SIAM J. Numer. Anal.
Vol. 53, No. 1, pp. 405–420

DIV FIRST-ORDER SYSTEM LL* (FOSLL*) FOR SECOND-ORDER ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS*

ZHIQIANG CAI†, ROB FALGOUT‡, and SHUN ZHANG§

Abstract. The first-order system LL* (FOSLL*) approach for general second-order elliptic partial differential equations was proposed and analyzed in [Z. Cai et al., *SIAM J. Numer. Anal.*, 39 (2001), pp. 1418–1445], in order to retain the full efficiency of the L^2 norm first-order system least-squares (FOSLS) approach while exhibiting the generality of the inverse-norm FOSLS approach. The FOSLL* approach of Cai et al. was applied to the div-curl system with added slack variables, and hence it is quite complicated. In this paper, we apply the FOSLL* approach to the div system and establish its well-posedness. For the corresponding finite element approximation, we obtain a quasi-optimal a priori error bound under the same regularity assumption as the standard Galerkin method, but without the restriction to sufficiently small mesh size. Unlike the FOSLS approach, the FOSLL* approach does not have a free a posteriori error estimator. We then propose an explicit residual error estimator and establish its reliability and efficiency bounds.

Key words. LL* method, least-squares method, a priori error estimate, a posteriori error estimate, elliptic equations

AMS subject classifications. 65M60, 65M15

DOI. 10.1137/140971890

1. Introduction. There are substantial interests in the use of least-squares principles for the approximate solution of partial differential equations with applications in both solid and fluid mechanics. Many least-squares methods for the scalar elliptic partial differential equations have been proposed and analyzed [3, 15]. Their numerical properties depend on choices such as the first-order system and the least-squares norm. Loosely speaking, there are three types of least-squares methods: the inverse approach, the div approach, and the div-curl approach. The inverse approach employs an inverse norm that is further replaced by either the weighted mesh-dependent norm (see [2]) or the discrete H^{-1} norm (see [7]) for computational feasibility. Both the div and div-curl approaches use the L^2 norm and the corresponding homogeneous least-squares functionals are equivalent to the $H(\text{div}) \times H^1$ and $(H(\text{div}) \cap H(\text{curl})) \times H^1$ norms, respectively. The div approach based on the flux-pressure formulation has been studied by many researchers (see, e.g., [4, 8, 14, 16]). The div-curl approach [9] has also been well-studied.

In order to retain the full efficiency of the L^2 norm first-order system least-squares (FOSLS) approach while exhibiting the generality of the inverse-norm FOSLS approach, the first-order system LL* (FOSLL*) approach for general second-order

*Received by the editors June 9, 2014; accepted for publication (in revised form) October 1, 2014; published electronically February 10, 2015. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 (LLNL-JRNL-645325). The research was supported in part by the National Science Foundation under grant DMS-1217081 and the Research Grants Council of the Hong Kong SAR, China, under GRF grant project 11303914, CityU 9042090.

http://www.siam.org/journals/sinum/53-1/97189.html

†Department of Mathematics, Purdue University, West Lafayette, IN 47907-2067 (caiz@purdue.edu).

‡Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551-0808 (falgout2@llnl.gov).

§Department of Mathematics, City University of Hong Kong, Hong Kong (shun.zhang@cityu.edu.hk).

elliptic partial differential equations was proposed and analyzed in [10]. The FOSLL* approach was applied to the div-curl system, whose adjoint system is an underdetermined system and hence is not suitable for FOSLL*. This difficulty was overcome by carefully adding slack variables to the div-curl system. But the resulting approach is quite complicated.

Our purpose here is to study the FOSLL* approach applying to the div system. Without adding any slack variables to the div system, the resulting approach is much simpler than that in [10]. By showing that the bilinear form of the FOSLL* approach is coercive and bounded and that the linear form is bounded with respect to a weighted $H(\mathrm{div}) \times H^1$ norm, we establish the well-posedness of the FOSLL* approach. Under the same regularity assumption as the standard Galerkin method, but without the restriction to sufficiently small mesh size, we obtain a quasi-optimal a priori error bound for the corresponding finite element approximation. Note that this assumption is weaker than that for the div FOSLS [11]. Unlike the FOSLS approach, the FOSLL* approach does not have a free a posteriori error estimator; thus we study an explicit residual error estimator and establish its reliability and efficiency bounds.

The paper is organized as follows. In section 2 we introduce mathematical equations for the second-order scalar elliptic partial differential equations and its div first-order system, and we then derive the FOSLL* variational formulation and establish its well-posedness. In section 3, the FOSLL* finite element approximation is described. A priori and a posteriori error estimations are obtained in sections 4 and 5, respectively. In section 6, we present numerical results.

1.1. Notation. We use the standard notation and definitions for the Sobolev spaces $H^s(\Omega)^d$ and $H^s(\partial\Omega)^d$ for $s \geq 0$. The standard associated inner products are denoted by $(\cdot,\, \cdot)_{s,\Omega}$ and $(\cdot,\, \cdot)_{s,\partial\Omega}$, and their respective norms are denoted by $\|\cdot\|_{s,\Omega}$ and $\|\cdot\|_{s,\partial\Omega}$. (We suppress the superscript d because the dependence on dimension will be clear by context. We also omit the subscript Ω from the inner product and norm designation when there is no risk of confusion.) For $s = 0$, $H^s(\Omega)^d$ coincides with $L^2(\Omega)^d$. In this case, the inner product and norm will be denoted by $\|\cdot\|$ and $(\cdot,\, \cdot)$, respectively. Set

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

When $\Gamma_D = \partial\Omega$, denote $H^1_D(\Omega)$ by $H^1_0(\Omega)$. Finally, set

$$H(\mathrm{div};\Omega) = \{\mathbf{v} \in L^2(\Omega)^d \,:\, \nabla\cdot\mathbf{v} \in L^2(\Omega)\},$$

which is a Hilbert space under the norm

$$\|\mathbf{v}\|_{H(\mathrm{div};\,\Omega)} = \left(\|\mathbf{v}\|^2 + \|\nabla\cdot\mathbf{v}\|^2\right)^{1/2},$$

and define the subspace

$$H_N(\mathrm{div};\Omega) = \{\mathbf{v} \in H(\mathrm{div};\Omega) \,:\, \mathbf{n}\cdot\mathbf{v} = 0 \text{ on } \Gamma_N\}.$$

2. FOSLL* formulation. Let Ω be a bounded, open, connected subset of $\Re^d$ ($d = 2$ or 3) with a Lipschitz continuous boundary $\partial\Omega$. Denote by $\mathbf{n} = (n_1, \ldots, n_d)$ the outward unit vector normal to the boundary. We partition the boundary of the domain Ω into two open subsets Γ_D and Γ_N such that $\partial\Omega = \bar\Gamma_D \cup \bar\Gamma_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$. For simplicity, we assume that Γ_D is not empty (i.e., meas$\,(\Gamma_D) \neq 0$) and is connected.

2.1. Second-order elliptic problem. Consider the following second-order elliptic boundary value problem:

$$-\nabla \cdot (A\nabla u) + \mathbf{b} \cdot \nabla u + a\, u = f \quad \text{in } \Omega \tag{2.1}$$

with boundary conditions

$$u = g_{D} \quad \text{on } \Gamma_D \quad \text{and} \quad - A\nabla u \cdot \mathbf{n} = g_{N} \quad \text{on } \Gamma_N, \tag{2.2}$$

where the symbols $\nabla\cdot$ and ∇ stand for the divergence and gradient operators, respectively; $A \in L^\infty(\Omega)^{d\times d}$ is a given $d \times d$ tensor-valued function; $\mathbf{b} \in L^\infty(\Omega)^d$ and $a \in L^\infty(\Omega)$ are given vector- and scalar-valued functions, respectively; and f is a given scalar function. Assume that A is uniformly symmetric positive definite: there exist positive constants $0 < \Lambda_0 \le \Lambda_1$ such that

$$\Lambda_0 \boldsymbol{\xi}^T \boldsymbol{\xi} \le \boldsymbol{\xi}^T A \boldsymbol{\xi} \le \Lambda_1 \boldsymbol{\xi}^T \boldsymbol{\xi}$$

for all $\boldsymbol{\xi} \in \Re^d$ and almost all $x \in \overline{\Omega}$. The corresponding variational form of system (2.1)–(2.2) is to find $u \in H^1(\Omega)$ such that $u|_{\Gamma_D} = g_{D}$ and that

$$a(u, v) = (f, v) - \int_{\Gamma_N} g_{N} v\, ds \quad \forall\, v \in H^1_D(\Omega), \tag{2.3}$$

where the bilinear form is defined by

$$a(u, v) = (A\nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u + a\, u, v).$$

The dual problem of (2.3) is to find $z \in H^1(\Omega)$ such that $z|_{\Gamma_D} = g_{D}$ and that

$$\hat{a}(\phi, z) = (f, \phi) - \int_{\Gamma_N} g_{N} \phi\, ds \quad \forall\, \phi \in H^1_D(\Omega), \tag{2.4}$$

where the bilinear form is defined by

$$\hat{a}(\phi, z) = (\nabla \phi, (A\, \nabla + \mathbf{b})\, z) + (\phi, a\, z).$$

Assume that both problems (2.3) and (2.4) have unique solutions.

2.2. First-order system. Introducing the flux (vector) variable

$$\boldsymbol{\sigma} = -A\, \nabla u,$$

the scalar elliptic problem in (2.1)–(2.2) may be rewritten as the following first-order partial differential system:

$$\begin{cases} A^{-1}\boldsymbol{\sigma} + \nabla u & = \mathbf{0} \quad \text{in } \Omega, \\ \nabla \cdot \boldsymbol{\sigma} - \mathbf{b} \cdot A^{-1}\boldsymbol{\sigma} + a\, u & = f \quad \text{in } \Omega \end{cases} \tag{2.5}$$

with boundary conditions

$$u = g_{D} \quad \text{on } \Gamma_D \quad \text{and} \quad \mathbf{n} \cdot \boldsymbol{\sigma} = g_{N} \quad \text{on } \Gamma_N. \tag{2.6}$$

Let

$$\mathcal{L} = \begin{pmatrix} A^{-1} & \nabla \\ \nabla\cdot - \mathbf{b} \cdot A^{-1} & a \end{pmatrix}, \quad \mathcal{U} = \begin{pmatrix} \boldsymbol{\sigma} \\ u \end{pmatrix}, \quad \text{and} \quad \mathcal{F} = \begin{pmatrix} \mathbf{0} \\ f \end{pmatrix};$$

then (2.5) may be rewritten as

$$\mathcal{L}\mathcal{U} = \mathcal{F}. \tag{2.7}$$

2.3. Div FOSLL* variational formulation. Multiplying test function $\mathcal{V} = (\boldsymbol{\tau},\, v)^t \in H_N(\text{div};\Omega) \times H_D^1(\Omega)$, integrating over the domain Ω, and using integration by parts, we have

$$\begin{aligned}(f,\, v) &= \langle \mathcal{F},\, \mathcal{V}\rangle = \langle \mathcal{L}\mathcal{U},\, \mathcal{V}\rangle\\ &= (A^{-1}\boldsymbol{\sigma} + \nabla u,\, \boldsymbol{\tau}) + (\nabla\cdot\boldsymbol{\sigma} - \mathbf{b}\cdot A^{-1}\boldsymbol{\sigma} + a\,u,\, v)\\ &= (\boldsymbol{\sigma},\, A^{-1}\boldsymbol{\tau}) - (\boldsymbol{\sigma},\, \nabla v) + \int_{\partial\Omega} (\boldsymbol{\sigma}\cdot\mathbf{n})\, v\, ds - (\boldsymbol{\sigma},\, A^{-1}\mathbf{b}\, v)\\ &\quad -(u,\, \nabla\cdot\,\boldsymbol{\tau}) + \int_{\partial\Omega} (\boldsymbol{\tau}\cdot\mathbf{n})\, u\, ds + (u,\, a\, v)\\ &= \big(\boldsymbol{\sigma},\, A^{-1}\boldsymbol{\tau} - (\nabla + A^{-1}\mathbf{b})\, v\big) + \big(u,\, a\, v - \nabla\cdot\boldsymbol{\tau}\big)\\ &\quad + \int_{\Gamma_N} g_{_N} v\, ds + \int_{\Gamma_D} g_{_D}(\boldsymbol{\tau}\cdot\mathbf{n})\, ds\\ &= \langle \mathcal{U},\, \mathcal{L}^*\,\mathcal{V}\rangle + g(\boldsymbol{\tau},\, v),\end{aligned}$$

where the formal adjoint of $\mathcal{L}$ and the boundary functional are defined by

$$\mathcal{L}^* = \begin{pmatrix} A^{-1} & -(\nabla + A^{-1}\mathbf{b}) \\ -\nabla\cdot & a \end{pmatrix} \quad \text{and} \quad g(\boldsymbol{\tau},\, v) = \int_{\Gamma_N} g_{_N} v\, ds + \int_{\Gamma_D} g_{_D}(\boldsymbol{\tau}\cdot\mathbf{n})\, ds,$$

respectively.

Let

$$\mathcal{A} = \begin{pmatrix} A & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \quad \text{where} \quad \alpha = \begin{cases} |a|, & a \neq 0, \\ 1, & a = 0. \end{cases}$$

Let $\mathcal{W} = (\boldsymbol{\eta},\, w)^t$ satisfy

$$\mathcal{U} = \mathcal{A}\,\mathcal{L}^*\,\mathcal{W} = \begin{pmatrix} \boldsymbol{\eta} - (A\nabla + \mathbf{b})\, w \\ -\alpha^{-1}\,(\nabla\cdot\boldsymbol{\eta} - aw) \end{pmatrix}; \tag{2.8}$$

then we have

$$\langle \mathcal{A}\,\mathcal{L}^*\,\mathcal{W},\, \mathcal{L}^*\,\mathcal{V}\rangle = (f,\, v) - g(\boldsymbol{\tau},\, v) \equiv l(\boldsymbol{\tau},\, v).$$

Now, our div FOSLL* variational formulation is to find $(\boldsymbol{\eta},\, w) \in H_N(\text{div};\Omega)\times H_D^1(\Omega)$ such that

$$b(\boldsymbol{\eta},\, w;\, \boldsymbol{\tau},\, v) = l(\boldsymbol{\tau},\, v) \quad \forall\ (\boldsymbol{\tau},\, v) \in H_N(\text{div};\Omega) \times H_D^1(\Omega), \tag{2.9}$$

where the bilinear form $b(\cdot,\, \cdot)$ is defined by

$$\begin{aligned}&b(\boldsymbol{\eta},\, w;\, \boldsymbol{\tau},\, v) = \langle \mathcal{A}\,\mathcal{L}^*\,\mathcal{W},\, \mathcal{L}^*\,\mathcal{V}\rangle\\ &= \big(\boldsymbol{\eta} - (A\,\nabla + \mathbf{b})\, w,\, A^{-1}\boldsymbol{\tau} - (\nabla + A^{-1}\mathbf{b})\, v\big) + \big(\alpha^{-1}(\nabla\cdot\,\boldsymbol{\eta} - a\, w),\, \nabla\cdot\,\boldsymbol{\tau} - a\, v\big).\end{aligned}$$

Note that both nonhomogeneous Dirichlet and Neumann boundary conditions are imposed weakly.

Remark 2.1. For any $(\boldsymbol{\eta},\, w),\ (\boldsymbol{\tau},\, v) \in H_N(\text{div};\Omega) \times H_D^1(\Omega)$, integration by parts gives

$$(\nabla w, \boldsymbol{\tau}) + (w, \nabla\cdot\boldsymbol{\tau}) = (\nabla v, \boldsymbol{\eta}) + (v, \nabla\cdot\boldsymbol{\eta}) = 0.$$

Hence, when $a > 0$, the bilinear form $b(\cdot,\,\cdot)$ has the form

$$\begin{aligned} b(\boldsymbol{\eta},\, w;\, \boldsymbol{\tau},\, v) = (A^{-1}\boldsymbol{\eta},\, \boldsymbol{\tau}) + (a^{-1}\nabla\cdot\boldsymbol{\eta},\, \nabla\cdot\boldsymbol{\tau}) + (A\nabla w,\, \nabla v) + (a\, w,\, v) \\ - (\mathbf{b}\, w,\, A^{-1}\boldsymbol{\tau}) - (A^{-1}\boldsymbol{\eta},\, \mathbf{b}\, v) + (\mathbf{b}\, w,\, \nabla v) \\ + (\nabla w,\, \mathbf{b}\, v) + (A^{-1}\mathbf{b}\, w,\, \mathbf{b}\, v). \end{aligned}$$

If, in addition, $\mathbf{b} = \mathbf{0}$, i.e., the diffusion-reaction problem with positive reaction coefficient, the div FOSLL* problem in (2.9) is decoupled. More specifically, $w \in H^1_D(\Omega)$ is the solution of

$$(A\nabla w,\, \nabla v) + (a\, w,\, v) = (f,\, v) - \int_{\Gamma_N} g_{_N} v\, ds \quad \forall\, v \in H^1_D(\Omega),$$

and $\boldsymbol{\eta} \in H_N(\mathrm{div};\Omega)$ satisfies

$$(A^{-1}\boldsymbol{\eta},\, \boldsymbol{\tau}) + (a^{-1}\nabla\cdot\boldsymbol{\eta},\, \nabla\cdot\boldsymbol{\tau}) = -\int_{\Gamma_D} g_{_D}(\boldsymbol{\tau}\cdot\mathbf{n})\, ds.$$

Note that the problem for w is similar to the standard variational formulation for the diffusion-reaction problem, but the nonhomogeneous Dirichlet boundary condition is weakly imposed here.

2.4. Well-posedness. Denote by

$$\begin{aligned} |\!|\!| v |\!|\!|_1 = \left(\|\alpha^{1/2}v\|^2 + \|A^{1/2}\nabla v\|^2\right)^{1/2} \text{ and } |\!|\!|\boldsymbol{\tau}|\!|\!|_{H(\mathrm{div})} \\ = \left(\|A^{-1/2}\boldsymbol{\tau}\|^2 + \|\alpha^{-1/2}\nabla\cdot\boldsymbol{\tau}\|^2\right)^{1/2} \end{aligned}$$

the weighted $H^1(\Omega)$ and $H(\mathrm{div};\Omega)$ norms, respectively. Let

$$|\!|\!|(\boldsymbol{\tau},\, v)|\!|\!| = \left(|\!|\!|v|\!|\!|_1^2 + |\!|\!|\boldsymbol{\tau}|\!|\!|^2_{H(\mathrm{div})}\right)^{1/2}.$$

The following theorem establishes the coercivity and continuity of the bilinear form.

THEOREM 2.2. *The bilinear form $b(\cdot,\,\cdot)$ is coercive and continuous in $H_N(\mathrm{div};\Omega) \times H^1_D(\Omega)$, i.e., there exist positive constant α_0 and C, depending on bounds of the coefficients $(A,\, \mathbf{b},$ and $a)$, such that*

$$\alpha_0\, |\!|\!|(\boldsymbol{\tau},\, \mathbf{v})|\!|\!|^2 \le b(\boldsymbol{\tau},\, v;\, \boldsymbol{\tau},\, v) \tag{2.10}$$

and that

$$b(\boldsymbol{\eta},\, w;\, \boldsymbol{\tau},\, v) \le C\, |\!|\!|(\boldsymbol{\eta},\, w)|\!|\!|\, |\!|\!|(\boldsymbol{\tau},\, v)|\!|\!| \tag{2.11}$$

for all $(\boldsymbol{\eta},\, u),\ (\boldsymbol{\tau},\, v) \in H_N(\mathrm{div};\Omega) \times H^1_D(\Omega)$.

Here and hereafter, we use C with or without subscripts in this paper to denote a generic positive constant, possibly different at different occurrences, that is independent of the mesh size h but may depend on the domain Ω. A result similar to that of Theorem 2.2 was proved in [8]. For the convenience of readers, we provide a comprehensive proof here.

Proof. Without loss of generality, we only prove the theorem when $a > 0$. In this case, we have $\alpha = a$.

Equation (2.11) is a direct consequence of the Cauchy–Schwarz and triangle inequalities and the bounds of the coefficients (A, $\mathbf{b}$, and a) of the underlying problem. To show the validity of (2.10), we first establish that

$$\|\!|(\boldsymbol{\tau},\, v)|\!\|^2 \le C\, \left(b(\boldsymbol{\tau},\, v;\, \boldsymbol{\tau},\, v) + \|v\|^2\right) \tag{2.12}$$

for all $(\boldsymbol{\tau},\, v) \in H_N(\mathrm{div};\Omega) \times H^1_D(\Omega)$. To this end, integrating by parts gives

$$(\boldsymbol{\tau},\, \nabla v) = (-a^{-1/2}\nabla\cdot\boldsymbol{\tau} + a^{1/2}v,\, a^{1/2}v) - (a\,v,\, v).$$

It then follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}
&\|A^{1/2}\nabla v\|^2 + \|a^{1/2}v\|^2 \\
&\quad = \left(A^{1/2}\nabla v - A^{-1/2}(\boldsymbol{\tau} - \mathbf{b}\,v),\, A^{1/2}\nabla v\right) \\
&\qquad + \left(a^{1/2}v - a^{-1/2}\nabla\cdot\boldsymbol{\tau},\, a^{1/2}v\right) - (\mathbf{b}\,v,\, \nabla v) \\
&\quad \le \left(\|A^{1/2}\nabla v - A^{-1/2}(\boldsymbol{\tau} - \mathbf{b}\,v)\| + C\,\|v\|\right)\|A^{1/2}\nabla v\| \\
&\qquad + \|a^{-1/2}\nabla\cdot\boldsymbol{\tau} - a^{1/2}v\|\,\|a^{1/2}v\|,
\end{aligned}$$

which implies

$$\|A^{1/2}\nabla v\|^2 + \|a^{1/2}v\|^2 \le C\, \left(b(\boldsymbol{\tau},\, v;\, \boldsymbol{\tau},\, v) + \|v\|^2\right). \tag{2.13}$$

By the triangle inequality and (2.13), we have that

$$\begin{aligned}
\|A^{-1/2}\boldsymbol{\tau}\| &\le \left(\|A^{-1/2}(\boldsymbol{\tau} - \mathbf{b}\,v) - A^{1/2}\nabla v\| + \|A^{1/2}\nabla v\| + C\,\|v\|\right) \\
&\le C\, \left(b(\boldsymbol{\tau},\, v;\, \boldsymbol{\tau},\, v)^{1/2} + \|v\|\right)
\end{aligned}$$

and that

$$\|a^{-1/2}\nabla\cdot\boldsymbol{\tau}\| \le \|a^{-1/2}\nabla\cdot\boldsymbol{\tau} - a^{1/2}v\| + \|a^{1/2}v\| \le C\, \left(b(\boldsymbol{\tau},\, v;\, \boldsymbol{\tau},\, v) + \|v\|^2\right).$$

Combining the above three inequalities yields (2.12).

With (2.12), we now show the validity of (2.10) by the standard compactness argument. To this end, assume that (2.10) is not true. This implies that there exists a sequence $\{\boldsymbol{\tau}_n,\, v_n\} \in H_N(\mathrm{div};\, \Omega) \times H^1_D(\Omega)$ such that

$$\|\!|\boldsymbol{\tau}_n|\!\|^2_{H(\mathrm{div})} + \|\!|v_n|\!\|^2_1 = 1 \quad \text{and} \quad b(\boldsymbol{\tau}_n,\, v_n;\, \boldsymbol{\tau}_n,\, v_n) \le \frac{1}{n}. \tag{2.14}$$

Since $H^1_D(\Omega)$ is compactly contained in $L^2(\Omega)$, there exists a subsequence $\{v_{n_k}\} \in H^1_D(\Omega)$ which converges in $L^2(\Omega)$. For any $k,\ l$, and $(\boldsymbol{\tau}_{n_k},\, v_{n_k})$, $(\boldsymbol{\tau}_{n_l},\, v_{n_l}) \in H_N(\mathrm{div};\, \Omega) \times H^1_D(\Omega)$, it follows from (2.12) and the triangle inequality that

$$\begin{aligned}
&\|\!|\boldsymbol{\tau}_{n_k} - \boldsymbol{\tau}_{n_l}|\!\|^2_{H(\mathrm{div})} + \|\!|v_{n_k} - v_{n_l}|\!\|^2_{1,\,\Omega} \\
&\quad \le C\left(b(\boldsymbol{\tau}_{n_k} - \boldsymbol{\tau}_{n_l},\, v_{n_k} - v_{n_l};\, \boldsymbol{\tau}_{n_k} - \boldsymbol{\tau}_{n_l},\, v_{n_k} - v_{n_l}) + \|v_{n_k} - v_{n_l}\|^2\right) \\
&\quad \le C\left(b(\boldsymbol{\tau}_{n_k},\, v_{n_k};\, \boldsymbol{\tau}_{n_k},\, v_{n_k}) + b(\boldsymbol{\tau}_{n_l},\, v_{n_l};\, \boldsymbol{\tau}_{n_l},\, v_{n_l}) + \|v_{n_k} - v_{n_l}\|^2\right) \\
&\quad \to 0,
\end{aligned}$$

as $k,\ l \to \infty$. This implies that $(\boldsymbol{\tau}_{n_k},\, v_{n_k})$ is a Cauchy sequence in the complete space $H_N(\mathrm{div};\,\Omega) \times H^1_D(\Omega)$. Hence, there exists $(\boldsymbol{\tau},\, v) \in H_N(\mathrm{div};\,\Omega) \times H^1_D(\Omega)$ such that

$$\lim_{k\to\infty} \left(\|\boldsymbol{\tau}_{n_k} - \boldsymbol{\tau}\|_{H(\mathrm{div})} + \|v_{n_k} - v\|_1 \right) = 0.$$

Next, we show that

$$v = 0 \quad \text{and} \quad \boldsymbol{\tau} = \mathbf{0}, \tag{2.15}$$

which contradict with (2.14) that

$$0 = \|\boldsymbol{\tau}\|^2_{H(\mathrm{div})} + \|v\|_1^2 = \lim_{k\to\infty} \left(\|\boldsymbol{\tau}_{n_k}\|^2_{H(\mathrm{div})} + \|v_{n_k}\|_1^2 \right) = 1.$$

To this end, for any $\phi \in H^1_D(\Omega)$, integration by parts, the Cauchy–Schwarz inequality, and (2.14) give

$$\begin{aligned}
\hat{a}(\phi,\, v_{n_k}) &= (\nabla\,\phi,\, (A\,\nabla + \mathbf{b})\, v_{n_k}) + (\phi,\, a\, v_{n_k}) \\
&= (\nabla\phi,\, (A\,\nabla + \mathbf{b})\, v_{n_k} - \boldsymbol{\tau}_{n_k}) + (\phi,\, a\, v_{n_k} - \nabla\cdot\boldsymbol{\tau}_{n_k}) \\
&\le b(\boldsymbol{\tau}_{n_k},\, v_{n_k};\, \boldsymbol{\tau}_{n_k},\, v_{n_k})^{1/2} \|\phi\|_1 \le \left(\frac{1}{n_k}\right)^{1/2} \|\phi\|_1.
\end{aligned}$$

Since $\lim_{k\to\infty} v_{n_k} = v$ in $H^1(\Omega)$, we then have

$$|\hat{a}(\phi,\, v)| = \lim_{k\to\infty} |\hat{a}(\phi,\, v_{n_k})| \le \lim_{k\to\infty} \left(\frac{1}{n_k}\right)^{1/2} \|\phi\|_1 = 0.$$

Because (2.4) has a unique solution, we have that

$$v = 0.$$

Now, $\boldsymbol{\tau} = \mathbf{0}$ follows from (2.12):

$$\|\boldsymbol{\tau}\|^2_{H(\mathrm{div})} = \lim_{k\to\infty} \|\boldsymbol{\tau}_{n_k}\|^2_{H(\mathrm{div})} \le C \lim_{k\to\infty} \left(b(\boldsymbol{\tau}_{n_k},\, v_{n_k};\, \boldsymbol{\tau}_{n_k},\, v_{n_k}) + \|v_{n_k}\|^2 \right) = 0.$$

This completes the proof of (2.15) and, hence, the theorem. ∎

THEOREM 2.3. *The variational formulation in* (2.9) *has a unique solution* $(\boldsymbol{\eta},\, w) \in H_N(\mathrm{div};\Omega) \times H^1_D(\Omega)$ *satisfying the following a priori estimate:*

$$\|(\boldsymbol{\eta},\, w)\| \le C \left(\|f\|_{-1,\Omega} + \|g_D\|_{1/2,\Gamma_D} + \|g_N\|_{-1/2,\Gamma_N} \right). \tag{2.16}$$

Proof. For all $(\boldsymbol{\tau},\, v) \in H_N(\mathrm{div};\Omega) \times H^1_D(\Omega)$, it follows from the definition of the dual norms and the trace theorem that

$$\begin{aligned}
|l(\boldsymbol{\tau},\, v)| &\le \|f\|_{-1,\Omega} \|v\|_{1,\Omega} + \|g_D\|_{1/2,\Gamma_D} \|\boldsymbol{\tau}\cdot\mathbf{n}\|_{-1/2,\Gamma_D} + \|g_N\|_{-1/2,\Gamma_N} \|v\|_{1/2,\Gamma_D} \\
&\le C \left(\|f\|_{-1,\Omega} + \|g_D\|_{1/2,\Gamma_D} + \|g_N\|_{-1/2,\Gamma_N} \right) \|(\boldsymbol{\tau},\, v)\|.
\end{aligned}$$

Now, by the Lax–Milgram lemma, the well-posedness of (2.9) and the a priori estimate in (2.16) follow directly from Theorem 2.2. ∎

3. Div FOSLL* finite element approximation. Theorem 2.2 guarantees that conforming finite element spaces of $H_N(\mathrm{div};\Omega) \times H_D^1(\Omega)$ for the vector and scalar variables, $\boldsymbol{\eta}$ and w, may be chosen independently. However, the only finite element spaces having optimal approximations in terms of both the regularity and the approximation property are the continuous piecewise polynomials for the scalar variable and the Raviart–Thomas (or Brezzi–Douglas–Marini) elements for the vector variable. (The Brezzi–Douglas–Marini element has slightly more degrees of freedom than that of the Raviart–Thomas element.) Moreover, the system of algebraic equations resulting from these elements can be solved efficiently by fast multigrid methods. For the above reasons, only these elements are analyzed in this paper. But it is easy to see that our analysis does apply to any other conforming finite element spaces with no essential modifications.

For simplicity of presentation, we consider only triangular and tetrahedra elements for the respective two and three dimensions. Assuming that the domain Ω is polygonal, let $\mathcal{T}_h$ be a regular triangulation of Ω (see [13]) with triangular/tetrahedra elements of size $\mathcal{O}(h)$. Let $P_k(K)$ be the space of polynomials of degree k on triangle K and denote the local Raviart–Thomas space of order k on K:

$$RT_k(K) = P_k(K)^d + \mathbf{x}\, P_k(K)$$

with $\mathbf{x} = (x_1, \ldots, x_d)$. Then the standard $H(\mathrm{div};\,\Omega)$ conforming Raviart–Thomas space of index k [17] and the standard (conforming) continuous piecewise polynomials of degree $k+1$ are defined, respectively, by

$$\Sigma_h^k = \{\boldsymbol{\tau} \in H_N(\mathrm{div};\Omega) \,:\, \boldsymbol{\tau}|_K \in RT_k(K) \;\; \forall\, K \in \mathcal{T}_h\}$$

$$\text{and} \qquad V_h^{k+1} = \{v \in H_D^1(\Omega) \,:\, v|_K \in P_{k+1}(K) \;\; \forall\, K \in \mathcal{T}_h\}.$$

It is well-known (see [13]) that V_h^{k+1} has the following approximation property: let $k \ge 0$ be an integer and let $l \in [0,\, k+1]$,

$$\inf_{v \in V_h^{k+1}} \|u - v\|_1 \le C\, h^l\, \|u\|_{l+1} \tag{3.1}$$

for $u \in H^{l+1}(\Omega) \cap H_D^1(\Omega)$. It is also well-known (see [17]) that Σ_h^k has the following approximation property: let $k \ge 0$ be an integer and let $l \in [1,\, k+1]$,

$$\inf_{\boldsymbol{\tau} \in \Sigma_h^k} \|\boldsymbol{\sigma} - \boldsymbol{\tau}\|_{H(\mathrm{div};\,\Omega)} \le C\, h^l\, (\|\boldsymbol{\sigma}\|_l + \|\nabla\cdot\boldsymbol{\sigma}\|_l) \tag{3.2}$$

for $\boldsymbol{\sigma} \in H^l(\Omega)^d \cap H_N(\mathrm{div};\Omega)$ with $\nabla\cdot\boldsymbol{\sigma} \in H^l(\Omega)^m$. Since $\boldsymbol{\sigma}$ and $\nabla\cdot\boldsymbol{\sigma}$ are one order less smooth than u, we will choose k to be the smallest integer greater than or equal to $l-1$.

The finite element discretization of the FOSLL* variational problem is to find $(\boldsymbol{\eta}_h,\, w_h) \in \Sigma_h^k \times V_h^{k+1}$ such that

$$b(\boldsymbol{\eta}_h,\, w_h;\, \boldsymbol{\tau},\, v) = l(\boldsymbol{\tau},\, v) \quad \forall\, (\boldsymbol{\tau},\, v) \in \Sigma_h^k \times V_h^{k+1}. \tag{3.3}$$

Since $\Sigma_h^k \times V_h^{k+1}$ is a subspace of $H_N(\mathrm{div};\Omega) \times H_D^1(\Omega)$, the div FOSLL* problem in (3.3) is well-posed and the solution continuously depends on the data.

THEOREM 3.1. *The variational formulation in* (3.3) *has a uniques solution* $(\boldsymbol{\eta}_h,\, w_h) \in \Sigma_h^k \times V_h^{k+1}$ *satisfying the following a priori estimate:*

$$|\!|\!|(\boldsymbol{\eta}_h,\, w_h)|\!|\!| \le C\, \left(\|f\|_{-1,\Omega} + \|g_D\|_{1/2,\Gamma_D} + \|g_N\|_{-1/2,\Gamma_N}\right). \tag{3.4}$$

Now, the finite element approximation to $(\boldsymbol{\sigma},\, u)$ is defined as follows:

$$\boldsymbol{\sigma}_h = \boldsymbol{\eta}_h - A\,\nabla\, w_h - \mathbf{b}\, w_h \quad \text{and} \quad u_h = -\alpha^{-1}\,(\nabla\cdot\,\boldsymbol{\eta}_h - a\, w_h). \tag{3.5}$$

Remark 3.2. When the coefficients *(A*, **b**, and *a)* are not polynomials, they can be replaced by their approximations of appropriate polynomials locally, if piecewise polynomial approximation to $(\boldsymbol{\sigma},\, u)$ is desirable.

Remark 3.3. The FOSLL* approximation to the solution u is not continuous. To obtain a continuous approximation, one can simply project u_h onto appropriate continuous finite element space.

4. A priori error estimate. The difference between equations in (2.9) and (3.3) gives the error equation

$$b(\boldsymbol{\eta}-\boldsymbol{\eta}_h,\, w-w_h;\, \boldsymbol{\tau},\, v)=0 \quad \forall\ (\boldsymbol{\tau},\, v)\in \Sigma_h^k\times V_h^{k+1}. \tag{4.1}$$

THEOREM 4.1. *Assume that the solution* $(\boldsymbol{\sigma},\, u)$ *of* (2.5)–(2.6) *is in* $H^l(\Omega)^d\times H^{l+1}(\Omega)$ *and that the solution* $(\boldsymbol{\eta},\, w)$ *of* (2.9) *satisfies*

$$\|\nabla\cdot\boldsymbol{\eta}\|_l+\|\boldsymbol{\eta}\|_l+\|w\|_{l+1}\le C\,\left(\|A^{-1/2}\boldsymbol{\sigma}\|_l+\|\alpha^{1/2}u\|_l\right). \tag{4.2}$$

Let k be the smallest integer greater than or equal to $l-1$. Then the FOSLL approximation* $(\boldsymbol{\sigma}_h,\, u_h)$ *defined in* (3.5) *has the following error estimate:*

$$\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|+\|\alpha^{1/2}(u-u_h)\|\le C\,h^l\big(\|\boldsymbol{\sigma}\|_l+\|u\|_l\big)\le C\,h^l\|u\|_{l+1}. \tag{4.3}$$

Proof. Let $(\boldsymbol{\eta}_h,\, w_h)$ be the solution of (3.3). It follows from Theorem 2.2 and the error equation in (4.1) that for any $(\boldsymbol{\tau},\, v)\in\Sigma_h^k\times V_h^{k+1}$

$$\begin{aligned}\alpha_0|\!|\!|(\boldsymbol{\eta}-\boldsymbol{\eta}_h,\, w-w_h)|\!|\!|^2 &\le b(\boldsymbol{\eta}-\boldsymbol{\eta}_h,\, w-w_h;\, \boldsymbol{\eta}-\boldsymbol{\tau},\, w-v)\\ &\le C\,|\!|\!|(\boldsymbol{\eta}-\boldsymbol{\eta}_h,\, w-w_h)|\!|\!|\;|\!|\!|(\boldsymbol{\eta}-\boldsymbol{\tau},\, w-v)|\!|\!|,\end{aligned}$$

which, together with the approximation properties in (3.1) and (3.2), implies

$$\begin{aligned}|\!|\!|(\boldsymbol{\eta}-\boldsymbol{\eta}_h,\, w-w_h)|\!|\!| &\le C\,\left(\inf_{\boldsymbol{\tau}\in\Sigma_h^k}\|\boldsymbol{\eta}-\boldsymbol{\tau}\|_{H(\mathrm{div};\,\Omega)}+\inf_{v\in V_h^{k+1}}\|w-v\|_1\right)\\ &\le C\,h^l\,(\|\nabla\cdot\boldsymbol{\eta}\|_l+\|\boldsymbol{\eta}\|_l+\|w\|_{l+1}).\end{aligned}$$

Now, (4.3) is a direct consequence of (4.2). This completes the proof of the theorem. ∎

Remark 4.2. When $l=0$, assumption (4.2) is the coercivity bound in (2.10).

5. A posteriori error estimate. Unlike the FOSLS approach, the FOSLL* approach does not have a free a posteriori error estimator; thus in this section we study an explicit residual error estimator and establish its reliability and efficiency bounds.

5.1. Local indicator and global estimator. Since the bilinear form $b(\cdot,\,\cdot)$ is coercive and continuous in $H_N(\mathrm{div};\Omega)\times H_D^1(\Omega)$ (see Theroem 2.2), the explicit residual a posteriori error estimator to be derived in this paper is a combination of those for the $H(\mathrm{div})$ and the elliptic problems (see [12, 1, 18, 19]).

To this end, we first introduce some notation. For the sake of simplicity, we use terms in three dimensions. Denote by $\mathcal{E}_K$ the set of faces of element $K\in\mathcal{T}_h$ and the set of faces of the triangulation $\mathcal{T}_h$ by $\mathcal{E}_h:=\mathcal{E}_I\cup\mathcal{E}_D\cup\mathcal{E}_N$, where $\mathcal{E}_I$ is the set of interior element faces, and $\mathcal{E}_D$ and $\mathcal{E}_N$ are the sets of boundary faces belonging to the respective Γ_D and Γ_N. For each $e \in \mathcal{E}$, denote by h_e the diameter of the face e and by $\mathbf{n}_e$ a unit vector normal to e. Let K_e^- and K_e^+ be the two elements sharing the common face e such that the unit outward normal vector of K_e^- coincides with $\mathbf{n}_e$. When $e \in \mathcal{E}_D \cup \mathcal{E}_N$, $\mathbf{n}_e$ is the unit outward vector normal to $\partial\Omega$ and denote by K_e^- the element having the face e. For a function v defined on $K_e^- \cup K_e^+$, denote its traces on $e \in \mathcal{E}$ by $v|_e^-$ and $v|_e^+$, respectively. The jump over the face e is denoted by

$$[\![v]\!]_e := \begin{cases} v|_e^- - v|_e^+, & e \in \mathcal{E}_I, \\ v|_e^-, & e \in \mathcal{E}_D \cup \mathcal{E}_N. \end{cases}$$

(When there is no ambiguity, the subscript or superscript e in the designation of the jump will be dropped.) For any $K \in \mathcal{T}_h$, denote by h_K the diameter of the element K. For a function v, we will use the following notation on the weighted L^2 norms:

$$\|h\,v\|_{\mathcal{T}_h} = \left(\sum_{K\in\mathcal{T}_h} \|h\,v\|_K^2\right)^{1/2}, \qquad \text{where } \|h\,v\|_K = \|h_K\,v\|_{0,K} \quad \forall\, K \in \mathcal{T}_h,$$

$$\text{and} \quad \|h\,v\|_{\mathcal{E}_h} = \left(\sum_{e\in\mathcal{E}_h} \|h\,v\|_e^2\right)^{1/2}, \qquad \text{where } \|h\,v\|_e = \|h_e\,v\|_{0,e} \quad \forall\, e \in \mathcal{E}_h.$$

Let $(\boldsymbol{\eta}_h, w_h)$ be the solution of (3.3) and let $(\boldsymbol{\sigma}_h, u_h)$ be the finite element approximation to $(\boldsymbol{\sigma}, u)$ defined in (3.5). On each element $K \in \mathcal{T}_h$, denote the following element residuals by

$$r_1|_K = \left[f - \nabla\cdot\boldsymbol{\sigma}_h + \mathbf{b}\cdot A^{-1}\boldsymbol{\sigma}_h - a\,u_h\right]_K, \qquad \mathbf{r}_2|_K = \left[A^{-1}\boldsymbol{\sigma}_h + \nabla u_h\right]_K,$$

$$\text{and} \quad r_3|_K = \nabla\times(A^{-1}\boldsymbol{\sigma}_h)|_K.$$

Denote the following face jumps by

$$\begin{array}{llll} J_1|_e = [\![\boldsymbol{\sigma}_h\cdot\mathbf{n}]\!]_e, & J_2|_e = [\![u_h]\!]_e, & J_3|_e = [\![A^{-1}\boldsymbol{\sigma}_h\cdot\mathbf{t}]\!]_e & \text{on } e \in \mathcal{E}_I, \\ J_1|_e = 0, & J_2|_e = [u_h - g_D]_e, & J_3|_e = \left[\nabla g_D\cdot\mathbf{t} + A^{-1}\boldsymbol{\sigma}_h\cdot\mathbf{t}\right]_e & \text{on } e \in \mathcal{E}_D, \\ J_1|_e = [\boldsymbol{\sigma}_h\cdot\mathbf{n} - g_N]_e, & J_2|_e = 0, & J_3|_e = 0 & \text{on } e \in \mathcal{E}_N. \end{array}$$

Let $\bar{\mathbf{r}}_2|_K$, $\bar{r}_1|_K$ and $\bar{r}_3|_K$, and $\bar{J}_i|_e$ $(i = 1, 2, 3)$ be the L^2-projections of the respective $\mathbf{r}_2|_K$, $r_1|_K$ and $r_3|_K$, and $J_i|_e$ $(i = 1, 2, 3)$ onto $P_k(K)^2$, $P_k(K)$, and $P_k(e)$, respectively. Now, the local error indicator on each element $K \in \mathcal{T}_h$ is defined by

$$\eta_K^2 = \|h\,\bar{r}_1\|_K^2 + \|h\,\bar{\mathbf{r}}_2\|_K^2 + \|h\,\bar{r}_3\|_K^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_I\cap\mathcal{E}_K} \left(\|h^{1/2}\bar{J}_1\|_e^2 + \|h^{1/2}\bar{J}_3\|_e^2\right)$$

$$+ \sum_{e\in\mathcal{E}_D\cap\mathcal{E}_K} \|h^{1/2}\bar{J}_3\|_e^2 + \sum_{e\in\mathcal{E}_N\cap\mathcal{E}_K} \|h^{1/2}\bar{J}_1\|_e^2, \tag{5.1}$$

and the global error estimator is defined by

$$\eta^2 = \sum_{K\in\mathcal{T}_h} \eta_K^2 = \|h\,\bar{r}_1\|_{\mathcal{T}_h}^2 + \|h\,\bar{\mathbf{r}}_2\|_{\mathcal{T}_h}^2 + \|h\,\bar{r}_3\|_{\mathcal{T}_h}^2 + \|h^{1/2}\bar{J}_1\|_{\mathcal{E}_h}^2 + \|h^{1/2}\bar{J}_3\|_{\mathcal{E}_h}^2. \tag{5.2}$$

The terms r_1 and $\mathbf{r}_2$ are the residuals of the equations in (2.5). The term r_3 measures the violation of the fact that the exact quantity $-A^{-1}\boldsymbol{\sigma} = \nabla u$ is in the kernel of the $\nabla\times$ operator. The terms J_1, J_2, and J_3 are due to the fact that the numerical flux $\boldsymbol{\sigma}_h$, the numerical solution u_h, and the numerical gradient $-A^{-1}\boldsymbol{\sigma}_h$ are not in $H(\mathrm{div};\Omega)$, $H^1(\Omega)$, and $H(\mathrm{curl};\Omega)$, respectively.

5.2. Reliability and efficiency bounds. For simplicity, we analyze only two dimensions here since there are no essential difficulties for three dimensions. For a vector field $\boldsymbol{\tau} = (\tau_1, \tau_2)^t$ and a scalar-valued function v, define the respective curl operator and its formal adjoint by

$$\nabla\times\boldsymbol{\tau} := \frac{\partial \tau_2}{\partial x_1} - \frac{\partial \tau_1}{\partial x_2} \quad \text{and} \quad \nabla^{\perp} v := \left(\frac{\partial v}{\partial x_2}, -\frac{\partial v}{\partial x_1}\right)^t.$$

Denote by $\Pi_h : H_N(\mathrm{div};\Omega) \cap L^s(\Omega)^2 \to \Sigma_h^0$ with $s > 2$ the RT_0 interpolation operator; i.e., for all $\boldsymbol{\tau} \in H^1(\Omega)^2$, one has [5]

$$\langle(\boldsymbol{\tau} - \Pi_h\boldsymbol{\tau}) \cdot \mathbf{n}, v\rangle_e = 0 \quad \forall\, v \in P_0(e) \quad \text{and} \quad \forall\, e \in \mathcal{E}. \tag{5.3}$$

Let S_h be the standard continuous piecewise linear finite element space on the triangulation $\mathcal{T}_h$. For $B = D$ or N, denote by $I_{_B} : H_B^1(\Omega) \to H_B^1(\Omega) \bigcap S_h$ the Clement interpolation operator which satisfies the following local approximation property [6]:

$$\|h^{-1}(v - I_{_B} v)\|_{\mathcal{T}_h} \le C\, \|\nabla v\| \quad \forall\, v \in H_B^1(\Omega),$$

where $B = D$ or N. It is easy to check that $\nabla^{\perp}(I_{_N} v) \in \Sigma_h^0$.

Let $(\boldsymbol{\eta},\, w)$ be the solution of (2.9), and let

$$\mathbf{E} = \boldsymbol{\eta} - \boldsymbol{\eta}_h \in H_N(\mathrm{div};\Omega) \quad \text{and} \quad e = w - w_h \in H_D^1(\Omega). \tag{5.4}$$

By Lemmas 5.1 in [12], the $\mathbf{E}$ has the following quasi-Helmholtz decomposition:

$$\mathbf{E} = \boldsymbol{\phi} + \nabla^{\perp}\psi \quad \text{in} \ \ \Omega, \tag{5.5}$$

where $\boldsymbol{\phi} \in H_N^1(\Omega)^2$ and $\psi \in H_N^1(\Omega)$. Moreover, there exists a constant $C > 0$ such that

$$\|\nabla\boldsymbol{\phi}\| \le C\, \|\mathbf{E}\|_{H(\mathrm{div})} \quad \text{and} \quad \|\nabla\psi\| \le C\, \|\mathbf{E}\|_{H(\mathrm{div})}. \tag{5.6}$$

(In three dimensions, a decomposition similar to that in (5.5) was also established in Lemma 7.1 of [12] under the assumption that the domain Ω is topologically equivalent to a ball.) Let

$$\boldsymbol{\phi}_h := \Pi_h\boldsymbol{\phi}, \quad \psi_h := I_{_N}\psi, \quad \text{and} \quad e_h := I_{_D} e$$

and let

$$\tilde{e} = e - e_h, \quad \tilde{\boldsymbol{\phi}} = \boldsymbol{\phi} - \boldsymbol{\phi}_h, \quad \text{and} \quad \tilde{\psi} = \psi - \psi_h.$$

By the approximation properties of the interpolation operators and (5.6), we have

$$\|h^{-1}\tilde{e}\|_{\mathcal{T}_h} + \|h^{-1/2}\tilde{e}\|_{\mathcal{E}_h} \le C\, \|\nabla e\| \le C\, \|e\|_1,$$

$$\|h^{-1}\tilde{\boldsymbol{\phi}}\|_{\mathcal{T}_h} + \|h^{-1/2}\tilde{\boldsymbol{\phi}}\|_{\mathcal{E}_h} \le C\, \|\nabla\boldsymbol{\phi}\| \le C\, \|\mathbf{E}\|_{H(\mathrm{div})},$$

$$\text{and} \quad \|h^{-1}\tilde{\psi}\|_{\mathcal{T}_h} + \|h^{-1/2}\tilde{\psi}\|_{\mathcal{E}_h} \le C\, \|\nabla\psi\| \le C\, \|\mathbf{E}\|_{H(\mathrm{div})}.$$

LEMMA 5.1. *We have the following error representation:*

$$\begin{aligned} b(\mathbf{E}, e; \mathbf{E}, e) = &\sum_{K\in\mathcal{T}_h} \left\{(r_1,\, \tilde{e})_{_K} - (\mathbf{r}_2,\, \tilde{\boldsymbol{\phi}})_{_K} - (r_3,\, \tilde{\psi})_{_K}\right\} \\ &+ \sum_{e\in\mathcal{E}_h} \left\{\langle J_1,\, \tilde{e}\rangle_e + \langle J_2, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n}\rangle_e + \langle J_3, \tilde{\psi}\rangle_e\right\}. \end{aligned} \tag{5.7}$$

Proof. Since $\mathbf{E}_h := \boldsymbol{\phi}_h + \nabla^\perp \psi_h \in \Sigma_h^0 \subset \Sigma_h^k$ and $e_h \in V_h^1 \subset V_h^{k+1}$, the error equation in (4.1) gives

$$b(\mathbf{E}, e; \mathbf{E}, e) = b(\mathbf{E}, e; \mathbf{E} - \mathbf{E}_h, e - e_h) = b(\mathbf{E}, e; \tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}, \tilde{e}). \tag{5.8}$$

By the fact that $\nabla \cdot \nabla^\perp \tilde{\psi} = 0$, the definitions of the bilinear form $b(\cdot, \cdot)$, and the FOSLL* finite element approximation $(\boldsymbol{\sigma}_h, u_h)$, we have

$$\begin{aligned} b(\boldsymbol{\eta}_h, w_h; \tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}, \tilde{e}) &= (\boldsymbol{\sigma}_h,\, A^{-1}(\tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}) - \nabla \tilde{e} - A^{-1}\mathbf{b}\, \tilde{e}) - (u_h, \nabla \cdot \tilde{\boldsymbol{\phi}} - a\, \tilde{e}) \\ &= (A^{-1}\boldsymbol{\sigma}_h,\, \tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}) - (u_h,\, \nabla \cdot \tilde{\boldsymbol{\phi}}) - (\boldsymbol{\sigma}_h,\, \nabla \tilde{e}) - (\mathbf{b}^t A^{-1}\boldsymbol{\sigma}_h - a\, u_h,\, \tilde{e}). \end{aligned}$$

It follows from integration by parts and the boundary conditions that

$$\begin{aligned} (A^{-1}\boldsymbol{\sigma}_h,\, \nabla^\perp \tilde{\psi}) =& \sum_{K \in \mathcal{T}_h} \big(\nabla \times (A^{-1}\boldsymbol{\sigma}_h),\, \tilde{\psi}\big)_K \\ & - \sum_{e \in \mathcal{E}_I} \langle [\![A^{-1}\boldsymbol{\sigma}_h \cdot \mathbf{t}]\!],\, \tilde{\psi} \rangle_e - \sum_{e \in \mathcal{E}_D} \langle A^{-1}\boldsymbol{\sigma}_h \cdot \mathbf{t},\, \tilde{\psi} \rangle_e, \\ (u_h,\, \nabla \cdot \tilde{\boldsymbol{\phi}}) =& - \sum_{K \in \mathcal{T}_h} (\nabla u_h,\, \tilde{\boldsymbol{\phi}})_K + \sum_{e \in \mathcal{E}_I} \langle [\![u_h]\!],\, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n} \rangle_e + \sum_{e \in \mathcal{E}_D} \langle u_h,\, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n} \rangle_e, \\ (\boldsymbol{\sigma}_h,\, \nabla \tilde{e}) =& - \sum_{K \in \mathcal{T}_h} (\nabla \cdot \boldsymbol{\sigma}_h,\, \tilde{e})_K + \sum_{e \in \mathcal{E}_I} \langle [\![\boldsymbol{\sigma}_h \cdot \mathbf{n}]\!],\, \tilde{e} \rangle_e + \sum_{e \in \mathcal{E}_N} \langle \boldsymbol{\sigma}_h \cdot \mathbf{n},\, \tilde{e} \rangle_e, \end{aligned}$$

which, together with the definitions of the residuals and the jumps, lead to

$$\begin{aligned} & b(\boldsymbol{\eta}_h, w_h; \tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}, \tilde{e}) \\ &= \sum_{K \in \mathcal{T}_h} \Big\{ \big(r_3,\, \tilde{\psi}\big)_K + \big(\mathbf{r}_2,\, \tilde{\boldsymbol{\phi}}\big)_K - \big(r_1,\, \tilde{e}\big)_K \Big\} + \sum_{K \in \mathcal{T}_h} \big(f,\, \tilde{e}\big)_K - \sum_{e \in \mathcal{E}_I} \langle J_3,\, \tilde{\psi} \rangle_e \\ & \quad - \sum_{e \in \mathcal{E}_D} \langle A^{-1}\boldsymbol{\sigma}_h \cdot \mathbf{t},\, \tilde{\psi} \rangle_e - \sum_{e \in \mathcal{E}} \langle J_2,\, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n} \rangle_e \\ & \quad - \sum_{e \in \mathcal{E}} \langle J_1,\, \tilde{e} \rangle_e - \sum_{e \in \mathcal{E}_D} \langle g_D,\, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n} \rangle_e - \sum_{e \in \mathcal{E}_N} \langle g_N,\, \tilde{e} \rangle_e. \end{aligned}$$

By (2.9), integration by parts, and the boundary condition of $\tilde{\psi} \in H_N^1(\Omega)$, we have

$$\begin{aligned} b(\boldsymbol{\eta}, w; \tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}, \tilde{e}) &= (f,\, \tilde{e}) - \langle g_N, \tilde{e} \rangle_{\Gamma_N} - \langle g_D, (\tilde{\boldsymbol{\phi}} + \nabla^\perp \tilde{\psi}) \cdot \mathbf{n} \rangle_{\Gamma_D} \\ &= (f,\, \tilde{e}) - \langle g_N, \tilde{e} \rangle_{\Gamma_N} - \langle g_D,\, \tilde{\boldsymbol{\phi}} \cdot \mathbf{n} \rangle_{\Gamma_D} + \langle \nabla g_D \cdot \mathbf{t},\, \tilde{\psi} \rangle_{\Gamma_D}. \end{aligned}$$

Now, (5.7) is a direct consequence of (5.8) and the difference of the above two equalities. This completes the proof of the lemma. ∎

Define the local and global oscillations as follows:

$$\begin{aligned} \operatorname{osc}_k^2(K) =& \|h\,(r_1 - \bar{r}_1)\|_K^2 + \|h\,(\mathbf{r}_2 - \bar{\mathbf{r}}_2)\|_K^2 + \|h\,(r_3 - \bar{r}_3)\|_K^2 + \|h^{1/2}(J_1 - \bar{J}_1)\|_{\partial K}^2 \\ & + \|h^{1/2}(J_2 - \bar{J}_2)\|_{\partial K}^2 + \|h^{1/2}(J_3 - \bar{J}_3)\|_{\partial K}^2 \end{aligned} \tag{5.9}$$

$$\text{and} \quad \operatorname{osc}_k^2(\mathcal{T}_h) = \sum_{K \in \mathcal{T}_h} \operatorname{osc}_k^2(K), \tag{5.10}$$

respectively.

THEOREM 5.2 (reliability bound). *The global estimator* η *defined in* (5.2) *is reliable; i.e., there exists a positive constant* C *such that*

$$\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\| + \|\alpha^{1/2}(u-u_h)\| \le C\,|\!|\!|(\mathbf{E},e)|\!|\!| \le C\,(\eta + \operatorname{osc}_k(\mathcal{T}_h)). \tag{5.11}$$

Proof. The first inequality in (5.11) is a direct consequence of the definition of $(\boldsymbol{\sigma}_h,\, u_h)$ and the triangle inequality.

To show the validity of the second inequality in (5.11), by the coercivity in (2.10), it suffices to show that

$$b(\mathbf{E},e;\,\mathbf{E},e) \le C\,(\eta + \operatorname{osc}_k(\mathcal{T}_h))\,|\!|\!|(\mathbf{E},e)|\!|\!|. \tag{5.12}$$

To this end, first notice that by the property in (5.3) and the definition of $\bar{J}_2$, we have

$$\langle J_2, (\boldsymbol{\phi}-\boldsymbol{\phi}_h)\cdot\mathbf{n}\rangle_e = \langle J_2-\bar{J}_2, (\boldsymbol{\phi}-\boldsymbol{\phi}_h)\cdot\mathbf{n}\rangle_e.$$

Now, it follows from Lemma 5.1, the Cauchy–Schwarz inequality, the approximation properties of $\tilde{e}$, $\tilde{\boldsymbol{\phi}}$, and $\tilde{\psi}$, and the triangle inequality that

$$\begin{aligned}
&b(\mathbf{E},e;\,\mathbf{E},e)\\
&= \sum_{K\in\mathcal{T}_h}\big((r_1,\,\tilde{e})_K + (\mathbf{r}_2,\,\tilde{\boldsymbol{\phi}})_K + (r_3,\,\tilde{\psi})_K\big)\\
&\quad + \sum_{e\in\mathcal{E}}\big(\langle J_1,\,\tilde{e}\rangle_e + \langle J_2-\bar{J}_2, \tilde{\boldsymbol{\phi}}\cdot\mathbf{n}\rangle_e + \langle J_3, \tilde{\psi}\rangle_e\big)\\
&\le \sum_{K\in\mathcal{T}_h}\big(\|h\,r_1\|_K\,\|h^{-1}\tilde{e}\|_K + \|h\,\mathbf{r}_2\|_K\,\|h^{-1}\tilde{\boldsymbol{\phi}}\|_K + \|h\,r_3\|_K\,\|h^{-1}\tilde{\psi}\|_K\big)\\
&\quad + \sum_{e\in\mathcal{E}_h}\Big(\|h^{1/2}J_1\|_e\,\|h^{-1/2}\tilde{e}\|_e + \|h^{1/2}(J_2-\bar{J}_2)\|_e\,\|h^{-1/2}\tilde{\boldsymbol{\phi}}\cdot\mathbf{n}\|_e\\
&\quad + \|h^{1/2}J_3\|_e\,\|h^{-1/2}\tilde{\psi}\|_e\Big)\\
&\le C\left(\sum_{i=1,3}\big(\|h\,r_i\|^2_{\mathcal{T}_h} + \|h^{1/2}J_i\|^2_{\mathcal{E}_h}\big) + \|h\,\mathbf{r}_2\|^2_{\mathcal{T}_h} + \|h^{1/2}(J_2-\bar{J}_2)\|^2_{\mathcal{E}_h}\right)^{1/2} |\!|\!|(\mathbf{E},e)|\!|\!|\\
&\le C\,(\eta + \operatorname{osc}_k(\mathcal{T}_h))\,|\!|\!|(\mathbf{E},e)|\!|\!|,
\end{aligned}$$

which proves (5.12) and, hence, the theorem. ∎

THEOREM 5.3 (local efficiency bound). *For all* $K\in\mathcal{T}_h$*, the local error indicator* η_K *defined in* (5.1) *is efficient; i.e., there exists a positive constant* C *such that*

$$C\,\eta_K \le \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\omega_K} + \|u-u_h\|_{\omega_K} + \operatorname{osc}_k(\omega_K), \tag{5.13}$$

where ω_K *is the union of elements in* $\mathcal{T}_h$ *sharing an edge with* K.

The proof of the local efficiency bound in Theorem 5.3 is standard; i.e., it is proved by using local edge and element bubble functions, ϕ_e and ϕ_K (see [18] for their definitions and properties). For simplicity, we only sketch the proof below.

Proof. For any $(\boldsymbol{\tau}, v) \in H_N(\text{div};\Omega) \times H_D^1(\Omega)$, by the quasi-Helmholtz decomposition, we have

$$\boldsymbol{\tau} = \boldsymbol{\phi} + \nabla^\perp \psi \in H_N(\text{div};\Omega),$$

where $\boldsymbol{\phi} \in H_N^1(\Omega)$ and $\psi \in H_N^1(\Omega)$. The same argument as in the proof of Lemma 5.1 gives

$$\begin{aligned} b(\mathbf{E}, e; \boldsymbol{\tau}, v) = &\sum_{K \in \mathcal{T}_h} \{(r_1, v)_K + (\mathbf{r}_2, \boldsymbol{\phi})_K + (r_3, \psi)_K\} \\ &+ \sum_{e \in \mathcal{E}_h} \{\langle J_1, v\rangle_e + \langle J_2, \boldsymbol{\phi}\cdot\mathbf{n}\rangle_e + \langle J_3, \psi\rangle_e\}, \end{aligned}$$

which, together with the definitions of $\boldsymbol{\sigma}$ and u, yields

$$b(\mathbf{E}, e; \boldsymbol{\tau}, v) = \big(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h,\ A^{-1}\boldsymbol{\tau} - (\nabla + A^{-1}\mathbf{b})\, v\big) - (u - u_h,\ \nabla\cdot\boldsymbol{\tau} - a\, v).$$

Hence,

$$\begin{aligned} &\sum_{K \in \mathcal{T}_h} ((r_1, v)_K + (\mathbf{r}_2, \boldsymbol{\phi})_K + (r_3, \psi)_K) + \sum_{e \in \mathcal{E}_h} (\langle J_1, v\rangle_e + \langle J_2, \boldsymbol{\phi}\cdot\mathbf{n}\rangle_e + \langle J_3, \psi\rangle_e) \\ &= \big(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h,\ A^{-1}(\boldsymbol{\phi} + \nabla^\perp\psi) - (\nabla + A^{-1}\mathbf{b})\, v\big) - (u - u_h,\ \nabla\cdot\ \boldsymbol{\phi} - a\, v). \qquad (5.14) \end{aligned}$$

In (5.14), by choosing (1) $\boldsymbol{\phi} = 0$, $\psi = 0$, and $v = \bar{r}_1\, \phi_K$, (2) $\boldsymbol{\phi} = \bar{\mathbf{r}}_2\, \phi_K$, $\psi = 0$, and $v = 0$, and (3) $\boldsymbol{\phi} = 0$, $\psi = \bar{r}_3\, \phi_K$, and $v = 0$, and by the standard argument, we can then establish upper bounds for the element residuals, $\|h\, \bar{r}_1\|_K$, $\|h\, \bar{\mathbf{r}}_2\|_K$, and $\|h\, \bar{r}_3\|_K$, respectively. In a similar fashion, to bound the edge jumps $\|h\, \bar{J}_1\|_e$ and $\|h\, \bar{J}_3\|_e$ above, we choose (1) $\boldsymbol{\phi} = 0$, $\psi = 0$, and $v = \bar{J}_1\, \phi_e$ and (2) $\boldsymbol{\phi} = 0$, $\psi = \bar{J}_3\, \phi_e$, and $v = 0$ in (5.14), respectively. ∎

6. Numerical results. In this section, numerical results for a second-order elliptic partial differential equation are presented.

We begin with discretizations of a test problem on a sequence of uniform meshes to verify the a priori error estimation. The test problem is defined on $\Omega = (0,1)^2$ with coefficients $A = I$, $\mathbf{b} = (3,2)^t$, and $a = 2$. The exact solution of this problem is $u = \sin(\pi x)\sin(\pi y)$ with homogeneous boundary condition on $\partial\Omega$. Finite element spaces Σ_h^0 and V_h^1 are used to approximate $\boldsymbol{\tau}$ and w, respectively. Table 1 shows that the convergence rates of the errors for the original variables $\boldsymbol{\sigma}$ and u in L^2 norms are optimal.

TABLE 1
Errors $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|$, $\|u - u_h\|$, $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\| + \|u - u_h\|$, and convergence rates.

h	$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|$	Rate	$\|u - u_h\|$	Rate	$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\| + \|u - u_h\|$	Rate
1/8	5.859E-1		4.351E-2		6.294E-1	
1/16	2.972E-1	1.971	1.601E-2	2.718	3.132E-1	2.010
1/32	1.492E-1	1.992	7.013E-3	2.283	1.562E-1	2.005
1/64	7.466E-2	1.998	3.367E-3	2.013	7.802E-2	2.002
1/128	3.734E-2	2.000	1.666E-3	2.201	3.900E-2	2.006

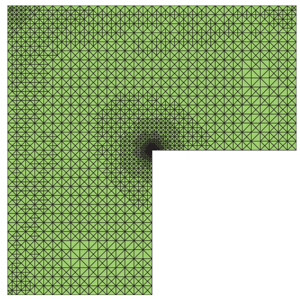

Fig. 1. *Mesh generated by error estimator.*

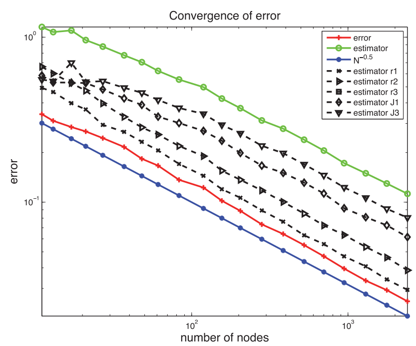

Fig. 2. *Error* $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\| + \|u - u_h\|$ *and estimator.*

The next example is to test the a posteriori error estimator. The test problem is the Laplace equation $-\Delta u = 0$ defined on an L-shaped domain $\Omega := (-1,1)^2 \backslash [0,1) \times (-1,0]$ with a reentrant corner at the origin. The Dirichlet boundary condition $u|_{\partial\Omega} = g_D$ is chosen such that the exact solution is $u(r,\theta) = r^{2/3}\sin(2\theta)$ in polar coordinates. Starting with the coarsest triangulation $\mathcal{T}_0$ obtained from halving 12 uniform squares, a sequence of meshes is generated by using the standard adaptive meshing algorithm that adopts the bulk marking strategy with bulk marking parameter 0.5. Marked triangles are refined by bisection.

Mesh generated by η is shown in Figure 1. The refinement is mainly around the reentrant corner. Comparison of the true error and the η is shown in Figure 2. Moreover, the slope of the log(dof)-log(error) for the η and the true error is $-1/2$, which indicates the optimal decay of the error with respect to the number of unknowns. Comparison of the true error and the η is shown in Figure 2. We also plot contributions of r_1, r_2, r_3 and J_1, J_3 in the figure. In this case, the contribution of r_3 is 0.

REFERENCES

[1] M. Ainsworth and J. T. Oden, *A Posteriori Error Estimation in Finite Element Analysis*, John Wiley & Sons, New York, 2000.

[2] A. Aziz, R. Kellogg, and A. Stephens, *Least-squares methods for elliptic systems*, Math. Comp., 44 (1985), pp. 53–70.

[3] P. B. Bochev and M. D. Gunzburger, *Least Squares Finite Element Methods*, Springer, Berlin, 2009.

[4] P. B. Bochev and M. D. Gunzburger, *On least-squares finite element methods for the Poisson equation and their connection to the Dirichlet and Kelvin principles*, SIAM J. Numer. Anal., 43 (2005), pp. 340–362.

[5] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer, Berlin, 2013.

[6] D. Braess, *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*, 3rd ed., Cambridge University Press, Cambridge, UK, 2007.

[7] 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), pp. 935–955.

[8] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* I, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799.

[9] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* II, SIAM J. Numer. Anal., 34 (1997), pp. 425–454.

[10] Z. Cai, T. A. Manteuffel, S. F. McCormick, and J. Ruge, *First-order system LL* (FOSLL*) for scalar partial differential equations*, SIAM J. Numer. Anal., 39 (2001), pp. 1418–1445.

[11] Z. Cai and Ku, *Optimal error estimates for the div least-squares method with data* $f \in L^2$ *and application to nonlinear problems*, SIAM J. Numer. Anal., 47 (2010), pp. 4098–4111.

[12] J. M. Cascon, R. H. Nochetto, and K. G. Siebert, *Design and convergence of AFEM in H(div)*, Math. Models Methods Appl. Sci., 17 (2007), pp. 1849–1881.

[13] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[14] D. C. Jespersen, *A least-square decomposition method for solving elliptic systems*, Math. Comp., 31 (1977), pp. 873–880.

[15] B. N. Jiang, *The Least-Squares Finite Element Method: Theory and Applications in Computational Fluid Dynamics and Electromagnectics*, Springer, Berlin, 1998.

[16] 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), pp. 1368–1377.

[17] P. A. Raviart and J. M. Thomas, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, Lecture Notes in Math. 606, I. Galligani and E. Magenes, eds., Springer-Verlag, New York, 1977, pp. 292–315.

[18] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley–Teubner Ser. Adv. Numer. Math., Wiley–Teubner, New York, 1996.

[19] R. Verfürth, *A Posteriori Error Estimation Techniques for Finite Element Methods*, Numer. Math. Sci. Comput., Oxford University Press, Oxford, UK, 2013.