
# Finite Element Methods for Least-Squares Problems

**Susanne C. Brenner**

**Blaise Pascal Medal 2025**

**Computational and Information Sciences**

Joint work with

Christopher B. Davis, José C. Garay, Joscha Gedicke, Thirupathi Gudi, SeongHee Jeong, Kamana Porwal, Li-yeng Sung, Zhiyu Tan, Hongchao Zhang, Yi Zhang


Supported by NSF

# Least-Squares Problems

## Linear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\left(\|Ax - b\|^2_{\mathbb{R}^m} + \gamma \|x\|^2_{\mathbb{R}^n}\right)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\gamma$ is a positive constant.

## Linear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\left(\|Ax - b\|^2_{\mathbb{R}^m} + \gamma \|x\|^2_{\mathbb{R}^n}\right)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\gamma$ is a positive constant.

### Data Fitting

Find the parameters $x$ that best fit the data given by $A$ and $b$. The regularization term involving $\gamma$ is included to prevent over-fitting.

This was investigated by Legendre in 1805 and Gauss in 1809.

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\left(\|Ax - b\|^2_{\mathbb{R}^m} + \gamma \|x\|^2_{\mathbb{R}^n}\right)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\gamma$ is a positive constant.

## Data Fitting

Find the parameters $x$ that best fit the data given by $A$ and $b$. The regularization term involving $\gamma$ is included to prevent over-fitting.

This was investigated by Legendre in 1805 and Gauss in 1809.

Constraints such as $x \geq 0$ or $y \geq 0$ (in the componentwise sense) can also be included.

# Linear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\left(\|Ax - b\|_{\mathbb{R}^m}^2 + \gamma \|x\|_{\mathbb{R}^n}^2\right)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\gamma$ is a positive constant.

Equivalently, find

$$(\bar{x}, \bar{y}) = \underset{\substack{y = Ax \\ x \in \mathbb{R}^n}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - b\|_{\mathbb{R}^m}^2 + \gamma \|x\|_{\mathbb{R}^n}^2\right)$$

# Linear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\left(\|Ax - b\|^2_{\mathbb{R}^m} + \gamma \|x\|^2_{\mathbb{R}^n}\right)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\gamma$ is a positive constant.

Equivalently, find

$$(\bar{x}, \bar{y}) = \underset{\substack{y = Ax \\ x \in \mathbb{R}^n}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - b\|^2_{\mathbb{R}^m} + \gamma \|x\|^2_{\mathbb{R}^n}\right)$$

## An Infinite Dimensional Analog

Find $$(\bar{u}, \bar{y}) = \underset{\substack{y = (-\Delta)^{-1} u \\ u \in L_2(\Omega)}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma \|u\|^2_{L_2(\Omega)}\right)$$

## Linear Least-Squares

Find

$$(\bar{u}, \bar{y}) = \underset{\substack{y=(-\Delta)^{-1}u \\ u \in L_2(\Omega)}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma \|u\|^2_{L_2(\Omega)}\right)$$

where $\Omega \subset \mathbb{R}^d$ and $y_d \in L_2(\Omega)$, is an elliptic distributed optimal control problem where the state $y \in H^1_0(\Omega)$ and the control $u \in L_2(\Omega)$ are related by

$$\int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \qquad \forall \, z \in H^1_0(\Omega)$$

This problem was first systematically studied by Lions in the 1960's.

Constraints on the control $u$ or the state $y$ can also be included.

# Elliptic Distributed Optimal Control Problem

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H_0^1(\Omega)$ if and only if

PDE Constraint

$$\int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \qquad \forall\, z \in H_0^1(\Omega)$$

Control Constraints

$$\phi_1 \leq u \leq \phi_2$$

State Constraints

$$\psi_1 \leq y \leq \psi_2$$

## Nonlinear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \sum_{j=1}^{m} r_j^2(x)$$

where $r_j$ is a (nonlinear) function from $\mathbb{R}^n$ to $\mathbb{R}$.

## Nonlinear Least-Squares

Find

$$\bar{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \sum_{j=1}^{m} r_j^2(x)$$

where $r_j$ is a (nonlinear) function from $\mathbb{R}^n$ to $\mathbb{R}$.

### Nonlinear Equations

Find a solution $\bar{x}$ of the system of equations

$$r_j(x) = 0 \qquad j = 1, \ldots, m$$

This was an idea first proposed by Cauchy in 1847.

Find

$$\bar{x} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{j=1}^{m} r_j^2(x)$$

where $r_j$ is a (nonlinear) function from $\mathbb{R}^n$ to $\mathbb{R}$.

## Nonlinear Equations

Find a solution $\bar{x}$ of the system of equations

$$r_j(x) = 0 \qquad j = 1, \ldots, m$$

This was an idea first proposed by Cauchy in 1847.

One can include a (componentwise) constraint

$$Ax \geq 0$$

to guarantee the uniqueness of the solution.

## Concise Form

$$\text{Find} \quad \bar{x} = \underset{\substack{x \in \mathbb{R}^n \\ Ax \geq 0}}{\operatorname{argmin}} \frac{1}{2} \|r(x)\|^2_{\mathbb{R}^m}$$

where $r(x) = [r_1(x), \ldots, r_m(x)]$.

## Concise Form

$$\text{Find} \quad \bar{x} = \underset{\substack{x \in \mathbb{R}^n \\ Ax \geq 0}}{\operatorname{argmin}} \frac{1}{2} \|r(x)\|_{\mathbb{R}^m}^2$$

where $r(x) = [r_1(x), \ldots, r_m(x)]$.

## An Infinite Dimensional Analog

$$\text{Find} \quad u = \underset{\substack{v \in H^4(\Omega) \cap H_0^1(\Omega) \\ \Delta v \geq 0}}{\operatorname{argmin}} \frac{1}{2} \| \det D^2 v - \psi \|_{L_2(\Omega)}^2$$

where $\Omega$ is a convex domain in $\mathbb{R}^2$ and $\psi \in H^2(\Omega)$.

# Dirichlet Boundary Value Problem for the Monge-Ampère Equation

$$\text{Find} \quad u = \underset{\substack{v \in H^4(\Omega) \cap H_0^1(\Omega) \\ \Delta v \geq 0}}{\operatorname{argmin}} \frac{1}{2} \| \det D^2 v - \psi \|^2_{L_2(\Omega)}$$

is the least-squares formulation of the Dirichlet boundary value problem for the Monge-Amp`ere equation:

Find a convex function $u$ such that

$$\begin{aligned} \det D^2 u &= \psi \qquad \text{in } \Omega \\ u &= 0 \qquad \text{on } \partial\Omega \end{aligned}$$

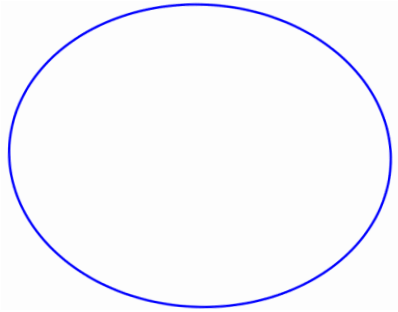
where $D^2 u$ is the Hessian of $u$.

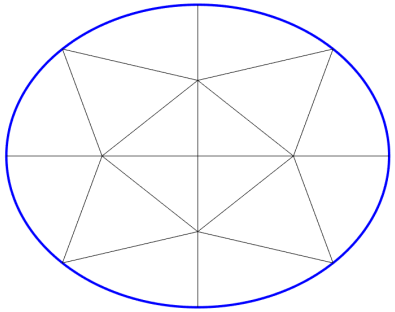
This was investigated by Monge in 1784 and Ampère in 1820.

## Finite Element Methods

I will present finite element methods for these least-squares problems.

I will present finite element methods for these least-squares problems.





In the finite element approach, given a domain $\Omega$, we create a triangulation and look for approximate solutions (finite element functions) that are polynomials on each triangle/element. The approximation is very accurate when the triangles/elements are small by Taylor's Theorem. They are the preferred numerical schemes for solving partial differential equations on domains with complex geometry and are very well developed for elliptic boundary value problems.

## Sobolev Spaces

I will use standard notation of Sobolev spaces throughout the presentation.

$H^k(\Omega)$ is the space of square integrable functions on $\Omega$ whose derivatives up to order $k$ are also square integrable.

$H_0^1(\Omega)$ is the subspace of $H^1(\Omega)$ whose members vanish on $\partial\Omega$.

# Elliptic Distributed Optimal Control Problems with State Constraints

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

PDE Constraint

$$\int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \qquad \forall\, z \in H^1_0(\Omega)$$

State Constraints

$$\psi_1 \leq y \leq \psi_2$$

This problem was first studied by Casas in 1986.

For simplicity, we assume $\Omega$ is convex.

# The Continuous Problem

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where

$$\mathbb{K} = \Big\{(u, y) \in L_2(\Omega) \times H^1_0(\Omega) : \int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \quad \forall\, z \in H^1_0(\Omega)$$
$$\text{and} \quad \psi_1 \le y \le \psi_2\Big\}$$

A traditional approach is to treat $y$ as a function of $u$ to reduce the optimal control problem into a minimization problem for $u$ and discretize accordingly.

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where

$$\mathbb{K} = \Big\{(u, y) \in L_2(\Omega) \times H^1_0(\Omega) : \int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \quad \forall\, z \in H^1_0(\Omega)$$

$$\text{and} \quad \psi_1 \le y \le \psi_2\Big\}$$

A traditional approach is to treat $y$ as a function of $u$ to reduce the optimal control problem into a minimization problem for $u$ and discretize accordingly.

But we can also treat $u$ as a function of $y$ and reduce the optimal control problem into a minimization problem for $y$.

Pierre-Sokolowski 1994
Liu-Gong-Yan 2009

# The Continuous Problem

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where

$$\mathbb{K} = \Big\{(u, y) \in L_2(\Omega) \times H^1_0(\Omega) : \int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \quad \forall\, z \in H^1_0(\Omega)$$

$$\text{and} \quad \psi_1 \le y \le \psi_2\Big\}$$

Since $\Omega$ is convex, we have $y \in H^2(\Omega)$ by elliptic regularity and $u = -\Delta y$ in the strong sense.

## The Continuous Problem

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where

$$\mathbb{K} = \Big\{(u, y) \in L_2(\Omega) \times H^1_0(\Omega) : \int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \quad \forall\, z \in H^1_0(\Omega)$$
$$\text{and} \quad \psi_1 \le y \le \psi_2\Big\}$$

Since $\Omega$ is convex, we have $y \in H^2(\Omega)$ by elliptic regularity and $u = -\Delta y$ in the strong sense.

# The Continuous Problem

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where

$$\mathbb{K} = \Big\{(u, y) \in L_2(\Omega) \times H^1_0(\Omega) : \int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \quad \forall\, z \in H^1_0(\Omega)$$

$$\text{and} \quad \psi_1 \le y \le \psi_2\Big\}$$

Equivalent problem

$$\text{Find} \quad \bar{y} = \underset{y\in K}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|\Delta y\|^2_{L_2(\Omega)}\right)$$

where

$$K = \{y \in H^2(\Omega) \cap H^1_0(\Omega) : \ \psi_1 \le y \le \psi_2\}$$

This convex minimization problem

$$\text{Find} \quad \bar{y} = \underset{y \in K}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma \|\Delta y\|^2_{L_2(\Omega)}\right)$$

where

$$K = \{y \in H^2(\Omega) \cap H^1_0(\Omega) : \ \psi_1 \leq y \leq \psi_2\}$$

has a unique solution characterized by the first order optimality condition

$$\int_\Omega (\bar{y} - y_d)(y - \bar{y}) dx + \gamma \int_\Omega (\Delta \bar{y}) \Delta (y - \bar{y}) dx \geq 0 \qquad \forall\, y \in K$$

which is a fourth order variational inequality.

The fourth order variational inequality

$$\int_\Omega (\bar{y} - y_d)(y - \bar{y})dx + \gamma \int_\Omega (\Delta\bar{y})\Delta(y - \bar{y})dx \geq 0 \qquad \forall\, y \in K$$

can be written concisely as

$$a(\bar{y}, y - \bar{y}) \geq \int_\Omega y_d(y - \bar{y})dx \qquad \forall\, y \in K$$

where

$$\begin{aligned} a(y, z) &= \gamma \int_\Omega (\Delta y)(\Delta z)dx + \int_\Omega yz\, dx \\ &= \gamma \int_\Omega (D^2 y : D^2 z)dx + \int_\Omega yz\, dx \end{aligned}$$

## Discretization

The fourth order variational inequality

$$\int_\Omega (\bar{y} - y_d)(y - \bar{y})dx + \gamma \int_\Omega (\Delta\bar{y})\Delta(y - \bar{y})dx \geq 0 \qquad \forall\, y \in K$$

can be written concisely as

$$a(\bar{y}, y - \bar{y}) \geq \int_\Omega y_d(y - \bar{y})dx \qquad \forall\, y \in K$$

In the absence of constraints for $y$ $(\psi_1 = -\infty, \psi_2 = +\infty)$, the variational inequality becomes a variational equality

$$a(\bar{y}, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in H^2(\Omega) \cap H_0^1(\Omega)$$

which is a boundary value problem for simply supported plates that can be discretized by many finite element methods.

Find $\bar{y} \in H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$a(\bar{y}, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in H^2(\Omega) \cap H_0^1(\Omega)$$

Find $\bar{y} \in H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$a(\bar{y}, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in H^2(\Omega) \cap H_0^1(\Omega)$$

Finite Element Method Find $\bar{y}_h \in V_h$ such that

$$a_h(\bar{y}_h, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in V_h$$

where

- $V_h$ is a finite element space that approximates the continuous space $H^2(\Omega) \cap H_0^1(\Omega)$.
- $a_h(\cdot, \cdot)$ is a bilinear form that approximates $a(\cdot, \cdot)$.

Find $\bar{y} \in H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$a(\bar{y}, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in H^2(\Omega) \cap H_0^1(\Omega)$$

Finite Element Method Find $\bar{y}_h \in V_h$ such that

$$a_h(\bar{y}_h, z) = \int_\Omega y_d z\, dx \qquad \forall\, z \in V_h$$

- Conforming $C^1$ Finite Element Methods
- Mixed Finite Element Methods
- Classical Nonconforming Finite Element Methods
- Discontinuous Galerkin Finite Element Methods

## Variational Inequality

Find $\bar{y} \in K$ such that

$$a(\bar{y}, y - \bar{y}) \geq \int_\Omega y_d(y - \bar{y})dx \qquad \forall\, y \in K$$

where $K = \{y \in H^2(\Omega) \cap H_0^1(\Omega) : \psi_1 \leq y \leq \psi_2\}$.

Discretization Find $\bar{y}_h \in K_h$ such that

$$a_h(\bar{y}_h, y_h - \bar{y}_h) \geq \int_\Omega y_d(y_h - \bar{y}_h)dx \qquad \forall\, y_h \in K_h$$

where

$K_h = \{y_h \in V_h : \psi_1(p) \leq y_h(p) \leq \psi_2(p)$ at all the vertices $p$ of $\mathcal{T}_h\}$

$V_h$ and $a_h(\cdot, \cdot)$ can come from conforming methods, mixed methods, nonconforming methods and DG methods.

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.

### Conforming $C^1$ Finite Element Method

B., C.B. Davis and L.-Y. Sung

A partition of unity method for a class of fourth order variational inequalities

Computer Methods in Applied Mechanics and Engineering (2014)

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.

Mixed Finite Element Method

B., J. Gedicke and L.-Y. Sung

$P_1$ finite element methods for an elliptic optimal control problem with pointwise state constraints

IMA Journal of Numerical Analysis (2020)

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.

### Classical Nonconforming Finite Element Method

B., T. Gudi, K. Porwal and L.-Y. Sung

A Morley finite element method for an elliptic distributed optimal control problem with pointwise state and control constraints

ESAIM: Control, Optimisation and Calculus of Variations (2018)

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.

### Discontinuous Galerkin Finite Element Methods

B., L.-Y. Sung and Y.-Zhang

A quadratic $C^0$ interior penalty method for an elliptic optimal control problem with state constraints

IMA Volumes in Mathematics and Its Applications (2014)

B., J. Gedicke and L.-Y. Sung

A symmetric interior penalty method for an elliptic distributed optimal control problem with pointwise state constraints

Computational Methods in Applied Mathematics (2023)

## Summary

- We can solve elliptic optimal control problems as fourth order variational inequalities by using finite element methods originally designed for fourth order elliptic boundary value problems. The traditional approach of reducing the problem to a minimization problem for the control corresponds to the special case of a mixed finite element method.
- Since the unknown is the optimal state $\bar{y}$, the pointwise state constraints become box constraints for the discrete problems so that the resulting quadratic programs can be solved efficiently by a primal-dual active set method.

Bergounioux-Kunisch 2002
Hintermüller-Ito-Kunisch 2003

## Summary

- The error analysis is based on the work of Frehse, Caffarelli and Friedman on fourth order variational inequalities, standard finite element techniques for fourth order elliptic boundary value problems and a clever application of the Karush-Kuhn-Tucker conditions.

B. and L.-Y. Sung

A new convergence analysis of finite element methods for elliptic distributed optimal control problems with pointwise state constraints

SIAM Journal on Control and Optimization (2017)

## Summary

- This approach has been extended to optimal control problems with pointwise state and pointwise control constraints.

B., T. Gudi, K. Porwal and L.-Y. Sung

A Morley finite element method for an elliptic distributed optimal control problem with pointwise state and control constraints

ESAIM: Control, Optimisation and Calculus of Variations (2018)

B., L.-Y. Sung and Z. Tan

A cubic $C^0$ interior penalty method for elliptic distributed optimal control problems with pointwise state and control constraints

Results in Applied Mathematics (2020)

## Summary

- We can also handle optimal control problems with general cost functions that include the point-tracking problem.

B., S. Jeong, L.-Y. Sung and Z. Tan

$C^0$ interior penalty methods for an elliptic distributed optimal control problem with general tracking and pointwise state constraints

Computers and Mathematics with Applications (2024)

## Summary

- There are finite element methods (virtual element methods, HDG methods, weak Galerkin methods, HHO methods, ...) that can be implemented on polygonal meshes. The new approach can also be applied to these methods.

B., L.-Y. Sung and Z. Tan

A $C^1$ virtual element method for an elliptic distributed optimal control problem with pointwise state constraints

Mathematical Models & Methods in Applied Sciences (2021)

# Elliptic Distributed Optimal Control Problems with Control Constraints

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

PDE Constraint

$$\int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \qquad \forall\, z \in H^1_0(\Omega)$$

Control Constraints

$$\phi_1 \leq u \leq \phi_2$$

This is a simpler problem that was first analyzed by Falk in 1973.

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

PDE Constraint

$$\int_\Omega \nabla y \cdot \nabla z \, dx = \int_\Omega uz \, dx \qquad \forall\, z \in H^1_0(\Omega)$$

Control Constraints

$$\phi_1 \leq u \leq \phi_2$$

This is a simpler problem that was first analyzed by Falk in 1973.

To make it more interesting, we will consider a more general PDE constraint.

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

PDE Constraint

$$a(y, z) = \int_\Omega uz\, dx \qquad \forall\, z \in H^1_0(\Omega)$$

where the symmetric bilinear form $a(\cdot, \cdot)$ satisfies

$$\alpha|y|^2_{H^1(\Omega)} \leq a(y, y) \leq \beta|y|^2_{H^1(\Omega)} \qquad \forall\, y \in H^1(\Omega)$$

for positive constants $\alpha \leq \beta$.

Control Constraints

$$\phi_1 \leq u \leq \phi_2$$

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

PDE Constraint

$$a(y, z) = \int_\Omega uz\, dx \qquad \forall\, z \in H^1_0(\Omega)$$

where the symmetric bilinear form $a(\cdot, \cdot)$ satisfies

$$\alpha|y|^2_{H^1(\Omega)} \leq a(y, y) \leq \beta|y|^2_{H^1(\Omega)} \qquad \forall\, y \in H^1(\Omega)$$

for positive constants $\alpha \leq \beta$.

Remark The assumption on $a(\cdot, \cdot)$ means that the PDE has rough coefficients, which appears in many multiscale problems.

# Examples of PDE Constraints with Rough Coefficients

$$a(y,z) = \int_\Omega \mathcal{A}(x)\nabla y \cdot \nabla z \, dx$$

# Examples of PDE Constraints with Rough Coefficients

$$a(y,z) = \int_\Omega \mathcal{A}(x)\nabla y \cdot \nabla z \, dx$$

## Highly Oscillatory Coefficients

$$\mathcal{A}(x) = c(x)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where $(\epsilon \ll 1)$

$$c(x) = \frac{2 + 1.8\sin\left(\frac{2\pi x_1}{\epsilon}\right)}{2 + 1.8\cos\left(\frac{2\pi x_2}{\epsilon}\right)} + \frac{2 + \sin\left(\frac{2\pi x_2}{\epsilon}\right)}{2 + 1.8\sin\left(\frac{2\pi x_1}{\epsilon}\right)}.$$

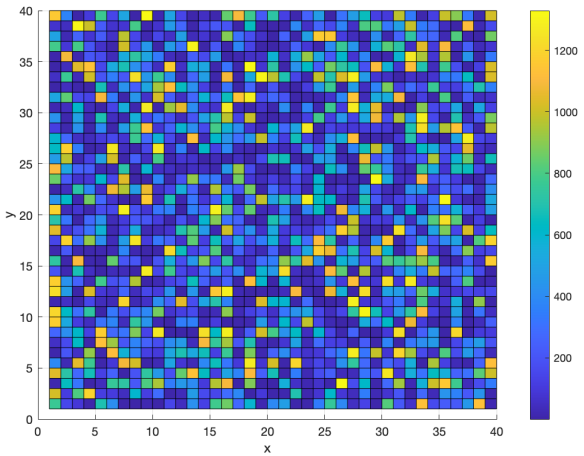
(Hou-Wu 1997)

# Examples of PDE Constraints with Rough Coefficients

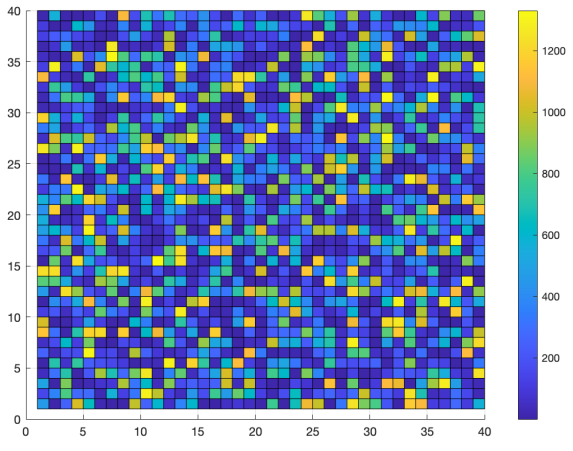
$$a(y,z) = \int_\Omega \mathcal{A}(x)\nabla y \cdot \nabla z \, dx$$

## Highly Heterogeneous Coefficients

$$\mathcal{A}(x) = \begin{bmatrix} \mathcal{A}_{11}(x) & 0 \\ 0 & \mathcal{A}_{22}(x) \end{bmatrix}$$




$\mathcal{A}_{11}(x)$




$\mathcal{A}_{22}(x)$

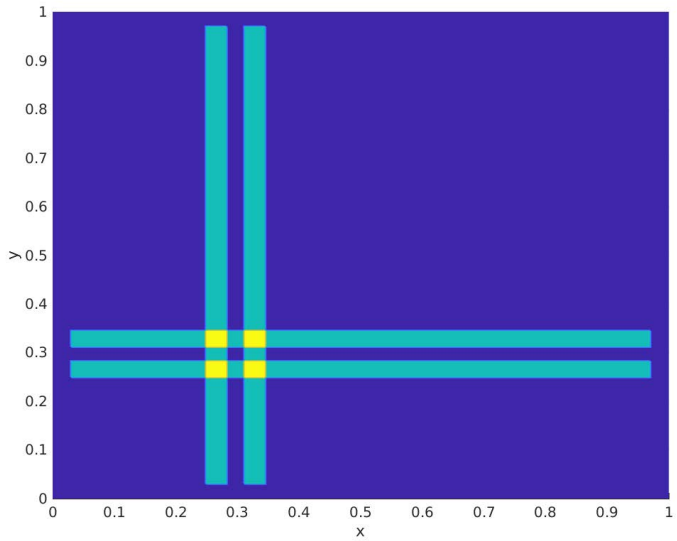
$\mathcal{A}_{11}(x)$ and $\mathcal{A}_{22}(x)$ are randomly generated piecewise constant functions with values in $[1, 1331]$.

# Examples of PDE Constraints with Rough Coefficients

$$a(y,z) = \int_\Omega \mathcal{A}(x)\nabla y \cdot \nabla z \, dx$$

## High Contrast Coefficients

$$\mathcal{A}(x) = c(x)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$




$$c(x) = \kappa(x_1, x_2) + \kappa(x_2, x_1)$$

$(\beta \gg 1)$

$$\kappa(x) = \begin{cases} \beta/2 & \text{for } x \in [\frac{8}{32}, \frac{9}{32}] \times [\frac{1}{32}, \frac{31}{32}] \cup [\frac{10}{32}, \frac{11}{32}] \times [\frac{1}{32}, \frac{31}{32}] \\ 1 & \text{elsewhere} \end{cases}$$

(Peterseim-Scheichl 2016)

Find

$$(\bar{u}, \bar{y}) = \underset{(u,y)\in\mathbb{K}}{\operatorname{argmin}} \frac{1}{2}\left(\|y - y_d\|^2_{L_2(\Omega)} + \gamma\|u\|^2_{L_2(\Omega)}\right)$$

where $(u, y) \in \mathbb{K} \subset L_2(\Omega) \times H^1_0(\Omega)$ if and only if

$$a(y, z) = \int_\Omega uz\, dx \qquad \forall\, z \in H^1_0(\Omega)$$

$$\phi_1 \le u \le \phi_2 \qquad \text{a.e. in } \Omega$$

We assume that

$$\phi_1 \text{ and } \phi_2 \text{ belong to } H^1(\Omega)$$

$$\phi_1 \le \phi_2 \quad \text{a.e. in } \Omega$$

This convex minimization problem has a unique solution.

## The Continuous Problem

The solution of the optimal control problem is characterized by

- The state equation for $\bar{y} \in H_0^1(\Omega)$

$$a(\bar{y}, z) = \int_\Omega \bar{u} z \, dx \qquad \forall\, z \in H_0^1(\Omega)$$

- The adjoint state equation for $\bar{p} \in H_0^1(\Omega)$

$$a(q, \bar{p}) = \int_\Omega (\bar{y} - y_d) q \, dx \qquad \forall\, q \in H_0^1(\Omega)$$

- The relation between the adjoint state and the optimal control

$$\bar{u} = \max\left(\phi_1, \min(\phi_2, -\gamma^{-1}\bar{p})\right) \qquad (\phi_1, \phi_2 \in H^1(\Omega))$$

The solution of the optimal control problem is characterized by

- The state equation for $\bar{y} \in H_0^1(\Omega)$

$$a(\bar{y}, z) = \int_\Omega \bar{u} z \, dx \qquad \forall\, z \in H_0^1(\Omega)$$

- The adjoint state equation for $\bar{p} \in H_0^1(\Omega)$

$$a(q, \bar{p}) = \int_\Omega (\bar{y} - y_d) q \, dx \qquad \forall\, q \in H_0^1(\Omega)$$

- The relation between the adjoint state and the optimal control

$$\bar{u} = \max\left(\phi_1, \min(\phi_2, -\gamma^{-1}\bar{p})\right) \qquad (\phi_1, \phi_2 \in H^1(\Omega))$$

Since $H^1(\Omega)$ is preserved by the operators max and min, we have $\bar{u} \in H^1(\Omega)$ which improves on the assumption that $\bar{u} \in L_2(\Omega)$, i.e., we have a gain of elliptic regularity for $\bar{u}$.

The solution of the optimal control problem is characterized by

- The state equation for $\bar{y} \in H_0^1(\Omega)$

$$a(\bar{y}, z) = \int_\Omega \bar{u} z \, dx \qquad \forall\, z \in H_0^1(\Omega)$$

- The adjoint state equation for $\bar{p} \in H_0^1(\Omega)$

$$a(q, \bar{p}) = \int_\Omega (\bar{y} - y_d) q \, dx \qquad \forall\, q \in H_0^1(\Omega)$$

- The relation between the adjoint state and the optimal control

$$\bar{u} = \max\left(\phi_1, \min(\phi_2, -\gamma^{-1}\bar{p})\right) \qquad (\phi_1, \phi_2 \in H^1(\Omega))$$

On the other hand we have $\bar{y}, \bar{p} \in H_0^1(\Omega)$ and nothing more under the assumption of rough coefficients, i.e., we do not have any gain of elliptic regularity for $\bar{y}$ and $\bar{p}$.

## Discretization

Let $\mathcal{T}_\rho$ be a triangulation of $\Omega$, $W_\rho$ be the space of piecewise constant functions with respect to $\mathcal{T}_\rho$ and $V_*$ be a finite dimensional subspace of $H_0^1(\Omega)$.

We use the piecewise constant functions in $W_\rho$ to approximate the optimal control $\bar{u}$ and the functions in $V_*$ to approximate the optimal state $\bar{y}$.

We allow $V_*$ to be a standard finite element space or a multiscale finite element space.

## Discretization

Let $\mathcal{T}_\rho$ be a triangulation of $\Omega$, $W_\rho$ be the space of piecewise constant functions with respect to $\mathcal{T}_\rho$ and $V_*$ be a finite dimensional subspace of $H_0^1(\Omega)$.

The discrete problem is to find

$$(\bar{u}_{\rho,*}, \bar{y}_{\rho,*}) = \underset{(u_\rho, y_*) \in \mathbb{K}_{\rho,*}}{\operatorname{argmin}} \frac{1}{2}\left[\|y_* - y_d\|^2_{L_2(\Omega)} + \gamma \|u_\rho\|^2_{L_2(\Omega)}\right]$$

where the closed convex subset $\mathbb{K}_{\rho,*}$ of $W_\rho \times V_*$ is defined by the following conditions:

$$a(y_*, z_*) = \int_\Omega u_\rho z_* dx \qquad \forall\, z_* \in V_*$$

$$Q_\rho \phi_1 \le u_\rho \le Q_\rho \phi_2 \qquad \text{a.e. in } \Omega$$

and $Q_\rho$ is the orthogonal projection from $L_2(\Omega)$ onto $W_\rho$.

This finite dimensional convex minimization problem has a unique solution.

# Discretization

The solution of the discrete problem is characterized by

- The discrete state equation for $\bar{y}_{\rho,*} \in V_*$

$$a(\bar{y}_{\rho,*}, z_*) = \int_\Omega \bar{u}_{\rho,*} z_* \, dx \qquad \forall \, z_* \in V_*$$

- The discrete adjoint state equation for $\bar{p}_{\rho,*} \in V_*$

$$a(q_*, \bar{p}_{\rho,*}) = \int_\Omega (\bar{y}_{\rho,*} - y_d) q_* \, dx \qquad \forall \, q_* \in V_*$$

- The relation between the $\bar{p}_{\rho,*}$ and $\bar{u}_{\rho,*}$

$$\int_\Omega (\bar{p}_{\rho,*} + \gamma \bar{u}_{\rho,*})(u_\rho - \bar{u}_{\rho,*}) dx \geq 0 \qquad \forall \, u_\rho \in U_{\text{ad}}^\rho$$

where

$$U_{\text{ad}}^\rho = \{u_\rho \in W_\rho : \, Q_\rho \phi_1 \leq u_\rho \leq Q_\rho \phi_2\}$$

In the case of a classical PDE constraint where

$$a(y,z) = \int_\Omega \nabla y \cdot \nabla z \, dx$$

the error analysis for standard finite element methods was already carried out in the 1970's.

However this error analysis is not directly applicable to the optimal control problem where the PDE constraint has rough coefficients due to the lack of elliptic regularity pick up for the optimal state $\bar{y}$ and the adjoint state $\bar{p}$.

A new error analysis is needed.

**Theorem** There exist constants $C_{\clubsuit}$ and $C_{\spadesuit}$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

where $R_* : H_0^1(\Omega) \longrightarrow V_*$ is the Ritz projection operator defined by

$$a(R_*\zeta, v_*) = a(\zeta, v_*) \qquad \forall\, v_* \in V_*$$

and

$$\|v\|_a = \sqrt{a(v,v)}$$

## New Error Estimates

**Theorem** There exist constants $C_\clubsuit$ and $C_\spadesuit$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_\clubsuit(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_\spadesuit(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

**Remark** The derivations of these estimates are based on the characterizations of the solutions of the continuous and discrete problems. The $O(\rho)$ term appears because we are approximating the optimal control $\bar{u} \in H^1(\Omega)$ by piecewise constant functions associated with $\mathcal{T}_\rho$.

**Theorem** There exist constants $C_\clubsuit$ and $C_\spadesuit$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_\clubsuit(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_\spadesuit(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

**Remark** When we solve the boundary value problem

$$a(\zeta, v) = \int_\Omega f v \, dx \qquad \forall\, v \in H_0^1(\Omega)$$

by the Ritz-Galerkin method based on $V_*$, the $L_2(\Omega)$ and energy norm errors are defined in terms of $R_*$.

## New Error Estimates

**Theorem** There exist constants $C_\clubsuit$ and $C_\spadesuit$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_\clubsuit(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_\spadesuit(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

Therefore these estimates reduce the error analysis of the optimal control problem to the error analysis of the boundary value problem.

## New Error Estimates

**Theorem** There exist constants $C_\clubsuit$ and $C_\spadesuit$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_\clubsuit(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_\spadesuit(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

These are new error estimates for the elliptic optimal control problem that are suitable for the analysis of multiscale finite element methods for PDE constraints with rough coefficients.

## New Error Estimates

**Theorem** There exist constants $C_\clubsuit$ and $C_\spadesuit$ such that

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_\clubsuit(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a$$
$$\leq C_\spadesuit(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

These are new error estimates for the elliptic optimal control problem that are suitable for the analysis of multiscale finite element methods for PDE constraints with rough coefficients.

We can also apply these estimates to the case of classical PDE constraints.

## Classical PDE Constraint

$$a(y,z) = \int_\Omega \nabla y \cdot \nabla z \, dx$$

where $\Omega$ is a convex polygonal domain, and $\|\cdot\|_a = |\cdot|_{H^1(\Omega)}$.

$$a(y,z) = \int_\Omega \nabla y \cdot \nabla z \, dx$$

where $\Omega$ is a convex polygonal domain, and $\|\cdot\|_a = |\cdot|_{H^1(\Omega)}$.

In this case $\bar{y}$ and $\bar{p}$ belong to $H^2(\Omega)$ by elliptic regularity for the state equation and the adjoint state equation.

$$\int_\Omega \nabla \bar{y} \cdot \nabla z \, dx = \int_\Omega \bar{u} z \, dx \qquad \forall\, z \in H_0^1(\Omega)$$

$$\int_\Omega \nabla q \cdot \nabla \bar{p} \, dx = \int_\Omega (\bar{y} - y_d) q \, dx \qquad \forall\, q \in H_0^1(\Omega)$$

$$a(y,z) = \int_\Omega \nabla y \cdot \nabla z \, dx$$

where $\Omega$ is a convex polygonal domain, and $\|\cdot\|_a = |\cdot|_{H^1(\Omega)}$.

In this case $\bar{y}$ and $\bar{p}$ belong to $H^2(\Omega)$ by elliptic regularity for the state equation and the adjoint state equation.

We can take $V_* \subset H_0^1(\Omega)$ to be a standard $P_1$ finite element space $V_h$ associated with a triangulation $\mathcal{T}_h$ of $\Omega$, and we can write the solution of the discrete problem as

$$\bar{y}_{h,\rho} \in V_h, \quad \bar{p}_{h,\rho} \in V_h \quad \text{and} \quad \bar{u}_{h,\rho} \in W_\rho$$

The Ritz projection operator $R_h : H_0^1(\Omega) \longrightarrow V_h$ is defined by

$$\int_\Omega \nabla(R_h\zeta) \cdot \nabla v \, dx = \int_\Omega \nabla\zeta \cdot \nabla v \, dx \qquad \forall\, v \in H_0^1(\Omega)$$

# Classical PDE Constraint

The estimates from the theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)} \leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

become the concrete estimates

$$\|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} \leq C(h^2 + \rho)$$
$$|\bar{y} - \bar{y}_{\rho,h}|_{H^1(\Omega)} + |\bar{p} - \bar{p}_{\rho,h}|_{H^1(\Omega)} \leq C(h + \rho)$$

that come from standard finite element estimates

$$|\zeta - R_h\zeta|_{H^1(\Omega)} \leq Ch|\zeta|_{H^2(\Omega)} \qquad \forall\, \zeta \in H^2(\Omega) \cap H_0^1(\Omega)$$
$$\|\zeta - R_h\zeta\|_{L_2(\Omega)} \leq Ch^2|\zeta|_{H^2(\Omega)} \qquad \forall\, \zeta \in H^2(\Omega) \cap H_0^1(\Omega)$$

$(\bar{y}, \bar{p} \in H^2(\Omega) \cap H_0^1(\Omega))$

# Classical PDE Constraint

The estimates from the theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)} \leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

become the concrete estimates

$$\|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} \leq C(h^2 + \rho)$$

$$|\bar{y} - \bar{y}_{\rho,h}|_{H^1(\Omega)} + |\bar{p} - \bar{p}_{\rho,h}|_{H^1(\Omega)} \leq C(h + \rho)$$

We have recovered the error estimates of Falk 1973 for the elliptic optimal control problem with smooth data.

# PDE Constraints with Rough Coefficients

# PDE Constraints with Rough Coefficients

For the standard finite element method, the estimates from the theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)} \leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

now take this form

$$\|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_a + \|\bar{p} - \bar{p}_{\rho,h}\|_a \leq C(\|\bar{y} - R_h\bar{y}\|_a + \|\bar{p} - R_h\bar{p}\|_a + \rho)$$

# PDE Constraints with Rough Coefficients

For the standard finite element method, the estimates from the theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)} \leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

now take this form

$$\|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_a + \|\bar{p} - \bar{p}_{\rho,h}\|_a \leq C(\|\bar{y} - R_h\bar{y}\|_a + \|\bar{p} - R_h\bar{p}\|_a + \rho)$$

Note that

$$\lim_{h\downarrow 0} \|\bar{y} - R_h\bar{y}\|_a = \lim_{h\downarrow 0} \inf_{z_h \in V_h} \|\bar{y} - z_h\|_a = 0$$

$$\lim_{h\downarrow 0} \|\bar{p} - R_h\bar{p}\|_a = \lim_{h\downarrow 0} \inf_{q_h \in V_h} \|\bar{p} - q_h\|_a = 0$$

For the standard finite element method, the estimates from the theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)} \\ \leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

now take this form

$$\|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} \\ + \|\bar{y} - \bar{y}_{\rho,h}\|_a + \|\bar{p} - \bar{p}_{\rho,h}\|_a \leq C(\|\bar{y} - R_h\bar{y}\|_a + \|\bar{p} - R_h\bar{p}\|_a + \rho)$$

Therefore we have convergence as $h, \rho \downarrow 0$. But the convergence in $h$ can be arbitrarily slow.

(Babuška-Osborn 2000)

We will need a very fine mesh $\mathcal{T}_h$ in order to get a good approximate solution by using a standard finite element space $V_h$.

Suppose we have a good approximate solution $(\bar{y}_{\rho,h}, \bar{u}_{\rho,h}, \bar{p}_{\rho,h})$ obtained by a standard finite element method on a sufficiently fine mesh $\mathcal{T}_h$.

We can improve the performance on coarser meshes by switching to a multiscale finite element space.

Suppose we have a good approximate solution $(\bar{y}_{\rho,h}, \bar{u}_{\rho,h}, \bar{p}_{\rho,h})$ obtained by a standard finite element method on a sufficiently fine mesh $\mathcal{T}_h$.

We can improve the performance on coarser meshes by switching to a multiscale finite element space.

For example, we can take

$$V_* = V_H^{\mathrm{ms},h}$$

where $V_H^{\mathrm{ms},h}$ is the multiscale finite element space of Målqvist and Peterseim.

Målqvist and Peterseim 2014

Suppose we have a good approximate solution $(\bar{y}_{\rho,h}, \bar{u}_{\rho,h}, \bar{p}_{\rho,h})$ obtained by a standard finite element method on a sufficiently fine mesh $\mathcal{T}_h$.

We can improve the performance on coarser meshes by switching to a multiscale finite element space.

For example, we can take

$$V_* = V_H^{\mathrm{ms},h}$$

This is a reduced order method. The time consuming construction of $V_H^{\mathrm{ms},h}$, which involves solving problems on the fine mesh $\mathcal{T}_h$, is carried out off-line. The on-line computation using $V_H^{\mathrm{ms},h}$ is very fast because the dimension of $V_H^{\mathrm{ms},h}$ is identical to the dimension of a standard finite element space on a coarse mesh $\mathcal{T}_H$.

Suppose we have a good approximate solution $(\bar{y}_{\rho,h}, \bar{u}_{\rho,h}, \bar{p}_{\rho,h})$ obtained by a standard finite element method on a sufficiently fine mesh $\mathcal{T}_h$.

We can improve the performance on coarser meshes by switching to a multiscale finite element space.

For example, we can take

$$V_* = V_H^{\mathrm{ms},h}$$

The salient feature is that, up to the error of a standard finite element method on the fine mesh, the performance of this method is similar to the performance of a standard finite element method on the coarse mesh for a classical elliptic problem on a convex domain even if the coefficients are rough and the domain is nonconvex.

## Error Estimates

Let $(\bar{u}^{\mathrm{ms},h}_{\rho,H,}, \bar{y}^{\mathrm{ms},h}_{\rho,H}, \bar{p}^{\mathrm{ms},h}_{\rho,H})$ be the solution of the optimal control problem where $V_* = V_H^{\mathrm{ms},h}$.

## Error Estimates

Let $(\bar{u}_{\rho,H,}^{\mathrm{ms},h}, \bar{y}_{\rho,H}^{\mathrm{ms},h}, \bar{p}_{\rho,H}^{\mathrm{ms},h})$ be the solution of the optimal control problem where $V_* = V_H^{\mathrm{ms},h}$.

### Error Estimate in $L_2(\Omega)$

$$\|\bar{u} - \bar{u}_{\rho,H,}^{\mathrm{ms},h}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,H}^{\mathrm{ms},h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,H}^{\mathrm{ms},h}\|_{L_2(\Omega)}$$
$$\leq C_{\clubsuit}(\|\bar{y} - R_H^{\mathrm{ms},h}\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_H^{\mathrm{ms},h}\bar{p}\|_{L_2(\Omega)} + \rho)$$

where $R_H^{\mathrm{ms},h}$ is the Ritz projection operator from $H_0^1(\Omega)$ onto $V_H^{\mathrm{ms},h}$.

### Theorem

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

# Error Estimates

Let $(\bar{u}^{\mathrm{ms},h}_{\rho,H,}, \bar{y}^{\mathrm{ms},h}_{\rho,H}, \bar{p}^{\mathrm{ms},h}_{\rho,H})$ be the solution of the optimal control problem where $V_* = V_H^{\mathrm{ms},h}$.

## Error Estimate in $L_2(\Omega)$

$$\begin{aligned}
&\|\bar{u} - \bar{u}^{\mathrm{ms},h}_{\rho,H,}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}^{\mathrm{ms},h}_{\rho,H}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}^{\mathrm{ms},h}_{\rho,H}\|_{L_2(\Omega)} \\
&\qquad \leq C_{\clubsuit}(\|\bar{y} - R_H^{\mathrm{ms},h}\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_H^{\mathrm{ms},h}\bar{p}\|_{L_2(\Omega)} + \rho) \\
&\qquad \leq C\big(\|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} \\
&\qquad\qquad + H^2 + \rho\big)
\end{aligned}$$

# Error Estimates

Let $(\bar{u}^{\mathrm{ms},h}_{\rho,H,}, \bar{y}^{\mathrm{ms},h}_{\rho,H}, \bar{p}^{\mathrm{ms},h}_{\rho,H})$ be the solution of the optimal control problem where $V_* = V_H^{\mathrm{ms},h}$.

## Error Estimate in $L_2(\Omega)$

$$\begin{aligned}
&\|\bar{u} - \bar{u}^{\mathrm{ms},h}_{\rho,H,}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}^{\mathrm{ms},h}_{\rho,H}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}^{\mathrm{ms},h}_{\rho,H}\|_{L_2(\Omega)} \\
&\qquad \leq C_{\clubsuit}(\|\bar{y} - R_H^{\mathrm{ms},h}\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_H^{\mathrm{ms},h}\bar{p}\|_{L_2(\Omega)} + \rho) \\
&\qquad \leq C\big(\|\bar{y} - \bar{y}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{u} - \bar{u}_{\rho,h}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,h}\|_{L_2(\Omega)} \\
&\qquad\qquad + H^2 + \rho\big)
\end{aligned}$$

## Classical PDE Constraint on Convex Domains

$$\|\bar{u} - \bar{u}_{\rho,H}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,H}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,H}\|_{L_2(\Omega)} \leq C(H^2 + \rho)$$

## Error Estimate in the Energy Norm

$$\|\bar{y}-\bar{y}^{\mathrm{ms},h}_{\rho,H}\|_a+\|\bar{p}-\bar{p}^{\mathrm{ms},h}_{\rho,H}\|_a \leq C\big(\|\bar{y}-\bar{y}_{\rho,h}\|_a+\|\bar{p}-\bar{p}_{\rho,h}\|_a+H+\rho\big)$$

## Classical PDE Constraint on Convex Domains

$$\|\bar{y}-\bar{y}_{\rho,H}\|_a+\|\bar{p}-\bar{p}_{\rho,H}\|_a \leq C(H+\rho)$$

# Summary

- The new error estimates

$$\|\bar{u} - \bar{u}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{y} - \bar{y}_{\rho,*}\|_{L_2(\Omega)} + \|\bar{p} - \bar{p}_{\rho,*}\|_{L_2(\Omega)}$$
$$\leq C_{\clubsuit}(\|\bar{y} - R_*\bar{y}\|_{L_2(\Omega)} + \|\bar{p} - R_*\bar{p}\|_{L_2(\Omega)} + \rho)$$

$$\|\bar{y} - \bar{y}_{\rho,*}\|_a + \|\bar{p} - \bar{p}_{\rho,*}\|_a \leq C_{\spadesuit}(\|\bar{y} - R_*\bar{y}\|_a + \|\bar{p} - R_*\bar{p}\|_a + \rho)$$

reduce the error analysis of finite element methods for the elliptic distributed optimal control problems with pointwise control constraints to the error analysis for elliptic boundary value problems.

They are useful for the analysis of multiscale finite element methods for PDE constraints with rough coefficients.

## New Error Analysis

B. and L.-Y. Sung

New error estimates for an elliptic distributed optimal control problem with pointwise control constraints

Beijing Journal of Pure and Applied Mathematics (2025)

# Summary

## New Error Analysis

B. and L.-Y. Sung

New error estimates for an elliptic distributed optimal control problem with pointwise control constraints

Beijing Journal of Pure and Applied Mathematics (2025)

## Numerical Results

B. , J. Garay and L.-Y. Sung

A multiscale finite element method for an elliptic distributed optimal control problem with rough coefficients and control constraints

Journal of Scientific Computing (2024)

## Summary

- The new error analysis has been extended to elliptic Neumann boundary control problems.

B. and L.-Y. Sung

A new error analysis for finite element methods for elliptic Neumann boundary control problems with pointwise control constraints

Results in Applied Mathematics (2025)

# Monge-Ampère Equation

# Two Dimensional Monge-Ampère Equation

$$\det D^2 u = f(x, u, Du)$$

where

$$D^2 u = \begin{bmatrix} u_{x_1 x_1} & u_{x_1 x_2} \\ u_{x_2 x_1} & u_{x_2 x_2} \end{bmatrix} \quad \text{and} \quad Du = \begin{bmatrix} u_{x_1} & u_{x_2} \end{bmatrix}$$

or

$$u_{x_1 x_1} u_{x_2 x_2} - u_{x_1 x_2} u_{x_2 x_1} = f(x, u, u_{x_1}, u_{x_2})$$

# Two Dimensional Monge-Ampère Equation

$$\det D^2 u = f(x, u, Du)$$

where

$$D^2 u = \begin{bmatrix} u_{x_1 x_1} & u_{x_1 x_2} \\ u_{x_2 x_1} & u_{x_2 x_2} \end{bmatrix} \quad \text{and} \quad Du = \begin{bmatrix} u_{x_1} & u_{x_2} \end{bmatrix}$$

or

$$u_{x_1 x_1} u_{x_2 x_2} - u_{x_1 x_2} u_{x_2 x_1} = f(x, u, u_{x_1}, u_{x_2})$$

It is a second order partial differential equation that is nonlinear in the second order derivatives, i.e., it is a fully nonlinear partial differential equation.

# Two Dimensional Monge-Ampère Equation

$$\det D^2 u = f(x, u, Du)$$

where

$$D^2 u = \begin{bmatrix} u_{x_1 x_1} & u_{x_1 x_2} \\ u_{x_2 x_1} & u_{x_2 x_2} \end{bmatrix} \quad \text{and} \quad Du = \begin{bmatrix} u_{x_1} & u_{x_2} \end{bmatrix}$$

or

$$u_{x_1 x_1} u_{x_2 x_2} - u_{x_1 x_2} u_{x_2 x_1} = f(x, u, u_{x_1}, u_{x_2})$$

It appears in

- the prescribed Gaussian curvature problem
- the Minkowski problem
- optimal transport
-
-
-

# Linearization of the Monge-Amp`ere operator

$$\frac{d}{dt} \det D^2(u+tv)\Big|_{t=0} = \operatorname{Cof} D^2 u : D^2 v$$

Frobenius inner product

where

$$\operatorname{Cof} D^2 u = \begin{bmatrix} u_{x_2x_2} & -u_{x_2x_1} \\ -u_{x_1x_2} & u_{x_1x_1} \end{bmatrix}$$

The linearized Monge-Amp`ere operator is elliptic if and only if $u$ is strictly convex ($\Rightarrow \det D^2 u > 0$).

# Linearization of the Monge-Amp`ere operator

$$\frac{d}{dt} \det D^2(u+tv)\Big|_{t=0} = \operatorname{Cof} D^2u : D^2v$$

Frobenius inner product

where

$$\operatorname{Cof} D^2u = \begin{bmatrix} u_{x_2x_2} & -u_{x_2x_1} \\ -u_{x_1x_2} & u_{x_1x_1} \end{bmatrix}$$

The linearized Monge-Amp`ere operator is elliptic if and only if $u$ is strictly convex ($\Rightarrow \det D^2u > 0$).

Therefore we assume $f$ is strictly positive and look for strictly convex solutions of

$$\det D^2u = f(x,u,Du)$$

## A Dirichlet Boundary Value Problem

$$\begin{aligned} \det D^2 u &= \psi \qquad \text{in } \Omega \\ u &= \phi \qquad \text{on } \partial\Omega \end{aligned}$$

where $\Omega$ is a convex polygonal domain.

We assume that there exists a smooth strictly convex solution (which is necessarily unique).

# A Dirichlet Boundary Value Problem

$$\begin{aligned} \det D^2 u &= \psi \qquad \text{in } \Omega \\ u &= \phi \qquad \text{on } \partial\Omega \end{aligned}$$

where $\Omega$ is a convex polygonal domain.

We assume that there exists a smooth strictly convex solution (which is necessarily unique).

This is a challenging problem for two reasons.

- It is a fully nonlinear PDE and therefore the usual finite element approach based on integration by parts is not applicable.
- We have to compute a strictly convex solution.

$$\begin{aligned} \det D^2 u &= \psi \qquad \text{in } \Omega \\ u &= \phi \qquad \text{on } \partial\Omega \end{aligned}$$

Observation

If

$$\det D^2 u = \psi > 0$$

then $u$ is strictly convex if and only if $\Delta u \geq 0$.

$$\lambda_1, \lambda_2 > 0 \quad \Leftrightarrow \quad \begin{cases} \lambda_1 \lambda_2 & > 0 \\ \lambda_1 + \lambda_2 & \geq 0 \end{cases}$$

## A Dirichlet Boundary Value Problem

$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \end{aligned}$$

### Observation

If

$$\det D^2 u = \psi > 0$$

then $u$ is strictly convex if and only if $\Delta u \geq 0$.

### New Formulation

$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \\ \Delta u &\geq 0 && \text{in } \Omega \end{aligned}$$

## A Dirichlet Boundary Value Problem

$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \end{aligned}$$

### Observation

If

$$\det D^2 u = \psi > 0$$

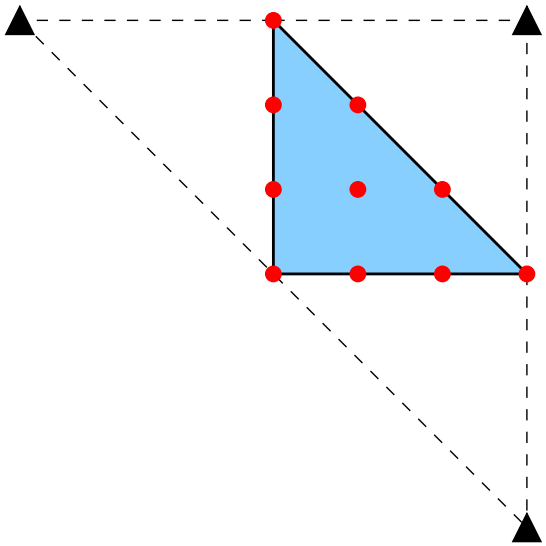
then $u$ is strictly convex if and only if $\Delta u \geq 0$.

### New Formulation

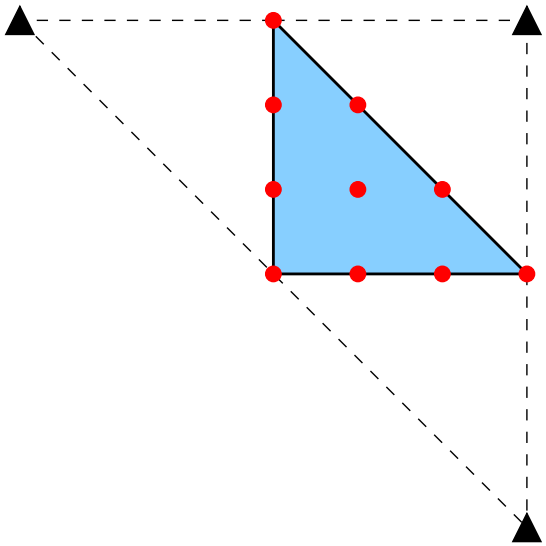
$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \\ \Delta u &\geq 0 && \text{in } \Omega \end{aligned}$$

Therefore we need to enforce the constraint $\Delta u \geq 0$.

# An Exotic Finite Element (enhanced cubic Lagrange element)

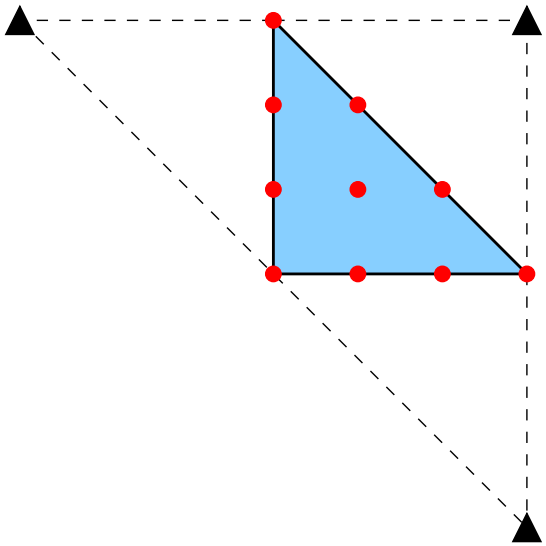


# An Exotic Finite Element (enhanced cubic Lagrange element)



We use cubic polynomials plus the linear combinations of three higher order polynomials on the triangle $T$ (shaded in blue).

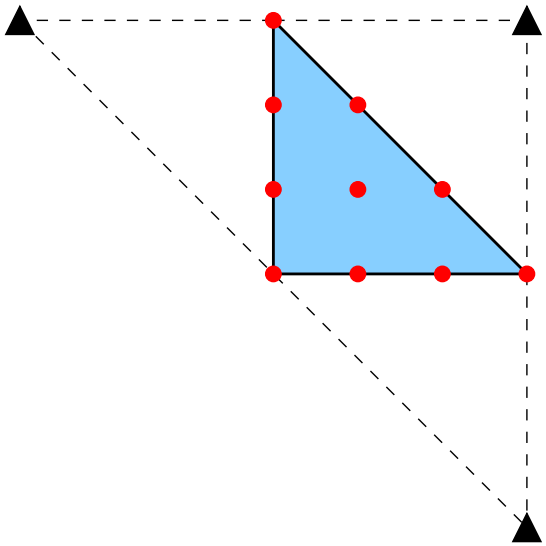
# An Exotic Finite Element (enhanced cubic Lagrange element)



We use cubic polynomials plus the linear combinations of three higher order polynomials on the triangle $T$ (shaded in blue).

Degrees of Freedom: In addition to the standard dofs of $v$ for the cubic polynomials (represented by the red dots), we also include $\Delta v$ at the vertices of the larger triangle $T_\dagger$ (represented by the black triangles), where $T_\dagger$ is the triangle whose midpoints are the vertices of $T$.

# An Exotic Finite Element (enhanced cubic Lagrange element)

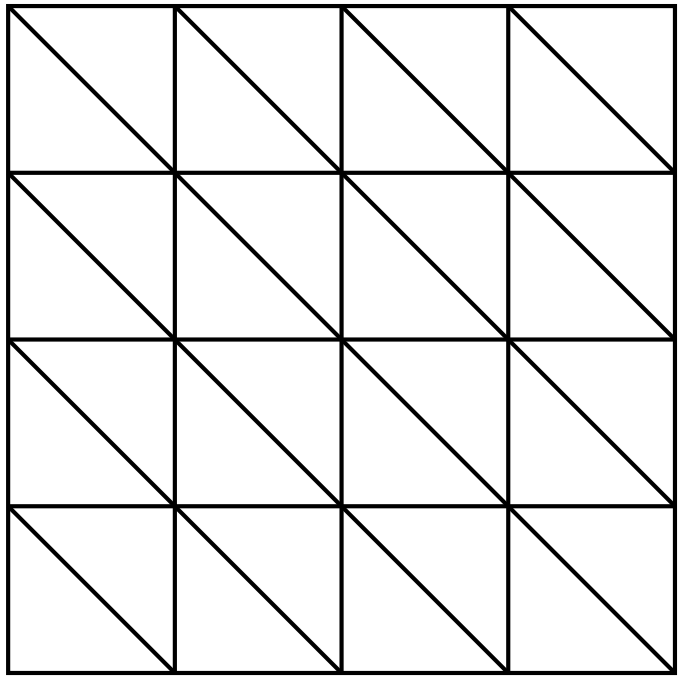


We use cubic polynomials plus the linear combinations of three higher order polynomials on the triangle $T$ (shaded in blue).

Degrees of Freedom: In addition to the standard dofs of $v$ for the cubic polynomials (represented by the red dots), we also include $\Delta v$ at the vertices of the larger triangle $T_\dagger$ (represented by the black triangles), where $T_\dagger$ is the triangle whose midpoints are the vertices of $T$.

The three additional dofs enable us to enforce the constraint $\Delta u \geq 0$.

# An Exotic Finite Element (enhanced cubic Lagrange element)



Given a triangulation $\mathcal{T}_h$ of $\Omega$, we can glue the enhanced cubic Lagrange elements on each triangle together to form a finite element space $V_h$.

$$\det D^2 u = \psi \quad \text{in } \Omega, \quad u = \phi \quad \text{on } \partial\Omega, \quad \Delta u \geq 0 \quad \text{in } \Omega$$

We will find an approximation $u_h$ of $u$ by solving a nonlinear least-squares problem on the constraint set

$$L_h = \{v \in V_h : v = \phi_h \text{ on } \partial\Omega \text{ and } \Delta v_T \geq 0 \text{ at the vertices of } T_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

where $\phi_h$ is the one-dimensional cubic Lagrange interpolant of $\phi$ along $\partial\Omega$ and $v_T$ is the restriction of $v$ to $T$.

$$\det D^2 u = \psi \quad \text{in } \Omega, \quad u = \phi \quad \text{on } \partial\Omega, \quad \Delta u \geq 0 \quad \text{in } \Omega$$

We will find an approximation $u_h$ of $u$ by solving a nonlinear least-squares problem on the constraint set

$$L_h = \{v \in V_h : v = \phi_h \text{ on } \partial\Omega \text{ and } \Delta v_T \geq 0 \text{ at the vertices of } T_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

Find

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2} \|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2} \sum_{T \in \mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2 + \frac{1}{2} \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial v / \partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2} \|\det D_h^2 v - \psi\|_{L_2(\Omega)}^2$$

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2}\|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2$$
$$+ \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 v - \psi\|_{L_2(\Omega)}^2$$

- $(D_h^2 v)\big|_T = D^2(v\big|_T)$ and $(\Delta_h v)\big|_T = \Delta(v\big|_T) \qquad \forall\, T \in \mathcal{T}_h$

  are the piecewise defined Hessian and Laplacian of $v \in V_h$.

- $\mathcal{E}_h^i$ is the set of the interior edges of $\mathcal{T}_h$.

- $[\![\partial v/\partial n]\!]$ is the jump of the normal derivative of $v$.

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2} \|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2} \sum_{T \in \mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2$$
$$+ \frac{1}{2} \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial v / \partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2} \| \det D_h^2 v - \psi \|_{L_2(\Omega)}^2$$

The last term

$$\frac{1}{2} \| \det D_h^2 v - \psi \|_{L_2(\Omega)}^2$$

is the least-squares discretization of the Monge-Ampère equation

$$\det D^2 u = \psi$$

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2}\|D_h^2 v\|^2_{L_2(\Omega)} + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|^2_{H^2(T)}$$
$$+ \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|^2_{L_2(e)} + \frac{1}{2}\|\det D_h^2 v - \psi\|^2_{L_2(\Omega)}$$

The third term ($|e|$ is the length of the edge $e$)

$$\frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|^2_{L_2(e)}$$

is a penalty term that compensates for $V_h \not\subset H^2(\Omega)$ where the strong solution of the Monge-Ampère equation is posed.

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2}\|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2$$
$$+ \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[[\partial v/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 v - \psi\|_{L_2(\Omega)}^2$$

The first term

$$\frac{h^4}{2}\|D_h^2 v\|_{L_2(\Omega)}^2$$

is a regularization term which together with the penalty term ensures that the discrete problem has a global minimizer.

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2}\|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 v - \psi\|_{L_2(\Omega)}^2$$

We have

$$\|v\|_{L_2(\Omega)} \le C\left(\|D_h^2 v\|_{L_2(\Omega)} + \sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2\right)$$

for continuous piecewise $H^2$ functions $v$ that vanish on $\partial\Omega$, where the positive constant $C$ depends only on $\Omega$ and the shape regularity of $\mathcal{T}_h$.

B.-Wang-Zhao 2004

$$u_h = \operatorname*{argmin}_{v \in L_h} \mathcal{J}_h(v)$$

where

$$\mathcal{J}_h(v) = \frac{h^4}{2}\|D_h^2 v\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2$$
$$+ \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 v - \psi\|_{L_2(\Omega)}^2$$

The second term

$$\frac{h^4}{2}\sum_{T\in\mathcal{T}_h} |\Delta_h v|_{H^2(T)}^2$$

is a regularization term that enforces the elementwise convexity of the discrete solution.

The interpolation operator $\Pi_h : H^4(\Omega) \longrightarrow V_h$ is defined by the condition that $\zeta$ and $\Pi_h \zeta$ have identical dofs.

The interpolation operator $\Pi_h : H^4(\Omega) \longrightarrow V_h$ is defined by the condition that $\zeta$ and $\Pi_h \zeta$ have identical dofs.

Let $u$ be the solution of the boundary value problem

$$\begin{aligned} \det D^2 u &= \psi \qquad \text{in } \Omega \\ u &= \phi \qquad \text{on } \partial\Omega \end{aligned}$$

By design $\Pi_h u$ belongs to the constraint set

$$L_h = \{v \in V_h : v = \phi_h \text{ on } \partial\Omega \text{ and } \operatorname{tr}(J_T^t D^2 v_T J_T) \geq 0 \text{ at the vertices of } T_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

$$\frac{h^4}{2}\|D_h^2 u_h\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h u_h|_{H^2(T)}^2$$
$$+\frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[[\partial u_h/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)}^2$$
$$= \mathcal{J}_h(u_h)$$

$$\frac{h^4}{2}\|D_h^2 u_h\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h u_h|_{H^2(T)}^2$$

$$+\frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[[\partial u_h/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)}^2$$

$$= \mathcal{J}_h(u_h)$$

$$\leq \mathcal{J}_h(\Pi_h u)$$

$$= \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}\|D_h^2\Pi_h u\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h\Pi_h u|_{H^2(T)}^2$$

$$+\frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[[\partial\Pi_h u/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2\Pi_h u - \psi\|_{L_2(\Omega)}^2$$

$$\frac{h^4}{2}\|D_h^2 u_h\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h u_h|_{H^2(T)}^2$$

$$+\frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)}^2$$

$$= \mathcal{J}_h(u_h)$$

$$\leq \mathcal{J}_h(\Pi_h u)$$

$$= \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}\|D_h^2\Pi_h u\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h\Pi_h u|_{H^2(T)}^2$$

$$+\frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h u/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2\Pi_h u - \psi\|_{L_2(\Omega)}^2$$

$$\leq Ch^4$$

interpolation error estimates

$$\begin{aligned}
&\frac{h^4}{2}\|D_h^2 u_h\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h u_h|_{H^2(T)}^2 \\
&\quad + \frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[[\partial u_h/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)}^2 \\
&\quad = \mathcal{J}_h(u_h) \\
&\quad \leq \mathcal{J}_h(\Pi_h u) \\
&\quad = \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}\|D_h^2\Pi_h u\|_{L_2(\Omega)}^2 + \frac{h^4}{2}\sum_{T\in\mathcal{T}_h}|\Delta_h\Pi_h u|_{H^2(T)}^2 \\
&\qquad + \frac{1}{2}\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[[\partial\Pi_h u/\partial n]]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2\Pi_h u - \psi\|_{L_2(\Omega)}^2 \\
&\quad \leq Ch^4
\end{aligned}$$

$$\|D_h^2 u_h\|_{L_2(\Omega)} \leq C$$

$$\Big( \sum_{T \in \mathcal{T}_h} |\Delta_h u_h|^2_{H^2(T)} \Big)^{\frac{1}{2}} \leq C$$

$$\Big( \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial u_h / \partial n]\!]\|^2_{L_2(e)} \Big)^{\frac{1}{2}} \leq Ch^2,$$

$$\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)} \leq Ch^2$$

$$\|D_h^2 u_h\|_{L_2(\Omega)} \leq C$$

$$\Big( \sum_{T \in \mathcal{T}_h} |\Delta_h u_h|^2_{H^2(T)} \Big)^{\frac{1}{2}} \leq C$$

$$\Big( \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial u_h / \partial n]\!]\|^2_{L_2(e)} \Big)^{\frac{1}{2}} \leq Ch^2,$$

$$\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)} \leq Ch^2$$

These *a priori* bounds are crucial for the error analysis of the nonlinear least-squares method. They are based on the important fact that $\Pi_h u$ belongs to the constraint set $L_h$.

$$\|D_h^2 u_h\|_{L_2(\Omega)} \leq C$$

$$\Big( \sum_{T \in \mathcal{T}_h} |\Delta_h u_h|^2_{H^2(T)} \Big)^{\frac{1}{2}} \leq C$$

$$\Big( \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial u_h / \partial n]\!]\|^2_{L_2(e)} \Big)^{\frac{1}{2}} \leq Ch^2,$$

$$\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)} \leq Ch^2$$

These *a priori* bounds are crucial for the error analysis of the nonlinear least-squares method. They are based on the important fact that $\Pi_h u$ belongs to the constraint set $L_h$.

In particular, one can use these estimates to show that $u_h$ is strictly convex on each triangle in $\mathcal{T}_h$ if $h$ is sufficiently small.

By the fundamental theorem of calculus, we have

$$(*) \quad \det D^2 u - \det D_h^2 u_h = A_h : D_h^2 (u - u_h)$$

where

$$A_h(x) = \int_0^1 \Big[ \mathrm{Cof} D_h^2 (tu + (1-t)u_h) \Big](x) dt$$

is defined at all the points in $\Omega$ except those on the edges of $\mathcal{T}_h$.

$$\Big[ \frac{d}{dt} \det(M + tN) \Big]_{t=0} = \mathrm{Cof}\, M : N \qquad \forall\, M, N \in \mathbb{R}^{n \times n}$$

By the fundamental theorem of calculus, we have

$$(*) \quad \det D^2 u - \det D_h^2 u_h = A_h : D_h^2 (u - u_h)$$

where

$$A_h(x) = \int_0^1 \left[ \mathrm{Cof} D_h^2 (tu + (1-t)u_h) \right](x) dt$$

is defined at all the points in $\Omega$ except those on the edges of $\mathcal{T}_h$.

Since $u$ is strictly convex on $\bar{\Omega}$ and $u_h$ is strictly convex on each triangle, the symmetric matrix $A_h(x)$ is also positive definite on each triangle

By the fundamental theorem of calculus, we have

$$(*) \quad \det D^2 u - \det D_h^2 u_h = A_h : D_h^2(u - u_h)$$

where

$$A_h(x) = \int_0^1 \left[\mathrm{Cof} D_h^2(tu + (1-t)u_h)\right](x)dt$$

is defined at all the points in $\Omega$ except those on the edges of $\mathcal{T}_h$.

Since $u$ is strictly convex on $\bar{\Omega}$ and $u_h$ is strictly convex on each triangle, the symmetric matrix $A_h(x)$ is also positive definite on each triangle

Therefore $(*)$ is an elliptic equation in nondivergence form for $u - u_h$.

## Error Estimates

Combining the techniques for the error analysis of discontinuous Galerkin methods for elliptic equations in nondivergence form

Smears-Süli 2013
Neilan-Wu 2019

and the *a priori* bounds for $u_h$, we can derive the error estimate

$$\|u - u_h\|_h \leq Ch^2$$

where $u_h = \underset{v \in L_h}{\operatorname{argmin}} \mathcal{J}_h(v)$ and

$$\|v\|_h = \left( \|D_h^2 v\|_{L_2(\Omega)}^2 + \sum_{e \in \mathcal{E}_h^i} |e|^{-1} \|[\![\partial v / \partial n]\!]\|_{L_2(e)}^2 \right)^{\frac{1}{2}}$$

## Error Estimates

Combining the techniques for the error analysis of discontinuous Galerkin methods for elliptic equations in nondivergence form

Smears-Süli 2013
Neilan-Wu 2019

and the *a priori* bounds for $u_h$, we can derive the error estimate

$$\|u - u_h\|_h \leq Ch^2$$

Corollary

$$\|u - u_h\|_{L_2(\Omega)} + |u - u_h|_{H^1(\Omega)} + \|u - u_h\|_{L_\infty(\Omega)} \leq Ch^2$$

Poincaré-Friedrichs and Sobolev inequalities for piecewise $H^2$ functions

B.-Wang-Zhao 2004
B.-Neilan-Reiser-Sung 2017

## Summary

- By introducing an exotic convexity enforcing finite element, we can solve the constrained Dirichlet boundary value problem for the Monge-Ampère equation

$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \\ \Delta u &\geq 0 && \text{in } \Omega \end{aligned}$$

as a nonlinear least-squares problem.

The method is efficient due to the $O(h^2)$ convergence.

## Summary

- By introducing an exotic convexity enforcing finite element, we can solve the constrained Dirichlet boundary value problem for the Monge-Ampère equation

$$\begin{aligned} \det D^2 u &= \psi && \text{in } \Omega \\ u &= \phi && \text{on } \partial\Omega \\ \Delta u &\geq 0 && \text{in } \Omega \end{aligned}$$

  as a nonlinear least-squares problem.

  The method is efficient due to the $O(h^2)$ convergence.

B., L.-Y. Sung, Z. Tan and H. Zhang

A convexity enforcing $C^0$ interior penalty method for the Monge-Ampère equation on convex polygonal domains

Numerische Mathematik (2021)

## Summary

- The results for convex polygonal domains have been extended to smooth strictly convex domains by using a Hermite isoparametric finite element space with exotic degrees of freedom.

A nonlinear least-squares convexity enforcing $C^0$ interior penalty method for the Monge–Ampère equation on strictly convex smooth planar domains

Communications of the AMS (2024)

## Summary

Our approach can be applied to other fully nonlinear elliptic partial differential equations such as

- the Pucci equation
- the 2-Hessian equation
- 
- 
- 

B., L-Y. Sung and Z. Tan

A finite element method for a two-dimensional Pucci equation

Comptes Rendus Mécanique (2023)

An adaptive nonlinear least-squares finite element method for a Pucci equation in two dimensions

East Asian Journal on Applied Mathematics (2024)