) = (f, v_h),
\]
while the approximation \((\phi_e, u_e, \lambda_e, \psi_e)\) satisfies
\[
a_1(\phi_e, \zeta_e) + b_1(\psi_e, \zeta_e) = 0,
\]
\[
a_1(\phi_e - \phi_h, \zeta_e - \zeta_h) + b_2(\psi_e - \psi_h, (v_h, \eta_h)) = (f, v_h),
\]
for all \((\zeta_h, v_h, \eta_h, \tau_h) \in T_h^0 \times U_h \times R \times S_h \). For ease of notation, set \( \bar{u}_e = (u_e, \lambda_e) \) and \( \bar{u}_h = (u_h, \lambda_h) \). Subtract (4.8) from (4.9) and set \( \zeta_h = \phi_e - \phi_h, \bar{v}_h = \bar{u}_e - \bar{u}_h, \) and \( \tau_h = \psi_e - \psi_h \). Then
\[
a_1(\phi_e - \phi_h, \phi_e - \phi_h) + b_1(\psi_e - \psi_h, \phi_e - \phi_h) = 0,
\]
\[
a_1(\bar{u}_e, \bar{u}_e - \bar{u}_h) + b_2(\psi_e - \psi_h, \bar{u}_e - \bar{u}_h) = 0,
\]
Now (3.15), the continuity of \( b_1 \), and the first and third equations of (4.10) imply that
\[
\beta_2 h \|\bar{u}_e - \bar{u}_h\|_{U \times R} \leq \frac{b_2(\psi_e - \psi_h, \bar{u}_e - \bar{u}_h)}{\|\psi_e - \psi_h\|_S} \leq \frac{-b_1(\psi_e - \psi_h, \phi_e - \phi_h)}{\|\psi_e - \psi_h\|_S} \leq \|b_1\| \|\phi_e - \phi_h\|_T.
\]
Adding the first two equations of (4.10) and subtracting the third gives the relation
\[ a_1(\phi_e - \phi_h, \phi_e - \phi_h) + e_1(\bar{u}_e, \bar{u}_e - \bar{u}_h) = 0, \]
or
\[ a_1(\phi_e - \phi_h, \phi_e - \phi_h) + e_1\|\bar{u}_e - \bar{u}_h\|_{U^\infty;\Omega}^2 = e_1(\bar{u}_h, \bar{u}_h - \bar{u}_e), \]
which through coercivity of \( a_1 \) and Hölder’s inequality gives
\[ v\|\phi_e - \phi_h\|_T^2 \leq e_1\|\bar{u}_h\|_{U^\infty;\Omega}\|\bar{u}_h - \bar{u}_e\|_{U^\infty;\Omega}. \tag{4.12} \]
Together (4.11) and (4.12) imply
\[ \|\phi_e - \phi_h\|_T \leq \left( \frac{||b_1||_1}{v\beta_2h} \right) e_1\|\bar{u}_h\|_{U^\infty;\Omega}, \tag{4.13} \]
which also bounds \( \bar{u}_e - \bar{u}_h \):
\[ \|\bar{u}_e - \bar{u}_h\|_{U^\infty;\Omega} \leq \left( \frac{||b_1||_2^2}{v\beta_2h^2} \right) e_1\|\bar{u}_h\|_{U^\infty;\Omega}. \tag{4.14} \]
It remains to bound \( \psi_e - \psi_h \). The inf-sup condition (4.4) and (4.10) imply that
\[
\beta_h \|\psi_h - \psi_e\|_{S} \leq \frac{b_1(\psi_h - \psi_e, \phi_e - \phi_h) + b_2(\psi_h - \psi_e, \bar{u}_e - \bar{u}_h)}{||\phi_e - \phi_h, \bar{u}_e - \bar{u}_h||_{T\times U^\infty;\Omega}} \\
= -\frac{b_1(\psi_h - \psi_e, \phi_e - \phi_h)}{||\phi_e - \phi_h, \bar{u}_e - \bar{u}_h||_{T\times U^\infty;\Omega}} + \frac{-b_2(\psi_h - \psi_e, \bar{u}_e - \bar{u}_h)}{||\phi_e - \phi_h, \bar{u}_e - \bar{u}_h||_{T\times U^\infty;\Omega}} \\
\leq \frac{-b_1(\psi_h - \psi_e, \phi_e - \phi_h)}{||\phi_e - \phi_h||_T} + \frac{e_1(\bar{u}_e, \bar{u}_e - \bar{u}_h)}{||\bar{u}_e - \bar{u}_h||_{U^\infty;\Omega}} \\
= \frac{a_1(\phi_e - \phi_h, \phi_e - \phi_h)}{||\phi_e - \phi_h||_T} + \frac{e_1(\bar{u}_e, \bar{u}_e - \bar{u}_h)}{||\bar{u}_e - \bar{u}_h||_{U^\infty;\Omega}} + \frac{e_1(\bar{u}_h, \bar{u}_h - \bar{u}_e)}{||\bar{u}_e - \bar{u}_h||_{U^\infty;\Omega}} \\
\leq v\|\phi_e - \phi_h\|_T + e_1\|\bar{u}_e - \bar{u}_h\|_{U^\infty;\Omega} + e_1\|\bar{u}_h\|_{U^\infty;\Omega}. \tag{4.15} \]
Application of (4.13) and (4.14) yield the desired result.

**Remark 4.1.** When the finite element subspaces \( S_h, T^0_h, \) and \( U_h \) are Raviart–Thomas elements of order \( k \geq 0 \) and discontinuous piecewise polynomials of degree \( k \), then it suffices to take \( e_1 \approx O(h^{k+1}) \) to preserve the error estimate given in Theorem 4.3 of [20].

### 4.2. Penalty Method 2

In this section, a constant multiple of the inner product on \( S_h \) is added into the variational formulation. If efficient solvers for the \( H(div;\Omega) \) inner product are available, then the pseudostress variable can be eliminated from the problem, resulting in a system with unknowns of velocity, velocity gradient, and Lagrange multiplier. This may be useful when higher-order finite elements are used, as the use of discontinuous piecewise polynomials for the velocity and velocity gradient unknowns yield matrices that are block-diagonal with small size blocks.
Let $\epsilon_2 > 0$ and let $(\psi, \tau)_S = (\psi, \tau) + (\operatorname{div} \psi, \operatorname{div} \tau)$ represent the $H(\operatorname{div}; \Omega)$ inner product. Then with the usual definitions of $a_1, b_1, b_2$, the variational formulation of Penalty Method 2 is:

$$
\begin{align*}
& a_1(\phi_e, \xi_h) + b_1(\phi_e, \eta_h) = 0, \\
& b_1(\tau_h, \phi_e) + b_2(\psi_e, (v_h, \eta_h)) = (f, v_h), \\
& b_1(\tau_h, \phi_e) + b_2(\tau_h, (u_e, \lambda_e)) - \epsilon_2(\tau_h, \tau_h)_S = (u_f, \tau_h \cdot n)_T,
\end{align*}
$$

(4.16)

for all $(\xi_h, v_h, \eta_h, \tau_h) \in T_h^0 \times U_h \times \mathbb{R} \times S_h$. In [20], (4.8) was considered a twofold saddle point problem and the conditions (3.14) and (3.15) were used to show well-posedness. However, Lemma 3.1 and Theorem 3.2 of [9] show that problem (4.16) can also be treated as the single saddle point problem

$$
\begin{bmatrix}
\tilde{A} & B^T \\
B & 0
\end{bmatrix}
\begin{bmatrix}
\phi_h \\
\psi_h
\end{bmatrix}
= \begin{bmatrix}
-f \\
\epsilon_2(\tau_h \cdot n)_T
\end{bmatrix}
\quad \text{where} \quad \tilde{A} = \begin{bmatrix}
A_1 \\
0
\end{bmatrix}
$$

and $B$ is as defined in (2.32). Thus classical results, such as Theorem I.5.3 in [1] or Theorem II.1.2 and Proposition II.2.12 in [5] can be applied to the penalized problem (4.16) to show existence, uniqueness, and error estimates.

**Theorem 4.2.** Let $(\phi_h, u_h, \lambda_h, \psi_h) \in T_h^0 \times U_h \times \mathbb{R} \times S_h$ be the solution to (4.1)–(4.3) and let $(\phi_e, u_e, \lambda_e, \psi_e) \in T_h^0 \times U_h \times \mathbb{R} \times S_h$ be the solution to (4.16). Then there is a constant $C > 0$, independent of $\epsilon_2$, such that

$$
\|\phi_h - \phi_e\|_T + \|u_h - u_e\|_U + \|\lambda_h - \lambda_e\|_G + \|\psi_h - \psi_e\|_S \leq C \epsilon_2 \|\psi_h\|_S.
$$

(4.17)

5. **COMPUTATIONAL ASPECTS OF THE METHODS**

In this section, numerical approaches to solving the linear systems that arise in the above methods are briefly discussed. A comprehensive discussion of the numerical solution of saddle point problems is given in [37] and is beyond the scope of this work. The similarities between the methods described here and existing pseudostress methods [23, 24, 25] are also described.

5.1. **Solution of Linear Systems Arising in the Methods**

The linear system of equations arising from the dual-mixed method for the generalized Stokes problem (3.12) and Penalty Method 1 for the steady Stokes system (4.9) has a structure that can be exploited. The linear systems will have the form

$$
\begin{bmatrix}
A_1 & B_1^T \\
0 & A_2 \\
B_1 & B_2
\end{bmatrix}
\begin{bmatrix}
\Phi_h \\
U_h \\
\Psi_h
\end{bmatrix}
= \begin{bmatrix}
F \\
0
\end{bmatrix}
$$

(5.1)

where $\Phi_h$, $U_h$, and $\Psi_h$ represent the coefficients of the unknowns $\phi_h$, $\bar{u}_h = (u_h, \lambda_h)$, and $\psi_h$, respectively, in the appropriate finite element bases, and $F$ and $G$ represent the products $(f, v_h)$ and $(\tau_h \cdot n)_T$ in terms of the appropriate bases. In (5.1) the coefficient matrix is symmetric positive definite (SPD) and both $B_1 (\ell \times m)$ and $B_2 (\ell \times n)$ have full rank, i.e., rank$(B_1) = m$ and rank$(B_2) = n$. The matrix $A ((m+n) \times (m+n))$ is SPD, and using Sylvester’s Law of Inertia [37] it can be shown that the coefficient matrix of (5.1) has $m+n$ positive eigenvalues and $\ell$ negative eigenvalues.
In a variational sense, the reduction to the system (5.2) can be seen by letting \( v \) be of size 3 triangles and faces, respectively, in two space dimensions. In three space dimensions, the system will order case (\( \tau \) for all \( \eta \)) resulting in a Schur complement system for the pseudostress only: 

\[
\Phi \text{ immediately available as well. This then leads to the elimination of the variables } (\Phi_h, U_h) \text{ from (5.1), resulting in a Schur complement system for the pseudostress only:}
\]

\[
BA^{-1}B^T\Psi_h = BA^{-1}F - G
\]

In a variational sense, the reduction to the system (5.2) can be seen by letting \( v_h := -\text{div} \tau_h, \eta_h := \int_\Omega \text{tr}(\tau_h) \, d\Omega, \) and \( \varsigma := \tau^0 = \tau - (1/d)\text{tr}(\tau)I \) in (3.9)–(3.10). Note that

\[
\psi_h : \tau^0_h = \psi_h : \tau_h - \frac{1}{d} \psi_h : \text{tr}(\tau_h)I = \psi_h : \tau_h - \frac{1}{d} \sum_{i=1}^{d} \psi_{ii} \left( \sum_{j=1}^{d} \tau_{jj} \right)
\]

Then from (3.9),

\[
\int_\Omega \phi_h : \tau^0_h \, d\Omega = \frac{1}{\nu} \int_\Omega \psi_h : \tau^0_h \, d\Omega = \frac{1}{\nu} \int_\Omega \psi^0_h : \tau_h \, d\Omega
\]

and (3.10) implies

\[
- \int_\Omega \upsilon_h \cdot \text{div} \tau_h \, d\Omega + \lambda_h \int_\Omega \text{tr}(\tau_h) \, d\Omega = \frac{1}{\alpha} \left( \int_\Omega f \cdot \text{div} \tau_h \, d\Omega - \int_\Omega \text{div} \tau_h \cdot \text{div} \psi_h \, d\Omega + \left( \int_\Omega \text{tr}(\tau_h) \, d\Omega \right) \left( \int_\Omega \text{tr}(\psi_h) \, d\Omega \right) \right).
\]

Substituting these into (3.11) reduces to the problem

\[
\frac{1}{\nu} \int_\Omega \psi^0_h : \tau_h \, d\Omega + \frac{1}{\alpha} \int_\Omega \text{div} \tau_h \cdot \text{div} \psi_h \, d\Omega + \frac{1}{\alpha} \left( \int_\Omega \text{tr}(\tau_h) \, d\Omega \right) \left( \int_\Omega \text{tr}(\psi_h) \, d\Omega \right) = \int_\Gamma u_\Gamma : (\tau_h \cdot n) \, d\Gamma - \frac{1}{\alpha} \int_\Omega f \cdot \text{div} \tau_h \, d\Omega,
\]

for all \( \tau_h \in S_h \).

The Schur complement matrix \( BA^{-1}B^T \) of (5.2) is symmetric and positive definite [37]. In the lowest-order case (\( k = 0 \)), as \( h \to 0 \) this matrix will be of size \( 3N_t = 2N_f \) where \( N_t \) and \( N_f \) are the number of triangles and faces, respectively, in two space dimensions. In three space dimensions, the system will be of size \( 3N_f \).

Once the system (5.2) has been solved for \( \Psi_h \), the solution components \( (\Phi_h, U_h) \) can be recovered through the formula

\[
\begin{bmatrix} \Phi_h \\ U_h \end{bmatrix} = A^{-1} \begin{bmatrix} F - B^T \Psi_h \end{bmatrix} = \begin{bmatrix} A_1^{-1} & 0 \\ 0 & A_2^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ B_1^T \\ B_2^T \end{bmatrix} \Psi_h.
\]

Copyright © 2009 John Wiley & Sons, Ltd.

It may be the case that the system (5.2) is dense and/or ill-conditioned. Conditioning of this system, preconditioning approaches, and numerical algorithms tailored to this system will be studied in subsequent work.

5.2. Relation to Existing Pseudostress Methods

The methods (3.12) and (4.9) share some similarities with the pseudostress methods presented in [23, 24, 25]. In these works, the variational formulation is obtained without the introduction of the velocity gradient as an additional variable, and the constitutive law (2.9) is inverted to express $\nabla u$ in terms of $\psi$. The resulting variational formulation for the generalized Stokes problem is: Given $f \in U$ and $u_\Gamma \in H^{1/2}(\Gamma)$, find $(\psi_h, u_h) \in S_h^0 \times U_h$ such that

$$\int_{\Omega} A(\psi_h) : \tau_h \, d\Omega + \int_{\Omega} u_h \cdot \text{div} \tau_h \, d\Omega = \int_{\Gamma} u_\Gamma \cdot (\tau_h \cdot n) \, d\Gamma, \quad (5.5)$$

$$\int_{\Omega} v_h \cdot \text{div} \psi_h \, d\Omega - \frac{1}{\alpha} \int_{\Omega} u_h \cdot v_h \, d\Omega = \int_{\Omega} f \cdot v_h \, d\Omega. \quad (5.6)$$

Here $S_h^0$ is the space of tensors in $S_h$ with mean trace zero and $A : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d}$ is the linear (singular) map

$$A \tau = \tau - \frac{1}{d} \text{tr}(\tau) I.$$ 

The parameter $\alpha$ can be replaced with $\epsilon_1$ for a penalty method for the steady Stokes problem. It is shown in [25] that $RT_k - P_k$ elements can be used for finite element approximation of the pseudostress and velocity. Following the discussion in Section 4 of [25], the velocity is eliminated from the system by choosing $v_h = \text{div} \tau_h$ in (5.6) and the resulting problem for the pseudostress has the form

$$\int_{\Omega} (A \psi_h, \tau_h) + \frac{1}{\alpha} (\text{div} \psi_h, \text{div} \tau_h) = (u_\Gamma, \tau_h \cdot n)_{\Gamma} - \frac{1}{\alpha} (f, \text{div} \tau_h), \quad \forall \tau \in S_h^0. \quad (5.7)$$

It can be seen that, with the exception of the spaces $S_h$ and $S_h^0$ and the inclusion of the Lagrange multiplier $\lambda$ (which is optional, as noted in Remark 2.2), the variational problem (5.3) and linear system (5.2) arising from the methods (3.12) and (4.9) is the same as problem (5.7) and the linear system produced by it. The main distinctions between the two approaches (5.5)–(5.6) and (3.12) thus lie in

- the treatment of the constitutive law (2.9), which must be inverted for (5.5)–(5.6) but not for (3.12) and (4.9),
- the analysis required to show well-posedness and error estimates,
- and the inclusion of $\nabla u$ as an auxiliary variable in (3.12) and (4.9).

Thus the multigrid methods and preconditioners in [24, 25] can be easily adapted for the methods (3.12) and (4.9).

6. NUMERICAL EXPERIMENTS

In this section, numerical experiments that support the theoretical results outlined in Sections 3 and 4 are presented. The experiments are divided into two main categories: examples that illustrate the
theoretical rate of convergence of the dual-mixed method in Section 3, and examples that investigate
the application of the penalty methods in Section 4.
Computations are performed using the FreeFEM++ finite element software package [38]. All
computations below are performed in the lowest-order finite element space case $(k = 0)$ in two space
dimensions. The data $f$ and $u_B$ are chosen such that the solutions $u, \psi$ to (2.8)–(2.12) can be
described analytically, and errors are computed using these. Three different example problems are
used: the first two are taken from [19] and the third from [25]. Details of $\Omega, u$, and $p$ are given in Table
6.1. The true values of $\phi$ and $\psi$ can be computed from $u$ and $p$ using (2.8)–(2.9).

6.1. Generalized Stokes Problems
In this section, numerical results for the method (3.12) are given for Examples I and II. The method
(3.12) will be denoted the “Traceless Gradient Method.” Additionally, numerical results for these
problems in [19] using the method derived there (“Enriched Method”) are presented for comparison.
The computations are performed under uniform mesh refinement and a global rate of convergence $\theta$
is computed from the global solution norm

$$\|\phi - \phi_h\|_P + \|u - u_h\|_{L^2} + \|\psi - \psi_h\|_{L^2} + \|p - p_h\|_P + |\lambda - \lambda_h|,$$

where $P = L^2(\Omega)$. For ease of notation, let $e = \|\phi - \phi_h\|_P + \|u - u_h\|_{L^2} + \|\psi - \psi_h\|_{L^2} + \|p - p_h\|_P + |\lambda - \lambda_h|$, $e(u) = \|u - u_h\|_{L^2}$, $e(\psi) = \|\psi - \psi_h\|_{L^2}$, and $e(p) = \|p - p_h\|_P$.

6.1.1. Example I For this example, computations are performed for two different choices of the
parameters $\alpha$ and $\rho$: Case 1 $(\alpha = 10)$ and Case 2 $(\alpha = 100)$. The results are given in Tables
6.2 (Case 1) and 6.3 (Case 2). In Case 1, the error computations of both methods are very
similar across all mesh sizes. For Case 2, while the approximation error of the Traceless Gradient
Method is significantly larger for coarse meshes, as $h$ decreases the errors of the Traceless Gradient
Method approach the same values generated by the Enriched Method, due to the larger convergence
rate at coarse meshes. The convergence behavior of both methods is displayed in Figure 6.1.

6.1.2. Example II For this example, computations are also performed for two different choices of the
parameters $\alpha$ and $\rho$: Case 1 $(\alpha = 100)$ and Case 2 $(\alpha = 1000)$. The results are given in Tables
6.4 (Case 1) and 6.5 (Case 2). For Example II, the global error in both methods is dominated
by the error in approximating $\psi$. This is consistent with the theoretical error estimate in Section (3), as
the constant in the error estimate for $\psi$ will be $O(\alpha)$ larger than the constant in the error estimates for
the other solution components (see Remark 3.3). Again it is observed that the approximation error of
the Traceless Gradient Method is significantly larger for coarse meshes, but as $h$ decreases the errors

<table>
<thead>
<tr>
<th>Example</th>
<th>$\Omega$</th>
<th>$u_1(x,y)$</th>
<th>$u_2(x,y)$</th>
<th>$p(x,y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$[-1, 1] \times [-1, 1]$</td>
<td>$e^y(y \cos y + \sin y)$</td>
<td>$e^y \sin(y)$</td>
<td>$2e^y \sin y$</td>
</tr>
<tr>
<td>II</td>
<td>$[0, 1] \times [0, 1]$</td>
<td>$-\sqrt{1-e^{-\sqrt{x+y}}}$</td>
<td>$-u_1(x,y)$</td>
<td>$2e^{2x-1} \sin(2y - 1)$</td>
</tr>
<tr>
<td>III</td>
<td>$[0, 1] \times [0, 1]$</td>
<td>$\sin(2\pi x) \cos(2\pi y)$</td>
<td>$-\cos(2\pi x) \sin(2\pi y)$</td>
<td>$x^2 + y^2$</td>
</tr>
</tbody>
</table>
Table 6.2. Error results for Example I, Case 1 ($\nu = 1, \alpha = 10$).

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2664</td>
<td>2.2598</td>
<td>2.3904</td>
<td>4.6497</td>
<td>5.8348</td>
<td>–</td>
</tr>
<tr>
<td>1/2</td>
<td>0.6482</td>
<td>1.2970</td>
<td>1.4075</td>
<td>2.5675</td>
<td>3.2673</td>
<td>0.84</td>
</tr>
<tr>
<td>1/4</td>
<td>0.3246</td>
<td>0.6504</td>
<td>0.7900</td>
<td>1.2917</td>
<td>1.6796</td>
<td>0.96</td>
</tr>
<tr>
<td>1/8</td>
<td>0.1616</td>
<td>0.2797</td>
<td>0.4213</td>
<td>0.953</td>
<td>0.7976</td>
<td>1.07</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0806</td>
<td>0.1198</td>
<td>0.2164</td>
<td>0.2776</td>
<td>0.3805</td>
<td>1.07</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0403</td>
<td>0.0548</td>
<td>0.1092</td>
<td>0.1343</td>
<td>0.1860</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Table 6.3. Error results for Example I, Case 2 ($\nu = 1, \alpha = 100$).

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2494</td>
<td>5.8447</td>
<td>2.3488</td>
<td>16.7981</td>
<td>17.9838</td>
<td>–</td>
</tr>
<tr>
<td>1/2</td>
<td>0.6407</td>
<td>2.9519</td>
<td>1.2689</td>
<td>7.3661</td>
<td>8.0619</td>
<td>1.16</td>
</tr>
<tr>
<td>1/4</td>
<td>0.3226</td>
<td>1.4973</td>
<td>0.7086</td>
<td>3.6158</td>
<td>3.9903</td>
<td>1.01</td>
</tr>
<tr>
<td>1/8</td>
<td>0.1614</td>
<td>0.6203</td>
<td>0.3964</td>
<td>1.4834</td>
<td>1.6638</td>
<td>1.26</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0806</td>
<td>0.2109</td>
<td>0.2118</td>
<td>0.5071</td>
<td>0.5942</td>
<td>1.49</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0403</td>
<td>0.0722</td>
<td>0.1086</td>
<td>0.1767</td>
<td>0.2232</td>
<td>1.41</td>
</tr>
</tbody>
</table>

Table 6.2. Error results for Example I, Case 1 ($\nu = 1, \alpha = 10$).

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2664</td>
<td>2.2598</td>
<td>2.3904</td>
<td>4.6497</td>
<td>5.8348</td>
<td>–</td>
</tr>
<tr>
<td>1/2</td>
<td>0.6482</td>
<td>1.2970</td>
<td>1.4075</td>
<td>2.5675</td>
<td>3.2673</td>
<td>0.84</td>
</tr>
<tr>
<td>1/4</td>
<td>0.3246</td>
<td>0.6504</td>
<td>0.7900</td>
<td>1.2917</td>
<td>1.6796</td>
<td>0.96</td>
</tr>
<tr>
<td>1/8</td>
<td>0.1616</td>
<td>0.2797</td>
<td>0.4213</td>
<td>0.953</td>
<td>0.7976</td>
<td>1.07</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0806</td>
<td>0.1198</td>
<td>0.2164</td>
<td>0.2776</td>
<td>0.3805</td>
<td>1.07</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0403</td>
<td>0.0548</td>
<td>0.1092</td>
<td>0.1343</td>
<td>0.1860</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Table 6.4. Error results for Example II, Case 1 ($\nu = 0.5, \alpha = 100$).

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>0.5701</td>
<td>5.8974</td>
<td>7.7005</td>
<td>57.8684</td>
<td>58.6784</td>
<td>-</td>
</tr>
<tr>
<td>1/4</td>
<td>0.5265</td>
<td>3.3068</td>
<td>7.3710</td>
<td>52.9618</td>
<td>53.5770</td>
<td>0.13</td>
</tr>
<tr>
<td>1/8</td>
<td>0.3326</td>
<td>1.3674</td>
<td>4.7051</td>
<td>33.630</td>
<td>33.7226</td>
<td>0.67</td>
</tr>
<tr>
<td>1/16</td>
<td>0.1771</td>
<td>0.5061</td>
<td>2.5592</td>
<td>17.7459</td>
<td>17.9375</td>
<td>0.91</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0898</td>
<td>0.1981</td>
<td>1.3160</td>
<td>8.9999</td>
<td>9.0982</td>
<td>0.98</td>
</tr>
<tr>
<td>1/64</td>
<td>0.0451</td>
<td>0.0872</td>
<td>0.6638</td>
<td>4.5145</td>
<td>4.5641</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>0.5032</td>
<td>3.2059</td>
<td>5.4391</td>
<td>50.7765</td>
<td>51.1700</td>
<td>-</td>
</tr>
<tr>
<td>1/4</td>
<td>0.5047</td>
<td>2.4017</td>
<td>4.7510</td>
<td>50.7097</td>
<td>50.9908</td>
<td>0.01</td>
</tr>
<tr>
<td>1/8</td>
<td>0.3270</td>
<td>1.1490</td>
<td>2.7950</td>
<td>32.8015</td>
<td>32.9420</td>
<td>0.63</td>
</tr>
<tr>
<td>1/16</td>
<td>0.1760</td>
<td>0.4435</td>
<td>1.5002</td>
<td>17.6482</td>
<td>17.7182</td>
<td>0.89</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0896</td>
<td>0.1716</td>
<td>0.7768</td>
<td>8.9857</td>
<td>9.0213</td>
<td>0.97</td>
</tr>
<tr>
<td>1/64</td>
<td>0.0450</td>
<td>0.0740</td>
<td>0.3937</td>
<td>4.5126</td>
<td>4.5305</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 6.5. Error results for Example II, Case 2 ($\nu = 0.5, \alpha = 1000$).

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>0.3132</td>
<td>31.2277</td>
<td>5.8181</td>
<td>317.2370</td>
<td>318.8240</td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>0.4149</td>
<td>19.6693</td>
<td>18.8422</td>
<td>416.1100</td>
<td>417.0010</td>
<td>-0.39</td>
</tr>
<tr>
<td>1/8</td>
<td>0.5765</td>
<td>14.5713</td>
<td>25.8863</td>
<td>577.0040</td>
<td>577.7680</td>
<td>-0.47</td>
</tr>
<tr>
<td>1/16</td>
<td>0.4645</td>
<td>7.3428</td>
<td>20.6140</td>
<td>464.7110</td>
<td>465.2270</td>
<td>0.31</td>
</tr>
<tr>
<td>1/32</td>
<td>0.2712</td>
<td>2.8710</td>
<td>12.2047</td>
<td>271.2680</td>
<td>271.5570</td>
<td>0.78</td>
</tr>
<tr>
<td>1/64</td>
<td>0.1410</td>
<td>1.0880</td>
<td>6.4634</td>
<td>141.0060</td>
<td>141.1590</td>
<td>0.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>0.0643</td>
<td>5.0629</td>
<td>7.0383</td>
<td>65.0020</td>
<td>65.5777</td>
<td>-</td>
</tr>
<tr>
<td>1/4</td>
<td>0.3337</td>
<td>7.4337</td>
<td>12.9106</td>
<td>334.0185</td>
<td>334.3507</td>
<td>-2.35</td>
</tr>
<tr>
<td>1/8</td>
<td>0.5382</td>
<td>8.1264</td>
<td>17.7197</td>
<td>538.4277</td>
<td>538.7808</td>
<td>-0.69</td>
</tr>
<tr>
<td>1/16</td>
<td>0.4502</td>
<td>5.3283</td>
<td>12.7694</td>
<td>450.3525</td>
<td>450.5652</td>
<td>0.26</td>
</tr>
<tr>
<td>1/32</td>
<td>0.2680</td>
<td>2.3559</td>
<td>7.1160</td>
<td>268.0732</td>
<td>268.1781</td>
<td>0.75</td>
</tr>
<tr>
<td>1/64</td>
<td>0.1404</td>
<td>0.9185</td>
<td>3.7577</td>
<td>140.4863</td>
<td>140.5396</td>
<td>0.93</td>
</tr>
</tbody>
</table>
of the Traceless Gradient Method approach the same values generated by the Enriched Method. The convergence behavior of both methods is displayed in Figure 6.2.

The Traceless Gradient Method reduces the overall number of degrees of freedom required for the discrete problem (3.12) by the Enriched Method. On a single triangle $K \in T_h$, the Enriched Method requires 15 degrees of freedom [19], while the Traceless Gradient Method only requires 11: 6 for the tensor $\psi_h$, 3 for the tensor $\phi_h$, and 2 for the vector $u_h$. For Examples I and II, the total number of degrees of freedom required for different values of $h$ for each method are given in Table 6.6, in which...
Table 6.6. Comparison of number of degrees of freedom required for the methods.

<table>
<thead>
<tr>
<th>$N_T$</th>
<th>$h$</th>
<th>$h$</th>
<th>Enriched Method</th>
<th>Traceless Method</th>
<th>% Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1</td>
<td>1/2</td>
<td>105</td>
<td>73</td>
<td>30.48</td>
</tr>
<tr>
<td>32</td>
<td>1/2</td>
<td>1/4</td>
<td>401</td>
<td>273</td>
<td>31.92</td>
</tr>
<tr>
<td>128</td>
<td>1/4</td>
<td>1/8</td>
<td>1569</td>
<td>1057</td>
<td>32.63</td>
</tr>
<tr>
<td>512</td>
<td>1/8</td>
<td>1/16</td>
<td>6209</td>
<td>4161</td>
<td>32.98</td>
</tr>
<tr>
<td>2048</td>
<td>1/16</td>
<td>1/32</td>
<td>24705</td>
<td>16513</td>
<td>33.16</td>
</tr>
<tr>
<td>8192</td>
<td>1/32</td>
<td>1/64</td>
<td>98561</td>
<td>65793</td>
<td>33.25</td>
</tr>
</tbody>
</table>

$N_T$ denotes the number of triangles in the mesh and $N$ is the total overall number of degrees of freedom required by the method.

6.2. Penalty Methods for the Steady Stokes Problem

In this section it is demonstrated that the two penalty methods described in Section 4 produce approximations that satisfy the results of Theorems 4.1 and 4.2. Error computations of these methods are also compared to results of the Traceless Gradient Method [20] for the steady Stokes problem (4.1)–(4.3).

For Example III with the parameter $\nu = 1$, error computations for each of the three methods are presented in Table 6.7. It is observed that as $h \to 0$ with $\epsilon_1 = \epsilon_2 = h$, the global error in the approximation produced by both penalty methods approaches the global error of the unpenalized method, with all three methods attaining the theoretical rate of convergence. The only significant difference in the errors generated by the three methods is that $e(u)$ is much larger in magnitude for Penalty Method 2 compared to the other two methods. However, the rate of convergence of $e(u)$ approaches 1 for all three methods. The relative performance of the penalty methods for differing values of $\nu$ is also investigated. Figure 6.3 gives the global errors against mesh size of each method for the choices $\nu = 0.001$, $\nu = 1$, and $\nu = 100$. The baseline choice for $\epsilon_1$ and $\epsilon_2$ is $h$, however, other choices of $h$ that improved performance were considered. It is observed in Figure 6.3(a) that the choice $\epsilon_1 = h$ in Penalty Method 1 produces errors that are smaller in magnitude but the rate of convergence is not satisfactory on coarser meshes. Errors and convergence rates for Penalty Method 1 are improved when the penalty parameter is scaled using the viscosity $\nu$, i.e., taking $\epsilon_1 = vh$. In Figure 6.3(c), the choice $\epsilon_2 = h$ for Penalty Method 2 is unsatisfactory as the rate of convergence degrades as $h \to 0$. However, the choice $\epsilon_2 = h/\nu$ improves the performance of Penalty Method 2. In further experimentation not included here for brevity, it has been observed that generally taking $\epsilon_1 = vh$ and $\epsilon_2 = h/\nu$ often results in satisfactory performance of both penalty methods.

ACKNOWLEDGEMENT

This material is based upon work supported by the Center for Nonlinear Analysis (CNA) under the National Science Foundation Grant No. DMS 0635983.
Table 6.7. Error results for unpenalized and penalty methods, Example III, $\nu = 0.5$.

<table>
<thead>
<tr>
<th>h</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5749</td>
<td>1.4231</td>
<td>5.9517</td>
<td>40.3601</td>
<td>40.8254</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>0.3687</td>
<td>0.9956</td>
<td>3.3696</td>
<td>27.2694</td>
<td>27.4973</td>
<td>0.57</td>
</tr>
<tr>
<td>1/4</td>
<td>0.1853</td>
<td>0.5111</td>
<td>1.7646</td>
<td>14.4115</td>
<td>14.5293</td>
<td>0.92</td>
</tr>
<tr>
<td>1/8</td>
<td>0.0926</td>
<td>0.2542</td>
<td>0.8932</td>
<td>7.3084</td>
<td>7.3678</td>
<td>0.98</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0463</td>
<td>0.1266</td>
<td>0.4480</td>
<td>3.6672</td>
<td>3.6970</td>
<td>0.99</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0231</td>
<td>0.0632</td>
<td>0.2242</td>
<td>1.8355</td>
<td>1.8501</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Traceless Gradient Method (unpenalized)

<table>
<thead>
<tr>
<th>h</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5738</td>
<td>1.4148</td>
<td>5.9425</td>
<td>40.3588</td>
<td>40.8225</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>0.3689</td>
<td>0.9945</td>
<td>3.3700</td>
<td>27.2697</td>
<td>27.4976</td>
<td>0.57</td>
</tr>
<tr>
<td>1/4</td>
<td>0.1854</td>
<td>0.5112</td>
<td>1.7648</td>
<td>14.4118</td>
<td>14.5296</td>
<td>0.92</td>
</tr>
<tr>
<td>1/8</td>
<td>0.0926</td>
<td>0.2543</td>
<td>0.8932</td>
<td>7.3086</td>
<td>7.3680</td>
<td>0.98</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0463</td>
<td>0.1266</td>
<td>0.4480</td>
<td>3.6672</td>
<td>3.6970</td>
<td>0.99</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0231</td>
<td>0.0632</td>
<td>0.2242</td>
<td>1.8355</td>
<td>1.8502</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Penalty Method 1 ($\epsilon_1 = h$)

<table>
<thead>
<tr>
<th>h</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.0336</td>
<td>1.5117</td>
<td>5.9845</td>
<td>40.3714</td>
<td>44.6448</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>12.2204</td>
<td>0.9438</td>
<td>3.4206</td>
<td>27.2732</td>
<td>30.0968</td>
<td>0.57</td>
</tr>
<tr>
<td>1/4</td>
<td>6.7737</td>
<td>0.5347</td>
<td>1.8157</td>
<td>14.4195</td>
<td>16.0433</td>
<td>0.91</td>
</tr>
<tr>
<td>1/8</td>
<td>3.4797</td>
<td>0.3002</td>
<td>0.9268</td>
<td>7.3161</td>
<td>8.1598</td>
<td>0.98</td>
</tr>
<tr>
<td>1/16</td>
<td>1.7523</td>
<td>0.1630</td>
<td>0.4674</td>
<td>3.6725</td>
<td>4.0991</td>
<td>0.99</td>
</tr>
<tr>
<td>1/32</td>
<td>0.8779</td>
<td>0.0854</td>
<td>0.2347</td>
<td>1.8383</td>
<td>2.0524</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Penalty Method 2 ($\epsilon_2 = h$)

<table>
<thead>
<tr>
<th>h</th>
<th>$e(u)$</th>
<th>$e(p)$</th>
<th>$e(\phi)$</th>
<th>$e(\psi)$</th>
<th>$e$</th>
<th>$\theta$</th>
</tr>
</thead>
</table>

REFERENCES

18. Figueroa LE, Gatica GN, Márquez A. Augmented mixed finite element methods for the stationary Stokes equations. *SIAM*


