LEAST-SQUARES FINITE ELEMENT METHODS FOR THE POISSON EQUATION AND THEIR CONNECTION TO THE DIRICHLET AND KELVIN PRINCIPLES

Max Gunzburger

*School of Computational Science*
*Florida State University*

*Joint work with*

Pavel Bochev

*Computational Mathematics and Algorithms Department*
*Sandia National Laboratories*

Last updated in 2007
WHY LEAST SQUARES?

- Finite element methods were first developed and analyzed in the Rayleigh-Ritz setting
  - i.e., problems whose solutions can be characterized as minimizers of convex, quadratic functionals
  - examples include
    - the equations of linear elasticity
    - the Poisson equation
      \[-\Delta \phi = f \text{ in } \Omega \quad \text{and} \quad \phi = 0 \text{ on } \Gamma = \partial \Omega\]
      whose solution \( \phi \) can be characterized as the minimizer, over a suitable class of functions, of the functional
      \[
      \frac{1}{2} \int_{\Omega} |\nabla \phi|^2 \, d\Omega - \int_{\Omega} f \phi \, d\Omega
      \]
• The Rayleigh-Ritz setting results in finite element methods having the following desirable features

1. general regions and boundary conditions are relatively easy to handle and higher-order accuracy is relatively easy to achieve

2. the conformity of the finite element spaces are sufficient to guarantee stability and optimal accuracy

3. all variables can be approximated using a single type of finite element space, i.e., based on the same grid and same degree polynomials

4. the resulting linear systems are
   a) sparse;  b) symmetric;  c) positive definite

• FEMs were quickly applied in other settings,
  – motivated by the fact that properties 1 and 4a are retained for all FEMs
• **Mixed** FEMs arose from minimization problems constrained by PDEs
  – result in saddle-point problems
  – the only other property retained from the Rayleigh-Ritz setting is *symmetry*
  – onerous compatibility conditions between finite element spaces often arise

• More generally, **Galerkin** FEMs are defined by forcing the *residual of the PDE* to be *orthogonal* to the finite element subspace
  – *none* of the other advantages of the Rayleigh-Ritz setting are retained

• It is a testament to the importance of advantage 1 that despite the loss of other advantages, mixed and Galerkin FEMs are in widespread use
• Not surprisingly, despite the success of mixed and Galerkin FEMs, there has been substantial interest and effort devoted to developing finite element approaches that recover at least some of the advantages of the Rayleigh-Ritz setting
  – stabilized FEMs
  – penalty FEMs

• Least-squares finite element (LSFEMs) methods can be viewed as another attempt at retaining the advantages of the Rayleigh-Ritz setting even for much more general problems
  – in fact, they offer the possibility of, in principle, retaining all of the advantages of that setting for practically any PDE problem
• LSFEMs possess two additional advantages over other FEMs, even in the Rayleigh-Ritz setting
  – easily computable residual error indicators are available
    - the least-squares functional provides element-wise equation residual estimators
  – the treatment of general, inhomogeneous boundary conditions is facilitated
    - all boundary conditions can be made “natural” to the variational principle

• Essential difference between Galerkin and least-squares methods
  – Galerkin finite element methods ⇐ residual orthogonalization
  – least-squares finite element methods ⇐ residual minimization
Straightforward LSFEM

- Consider the problem

\[ \mathcal{L}u = f \quad \text{in } \Omega \quad \text{and} \quad \mathcal{R}u = g \quad \text{on } \Gamma \]

- We assume nothing about this problem other than it is well posed

  - there exists

    a solution Hilbert space \( S \)
    
    data Hilbert spaces \( H_\Omega \) and \( H_\Gamma \)
    
    positive constants \( \alpha_1 \) and \( \alpha_2 \)
    
    such that

\[ \alpha_1 \| u \|^2_S \leq \| \mathcal{L}u \|^2_{H_\Omega} + \| \mathcal{R}u \|^2_{H_\Gamma} \leq \alpha_2 \| u \|^2_S \quad \forall u \in S \]
Now, consider the least-squares functional

\[ J(u; f, g) = \| \mathcal{L}u - f \|_{H_\Omega}^2 + \| \mathcal{R}u - g \|_{H_\Gamma}^2 \]

and the unconstrained minimization problem

\[ \min_{u \in S} J(u; f, g) \]

so that

\[ (\mathcal{L}v, \mathcal{L}u)_{H_\Omega} + (\mathcal{R}v, \mathcal{R}u)_{H_\Gamma} = (\mathcal{L}v, f)_{H_\Omega} + (\mathcal{R}v, g)_{H_\Gamma} \quad \forall v \in S \]

– note that

the spaces used to measure the sizes of the residuals are the data spaces \( H_\Omega \) and \( H_\Gamma \)

and

the space in which candidate minimizers are sought is the solution space \( S \)

for which the PDE problem is well posed.
• A LSFEM can be defined by
  - choosing a finite element subspace \( S^h \subset S \)
  - then restricting the minimization problem to the subspace

  – thus, the LSFEM approximation \( u^h \in S^h \) is the solution of the problem
  \[
  \min_{u^h \in S^h} J(u^h; f, g)
  \]

• Since we have recast the problem as an unconstrained quadratic optimization problem,
  
  all the desirable properties of the Rayleigh-Ritz setting are recovered!

• The key is the **norm equivalence** of the functional

  \[
  \alpha_1 \|u\|_S^2 \leq J(u; 0, 0) = \|\mathcal{L}u\|_{H_\Omega}^2 + \|\mathcal{R}u\|_{H_\Gamma}^2 \leq \alpha_2 \|u\|_S^2 \quad \forall u \in S
  \]
• This is not the whole story: additional practicality criteria should be met
  A. bases are easily constructed
  B. linear systems are easily assembled
  C. linear systems are relatively well conditioned

• Two keys to practicality
  – use first-order system form of PDEs
  – use $L^2$ norms in the functional

• The choices (operators and spaces) that go into defining the straightforward LSFEM are often in conflict with practicality
  – i.e., there is often a conflict between norm equivalence and practicality
  – this conflict has been largely resolved; see our upcoming Springer book
Today’s talk

• We consider mixed and least-squares finite element methods for the first-order system

\[ \mathbf{u} + \nabla \phi = 0 \quad \text{and} \quad \nabla \cdot \mathbf{u} = f \]

which, of course, is equivalent to the Poisson equation

\[ -\Delta \phi = f \]

• This is a prototype for more general elliptic equations and Darcy flows

• Reasons for considering the first-order formulation include

  – the flux variable \( \mathbf{u} \) is often the primary variable of interest

  – it may be easier to apply Dirichlet boundary conditions
**Notation**

- \( \Omega \) denotes a bounded region in \( \mathbb{R}^n, n = 2, 3 \), with boundary \( \Gamma \)
  - \( \Gamma \) consists of two disjoint parts denoted by \( \Gamma_D \) and \( \Gamma_N \)

![Diagram of \( \Omega \) with \( \Gamma_D \) and \( \Gamma_N \) as disjoint parts]

- Usual Sobolev spaces and some not so usual spaces

\[
H_N(\Omega, \text{div}) = \{ u \in L^2(\Omega) \mid \nabla \cdot u \in L^2(\Omega), \ u \cdot n = 0 \text{ on } \Gamma_N \}
\]

\[
H_D(\Omega, \text{grad}) = \{ \phi \in L^2(\Omega) \mid \nabla \phi \in L^2(\Omega), \ \phi = 0 \text{ on } \Gamma_D \}
\]

- more common notation:

\[
H_D(\Omega, \text{grad}) = H^1_D(\Omega) = \{ \phi \in H^1(\Omega) \mid \phi = 0 \text{ on } \Gamma_D \}
\]
• Norms

\[ \|\phi\|_{H_N(\Omega, \text{div})} = \left( \|u\|_0^2 + \|\nabla \cdot u\|_0^2 \right)^{1/2} \quad \text{for } u \in H_N(\Omega, \text{div}) \]

\[ \|u\|_{H_D(\Omega, \text{grad})} = \left( \|\phi\|_0^2 + \|\nabla \phi\|_0^2 \right)^{1/2} \quad \text{for } \phi \in H_D(\Omega, \text{grad}) \]

• We will also use the spaces

\[ \nabla \cdot \left( H_N(\Omega, \text{div}) \right) = L^2(\Omega) \]

\[ \nabla \left( H_D(\Omega, \text{grad}) \right) = \{ u \in L^2(\Omega) \mid \nabla \times u = 0 \} \subset L^2(\Omega) \]

• With \( \gamma \geq 0 \) a given function belonging to \( L^\infty(\Omega) \), we will refer to the problem

\[ -\Delta \phi + \gamma \phi = f \quad \text{in } \Omega, \quad \phi = 0 \quad \text{on } \Gamma_D, \quad \text{and} \quad \partial \phi / \partial n = 0 \quad \text{on } \Gamma_N \]

as the Poisson problem
The Dirichlet and Kelvin principles arise in a variety of applications.

Mathematically, they provide two variational formulations for the Poisson problem.

They also form the basis for defining mixed Galerkin finite element methods for approximations of the solution of the Poisson problem.
The generalized Dirichlet principle

\[
\begin{cases}
\min_{(\phi, w) \in H_D(\Omega, \text{grad}) \times L^2(\Omega)} D(\phi, w; f) \quad \text{subject to} \quad w + \nabla \phi = 0,
\end{cases}
\]

where

\[
D(\phi, w; f) = \frac{1}{2} \int_{\Omega} \left( |w|^2 + \gamma |\phi|^2 \right) d\Omega - \int_{\Omega} f \phi d\Omega
\]

• For \( f = 0, \gamma = 0 \), and appropriate boundary conditions, the Dirichlet principle in the inviscid fluid mechanics setting implies that among all irrotational velocity fields, the one that minimizes the kinetic energy is the solenoidal one.

• In the solid mechanics setting, \( \phi \) is a vector and \( w \) is a tensor and a simplified version of the Dirichlet principle is the potential energy minimization principle.
• \( \phi = 0 \) on \( \Gamma_D \) is an essential boundary condition for the Dirichlet principle

• \( u \cdot n = 0 \) on \( \Gamma_N \) is a natural boundary condition for the Dirichlet principle

• The constraint \( w + \nabla \phi = 0 \) can be directly substituted into the functional to eliminate the flux \( w \) to yield the certainly more familiar form of the (generalized) Dirichlet principle:

\[
\begin{align*}
\min_{\phi \in H^1_D(\Omega)} & \quad \tilde{D}(\phi; f), \\
\text{where} & \quad \tilde{D}(\phi; f) = \frac{1}{2} \int_{\Omega} \left( |\nabla \phi|^2 + \gamma |\phi|^2 \right) \, d\Omega - \int_{\Omega} f \phi \, d\Omega
\end{align*}
\]

• Although one can eliminate the vector field \( w \), it will be more profitable for our discussions to continue to consider the first form of the Dirichlet principle

• We apply the Lagrange multiplier rule to enforce the constraint \( w + \nabla \phi = 0 \) appearing in the Dirichlet principle; \( u \) denotes the Lagrange multiplier
• The first-order necessary conditions reduce to \( w = u \) and:

seek \((\phi, u) \in H_D(\Omega, \text{grad}) \times L^2(\Omega)\) such that

\[
\begin{aligned}
\int_{\Omega} u \cdot v \, d\Omega + \int_{\Omega} \nabla \phi \cdot v \, d\Omega &= 0 \quad \forall v \in L^2(\Omega) \\
\int_{\Omega} \nabla \psi \cdot u \, d\Omega - \int_{\Omega} \gamma \psi \phi \, d\Omega &= -\int_{\Omega} f \psi \, d\Omega \quad \forall \psi \in H_D(\Omega, \text{grad}) = H_D^{1}(\Omega)
\end{aligned}
\]

• We may choose \( v = \nabla \psi \in L^2(\Omega) \) to obtain the more familiar weak formulation

\[
\int_{\Omega} \nabla \phi \cdot \nabla \psi \, d\Omega + \int_{\Omega} \gamma \psi \phi \, d\Omega = \int_{\Omega} f \psi \, d\Omega \quad \forall \psi \in H_D^{1}(\Omega)
\]

for the Poisson problem
If solutions of the Dirichlet principle are sufficiently smooth, then without much difficulty one obtains that

\[
\begin{align*}
\nabla \cdot \mathbf{u} + \gamma \phi &= f \quad \text{and} \quad \mathbf{u} + \nabla \phi = 0 \quad \text{in } \Omega \\
\phi &= 0 \quad \text{on } \Gamma_D \quad \text{and} \quad \mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_N
\end{align*}
\]

Eliminating the flux \( \mathbf{u} \) (again assuming that sufficient smoothness is available), one obtains the second-order Poisson problem for \( \phi \).
The generalized Kelvin principle

\[
\begin{align*}
\min_{(u,\lambda) \in H^N(\Omega,\text{div}) \times L^2(\Omega)} & K(\lambda, v) \quad \text{subject to} \quad \nabla \cdot u + \gamma \lambda = f, \\
\text{where} & \\
K(\lambda, u) &= \frac{1}{2} \int_{\Omega} (|u|^2 + \gamma |\lambda|^2) \, d\Omega.
\end{align*}
\]

- For \( f = 0, \gamma = 0, \) and appropriate boundary condition, the Kelvin principle for inviscid flows implies that among all incompressible velocity fields, the one that minimizes the kinetic energy is irrotational.

- In structural mechanics (with \( u \) is a tensor and \( \lambda \) is a vector), a simplified version of the Kelvin principle is known as the minimum complimentary energy principle.
\( \mathbf{u} \cdot \mathbf{n} = 0 \) on \( \Gamma_N \) is an essential boundary condition for the Kelvin principle.

\( \phi = 0 \) on \( \Gamma_D \) is a natural boundary condition for the Kelvin principle.

We can apply the Lagrange multiplier rule to enforce the constraint \( \nabla \cdot \mathbf{u} + \gamma \lambda = 0 \); now \( \phi \) is the Lagrange multiplier.

The first-order necessary conditions reduce to \( \lambda = \phi \) and:

seek \( (\mathbf{u}, \phi) \in \mathbf{H}_N(\Omega, \text{div}) \times L^2(\Omega) \) such that

\[
\begin{cases}
- \int_\Omega \psi \nabla \cdot \mathbf{u} \, d\Omega - \int_\Omega \gamma \psi \phi \, d\Omega = - \int_\Omega f \psi \, d\Omega & \forall \psi \in L^2(\Omega) \\
\int_\Omega \mathbf{u} \cdot \mathbf{v} \, d\Omega - \int_\Omega \phi \nabla \cdot \mathbf{v} \, d\Omega = 0 & \forall \mathbf{v} \in \mathbf{H}_N(\Omega, \text{div})
\end{cases}
\]

We again have that if solutions of the Kelvin principle are sufficiently smooth, then they are also solutions of the Poisson problem.
To define a discretization of the Dirichlet principle, one first chooses finite dimensional spaces $S_D^h$ and $V_D^h$ then one seeks $\phi^h \in S_D^h$ and $u^h \in V_D^h$ such that

$$\begin{align*}
\int_{\Omega} u^h \cdot v^h d\Omega + \int_{\Omega} \nabla \phi^h \cdot v^h d\Omega &= 0 \quad \forall v^h \in V_D^h \\
\int_{\Omega} \nabla \psi^h \cdot u^h d\Omega - \int_{\Omega} \gamma \psi^h \phi^h d\Omega &= -\int_{\Omega} f \psi^h d\Omega \quad \forall \psi^h \in S_D^h
\end{align*}$$

there remains the issue of choosing $S_D^h$ and $V_D^h$ so that

- we have a stable discretization method
- we obtain optimally accurate approximate solutions
• For the Dirichlet principle, it is natural to choose conforming subspaces

\[ S^h_D \subset H_D(\Omega, \text{grad}) = H^1_D(\Omega) \quad \text{and} \quad V^h_D \subset L^2(\Omega) \]

– however, in general, these inclusions are not by themselves sufficient to guarantee that the approximations are stably defined

– \( S^h_D \) and \( V^h_D \) also must satisfy

\[
\inf_{u^h \in V^h_D} \sup_{\phi^h \in S^h_D} \frac{\int_{\Omega} u^h \cdot \nabla \phi^h \, d\Omega}{\|\phi^h\|_1 \|u^h\|_0} \geq \alpha_D > 0
\]
• To define a discretization of the Kelvin principle,

- one first chooses finite dimensional spaces $V^h_K$ and $S^h_K$

- then one seeks $u^h \in V^h_K$ and $\phi^h \in S^h_K$ such that

\[
\begin{aligned}
\begin{cases}
\int_{\Omega} u^h \cdot v^h \, d\Omega - \int_{\Omega} \phi^h \nabla \cdot v^h \, d\Omega = 0 & \forall v^h \in V^h_K \\
\int_{\Omega} \psi^h \nabla \cdot u^h \, d\Omega + \int_{\Omega} \gamma \psi^h \phi^h \, d\Omega = \int_{\Omega} f \psi^h \, d\Omega & \forall \psi^h \in S^h_K
\end{cases}
\end{aligned}
\]

- there again remains the issue of choosing $V^h_K$ and $S^h_K$ so that
  - we have a stable discretization method
  - we obtain optimally accurate approximate solutions
• For the Kelvin principle, it is natural to choose conforming subspaces

\[ V_h^K \subset H_N(\Omega, \text{div}) \quad \text{and} \quad S_h^K \subset L^2(\Omega) \]

– however, in general, these inclusions are not by themselves sufficient to guarantee that the approximations are stably defined

– \( V_h^K \) and \( S_h^K \) also must satisfy

\[
\inf_{\phi^h \in S_h^K} \sup_{u^h \in V_h^K} \frac{\int_{\Omega} \phi^h \nabla \cdot u^h \, d\Omega}{\|u^h\|_{H_N(\Omega, \text{div})} \|\phi^h\|_0} \geq \alpha_K > 0
\]
Stable finite element methods for the Dirichlet principle

• Let

\[ \mathcal{T}_h \] denote a simplicial triangulation of \( \Omega \) into elements

\[ \mathcal{K} \] denote an element of \( \mathcal{T}_h \)

\[ P_k(\mathcal{K}) \] denote the space of all polynomials of degree less than or equal to \( k \)

defined on \( \mathcal{K} \)

• For the Dirichlet principle, we choose

\[ S_h^D = \mathcal{W}_m^0 = \{ \psi \in C^0(\Omega) \mid \psi|_{\mathcal{K}} \in P_m(\mathcal{K}) \} \]

\[ \mathbf{V}_h^D = \mathcal{W}_m^1 = \nabla (\mathcal{W}_m^0) \]
• The space $W_1^0$ restricted to an element
  – the degrees of freedom are the nodal values

• The space $W_1^1 = \nabla (W_1^0)$ restricted to an element
  – the degrees of freedom are vectors tangent to the edges
• It can be shown that if $\phi \in H^{m+1}(\Omega) \cap H_D^1(\Omega)$, then
  \[ \|\phi - \phi^h\|_0 \leq C h^{m+1} \|\phi\|_{m+1} \]

  \[ \|u - u^h\|_0 = \|\nabla (\phi - \phi^h)\|_0 \leq C h^m \|\phi\|_{m+1} \]

• Note that, since $W_1^m \equiv \nabla(W_0^m)$, even at the discrete level, we may again eliminate the flux approximation to obtain the equivalent discrete problem
  \[ \int_{\Omega} \nabla \phi^h \cdot \nabla \psi^h \, d\Omega + \int_{\Omega} \gamma \psi^h \phi^h \, d\Omega = \int_{\Omega} f \psi^h \, d\Omega \quad \forall \psi \in W_0^m \]
  that we recognize as the standard Galerkin discretization of the second-order Poisson problem

  – in fact, solutions of the first and second-order problems are identical

  – in this way we see that for discretizations of the Dirichlet principle, the required inf-sup condition is completely benign
Stable finite element methods for the Kelvin principle

• For each finite element $\mathcal{K} \in \mathcal{T}_h$, we let\(^1\)

\[
BDM_k(\mathcal{K}) = (P_k(\mathcal{K}))^n \quad \text{and} \quad RT_k(\mathcal{K}) = (P_k(\mathcal{K}))^n + xP_k(\mathcal{K})
\]

and then we define, for $k \geq 1$,

\[
\mathcal{W}_k^2(\mathcal{K}) = \begin{cases} 
RT_{k-1}(\mathcal{K}) \\
\text{or} \\
BDM_k(\mathcal{K})
\end{cases}
\]

• Then, we define the finite element spaces for the Kelvin principle

\[
\mathbf{V}_K^h = \mathcal{W}_k^2 = \{ \mathbf{v} \in H_N(\Omega, \text{div}) \mid \mathbf{v}|_\mathcal{K} \in \mathcal{W}_k^2(\mathcal{K}) \}
\]

\[
S_K^h = \mathcal{W}_k^3 = \nabla \cdot (\mathcal{W}_k^2)
\]

\(^1\)BDM = Brezzi, Douglas, Marini (1985); RT = Raviart, Thomas (1977)
- The space $\mathcal{W}_1^2 = RT_0(\Omega)$ restricted to an element
  - the degrees of freedom are vectors normal to the faces

- The space $\mathcal{W}_1^3 = \nabla \cdot (\mathcal{W}_1^2)$ restricted to an element
  - the degrees of freedom are averages over the simplices
• The space $W^2_1 = BDM_1(\Omega)$ restricted to an element
  – on each face, we have a linear polynomial times a constant normal vector
    (two degrees of freedom)

• The space $W^3_1 = \nabla \cdot (W^2_1) = P_0(\mathcal{K})$ restricted to an element
  – the degrees of freedom are averages over the simplices
• The inf-sup condition in the case of the Kelvin principle is definitely not benign

• It can be shown that if the exact solution is sufficiently smooth, then

\[
\|u - u^h\|_0 \leq \begin{cases} 
    C h^m \|u\|_m & \text{if } W_m^2(K) = RT_{m-1}(K) \\
    C h^{m+1} \|u\|_{m+1} & \text{if } W_m^2(K) = BDM_m(K)
\end{cases}
\]

\[
\|\nabla \cdot (u - u^h)\|_0 \leq C h^m \|\nabla \cdot u\|_m
\]

\[
\|\phi - \phi^h\|_0 \leq C h^m \left( \|\phi\|_m + \|u\|_m \right)
\]
• It is important to note that

  if one uses $C^0(\Omega)$ finite element spaces for both the scalar and the flux,
then the discrete Dirichlet and Kelvin problems are identical

  – it is well known that this leads to unstable approximations so that one
    cannot use (without introducing a stabilization techniques) such pairs of
    finite element spaces in mixed methods derived from either the Dirichlet
    or Kelvin principles

  – in particular, for such methods, one cannot use the same finite element
    space for approximating both the scalar and flux variables

• In passing, we note that both the discrete Dirichlet and Kelvin are, in general, symmetric and indefinite
LEAST-SQUARES FINITE ELEMENT METHODS

- A least-squares finite element\(^2\) method for the Poisson equation replaces the search for saddle-points of Lagrangian functionals by a search for the unconstrained global minimizer of the quadratic functional

\[ J(\phi, u; f) = \| \nabla \cdot u + \gamma \phi - f \|^2_0 + \| \nabla \phi + u \|^2_0 \]

- we have chosen the solution space \( S = H_D(\Omega, \text{grad}) \times H_N(\Omega, \text{div}) \)
- the data space \( H_\Omega = L^2(\Omega) \times L^2(\Omega) \)
- the boundary conditions are being enforced strongly, so that we need not include the residual of the boundary conditions in the least-squares functional

\(^2\)Jespersen 1977; Fix, G., Nicolaides 1977–1985; see also Cai, Chang, Carey, Jiang, Lazarov, Manteuffel, and others (94-00)
A least-squares variational principle is therefore given by

$$\min_{(\phi, u) \in H_D(\Omega, \text{grad}) \times H_N(\Omega, \text{div})} J(\phi, u; f)$$

Its solution minimizes the $L^2$-norm of the residuals of the first-order system

$$\nabla \cdot u + \gamma \phi = f \quad \text{and} \quad u + \nabla \phi = 0$$
• To define a least-squares finite element method,
  
  – we first choose any conforming finite element spaces
    
    \[ S^h_{LS} \subset H_D(\Omega, \text{grad}) = H^1_D(\Omega) \quad \text{and} \quad V^h_{LS} \subset H_N(\Omega, \text{div}) \]

  – then, we restrict the least-squares variational principle to the subspaces
    
    \[ \min_{(\phi^h, u^h) \in S^h_{LS} \times V^h_{LS}} J(\phi^h, u^h; f) \]

  – the resulting linear algebraic system has a symmetric, positive definite coefficient matrix

• The least squares functional is norm equivalent
  
  – there exist positive constants \( \alpha_1 \) and \( \alpha_2 \) such that
    
    \[ \alpha_1 \left( \| \psi \|_1^2 + \| v \|_{H_N(\Omega, \text{div})}^2 \right) \leq J(\psi, v; 0, 0) \leq \alpha_2 \left( \| \psi \|_1^2 + \| v \|_{H_N(\Omega, \text{div})}^2 \right) \]

  for any \( (\psi, v) \in H_D(\Omega, \text{grad}) \times H_N(\Omega, \text{div}) \)
Basic error estimates

• We first see how far we can get by merely assuming conformity

\[ S_{LS}^h \subset H_D(\Omega, \text{grad}) = H_D^1(\Omega) \quad \text{and} \quad V_{LS}^h \subset H_N(\Omega, \text{div}) \]

• We can do pretty well
  – we get an optimal estimate with respect to the data space norms, i.e., for

\[ \| \phi - \phi^h \|_1 + \| u - u^h \|_{H_N(\Omega, \text{div})} \]

  – the finite element spaces are not required to satisfy an inf-sup condition
• Since the only restriction imposed on the finite element spaces is conformity, we can choose either

- standard nodal, continuous, piecewise polynomial finite element spaces for both the scalar and flux variables:

or

- the finite element spaces used for the scalar and flux variables for the Dirichlet and Kelvin principles, respectively

\[ S^h_{LS} = \mathcal{W}_m^0 \quad \text{and} \quad V^h_{LS} = \mathcal{W}_m^2 \]

- the two choices use the same finite element space for the scalar variable

- for both choices we have that if \( \phi \) and \( u \) are smooth enough,

\[ \| \phi - \phi^h \|_1 + \| u - u^h \|_{H^N(\Omega, \text{div})} \leq C h^m \]

- this is an optimal convergence result
• Since the (asymptotic) accuracy is the same for either choice,
  
  – in the implementation of the LSFEM, it is most common to choose standard $C^0$ nodal elements for all variables
  
  – indeed, the ability to use the same degree polynomials defined with respect to same grid for approximating both the scalar and flux variables is often cited as a primary reason for choosing to use LSFEMs

• What about error estimates with respect to $L^2$ norms?
  
  – using either of the choices for the finite element spaces, we obtain the optimal convergence result for the scalar variable
    \[ \| \phi - \phi^h \|_0 \leq C h^{m+1} \]
  
  – again, the finite element spaces need only be conforming
    - they are not required to satisfy an inf-sup condition

• What about $L^2(\Omega)$ error estimates for the flux variable $u$?
The inclusions
\[ S^h_{LS} \subset H_D(\Omega, \text{grad}) = H^1_D(\Omega) \quad \text{and} \quad V^h_{LS} \subset H_N(\Omega, \text{div}) \]
are not sufficient for achieving optimal \( L^2 \)-accuracy for the least-squares finite element approximation of the flux variable

– for example, computational studies indicate that the \( L^2 \) convergence of the flux is suboptimal if \( C^0 \) finite element spaces are used for both the scalar and flux variables
Convergence rates for least-squares finite element (LS) and best (BA) approximations when continuous linear and quadratic finite element spaces are used to approximate both the scalar and flux variables.
• However, if

\[ S^h_{LS} = \mathcal{W}^0_k \subset H_D(\Omega, \text{grad}) = H^1_D(\Omega) \quad \text{and} \quad V^h_{LS} = \mathcal{W}^2_k \subset H_N(\Omega, \text{div}) \]

i.e., if

- the finite element space for the scalar is a stable space for the Dirichlet principle

- the finite element space for the flux is a stable space for the Kelvin principle

then, optimal \( L^2(\Omega) \) error estimates for both the scalar and flux variables are obtained
Convergence rates for least-squares finite element (LS) and best (BA) approximations when the subspace for the flux satisfies the GDP condition.

<table>
<thead>
<tr>
<th></th>
<th>scalar</th>
<th>vector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L^2$ error</td>
<td>$H$(grad) error</td>
</tr>
<tr>
<td>$P_1-RT_0$</td>
<td>LS: 2, 1</td>
<td>LS: 1, 1</td>
</tr>
<tr>
<td>$P_1-BDM_1$</td>
<td>LS: 2, 1</td>
<td>LS: 2, 1</td>
</tr>
</tbody>
</table>
Redundant equation approach

• For $C^0$ finite element pairs $S_{LS}^h$ and $V_{LS}^h$,

  – optimal $L^2$-norm error estimates for the flux approximation $\mathbf{u}^h$ can be obtained if one augments the first order system
    \[
    \nabla \cdot \mathbf{u} + \gamma \phi = f \quad \text{and} \quad \mathbf{u} = -\nabla \phi
    \]
    with an additional, redundant curl-constraint equation
    \[
    \nabla \cdot \mathbf{u} + \gamma \phi = f, \quad \mathbf{u} + \nabla \phi = 0, \quad \text{and} \quad \nabla \times \mathbf{u} = 0
    \]

  – the additional curl constraint makes the least-squares functional
    \[
    \hat{J}(\phi, \mathbf{u}; f) = \| \nabla \cdot \mathbf{u} + \gamma \phi - f \|_0^2 + \| \nabla \phi + \mathbf{u} \|_0^2 + \| \nabla \times \mathbf{u} \|_0^2
    \]
    norm-equivalent on $H^1_D(\Omega) \times H^1_N(\Omega)$

  – however, the addition of the curl equation may unduly restrict the range of the differential operator and should be avoided
We have seen that a least-squares finite element method implemented using equal-order $C^0$ finite element spaces

- approximates the scalar variable with the same accuracy as a mixed method for the Dirichlet principle

- however, the approximation properties of the Kelvin principle are only partially inherited $\Rightarrow$ the accuracy in the approximation of
  - the divergence of the flux is recovered
  - but not that for the flux itself

- perhaps this this is not surprising since $C^0$ elements
  - provide stable discretization for the scalar in the Dirichlet principle
  - but not that for the flux in the Kelvin principle
• It seems that while least-squares minimization
  – is stable enough to allow for the approximation of the scalar variable and the flux by equal-order $C^0$ finite element spaces,
  – it cannot completely recover from the fact that such spaces are unstable for the Kelvin principle

• The key observation is that a least-squares finite element method can inherit the computational properties of both the Dirichlet and the Kelvin principles
  – provided the scalar and flux variables are respectively approximated by finite element spaces that are stable with respect to those two principles
  – then, least-squares finite element solutions recover
    - the accuracy of the Dirichlet principle for the scalar variable
    - the accuracy of the Kelvin principle for the flux

• Thus, there seems to be some deep connections between mixed methods and least-squares finite element methods
Actually, the connection is very direct (for $\gamma \neq 0$)

– the Euler-Lagrange equations corresponding to the discrete least-squares principle

$$\min_{(\phi^h, u^h) \in S^h_{LS} \times V^h_{LS}} J(\phi^h, u^h; f)$$

can be shown to be equivalent to

$$\int_{\Omega} \nabla \phi^h \cdot \nabla \psi^h d\Omega + \int_{\Omega} \gamma \psi^h \phi^h d\Omega = \int_{\Omega} f \psi^h d\Omega \quad \forall \psi^h \in S^h_{LS}$$

and

$$\int_{\Omega} \gamma^{-1} \nabla \cdot u^h \nabla \cdot v^h d\Omega + \int_{\Omega} u^h \cdot v^h d\Omega = \int_{\Omega} \gamma^{-1} f \nabla \cdot v^h d\Omega \quad \forall v^h \in V^h_{LS}$$
– if we choose

\[ S^h_{LS} = \mathcal{W}^0_m = \text{a stable space for the scalar for the Dirichlet principle} \]
\[ \mathbf{V}^h_{LS} = \mathcal{W}^2_m = \text{a stable space for the vector for the Kelvin principle} \]

it can be shown that these equations are also satisfied by

the scalar approximation \( \phi^h \) in the Dirichlet principle

the vector approximation \( \mathbf{v}^h \) for the Kelvin principle

– thus,

- if the least-squares finite element method is implemented using stable spaces for the Dirichlet and Kelvin principles

then

- the approximations obtained by the least-squares finite element method are exactly the same as those obtained by the Dirichlet and Kelvin principles!
• So, what is the advantage of LSFEMs over mixed Galerkin FEMs?

  – one can recover the accuracy of the scalar variable in the Dirichlet principle and the vector variable in the Kelvin principle
    - one does not have to choose which one to approximate better

  – one can do this by solving a symmetric, positive definite matrix problem

• Of course, if one does not care about the optimal $L^2(\Omega)$ accuracy of the vector variable, then one can use equal-order interpolation for all variables in the LSFEM
• Using some simple computational experiments, we compare the $P_0 - RT_0$ mixed-Galerkin (Kelvin-based) finite element method with the $P_1 - RT_0$ least-squares finite element method.

• We focus on
  – convergence rates
  – mass conservation $\Rightarrow$ how well is $\nabla \cdot \mathbf{u} = 0$ satisfied
  – non-uniform grid performance $\Rightarrow$ two types of nonuniform grids used

• We note that least-squares finite element methods are often criticized due to their lack of mass conservation.
Non-uniform grids used in computational experiments

left: randomly perturbed grid
right: smooth, structured grid
Example 1

• $\Omega = (0, 1) \times (0, 1)$

• Exact solution: $\phi = \cos(\pi x^2) \cos(2\pi y)$ and $\mathbf{u} = -\nabla \phi$ in $\Omega$

  – for this solution, $\mathbf{u} \cdot \mathbf{n} = 0$ on $\Gamma$
### $L^2$ errors vs. number of grid points for uniform grids

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>LS</td>
<td>1.95898E+00</td>
<td>7.14389E−01</td>
<td>3.20360E−01</td>
<td>1.53575E−01</td>
<td>7.54023E−02</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>1.86215E+00</td>
<td>7.03207E−01</td>
<td>3.19134E−01</td>
<td>1.53434E−01</td>
<td>7.53854E−02</td>
</tr>
<tr>
<td>divu</td>
<td>LS</td>
<td>1.97815E+01</td>
<td>8.97653E+00</td>
<td>4.24617E+00</td>
<td>2.06086E+00</td>
<td>1.01480E+00</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>1.97809E+01</td>
<td>8.97644E+00</td>
<td>4.24616E+00</td>
<td>2.06086E+00</td>
<td>1.01480E+00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>LS</td>
<td>3.19383E−01</td>
<td>7.37150E−02</td>
<td>1.67233E−02</td>
<td>3.94968E−03</td>
<td>9.58253E−04</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>3.65643E−01</td>
<td>1.63230E−01</td>
<td>7.62731E−02</td>
<td>3.69037E−02</td>
<td>1.81583E−02</td>
</tr>
</tbody>
</table>

Example 1: $L^2$ errors vs. number of grid points in each direction for $P_1-RT_0$ least-squares and $P_0-RT_0$ mixed-Galerkin finite element approximations.
Example 1: $L^2$ errors vs. number of grid points in each direction for $P_1$-$RT_0$ least-squares finite element approximations
Example 1: $L^2$ errors vs. number of grid points in each direction for $P_0$-$RT_0$ mixed-Galerkin finite element approximations
Example 1: $L^2$ errors for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations.
Example 1: $P_1$-$RT_0$ least-squares finite element approximation on a $25 \times 25$ randomly perturbed nonuniform grid

left: $x$-component of $u$

center: $y$-component of $u$

right: $\phi$
Example 1: $P_0$-$RT_0$ mixed-Galerkin finite element approximation on a $25 \times 25$ randomly perturbed nonuniform grid

- left: $x$-component of $\mathbf{u}$
- center: $y$-component of $\mathbf{u}$
- right: $\phi$
Example 1: $L^2$ errors for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>MG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.270278</td>
<td>0.269682</td>
</tr>
<tr>
<td>div$u$</td>
<td>3.58186</td>
<td>3.58185</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.012129</td>
<td>0.064601</td>
</tr>
<tr>
<td>grad$\phi$</td>
<td>0.387836</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Example 1: $P_1$-$RT_0$ least-squares finite element approximation on a $25 \times 25$ smooth, structured nonuniform grid

left: $x$-component of $\mathbf{u}$

center: $y$-component of $\mathbf{u}$

right: $\phi$
Example 1: $P_0$-$RT_0$ mixed-Galerkin finite element approximation on a $25 \times 25$ smooth, structured nonuniform grid

left: $x$-component of $\mathbf{u}$

center: $y$-component of $\mathbf{u}$

right: $\phi$
Example 1: $P_1$-$RT_0$ least-squares (top) and $P_0$-$RT_0$ mixed-Galerkin (bottom) finite element approximations of $\phi$ on a $25 \times 25$ randomly perturbed nonuniform grid.
Example 2

• $\Omega = (0, 1) \times (0, 1)$

• Exact solution: $\phi = -e^x \sin y$ and $u = -\nabla \phi$ in $\Omega$

  – for this solution, $\nabla \cdot u = 0$ in $\Omega$

• Let us see how well approximations satisfy the “conservation” relation $\nabla \cdot u = 0$

  – there are those for whom achieving $\nabla \cdot u^h \rightarrow 0$ as $h \rightarrow 0$ is not good enough
    - they are fixated on achieving “exact” conservation
Example 2: $L^2$ norm of the $\nabla \cdot u^h$ in each element (left) and histogram of those element norms (right) for the $P_1$-$RT_0$ least-squares (top) and $P_0$-$RT_0$ mixed-Galerkin (bottom) finite element methods with a uniform $30 \times 30$ grid.
Example 2: $L^2$ errors for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations

- Although $\| \nabla \cdot u^h \|_0$ is small for the least-squares approximation, it is not within round-off error of the exact relation $\nabla \cdot u = 0$ as is the mixed-Galerkin approximation.
Fortunately, a simple, local, explicit frontal post-processing procedure results in an exactly conservative least-squares approximation – one begins by choosing an element, and then subtracting the (weighted) averages of the fluxes across the boundary of that element from the computed fluxes

\[ u_{ni} \leftarrow u_{ni} - \frac{\left( u_{n1}h_1 + u_{n2}h_2 + u_{n3}h_3 \right)}{(h_1 + h_2 + h_3)} \quad \text{for} \; i = 1, 2, 3 \]
— one then repeats this process in all the triangles, one at a time, except that if a flux has been previously been updated, then it is left unchanged e.g., if \( u_{n1} \) has already been updated, we have that

\[
    u_{n1} \leftarrow u_{n1}
\]

and

\[
    u_{ni} \leftarrow u_{ni} - \frac{\left( u_{n1}h_1 + u_{n2}h_2 + u_{n3}h_3 \right)}{(h_2 + h_3)} \quad \text{for } i = 2, 3
\]

• The post-processing procedure results in a new flux approximation \( u^h \) for which

\[
    u_{n1}h_1 + u_{n2}h_2 + u_{n3}h_3 = 0
\]

in every triangle

• The accuracy of \( u^h \) itself and of \( \phi^h \) is not affected
Example 2: $L^2$ norm of the $\nabla \cdot \mathbf{u}^h$ in each element (left) and histogram of those element norms (right) for the $P_1$-$RT_0$ least-squares (top) and post-processed least-squares (bottom) finite element methods with a uniform $30 \times 30$ grid.
Example 2: $L^2$ norm of the $\nabla \cdot u^h$ in each element (left) and histogram of those element norms (right) for the $P_0$-$RT_0$ mixed-Galerkin (top) and the $P_1$-$RT_0$ post-processed least-squares (bottom) finite element methods
Example 2: \( L^2 \) errors for \( P_1 - RT_0 \) least-squares, \( P_1 - RT_0 \) post-processed least-squares, and \( P_0 - RT_0 \) mixed-Galerkin finite element approximations

- We see that, indeed, the post-processed least-squares approximation achieves “exact” conservation without compromising the accuracy of the flux approximations
  - all least-squares calculations reported on below are for the post-processed version
<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>u LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>divu LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>φ LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 2: $L^2$ errors vs. number of grid points in each direction for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations
Example 2: $L^2$ errors vs. number of grid points in each direction for $P_1$-$RT_0$ least-squares finite element approximations
Example 2: $L^2$ errors vs. number of grid points in each direction for $P_0$-$RT_0$ mixed-Galerkin finite element approximations
Example 2: $L^2$ errors for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>MG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.356537E–01</td>
<td>0.356303E–01</td>
</tr>
<tr>
<td>div$u$</td>
<td>0.34923E–14</td>
<td>0.45140E–14</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.316519E–03</td>
<td>0.110673E+00</td>
</tr>
<tr>
<td>grad$\phi$</td>
<td>0.239616E–01</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Example 2: $P_1$-$RT_0$ least-squares finite element approximation on a $25 \times 25$ randomly perturbed nonuniform grid

left: $x$-component of $u$

center: $y$-component of $u$

right: $\phi$
Example 2: $P_0$-$RT_0$ mixed-Galerkin finite element approximation on a $25 \times 25$ randomly perturbed nonuniform grid

left: $x$-component of $\mathbf{u}$  
center: $y$-component of $\mathbf{u}$  
right: $\phi$
Example 2: $L^2$ errors for $P_1$-$RT_0$ least-squares and $P_0$-$RT_0$ mixed-Galerkin finite element approximations

<table>
<thead>
<tr>
<th></th>
<th>$L^2$ errors for a $25 \times 25$ smooth, structured nonuniform grid</th>
<th></th>
<th>MG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.344941E–01</td>
<td>0.344942E–01</td>
<td></td>
</tr>
<tr>
<td>$\text{div} u$</td>
<td>0.32852E–14</td>
<td>0.68958E–14</td>
<td></td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.235850E–03</td>
<td>0.352657E–01</td>
<td></td>
</tr>
<tr>
<td>$\text{grad} \phi$</td>
<td>0.224798E–01</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
Example 2: $P_1$-$RT_0$ least-squares finite element approximation on a $25 \times 25$ smooth, structured nonuniform grid

left: $x$-component of $\mathbf{u}$

center: $y$-component of $\mathbf{u}$

right: $\phi$
Example 2: $P_0$-$RT_0$ mixed-Galerkin finite element approximation on a $25 \times 25$ smooth, structured nonuniform grid

left: $x$-component of $\mathbf{u}$

center: $y$-component of $\mathbf{u}$

right: $\phi$
Example 2: $P_1$-$RT_0$ least-squares (top) and $P_0$-$RT_0$ mixed-Galerkin (bottom) finite element approximations of $\phi$ on a $25 \times 25$ randomly perturbed nonuniform grid
Example 2: $P_1$-$RT_0$ least-squares (top) and $P_0$-$RT_0$ mixed-Galerkin (bottom) finite element approximations of $\phi$ on a $25 \times 25$ smooth, structured nonuniform grid