SEVERAL COERCIVITY PROOFS OF FIRST-ORDER SYSTEM LEAST-SQUARES METHODS FOR SECOND-ORDER ELLIPTIC PDES

YUXIANG LIANG AND SHUN ZHANG

Abstract. In this paper, we present three versions of proofs of the coercivity for first-order system least-squares methods for second-order elliptic PDEs. The first version is based on the a priori error estimate of the PDEs, which has the weakest assumption. For the second and third proofs, a sufficient condition on the coefficients ensuring the coercivity of the standard variational formulation is assumed. The second proof is a simple direct proof and the third proof is based on a lemma introduced in the discontinuous Petrov-Galerkin method. By pointing out the advantages and limitations of different proofs, we hope that the paper will provide a guide for future proofs. As an application, we also discuss least-squares finite element methods for problems with $H^{-1}$ righthand side.

1. Introduction

The least-squares variational principle and the corresponding least-squares finite element methods based on a first-order system reformulation have been widely used in numerical solutions of partial differential equations, see for example [16, 18, 6, 32, 19, 17, 33, 22, 13]. Compared to the standard variational formulation and the related finite element methods, the first-order system least-squares finite element methods have several known advantages: (1) The least-squares continuous problem is well-posed as long as the linear PDE is well-posed. (2) The discrete problem is stable and optimal accurate as long as the discrete spaces are the subspaces of the corresponding abstract solution spaces. (3) The resulting linear systems are symmetric positive definite. Fast-solvers like multigrid can be used to solve the discrete problem. (4) The least-squares functional itself is a good error indicator/estimator for the finite element mesh refinement and the error control. (5) The boundary condition can be easily handled in the least-squares formulation with strong or weak enforcement. (6) The least-squares functional is a natural loss function for deep neuron network learning algorithms, see [12, 11].

The key to establishing the well-posedness and the a priori and a posteriori error estimates of the first-order system least-squares finite element methods is the coercivity of the corresponding bilinear forms. The simplest way uses the least-squares graph norm as the norm of choice, then the uniqueness of the PDE can be used to show the well-posedness of the least-squares system. Such analysis is too crude and often only be used for least-squares bilinear forms we have less understandings only, see for example, [38, 37, 40].

For many problems like second-order elliptic PDEs, a more refined analysis is often needed to establish the norm equivalence: the equivalence of the least-squares norm and the standard Sobolev norms of the unknowns. For the $L^2$-based first-order system least-squares method for the second-order elliptic equation with a new variable flux in $H$-(div) space, the norm equivalence of the least-squares norm and the $H$-(div) × $H^1$-norm is needed. For the $H^{-1}$-based first-order system least-squares method, the norm equivalence of the least-squares norm and the $L^2(\Omega)^d$ × $H^1$-norm is needed. The continuity of the least-squares bilinear form is often very easy to check, so the main task is to prove the coercivity of the least-squares bilinear form in the corresponding norms. For self-adjoint second-order elliptic problems with a coercive standard variational formulation, the coercivity proof is often quite simple with the help of the Poincaré inequality.

Date: January 21, 2022.

This work was supported in part by Research Grants Council of the Hong Kong SAR, China under the GRF Grant Project No. CityU 11302519, CityU 11300517, and CityU 11305319.
For the general second-order elliptic equation, there are various coercivity proofs available in the past thirty years for first-order system least-squares methods. In [16], in a quite general setting which includes a real Helmholtz problem (whose standard variational formulation is naturally non-coercive), the coercivity is established for the \( L^2 \)-based first-order system least-squares method. In the proof of [16], the key tool is a norm equivalence (3.5) in the paper of "two uniformly elliptic operators which are invertible in \( H^1 \) and have the same leading part and boundary conditions". Later in [7], in the same general setting, based on a compactness argument and the uniqueness of the PDE, the coercivity of the \( H^{-1} \)-based first-order system least-squares method is established. In Cai’s lecture notes [10], based on the [7]’s arguments, the coercivity of the \( L^2 \)-based first-order system least-squares method is established in the same general setting, see also [13]. In [33], the coercivity of the \( L^2 \)-based first-order system least-squares method is established based on the a priori error estimate of the general elliptic PDE. One way to prove the uniqueness of the PDE is the Fredholm alternative. Note that the compactness and contradiction arguments are also used in the proof of the Fredholm alternative, see for example [29]. Due to similar assumptions, the above three proofs are in some sense equivalent. One possible shortcoming of [7, 10, 13] (also the proofs of [16, 33]) is that it is a proof by contradiction. When applied to problems with mesh-dependent settings, the indirect proof can not show the coercivity constant is independent of the mesh.

In a more restrictive setting, where some assumptions on the coefficients (similar to (4.1)) are assumed to ensure the standard variational formulation is coercive, the coercivity of the \( L^2 \)-based first-order system least-squares method was proved in [39]. Such a setting excludes problems like the real Helmholtz problems. The advantage of such proof is it is a direct proof.

In recent years, there are several new ideas emerged to prove the coercivity of first-order system least-squares methods. In [24], based on a priori error estimates of two separated problems with the terms of the least-squares functionals as right-hand sides, the coercivity is established for the Helmholtz equation in a complex setting. Such an idea originally appeared in \( L^2 \)-analysis of the first-order system least-squares method [15] and analysis of discontinuous Petrov-Galerkin for Helmholtz problems [27]. The same argument is used in [4] for the self-adjoint diffusion-reaction problems. In [23], a new proof technique based on a lemma introduced in the discontinuous Petrov-Galerkin method [30] is used to prove the coercivity of specially designed least-squares finite element methods with weakly enforced boundary conditions. Compared to the standard arguments, this new technique has more flexibility when choosing test functions.

In this paper, we present three new versions of proofs for the coercivity of two formulations of the first-order system least-squares methods for second-order elliptic PDEs. One formulation is based on the diffusion flux reformulation and the other one is based on a total physical flux reformulation. Both the \( L^2 \) and \( H^{-1} \) first-order system least-squares methods are discussed. The first proof is based on the a priori error estimate of the PDEs and the proof idea from [33, 24, 4]. As [7, 10, 13, 33], it has the weakest assumption on the PDE but may suffer from indirectness. For the second and third proofs, a sufficient condition on the coefficients (4.1) ensuring the coercivity of the standard variational formulation is assumed. The second proof is a simple direct proof and the third proof is based on a lemma introduced in the discontinuous Petrov-Galerkin method [30, 23]. Although the assumption is more restrictive in the second and third proofs, they are direct proofs and have the potential to be applied to some cases that the PDE theory is not applicable directly or the proof of contradiction will only give mesh-dependent coercivity constant. By pointing out the advantages and limitations of different proofs, we hope that the paper will provide a guide for the future proofs of coercivity of least-squares methods.

As an application, we also discuss least-squares finite element methods for the second-order elliptic equation with an \( H^{-1} \) right-hand side in the paper. The key to setting up a first-order system is to choose the new vector variable carefully so it is in the \( H(\text{div}) \) space. Contrary to the usual global regularity assumption, we derive the local optimal a priori and a posteriori error estimates in the spirit of [13, 33].

The paper is organized as follows. In sub-section 1.1, we discuss notations and the function spaces, especially the \( H^{-1} \) space. In section 2 we introduce mathematical equations for the second-order scalar
elliptic partial differential equations and two variants of the first-order systems. Then we set up the corresponding $L^2$ and $H^1$ least-squares functionals and some sufficient conditions to ensure the coercivity. In section 3, the first proof based on the a priori error estimate is presented. Assumptions on the coefficients are given in section 4. The second and third proofs are presented in sections 5 and 6, respectively. In section 7, we discuss the least-squares finite element method for the elliptic equation with an $H^{-1}$ righthand side.

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

We use the standard notations and definitions for the Sobolev spaces $H^s(\Omega)^d$ for $s \geq 0$. The standard associated inner product is denoted by $(\cdot, \cdot)_{s, \Omega}$, and their respective norms are denoted by $\| \cdot \|_{s, \Omega}$ and $\| \cdot \|_{s, \partial \Omega}$ The notation $| \cdot |_{s, \Omega}$ is used for semi-norms. (We suppress the superscript $d$ because the dependence on dimension will be clear by context. We also omit the subscript $\Omega$ from the inner product and norm designation when there is no risk of confusion.) For $s = 0$, $H^s(\Omega)^d$ coincides with $L^2(\Omega)^d$. The symbols $\nabla \cdot$ and $\nabla$ stand for the divergence and gradient operators, respectively. Set $H^1_D(\Omega) := \{ v \in H^1(\Omega) : v = 0$ on $\Gamma_D \}$ and $H^1_0(\Omega) := \{ v \in H^1(\Omega) : v = 0$ on $\partial \Omega \}$. Define $H(\text{div};\Omega) = \{ \tau \in L^2(\Omega)^d : \nabla \cdot \tau \in L^2(\Omega) \}$, which is a Hilbert space under the norm $\| \tau \|_{H(\text{div};\Omega)} = (\| \nabla \tau \|^2 + \| \nabla \cdot \tau \|^2)^{1/2}$, and define its subspace $H_N(\text{div};\Omega) = \{ \tau \in H(\text{div};\Omega) : \mathbf{n} \cdot \tau = 0$ on $\Gamma_N \}$. Since most of our discussion will be in $H_N(\text{div};\Omega) \times H^1_D(\Omega)$, we introduce the notation for simplicity,

$$X := H_N(\text{div};\Omega) \times H^1_D(\Omega).$$

The space $(H^1_D(\Omega))'$ is defined as the dual space of $H^1_D(\Omega)$ and consists of the functionals $v$ for which the norm

$$(1.1) \quad \| v \|_{-1} = \sup_{w \in H^1_D(\Omega)} \frac{(v, w)_{(H^1_D(\Omega))' \times H^1_D(\Omega)}}{\| \nabla w \|_0},$$

is finite. When $\partial \Omega = \Gamma_D$, then it is the usual $H^{-1}(\Omega)$, see for example, discussions in [29]. Here, we use $\| \nabla \cdot \|_0$ instead of the standard $\| \cdot \|_1$ norm due to the Poincaré inequality. The space $(H^1_D(\Omega))'$ is a Hilbert space. Let $S : (H^1_D(\Omega))' \to H^1_D(\Omega)$ to be the solution operator of the following problem:

$$(1.3) \quad z \in H^1_D(\Omega) \quad \text{with} \quad (\nabla z, \nabla v) = (f, v) \quad \forall v \in H^1_D(\Omega).$$

That is, for $f \in (H^1_D(\Omega))'$, $Sf = z \in H^1_D(\Omega)$ is the solution of (1.3). Again, we modify the setting in [7] and use the Poisson problem in weak form only. In [7], the PDE version of $-\Delta u + u = f$ with boundary conditions is used. It is easy to see that the two settings are equivalent. We have the following result (Lemma 2.1 of [7]): The inner product on $(H^1_D(\Omega))' \times (H^1_D(\Omega))'$ is given by

$$(1.4) \quad (v, q) \quad \forall v, q \in (H^1_D(\Omega))'.$$

We have $\| q \|^2_{-1} = (q, Sq)$, for $q \in (H^1_D(\Omega))'$. Let $f = S^{-1}z$ and $v = z$ in (1.3), we have

$$(1.5) \quad \| \nabla z \|^2_0 = (S^{-1}z, z), \quad \forall z \in H^1_D(\Omega).$$

Thus, for $v \in H^1_D(\Omega)$ and $w = S^{-1}v \in (H^1_D(\Omega))'$, we have

$$(1.6) \quad \| w \|_{-1}^2 = (w, Sw) = (S^{-1}v, SS^{-1}v) = (S^{-1}v, v) = \| \nabla v \|^2_0.$$
For $\mathbf{\tau} \in H_N(\text{div}; \Omega)$, we also have
\begin{equation}
\|\nabla \cdot \mathbf{\tau}\|_1 = \sup_{\mathbf{w} \in H_0^1(\Omega)} \frac{(\nabla \cdot \mathbf{\tau}, \mathbf{w})}{\|\nabla \mathbf{w}\|_0} = \sup_{\mathbf{w} \in H_0^1(\Omega)} \frac{(\mathbf{\tau}, \nabla \mathbf{w})}{\|\nabla \mathbf{w}\|_0} \leq \|\mathbf{\tau}\|_0.
\end{equation}

2. General second-order elliptic equations and their least-squares methods

Consider the general second-order elliptic equation
\begin{equation}
-\nabla \cdot (A\nabla u) + b \cdot \nabla u + cu = f \text{ in } \Omega,
\end{equation}
(2.1)
\begin{equation*}
u = 0 \text{ on } \Gamma_D,
\end{equation*}
\begin{equation*}
A\nabla u \cdot \mathbf{n} = 0 \text{ on } \Gamma_N,
\end{equation*}
and its adjoint problem
\begin{equation}
-\nabla \cdot (A\nabla u) - \nabla \cdot (bu) + cu = f \text{ in } \Omega,
\end{equation}
(2.2)
\begin{equation*}
u = 0 \text{ on } \Gamma_D,
\end{equation*}
\begin{equation*}
(A\nabla u + bu) \cdot \mathbf{n} = 0 \text{ on } \Gamma_N.
\end{equation*}
The diffusion coefficient matrix $A \in L^\infty(\Omega)^{d \times d}$ is a given $d \times d$ tensor-valued function; $b \in L^\infty(\Omega)^d$ and $c \in L^\infty(\Omega)$ are given vector- and scalar-valued functions, respectively; and $f \in L^2(\Omega)$ is a given scalar function. Assume that $A$ is uniformly symmetric positive definite: there exist positive constants $0 < \Lambda_0 \leq \Lambda_1$ such that
\begin{equation}
\Lambda_0 \mathbf{y}^T \mathbf{y} \leq \mathbf{y}^T A \mathbf{y} \leq \Lambda_1 \mathbf{y}^T \mathbf{y}
\end{equation}
for all $\mathbf{y} \in \mathbb{R}^d$ and almost all $x \in \Omega$.

The variational problem of (2.1) is: Find $u \in H_0^1(\Omega)$, such that,
\begin{equation}
(A\nabla u, \nabla v) + (b \cdot \nabla u + cu, v) = (f, v) \quad \forall v \in H_0^1(\Omega).
\end{equation}
And the variational problem of (2.2) is: Find $u \in H_D^1(\Omega)$, such that,
\begin{equation}
(A\nabla u, \nabla v) + (u, b \cdot \nabla v + cv) = (f, v) \quad \forall v \in H_D^1(\Omega).
\end{equation}
It is easy to see that (2.2) is the adjoint problem of (2.1).

For the equation (2.1), let the flux $\mathbf{\sigma} = -A\nabla u$, and define
\begin{equation}
Y_{st}(\mathbf{\tau}, v) = b \cdot (s\nabla v - tA^{-1}\mathbf{\tau}) + cv,
\end{equation}
where $0 \leq t \leq 1$ and $s + t = 1$. We have the first-order system:
\begin{equation}
\begin{cases}
\mathbf{\sigma} + A\nabla u = 0 & \text{in } \Omega, \\
\nabla \cdot \mathbf{\sigma} + Y_{st}(\mathbf{\sigma}, u) = f & \text{in } \Omega, \\
u = 0 & \text{on } \Gamma_D, \\
\mathbf{\sigma} \cdot \mathbf{n} = 0 & \text{on } \Gamma_N.
\end{cases}
\end{equation}

For $u \in H_D^1(\Omega)$, we have $\mathbf{\sigma} = -A\nabla u \in L^2(\Omega)^d$ and $\nabla \cdot \mathbf{\sigma} = f - Y_{st}(\mathbf{\sigma}, u) \in L^2(\Omega)$, so $(\mathbf{\sigma}, u) \in X = H_N(\text{div}; \Omega) \times H_D^1(\Omega)$.

For $(\mathbf{\tau}, v) \in X$, define least-squares functionals for the system (2.7),
\begin{equation}
L(\mathbf{\tau}, v; f) := \|A^{-1/2}\mathbf{\tau} + A^{1/2}\nabla v\|^2_0 + \|\nabla \cdot \mathbf{\tau} + Y_{st}(\mathbf{\tau}, v) - f\|^2_0,
\end{equation}
\begin{equation}
L_{-1}(\mathbf{\tau}, v; f) := \|A^{-1/2}\mathbf{\tau} + A^{1/2}\nabla v\|^2_0 + \|\nabla \cdot \mathbf{\tau} + Y_{st}(\mathbf{\tau}, v) - f\|^2_1.
\end{equation}
The corresponding least-squares minimization problems are:
\begin{equation}
\text{Find } (\mathbf{\sigma}, u) \in X \quad \text{s.t.} \quad L(\mathbf{\sigma}, u; f) = \inf_{(\mathbf{\tau}, v) \in X} L(\mathbf{\tau}, v; f)
\end{equation}
\begin{equation}
\text{and find } (\mathbf{\sigma}, u) \in X \quad \text{s.t.} \quad L_{-1}(\mathbf{\sigma}, u; f) = \inf_{(\mathbf{\tau}, v) \in X} L_{-1}(\mathbf{\tau}, v; f).
\end{equation}
For the equation (2.2), define the total flux as \( \sigma = -A\nabla u - bu \), then

\[
\begin{aligned}
\sigma + A\nabla u + bu &= 0 \quad \text{in } \Omega, \\
\nabla \cdot \sigma + cu &= f \quad \text{in } \Omega, \\
u &= 0 \quad \text{on } \Gamma_D, \\
\sigma \cdot n &= 0 \quad \text{on } \Gamma_N.
\end{aligned}
\]

(2.12)

Similarly, we have \((\sigma, u) \in X\).

For \((\tau, v) \in X\), define least-squares functionals for the system (2.12),

\[
\begin{aligned}
J(\tau, v; f) &= \|\tau + A\nabla v + bv\|^2_0 + \|\nabla \cdot \tau + cv - f\|^2_0, \\
J_{-1}(\tau, v; f) &= \|\tau + A\nabla v + bv\|^2_0 + \|\nabla \cdot \tau + cv - f\|^2_1.
\end{aligned}
\]

(2.13) (2.14)

Remark 2.1. We discuss the problem (2.2) independently since we can choose the total physical flux which is different from (2.1) and has many applications. The introduction of the total flux in the least-squares context can be found in [34].

If only the boundary condition \(A\nabla u \cdot n = 0\) is given for (2.2), and we still choose \(\sigma = -A\nabla u - bu\), the condition \(A\nabla u \cdot n = 0\) can also be viewed as a Robin boundary condition \(\sigma \cdot n + b \cdot nu = 0\). The discussion of FOSLS for elliptic problems with Robin boundary condition can be found in [35].

For any \((\tau, v) \in L^2(\Omega)^d \times H^1_D(\Omega)\) and \((\rho, w) \in X\), define the following two norms respectively:

\[
\|(\tau, v)\|^2 := \|\tau\|_0^2 + \|\nabla v\|_0^2
\]

and

\[
\|(\rho, w)\|^2 := \|\rho\|_0^2 + \|\nabla \cdot \rho\|_0^2 + \|\nabla w\|_0^2.
\]

(2.15)

Remark 2.2. By the Poincaré inequality, we can also replace \(\|\nabla v\|_0\) by a standard \(\|v\|_1\) in the above norms. We use the norm \(\|\nabla v\|_0\) for simplicity.

From the Cauchy-Schwarz, triangle, and Poincaré inequalities and the bounds of the coefficients \((A, b,\) and \(c)\) of the underlying problems, we immediately have the following up-bounds: There exists \(C > 0\), such that,

\[
L(\tau, v; 0) \leq C\|(\tau, v)\|^2, \quad J(\tau, v; 0) \leq C\|(\tau, v)\|^2, \quad \forall (\tau, v) \in X.
\]

(2.16)

Together with (1.7) and (1.8), we can show that there exists \(C > 0\), such that,

\[
L_{-1}(\tau, v; 0) \leq C\|(\tau, v)\|^2, \quad J_{-1}(\tau, v; 0) \leq C\|(\tau, v)\|^2, \quad \forall (\tau, v) \in X.
\]

(2.17)

In this paper, we want to prove the following coercivity: For all \((\tau, v) \in X\), there exists \(C > 0\), such that,

\[
C\|(\tau, v)\|^2 \leq L(\tau, v; 0), \quad C\|(\tau, v)\|^2 \leq L_{-1}(\tau, v; 0),
\]

(2.18) (2.19)

Remark 2.3. If both the coercivity results (2.18) and (2.19) hold, then we have the following equivalence:

\[
C_1L(\tau, v; 0) \leq J(\tau, v; 0) \leq C_2L(\tau, v; 0) \quad (\tau, v) \in X;
\]

(2.20) (2.21)

Before discussing the full coercivity proofs, we introduce some conditions that ensure coercivity. These conditions are elementary and well-known, we list them here to avoid repeating in future proofs.

Theorem 2.4. For \((\tau, v) \in X\), assume that

\[
C\|\nabla v\|_0 \leq \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_0 + \|\tau + A\nabla v\|_0 \text{ or } C\|\tau\|_0 \leq \|\nabla \cdot \tau + X_{st}(\tau, v)\|_0 + \|\tau + A\nabla v\|_0
\]

is true, then the following coercivity is true,

\[
C\|(\tau, v)\|^2 \leq L(\tau, v; 0) \quad \forall (\tau, v) \in X.
\]

(2.22) (2.23)
Proof. From the triangle and Poincaré inequalities, we have
\[ \| \nabla \cdot \tau \|_0 \leq \| \nabla \cdot \tau + Y_{st}(\tau, v) \|_0 + \| Y_{st}(\tau, v) \|_0 \leq \| \nabla \cdot \tau + Y_{st}(\tau, v) \|_0 + C\| \nabla v \|_0 + C\| \tau \|_0, \]
and \( \| \tau \|_0 \leq \| \tau + A\nabla v \|_0 + C\| \nabla v \|_0. \)
Thus, if the first inequality of (2.22) is true, we have coercivity.

On the other hand, \( C\| \nabla v \|_0 \leq \| \tau + A\nabla v \|_0 + \| \tau \|_0, \)
thus if the second inequality of (2.22) is true, we have the desired coercivity. \( \square \)

We have the similar result for the \( J \) functional.

**Theorem 2.5.** Assume that
\[ (2.24) \quad C\| \nabla v \|_0 \leq \| \nabla \cdot \tau + cv \|_0 + \| \tau + A\nabla v + bv \|_0 \quad \forall (\tau, v) \in X \]
holds, then we have coercivity,
\[ (2.25) \quad C\| (\tau, v) \|_2^2 \leq J(\tau, v; 0) \quad \forall (\tau, v) \in X. \]
Proof. From the triangle and Poincaré inequalities, we get
\[ \| \nabla \cdot \tau \|_0 \leq \| \nabla \cdot \tau + cv \|_0 + \| cv \|_0 \leq \| \nabla \cdot \tau + cv \|_0 + C\| \nabla v \|_0 \]
and \( \| \tau \|_0 \leq \| \tau + A\nabla v + bv \|_0 + C\| \nabla v \|_0. \)
Thus, if (2.24) is true, we have coercivity. \( \square \)

With the help of triangle and Poincaré inequalities, we can immediately prove the following theorem in a similar fashion.

**Theorem 2.6.** For \( (\tau, v) \in X \), assume that
\[ (2.26) \quad C\| \nabla v \|_0 \leq \| \nabla \cdot \tau + Y_{st}(\tau, v) \|_{-1} + \| \tau + A\nabla v \|_0 \]
is true, then we have coercivity,
\[ (2.27) \quad C\| (\tau, v) \|_2^2 \leq L_{-1}(\tau, v; 0) \quad \forall (\tau, v) \in X. \]
For \( (\tau, v) \in X \), assume that
\[ (2.28) \quad C\| \nabla v \|_0 \leq \| \nabla \cdot \tau + cv \|_{-1} + \| \tau + A\nabla v + bv \|_0 \]
holds, then we have coercivity,
\[ (2.29) \quad C\| (\tau, v) \|_2^2 \leq J_{-1}(\tau, v; 0) \quad \forall (\tau, v) \in X. \]

3. **Proof I: Proofs Based on A Priori Error Estimate**

In this section, we assume the following a priori estimates for both versions of equations (2.1) and (2.2) with an \( H^{-1} \) righthand side: For a \( g \in (H^1_D(\Omega))' \), assume that \( u \in H^1_D(\Omega) \) is the solution of the weak problem
\[ (A\nabla u, \nabla v) + (b \cdot \nabla u + cu, v) = \langle g, v \rangle_{(H^1_D(\Omega))' \times H^1_D(\Omega)} \quad \forall v \in H^1_D(\Omega). \]
or the adjoint weak problem
\[ (A\nabla u, \nabla v) + (u, b \cdot \nabla v + cv) = \langle g, v \rangle_{(H^1_D(\Omega))' \times H^1_D(\Omega)} \quad \forall v \in H^1_D(\Omega). \]
We assume the following a priori error estimate holds:
\[ (3.3) \quad \| \nabla u \|_0 \leq C\| g \|_{-1} \quad \forall g \in (H^1_D(\Omega))'. \]
We list two important cases. The first one is that
\[ (3.4) \quad \langle g, v \rangle_{(H^1_D(\Omega))' \times H^1_D(\Omega)} := (f_1, v) \]
for some $f_1 \in L^2(\Omega)$. It is obvious that $f_1 \in (H^1_0(\Omega))'$, then by (3.3) and (1.2), we have
\begin{equation}
(3.5) \quad \|\nabla u\|_0 \leq C\|f_1\|_{-1} \leq C\|f_1\|_0.
\end{equation}
The second one is that
\begin{equation}
(3.6) \quad \langle g, v \rangle_{(H^1_0(\Omega))' \times H^1_0(\Omega)} := \langle f_2, \nabla v \rangle
\end{equation}
for some $f_2 \in L^2(\Omega)^d$. Then by (3.3) and (1.2), we have
\begin{equation}
(3.7) \quad \|\nabla u\|_0 \leq C\|g\|_{-1} = C \sup_{v \in H^1_0(\Omega)} \frac{\langle f_2, \nabla v \rangle}{\|\nabla v\|_0} \leq C\|f_2\|_0.
\end{equation}
Note that, for the second case and $\Gamma_D = \partial \Omega$, we have $(H^1_0(\Omega))' = H^{-1}(\Omega)$, $g$ can be (at least formally) written as $-\nabla \cdot f_2$, see the discussion in [29, 9]. For the general case $\Gamma_D \neq \partial \Omega$, we can not write $g$ as $-\nabla \cdot f_2$, thus the weak forms (2.4) and (3.2) instead of the strong PDE forms are preferred in our presentations.

The a priori error estimate for the $\Gamma_D = \partial \Omega$ case can be found in standard PDE books, for example, [29]. For the mixed boundary condition case, a proof can be found in [3] for an even more general boundary condition. In the proof of [4], the proof of contradiction is used with the compactness argument and the uniqueness assumption of the solution. Another possible way to prove the a priori error estimate is the Fredholm alternative where the compactness and contradiction arguments are also used, see [29].

The assumption (3.3) is very general. The a priori estimate is also true for Helmholtz equations with a unique solution, where the usual coercivity of the standard variational problem does not hold.

Before we dive into the proof, we first present a priori error estimates for four auxiliary problems.

**Problem 1.** Let $f \in L^2(\Omega)^d$. Consider the following first-order system
\begin{equation}
(3.8) \quad \begin{cases}
\mathbf{r} + A\nabla w &= f \quad \text{in } \Omega, \\
\nabla \cdot \mathbf{r} + Y_{st}(\mathbf{r}, w) &= 0 \quad \text{in } \Omega, \\
w &= 0 \quad \text{on } \Gamma_D, \\
\mathbf{r} \cdot \mathbf{n} &= 0 \quad \text{on } \Gamma_N.
\end{cases}
\end{equation}

**Lemma 3.1.** Assuming the a priori error estimate (3.3) for (3.1) holds, the following a priori error estimates for (3.8) are true:
\begin{equation}
(3.9) \quad \|\nabla u\|_0 \leq C\|f\|_0 \quad \text{and} \quad \|\mathbf{r}\|_0 \leq C\|f\|_0.
\end{equation}

**Proof.** First, we construct a unique set of solutions for (3.8). Let $w \in H^1_D(\Omega)$ be the unique solution of
\begin{equation}
(3.10) \quad (A\nabla w, \nabla v) + (\mathbf{b} \cdot \nabla w + cw, v) = (\mathbf{b} \cdot A^{-1}f, v) + (f, \nabla v), \quad \forall v \in H^1_D(\Omega).
\end{equation}
The uniqueness of the solution $w$ of (3.10) is guaranteed by the assumption (3.3). Let $\mathbf{r} = f - A\nabla w$, which is the first equation of the system (3.8), and substitute it into (3.10), we have
\begin{equation}
(3.11) \quad - (\mathbf{r}, \nabla v) + (Y_{st}(\mathbf{r}, w), v) = 0 \quad \forall v \in H^1_D(\Omega).
\end{equation}
Following the arguments in [3], picking $v \in C_0^\infty(\Omega)$ and integrating by parts on the term $(\mathbf{r}, \nabla v)$, we get the second equation of the system (3.8). For any $v \in H^1_D(\Omega)$,
\begin{equation}
(\mathbf{r} \cdot \mathbf{n}, v)_{H^{-1/2}(\Gamma_N) \times H^{1/2}(\Gamma_N)} = (\mathbf{r}, \nabla v) + (\nabla \cdot \mathbf{r}, v) = (\nabla \cdot \mathbf{r} + Y_{st}(\mathbf{r}, w), v) = 0,
\end{equation}
thus the boundary condition of $\mathbf{r}$ of the system (3.8) is satisfied. Thus, we have completed our construction of the solutions of (3.8).

For the equation (3.10), based on our a priori assumptions (3.5) and (3.7), we have the following a priori error estimate,
\begin{equation}
\|\nabla w\|_0 \leq C\|f\|_0 \quad \text{and} \quad \|\mathbf{r}\|_0 \leq \|A\nabla w\|_0 + \|f\|_0 \leq C\|f\|_0.
\end{equation}

The lemma is proved. \qed
Problem 2. For $g \in L^2(\Omega)$, consider the following system

\begin{equation}
\begin{aligned}
s + A\nabla z &= 0, \quad \text{in } \Omega, \\
\nabla \cdot s + Y_A(s, z) &= g, \quad \text{in } \Omega, \\
z &= 0, \quad \text{on } \Gamma_D, \\
s \cdot n &= 0, \quad \text{on } \Gamma_N.
\end{aligned}
\end{equation}

(3.12)

Lemma 3.2. Assuming that the a priori error estimate (3.3) for (3.1) holds, the following a priori error estimates for (3.12) are true:

\begin{equation}
\|s\|_0 \leq C\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0.
\end{equation}

(3.13)

Proof. Similar to Problem 1, we construct solutions of (3.12) first. Let $s \in H^1_D(\Omega)$ be the unique solution of the following problem:

\begin{equation}
(A\nabla z, \nabla v) + (b \cdot \nabla z + cz, v) = (g, v) \quad \forall v \in H^1_D(\Omega).
\end{equation}

(3.14)

Let $s = -A\nabla z$ and substitute it into (3.14), and use the same arguments as in Problem 1, we get the second equation and the boundary condition of $s$ of (3.12). Thus, we have constructed $z$ and $s$ to be the solution of (3.12).

For the equation (3.14), based on the assumption (3.3), we have the following a priori error estimate,

\[\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0 \quad \text{and} \quad \|s\|_0 = \|A\nabla z\|_0 \leq C\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0.\]

The a priori error estimate (3.13) for (3.12) is now proved. \hfill \Box

Problem 3. Let $f \in L^2(\Omega)$. Consider the following system

\begin{equation}
\begin{aligned}
r + A\nabla w + bw &= f, \quad \text{in } \Omega, \\
\nabla \cdot r + cw &= 0, \quad \text{in } \Omega, \\
w &= 0, \quad \text{on } \Gamma_D, \\
r \cdot n &= 0, \quad \text{on } \Gamma_N.
\end{aligned}
\end{equation}

(3.15)

Lemma 3.3. Assuming that the a priori error estimate (3.3) for (3.2) holds, the following a priori error estimates for (3.15) are true:

\begin{equation}
\|\nabla w\|_0 \leq C\|f\|_0 \quad \text{and} \quad \|r\|_0 \leq C\|f\|_0.
\end{equation}

(3.16)

Proof. Let $w \in H^1_D(\Omega)$ be the unique solution of the following problem:

\begin{equation}
(A\nabla w, \nabla v) + (bw, \nabla v) + (cw, v) = (f, \nabla v) \quad \forall v \in H^1_D(\Omega).
\end{equation}

(3.17)

Then let $r = f - A\nabla w - bw$ and substitute it into (3.17), we have

\begin{equation}
-(r, \nabla v) + (cw, v) = 0.
\end{equation}

(3.18)

Use the same arguments as in the Problem 1, we get the second equation and the boundary condition of $r$ of (3.15).

For the equation (3.17), based on the assumption (3.7), we have the following a priori error estimate,

\[\|\nabla w\|_0 \leq C\|f\|_0 \quad \text{and} \quad \|r\|_0 \leq \|f\|_0 + \|A\nabla w\|_0 + \|bw\|_0 \leq C\|f\|_0.\]

The proof of the lemma is completed. \hfill \Box

Problem 4. For $g \in L^2(\Omega)$, consider the following system

\begin{equation}
\begin{aligned}
s + A\nabla z + bz &= 0, \quad \text{in } \Omega, \\
\nabla \cdot s + cz &= g, \quad \text{in } \Omega, \\
z &= 0, \quad \text{on } \Gamma_D, \\
s \cdot n &= 0, \quad \text{on } \Gamma_N.
\end{aligned}
\end{equation}

(3.19)
Lemma 3.4. Assuming the a priori error estimate (3.3) for (3.1), the following a priori error estimates for (3.19) are true:

\[(3.20)\] \[\|s\|_0 \leq C\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0.\]

Proof. Let \(z \in H^1_D(\Omega)\) be the unique solution of the following problem:

\[(3.21)\] \[(Av, \nabla v) + (bz, \nabla v) + (cz, v) = (g, v) \quad \forall v \in H^1_D(\Omega).\]

Let \(s = -A\nabla z - bz\), substitute into (3.21), and use the same argument as in Problem 1, we get the second equation and the boundary condition for \(s\). For the equation (3.22), based on the assumption (3.5), we have the following a priori error estimate:

\[(3.22)\] \[\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0.\]

and by the triangle and Poincaré inequalities, we get

\[(3.23)\] \[\|s\|_0 \leq \|A\nabla z\|_0 + \|bz\|_0 \leq C\|\nabla z\|_0 \leq C\|g\|_{-1} \leq C\|g\|_0.\]

Now, we are in a position to prove the coercivity results.

Theorem 3.5. Assuming the a priori error estimate (3.3) for (3.1) is true, the coercivity results in (2.18) hold.

Proof. For \((\tau, v) \in X\), let

\[f = \tau + A\nabla v \quad \text{and} \quad g = \nabla \cdot \tau + Y_s(\tau, v)\]

for Problems 1 (3.3) and 2 (3.12), respectively. By the linearity of Problems 1 and 2, we have \(\tau = r + s\) and \(v = w + z\). By the a priori error estimates (3.9) and (3.13), we have

\[(3.24)\] \[C\|\nabla v\|_0 \leq \|\tau + A\nabla v\|_0 + \|\nabla \cdot \tau + Y_s(\tau, v)\|_{-1}, \quad C\|\nabla v\|_0 \leq \|\tau + A\nabla v\|_0 + \|\nabla \cdot \tau + Y_s(\tau, v)\|_0\]

and

\[(3.25)\] \[C\|\tau\|_0 \leq \|\tau + A\nabla v\|_0 + \|\nabla \cdot \tau + Y_s(\tau, v)\|_{-1}, \quad C\|\tau\|_0 \leq \|\tau + A\nabla v\|_0 + \|\nabla \cdot \tau + Y_s(\tau, v)\|_0.\]

Thus (2.22) of Theorem 2.4 is satisfied, the coercivity results in (2.18) are proved. □

Theorem 3.6. Assuming that the a priori error estimate (3.3) for (3.2) is true, the coercivity results in (2.19) hold.

Proof. For \((\tau, v) \in X\), let

\[f = \tau + A\nabla v + bv \quad \text{and} \quad g = \nabla \cdot \tau + cv\]

for Problems 3 (3.16) and 4 (3.19), respectively. By the linearity of Problems 3 and 4, we have \(\tau = r + s\) and \(v = w + z\). Then by the a priori error estimates (3.10) and (3.20),

\[(3.26)\] \[C\|\nabla v\|_0 \leq \|\tau + A\nabla v + bv\|_0 + \|\nabla \cdot \tau + cv\|_{-1} \leq \|\tau + A\nabla v + bv\|_0 + C\|\nabla \cdot \tau + cv\|_0.\]

Thus (2.24) of Theorem 2.5 is satisfied and the coercivity results in (2.19) are proved. □

Remark 3.7. As pointed out in [33], the a priori estimate (3.3) is equivalent to the coercivity results of the \(L\) and \(J\) (for \(\|u\|_0 \leq C\|f\|_0\)), and \(L_1\) and \(J_1\) (for \(\|u\|_0 \leq C\|f\|_{-1}\)). It can be easily proved by letting \(\tau = -A\nabla v\) and \(s = 0\) for \(L\) and \(L_1\) and letting \(\tau = -A\nabla v - bv\) for \(J\) and \(J_1\).

One of the restrictions of the above proof (also the proofs in [16, 7, 10, 13, 33]) is that when the discrete space is non-conforming or the least-squares functional is mesh-size dependent, then it is not applicable, see our discussion for the Crouzeix-Raviart non-conforming least-squares finite element [36].
4. SOME ASSUMPTIONS ON THE COEFFICIENTS

We assume \( b \in W^{1,\infty}(\Omega) \) so \( \nabla \cdot b \) is well-defined, and assume that

\[
e - \frac{1}{2} \nabla \cdot b \geq 0 \text{ in } \Omega \quad \text{and} \quad b \cdot n \geq 0 \text{ on } \Gamma_N.
\]

The last condition means that the outflow boundary is a part of the Neumann boundary. We have the following result from the integration by parts:

\[
(b \cdot \nabla v, v) = -(v, \nabla \cdot (bv)) + (b \cdot n, v^2)_{\Gamma_N} = -(v, (\nabla \cdot b)v) - (b \cdot \nabla v, v) + (b \cdot n, v^2)_{\Gamma_N}.
\]

Thus

\[
(b \cdot \nabla v, v) = -\frac{1}{2}(\nabla \cdot b, v^2) + \frac{1}{2}(b \cdot n, v^2)_{\Gamma_N}.
\]

Then, for \( v \in H^1_D(\Omega) \) and coefficients satisfying (4.1), we have

\[
(cv, v) + (b \cdot \nabla v, v) = ((c - \frac{1}{2} \nabla \cdot b), v^2) + \frac{1}{2}(b \cdot n, v^2)_{\Gamma_N} \geq 0,
\]

and (4.3)

\[
(Y_{st}(\tau, v), v) \geq -t((A^{-1} \tau + \nabla v), bv).
\]

Let \( u = v \) for both weak problems (2.4) and (2.5), we have

\[
(A\nabla v, \nabla v) + (b \cdot \nabla v + cv, v) = (A\nabla v, \nabla v) + ((c - \frac{1}{2} \nabla \cdot b), v^2) + \frac{1}{2}(b \cdot n, v^2)_{\Gamma_N} \geq \|A^{1/2}\nabla v\|_0^2.
\]

Thus, by the Lax-Milgram Theorem, we have the existence and uniqueness of the solution for both (2.4) and (2.2) under the assumption of (4.1) with a direct proof.

5. PROOF II: SIMPLE DIRECT PROOFS WITH ASSUMPTIONS ON COEFFICIENTS

In this section, we present simple and short direct proofs based on the assumptions on coefficients (4.1).

Theorem 5.1. Assuming the coefficients satisfying the condition \((4.1)\) for \((2.4)\), the coercivity results in \((2.18)\) hold.

Proof. It follows from integration by parts, the Cauchy-Schwarz inequality, the Poincaré inequality, assumption on \(A, (1.7)\), and (1.3),

\[
C\|\nabla v\|_0^2 \leq \|A^{1/2}\nabla v\|_0^2 = (A\nabla v + \tau, \nabla v) - (\tau, \nabla v) = (A\nabla v + \tau, \nabla v) - (\nabla \cdot \tau, v)
\]

\[
= (A\nabla v + \tau, \nabla v) + (\nabla \cdot \tau + Y_{st}(\tau, v), v) - (Y_{st}(\tau, v), v)
\]

\[
\leq (A\nabla v + \tau, \nabla v) + (\nabla \cdot \tau + Y_{st}(\tau, v), v) + |t((A^{-1} \tau + \nabla v), bv)|
\]

\[
\leq C\|A^{1/2}\nabla v + A^{-1/2}\tau\|_0\|\nabla v\|_0 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_{-1}\|\nabla v\|_0
\]

\[
\leq C\|A^{1/2}\nabla v + A^{-1/2}\tau\|_0 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_0\|\nabla v\|_0.
\]

We then have

\[
(5.1) \quad C\|\nabla v\|_0 \leq \|A^{1/2}\nabla v + A^{-1/2}\tau\|_0 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_{-1} \leq \|A^{1/2}\nabla v + A^{-1/2}\tau\|_0 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_0.
\]

Thus (2.22) of Theorem 2.4 is satisfied, the coercivity results in (2.18) are proved. □

Theorem 5.2. Assuming the coefficients satisfying the condition \(4.1\) for \(2.2\), the coercivity results in \(2.19\) hold.
Proof. It follows from integration by parts, the Cauchy-Schwarz inequality, the Poicaré inequality, assumption on $A$, (1.7), and (1.2),
\[
C\|\nabla v\|_0^2 \leq \|A^{1/2}\nabla v\|_0^2 = (A\nabla v + \tau +bv, \nabla v) - (v, \nabla v) - (b, \nabla v)
\]
\[
= (A\nabla v + \tau +bv, \nabla v) + (\nabla \cdot \tau, v) + \frac{1}{2}(\nabla \cdot b, v^2) - \frac{1}{2}(b \cdot n, v^2)_{\Gamma_N}.
\]
\[
= (A\nabla v + \tau +bv, \nabla v) + (\nabla \cdot \tau + cv, v) - (c - \nabla \cdot b/2, v^2) - \frac{1}{2}(b \cdot n, v^2)_{\Gamma_N}.
\]
\[
\leq (A\nabla v + \tau +bv, \nabla v) + (\nabla \cdot \tau + cv, v)
\]
\[
\leq \|A\nabla v + \tau +bv\|_0\|\nabla v\|_0 + \|\nabla \cdot \tau + cv\|_{-1}\|\nabla v\|_0
\]
\[
\leq C(\|A\nabla v + \tau +bv\|_0 + \|\nabla \cdot \tau + cv\|_0)\|\nabla v\|_0.
\]

We then have
\[
(5.2) \quad C\|\nabla v\|_0 \leq \|A\nabla v + \tau -bv\|_0 + \|\nabla \cdot \tau + cv\|_{-1} \leq \|A\nabla v + \tau -bv\|_0 + \|\nabla \cdot \tau + cv\|_0.
\]

Thus (2.24) of Theorem 2.5 is satisfied, the coercivity results in (2.19) are proved. ∎

6. Proof III: Proofs Based on an Equivalence Theorem and Assumptions on Coefficients

The major tool in the proof of this section is an equivalence theorem from the discontinuous Petrov-Galerkin method [30]. Theorem 6.1 can be found in Theorem 2.1 of [30] and Theorem 3.1 of [23]. The first two equalities in (6.1) are trivial and the proof of the last equality in (6.1) is based on a simple fact that the last two supremums in the theorem are achieved for $v = Tw$.

**Theorem 6.1. An Equivalence Theorem.** Let $U$ be a normed linear space, and $V$ be a Hilbert space with associate bilinear form $(\cdot , \cdot)_V$ and norm $\|v\|_V$. The operator $T$ is a linear map from $U$ to $V$. Then, for any $w \in U$
\[
(6.1) \quad \|Tw\|_V = \sup_{z \in U} \left( \frac{(Tw, z)_V}{\|z\|_V} \right) = \sup_{v \in T(U)} \left( \frac{(Tw, v)_V}{\|v\|_V} \right) = \sup_{v \in V} \left( \frac{(Tw, v)_V}{\|v\|_V} \right).
\]

In most first-order least-squares methods, we want to prove
\[
C\|w\|_U \leq \|Tw\|_V \quad \forall w \in U,
\]
for the bilinear form $(Tw, z)_V$ with $C > 0$. Theorem 6.1 enables us to choose the test function in a larger space $V$ instead of $z \in U$ or $v \in T(U)$. In most $L^2$-based first-order least-squares methods, $V$ is a simple $L^2$-space on some domain or some domain’s boundary. The proof usually follows a simple procedure: we pick a $v \in V$ such that $\|v\|_V \leq C\|w\|_U$ and $(Tw, v)_V \geq C\|w\|_V^2$, then $\|Tw\|_V \geq C\|w\|_V^2 \geq C\|w\|_U$.

**Theorem 6.2.** Assuming the coefficients satisfying the condition (4.1) for (2.1), the $L^2$ version of the coercivity result in (2.18) holds.

Proof. Let $U = X$. We choose $V$ to be simple $L^2$ spaces $V = L^2(\Omega)^d \times L^2(\Omega)$. The operator $T$ from $U$ to $V$ is defined as:
\[
T(\tau, v) = \left( A^{-1/2}\tau + A^{1/2}\nabla v \right).
\]
It follows that $\|T(\tau, v)\|_0^2 = \|A^{-1/2}\tau + A^{1/2}\nabla v\|_0^2 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_0^2$. From Theorem 6.1
\[
(6.2) \quad \|T(\tau, v)\|_0 = \sup_{(\tau, w) \in L^2(\Omega)^d \times L^2(\Omega)} \left( \frac{T(\tau, v), (\tau, w)}{\|(\tau, w)\|_0} \right).
\]
Let
\[
r_1 = A^{-1/2}\tau + tA^{-1/2}b, \quad r_2 = A^{1/2}\nabla v + tA^{-1/2}b, \quad r = r_1 + r_2, \quad \text{and} \quad w = v.
\]
Then by integration by parts and (13),

\[
(T(\tau, v), (r_1, w)) = (A^{-1}\tau + \nabla v, \tau + t bv) + (\nabla \cdot \tau + Y_{st}(\tau, v), v) \\
\geq (A^{-1}\tau + \nabla v, \tau + t((A^{-1}\tau + \nabla v), bv)) + (\nabla \cdot \tau, v) - t((A^{-1}\tau + \nabla v, bv) \\
\geq C\|\tau\|_0^2.
\]

and

\[
(T(\tau, v), (r_2, w)) = (A^{-1}\tau + \nabla v, A\nabla v + tbv) + (\nabla \cdot \tau + Y_{st}(\tau, v), v) \\
\geq (A\nabla v, \nabla v) + (\nabla, \tau + t((A^{-1}\tau + \nabla v, bv)) + (\nabla \cdot \tau, v) - t((A^{-1}\tau + \nabla v, bv) \\
\geq C\|\tau\|^2_0.
\]

We also have \(|r|_0 = |r_1 + r_2|_0 \leq C(\|\nabla v\|_0 + |\tau|_0)\) and \(|v|_0 \leq C\|\nabla v\|_0\). Thus

(6.3) \(\big|T(\tau, v)\big|_0^2 \geq (T(\tau, v), (r_1 + r_2, w)) \geq C(\|\nabla v\|_0^2 + |\tau|_0^2) \geq C(\|\nabla v\|_0^2 + |\tau|_0^2)(|r|_0 + |v|_0).
\)

So \(|T(\tau, v)|_0 \geq C(\|\nabla v\|_0^2 + |\tau|_0^2). The conditions (2.22) of Theorem 2.4 are satisfied, the \(L^2\)-version coercivity result in (2.18) is proved.

\[\square\]

**Theorem 6.3.** Assuming the coefficients satisfying the condition (1.1) for (2.1), the \(H^{-1}\)-version of the coercivity result in (2.18) holds.

**Proof.** Let \(U = \{\tau \in L^2(\Omega)^d : \tau \cdot n = 0 \text{ on } \Gamma_N\} \times H_D^1(\Omega)\) and let \(V = L^2(\Omega)^d \times (H_D^1(\Omega))'.\) The operator \(T\) from \(U\) to \(V\) is defined as:

\[T(\tau, v) = \left(\begin{array}{c}
A^{-1/2}\tau + A^{1/2}\nabla v \\
\nabla \cdot \tau + Y_{st}(\tau, v)
\end{array}\right)\bigg|_V.
\]

Then we have

(6.4) \(\|T(\tau, v)\|_V^2 = \|A^{-1/2}\tau + A^{1/2}\nabla v\|^2_0 + \|\nabla \cdot \tau + Y_{st}(\tau, v)\|_{-1}^2.
\)

Let

\[r_1 = A^{-1/2}\tau + tA^{-1/2}bv, \quad r_2 = A^{1/2}\nabla v + tA^{-1/2}bv, \quad \tau = r_1 + r_2, \quad \text{and} \quad w = S^{-1}v.
\]

Since \(v \in H_D^1(\Omega),\) thus \(S^{-1}\) is well-defined. By (1.6), we have \(|v|_{-1} = |\nabla v|_0\). By the definition of the inner product in \((H_D^1(\Omega))'\) in (1.4) and the definitions of \(r\) and \(w,\) we have

\[
(T(\tau, v), (r, w))_V = (A^{-1}\tau + \nabla v, \tau + t bv) + (A^{-1}\tau + \nabla v, A\nabla v + t bv) + 2(\nabla \cdot \tau + Y_{st}(\tau, v), SS^{-1}v), \\
= (A^{-1}\tau + \nabla v, \tau + t bv) + (A^{-1}\tau + \nabla v, A\nabla v + t bv) + 2(\nabla \cdot \tau + Y_{st}(\tau, v), v).
\]

Then with the almost same calculations in Theorem 2.4 we have

\[
(T(\tau, v), (r, w))_V \geq C(\|\tau\|_0^2 + |\tau|_0^2).
\]

We also have \(|r|_0 \leq C(\|\nabla v\|_0 + |\tau|_0)\) and \(|w|_{-1} = |\nabla v|_0\). Thus

(6.5) \(\|T(\tau, v)\|_V^2 \geq (T(\tau, v), (r_1 + r_2, w)) \geq C(\|\nabla v\|_0^2 + |\tau|_0^2) \geq C(\|\nabla v\|_0^2 + |\tau|_0^2)(|r|_0 + |w|_{-1}).
\)

So \(|T(\tau, v)|_0 \geq C(\|\nabla v\|_0 + |\tau|_0). The conditions (2.22) of Theorem 2.4 are satisfied, the \(H^{-1}\)-version coercivity result in (2.18) is proved.

\[\square\]

**Theorem 6.4.** Assuming the coefficients satisfying the condition (4.1) for (2.2), the \(L^2\)-version of the coercivity result in (2.19) holds.

**Proof.** Let \(U = X\) and \(V = L^2(\Omega)^d \times L^2(\Omega).\) The operator \(T\) from \(U\) to \(V\) is defined as:

\[T(\tau, v) = \left(\begin{array}{c}
\tau + A\nabla v + bv \\
\nabla \cdot \tau + cv
\end{array}\right).
\]
Let \( r = \nabla v \) and \( w = v \). We have \( \| r \|_0 + \| w \|_0 \leq C \| \nabla v \|_0 \). Then
\[
(T(\tau), (r, w)) = (\tau + A\nabla v + bv, \nabla v) + (\nabla \cdot \tau + cv, v)
\]
\[
\geq (A\nabla v, \nabla v) + (\nabla \cdot \tau, v) - (\tau, \nabla v)
\]
\[
\geq C\| \nabla v \|_0^2 \geq C\| \nabla v \|_0 (\| r \|_0 + \| w \|_0).
\]
Thus \( \| T(\tau, v) \|_0 \geq C \| \nabla v \|_0 \) and (2.24) of Theorem 2.5 is satisfied, the \( L^2 \)-version coercivity result in (2.19) is proved.

**Theorem 6.5.** Assuming the coefficients satisfying the condition \( \text{(2.1)} \) for \( \text{(2.22)} \), the \( H^{-1} \) version of the coercivity result in (2.19) holds.

**Proof.** Let \( U = \{ \tau \in L^2(\Omega)^d : \tau \cdot n = 0 \text{ on } \Gamma_N \} \times H^1_D(\Omega) \) and \( V = L^2(\Omega)^d \times (H^1_D(\Omega))^\prime \). The operator \( T \) from \( U \) to \( V \) is defined as:
\[
T(\tau, v) = \left( \tau + A\nabla v + bv, \nabla \cdot \tau + cv \right).
\]
Let \( r = \nabla v \) and \( w = S^{-1}v \). We have \( \| r \|_0 = \| \nabla v \|_0 \) and \( \| w \|_1 = \| \nabla v \|_0 \) as before.

Then
\[
(T(\tau, v), (r, w))_V = (\tau + A\nabla v + bv, \nabla v) + (\nabla \cdot \tau + cv, SS^{-1}v)
\]
\[
= (\tau + A\nabla v + bv, \nabla v) + (\nabla \cdot \tau + cv, v)
\]
\[
\geq C\| \nabla v \|_0^2 \geq C\| \nabla v \|_0 (\| r \|_0 + \| w \|_1).
\]
Thus \( \| T(\tau, v) \|_V \geq C \| \nabla v \|_0 \) and (2.24) of Theorem 2.5 is satisfied, the \( H^{-1} \)-version coercivity result in (2.19) is proved.

**Remark 6.6.** The proofs in this section are more complicated than those in Proof II section. The advantage of the proofs here is that we have larger freedom to choose the test function, so that a more refined analysis is possible. In fact, in [23], test functions similar to those in [2] are chosen.

7. **Extension to least-squares finite element methods with \( H^{-1} \) righthand side**

In this section, we discuss the least-squares finite element methods for the second-order elliptic equation with an \( H^{-1} \) righthand side. Such a problem appears in many situations, for example, in the goal-oriented a posteriori error estimate [31].

For simplicity, we only consider the \( \text{(2.1)} \)-type of equation with a pure Dirichlet boundary condition and the \( L^2 \)-version of the least-squares finite element method. As pointed out in [29] [9], any functional in \( H^{-1} := (H^1_0(\Omega))^\prime \) can be written as \( f - \nabla \cdot g \) for \( f \in L^2(\Omega) \) and \( g \in L^2(\Omega)^d \). Thus, we consider the following problem:

\[
(7.1) \quad -\nabla \cdot (A\nabla u) + b \cdot \nabla u + cu = f_1 - \nabla \cdot (Af_2) \text{ in } \Omega, \quad u = 0 \text{ on } \partial \Omega.
\]

Here \( f_1 \in L^2(\Omega) \) and \( f_2 \in L^2(\Omega)^d \) are given functions. We have \( f_1 + \nabla \cdot (Af_2) \in H^{-1}(\Omega) \). The divergence of \( Af_2 \) should be understood in the distributional sense, i.e.,
\[
(\nabla \cdot (Af_2), v) = -(Af_2, \nabla v) \quad \forall v \in H^1_0(\Omega).
\]
A typical \( Af_2 \) can be \( A\nabla v_h \) for \( v_h \) being a function in the conforming finite element space. Such a righthand side appears in the recovery based error estimators, see [21]. The standard variational problem reads: Find \( u \in H^1_0(\Omega) \), such that

\[
(7.2) \quad (A\nabla u, \nabla v) + (b \cdot \nabla u + cu, v) = (f_1, v) + (Af_2, \nabla v) \quad \forall v \in H^1_0(\Omega).
\]
For the equation (7.1), the quantity \( A\nabla u \) is not in \( H(\text{div}; \Omega) \). Since \( \nabla \cdot (Af_2) \) is only in \( H^{-1}(\Omega) \), not \( L^2(\Omega) \), thus \( \nabla \cdot (A\nabla u) = b \cdot \nabla u + cu - f_1 - \nabla \cdot (Af_2) \) is only in \( H^{-1}(\Omega) \). Taking the simplest example, let \( b = 0, c = 0, f_1 = 0, A = I \), and \( f_2 = \nabla v_h \), where \( v_h \) is a continuous piecewise linear finite element function. It is
well-known that we usually do not have $[\nabla v_h \cdot \mathbf{n}_e]_F = 0$ across an internal edge $F$ of a finite element mesh. In fact, this term often appears in the residual type of a posteriori error estimator [11,12] and recovering it in the $H(\text{div})$-conforming space is the foundation of recovery-based error estimator [20]. Then for this simplest example, $A \nabla u = -A \nabla v_h \notin H(\text{div}; \Omega)$.

On the other hand, let the flux $\boldsymbol{\sigma} = -A \nabla u + A f_2 \in L^2(\Omega)^d$, then $\nabla \cdot \boldsymbol{\sigma} = f_1 - b \cdot \nabla u - cu \in L^2(\Omega)$, thus we have $\boldsymbol{\sigma} \in H(\text{div}; \Omega)$. We use the fact $\nabla u = f_2 - A^{-1} \boldsymbol{\sigma}$ and get a first-order system:

$$
\begin{align*}
\boldsymbol{\sigma} + A \nabla u &= Af_2 & \text{in } \Omega, \\
\nabla \cdot \boldsymbol{\sigma} + Y_{st}(\sigma, u) &= f_1 - t b \cdot f_2 & \text{in } \Omega, \\
u &= 0 & \text{on } \partial \Omega.
\end{align*}
$$

(7.3)

The definition of $Y_{st}$ can be found in (2.6). For $(\tau, v) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega)$, define the least-squares functional for the system (7.3).

$$
M(\tau, v; f_1, f_2) := \|A^{-1/2} \tau + A^{1/2} \nabla v - A^{1/2} f_2\|_{0, \Omega}^2 + \|\nabla \cdot \tau + Y_{st}(\tau, v) - f_1 + t b \cdot f_2\|_{0, \Omega}^2.
$$

(7.4)

The corresponding least-squares minimization problem is:

$$
\text{Find } (\sigma, u) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega) \text{ s.t. } M(\sigma, u; f_1, f_2) = \inf_{(\tau, v) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega)} M(\tau, v; f_1, f_2).
$$

(7.5)

The Euler-Lagrange weak problem is: Find $(\sigma, u) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega)$ such that

$$
a(\sigma, u; \tau, v) = F(\tau, v) \quad \forall (\tau, v) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega),
$$

(7.6)

where the bilinear form $a$ and linear form $F$ are defined for all $(\rho, w)$ and $(\tau, v) \in H(\text{div}; \Omega) \times H_{00}^1(\Omega)$ as:

$$
a(\rho, w; \tau, v) = (\rho + A \nabla w, A^{-1} \tau + \nabla v) + (\nabla \cdot \rho + Y_{st}(\rho, w), \nabla \cdot \tau + Y_{st}(\tau, v)),
$$

(7.7)

$$
F(\tau, v) = (Af_2, A^{-1} \tau + \nabla v) + (f_1 - t b \cdot f_2, \nabla \cdot \tau + Y_{st}(\tau, v)).
$$

(7.8)

Let $\mathcal{T} = \{K\}$ be a triangulation of $\Omega$ using simplicial elements. The mesh $\mathcal{T}$ is assumed to be regular. For an element $K \in \mathcal{T}$ and an integer $k \geq 0$, let $P_k(K)$ be the space of polynomials with degrees less than or equal to $k$. Define the finite element spaces $RT_k, S_k$, and $S_{k,0}$ as follows:

$$
RT_k := \{\tau \in H(\text{div}; \Omega): \tau|_K \in P_k(K)^d + xP_k(K), \forall K \in \mathcal{T}\},
$$

$$
S_k := \{v \in C^0(\Omega): v|_K \in P_k(K), \forall K \in \mathcal{T}\} \quad \text{and} \quad S_{k,0} := S_k \cap H_{00}^1(\Omega).
$$

Then for $k \geq 0$ being an integer, the corresponding first-order system least-squares minimization problem and finite element problem are: Find $(\sigma_h, u_h) \in RT_k \times S_{k+1,0}$, such that

$$
M(\sigma_h, u_h; f_1, f_2) = \inf_{(\tau, v) \in RT_k \times S_{k+1,0}} M(\tau, v; f_1, f_2),
$$

(7.9)

and find $(\sigma_h, u_h) \in RT_k \times S_{k+1,0}$, such that

$$
a((\sigma_h, u_h), (\tau, v)) = F(\tau, v) \quad \forall (\tau, v) \in RT_k \times S_{k+1,0},
$$

(7.10)

respectively.

Since it is obvious that $M(\tau, v; 0, 0) = L(\tau, v; 0)$, we immediately have

$$
a((\tau, v), (\tau, v)) \geq C \|\tau, v\|^2.
$$

(7.11)

Thus the existence and uniqueness of both (7.3) and (7.10) are established.
7.1. A priori error estimate. It is easy to derive that the quasi-best approximation property holds:

\[(7.12) \quad \|\sigma - \sigma_h, u - u_h\| \leq C \inf_{(\tau_h,v_h) \in RT_k \times S_{k+1,0}} \|\sigma - \tau_h, u - v_h\|.
\]

Thus, we only need the approximation properties of the discrete spaces and the regularities of the solutions to finish the a priori error estimate.

First, we discuss the local approximation properties of \(S_{k+1,0}\). The space \(H^{1+s}(\Omega)\), with \(s > 0\) for two dimensions and \(s > 1/2\) for three dimensions, is embedded in \(C^0(\Omega)\) by the Sobolev’s embedding theorem, hence, we can define the nodal interpolation of the function \(v \in H^{1+s}(\Omega)\). It is proved in [28] that if \(v \in H^{1+s_k}(K)\) with \(s_k > 0\) in two dimensions and \(s_K > 1/2\) in three dimensions, then for \(s_K \leq k+1\), the following estimate holds for the nodal interpolation \(I_{no}\):

\[(7.13) \quad \|\nabla (v-I_{no}v)\|_{0,K} \leq Ch^{s_K} |v|_{1+s_k,K} \quad \forall K \in \mathcal{T}.
\]

For solutions with low regularities, the nodal interpolation is not well-defined, we can use the modified Clément interpolation [25, 3] or the Scott-Zhang interpolation [41]. For solutions with low regularities, the nodal interpolation is not well-defined, we can use the modified Clément interpolation [25, 3] or the Scott-Zhang interpolation [41]. For an element \(K \in \mathcal{T}\), let \(\Delta_K\) be the collection of elements in \(\mathcal{T}\) that share at least one vertex with \(K\). Assume that \(v \in H^1_0(\Omega)\) and \(|v|_{\Delta_k} \in H^{1+s_{\Delta_k}}(\Delta_K)\) for some \(0 < s_{\Delta_k} \leq k+1\), and let \(I_{sz} v\) be the Scott-Zhang interpolation into \(S_{k+1,0}\), we have

\[(7.14) \quad \|\nabla(v-I_{sz}v)\|_{0,K} \leq Ch^{s_{\Delta_k}} |v|_{1+s_{\Delta_k},\Delta_K}.
\]

Define \(\mathcal{T}_s\) to be the part of the mesh such that the local element-wise regularity \(s_K\) of \(H^{1+s_K}(K)\) is big enough to ensure the nodal interpolation:

\[(7.15) \quad \mathcal{T}_s := \{K \in \mathcal{T} : s_K > 0\} \quad \text{for } d = 2\text{ and } s_K > 1/2 \quad \text{for } d = 3\}.
\]

Combining the approximation properties of (7.13) and (7.14), we have an almost localized approximation result: assume that \(u \in H^1_0(\Omega)\), \(|u|_K \in H^{1+s_K}(K)\) for \(K \in \mathcal{T}_s\), and \(|u|_{\Delta_K} \in H^{1+s_{\Delta_K}}(\Delta_K)\) for \(K \in \mathcal{T}\setminus\mathcal{T}_s\), where \(\max_{K \in \mathcal{T}_s} \{s_K\} \leq k+1\) and \(\max_{K \in \mathcal{T}\setminus\mathcal{T}_s} \{s_{\Delta_K}\} \leq k+1\),

\[(7.16) \quad \inf_{v \in S_{k+1,0}} \|\nabla(u - v)\|_{0} \leq C \left( \sum_{K \in \mathcal{T}_s} h^{s_K}_K |u|_{1+s_K,K} + \sum_{K \in \mathcal{T}\setminus\mathcal{T}_s} h^{s_{\Delta_K}}_K |u|_{1+s_{\Delta_K},\Delta_K} \right).
\]

Assume that \(\sigma = A(f_2 - \nabla u)\) and \(\nabla \cdot \sigma = f_1 - b \cdot \nabla u - cu\) have the following local regularities, respectively:

\[(7.17) \quad \sigma|_K \in H^{f_K}(K) \quad \text{and} \quad \nabla \cdot \sigma|_K \in H^{f_K}(K) \quad K \in \mathcal{T}.
\]

For a fixed \(r > 0\), denote by \(I_{rt} : H(div; \Omega) \cap [H^r(\Omega)]^d \to RT_k\) the standard \(RT\) interpolation operator. We have the following local approximation property: for \(\tau \in H^{f_K}(K)\), \(0 < \ell_K \leq k+1\),

\[(7.18) \quad \|\tau - I_{rt} \tau\|_{0,K} \leq Ch^{f_K}_{\ell_K} |\tau|_{\ell_{K},K} \quad \forall K \in \mathcal{T}.
\]

The estimate in (7.18) is standard for \(\ell_K \geq 1\) and can be proved by the average Taylor series developed in [28] and the standard reference element technique with Piola transformation for \(0 < \ell_K < 1\). We also should notice that the interpolations and approximation properties are completely local.

Denote by \(Q_k : L^2(\Omega) \to D_k\) the \(L^2\)-projection onto \(D_k := \{v \in L^2(\Omega) : v|_K \in P_k(K), K \in \mathcal{T}\}\). The following commutativity property is well-known:

\[(7.19) \quad \nabla \cdot (I_{rt} \tau) = Q_k \nabla \cdot \tau \quad \forall \tau \in H(div; \Omega) \cap H^r(\Omega)^d \quad \text{with} \quad r > 0.
\]

Thus we have the following local approximation property: for \(\nabla \cdot \tau \in H^{t_K}(K)\), \(0 < t_K \leq k+1\),

\[(7.20) \quad \|\nabla \cdot (\tau - I_{rt} \tau)\|_{0,K} \leq Ch^{t_K}_{\ell_K} |\nabla \cdot \tau|_{\ell_{K},K} \quad \forall K \in \mathcal{T}.
\]

Combining the above approximation properties and (7.12), we have the following a priori error estimate:
Theorem 7.1. (A priori error estimate) Assume that \( u \in H^1_0(\Omega) \), \( u|_K \in H^{1+s_K}(K) \) for \( K \in \mathcal{T}_s \), and \( u|_{\Delta_K} \in H^{1+s_{\Delta_K}}(\Delta_K) \) for \( \Delta_K \in \mathcal{T}_\Delta \) where \( \max_{K \in \mathcal{T}_s}\{s_K\} \leq k+1 \) and \( \max_{K \in \mathcal{T}_\Delta}\{s_{\Delta_K}\} \leq k+1 \). Assume that \( \sigma|_K \in H^{1+\ell_K}(K) \) and \( \nabla \cdot \sigma|_K \in H^{1+k}(K) \), for \( 0 < \ell_K \leq k+1 \) and \( 0 < t_K \leq k+1 \). Then we have the following a priori error estimate:

\[
\left\| (\sigma - \sigma_h, u - u_h) \right\| \leq C \left( \sum_{K \in \mathcal{T}} h^{s_K}_K |u|_{1+s_K,K} + \sum_{K \in \mathcal{T} \setminus \mathcal{T}_s} h^{s_{\Delta_K}}_K |u|_{1+s_{\Delta_K} \cdot \Delta_K} + \sum_{K \in \mathcal{T}} (h^{\ell_K}_K |\tau|_{\ell_K,K} + h^{t_K}_K |\nabla \cdot \tau|_{t_K,K}) \right).
\]

Remark 7.2. We should notice that the smoothness \( s_K, \ell_K, \) and \( t_K \) can be independent. Compared to the standard analysis, the localized a priori does not require the global smoothness of \( \sigma \) and \( \nabla \cdot \sigma \), or in terms of the data, the global smoothness of \( M_2 \) and \( f_1 \). We only need the local (element-wise) smoothness of such data.

The discussion of (7.10) can be applied to the conforming \( C^0 \)-Lagrange approximation to the standard variational formulation too, although the analysis is not coefficient-robust comparing to the results in [14] for discontinuous and mixed approximations.

The a priori error estimate with local regularity is the base for adaptive finite element methods to achieve equal discretization error distribution. In a sense, the a priori results show us that what an optimal mesh should be. For example, assuming \( \mathcal{T}_s = \mathcal{T} \), then an error-equal-distributed mesh should have a similar size of \( h^{s_K}_K |u|_{1+s_K,K} + h^{\ell_K}_K |\tau|_{\ell_K,K} + h^{t_K}_K |\nabla \cdot \tau|_{t_K,K} \) for all \( K \in \mathcal{T} \).

7.2. A posteriori error estimate. Let

\[
E = \sigma - \sigma_h \quad \text{and} \quad e = u - u_h.
\]

Then by the first-order system (7.23), we have

\[
M(\sigma_h, u_h; f_1, f_2) = \left\| A^{-1/2} (\sigma_h + A\nabla u_h - \sigma - A\nabla u) \right\|_0^2 + \left\| \nabla \cdot \sigma + Y_st(\sigma, u) - \nabla \cdot \sigma_h + Y_st(\sigma_h, u_h) \right\|_0^2
\]

\[
= \left\| A^{-1/2} (E + A\nabla e) \right\|_0^2 + \left\| \nabla \cdot E + Y_st(E, v) \right\|_0^2 = M(E, e; 0, 0).
\]

Since \( E \in H_N(\text{div}; \Omega) \) and \( e \in H^1_0(\Omega) \), we have

\[
C_1 \| (E, e) \|_0^2 \leq M(\sigma_h, u_h; f_1, f_2) \leq C_2 \| (E, e) \|_0^2.
\]

Thus we can define a posteriori error estimator and local error indicator as:

\[
\eta^2_K = \left\| A^{-1/2} \tau + A^{1/2} \nabla v - A^{1/2} f_2 \right\|_0^2 + \left\| \nabla \cdot \tau + Y_st(\tau, v) - f_1 + t b \cdot f_2 \right\|_0^2,
\]

\[
\eta^2 = \sum_{K \in \mathcal{T}} \eta^2_K = M(\sigma_h, u_h; f_1, f_2).
\]

The equivalence (7.23) shows the reliability and efficiency of the error estimator.

By the triangle inequality, we also have the local efficiency bound,

\[
\eta_K \leq C(\| \nabla e \|_0 + \| e \|_0 + \| E \|_0 + \| \nabla \cdot E \|_0)
\]

\[
\leq C(h^{s_K}_K |u|_{1+s_K,K} + h^{\ell_K}_K |\tau|_{\ell_K,K} + h^{t_K}_K |\nabla \cdot \tau|_{t_K,K}).
\]

Thus, even though we do not have the local error exactness of the indicator, if the error indicators \( \eta_K \) are of a similar size, then we can achieve the local optimal error estimate (7.24), which means, the mesh obtained by equal \( \eta_K \) distribution for all \( K \in \mathcal{T} \) is an optimal mesh.

References

[1] M. Ainsworth and J. T. Oden, A Posteriori Error Estimation in Finite Element Analysis A Posteriori Error Estimation in Finite Element Analysis, Wiley, 2000.

[2] B. Ayuso and L. D. Marini, Discontinuous galerkin methods for advection-diffusion-reaction problems, SIAM J. Numer. Anal., 47 (2009), pp. 1391-1420.
[3] C. Bernardi and V. Girault, A local regularization operator for triangular and quadrilateral finite elements, SIAM J. Numer. Anal., 35 (1998), pp. 1893–1916.

[4] M. Bernkopf and J. M. Melenk, Optimal convergence rates in $L^2$ for a first order system least squares finite element method. Part I: Homogeneous boundary conditions for a first order system least squares finite element method. Part II: Homogeneous boundary conditions, arXiv:2012.12919, (2020).

[5] P. B. Bochev and M. D. Gunzburger, Least-squares finite element methods of least-squares type, SIAM Review, 40 (1998), pp. 789–837.

[6] P. B. Bochev and M. D. Gunzburger, Least-Squares Finite Element Methods, Applied Mathematical Sciences, 166, Springer, 2009.

[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 systems, Mathematics of Computation, 66 (1997), pp. 935–955.

[8] J. H. Bramble, R. D. Lazarov, and J. E. Pasciak, Least squares for second order elliptic problems, Computer Methods in Applied Mechanics and Engineering, 152 (1998), pp. 195–210.

[9] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Universitext, Springer, 2011.

[10] Z. Cai and S. Zhang, A first order system least squares method, Lecture notes of Purdue University, 2004.

[11] Z. Cai, J. Chen, and M. Liu, Least-squares relu neural network (lsnn) method for linear advection-reaction equation, Journal of Computational Physics, 443 (2021).

[12] Z. Cai, J. Chen, M. Liu, and X. Liu, Deep least-squares methods: an unsupervised learning-based numerical method for solving elliptic pdes, Journal of Computational Physics, 420 (2020).

[13] Z. Cai, B. Lee, and P. Wang, Deep least-squares methods for incompressible newtonian fluid flow: linear stationary problems, SIAM J. Numer. Anal., 42 (2004), pp. 843–859.

[14] Z. Cai, T. 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.

[15] Z. Cai, L. F. Demkowicz, J. Gopalakrishnan, I. Muga, and Z. J. H. Chaudhry, E. C. Cyr, K. Liu, T. A. Manteuffel, L. N. Olson, and L. Tang, Enhancing least-squares finite element methods through a quantity-of-interest, SIAM J. Numer. Anal., 52 (2014), pp. 3085–3105.

[16] H. Chen, G. Fu, J. Li, and W. Qu, First order least squares method with weakly imposed boundary condition for convection dominated diffusion problems, Computers and Mathematics with Applications, 68 (2014), pp. 1635–1652.

[17] H. Chen and W. Qu, A first order system least squares method for the helmholtz equation, Journal of Computational and Applied Mathematics, 309 (2017), pp. 145–162.

[18] P. Clément, Approximation by finite element functions using local regularization, RAIRO, Anal. Num., 9 (1975).

[19] A. Codd, T. Manteuffel, and S. McCormick, Multilevel first-order system least squares for nonlinear partial differential equations, SIAM J. Numer. Anal., 41 (2003), pp. 2197–2209.

[20] L. F. Demkowicz, J. Gopalakrishnan, I. Muga, and Z. J., Wavenumber explicit analysis of a DPG method for the multidimensional helmholtz equation, Comput. Methods Appl. Mech. Engrg., 213-216 (2012), pp. 126–138.

[21] T. Dupont and R. Scott, Polynomial approximation of functions in sobolev spaces, Math. Comp., 34 (1980), pp. 441–463.

[22] L. C. Evans, Partial Differential Equations, American Mathematical Society, 2 ed., 2010.

[23] J. Gopalakrishnan and W. Qu, An analysis of the practical dpg method, Mathematics of Computation, 83 (2014), pp. 537–552.

[24] M. Innerberger and D. Praetorius, Instance-optimal goal-oriented adaptivity, Comput. Methods Appl. Math., 21 (2021), pp. 109–126.

[25] B.-N. Jiang, The Least-Squares Finite Element Method Theory and Applications in Computational Fluid Dynamics and Electromagnetics, Scientific Computation, Springer, 1998.

[26] J. Ku, A remark on the coercivity for a first-order least-squares method, Numer. Methods Partial Differential Equations, 23 (2007), pp. 1577–1581.
[34] R. Lazarov, L. Tobiska, and P. Vassilevski, Streamline-diffusion least-squares mixed finite element methods for convection-diffusion problems, East-West J. Numer. Math., 5 (1997), pp. 249–264.

[35] B. Lee, First-order system least-squares for elliptic problems with robin boundary conditions, SIAM J. Numer. Anal., 37 (1999), pp. 70–104.

[36] Y. Liang and S. Zhang, Adaptive nonconforming least-squares finite element methods with crouzeix-raviart elements, preprint, (2021).

[37] Q. Liu and S. Zhang, Adaptive flux-only least-squares finite element methods for linear transport equations, Journal of Scientific Computing, 84:26 (2020), https://doi.org/10.1007/s10915-020-01269-y.

[38] Q. Liu and S. Zhang, Adaptive least-squares finite element methods for linear transport equations based on an $H(div)$ flux reformulation, Comput. Methods Appl. Mech. Engrg., 366 (2020), p. 113041, https://doi.org/10.1016/j.cma.2020.113041.

[39] A. Pehlivanov, G. Carey, and P. Vassilevski, Least-squares mixed finite element methods for non-selfadjoint elliptic problems: I. error estimates, Numer. Math., 72 (1996).

[40] W. Qiu and S. Zhang, Adaptive first-order system least-squares finite element methods for second order elliptic equations in non-divergence form, SIAM J. Numer. Anal., 58 (2020), pp. 3286–3308.

[41] L. R. Scott and S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions, Math. Comp, 54 (1990), pp. 483–493.

[42] R. Verfürth, A Posteriori Error Estimation Techniques for Finite Element Methods, Oxford University Press, 2013.

[43] S. Zhang, Robust and local optimal a priori error estimates for interface problems with low regularity: Mixed finite element approximations, Journal of Scientific Computing, 84 (2020), https://doi.org/10.1007/s10915-020-01284-w.

Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China
Email address: yuxiliang7-c@my.cityu.edu.hk, shun.zhang@cityu.edu.hk