Lecture Note III: Least-Squares Method

Zhiqiang Cai*

October 24, 2004

In this chapter, we shall present least-squares methods for second-order scalar partial differential equations, elastic equations of solids, incompressible Newtonian fluid flow, and Maxwell’s equations in electromagnetic.

1 A General Methodology

We give a general methodology for the design of least-squares methods applied to a first-order system of partial differential equations. Consider the following first-order partial differential system defined on a bounded domain $\Omega \subset R^d$ ($d = 2$ or $3$):

$$\begin{cases} \mathcal{L} U = F & \text{in } \Omega, \\ \mathcal{B} U = G & \text{on } \partial \Omega, \end{cases}$$

(1.1)

where $\mathcal{L} = (L_{ij})_{m \times n}$ is a block $m \times n$ matrix differential operator of at most first order, $\mathcal{B} = (B_{ij})_{l \times n}$ is a block $l \times n$ matrix operator, $U = (U_i)_{n \times 1}$ is unknown, $F = (F_i)_{m \times 1}$ is a given block vector-valued function defined in $\Omega$, $G = (G_i)_{l \times 1}$ is a given block vector-valued function defined on $\partial \Omega$. Assume that first-order system (1.1) has a unique solution $U$. Boundary conditions in a least-squares formulation can be imposed either strongly (in the solution space) or weakly (by adding boundary functionals). For simplicity of presentation, we impose them in the solution space $\Phi$. Assume that $\Phi$ is appropriately chosen so that least-squares functional is well defined.

Define the least-squares functional by

$$G(U; F) = \sum_{i=1}^m \left\lVert \sum_{j=1}^n L_{ij} U_j - F_i \right\rVert_{k(i), \Omega}^2,$$

(1.2)

where $\lVert \cdot \rVert_{k(i), \Omega}$ denotes a Sobolev norm and $k(i) = -1$ or $0$. If $k(i) = 0$ for all $i$, $G(U; F)$ is referred to as the $L^2$ norm least-squares functional, otherwise, it is referred to as the inverse norm least-squares functional. Denote the Laplace operator by $\Delta$ and the $L^2$
inner product by \((f, g) = \int_\Omega f g \, dx\). Then \(\| \cdot \|_{L^2(\Omega)} = \| \cdot \|_{H^0(\Omega)} = \sqrt{(\cdot, \cdot)}\) and \(\| \cdot \|_{H^{-1}(\Omega)} = \sqrt{(\cdot, \cdot)}\). Now, the least-squares minimization problem is to minimize the least-squares functional over \(\Phi\):

\[
G(U; F) = \min_{V \in \Phi} G(V; F).
\]

(1.3)

This is equivalent to solving the normal equation:

\[
(\mathcal{L}^* \mathcal{K} \mathcal{L}) U = (\mathcal{L}^* \mathcal{K}) F
\]

(1.4)

where \(\mathcal{L}^*\) is the adjoint operator of \(\mathcal{L}\) with respect to the \(L^2\) inner product and \(\mathcal{K}\) is a block diagonal operator with each block associated with the \(H^{k(i)}(\Omega)\) norm

\[
\mathcal{K} = \text{diag} \left( K_1, \cdots, K_m \right) \quad \text{where} \quad K_i = I \quad \text{or} \quad (-\Delta)^{-1}.
\]

That is, each diagonal block of \(\mathcal{K}\) is either the identity or the inverse of Laplacian. (For the \(L^2\) least squares, we have \(\mathcal{K} = I\).) The normal operator \(\mathcal{L}^* \mathcal{K} \mathcal{L}\) is a differential operator of at most second-order. The variational form of (1.4) is to find \(U \in \Phi\) such that

\[
b(U, V) \equiv (\mathcal{K} \mathcal{L} U, \mathcal{L} V) = (\mathcal{K} F, \mathcal{L} V) \equiv f(V) \quad \forall V \in \Phi.
\]

(1.5)

It is easy to check that

\[
G(V; 0) = b(V, V).
\]

The design of the least-squares method is to choose first-order system (1.1) and least-squares norms so that the least-squares problem (i.e., the normal equation or the minimization problem or the weak form) can be numerically solved effectively and efficiently; i.e.,

- the least-squares variables can be discretized with optimal accuracy;
- the resulting algebraic system can be solved with optimal complexity.

It is well known that problems with the identity or Laplace operators can be numerically solved with both optimal accuracy and complexity. Recently, it was shown that this is also true for problems involving the \(H(\text{div})\) and \(H(\text{curl})\) operators

\[
H_{\text{div}} = I - \nabla \text{div} \quad \text{and} \quad H_{\text{curl}} = I + \nabla \times \nabla \times
\]

provided Raviart-Thomas elements [17] are used for \(H(\text{div})\) and edge elements [3] are used for \(H(\text{curl})\). Hence, one wants to develop the least-squares method so that the normal operator \(\mathcal{L}^* \mathcal{K} \mathcal{L}\) is equivalent to a block diagonal operator whose diagonal block is either the identity, Laplacian, \(H_{\text{div}}\), or \(H_{\text{curl}}\) operators:

\[
\mathcal{D} = \text{diag} \left( D_1, \cdots, D_m \right) \quad \text{where} \quad D_i = I, -\Delta, H_{\text{div}}, \text{ or } H_{\text{curl}}.
\]

With such equivalence, the least-squares problem can then be numerically solved with optimal accuracy and optimal complexity. Moreover, finite element spaces for different
variables $U_i$ can be chosen independently and, hence, based solely on the approximation properties and implementation/computational costs.

The above equivalence between $\mathcal{L}^* \mathcal{K} \mathcal{L}$ and $\mathcal{D}$ means that there exist positive constants $\alpha_0$ and $\alpha_1$ such that

$$\alpha_0 \sum_{i=1}^{m} \|V_i\|_D^2 \leq b(V, V) \quad \text{and} \quad b(U, V) \leq \alpha_1 \left( \sum_{i=1}^{m} \|U_i\|_D^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^{m} \|V_i\|_D^2 \right)^{\frac{1}{2}}$$

(1.6)

for all $U, V \in \Phi$.

Here $\| \cdot \|_D$ denotes the $L^2$, $H^1$, $H(\text{div})$, or $H(\text{curl})$ norms. The main task of analyzing least-squares methods is to establish (1.6). Many physical models involve parameters such as the Lamé constants for solids and the viscosity parameters for fluids. It is then important to establish equivalence independent of these parameters. This is because parameter-independent equivalence implies robustness of the least-squares methods with respect to these parameters.

1.1 Least-Squares Approximation

Assume that $\Phi^h$ is a finite dimensional subspace of $\Phi$ satisfying the following approximation property:

$$\inf_{V^h \in \Phi^h} \left( \sum_{i=1}^{m} \|V_i - V_i^h\|_D^2 \right)^{\frac{1}{2}} \leq C_a h^s$$

(1.7)

for all $V = (V_i)_{n \times 1} \in \Phi$. Then least-squares approximation is to find $U^h \in \Phi^h$ such that

$$G(U^h; \mathcal{F}) = \min_{V \in \Phi^h} G(V; \mathcal{F}).$$

(1.8)

Equivalently, find $U^h \in \Phi^h$ such that

$$b(U^h, V) = f(V) \quad \forall \ V \in \Phi^h.$$  

(1.9)

**Theorem 1.1** Let $U$ and $U^h$ be the solutions of (1.5) and (1.9), respectively. Assume that equivalence (1.6) and approximation property (1.7) hold. Then we have the following error estimation:

$$\left( \sum_{i=1}^{m} \|U_i - U_i^h\|_D^2 \right)^{\frac{1}{2}} \leq C_a \alpha_1 \alpha_0^{-1} h^s.$$  

(1.10)

**Proof:** Difference of (1.5) and (1.9) gives the error equation:

$$b(U - U^h, V) = 0 \quad \forall \ V \in \Phi^h.$$  

(1.11)
It follows from (1.6) and (1.7) that for any $V \in \Phi^h$

$$\alpha_0 \sum_{i=1}^{m} ||U_i - U_i^h||_{D_i}^2 \leq b(U - U^h, U - U^h) = b(U - U^h, U - V)$$

$$\leq \alpha_1 \left( \sum_{i=1}^{m} ||U_i - U_i^h||_{D_i}^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^{m} ||V_i - U_i^h||_{D_i}^2 \right)^{\frac{1}{2}}$$

$$\leq \alpha_1 C_a h^s \left( \sum_{i=1}^{m} ||U_i - U_i^h||_{D_i}^2 \right)^{\frac{1}{2}}.$$

Dividing on both sides by $\alpha_0 \left( \sum_{i=1}^{m} ||U_i - U_i^h||_{D_i}^2 \right)^{\frac{1}{2}}$ yields (1.10) and, hence, the theorem.

1.2 Mesh Refinement Indicator

Let $U$ be the solution of (1.1) and $V \in \Phi$ be a computed approximation to $U$. Then (1.1) and (1.6) imply

$$G(V; F) = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} L_{ij} V_j - F_i \right)^2_{k(i), \Omega} = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} L_{ij} (V_j - U_j) \right)^2_{k(i), \Omega}$$

$$\sim \sum_{i=1}^{m} ||V_i - U_i||_{D_i}^2.$$

Since

$$G(0; F) = \sum_{i=1}^{m} ||F_i||_{k(i), \Omega}^2 = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} L_{ij} U_j \right)^2_{k(i), \Omega} \sim \sum_{i=1}^{m} ||U_i||_{D_i}^2,$$

combining with (1.12) gives

$$\frac{G(V; F)}{G(0; F)} \sim \frac{\sum_{i=1}^{m} ||V_i - U_i||_{D_i}^2}{\sum_{i=1}^{m} ||U_i||_{D_i}^2}.$$  (1.13)

(1.12) means that the value of the least-squares functional at $V$ gives certain measurement of absolute difference between the solution $U$ and an approximation $V$ in the functional induced norm. Therefore, the value of the least-squares functional at $V$ on each element probably gives a reasonable mesh refinement indicator. Especially, this is true for nonlinear problem.
2 Second-Order Scalar PDEs

Consider the following second-order elliptic boundary value problem:

\[
\begin{aligned}
-\nabla \cdot (A \nabla p) + Xp &= f, & \text{in } \Omega, \\
p &= 0, & \text{on } \Gamma_D, \\
n \cdot A \nabla p &= 0, & \text{on } \Gamma_N,
\end{aligned}
\]  

(2.1)

where \(A\) is a \(d \times d\) symmetric matrix of functions in \(L^2(\Omega)\) and \(X\) is an at most first-order linear differential operator. We assume that \(A\) is uniformly symmetric positive definite: there exist positive constants \(0 < \lambda \leq \Lambda\) such that

\[
\lambda \xi^T \xi \leq A \xi^T \xi \leq \Lambda \xi^T \xi
\]

(2.2)

for all \(\xi \in \mathbb{R}^d\) and almost all \(x \in \overline{\Omega}\). The corresponding variational form of system (2.1) is to find \(p \in V\) such that

\[
a(p, q) = f(q) \quad \forall q \in V
\]

(2.3)

where

\[
V = \begin{cases} 
H^1_{0, D}(\Omega) & \text{if } \text{mes}(\Gamma_D) \neq \emptyset \\
\hat{H}^1(\Omega) & \text{otherwise},
\end{cases}
\]

with \(\hat{H}^1(\Omega) = \{v \in H^1(\Omega) | \int_{\Omega} v \, dx = 0\}\) and the bilinear and linear forms are defined by

\[
a(p, q) = (A \nabla p, \nabla q) + (Xp, q) \quad \text{and} \quad f(q) = (f, q),
\]

respectively. Under appropriate assumptions on \(\Gamma_D\) and \(X\), problem (2.3) is uniquely solvable in \(H^1_{0, D}(\Omega)\) for any \(f \in H^{-1}(\Omega)\) or uniquely solvable in \(\hat{H}^1(\Omega)\) if and only if \(f\) satisfies the compatibility condition

\[
\int_{\Omega} f \, dx = 0.
\]

2.1 First-Order System of PDEs

For (2.1), we consider two first-order systems. To this end, introducing the flux variable

\[
u = -A \nabla p,
\]

problem (2.1) may be rewritten as a first-order system of partial differential equations as follows:

\[
\mathcal{L} \mathcal{U} \equiv \begin{pmatrix} A^{-\frac{1}{2}} & A^{\frac{1}{2}} \nabla \\
\text{div} & X \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} 0 \\ f \end{pmatrix} \equiv \mathcal{F} \quad \text{in } \Omega
\]

(2.4)

with boundary conditions

\[
p = 0 \quad \text{on } \Gamma_D \quad \text{and} \quad n \cdot u = 0 \quad \text{on } \Gamma_N.
\]

(2.5)
Based on this system, we will consider two functionals: (1) Div least-squares functional and (2) inverse norm least-squares functional.

Note that if \( \mathbf{u} \) is sufficiently smooth, then the properly scaled solution, \( A^{-1}\mathbf{u} \), of (2.4) is curl free, i.e., \( \nabla \times (A^{-1}\mathbf{u}) = \mathbf{0} \), and that the homogeneous Dirichlet boundary condition on \( \Gamma_D \) implies the tangential flux condition
\[
\mathbf{n} \times (A^{-1}\mathbf{u}) = \mathbf{0} \quad \text{on } \Gamma_D.
\]

We then have a redundant but consistent first-order system:
\[
\mathbf{L} \mathbf{u} \equiv \begin{pmatrix}
A^{-\frac{1}{2}} & A^\frac{1}{2} \nabla \\
\mathrm{div} & X \\
\nabla \times A^{-1} & 0
\end{pmatrix}
\begin{pmatrix}
\mathbf{u} \\
p
\end{pmatrix} =
\begin{pmatrix}
0 \\
f
\end{pmatrix}
\equiv \mathcal{F}
\quad (2.6)
\]
with boundary conditions
\[
p = 0, \quad \mathbf{n} \times (A^{-1}\mathbf{u}) = \mathbf{0} \quad \text{on } \Gamma_D, \quad \text{and} \quad \mathbf{n} \cdot \mathbf{u} = 0 \quad \text{on } \Gamma_N. \quad (2.7)
\]

Based on this system, we will consider Div-curl least-squares functional.

2.2 Div Least-Squares Functional

Let \( H_N(\mathrm{div}; \Omega) \) denote a subspace of \( H(\mathrm{div}; \Omega) \):
\[
H_N(\mathrm{div}; \Omega) = \{ \mathbf{v} \in H(\mathrm{div}; \Omega) : \mathbf{n} \cdot \mathbf{v} = 0 \text{ on } \Gamma_N \}.
\]

For any \( (\mathbf{v}, q) \in H_N(\mathrm{div}; \Omega) \times V \equiv \Phi \), consider the following div least-squares functional:
\[
G(\mathbf{v}, q; f) = \|A^{-\frac{1}{2}}(\mathbf{v} + A\nabla q)\|_{0,\Omega}^2 + \|\nabla \cdot \mathbf{v} + Xq - f\|_{0,\Omega}^2.
\quad (2.8)
\]

The corresponding normal operator is
\[
\mathcal{L}^* \mathcal{L} = \begin{pmatrix}
A^{-1} - \nabla \mathrm{div} & \nabla (I - X) \\
-(I - X^*) \mathrm{div} & -\mathrm{div} A \nabla + X^* X
\end{pmatrix}
\quad (2.9)
\]
with \( \mathcal{L} \) defined in (2.4) and the corresponding bilinear and linear forms are
\[
\begin{align*}
b(\mathbf{u}, p; \mathbf{v}, q) &= (A^{-1}(\mathbf{u} + A\nabla p), (\mathbf{v} + A\nabla q)) + (\nabla \cdot \mathbf{u} + Xp, \nabla \cdot \mathbf{v} + Xq), \\
f(\mathbf{v}, q) &= (f, \nabla \cdot \mathbf{v} + Xq).
\end{align*}
\quad (2.10)
\]

The main task of this section is to establish the following equivalence:
\[
\mathcal{L}^* \mathcal{L} \sim \begin{pmatrix}
I - \nabla \mathrm{div} & 0 \\
0 & -\Delta
\end{pmatrix}.
\]

6
Theorem 2.1 There exist positive constants $\alpha_0$ and $\alpha_1$ such that

$$\alpha_0 \left( \|\mathbf{v}\|^2_{H(\text{div}; \Omega)} + \|q\|^2_{1, \Omega} \right) \leq b(\mathbf{v}, q; \mathbf{v}, q) = G(\mathbf{v}, q; 0)$$  \hspace{1cm} (2.12)

for any $(\mathbf{v}, q) \in H_N(\text{div}; \Omega) \times V$ and

$$b(\mathbf{u}, p; \mathbf{v}, q) \leq \alpha_1 \left( \|\mathbf{u}\|^2_{H(\text{div}; \Omega)} + \|p\|^2_{1, \Omega} \right)^{\frac{1}{2}} \left( \|\mathbf{v}\|^2_{H(\text{div}; \Omega)} + \|q\|^2_{1, \Omega} \right)^{\frac{1}{2}}$$  \hspace{1cm} (2.13)

for any $(\mathbf{u}, p)$ and any $(\mathbf{v}, q)$ in $H_N(\text{div}; \Omega) \times V$.

Proof: (2.13) is a direct consequence of the Cauchy-Schwarz and triangle inequalities. To show the validity of (2.12), we first establish that

$$\left( \|\mathbf{v}\|^2_{H(\text{div}; \Omega)} + \|q\|^2_{1, \Omega} \right) \leq C \left( G(\mathbf{v}, q; 0) + \|q\|^2_{0, \Omega} \right).$$  \hspace{1cm} (2.14)

It follows from integration by parts, the Cauchy-Schwarz inequality, the Poincaré inequality, and (2.2) that

$$\|A^{-\frac{1}{2}} \nabla q\|^2_{0, \Omega} = (A^{-\frac{1}{2}} \nabla q + A^{\frac{1}{2}} \mathbf{v}, A^{-\frac{1}{2}} \nabla q) - (\mathbf{v}, \nabla q)
= (A^{-\frac{1}{2}} \nabla q + A^{\frac{1}{2}} \mathbf{v}, A^{-\frac{1}{2}} \nabla q) + (\nabla \cdot \mathbf{v}, q)
= (A^{-\frac{1}{2}} \nabla q + A^{\frac{1}{2}} \mathbf{v}, A^{-\frac{1}{2}} \nabla q) + (\nabla \cdot \mathbf{v} + Xq, q) - (Xq, q)
\leq \|A^{-\frac{1}{2}} \nabla q + A^{\frac{1}{2}} \mathbf{v}\|_{0, \Omega} \|A^{-\frac{1}{2}} \nabla q\|_{0, \Omega} + \|\nabla \cdot \mathbf{v} + Xq\|_{0, \Omega} \|q\|_{0, \Omega} + \|Xq\|_{0, \Omega} \|q\|_{0, \Omega}
\leq \left( \|A^{-\frac{1}{2}} \nabla q + A^{\frac{1}{2}} \mathbf{v}\|_{0, \Omega} + C \|q\|_{0, \Omega} \right) \|A^{-\frac{1}{2}} \nabla q\|_{0, \Omega} + \|\nabla \cdot \mathbf{v} + Xq\|_{0, \Omega} \|q\|_{0, \Omega}.

Combining the fact that $ab \leq \frac{1}{2} a^2 + \frac{1}{2} b^2$, we have

$$\|q\|^2_{1, \Omega} \leq C \|A^{-\frac{1}{2}} \nabla q\|^2_{0, \Omega} \leq C \left( G(\mathbf{v}, q; 0) + \|q\|^2_{0, \Omega} \right).$$  \hspace{1cm} (2.15)

(2.2), the triangle inequality, and (2.15) give

$$\|\mathbf{v}\|^2_{0, \Omega} \leq \frac{1}{\lambda} \|A^{\frac{1}{2}} \mathbf{v}\|^2_{0, \Omega} \leq \frac{2}{\lambda} \left( \|A^{\frac{1}{2}} \mathbf{v} + A^{-\frac{1}{2}} \nabla q\|^2_{0, \Omega} + \|A^{-\frac{1}{2}} \nabla q\|^2_{0, \Omega} \right)
\leq C \left( G(\mathbf{v}, q; 0) + \|q\|^2_{0, \Omega} \right).

By the triangle inequality and (2.15), we have

$$\|\nabla \cdot \mathbf{v}\|^2_{0, \Omega} \leq 2 \left( \|\nabla \cdot \mathbf{v} + Xq\|^2_{0, \Omega} + \|Xq\|^2_{0, \Omega} \right) \leq 2 \left( \|\nabla \cdot \mathbf{v} + Xq\|^2_{0, \Omega} + C \|q\|^2_{1, \Omega} \right)
\leq C \left( G(\mathbf{v}, q; 0) + \|q\|^2_{0, \Omega} \right).

Combining the above three inequalities yields (2.14).
With (2.14), we show the validity of (2.12) by the compactness argument. To this end, assume that (2.12) is not true. This implies that there exists a sequence \( \{v_n, q_n\} \in H_N(\text{div}; \Omega) \times V \) such that

\[
\|v_n\|_{H(\text{div})}^2 + \|q_n\|_{1, \Omega}^2 = 1 \quad \text{and} \quad G(v, q; 0) \leq \frac{1}{n} \quad (2.16)
\]

Since \( V \) is compactly contained in \( L^2(\Omega) \), there exists a subsequence \( \{p_{n_k}\} \subset V \) which converges in \( L^2(\Omega) \). For any \( k, l \) and \( (v_{n_k}, p_{n_k}), (v_{n_l}, p_{n_l}) \in H_N(\text{div}; \Omega) \times V \), it follows from (2.14) and the triangle inequality that

\[
\|v_{n_k} - v_{n_l}\|_{H(\text{div})}^2 + \|q_{n_k} - q_{n_l}\|_{1, \Omega}^2 \leq C (G(v_{n_k} - v_{n_l}, q_{n_k} - q_{n_l}; 0) + \|q_{n_k} - q_{n_l}\|_{0, \Omega}^2) \rightarrow 0.
\]

which implies that \( (v_{n_k}, p_{n_k}) \) is a Cauchy sequence in the complete space \( H_N(\text{div}; \Omega) \times V \). Hence, there exists \( (v, p) \in H_N(\text{div}; \Omega) \times V \) such that

\[
\lim_{k \to \infty} \left( \|v_{n_k} - v\|_{H(\text{div})} + \|q_{n_k} - q\|_{1, \Omega} \right) = 0.
\]

Next, we show that

\[
q = 0 \quad \text{and} \quad v = 0 \quad (2.17)
\]

which contradict with (2.16) that

\[
0 = \|v\|_{H(\text{div})}^2 + \|q\|_{1, \Omega}^2 = \lim_{k \to \infty} \|v_{n_k}\|_{H(\text{div})}^2 + \|q_{n_k}\|_{1, \Omega}^2 = 1.
\]

To this end, for any \( \phi \in V \), integration by parts and the Cauchy-Schwarz inequality give

\[
a(q_{n_k}, \phi) = (A\nabla q_{n_k}, \nabla \phi) + (Xq_{n_k}, \phi) = (A\nabla q_{n_k} + v_{n_k}, \nabla \phi) + (Xq_{n_k} + \nabla \cdot v_{n_k}, \phi)
\]

\[
\leq G(v_{n_k}, q_{n_k}; 0)^{\frac{1}{2}} \|\phi\|_{1, \Omega}.
\]

Since \( \lim q_{n_k} = q \) in \( V \), we then have

\[
|a(q, \phi)| = \lim_{k \to \infty} |a(q_{n_k}, \phi)| \leq \lim_{k \to \infty} G(v_{n_k}, q_{n_k}; 0)^{\frac{1}{2}} \|\phi\|_{1, \Omega} = 0.
\]

Because (2.3) has a unique solution, we have that

\[
q = 0.
\]

Now, \( v = 0 \) follows from (2.14):

\[
\|v\|_{H(\text{div})}^2 = \lim_{k \to \infty} \|v_{n_k}\|_{H(\text{div})}^2 \leq C \lim_{k \to \infty} (G(v_{n_k}, q_{n_k}; 0) + \|q_{n_k}\|_{0, \Omega}^2) = 0.
\]

This completes the proof of (2.17) and, hence, the theorem. ■
2.3 Inverse Norm Least-Squares Functional

For any \((v, q) \in H_N(\text{div}; \Omega) \times V \equiv \Phi\), consider the following least-squares functional:

\[
G(v, q; f) = \|A^{-\frac{1}{2}}(v + A\nabla q)\|_{0,\Omega}^2 + \|
abla \cdot v + Xq - f\|_{-1,D}^2.
\]  

(2.18)

Let

\[
\mathcal{K} = \begin{pmatrix}
I & 0 \\
0 & (-\Delta_D)^{-1}
\end{pmatrix}
\]  

(2.19)

where \((-\Delta_D)^{-1}\) is the solution operator of the Laplace equation with homogeneous Dirichlet boundary conditions on \(\Gamma_D\). Then the corresponding normal operator is

\[
\mathcal{L}^* \mathcal{K} \mathcal{L} = \begin{pmatrix}
A^{-1} - \nabla (-\Delta_D)^{-1} \text{div} & \nabla (I - (-\Delta_D)^{-1} X) \\
-(I - X^* (-\Delta_D)^{-1}) \text{div} & -\text{div} A \nabla + X^* (-\Delta_D)^{-1} X
\end{pmatrix}
\]  

(2.20)

with \(\mathcal{L}\) defined in (2.4) and the corresponding bilinear and linear forms are

\[
b(u, p; v, q) = (A^{-1}(u + A\nabla p), (v + A\nabla q)) \\
+ ((-\Delta_D)^{-1}(\nabla \cdot u + Xp), \nabla \cdot v + Xq)
\]  

(2.21)

\[
f(v, q) = ((-\Delta_D)^{-1} f, \nabla \cdot v + Xq),
\]  

(2.22)

respectively.

**Theorem 2.2** There exist positive constants \(\alpha_0\) and \(\alpha_1\) such that

\[
\alpha_0 \left(\|v\|^2 + \|q\|^2_{1,\Omega}\right) \leq b(v, q; v, q) = G(v, q; 0)
\]  

(2.23)

for any \((v, q) \in H_N(\text{div}; \Omega) \times V\) and

\[
b(u, p; v, q) \leq \alpha_1 \left(\|u\|^2 + \|p\|^2_{1,\Omega}\right)^{\frac{1}{2}} \left(\|v\|^2 + \|q\|^2_{1,\Omega}\right)^{\frac{1}{2}}
\]  

(2.24)

for any \((u, p)\) and any \((v, q)\) in \(H_N(\text{div}; \Omega) \times V\).

**Proof:** The theorem may be proved in a similar fashion as that of Theorem 2.1. 

This theorem gives the following equivalence:

\[
\mathcal{L}^* \mathcal{K} \mathcal{L} \sim \begin{pmatrix}
I & 0 \\
0 & -\Delta
\end{pmatrix}.
\]
2.4 Div-Curl Least-Squares Functional

We use the following space to define the div-curl least-squares functional for the extended system (2.6). Let

\[ H(\text{curl} A; \Omega) = \{ v \in L^2(\Omega)^d : \nabla \times (A^{-1}v) \in L^2(\Omega)^{2d-3} \}, \]  

which is a Hilbert space under the norm

\[ \|v\|_{H(\text{curl} A)} \equiv \left( \|v\|_{0,\Omega}^2 + \|\nabla \times (A^{-1}v)\|_{0,\Omega}^2 \right)^{\frac{1}{2}}. \]

When A is the identity matrix in (2.25), we use the simpler notation \( H(\text{curl}; \Omega) \). Define the subspaces

\[ H_D(\text{curl} A; \Omega) = \{ v \in H(\text{curl} A; \Omega) : n \times (A^{-1}v) = 0 \text{ on } \Gamma_D \}, \]

and

\[ W = H_N(\text{div}; \Omega) \cap H_D(\text{curl} A; \Omega). \]

For \( (v, q) \in W \times V = \Phi \), the div-curl least-squares functional is given by

\[ G(v, q; f) = \|A^{-\frac{1}{2}}(v - A\nabla q)\|_{0,\Omega}^2 + \|\nabla \cdot v + Xq - f\|_{0,\Omega}^2 + \|\nabla \times (A^{-1}v)\|_{0,\Omega}^2. \]  

The corresponding normal operator is

\[ L^* L = \begin{pmatrix} A^{-1} - \nabla \text{div} + A^{-1} \nabla \times \nabla \times A^{-1} & \nabla (I - X) \\ -(I - X^*) \text{div} & -\text{div} A \nabla + X^* X \end{pmatrix} \]  

(2.27)

with \( L \) defined in (2.6) and the corresponding normal operator and bilinear and linear forms are

\[ b(u, p; v, q) = (A^{-1}(u + A\nabla p), v + A\nabla q) + (\nabla \cdot u + Xp, \nabla \cdot v + Xq) \]

\[ + (\nabla \times (A^{-1}u), \nabla \times (A^{-1}v)) \]  

(2.28)

\[ f(v, q) = (f, \nabla \cdot v + Xq), \]  

(2.29)

respectively. It follows from Theorem 2.1 that we have the following equivalence:

\[ L^* L \sim \begin{pmatrix} I - \nabla \text{div} + A^{-1} \nabla \times \nabla \times A^{-1} & 0 \\ 0 & -\Delta \end{pmatrix} \sim \begin{pmatrix} -\Delta & 0 \\ 0 & -\Delta \end{pmatrix}. \]  

(2.30)

The second equivalence requires sufficient smoothness of coefficients and boundary (see [7] for the proof).
\textbf{Theorem 2.3} There exist positive constants $\alpha_0$ and $\alpha_1$ such that
\[
\alpha_0 \left( \|v\|_{H^1(\text{div},\Omega)}^2 + \|\nabla \times (A^{-1}v)\|_{0,\Omega}^2 + \|q\|_{1,\Omega}^2 \right) \leq b(v, q; v, q) \quad (2.31)
\]
for any $(v, q) \in W \times V$ and
\[
b(u, p; v, q) \leq \alpha_1 \left( \|u\|_{H^1(\text{div},\Omega)}^2 + \|\nabla \times (A^{-1}u)\|_{0,\Omega}^2 + \|p\|_{1,\Omega}^2 \right)^{\frac{1}{2}}
\cdot \left( \|v\|_{H^1(\text{div},\Omega)}^2 + \|\nabla \times (A^{-1}v)\|_{0,\Omega}^2 + \|q\|_{1,\Omega}^2 \right)^{\frac{1}{2}}\quad (2.32)
\]
for any $(u, p), (v, q) \in W \times V$.

\section*{2.5 Least-Squares Problems}

For the solution space $\Phi$, we have the following equivalent least-squares problems:
\textbf{minimization problem}: find $(u, p) \in \Phi$ such that
\[
G(u, p; f) = \min_{(v, q) \in \Phi} G(v, q; f); \quad (2.33)
\]
\textbf{variational problem}: find $(u, p) \in \Phi$ such that
\[
b(u, p; v, q) = f(v, q) \quad \forall (v, q) \in \Phi. \quad (2.34)
\]

\section*{2.6 Least-Squares Approximation}

In this subsection, we consider least-squares finite element approximation only based on the div least-squares functional. Approximation based on the div-curl least-squares functional may be studied in a similar fashion. There are two numerical approximations based on the inverse norm least-squares functional: (1) mesh-dependent norm approach in [1] and (2) the discrete $H^{-1}$ norm approach in [4].

Assume that $\Omega$ is a polygonal domain, let $T_h$ be a quasi-regular triangulation of $\Omega$ with (triangular/tetrahedra or rectangular) elements of size $O(h)$. Denote spaces of polynomials on an element $K \subset \mathcal{R}^d$:
\[
P_k(K) \text{ is the space of polynomials of degree } \leq k;
\]

\[
P_{k_1,k_2}(K) = \{p(x_1, x_2) : p(x_1, x_2) = \sum_{i \leq k_1, j \leq k_2} a_{ij} x_1^i x_2^j\}, \quad d = 2
\]

\[
P_{k_1,k_2,k_3}(K) = \{p(x_1, x_2, x_3) : p(x_1, x_2, x_3) = \sum_{i \leq k_1, j \leq k_2, k \leq k_3} a_{ijk} x_1^i x_2^j x_3^k\}, \quad d = 3.
\]

Denote the local Raviart-Thomas (RT) space of index $k \geq 0$ on an element $K$:
\[
RT_k(K) = \begin{cases} 
P_k(K)^d + (x_1, \ldots, x_d)P_k(K), & K = \text{triangle/tetrahedra} \\
P_{k+1,k}(K) \times P_{k,k+1}(K), & K = \text{rectangle }, d = 2 \\
P_{k+1,k,k}(K) \times P_{k,k+1,k}(K) \times P_{k,k,k+1}(K), & K = \text{rectangle }, d = 3
\end{cases}
\]
Degrees of freedom for \( RT_0(K) = (a + bx_1, c + bx_2) \) on triangle or \( RT_0(K) = (a + bx_1, c + dx_2) \) on rectangle are normal components of vector field on all edges (faces) of two- (three-) dimensional elements. See \([5]\) for the choice of degrees of freedom for the \( RT_k \) space of index \( k \geq 1 \). They are chosen for ensuring continuity of the normal component of vector field at interfaces of elements. Then one can define the \( H(\text{div}; \Omega) \) conforming Raviart-Thomas space of order \( k \geq 0 \) \([17]\) by

\[
RT_k = \{ \mathbf{v} \in H(\text{div}; \Omega) : \mathbf{v}|_K \in RT_k(K) \, \forall \, K \in \mathcal{T}_h \},
\]

which has the approximation property:

\[
\inf_{\phi \in RT_k} \| \mathbf{v} - \phi \|_{0,\Omega} \leq C h^r \| \mathbf{v} \|_{r,\Omega} \quad \text{for} \quad 1 \leq r \leq k + 1 \tag{2.35}
\]

\[
\inf_{\phi \in RT_k} \| \nabla \cdot (\mathbf{v} - \phi) \|_{0,\Omega} \leq C h^r \| \nabla \cdot \mathbf{v} \|_{r,\Omega} \quad \text{for} \quad 0 \leq r \leq k + 1. \tag{2.36}
\]

Denote the space of continuous piecewise polynomials of degree \( \leq k \) by

\[
S_k = \{ q \in H^1(\Omega) : q|_K \in P_k(K) \, \forall \, T \in \mathcal{T}_h \},
\]

which has the following approximation property:

\[
\inf_{\phi \in S_k} (\| q - \phi \|_{0,\Omega} + h \| q - \phi \|_{1,\Omega}) \leq C h^{r+1} \| q \|_{r+1,\Omega} \quad \text{for} \quad 0 \leq r \leq k + 1. \tag{2.37}
\]

Then least-squares approximation is to find \((\mathbf{u}^h, p^h) \in RT_k \times S_k\) such that

\[
b(\mathbf{u}^h, p^h; \mathbf{v}, q) = f(\mathbf{v}, q) \quad \forall \, (\mathbf{v}, q) \in RT_k \times S_k. \tag{2.38}
\]

**Theorem 2.4** Let \((\mathbf{u}, p)\) and \((\mathbf{u}^h, p^h)\) be the solutions of (2.34) and (2.38), respectively. Then we have the following error estimation:

\[
\| \mathbf{u} - \mathbf{u}^h \|_{H(\text{div})} + \| p - p^h \|_{1,\Omega} \leq C \frac{\alpha_1}{\alpha_0} h^r (\| p \|_{r+1,\Omega} + \| \mathbf{u} \|_{r,\Omega} + \| \nabla \cdot \mathbf{u} \|_{r,\Omega}) \leq C \frac{\alpha_1}{\alpha_0} h^r (\| p \|_{r+1,\Omega} + \| f \|_{r,\Omega}). \tag{2.39}
\]

**Proof:** (2.39) follows from Theorem 1.1, the approximation properties in (2.35), (2.36), and (2.37), and the facts that

\[
\| \mathbf{u} \|_{r,\Omega} \leq C \| p \|_{r+1,\Omega}
\]

and that

\[
\| \nabla \cdot \mathbf{u} \|_{r,\Omega} = \| f - Xp \|_{r,\Omega} \leq \| f \|_{r,\Omega} + C \| p \|_{r+1,\Omega}.
\]
2.7 Comparison of Least-Squares Methods

In this section, we make simple comparison of least-squares methods. The div least-squares method has the following numerical properties:

+ optimal finite element approximation;
+ optimal fast multigrid solver if Raviart-Thomas elements are used for the flux.

The mesh dependent least-squares method has the following properties:

+ optimal finite element approximation;
– unknown fast iterative solver.

The discrete $H^{-1}$ norm least-squares method has the following properties:

+ optimal finite element approximation;
+ uniformly well preconditioned by multigrid or domain decomposition;
– expensive evaluations of the discrete $H^{-1}$ norm.

The div-curl least-squares method has the following properties:

+ finite element approximations are $H^1$-optimally accurate in each variable (including new variables);
+ standard multigrid methods applied to the resulting discrete equation have optimal complexity;
– additional smoothness of the original problem is required for the second equivalence in (2.30).

2.8 Boundary Least-Squares Functional

Denote by $H^{-\frac{1}{2}}(\partial \Omega)$ the dual space of $H^{\frac{1}{2}}(\partial \Omega)$ with the dual norm

$$\|v\|_{-\frac{1}{2}, \partial \Omega} = \sup_{q \in H^{\frac{1}{2}}(\partial \Omega)} \frac{\langle v, q \rangle}{\|q\|_{\frac{1}{2}, \partial \Omega}},$$

where the bracket $\langle q, v \rangle$ denotes duality between $H^{-\frac{1}{2}}(\partial \Omega)$ and $H^{\frac{1}{2}}(\partial \Omega)$. When $\Gamma_N \neq \partial \Omega = \Gamma_N \cup \Gamma_D$, denote by $H^{-\frac{1}{2}}(\Gamma_N)$ the dual space of $H^{\frac{1}{2}}_0(\Gamma_N) = \{v|_{\Gamma_N} : v \in H^{\frac{1}{2}}_{0,0}(\Omega)\}$. In this section, we need the generalized Poincaré-Friedrichs inequality

$$\|q\|_{1, \Omega} \leq C (\|\nabla q\|_{0, \Omega} + \|q\|_{0, \Gamma_D}) \quad \forall q \in H^1(\Omega) \quad \text{if mes}(\Gamma_D) \neq 0,$$

$$\|q\|_{1, \Omega} \leq C \|\nabla q\|_{0, \Omega} \quad \forall q \in \hat{H}^1(\Omega) \quad \text{otherwise}; \quad (2.40)$$
and the trace inequalities for any subset $\Gamma \subset \partial \Omega$ with positive measure
\[
\|q\|_{\frac{1}{2},\Gamma} \leq \|q\|_{1,\Omega} \quad \forall q \in H^1(\Omega),
\]
\[
\|n \cdot v\|_{-\frac{1}{2},\Gamma} \leq \|v\|_{H(\text{div}; \Omega)} \quad \forall v \in H(\text{div}; \Omega). \tag{2.41}
\]

The first inequality in (2.41) follows from the definition. The second inequality in (2.41) follows from the definition, the Green’s formula, and the Cauchy-Schwarz inequality: for any $q \in H^1(\Omega)$ and $q = 0$ on $\Gamma' = \partial \Omega \setminus \Gamma$
\[
\frac{\int_{\Gamma} q \cdot n \cdot v \, ds}{\|q\|_{\frac{1}{2},\partial \Omega}} \leq \frac{\int_{\partial \Omega} q \cdot n \cdot v \, ds}{\|q\|_{1,\Omega}} = \frac{\int_{\Omega} \nabla \cdot v \, dx + \int_{\Omega} v \cdot \nabla q \, dx}{\|q\|_{1,\Omega}} \leq \|v\|_{H(\text{div})}.
\]

Considering non-homogeneous boundary conditions:

\[
p = g \quad \text{on } \Gamma_D \quad \text{and} \quad -n \cdot A \nabla p = h \quad \text{on } \Gamma_N
\]

and the following least-squares functional:

\[
G(v, q; \hat{f}) = \|A^{-\frac{1}{2}}(v + A \nabla q)\|_{0, \Omega}^2 + \|\nabla \cdot v + Xq - f\|_{0, \Omega}^2
\]
\[
+ \|p - g\|_{\frac{1}{2}, \Gamma_D}^2 + \|n \cdot v - h\|_{-\frac{1}{2}, \Gamma_N}^2 \tag{2.42}
\]

for $(v, q) \in H(\text{div}; \Omega) \times H^1(\Omega)$, where $\hat{f} = (f, g, h)$. Then the least-squares problem for (2.4) is to minimize this quadratic functional over $H(\text{div}; \Omega) \times H^1(\Omega)$: find $(u, p) \in H(\text{div}; \Omega) \times H^1(\Omega)$ such that

\[
G(u, p; \hat{f}) = \inf_{(v, q) \in H(\text{div}; \Omega) \times H^1(\Omega)} G(v, q; \hat{f}). \tag{2.43}
\]

It is easy to see that the variational form for (2.43) is to find $(u, p) \in H(\text{div}; \Omega) \times H^1(\Omega)$ such that

\[
b(u, p; v, q) = f(v, q), \quad \forall (v, q) \in H(\text{div}; \Omega) \times H^1(\Omega), \tag{2.44}
\]

where the bilinear form $b(\cdot, \cdot) : (H(\text{div}; \Omega) \times H^1(\Omega))^2 \rightarrow \mathbb{R}$ is defined by

\[
b(u, p; v, q) = \langle A^{-1}(u + A \nabla p), v + A \nabla q \rangle_{0, \Omega} + \langle \nabla \cdot u + Xp, \nabla \cdot v + Xq \rangle_{0, \Omega}
\]
\[
+ \langle p, q \rangle_{\frac{1}{2}, \Gamma_D} + \langle n \cdot u, n \cdot v \rangle_{-\frac{1}{2}, \Gamma_N}
\]

and the linear form $f(\cdot, \cdot) : H(\text{div}; \Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ is defined by

\[
f(v, q) = (f, \nabla \cdot v + Xq)_{0, \Omega} + \langle g, q \rangle_{\frac{1}{2}, \Gamma_D} + \langle h, n \cdot v \rangle_{-\frac{1}{2}, \Gamma_N}.
\]

**Theorem 2.5** Then there exist positive constants $\alpha_0$ and $\alpha_1$ such that

\[
\alpha_0 \left( \|v\|_{H(\text{div})}^2 + \|q\|_{1, \Omega}^2 \right) \leq b(v, q; v, q) \tag{2.45}
\]

for any $(v, q) \in H(\text{div}; \Omega) \times H^1(\Omega)$ and

\[
b(u, p; v, q) \leq \alpha_1 \left( \|u\|_{H(\text{div})}^2 + \|p\|_{1, \Omega}^2 \right)^{\frac{1}{2}} \left( \|v\|_{H(\text{div})}^2 + \|q\|_{1, \Omega}^2 \right)^{\frac{1}{2}} \tag{2.46}
\]

for any $(u, p), (v, q) \in H(\text{div}; \Omega) \times H^1(\Omega)$.
Proof: The continuity of the bilinear form $b(\cdot; \cdot)$ in (2.46) is an immediate consequence of the Cauchy-Schwarz and trace inequalities. To show the validity of the coercivity of the bilinear form in (2.45), it suffices to prove that

$$\|v\|_{H_0^{\text{div}}(\Omega)}^2 + \|q\|_{L^2(\Omega)}^2 \leq C \left(b(v, q; v, q) + \|q\|_{L^2(\Omega)}^2\right)$$

(2.47)

because (2.45) then follows from a standard compactness argument (see the proof of Theorem 2.1). To this end, first note that, using the triangle and trace inequalities,

$$|\int_{\partial\Omega} q \cdot v \, ds| \leq \int_{\Gamma_D} q \cdot v \, ds + \int_{\Gamma_N} q \cdot v \, ds$$

$$\leq \|q\|_{L^2(\Gamma_D)} \|n \cdot v\|_{H^{1/2}(\Gamma_D)} + \|q\|_{L^2(\Gamma_N)} \|n \cdot v\|_{H^{1/2}(\Gamma_N)}$$

$$\leq \|q\|_{L^2(\Gamma_D)} \|v\|_{H^{1/2}(\Gamma_D)} + \|q\|_{L^2(\Gamma_N)} \|n \cdot v\|_{H^{1/2}(\Gamma_N)}$$

$$\leq C \|q\|_{L^2(\Gamma_D)} \|A^{-1/2} v\|_{L^2(\Omega)} + C \|A^{1/2} \nabla q\|_{L^2(\Omega)} \left(\|q\|_{H^{1/2}(\Gamma_D)} + \|n \cdot v\|_{H^{-1/2}(\Gamma_N)}\right) + b(v, q; v, q).$$

The triangle inequality gives that

$$\|A^{-1/2} v\|_{L^2(\Omega)} \leq \|A^{-1/2} (v + A\nabla q)\|_{L^2(\Omega)} + \|A^{1/2} \nabla q\|_{L^2(\Omega)},$$

(2.48)

which, together with the above inequality, implies that

$$|\int_{\partial\Omega} q \cdot v \, ds| \leq C \left(\|q\|_{L^2(\Gamma_D)} + \|n \cdot v\|_{H^{-1/2}(\Gamma_N)}\right) \|A^{1/2} \nabla q\|_{L^2(\Omega)} + C b(v, q; v, q).$$

(2.49)

It follows from integration by parts, the Cauchy-Schwarz and Poincaré-Friedrichs inequalities, and (2.49) that

$$\|A^{1/2} \nabla q\|_{L^2(\Omega)}^2 = (A^{-1/2} (A \nabla q + v), A^{1/2} \nabla q)_{L^2(\Omega)} + (q, \nabla \cdot v)_{L^2(\Omega)} - \int_{\partial\Omega} q \cdot v \, ds$$

$$\leq \|A^{-1/2} (A \nabla q + v)\|_{L^2(\Omega)} \|A^{1/2} \nabla q\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)} \|\nabla \cdot v\|_{L^2(\Omega)} - \int_{\partial\Omega} q \cdot v \, ds$$

$$\leq \left(\|A^{-1/2} (A \nabla q + v)\|_{L^2(\Omega)} + \|\nabla \cdot v + X q\|_{L^2(\Omega)} + \|n \cdot v\|_{H^{-1/2}(\Gamma_N)} + \|q\|_{L^2(\Gamma_D)}\right) \|A^{1/2} \nabla q\|_{L^2(\Omega)}$$

$$+ \|q\|_{L^2(\Omega)} \|X q\|_{L^2(\Omega)} + C b(v, q; v, q).$$

Hence,

$$\|X q\|_{L^2(\Omega)}^2 \leq C \|A^{1/2} \nabla q\|_{L^2(\Omega)}^2 \leq C \left(b(v, q; v, q) + \|q\|_{L^2(\Omega)}^2\right).$$

Combining with (2.48) and the triangle inequality yields

$$\|A^{-1/2} v\|_{L^2(\Omega)}^2 + \|\nabla \cdot v\|_{L^2(\Omega)}^2 \leq C \left(b(v, q; v, q) + \|q\|_{L^2(\Omega)}^2\right).$$

This completes the proof of (2.47) and, hence, theorem.

For numerical approach based on the boundary least-squares functional, see [18].
References


Homework

Consider problem (2.1) with $Xp = b \cdot \nabla p + cp$, study dependence of constants $\alpha_0$ and $\alpha_1$ in Theorem 2.1 on the diffusion coefficients $A$, convection coefficients $b$, and reaction coefficient $c$ for the following cases:

- $A = a(x) I$, $b = 0$, and $c = 0$.
- $A = I$, $b = b \left( \begin{array}{c} 1 \\ 1 \end{array} \right)$, and $c = 0$, where $b$ is a constant.
- $A = I$, $b = 0$, and $c = -\omega$, where $\omega > 0$ is a constant.