

FINITE ELEMENT APPROXIMATION OF THE DISCRETE FIRST-ORDER SYSTEM LEAST SQUARES FOR ELLIPTIC PROBLEMS

BYEONG CHUN SHIN

ABSTRACT. In [Z. Cai and B. C. Shin, SIAM J. Numer. Anal. 40 (2002), 307–318], we developed the discrete first-order system least squares method for the second-order elliptic boundary value problem by directly approximating $H(\text{div}) \cap H(\text{curl})$ -type space based on the Helmholtz decomposition. Under general assumptions, error estimates were established in the L^2 and H^1 norms for the vector and scalar variables, respectively. Such error estimates are optimal with respect to the required regularity of the solution. In this paper, we study solution methods for solving the system of linear equations arising from the discretization of variational formulation which possesses discrete biharmonic term and focus on numerical results including the performances of multigrid preconditioners and the finite element accuracy.

1. Introduction

In recent, there are substantial interest in the use of least-squares methods for numerical approximation of partial differential equations and system. In [5], we recently developed a discrete first-order system least squares (FOSLS) for the following scalar second-order elliptic partial differential equations:

$$(1.1) \quad \begin{cases} -\nabla \cdot (A\nabla p) + \mathbf{b} \cdot \nabla p + cp & = f & \text{in } \Omega, \\ p & = 0 & \text{on } \Gamma_D, \\ \mathbf{n} \cdot (A\nabla p) & = 0 & \text{on } \Gamma_N \end{cases}$$

Received October 26, 2004.

2000 Mathematics Subject Classification: 65F10, 65F30.

Key words and phrases: least-squares method, multigrid, preconditioner.

This study was financially supported by Chonnam National University in the program, 2002.

where Ω is a bounded, open, and simply connected domain in \mathbb{R}^2 with Lipschitz boundary $\partial\Omega$; A is 2×2 uniformly symmetric positive definite matrix of functions in $L^\infty(\Omega)$; \mathbf{b} and c are the respective vector and scalar functions in $L^\infty(\Omega)$; $f \in L^2(\Omega)$ is a given scalar function; $\partial\Omega = \Gamma_D \cup \Gamma_N$ is the partition of the boundary of Ω ; and \mathbf{n} is the outward unit vector normal to the boundary. For simplicity, assume that both Γ_D and Γ_N are nonempty, with the obvious generalization to quotient spaces when one of them is empty.

The limitation of $L^2(\Omega)$ -norm FOSLS given in [4] is the requirement of sufficient smoothness of the underlying problem. Such smoothness implies the equivalence between homogeneous FOSLS functional and product $H^1(\Omega)$ -norm (See [4] for detail). But, when the domain Ω is not smooth or not convex or the coefficient A is not continuous, we can not guarantee such equivalence. The discrete first-order system least squares method for the second-order elliptic boundary value problem developed in [5] is using the direct approximating $H(\text{div}) \cap H(\text{curl})$ -type space based on the Helmholtz decomposition. Under general assumptions, error estimates were established in the L^2 and H^1 norms for the vector and scalar variables, respectively. Such error estimates are optimal with respect to the required regularity of the solution. In this paper, we study solution methods for solving the system of linear equations arising from the discretization of variational formulation for the discrete least squares method given in [5] which possesses discrete biharmonic term. We also focus on the numerical results including the performances of multigrid preconditioners and the finite element accuracy.

The paper is organized as follows. The L^2 -norm version of the FOSLS approach are introduced in section 2, along with some notations. The discrete FOSLS approach is developed in section 3. In section 4, we discuss the implementation issues. Finally, we report numerical experiment results in section 5.

2. First-order system least squares (FOSLS)

We assume that A is uniformly symmetric positive definite and scaled appropriately, that is, there exist positive constants

$$0 < \lambda \leq 1 \leq \Lambda$$

such that

$$(2.1) \quad \lambda \boldsymbol{\xi}^T \boldsymbol{\xi} \leq \boldsymbol{\xi}^T A \boldsymbol{\xi} \leq \Lambda \boldsymbol{\xi}^T \boldsymbol{\xi}$$

for all $\xi \in \mathfrak{R}^2$ and almost all $x \in \bar{\Omega}$.

We use standard notation and definitions for the Sobolev spaces $H^s(\Omega)^2$, associated inner products $(\cdot, \cdot)_s$, and respective norms $\|\cdot\|_s$, $s \geq 0$. (We suppress the designation Ω on the inner products and norms because dependence on region is clear by context.) $H^0(\Omega)^2$ coincides with $L^2(\Omega)^2$, in which case the norm and inner product will be denoted by $\|\cdot\|$ and (\cdot, \cdot) , respectively. Define subspaces of $H^1(\Omega)$

$$H_D^1(\Omega) = \{q \in H^1(\Omega) : q = 0 \text{ on } \Gamma_D\}$$

and

$$H_N^1(\Omega) = \{q \in H^1(\Omega) : q = 0 \text{ on } \Gamma_N\}.$$

Let $H_D^{-1}(\Omega)$ denote the dual of $H_D^1(\Omega)$ with the norm defined by

$$\|\phi\|_{H_D^{-1}(\Omega)} = \sup_{0 \neq \psi \in H_D^1(\Omega)} \frac{(\phi, \psi)}{\|\psi\|_1}.$$

Denote the curl operator in \mathfrak{R}^2 by

$$\nabla \times = (-\partial_2, \partial_1)$$

and its formal adjoint by

$$\nabla^\perp = \begin{pmatrix} \partial_2 \\ -\partial_1 \end{pmatrix}.$$

Let

$$H(\operatorname{div} A^{\frac{1}{2}}; \Omega) = \{\mathbf{v} \in L^2(\Omega)^2 : \nabla \cdot (A^{\frac{1}{2}} \mathbf{v}) \in L^2(\Omega)\}$$

and

$$H(\operatorname{curl} A^{-\frac{1}{2}}; \Omega) = \{\mathbf{v} \in L^2(\Omega)^2 : \nabla \times (A^{-\frac{1}{2}} \mathbf{v}) \in L^2(\Omega)\},$$

which are Hilbert spaces under norms

$$\|\mathbf{v}\|_{H(\operatorname{div} A^{\frac{1}{2}}; \Omega)} = \left(\|\mathbf{v}\|^2 + \|\nabla \cdot (A^{\frac{1}{2}} \mathbf{v})\|^2 \right)^{\frac{1}{2}}$$

and

$$\|\mathbf{v}\|_{H(\operatorname{curl} A^{-\frac{1}{2}}; \Omega)} = \left(\|\mathbf{v}\|^2 + \|\nabla \times (A^{-\frac{1}{2}} \mathbf{v})\|^2 \right)^{\frac{1}{2}},$$

respectively. When A is the identity matrix, we use the simpler notations $H(\operatorname{div}; \Omega)$ and $H(\operatorname{curl}; \Omega)$. Define the subspaces

$$H_0(\operatorname{div} A^{\frac{1}{2}}; \Omega) = \{\mathbf{v} \in H(\operatorname{div} A^{\frac{1}{2}}; \Omega) : \mathbf{n} \cdot (A^{\frac{1}{2}} \mathbf{v}) = 0 \text{ on } \Gamma_N\}$$

$$H_0(\operatorname{curl} A^{-\frac{1}{2}}; \Omega) = \{\mathbf{v} \in H(\operatorname{curl} A^{-\frac{1}{2}}; \Omega) : \boldsymbol{\tau} \cdot (A^{-\frac{1}{2}} \mathbf{v}) = 0 \text{ on } \Gamma_D\}$$

and denote by

$$\mathcal{U} = H_0(\operatorname{div} A^{\frac{1}{2}}; \Omega) \cap H_0(\operatorname{curl} A^{-\frac{1}{2}}; \Omega),$$

where $\boldsymbol{\tau}$ represents the unit vector tangent to the boundary oriented counterclockwise.

Introducing an independent vector variable

$$\mathbf{u} = A^{\frac{1}{2}} \nabla p,$$

by using the homogeneous Dirichlet boundary condition on Γ_D we have that

$$\nabla \times (A^{-\frac{1}{2}} \mathbf{u}) = 0 \quad \text{in } \Omega \quad \text{and} \quad \boldsymbol{\tau} \cdot (A^{-\frac{1}{2}} \mathbf{u}) = 0 \quad \text{on } \Gamma_D.$$

Then an equivalent extended system for problem (1.1) is

$$(2.2) \quad \left\{ \begin{array}{ll} \mathbf{u} - A^{\frac{1}{2}} \nabla p = \mathbf{0} & \text{in } \Omega, \\ -\nabla \cdot (A^{\frac{1}{2}} \mathbf{u}) + \mathbf{b} \cdot (A^{-\frac{1}{2}} \mathbf{u}) + cp = f & \text{in } \Omega, \\ \nabla \times (A^{-\frac{1}{2}} \mathbf{u}) = \mathbf{0} & \text{in } \Omega, \\ p = 0 & \text{on } \Gamma_D, \\ \mathbf{n} \cdot (A^{\frac{1}{2}} \mathbf{u}) = 0 & \text{on } \Gamma_N, \\ \boldsymbol{\tau} \cdot (A^{-\frac{1}{2}} \mathbf{u}) = 0 & \text{on } \Gamma_D. \end{array} \right.$$

Define the first-order system least-squares functional as follows:

$$G(\mathbf{v}, q; f) = \|f + \nabla \cdot (A^{\frac{1}{2}} \mathbf{v}) - \mathbf{b} \cdot (A^{-\frac{1}{2}} \mathbf{v}) - cq\|^2 + \|\mathbf{v} - A^{\frac{1}{2}} \nabla q\|^2 + \|\nabla \times (A^{-\frac{1}{2}} \mathbf{v})\|^2$$

for $(\mathbf{v}, q) \in \mathcal{U} \times H_D^1(\Omega)$. Then the FOSLS variational problem for (1.1) is to minimize the quadratic functional $G(\mathbf{v}, q; f)$ over $\mathcal{U} \times H_D^1(\Omega)$: find $(\mathbf{u}, p) \in \mathcal{U} \times H_D^1(\Omega)$ such that

$$(2.3) \quad G(\mathbf{u}, p; f) = \inf_{(\mathbf{v}, q) \in \mathcal{U} \times H_D^1(\Omega)} G(\mathbf{v}, q; f).$$

3. Discrete FOSLS

In [5], we proposed and analyzed the discrete first-order system least squares method for the second-order partial differential equations. We recall the formulation.

Let \mathcal{T}_h be a partition of the domain Ω into finite elements; i.e.,

$$\Omega = \cup_{K \in \mathcal{T}_h} K$$

with $h = \max\{\text{diam}(K) : K \in \mathcal{T}_h\}$. Assume that the triangulation \mathcal{T}_h is quasi-uniform; i.e., it is regular and satisfies the inverse assumption. Let \mathcal{P}_{m-1}^h be a finite-dimensional space consisting of continuous piecewise

polynomials of degree $m - 1$ with respect to the triangulation \mathcal{T}_h . Denote standard finite element spaces by

$$\mathcal{S}_D^h = H_D^1(\Omega) \cap \mathcal{P}_{m-1}^h \quad \text{and} \quad \mathcal{S}_N^h = H_N^1(\Omega) \cap \mathcal{P}_{m-1}^h$$

and define the approximation space for the vector variable as

$$\mathcal{U}^h = (A^{\frac{1}{2}} \nabla \mathcal{S}_D^h) \oplus (A^{-\frac{1}{2}} \nabla^\perp \mathcal{S}_N^h)$$

using the following Helmholtz decomposition, for any $\mathbf{u} \in \mathcal{U}$,

$$(3.1) \quad \mathbf{u} = A^{\frac{1}{2}} \nabla s + A^{-\frac{1}{2}} \nabla^\perp t,$$

where $s \in H_D^1(\Omega)$ and $t \in H_N^1(\Omega)$.

Define the discrete divergence operator and curl operator as follow.

$$\nabla_h \cdot : L^2(\Omega)^2 \rightarrow \mathcal{S}_D^h, \quad \text{for } \mathbf{v} \in L^2(\Omega)^2 \quad \text{by } \phi = \nabla_h \cdot \mathbf{v} \in \mathcal{S}_D^h$$

satisfying

$$(\phi, q) = -(\mathbf{v}, \nabla q), \quad \forall q \in \mathcal{S}_D^h.$$

$$\nabla_h \times : L^2(\Omega)^2 \rightarrow \mathcal{S}_N^h, \quad \text{for } \mathbf{v} \in L^2(\Omega)^2 \quad \text{by } \psi = \nabla_h \times \mathbf{v} \in \mathcal{S}_N^h$$

satisfying

$$(\psi, q) = (\mathbf{v}, \nabla^\perp q), \quad \forall q \in \mathcal{S}_N^h.$$

Denote by Q_h the L^2 -projection operator onto \mathcal{S}_D^h .

Now, we define the discrete FOSLS functional: for $(\mathbf{v}, q) \in \mathcal{U}^h \times \mathcal{S}_D^h$

$$(3.2) \quad G_h(\mathbf{v}, q; f) = \|f + \nabla_h \cdot (A^{\frac{1}{2}} \mathbf{v}) - Q_h(\mathbf{b} \cdot (A^{-\frac{1}{2}} \mathbf{v})) - cq\|^2 + \|\mathbf{v} - A^{\frac{1}{2}} \nabla q\|^2 + \|\nabla_h \times (A^{-\frac{1}{2}} \mathbf{v})\|^2.$$

Define a norm over $\mathcal{U}^h \times \mathcal{S}_D^h$ as

$$|||(\mathbf{v}, q)||| = \left(\|q\|_1^2 + \|\mathbf{v}\|^2 + \|\nabla_h \cdot (A^{\frac{1}{2}} \mathbf{v})\|^2 + \|\nabla_h \times (A^{-\frac{1}{2}} \mathbf{v})\|^2 \right)^{\frac{1}{2}}.$$

THEOREM 3.1. *The homogeneous functional $G_h(\cdot; 0)$ is uniformly elliptic and continuous in $\mathcal{U}^h \times \mathcal{S}_D^h$; i.e., for any $(\mathbf{v}, q) \in \mathcal{U}^h \times \mathcal{S}_D^h$, there exists a positive constant C such that*

$$(3.3) \quad \frac{1}{C} |||(\mathbf{v}, q)|||^2 \leq G_h(\mathbf{v}, q; 0) \leq C |||(\mathbf{v}, q)|||^2.$$

Then we have the following error estimate.

THEOREM 3.2. *Assume that the solution (\mathbf{u}, p) of the problem (2.2) is in $H^{m-1}(\Omega)^2 \times H^m(\Omega)$ with $m \geq 2$ and let $(\mathbf{u}_h, p_h) \in \mathcal{U}^h \times \mathcal{S}_D^h$ minimize the functional (3.2). Then the following error estimate holds:*

$$(3.4) \quad \|\mathbf{u} - \mathbf{u}_h\| + \|p - p_h\|_1 \leq C h^{m-1} (\|p\|_m + \|\mathbf{u}\|_{m-1}).$$

Instead of working with $\mathbf{v}_h \in \mathcal{U}^h$, we explicitly make use of its representation:

$$(3.5) \quad \mathbf{v} = A^{\frac{1}{2}} \nabla s + A^{-\frac{1}{2}} \nabla^\perp t \quad \text{where } s \in \mathcal{S}_D^h, t \in \mathcal{S}_N^h.$$

We introduce two discrete diffusion operators.

$\Delta_{h,A} : \mathcal{S}_D^h \rightarrow \mathcal{S}_D^h$, for given $s \in \mathcal{S}_D^h$, define $\Delta_{h,A}s \in \mathcal{S}_D^h$ to be the solution of

$$(3.6) \quad (\Delta_{h,A}s, q) = -(A\nabla s, \nabla q), \quad \forall q \in \mathcal{S}_D^h.$$

$\widehat{\Delta}_{h,A} : \mathcal{S}_N^h \rightarrow \mathcal{S}_N^h$, for given $t \in \mathcal{S}_D^h$, define $\widehat{\Delta}_{h,A}t \in \mathcal{S}_N^h$ to be the solution of

$$(3.7) \quad (\widehat{\Delta}_{h,A}t, q) = (A^{-1}\nabla^\perp t, \nabla^\perp q), \quad \forall q \in \mathcal{S}_N^h.$$

It is easy to see that

$$\Delta_{h,A} = \nabla_h \cdot A\nabla \quad \text{and} \quad \widehat{\Delta}_{h,A} = \nabla_h \times A^{-1}\nabla^\perp.$$

We can easily show that

$$(3.8) \quad \|(\mathbf{v}, p)\|^2 = \|(s, t, p)\|^2 = \|p\|_1^2 + \|s\|^2 + \|t\|^2,$$

where

$$\|s\|^2 = \|s\|^2 + \|A^{\frac{1}{2}}\nabla s\|^2 + \|\Delta_{h,A}s\|^2$$

and

$$\|t\|^2 = \|t\|^2 + \|A^{-\frac{1}{2}}\nabla^\perp t\|^2 + \|\widehat{\Delta}_{h,A}t\|^2.$$

Then the discrete least-squares functional can be restate in terms of functions (s, t, q) as

$$(3.9) \quad \begin{aligned} & G_h(s, t, q; f) \\ &= \|f + \Delta_{h,A}s - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1}\nabla^\perp t)) - cq\|^2 \\ & \quad + \|A^{\frac{1}{2}}\nabla s + A^{-\frac{1}{2}}\nabla^\perp t - A^{\frac{1}{2}}\nabla q\|^2 + \|\widehat{\Delta}_{h,A}t\|^2 \end{aligned}$$

and the minimization problem is to find $(\phi_h, \psi_h, p_h) \in \mathcal{S}_D^h \times \mathcal{S}_N^h \times \mathcal{S}_D^h$ such that

$$(3.10) \quad G_h(\phi_h, \psi_h, p_h; f) = \inf_{(s,t,q) \in \mathcal{S}_D^h \times \mathcal{S}_N^h \times \mathcal{S}_D^h} G_h(s, t, q; f)$$

with $\mathbf{u}_h = A^{\frac{1}{2}} \nabla \phi_h + A^{-\frac{1}{2}} \nabla^\perp \psi_h$. The corresponding variational problem is to find $(\phi_h, \psi_h, p_h) \in \mathcal{S}_D^h \times \mathcal{S}_N^h \times \mathcal{S}_D^h$ such that

$$(3.11) \quad b_h(\phi_h, \psi_h, p_h; s, t, q) = f_h(s, t, q), \quad \forall (s, t, q) \in \mathcal{S}_D^h \times \mathcal{S}_N^h \times \mathcal{S}_D^h,$$

where the bilinear and linear forms are given by

$$\begin{aligned} & b_h(\phi_h, \psi_h, p_h; s, t, q) \\ &= (\Delta_{h,A} \phi_h - Q_h(\mathbf{b} \cdot (\nabla \phi_h + A^{-1} \nabla^\perp \psi_h))) - cp, \\ & \quad \Delta_{h,As} - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1} \nabla^\perp t)) - cq) \\ &+ (A^{\frac{1}{2}} \nabla \phi_h + A^{-\frac{1}{2}} \nabla^\perp \psi_h - A^{\frac{1}{2}} \nabla p_h, A^{\frac{1}{2}} \nabla s + A^{-\frac{1}{2}} \nabla^\perp t - A^{\frac{1}{2}} \nabla q) \\ &+ (\widehat{\Delta_{h,A} \psi_h}, \widehat{\Delta_{h,A} t}) \end{aligned}$$

and

$$f_h(s, t, q) = (f, -\Delta_{h,As} + Q_h(\mathbf{b} \cdot (\nabla s + A^{-1} \nabla^\perp t)) + cq).$$

THEOREM 3.3. *For any $(s, t, q) \in \mathcal{S}_D^h \times \mathcal{S}_N^h \times \mathcal{S}_D^h$, there exists a positive constant C such that*

$$(3.12) \quad \frac{1}{C} |||(s, t, q)|||^2 \leq G_h(s, t, q; 0) \leq C |||(s, t, q)|||^2.$$

Theorem 3.3 indicates that the quadratic form $b_h(s, t, q; s, t, q)$ can be preconditioned well by the diagonal quadratic form $|||(s, t, q)|||^2$ because they are spectrally equivalent uniformly in the mesh size. We further replace these diagonal blocks of $|||(s, t, q)|||^2$ by some multigrid preconditioners (see [5] for details).

4. Implementation

From now on, we present three matrices corresponding to each term of the bilinear form $b_h(\cdot; \cdot)$ and the matrix associating with the linear form $f_h(\cdot)$. First, let us denote \mathcal{A}_1 by the matrix corresponding to the first term in the bilinear form $b_h(\cdot; \cdot)$. Let $\{\xi_i\}$ and $\{\eta_i\}$ be the nodal base for \mathcal{S}_D^h and \mathcal{S}_N^h , respectively. Then \mathcal{A}_1 can be easily assembled as the usual cases. Let

$$B_1 = ((A \nabla \xi_j, \nabla \xi_i)) \quad \text{and} \quad B_2 = ((A^{-1} \nabla^\perp \eta_j, \nabla^\perp \eta_i)).$$

Using the fact from the the orthogonality that

$$\begin{aligned} (A^{\frac{1}{2}}\nabla s, A^{-\frac{1}{2}}\nabla^\perp\psi) &= (A^{-\frac{1}{2}}\nabla^\perp t, A^{\frac{1}{2}}\nabla\phi - A^{\frac{1}{2}}\nabla q) \\ &= (A^{\frac{1}{2}}\nabla p, A^{-\frac{1}{2}}\nabla^\perp\psi) \\ &= 0, \end{aligned}$$

we have

$$\mathcal{A}_1 = \begin{pmatrix} B_1 & 0 & -B_1 \\ 0 & B_2 & 0 \\ -B_1 & 0 & B_1 \end{pmatrix}.$$

For a function $p = \sum p_i \xi_i \in \mathcal{S}_D^h$, denote by the coefficient vector $\hat{p} = (p_i)$ consisting of the nodal values and the data vector $\tilde{p} = ((p, \xi_i))$. Then $\tilde{p} = M \hat{p}$ where $M = ((\xi_j, \xi_i))$ denotes the mass matrix. It is easy check that $(p, q) = \hat{q}^T M \hat{p} = \tilde{q} M^{-1} \tilde{p}$.

Let S_1, S_2 and M_c be the matrices defined by $S_1 = ((Q_h(\mathbf{b} \cdot \nabla \xi_j), \xi_i))$, $S_2 = ((Q_h(\mathbf{b} \cdot A^{-1} \nabla^\perp \eta_j), \xi_i))$ and $M_c = ((c \xi_j, \xi_i))$, respectively. By (3.6), B_1 is also given by $B_1 = -((\Delta_{h,A} \xi_j, \xi_i))$. Then, the data vector of $\Delta_{h,As} - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1} \nabla^\perp t)) - cp$ is given by

$$\begin{aligned} &(\Delta_{h,As} - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1} \nabla^\perp t)) - cp, \xi_i) \\ &= -((B_1 + S_1)\hat{s} + S_2\hat{t} + M_c\hat{p})_i. \end{aligned}$$

Also, the data vector of $\Delta_{h,A\phi} - Q_h(\mathbf{b} \cdot (\nabla \phi + A^{-1} \nabla^\perp \psi)) - cq$ is similarly given by

$$\begin{aligned} &(\Delta_{h,A\phi} - Q_h(\mathbf{b} \cdot (\nabla \phi + A^{-1} \nabla^\perp \psi)) - cq, \xi_i) \\ &= -((B_1 + S_1)\hat{\phi} + S_2\hat{\psi} + M_c\hat{q})_i. \end{aligned}$$

Hence, the L^2 inner product of the second term in the bilinear form $b_h(\cdot; \cdot)$ can be represented by

$$\begin{aligned} &(\Delta_{h,As} - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1} \nabla^\perp t)) - cp, \\ &\quad \Delta_{h,A\phi} - Q_h(\mathbf{b} \cdot (\nabla \phi + A^{-1} \nabla^\perp \psi)) - cq) \\ &= ((B_1 + S_1)\hat{\phi} + S_2\hat{\psi} + M_c\hat{q})^T M^{-1} ((B_1 + S_1)\hat{s} + S_2\hat{t} + M_c\hat{p}). \end{aligned}$$

The existence of the inverse of the mass matrix M is not a pleasant thing to compute. However, it is well known that M^{-1} is spectrally equivalent to $h^{-2} I$, i.e.,

$$\frac{1}{C} h^{-2} \chi^T \chi \leq \chi^T M^{-1} \chi \leq C h^{-2} \chi^T \chi, \quad \forall \chi \in \mathfrak{R}^m.$$

Therefore, the last equation can be further switched by the discrete L^2 inner product $(\cdot; \cdot)_h$ such that

$$\begin{aligned} & (\Delta_{h,As} - Q_h(\mathbf{b} \cdot (\nabla s + A^{-1}\nabla^\perp t)) - cp, \\ & \quad \Delta_{h,A}\phi - Q_h(\mathbf{b} \cdot (\nabla\phi + A^{-1}\nabla^\perp\psi)) - cq)_h \\ &= h^{-2} ((B_1 + S_1)\hat{\phi} + S_2\hat{\psi} + M_c\hat{q})^T ((B_1 + S_1)\hat{s} + S_2\hat{t} + M_c\hat{p}). \end{aligned}$$

The matrix corresponding to the right hand side of the last equation is

$$A_2 = h^{-2} \begin{pmatrix} B_1 + S_1^T \\ S_2^T \\ M_c \end{pmatrix} (B_1 + S_1 \quad S_2 \quad M_c).$$

The computation of the third term $(\widehat{\Delta_{h,At}}, \widehat{\Delta_{h,A}\psi})$ in the bilinear form $b_h(\cdot; \cdot)$ similarly follows the case of the second term. Using the fact from (3.7) that B_2 is also given by $B_2 = -((\widehat{\Delta_{h,A}} \eta_j, \eta_i))$ and replacing the L^2 inner product $(\widehat{\Delta_{h,At}}, \widehat{\Delta_{h,A}\psi})$ by the discrete L^2 inner product $(\widehat{\Delta_{h,At}}, \widehat{\Delta_{h,A}\psi})_h$, we have $(\widehat{\Delta_{h,At}}, \widehat{\Delta_{h,A}\psi})_h = h^{-2} \hat{\psi}^T B_2^2 \hat{t}$ and the matrix corresponding to the third term of $b_h(\cdot; \cdot)$ is given by

$$A_3 = h^{-2} B_2^2.$$

Consequently, the matrix form \mathcal{A} corresponding to the bilinear form $b_h(\cdot; \cdot)$ is given by

$$\mathcal{A} = A_1 + A_2 + A_3.$$

Finally, we compute the linear form $f_h(\cdot)$. Let f_p be the L^2 projection of f into \mathcal{S}_D^h , i.e., $(f, \xi) = (f_p, \xi)$ for any $\xi \in \mathcal{S}_D^h$. Then, the data vector of f_p is given by $(f_p, \xi_i) = (M\hat{f}_p)_i$. Using the fact that f_p is the L^2 projection of f into \mathcal{S}_D^h , we have

$$\begin{aligned} & (f, \Delta_{h,A}\phi - Q_h(\mathbf{b} \cdot (\nabla\phi + A^{-1}\nabla^\perp\psi)) - cq) \\ &= ((B_1 + S_1)\hat{\phi} + S_2\hat{\psi} + M_c\hat{q})^T \hat{f}_p. \end{aligned}$$

The matrix form corresponding to the linear form $f_h(\cdot)$ is

$$F = \begin{pmatrix} B_1 + S_1^T \\ S_2^T \\ M_c \end{pmatrix} \hat{f}_p.$$

Now, we are led to the matrix problem associating with (3.10):

$$(4.13) \quad \mathcal{A} X = F,$$

where $X = (\hat{s}, \hat{t}, \hat{p})^T$.

5. Numerical experiments

In this section, we present the numerical experiments for the following elliptic partial differential equation:

$$(5.1) \quad \begin{cases} -\nabla \cdot (A\nabla p) + \mathbf{b} \cdot \nabla p + cp = f, & \text{in } \Omega, \\ p = 0, & \text{on } \partial\Omega, \end{cases}$$

where Ω is the unit square and $A = aI$, where I is the 2×2 identity matrix and a is the function defined for a given constant σ on the unit square by

$$a(x, y) = \begin{cases} 1, & x \leq \frac{1}{2}, \\ \sigma, & x > \frac{1}{2}. \end{cases}$$

The finite element approximation in this paper is performed as follows. The domain Ω is first partitioned into $2^j \times 2^j$ squares of size $h_j \times h_j$, with $h_j = 2^{-j}$. Then, each small square is divided into pairs of triangles by connecting the bottom right and upper left corners. We use the continuous piecewise linear finite element space for the approximation of all unknowns s , t , and p to solve the problem (3.11). The iteration method we used is the preconditioning conjugate gradient method with diagonal preconditioner

$$\mathcal{B} = \text{diag}[h^2 P_1^2, h^2 P_2^2, P_1],$$

where P_1 and P_2 are the standard multigrid $V(\eta_1, \eta_2)$ -cycle preconditioners of the operators $I - \Delta_{h,A}$ and $I - \widehat{\Delta}_{h,A}$, respectively. The coarsest grid size for multigrid V -cycle is $h_1 = 2^{-1}$.

We first study the performances of the preconditioner \mathcal{B} . To show the effectiveness of the preconditioner \mathcal{B} , we report the condition numbers of the preconditioned linear system along with various coefficients \mathbf{b} and c . For the convenience of the readers, we include a discussion about the relation between the iteration numbers and condition numbers which can be found in [2, 9].

Let X be the solution of $\mathcal{A}X = F$, X_m be the m -th iterates and $R_m = F - \mathcal{A}X_m$ be the residual. Then there is a constant C_0, C_1 , independent of h , such that

$$\begin{aligned} C_0(\mathcal{A}(X - X_m), X - X_m) &\leq (\mathcal{B}R_m, R_m) \\ &\leq C_1(\mathcal{A}(X - X_m), X - X_m), \end{aligned}$$

and so

$$\begin{aligned} \frac{C_0 (\mathcal{B} R_m, R_m)}{C_1 (\mathcal{B} R_0, R_0)} &\leq \frac{(\mathcal{A}(X - X_m), X - X_m)}{(\mathcal{A}(X - X_0), X - X_0)} \\ &\leq \frac{C_1 (\mathcal{B} R_m, R_m)}{C_0 (\mathcal{B} R_0, R_0)}. \end{aligned}$$

Therefore

$$\frac{(\mathcal{B} R_m, R_m)}{(\mathcal{B} R_0, R_0)} \leq \varepsilon$$

can be used to stop the iteration. In the preconditioning conjugate gradient method, $(\mathcal{B} R_m, R_m)$ is computed as part of the iteration, so the error estimator is free of cost. To reveal the real error reduction rate and condition number of the preconditioned system, we choose $\varepsilon = 10^{-8}$. The condition number of $\mathcal{B}\mathcal{A}$ can be estimated by

$$\kappa(\mathcal{B}\mathcal{A}) \leq \left(\frac{1+\theta}{1-\theta}\right)^2, \quad \text{with } \theta = \left(\frac{\varepsilon}{4}\right)^{\frac{1}{2m}},$$

where m is the iteration number.

We present iteration numbers and condition numbers of $\mathcal{B}\mathcal{A}$ for the problem (5.1) with the three values $\sigma = 1, 10, 100$. Tables 1 and 2 report the iteration numbers and condition numbers of $\mathcal{B}\mathcal{A}$ under the preconditioning conjugate gradient iterations when we use one sweep multigrid V(2,2)-cycle in Table 1 and one sweep multigrid V(3,3)-cycle in Table 2 for the preconditioners P_1 and P_2 . As expected, the condition numbers depend on the size of σ , convection \mathbf{b} and reaction c , but the degradation is fairly graceful. Comparing Table 1 with Table 2, we can observe that the use of multigrid V(2,2)-cycle preconditioners is more effective than the use of multigrid V(3,3)-cycle preconditioners except the case of $\sigma = 100$ and $\mathbf{b} = (6, 9)^t$, in which multigrid V(3,3)-cycle preconditioners are good.

We also present the discretization errors and their convergence rates. Let p be the exact solution to the problem (5.1). Then $\mathbf{u} = A^{\frac{1}{2}} \nabla p$ is the exact solution to the first-order problem (2.2). Let p_h, s_h and t_h be the approximation solutions to the problem (3.11). Then, the approximation solution \mathbf{u}_h is defined by

$$\mathbf{u}_h = A^{\frac{1}{2}} \nabla s_h + A^{-\frac{1}{2}} \nabla^\perp t_h.$$

Denote by

$$e_{u,h} = \|\mathbf{u} - \mathbf{u}_h\|, \quad \text{where } \mathbf{u} = A^{\frac{1}{2}} \nabla p,$$

and

$$e_{p,h} = \|p - p_h\| \quad \text{and} \quad e_{p,h}^1 = \|p - p_h\|_1.$$

Here, the norms of all errors were calculated approximately using the seven points quadrature rule in the triangles of triangulation \mathcal{T}_h . The convergence rates for discretization errors are measured by

$$\log_2 \frac{e_{u,h}}{e_{u,\frac{h}{2}}}, \quad \log_2 \frac{e_{p,h}}{e_{p,\frac{h}{2}}}, \quad \text{and} \quad \log_2 \frac{e_{p,h}^1}{e_{p,\frac{h}{2}}^1}.$$

We tested the problem (5.1) with several kinds of coefficients σ , \mathbf{b} and c ;

$$\mathbf{b}^t = (0, 0), (2, 3), (4, 6), (6, 9) \quad \text{and} \quad c = 0, -1, -10.$$

First, we chose A to be the identity matrix, i.e., $\sigma = 1$, and the smooth exact solution to be

$$p = x(x-1) \sin(\pi y).$$

The exact values of vector solution \mathbf{u} and right hand side f are defined consistently. The theoretically predicted discretization errors of p in H^1 and of \mathbf{u} in L^2 are $O(h)$, but the resulting errors in Table 3 appear to be $O(h^2)$. It is probably due to smooth exact solution and coefficients.

Finally, we present results for our method applied to a discontinuous coefficient problem. We again treat the problem (5.1) with the three values $\sigma = 1, 10, 100$, and constructed the exact solution p so that \mathbf{u} is not a product H^1 function:

$$p(x, y) = \begin{cases} ((2\sigma - 4)x^2 + (4 - \sigma)x) \sin(\pi y), & x \leq \frac{1}{2}, \\ (-6x^2 + 7x - 1) \sin(\pi y), & x > \frac{1}{2}. \end{cases}$$

Tables 4 shows the results for the case $\sigma = 100$. Also, the resulting errors of p in H^1 and \mathbf{u} in L^2 are apparently $O(h^2)$. We therefore appear to have obtained optimal convergence in the L^2 norm and super-convergence in the discrete H^1 norm. Our experimental results for the case $\sigma = 1$ and 10 also had the same fashions as the case $\sigma = 100$.

c	b ^t	σ = 1				σ = 10				σ = 100			
		h = 1/16	1/32	1/64	1/128	1/16	1/32	1/64	1/128	1/16	1/32	1/64	1/128
0	(0,0)	4	4	4	4	14	12	9	9	22	21	20	16
		1.6	1.6	1.6	1.6	10.9	8.2	4.9	4.9	25.9	23.7	21.6	14.0
	(2,3)	7	7	7	7	20	21	20	18	50	66	64	56
		3.2	3.2	3.2	3.2	21.6	23.7	21.6	17.6	131.2	228.1	214.6	164.4
(4,6)	16	16	16	16	35	37	33	30	95	177	170	143	
	14.0	14.0	14.0	14.0	64.6	72.2	57.5	47.7	471.9	1636.6	1509.8	1068.5	
(6,9)	31	31	36	34	46	57	54	47	215	300	300	300	
	50.9	50.9	68.3	61.0	111.2	170.3	152.9	116.0	2414.5	4700.3	4700.3	4700.3	
-1	(0,0)	4	4	4	4	14	13	10	9	23	22	21	18
		1.6	1.6	1.6	1.6	10.9	9.5	5.9	4.9	28.3	25.9	23.7	17.6
-10	(0,0)	5	6	6	7	52	19	20	17	73	86	66	44
		2.0	2.6	2.6	3.2	141.9	19.5	21.6	15.8	278.9	386.9	228.1	101.8

TABLE 1. Iteration numbers and condition numbers by V(2,2).

c	b ^t	σ = 1				σ = 10				σ = 100			
		h = 1/16	1/32	1/64	1/128	1/16	1/32	1/64	1/128	1/16	1/32	1/64	1/128
0	(0,0)	4	4	3	3	10	9	9	8	16	14	14	13
		1.6	1.6	1.2	1.2	5.9	4.9	4.9	4.0	14.0	10.9	10.9	9.5
	(2,3)	6	6	6	7	16	19	17	16	32	41	39	34
		2.6	2.6	2.6	3.2	14.0	19.5	15.8	14.0	54.1	88.4	80.1	61.0
(4,6)	16	15	15	15	24	30	28	25	52	63	65	58	
	14.0	12.4	12.4	12.4	30.7	47.7	41.6	33.3	141.9	207.9	221.3	176.3	
(6,9)	23	29	32	30	39	43	41	37	94	121	118	99	
	28.3	44.6	54.1	47.7	80.1	97.2	88.4	72.2	462.1	765.2	727.8	512.5	
-1	(0,0)	4	3	4	3	10	11	9	8	19	16	14	14
		1.6	1.2	1.6	1.2	5.9	7.0	4.9	4.0	19.5	14.0	10.9	10.9
-10	(0,0)	5	6	6	6	13	19	19	14	36	43	35	27
		2.0	2.6	2.6	2.6	9.5	19.5	19.5	10.9	68.3	97.2	64.6	38.7

TABLE 2. Iteration numbers and condition numbers by V(3,3).

c	\mathbf{b}	h	$e_{u,h}$		$e_{p,h}$		$e_{p,h}^1$	
0	(0,0)	$\frac{1}{4}$	6.607e-2		1.369e-2		6.747e-2	
		$\frac{1}{8}$	1.818e-2	1.86	3.959e-3	1.79	1.861e-2	1.86
		$\frac{1}{16}$	4.657e-3	1.96	1.027e-3	1.95	4.769e-3	1.96
		$\frac{1}{32}$	1.171e-3	1.99	2.593e-4	1.99	1.200e-3	1.99
		$\frac{1}{64}$	2.933e-4	2.00	6.497e-5	2.00	3.004e-4	2.00
		$\frac{1}{128}$	7.335e-5	2.00	1.625e-5	2.00	7.513e-5	2.00
0	(2,3)	$\frac{1}{4}$	6.433e-2		1.303e-2		6.563e-2	
		$\frac{1}{8}$	1.757e-2	1.87	3.719e-3	1.81	1.796e-2	1.87
		$\frac{1}{16}$	4.491e-3	1.97	9.618e-4	1.95	4.593e-3	1.97
		$\frac{1}{32}$	1.129e-3	1.99	2.425e-4	1.99	1.155e-3	1.99
		$\frac{1}{64}$	2.827e-4	2.00	6.076e-5	2.00	2.891e-4	2.00
		$\frac{1}{128}$	7.069e-5	2.00	1.520e-5	2.00	7.230e-5	2.00
0	(4,6)	$\frac{1}{4}$	6.030e-2		1.157e-2		6.140e-2	
		$\frac{1}{8}$	1.625e-2	1.89	3.225e-3	1.84	1.657e-2	1.89
		$\frac{1}{16}$	4.143e-3	1.97	8.305e-4	1.96	4.228e-3	1.97
		$\frac{1}{32}$	1.018e-3	2.02	2.027e-4	2.03	1.031e-3	2.04
		$\frac{1}{64}$	2.578e-4	1.98	5.160e-5	1.97	2.620e-4	1.98
		$\frac{1}{128}$	6.396e-5	2.01	1.290e-5	2.00	6.539e-5	2.00
0	(6,9)	$\frac{1}{4}$	5.586e-2		1.004e-2		5.677e-2	
		$\frac{1}{8}$	1.475e-2	1.92	2.708e-3	1.89	1.493e-2	1.93
		$\frac{1}{16}$	3.627e-3	2.02	6.554e-4	2.05	3.621e-3	2.04
		$\frac{1}{32}$	9.685e-4	1.91	1.880e-4	1.80	1.057e-3	1.78
		$\frac{1}{64}$	2.191e-4	2.14	3.341e-5	2.49	1.924e-4	2.46
		$\frac{1}{128}$	6.294e-5	1.80	9.930e-6	1.75	5.791e-5	1.73
-1	(0,0)	$\frac{1}{4}$	6.345e-2		1.209e-2		5.957e-2	
		$\frac{1}{8}$	1.743e-2	1.86	3.468e-3	1.80	1.629e-2	1.87
		$\frac{1}{16}$	4.461e-3	1.97	8.978e-4	1.95	4.164e-3	1.97
		$\frac{1}{32}$	1.122e-3	1.99	2.264e-4	1.99	1.047e-3	1.99
		$\frac{1}{64}$	2.809e-4	2.00	5.674e-5	2.00	2.621e-4	2.00
		$\frac{1}{128}$	7.024e-5	2.00	1.419e-5	2.00	6.555e-5	2.00
-10	(0,0)	$\frac{1}{4}$	5.113e-1		8.248e-2		7.317e-1	
		$\frac{1}{8}$	1.341e-1	1.93	2.142e-2	1.94	1.683e-1	2.12
		$\frac{1}{16}$	3.403e-2	1.98	3.609e-3	2.57	2.819e-2	2.58
		$\frac{1}{32}$	8.575e-3	1.99	9.218e-4	1.97	7.143e-3	1.98
		$\frac{1}{64}$	2.146e-3	2.00	2.066e-4	2.16	1.609e-3	2.15
		$\frac{1}{128}$	5.367e-4	2.00	5.106e-5	2.02	3.981e-4	2.02

TABLE 3. Discretization errors and convergence rates for the smooth solution.

c	\mathbf{b}	h	$e_{u,h}$		$e_{p,h}$		$e_{p,h}^1$	
0	(0,0)	$\frac{1}{4}$	4.844e+0		5.067e-1		4.334e+0	
		$\frac{1}{8}$	1.455e+0	1.74	1.719e-1	1.56	1.317e+0	1.72
		$\frac{1}{16}$	3.802e-1	1.94	4.624e-2	1.89	3.441e-1	1.94
		$\frac{1}{32}$	9.611e-2	1.98	1.180e-2	1.97	8.713e-2	1.98
		$\frac{1}{64}$	2.410e-2	2.00	2.962e-3	1.99	2.184e-2	2.00
		$\frac{1}{128}$	6.028e-3	2.00	7.405e-4	2.00	5.458e-3	2.00
0	(2,3)	$\frac{1}{4}$	9.298e+0		5.641e-1		4.790e+0	
		$\frac{1}{8}$	1.433e+0	2.70	1.692e-1	1.74	1.301e+0	1.88
		$\frac{1}{16}$	3.745e-1	1.94	4.566e-2	1.89	3.413e-1	1.93
		$\frac{1}{32}$	9.472e-2	1.98	1.162e-2	1.97	8.627e-2	1.98
		$\frac{1}{64}$	2.261e-2	2.07	2.770e-3	2.07	2.060e-2	2.07
		$\frac{1}{128}$	5.940e-3	1.93	7.304e-4	1.92	5.410e-3	1.93
0	(4,6)	$\frac{1}{4}$	1.042e+1		5.684e-1		4.884e+0	
		$\frac{1}{8}$	2.288e+0	2.19	1.779e-1	1.68	1.411e+0	1.79
		$\frac{1}{16}$	3.642e-1	2.65	4.326e-2	2.04	3.296e-1	2.10
		$\frac{1}{32}$	9.178e-2	1.99	1.100e-2	1.97	8.332e-2	1.98
		$\frac{1}{64}$	2.251e-2	2.03	2.677e-3	2.04	2.032e-2	2.04
		$\frac{1}{128}$	5.762e-3	1.97	6.931e-4	1.95	5.239e-3	1.96
0	(6,9)	$\frac{1}{4}$	1.382e+1		8.236e-1		7.046e+0	
		$\frac{1}{8}$	3.550e+0	1.96	1.849e-1	2.15	1.622e+0	2.12
		$\frac{1}{16}$	3.496e-1	3.34	4.008e-2	2.21	3.142e-1	2.37
		$\frac{1}{32}$	9.930e-2	1.82	9.212e-3	2.12	7.216e-2	2.12
		$\frac{1}{64}$	2.187e-2	2.18	2.508e-3	1.88	1.953e-2	1.89
		$\frac{1}{128}$	6.222e-3	1.81	7.191e-4	1.80	5.531e-3	1.82
-1	(0,0)	$\frac{1}{4}$	4.693e+0		4.592e-1		3.940e+0	
		$\frac{1}{8}$	1.405e+0	1.74	1.538e-1	1.58	1.182e+0	1.74
		$\frac{1}{16}$	3.668e-1	1.94	4.146e-2	1.89	3.093e-1	1.93
		$\frac{1}{32}$	9.272e-2	1.98	1.055e-2	1.97	7.814e-2	1.98
		$\frac{1}{64}$	2.324e-2	2.00	2.650e-3	1.99	1.959e-2	2.00
		$\frac{1}{128}$	5.815e-3	2.00	6.633e-4	2.00	4.900e-3	2.00
-10	(0,0)	$\frac{1}{4}$	1.908e+1		8.149e+0		6.714e+1	
		$\frac{1}{8}$	4.322e+0	2.14	2.833e+0	1.52	2.099e+1	1.68
		$\frac{1}{16}$	6.616e-1	2.71	4.850e-1	2.55	3.480e+0	2.59
		$\frac{1}{32}$	1.500e-1	2.14	1.129e-1	2.10	8.036e-1	2.11
		$\frac{1}{64}$	3.661e-2	2.03	2.775e-2	2.02	1.971e-1	2.03
		$\frac{1}{128}$	9.088e-3	2.01	6.901e-3	2.01	4.900e-2	2.01

TABLE 4. Discretization errors and convergence rates for $\sigma = 100$.

References

- [1] J. H. Bramble, R. D. Lazarov, and J. E. Pasciak, *A least-squares approach based on a discrete minus one inner product for first order system*, Math. Comp. **66** (1997), 935–955.
- [2] J. H. Bramble and T. Sun, *A negative-norm least squares method for Reissner-Mindlin plates*, Math. Comp. **67** (1998), 901–916.
- [3] Z. Cai, R. D. Lazarov, T. Manteuffel, and S. McCormick, *First-order system least squares for second-order partial differential equations: Part I*, SIAM J. Numer. Anal. **31** (1994), 1785–1799.
- [4] Z. Cai, T. Manteuffel, and S. McCormick, *First-order system least squares for second-order partial differential equations: Part II*, SIAM J. Numer. Anal. **34** (1997), 425–454.
- [5] Z. Cai and B. C. Shin, *The discrete first-order system least-squares: the second-order elliptic boundary value problem*, SIAM J. Numer. Anal. **40** (2002), 307–318.
- [6] C. L. Chang, *Finite element approximation for grad-div type systems in the plane*, SIAM J. Numer. Anal. **29** (1992), 452–461.
- [7] S. D. Kim and B. C. Shin, *H^{-1} least-squares methods for the velocity-pressure-stress formulation of Stokes equations*, Appl. Numer. Math. **40** (2002), no. 4, 451–465.
- [8] A. I. Pehlivanov, G. F. Carey, and R. D. Lazarov, *Least squares mixed finite elements for second order elliptic problems*, SIAM J. Numer. Anal. **31** (1994), 1368–1377.
- [9] A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations*, Springer-Verlag, Berlin Heidelberg, 1994.

Department of Mathematics
Chonnam National University
Kwangju 500-757, Korea
E-mail: bcshin@jnu.ac.kr