A nonlinear least-squares convexity enforcing finite element method for the Monge-Ampère equation in two dimensions

Susanne C. Brenner

Department of Mathematics
and
Center for Computation & Technology

Louisiana State University

PDEs and Geometry: Numerical Aspects

ICERM, March 11-15, 2024

A nonlinear least-squares convexity enforcing finite element method for the Monge-Ampère equation in two dimensions

Joint work with

Li-yeng Sung, Zhiyu Tan and Hongchao Zhang

Supported by NSF

Outline

Outline

- The Monge-Ampère Equation

Outline

- The Monge-Ampère Equation
- Convex Polygonal Domains

Outline

- The Monge-Ampère Equation
- Convex Polygonal Domains
- Convex Smooth Domains

Outline

- The Monge-Ampère Equation
- Convex Polygonal Domains
- Convex Smooth Domains
- Concluding Remarks

Outline

- The Monge-Ampère Equation
- Convex Polygonal Domains
- Convex Smooth Domains
- Concluding Remarks

B.-Sung-Tan-Zhang

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

Numerische Mathematik 2021

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

Preprint 2023

The Monge-Ampère Equation in Two Dimensions

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

where

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

or

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

The Monge-Ampère Equation in Two Dimensions

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

where

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

or

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

This 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.

Linearization of the Monge-Ampère Operator

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

Frobenius inner product

where

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

Linearization of the Monge-Ampère Operator

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

Frobenius inner product

where

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

Consequently the linearized Monge-Amp̀ere operator is elliptic if and only if $\operatorname{Cof} D^2u$ is SPD, which is equivalent to D^2u being SPD, i.e., u is strictly convex ($\Rightarrow \det D^2u > 0$).

Linearization of the Monge-Ampère Operator

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

Frobenius inner product

where

$$\operatorname{Cof} D^2 u = \begin{bmatrix} u_{x_2 x_2} & -u_{x_2 x_1} \\ -u_{x_1 x_2} & u_{x_1 x_1} \end{bmatrix}$$

Consequently the linearized Monge-Ampère operator is elliptic if and only if $\operatorname{Cof} D^2 u$ is SPD, which is equivalent to $D^2 u$ being SPD, i.e., u is strictly convex ($\Rightarrow \det D^2 u > 0$).

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

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

Prescribed Gaussian Curvature

Prescribed Gaussian Curvature

Let $\Sigma \subset \mathbb{R}^3$ be the graph of a function $u(x)$ of two variables. The Gaussian curvature $K(x)$ of Σ at the point $(x, u(x))$ over x is given by the formula

$$K = \frac{\det D^2 u}{(1 + |Du|^2)^2}$$

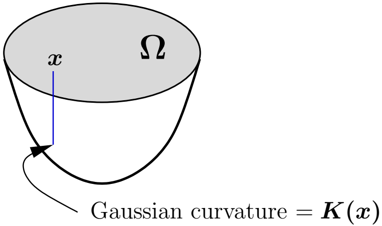
Prescribed Gaussian Curvature

Let $\Sigma \subset \mathbb{R}^3$ be the graph of a function $u(x)$ of two variables. The Gaussian curvature $K(x)$ of Σ at the point $(x, u(x))$ over x is given by the formula

$$K = \frac{\det D^2 u}{(1 + |Du|^2)^2}$$

Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain and $K(x)$ be a positive function defined on Ω. The prescribed Gaussian curvature problem is to find a surface Σ passing through $\partial\Omega$ so that Σ is the graph of a function and the Gaussian curvature of Σ over any point x equals $K(x)$.

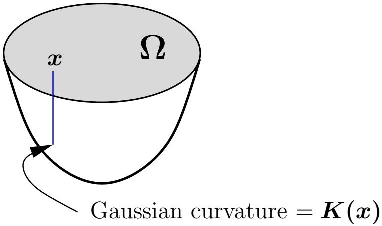
Prescribed Gaussian Curvature



$$K = \frac{\det D^2 u}{(1 + |Du|^2)^2}$$

Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain and $K(x)$ be a positive function defined on Ω. The prescribed Gaussian curvature problem is to find a surface Σ passing through $\partial\Omega$ so that Σ is the graph of a function and the Gaussian curvature of Σ over any point x equals $K(x)$.

Prescribed Gaussian Curvature



$$K = \frac{\det D^2 u}{(1 + |Du|^2)^2}$$

We can construct Σ as the graph of a function u by solving a Dirichlet boundary value problem for a Monge-Ampère equation.

$$\det D^2 u = K(1 + |Du|^2)^2 \qquad \text{in } \Omega$$
$$u = 0 \qquad \text{on } \partial\Omega$$

Optimal Transport Monge (1780's) Kantorovich (1940's)

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

Ω (original domain) and Ω_* (targeted domain) are bounded open connected subsets of $\mathbb{R}^2$.

ϕ is a positive measurable function on Ω bounded away from 0 and ∞ (mass density on Ω).

ϕ_* is a positive measurable function on Ω_* bounded away from 0 and ∞ (mass density on Ω_*).

A material originally distributed in Ω according to ϕ is to be transferred to Ω_* so that it is distributed according to ϕ_*.

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

Ω (original domain) and Ω_* (targeted domain) are bounded open connected subsets of $\mathbb{R}^2$.

ϕ is a positive measurable function on Ω bounded away from 0 and ∞ (mass density on Ω).

ϕ_* is a positive measurable function on Ω_* bounded away from 0 and ∞ (mass density on Ω_*).

A material originally distributed in Ω according to ϕ is to be transferred to Ω_* so that it is distributed according to ϕ_*.

This requires a necessary mass balance condition.

$$\int_{\Omega} \phi(x)dx = \int_{\Omega_*} \phi_*(y)dy$$

Optimal Transport Monge (1780's) Kantorovich (1940's)

An admissible mass transfer is a measurable one-to-one map $s : \Omega \longrightarrow \Omega_*$ such that the mass balance condition

$$\int_{s^{-1}(E)} \phi(x)dx = \int_E \phi_*(y)dy$$

is satisfied by every open subset E of Ω_*.

The set of all admissible mass transfers is denoted by $\mathcal{A}$.

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

An admissible mass transfer is a measurable one-to-one map $s : \Omega \longrightarrow \Omega_*$ such that the mass balance condition

$$\int_{s^{-1}(E)} \phi(x)dx = \int_E \phi_*(y)dy$$

is satisfied by every open subset E of Ω_*.

The set of all admissible mass transfers is denoted by $\mathcal{A}$.

For a sufficiently smooth mass transfer s, the mass balance condition becomes

$$\phi_*(s(x)) \det Ds(x) = \phi(x) \qquad \forall\, x \in X$$

by the change of variables formula.

(Ds is the Jacobian matrix of s.)

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

The optimal transport problem is to find

$$s = \underset{s \in \mathcal{A}}{\operatorname{argmin}} \int_{\Omega} c\big(x, s(x)\big)\phi(x)dx$$

where

$$c(x, y) : \Omega \times \Omega_* \longrightarrow (0, \infty)$$

represents the cost of transporting a unit mass from point x to point y and

$$\int_{\Omega} c\big(x, s(x)\big)\phi(x)dx$$

is the total cost of using the mass transfer s.

Optimal Transport Monge (1780's) Kantorovich (1940's)

In the case where

$$c(x, y) = |x - y|^2$$

the optimal transfer s is the gradient of a convex function u.

Knott–Smith (1984)
Brenier (1991)

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

In the case where

$$c(x,y) = |x-y|^2$$

the optimal transfer s is the gradient of a convex function u.

The condition that s is admissible, i.e.,

s maps Ω into Ω_*

$$\phi_*(s(x)) \det Ds(x) = \phi(x) \qquad \text{in } \Omega$$

can then be expressed in terms of u:

Du maps Ω into Ω_*

$$\phi_*(Du) \det D^2 u = \phi \qquad \text{in } \Omega$$

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

The optimal transport problem is therefore equivalent to the Monge-Ampère mapping problem consisting of a Monge-Ampère equation

$$\det D^2 u = \phi / \phi_*(Du) \quad \text{in } \Omega$$

and a state constraint

$$Du(x) \in \Omega_* \qquad \forall\, x \in \Omega$$

Optimal Transport   Monge (1780's)   Kantorovich (1940's)

The optimal transport problem is therefore equivalent to the Monge-Ampère mapping problem consisting of a Monge-Ampère equation

$$\det D^2 u = \phi / \phi_*(Du) \quad \text{in } \Omega$$

and a state constraint

$$Du(x) \in \Omega_* \qquad \forall\, x \in \Omega$$

Optimal transport with $c(x, y) = |x - y|^2$ has applications to

- Nonlinear interpolation
- Nonlinear diffusion
- Boltzmann equation
- Incompressible fluids

Numerical Methods for the Monge-Ampère Equation

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods

  Oliker-Prussner (1988), Neilan-Zhang (2018), Nochetto-Zhang (2019), Qiu-Tang (2020), Awanou (2023)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods

  Barles-Souganidis (1991), Kuo-Trudinger (1992), Oberman (2008), Benamou-Froese-Oberman (2010), Froese-Oberman (2011, 2013), Mirebeau (2015), Benamou-Collino-Mirebeau (2016), Hamfeldt-Salvador (2018), Hamfeldt-Lesniewski (2022), Bonnet-Mirebeau (2022), Hou-Chen-Xia (2022), Brusca-Hamfeldt (2023)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods

  Dean-Glowinski (2003, 2004, 2006), Caboussat-Glowinski-Sorensen (2013), Li-Yang (2023)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods
- finite element methods based on the vanishing moment approach

  Feng-Neilan (2009, 2011), Neilan (2010), Awanou (2015), Chen-Feng-Zhang (2021)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods
- finite element methods based on the vanishing moment approach
- finite element methods based on L_2 projection

  Böhmer (2008), B.-Gudi-Neilan-Sung (2011), B.-Neilan (2012), Neilan (2013), Lakkis-Pryer (2013), Davydov-Saeed (2013), Neilan (2014), Awanou-Li (2014), Awanou (2015, 2017), Böhmer and Schaback (2019), Awanou-Li-Malitz (2020)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods
- finite element methods based on the vanishing moment approach
- finite element methods based on L_2 projection
- finite element methods based on HJB reformulation

  Feng-Jensen (2017), B.-Kawecki (2020), Gallistl-Tran (2023), Gallistl-Tran (2024)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods
- finite element methods based on the vanishing moment approach
- finite element methods based on L_2 projection
- finite element methods based on HJB reformulation
- two-scale methods

  Li-Nochetto (2018), Nochetto-Ntogkas-Zhang (2019), Nochetto-Zhang (2019), Nochetto-Ntogkas (2019)

Numerical Methods for the Monge-Ampère Equation

- geometric finite difference methods
- monotone finite difference methods
- augmented Lagrangian and least-squares finite element methods
- finite element methods based on the vanishing moment approach
- finite element methods based on L_2 projection
- finite element methods based on HJB reformulation
- two-scale methods
- fixed point iteration methods

  Benamou-Froese-Oberman (2010), Awanou (2015), Lai-Lee (2023)

We will consider the Dirichlet boundary value problem of the simplest Monge-Ampère equation

$$(*) \quad \det D^2 u = f(x) \qquad (f > 0)$$

as a stepping stone to more complicated Monge-Ampère equations.

We will consider the Dirichlet boundary value problem of the simplest Monge-Ampère equation

$$(*) \quad \det D^2 u = f(x) \qquad (f > 0)$$

as a stepping stone to more complicated Monge-Ampère equations.

Our method for computing smooth strictly convex solutions is completely different from all the methods in the literature.

It is based on the simple observation that a solution of $(*)$ is strictly convex if and only if $\Delta u \geq 0$.

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

Convex Polygonal Domain Ω

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

Convex Polygonal Domain Ω

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

Assumptions

- $\psi \in H^2(\Omega)$ is positive on $\bar{\Omega}$.
- ϕ belongs to $H^4(\Omega)$.
- There exists a unique strictly convex solution u that belongs to $H^4(\Omega)$.

Convex Polygonal Domain Ω

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

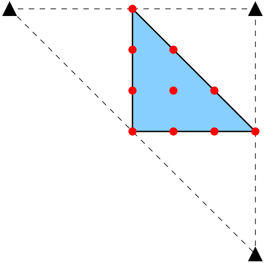
Assumptions

- $\psi \in H^2(\Omega)$ is positive on $\bar{\Omega}$.
- ϕ belongs to $H^4(\Omega)$.
- There exists a unique strictly convex solution u that belongs to $H^4(\Omega)$.

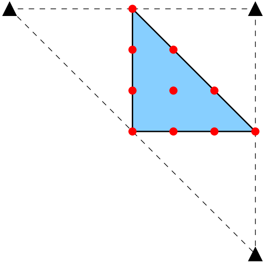
Remark In the case of a strictly convex smooth domain Ω, these assumptions are satisfied if $\psi \in C^3(\bar{\Omega})$ is positive on $\bar{\Omega}$ and $\phi \in C^{4,\delta}(\bar{\Omega})$ for some $\delta \in (0,1)$ $(\Rightarrow u \in C^{4,\delta}\bar{\Omega}))$.

Caffarelli-Nirenberg-Spruck (1984)

An Exotic Finite Element (enhanced cubic Lagrange element)



An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

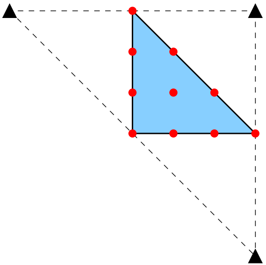
$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1 \hat{x}_2 (1 - \hat{x}_1 - \hat{x}_2)$$

An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

$\hat{T}_{\dagger}$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

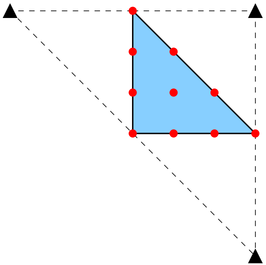
Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1 \hat{x}_2 (1 - \hat{x}_1 - \hat{x}_2)$$

Degrees of Freedom The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_{\dagger}$.

An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1\hat{x}_2(1-\hat{x}_1-\hat{x}_2)$$

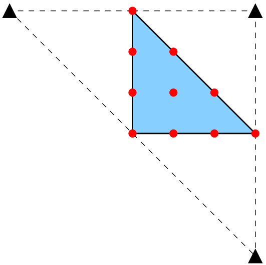
Degrees of Freedom The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

If v vanishes at the boundary nodes of $\hat{T}$, then v belongs to

$$\langle \varphi_{\hat{T}} \rangle \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

If additionally v vanishes at the center of $\hat{T}$ and Δv vanishes at the vertices of $\hat{T}_\dagger$, then $v=0$ (direct calculation).

An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

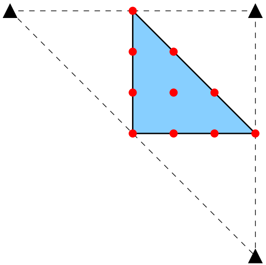
$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1 \hat{x}_2 (1 - \hat{x}_1 - \hat{x}_2)$$

Degrees of Freedom The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

Remark This is not a finite element in the classical sense of Ciarlet since three of the degrees of freedom are at nodes outside the element domain.

An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi^2_{\hat{T}} P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1\hat{x}_2(1-\hat{x}_1-\hat{x}_2)$$

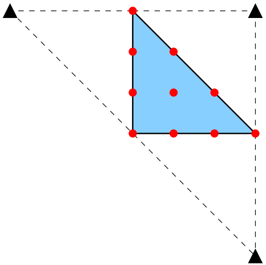
Degrees of Freedom

The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

Observation

If Δv is nonnegative at the three vertices of $\hat{T}_\dagger$, then the P_1 interpolant of Δv associated with $\hat{T}_\dagger$ is nonnegative on $\hat{T}_\dagger$ and hence also nonnegative on $\hat{T}$.

An Exotic Finite Element (enhanced cubic Lagrange element)



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1, 1)$, $(-1, 1)$ and $(1, -1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1 \hat{x}_2 (1 - \hat{x}_1 - \hat{x}_2)$$

Degrees of Freedom The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

Remark $\hat{T}_\dagger$ is the triangle whose midpoints are the vertices of the reference simplex $\hat{T}$. For a general triangle T, the corresponding triangle will be denoted by $T_\dagger$.

An Exotic Finite Element (enhanced cubic Lagrange element)

The enhanced cubic Lagrange element on an arbitrary triangle T is affine-equivalent to the element on the reference simplex.

An Exotic Finite Element (enhanced cubic Lagrange element)

The enhanced cubic Lagrange element on an arbitrary triangle T is affine-equivalent to the element on the reference simplex.

The space of shape functions is

$$P_3(T) \oplus \varphi_T^2 P_1(T)$$

where φ_T is the cubic bubble that vanishes on ∂T.

The dofs of v are given by the dofs of the standard cubic Lagrange element plus the values of $\mathrm{tr}(J_T^t D^2 v J_T)$ at the three vertices of $T_\dagger$, where $J_T \in \mathbb{R}^{2\times 2}$ is the Jacobian of an affine map Φ_T that maps the reference simplex to T

An Exotic Finite Element (enhanced cubic Lagrange element)

The enhanced cubic Lagrange element on an arbitrary triangle T is affine-equivalent to the element on the reference simplex.

The space of shape functions is

$$P_3(T) \oplus \varphi_T^2 P_1(T)$$

where φ_T is the cubic bubble that vanishes on ∂T.

The dofs of v are given by the dofs of the standard cubic Lagrange element plus the values of $\mathrm{tr}(J_T^t D^2 v J_T)$ at the three vertices of $T_\dagger$, where $J_T \in \mathbb{R}^{2\times 2}$ is the Jacobian of an affine map Φ_T that maps the reference simplex to T

Remark The values of $\mathrm{tr}(J_T^t D^2 v J_T)$ at the vertices of $T_\dagger$ are the values of $\Delta(v \circ \Phi_T)$ at the vertices of $\hat{T}_\dagger$.

An Exotic Finite Element (enhanced cubic Lagrange element)

The enhanced cubic Lagrange element on an arbitrary triangle T is affine-equivalent to the element on the reference simplex.

The space of shape functions is

$$P_3(T) \oplus \varphi_T^2 P_1(T)$$

where φ_T is the cubic bubble that vanishes on ∂T.

The dofs of v are given by the dofs of the standard cubic Lagrange element plus the values of $\mathrm{tr}(J_T^t D^2 v J_T)$ at the three vertices of $T_\dagger$, where $J_T \in \mathbb{R}^{2\times 2}$ is the Jacobian of an affine map Φ_T that maps the reference simplex to T

Remark The values of $\mathrm{tr}(J_T^t D^2 v J_T)$ at the vertices of $T_\dagger$ are the values of $\Delta(v \circ \Phi_T)$ at the vertices of $\hat{T}_\dagger$.

We will refer to $\mathrm{tr}(J_T^t D^2 v J_T)$ as the scaled Laplacian of v.

The Finite Element Space V_h

The Finite Element Space V_h

Let $\mathcal{T}_h$ be a quasi-uniform simplicial triangulation of the convex polygon Ω. A function v belongs to the finite element space V_h if and only if

- v belongs to $C(\Omega)$
- the restriction of v to $T \in \mathcal{T}_h$ belongs to $P_3(T) \oplus \varphi_T^2 P_1(T)$

The Finite Element Space V_h

Let $\mathcal{T}_h$ be a quasi-uniform simplicial triangulation of the convex polygon Ω. A function v belongs to the finite element space V_h if and only if

- v belongs to $C(\Omega)$
- the restriction of v to $T \in \mathcal{T}_h$ belongs to $P_3(T) \oplus \varphi_T^2 P_1(T)$

The global dofs of $v \in V_h$ are given by

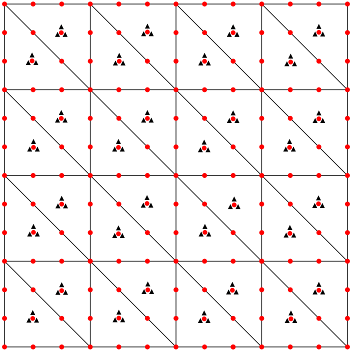
- the global dofs of the cubic Lagrange finite element space
- the values of $\mathrm{tr}(J_T^t D^2 v_T J_T)$ at the three vertices of $T_\dagger$ for each $T \in \mathcal{T}_h$, where v_T is the restriction of v to T

The Finite Element Space V_h

Let $\mathcal{T}_h$ be a quasi-uniform simplicial triangulation of the convex polygon Ω. A function v belongs to the finite element space V_h if and only if

- v belongs to $C(\Omega)$
- the restriction of v to $T \in \mathcal{T}_h$ belongs to $P_3(T) \oplus \varphi_T^2 P_1(T)$

The global dofs of $v \in V_h$ are given by



The Discrete Problem

The Discrete Problem

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

We will find an approximation u_h of u by solving a nonlinear least-squares problem with box constraints.

The Discrete Problem

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

We will find an approximation u_h of u by solving a nonlinear least-squares problem with box constraints.

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

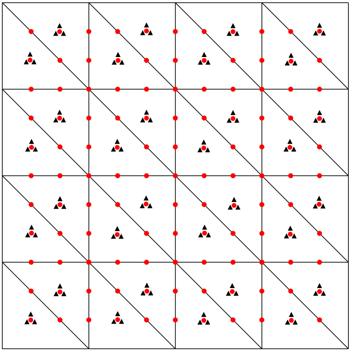
where ϕ_h is the one-dimensional cubic Lagrange interpolant of ϕ along $\partial\Omega$ and v_T is the restriction of v to T.

The Discrete Problem

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

We will find an approximation u_h of u by solving a nonlinear least-squares problem with box constraints.

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



The Discrete Problem

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

We will find an approximation u_h of u by solving a nonlinear least-squares problem with box constraints.

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

Find

$$u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$$

where

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

The Discrete Problem

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

where

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

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

  is the piecewise defined Hessian 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.

The Discrete Problem

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

where

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

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

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

The Discrete Problem

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

where

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

The third term

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

is a penalty term that compensates for $V_h \not\subset H^2(\Omega)$.

The Discrete Problem

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

where

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

The first term

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

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

The Discrete Problem

$$u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$$

where

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

We have

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

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

B.-Wang-Zhao (2004)

The Discrete Problem

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

where

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

The second term

$$\frac{1}{2} \sum_{T \in \mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 v J_T)|_{H^2(T)}^2$$

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

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

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

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

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

For the solution u of the boundary value problem, we have

- $\Pi_h u = \phi_h$ on $\partial\Omega$

where ϕ_h is the one-dimensional cubic Lagrange interpolant of ϕ along $\partial\Omega$.

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

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\operatorname{argmin}} \mathcal{J}_h(v)$

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

For the solution u of the boundary value problem, we have

- $\Pi_h u = \phi_h$ on $\partial\Omega$

where ϕ_h is the one-dimensional cubic Lagrange interpolant of ϕ along $\partial\Omega$.

Remark We can extend u to $\mathbb{R}^2$ (with the extension in $H^4(\mathbb{R}^2)$ still denoted by u) so that $\Pi_h u$ is well-defined at the nodes outside $\bar{\Omega}$. Note that the extended u is strictly convex in an open neighborhood of $\bar{\Omega}$.

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

For the solution u of the boundary value problem, we have

- $\Pi_h u = \phi_h$ on $\partial\Omega$
- For any $T \in \mathcal{T}_h$,

$$\mathrm{tr}\big(J_T^t D^2 \Pi_h u J_T\big) = \mathrm{tr}\big(J_T^t D^2 u J_T\big)$$

  is positive at the vertices of $T_\dagger$. $\qquad (h \ll 1)$

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

For the solution u of the boundary value problem, we have

- $\Pi_h u = \phi_h$ on $\partial\Omega$
- For any $T \in \mathcal{T}_h$,

$$\operatorname{tr}\big(J_T^t D^2 \Pi_h u J_T\big) = \operatorname{tr}\big(J_T^t D^2 u J_T\big)$$

  is positive at the vertices of $T_\dagger$. $\qquad (h \ll 1)$

Therefore $\Pi_h u$ belongs to L_h.

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

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

$$
\begin{aligned}
&\frac{h^4}{2}\|D_h^2 u_h\|_{L_2(\Omega)}^2 + \frac{1}{2}\sum_{T\in\mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|_{H^2(T)}^2 \\
&\quad + \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2 + \frac{1}{2}\|\det D_h^2 u_h - \psi\|_{L_2(\Omega)}^2 \\
&\quad = \mathcal{J}_h(u_h)
\end{aligned}
$$

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

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

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

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

interpolation error estimates

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

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

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

$$\Big(\sum_{T \in \mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|^2_{H^2(T)}\Big)^{\frac{1}{2}} \le Ch^2$$

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

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

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

$$\Big(\sum_{T \in \mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|^2_{H^2(T)}\Big)^{\frac{1}{2}} \leq Ch^2$$

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

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

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.

A Priori Bounds $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

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

$$\Big(\sum_{T \in \mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|^2_{H^2(T)}\Big)^{\frac{1}{2}} \leq Ch^2$$

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

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

Corollary

$$\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \leq Ch \qquad (\psi \in H^2(\Omega))$$

and hence

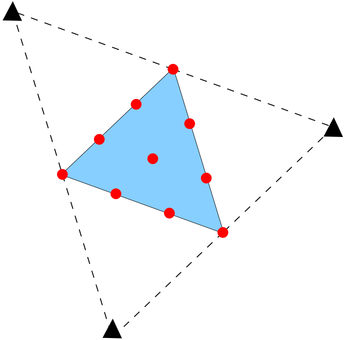
$$\det D_h^2 u_h(x) \geq \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega \text{ for } h \ll 1$$

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

Elementwise Convexity $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$



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

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$

Therefore on any $T \in \mathcal{T}_h$, we have

$$\mathrm{tr}\, D_h^2 u_h \geq C h^{-2} \mathrm{tr}(J_T^t D_h^2 u_h J_T)$$

$$|J_T| \approx h$$

Elementwise Convexity $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$

Therefore on any $T \in \mathcal{T}_h$, we have

$$\begin{aligned} \mathrm{tr}\, D_h^2 u_h &\geq C h^{-2} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \\ &\geq C h^{-2} \big[\mathrm{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \big] \end{aligned}$$

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

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$

Therefore on any $T \in \mathcal{T}_h$, we have

$$\begin{aligned}
\mathrm{tr}\, D_h^2 u_h &\geq C h^{-2} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \\
&\geq C h^{-2} \big[\mathrm{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \big] \\
&\geq -C h^{-2} \| \mathrm{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \|_{L_\infty(T)}
\end{aligned}$$

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

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \operatorname{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$

Therefore on any $T \in \mathcal{T}_h$, we have

$$\begin{aligned} \operatorname{tr} D_h^2 u_h &\geq C h^{-2} \operatorname{tr}(J_T^t D_h^2 u_h J_T) \\ &\geq C h^{-2} \big[\operatorname{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \operatorname{tr}(J_T^t D_h^2 u_h J_T)\big] \\ &\geq -C h^{-2} \|\operatorname{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \operatorname{tr}(J_T^t D_h^2 u_h J_T)\|_{L_\infty(T)} \\ &\geq -C h^{-1} |\operatorname{tr}(J_T^t D_h^2 u_h J_T)|_{H^2(T)} \end{aligned}$$

interpolation error estimate

Elementwise Convexity $u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

Let $I_{T_\dagger}$ be the P_1 nodal interpolation operator associated with $T_\dagger$. We have already observed that

$$I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \geq 0 \quad \text{on } T$$

Therefore on any $T \in \mathcal{T}_h$, we have

$$\begin{aligned}
\mathrm{tr}\, D_h^2 u_h &\geq Ch^{-2} \mathrm{tr}(J_T^t D_h^2 u_h J_T) \\
&\geq Ch^{-2} \big[\mathrm{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T)\big] \\
&\geq -Ch^{-2} \|\mathrm{tr}(J_T^t D_h^2 u_h J_T) - I_{T_\dagger} \mathrm{tr}(J_T^t D_h^2 u_h J_T)\|_{L_\infty(T)} \\
&\geq -Ch^{-1} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|_{H^2(T)} \\
&\geq -Ch
\end{aligned}$$

a priori bound

$$\Big(\sum_{T \in \mathcal{T}_h} |\mathrm{tr}(J_T^t D_h^2 u_h J_T)|_{H^2(T)}^2\Big)^{\frac{1}{2}} \leq Ch^2$$

Elementwise Convexity $u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

It follows from the *a priori* bound $\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \le Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \ge \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

Elementwise Convexity $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

It follows from the *a priori* bound $\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \le Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \ge \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

and we just saw that on each triangle $T \in \mathcal{T}_h$

$$\operatorname{tr} D_h^2 u_h \ge -Ch$$

Elementwise Convexity $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

It follows from the *a priori* bound $\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \leq Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \geq \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

and we just saw that on each triangle $T \in \mathcal{T}_h$

$$\operatorname{tr} D_h^2 u_h \geq -Ch$$

Two possibilities for the eigenvalues λ_1 and λ_2 of $D_h^2 u_h(x)$:

$$\lambda_1, \lambda_2 > 0 \quad \text{or} \quad \lambda_1, \lambda_2 < 0$$

Elementwise Convexity $\quad u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

It follows from the *a priori* bound $\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \le Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \ge \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

and we just saw that on each triangle $T \in \mathcal{T}_h$

$$\operatorname{tr} D_h^2 u_h \ge -Ch$$

Two possibilities for the eigenvalues λ_1 and λ_2 of $D_h^2 u_h(x)$:

$$\lambda_1, \lambda_2 > 0 \quad \text{or} \quad \lambda_1, \lambda_2 < 0$$

If $\lambda_1, \lambda_2 < 0$, then

$$0 > \lambda_1 + \lambda_2 \ge -Ch \quad \Rightarrow \quad |\lambda_1|, |\lambda_2| \le Ch$$

Elementwise Convexity $u_h = \underset{v \in L_h}{\text{argmin}}\, \mathcal{J}_h(v)$

It follows from the *a priori* bound $\| \det D_h^2 u_h - \psi \|_{L_\infty(\Omega)} \le Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \ge \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

and we just saw that on each triangle $T \in \mathcal{T}_h$

$$\operatorname{tr} D_h^2 u_h \ge -Ch$$

Two possibilities for the eigenvalues λ_1 and λ_2 of $D_h^2 u_h(x)$:

$$\lambda_1, \lambda_2 > 0 \quad \text{or} \quad \lambda_1, \lambda_2 < 0$$

If $\lambda_1, \lambda_2 < 0$, then

$$0 > \lambda_1 + \lambda_2 \ge -Ch \quad \Rightarrow \quad |\lambda_1|, |\lambda_2| \le Ch$$

and we have a contradiction.

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

It follows from the *a priori* bound $\|\det D_h^2 u_h - \psi\|_{L_\infty(\Omega)} \leq Ch$ that for $h \ll 1$

$$\det D_h^2 u_h(x) \geq \frac{1}{2} \min_{x \in \bar{\Omega}} \psi(x) > 0 \qquad \text{a.e. in } \Omega$$

and we just saw that on each triangle $T \in \mathcal{T}_h$

$$\operatorname{tr} D_h^2 u_h \geq -Ch$$

Two possibilities for the eigenvalues λ_1 and λ_2 of $D_h^2 u_h(x)$:

$$\lambda_1, \lambda_2 > 0 \quad \text{or} \quad \lambda_1, \lambda_2 < 0$$

Conclusion Both eigenvalues of $D_h^2 u_h(x)$ are positive and the function u_h is strictly convex on each $T \in \mathcal{T}_h$.

A Priori Error Analysis

A Priori Error Analysis

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 Ω except those on the edges of $\mathcal{T}_h$.

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

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

for a positive constants α.

$D^2 u$ and $\mathrm{Cof}\, D^2 u$ have identical eigenvalues

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

Therefore at any $x \in \Omega$ that is not on any edge of $\mathcal{T}_h$,

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

Cof is linear.

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha |\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

Therefore at any $x \in \Omega$ that is not on any edge of $\mathcal{T}_h$,

$$\begin{aligned} \xi^t A_h(x)\xi &= \int_0^1 t\xi^t [\mathrm{Cof}\, D_h^2 u(x)]\xi\,dt + \int_0^t (1-t)\xi^t [\mathrm{Cof}\, D_h^2 u_h(x)]\xi\,dt \\ &\geq \int_0^1 t\xi^t [\mathrm{Cof}\, D_h^2 u(x)]\xi\,dt \end{aligned}$$

u_h is elementwise convex.

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

Therefore at any $x \in \Omega$ that is not on any edge of $\mathcal{T}_h$,

$$\begin{aligned}
\xi^t A_h(x)\xi &= \int_0^1 t\xi^t[\mathrm{Cof}\; D_h^2 u(x)]\xi dt + \int_0^t (1-t)\xi^t[\mathrm{Cof}\; D_h^2 u_h(x)]\xi dt \\
&\geq \int_0^1 t\xi^t[\mathrm{Cof}\; D_h^2 u(x)]\xi dt \\
&\geq \alpha|\xi|^2 \int_0^1 t\, dt = \frac{\alpha}{2}|\xi|^2
\end{aligned}$$

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

In conjunction with the elementwise convexity of u_h, we have

$$\xi^t A_h(x)\xi \geq \frac{\alpha}{2}|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2, \text{ a.e.in } \Omega$$

A Priori Error Analysis

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

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

Since $u \in H^4(\Omega)$ is strictly convex on $\bar{\Omega}$, we have

$$\xi^t \mathrm{Cof}(D^2 u)\xi \geq \alpha|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2,\, x \in \bar{\Omega}$$

In conjunction with the elementwise convexity of u_h, we have

$$\xi^t A_h(x)\xi \geq \frac{\alpha}{2}|\xi|^2 \qquad \forall\, \xi \in \mathbb{R}^2, \text{ a.e.in } \Omega$$

Therefore $(*)$ is related to the error analysis of discontinuous Galerkin methods for elliptic equations in nondivergence form.

Smears-Süli (2013)
Neilan-Wu (2019)

A Stability Lemma

$$\|D_h^2(\zeta - v)\|_{L_2(\Omega)} \leq C\Big[\|A_h : D_h^2(\zeta - v)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]$$

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

A Stability Lemma

$$\|D_h^2(\zeta - v)\|_{L_2(\Omega)} \leq C\Big[\|A_h : D_h^2(\zeta - v)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|^2_{L_2(e)}\Big)^{\frac{1}{2}}\Big]$$

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

First Ingredient The relation

$$\|D^2\zeta\|_{L_2(\Omega)} = \|\Delta\zeta\|_{L_2(\Omega)} \qquad \forall\, \zeta \in H^2(\Omega) \cap H_0^1(\Omega)$$

that holds for any polygonal domains. Grisvard (1985)

A Stability Lemma

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

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

First Ingredient The relation

$$\|D^2\zeta\|_{L_2(\Omega)} = \|\Delta\zeta\|_{L_2(\Omega)} \qquad \forall\, \zeta \in H^2(\Omega) \cap H_0^1(\Omega)$$

that holds for any polygonal domains. Grisvard (1985)

Second Ingredient $E_h : V_h \cap H_0^1(\Omega) \longrightarrow H^2(\Omega) \cap H_0^1(\Omega)$

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

for all $v \in V_h \cap H_0^1(\Omega)$.

B.-Sung (2005)
Neilan-Wu (2019)

A Stability Lemma

$$\|D_h^2(\zeta - v)\|_{L_2(\Omega)} \leq C\Big[\|A_h : D_h^2(\zeta - v)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]$$

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

Third Ingredient Techniques for proving the existence and uniqueness of strong solutions for elliptic equations in nondivergence form that satisfy the Cordes condition.

Smears-Süli (2013)

A Stability Lemma

$$\|D_h^2(\zeta - v)\|_{L_2(\Omega)} \leq C\Big[\|A_h : D_h^2(\zeta - v)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]$$

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

We have $u - \phi \in H^2(\Omega) \cap H_0^1(\Omega)$ and $u_h - \Pi_h\phi \in V_h \cap H_0^1(\Omega)$.

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

and

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

A Stability Lemma

$$\|D_h^2(\zeta - v)\|_{L_2(\Omega)} \leq C\Big[\|A_h : D_h^2(\zeta - v)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]$$

for all $\zeta \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in V_h \cap H_0^1(\Omega)$, where the positive constant C is independent of h.

We have $u - \phi \in H^2(\Omega) \cap H_0^1(\Omega)$ and $u_h - \Pi_h\phi \in V_h \cap H_0^1(\Omega)$.

Therefore we can apply the stability lemma to

$$\zeta = u - \phi \quad \text{and} \quad v = u_h - \Pi_h\phi$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)} \\
&\quad\leq C\Big[\|A_h : D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)} \\
&\qquad\qquad+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial(u_h-\Pi_h\phi)/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\quad\leq C\Big[\|A_h:D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\qquad\qquad+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial(u_h-\Pi_h\phi)/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]\\
&\quad\leq C\Big[\|A_h:D_h^2(u-u_h)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\\
&\qquad\qquad+\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)} \\
&\quad\leq C\Big[\|A_h : D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)} \\
&\qquad\qquad + \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial(u_h-\Pi_h\phi)/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big] \\
&\quad\leq C\Big[\|A_h : D_h^2(u-u_h)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}} \\
&\qquad\qquad + \|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big] \\
&\quad= C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}} \\
&\qquad\qquad + \|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]
\end{aligned}
$$

Fudamental Theorem of Calculus

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

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\qquad\le C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\\
&\qquad\quad+\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)} \\
&\qquad \le C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12} \\
&\qquad\qquad +\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial \Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\Big]
\end{aligned}
$$

A Priori Bounds

$$
\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}=\|\psi-\det D_h^2u_h\|_{L_2(\Omega)}\le Ch^2
$$

$$
\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\le Ch^2
$$

Interpolation Error Estimates

$$
\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}\le Ch^2
$$

$$
\begin{aligned}
\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial \Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12} &= \Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial (\phi-\Pi_h\phi)/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12} \\
&\le Ch^2
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\qquad\le C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\\
&\qquad\qquad+\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\Big]\\
&\qquad\le Ch^2
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\qquad\le C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\\
&\qquad\quad+\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\Big]\\
&\qquad\le Ch^2
\end{aligned}
$$

Consequently

$$
\begin{aligned}
\|D_h^2(u-u_h)\|_{L_2(\Omega)}&\le\|D_h^2(u-\phi)-D_h^2(u_h-\Pi_h\phi)\|_{L_2(\Omega)}\\
&\qquad+\underbrace{\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}}_{O(h^2)}\\
&\le Ch^2
\end{aligned}
$$

$$
\begin{aligned}
&\|D_h^2\big((u-\phi)-(u_h-\Pi_h\phi)\big)\|_{L_2(\Omega)}\\
&\qquad\le C\Big[\|\det D^2u-\det D_h^2u_h\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\\
&\qquad\qquad+\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\Big]\\
&\qquad\le Ch^2
\end{aligned}
$$

Consequently

$$
\begin{aligned}
\|D_h^2(u-u_h)\|_{L_2(\Omega)}&\le\|D_h^2(u-\phi)-D_h^2(u_h-\Pi_h\phi)\|_{L_2(\Omega)}\\
&\qquad+\underbrace{\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}}_{O(h^2)}\\
&\le Ch^2
\end{aligned}
$$

From an *a priori* bound we also have

$$
\begin{aligned}
\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial(u-u_h)/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}&=\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial u_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac12}\\
&\le Ch^2
\end{aligned}
$$

In summary, we have established the following result.

Theorem

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

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

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

In summary, we have established the following result.

Theorem

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

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

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

Corollary

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

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

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

A Posteriori Error Analysis

A Posteriori Error Analysis

Recall

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

where

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

A Posteriori Error Analysis

Recall

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

where

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

Since $\mathcal{J}_h$ is non-convex, an optimization algorithm usually only produces an approximate stationary point $\tilde{u}_h$ of $\mathcal{J}_h$. The *a priori* error estimate for u_h does not guarantee that $\tilde{u}_h$ will converge to u.

A Posteriori Error Analysis

Suppose the approximation $\tilde{u}_h$ obtained by an optimization algorithm is elementwise convex.

(This can be verified numerically.)

A Posteriori Error Analysis

Suppose the approximation $\tilde{u}_h$ obtained by an optimization algorithm is elementwise convex.

(This can be verified numerically.)

By the fundamental theorem of calculus, we have

$$\det D^2 u - \det D_h^2 \tilde{u}_h = \tilde{A}_h : D_h^2(u - \tilde{u}_h)$$

where

$$\tilde{A}_h(x) = \int_0^1 \big[\mathrm{Cof}\, D_h^2(tu + (1-t)\tilde{u}_h)\big](x)\,dt$$

is uniformly elliptic.

By the same arguments in the *a priori* error analysis, we have

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

where C is independent of h and

$$\operatorname{Osc}(\phi)=\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}}\le Ch^2$$

By the same arguments in the *a priori* error analysis, we have

$$\|u-\tilde{u}_h\|_h \le C\Bigg[\|\det D_h^2\tilde{u}_h-\psi\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\tilde{u}_h/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}} + \mathrm{Osc}(\phi)\Bigg]^{\frac{1}{2}}$$

where C is independent of h and

$$\mathrm{Osc}(\phi)=\|D_h^2(\phi-\Pi_h\phi)\|_{L_2(\Omega)}+\Big(\sum_{e\in\mathcal{E}_h^i}|e|^{-1}\|[\![\partial\Pi_h\phi/\partial n]\!]\|_{L_2(e)}^2\Big)^{\frac{1}{2}} \le Ch^2$$

Therefore we can use the error estimator

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

to monitor the convergence of $\tilde{u}_h$ to u.

Numerical Results

Numerical Results

We will present the numerical results for two examples with known solutions. The first example satisfies all the assumptions of our theory. The second one has a less smooth solution.

The discrete problems are solved by an active set algorithm of Hager and Zhang.

Hager-Zhang (2006, 2016)

The elementwise convexity of the approximate solutions is verified numerically.

All the numerical experiments were carried out on a MacBook Pro laptop computer with a 2.8GHz Quad-Core Intel Core i7 processor and with 16GB 2133 MHz LPDDR3 memory. We use MATLAB (R2018b v.9.5.0) in our computations.

For each example we compute the relative errors

$$e^r_{2,h} = \frac{\|u - \tilde{u}_h\|_h}{\|u\|_{H^2(\Omega)}} \qquad e^r_{1,h} = \frac{\|u - \tilde{u}_h\|_{H^1(\Omega)}}{\|u\|_{H^1(\Omega)}}$$

$$e^r_{0,h} = \frac{\|u - \tilde{u}_h\|_{L_2(\Omega)}}{\|u\|_{L_2(\Omega)}} \qquad e^r_{\infty,h} = \frac{\max_{p\in\mathcal{V}_h} |u(p) - \tilde{u}_h(p)|}{\|u\|_{L_\infty(\Omega)}}$$

($\mathcal{V}_h$ is the set of the vertices of $\mathcal{T}_h$)

and the error estimator

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

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

We use uniform meshes in the computation.

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

Example 1

The domain is $\Omega = (0,1)^2$ and the exact solution is given by

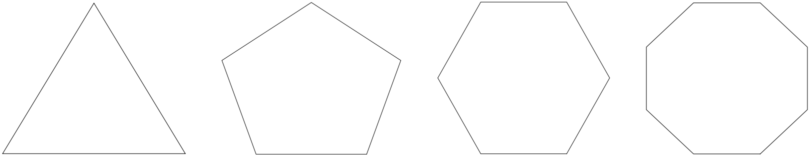
$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

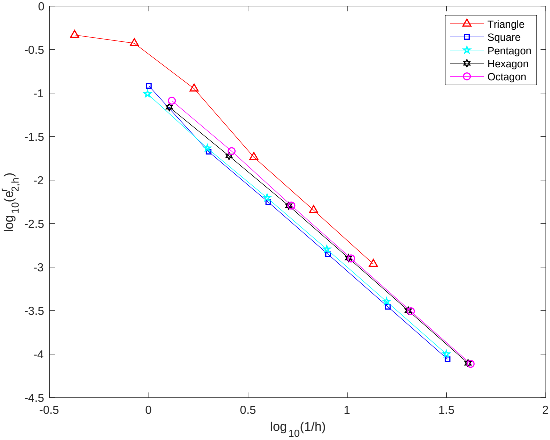
h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	1.21e-1	–	3.11e-2	–	6.71e-3	–	1.18e-2	–
2^{-1}	2.12e-2	2.51	2.91e-3	3.42	5.89e-4	3.51	7.43e-4	3.99
2^{-2}	5.58e-3	1.93	3.26e-4	3.16	3.05e-5	4.27	3.85e-5	4.27
2^{-3}	1.40e-3	1.99	3.95e-5	3.05	2.74e-6	3.47	3.71e-6	3.37
2^{-4}	3.50e-4	2.00	4.87e-6	3.02	2.93e-7	3.23	3.78e-7	3.30
2^{-5}	8.74e-5	2.00	6.53e-7	2.90	8.12e-8	1.85	9.23e-8	2.03

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	8.44e-1	1.96e-1	4.78e-2	1.16e-2	2.76e-3	6.53e-4
Order	–	2.11	2.03	2.04	2.08	2.08
CPU (s)	2.81e0	2.31e0	2.99e0	5.29e0	1.43e1	7.19e1

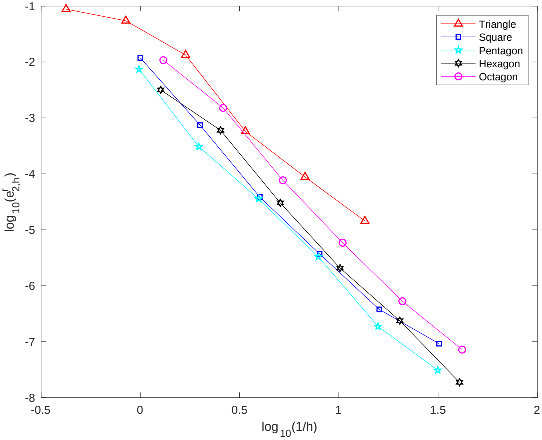
We have also tested the same problem on four other regular polygons.



The performance is similar.



Convergence histories of $e^r_{2,h}$ for the five regular polygons



Convergence histories of $e^r_{\infty,h}$ for the five regular polygons

Example 2

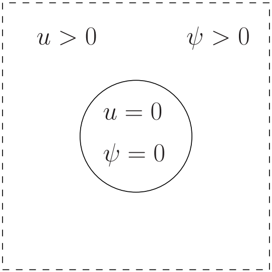
The domain $\Omega = (0,1)^2$, the exact solution is given by

$$u(x) = \frac{1}{2}(\max(|x - (0.5, 0.5)| - 0.2), 0)^2$$

and

$$\psi(x) = \max\left(1 - \frac{0.2}{|x - (0.5, 0.5)|}, 0\right)$$

Hamfeldt-Salvador (2018)
Nochetto-Ntogkas-Zhang (2019)



Example 2

The domain $\Omega = (0,1)^2$, the exact solution is given by

$$u(x) = \frac{1}{2}(\max(|x-(0.5,0.5)|-0.2),0)^2$$

and

$$\psi(x) = \max\left(1 - \frac{0.2}{|x-(0.5,0.5)|}, 0\right)$$

Hamfeldt-Salvador (2018)
Nochetto-Ntogkas-Zhang (2019)

We use uniform meshes in the computation. The approximate solutions are elementwise strictly convex outside the disc

$$|x-(0.5,0.5)| \leq 0.2$$

where $\psi = 0$.

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	3.47e-1	–	1.72e-1	–	1.46e-1	–	1.60e-1	–
2^{-1}	2.29e-1	0.60	8.51e-2	1.01	7.51e-2	0.96	7.53e-2	1.09
2^{-2}	1.44e-1	0.67	3.02e-2	1.49	2.28e-2	1.72	2.05e-2	1.88
2^{-3}	9.26e-2	0.64	8.97e-3	1.75	5.74e-3	1.99	4.84e-3	2.08
2^{-4}	7.96e-2	0.22	5.34e-3	0.75	3.05e-3	0.91	2.52e-3	0.94
2^{-5}	5.46e-2	0.54	2.07e-3	1.36	9.13e-4	1.74	7.48e-4	1.75

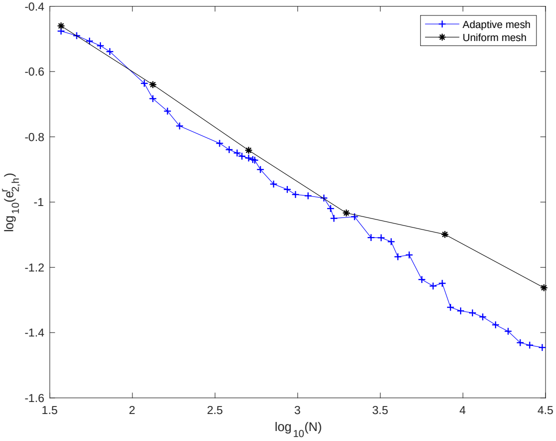
h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	1.37e-1	7.54e-2	2.96e-2	1.12e-2	4.30e-3	1.57e-3
Order	–	0.86	1.35	1.40	1.38	1.45
CPU (s)	1.12e0	1.97e0	3.51e0	7.78e0	2.94e1	2.47e2

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
2^0	3.47e-1	–	1.72e-1	–	1.46e-1	–	1.60e-1	–
2^{-1}	2.29e-1	0.60	8.51e-2	1.01	7.51e-2	0.96	7.53e-2	1.09
2^{-2}	1.44e-1	0.67	3.02e-2	1.49	2.28e-2	1.72	2.05e-2	1.88
2^{-3}	9.26e-2	0.64	8.97e-3	1.75	5.74e-3	1.99	4.84e-3	2.08
2^{-4}	7.96e-2	0.22	5.34e-3	0.75	3.05e-3	0.91	2.52e-3	0.94
2^{-5}	5.46e-2	0.54	2.07e-3	1.36	9.13e-4	1.74	7.48e-4	1.75

h	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}
$\eta_h(\tilde{u}_h)$	1.37e-1	7.54e-2	2.96e-2	1.12e-2	4.30e-3	1.57e-3
Order	–	0.86	1.35	1.40	1.38	1.45
CPU (s)	1.12e0	1.97e0	3.51e0	7.78e0	2.94e1	2.47e2

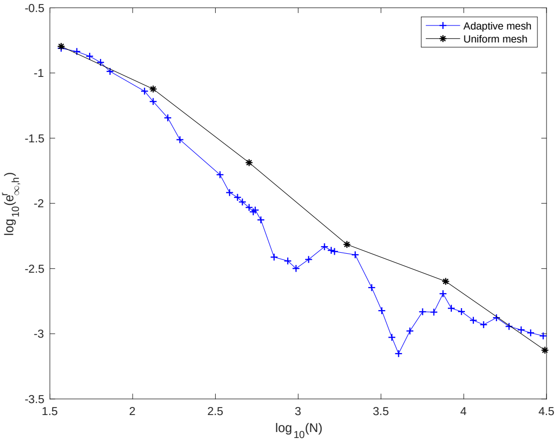
We also tested our method on adaptive meshes generated by the error estimator and a Dörfler marking strategy.

We also tested our method on adaptive meshes generated by the error estimator and a Dörfler marking strategy.

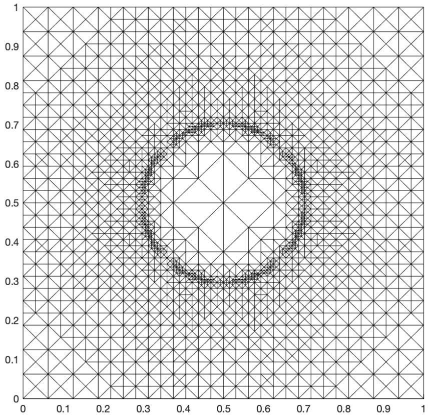


Convergence histories of $e^r_{2,h}$ on uniform and adaptive meshes

We also tested our method on adaptive meshes generated by the error estimator and a Dörfler marking strategy.



Convergence histories of $e^r_{\infty,h}$ on uniform and adaptive meshes



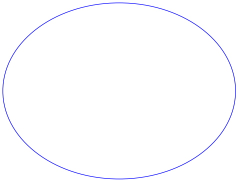
Adaptive mesh with 3385 dofs

Convex Smooth Domains

Convex Smooth Domains

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

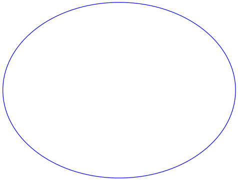
where Ω is a strictly convex smooth domain.



Convex Smooth Domains

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

where Ω is a strictly convex smooth domain.



We have a unique strictly convex solution $u \in H^4(\Omega)$ if ψ and ϕ are sufficiently smooth and $\psi > 0$ on $\bar{\Omega}$.

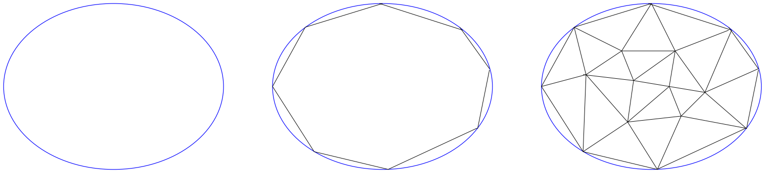
Caffarelli-Nirenberg-Spruck (1984)

An Isoparametric Mesh

An Isoparametric Mesh

We begin with a convex polygon $\tilde{\Omega}_h$ equipped with a quasi-uniform triangulation $\tilde{\mathcal{T}}_h$ such that

- The vertices of $\tilde{\Omega}_h$ belong to $\partial\Omega$.
- Each edge of $\tilde{\Omega}_h$ is also an edge of $\tilde{\mathcal{T}}_h$.
- Each triangle in $\tilde{\mathcal{T}}_h$ has at most two vertices on $\partial\Omega$.

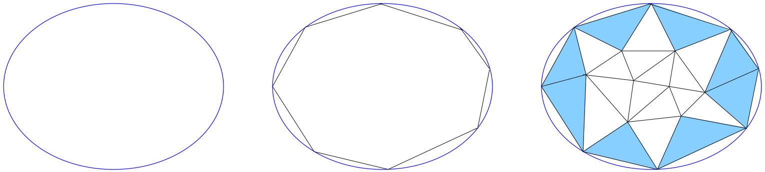


Ω $\qquad$ $\tilde{\Omega}_h$ $\qquad$ $\tilde{\mathcal{T}}_h$

An Isoparametric Mesh

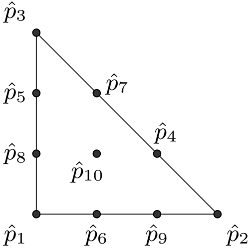
We begin with a convex polygon $\tilde{\Omega}_h$ equipped with a quasi-uniform triangulation $\tilde{\mathcal{T}}_h$ such that

- The vertices of $\tilde{\Omega}_h$ belong to $\partial\Omega$.
- Each edge of $\tilde{\Omega}_h$ is also an edge of $\tilde{\mathcal{T}}_h$.
- Each triangle in $\tilde{\mathcal{T}}_h$ has at most two vertices on $\partial\Omega$.

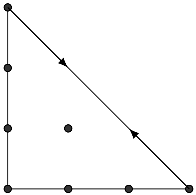


Ω $\quad$ $\tilde{\Omega}_h$ $\quad$ $\tilde{\mathcal{T}}_h$

We obtain the isoparametric mesh $\mathcal{T}_h$ by modifying the triangles in $\tilde{\mathcal{T}}_h$ that have two vertices on $\partial\Omega$.

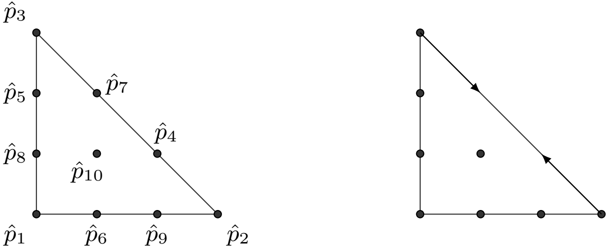


cubic Lagrange element



modified cubic Lagrange element

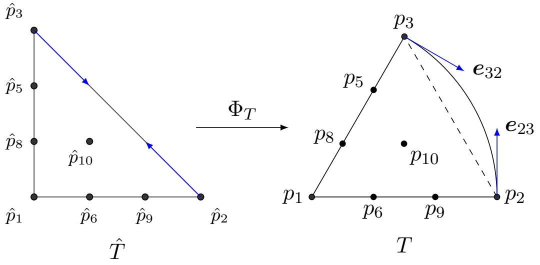
We modify the cubic Lagrange element by replacing the values at the nodes $\hat{p}_4$ and $\hat{p}_7$ on the reference triangle $\hat{T}$ by the directional derivatives at the node $\hat{p}_2$ in the direction of $\overrightarrow{\hat{p}_2\hat{p}_3}$ and the directional derivative at $\hat{p}_3$ in the direction of $\overrightarrow{\hat{p}_3\hat{p}_2}$.



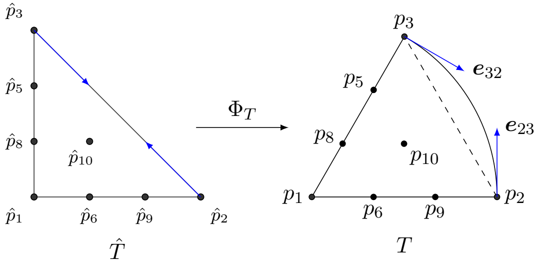
cubic Lagrange element   modified cubic Lagrange element

We modify the cubic Lagrange element by replacing the values at the nodes $\hat{p}_4$ and $\hat{p}_7$ on the reference triangle $\hat{T}$ by the directional derivatives at the node $\hat{p}_2$ in the direction of $\overrightarrow{\hat{p}_2\hat{p}_3}$ and the directional derivative at $\hat{p}_3$ in the direction of $\overrightarrow{\hat{p}_3\hat{p}_2}$.

We then modify a triangle $\tilde{T}$ in $\tilde{\mathcal{T}}_h$ with two vertices on $\partial\Omega$ into a triangle T with one curved edge by using a cubic polynomial map on the reference triangle $\hat{T}$ defined in terms of the degrees of freedom of the modified cubic Lagrange element.



Φ_T is the cubic polynomial map that maps the reference triangle $\hat{T}$ to the triangle T with one curved edge that is the modification of the triangle $\tilde{T}$ with vertices p_1, p_2 and p_3.



Definition of Φ_T Ciarlet-Raviart (1972)

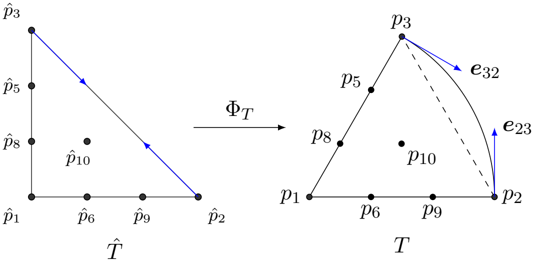
$$\Phi_T(\hat{p}_i) = p_i \qquad \text{for } i = 1, 2, 3, 5, 6, 8, 9$$

$$\Phi_T(\hat{p}_{10}) = p_{10} + \frac{1}{18}(|p_3 - p_2|\boldsymbol{e}_{23} + |p_2 - p_3|\boldsymbol{e}_{32})$$

$$D\Phi_T(\hat{p}_2)(\hat{p}_3 - \hat{p}_2) = |p_3 - p_2|\boldsymbol{e}_{23}$$

$$D\Phi_T(\hat{p}_3)(\hat{p}_2 - \hat{p}_3) = |p_2 - p_3|\boldsymbol{e}_{32}.$$

- $D\Phi_T$ is the Jacobian matrix of Φ_T.
- $\boldsymbol{e}_{23}$ is the unit tangent of $\partial\Omega$ at p_2 pointing towards p_3.
- $\boldsymbol{e}_{32}$ is the unit tangent of $\partial\Omega$ at p_3 pointing towards p_2.



Definition of Φ_T Ciarlet-Raviart (1972)

$$\Phi_T(\hat{p}_i) = p_i \qquad \text{for } i = 1, 2, 3, 5, 6, 8, 9$$

$$\Phi_T(\hat{p}_{10}) = p_{10} + \frac{1}{18}(|p_3 - p_2|\boldsymbol{e}_{23} + |p_2 - p_3|\boldsymbol{e}_{32})$$

$$D\Phi_T(\hat{p}_2)(\hat{p}_3 - \hat{p}_2) = |p_3 - p_2|\boldsymbol{e}_{23}$$

$$D\Phi_T(\hat{p}_3)(\hat{p}_2 - \hat{p}_3) = |p_2 - p_3|\boldsymbol{e}_{32}.$$

Key Properties

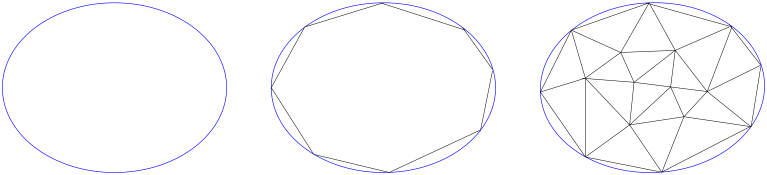
The curved edge of T is tangential to $\partial\Omega$ at the two vertices of T that belong to $\partial\Omega$ and it is a good approximation of $\partial\Omega$ between these two points.

The Isoparametric Mesh $\mathcal{T}_h$ T belongs to $\mathcal{T}_h$ if

- $T = \tilde{T}$ for a triangle $\tilde{T} \in \tilde{\mathcal{T}}_h$ that has at most one vertex on $\partial\Omega$.

or

- T with one curved edge is the modification of a triangle $\tilde{T} \in \tilde{\mathcal{T}}_h$ that has two vertices on $\partial\Omega$.



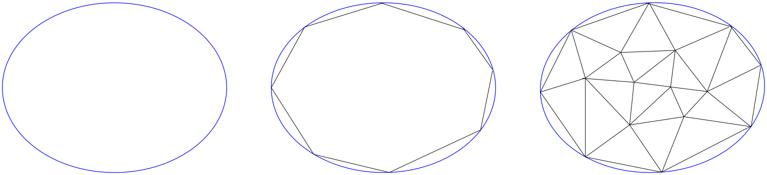
Ω $\tilde{\Omega}_h$ $\tilde{\mathcal{T}}_h$

The Isoparametric Mesh $\mathcal{T}_h$ 　 T belongs to $\mathcal{T}_h$ if

- $T = \tilde{T}$ for a triangle $\tilde{T} \in \tilde{\mathcal{T}}_h$ that has at most one vertex on $\partial\Omega$.

or

- T with one curved edge is the modification of a triangle $\tilde{T} \in \tilde{\mathcal{T}}_h$ that has two vertices on $\partial\Omega$.



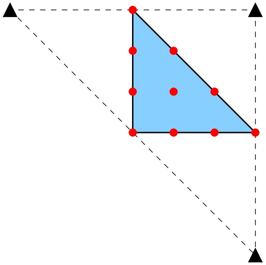
Ω 　 $\tilde{\Omega}_h$ 　 $\tilde{\mathcal{T}}_h$

The Domain Ω_h

Ω_h is the union of all the elements $T \in \mathcal{T}_h$. It is a strictly convex $C^{1,1}$ domain for $h \ll 1$. The discrete problem is posed on Ω_h.

Reference Element for a Straightedged Triangle in $\mathcal{T}_h$

We use the enhanced cubic Lagrange element from before.



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

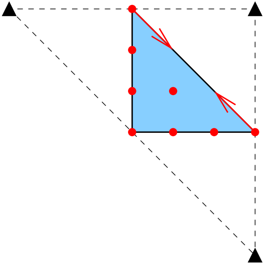
$$\varphi_{\hat{T}} = \hat{x}_1\hat{x}_2(1-\hat{x}_1-\hat{x}_2)$$

Degrees of Freedom

The dofs for v from the standard cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

Reference Element for a Triangle with a Curved Edge in $\mathcal{T}_h$

We use an enhanced modified cubic Lagrange element.



$\hat{T}$ is the reference simplex.

$\hat{T}_\dagger$ is the triangle with vertices $(1,1)$, $(-1,1)$ and $(1,-1)$.

Space of Shape Functions

$$P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$$

$$\varphi_{\hat{T}} = \hat{x}_1\hat{x}_2(1-\hat{x}_1-\hat{x}_2)$$

Degrees of Freedom

The dofs for v from the modified cubic Lagrange finite element together with Δv at the vertices of $\hat{T}_\dagger$.

An Isoparametric Finite Element Space V_h Associated with $\mathcal{T}_h$

An Isoparametric Finite Element Space V_h Associated with $\mathcal{T}_h$

For $T \in \mathcal{T}_h$,

$$\Phi_T : \hat{T} \longrightarrow T$$

is an affine map if T is straightedged and a cubic polynomial map if T has a curved edge.

An Isoparametric Finite Element Space V_h Associated with $\mathcal{T}_h$

For $T \in \mathcal{T}_h$,

$$\Phi_T : \hat{T} \longrightarrow T$$

is an affine map if T is straightedged and a cubic polynomial map if T has a curved edge.

A function v belongs to the finite element space V_h if

- v belongs to $C(\Omega_h)$.
- $v \circ \Phi_T$ belongs to $P_3(\hat{T}) \oplus \varphi_{\hat{T}}^2 P_1(\hat{T})$.
- The dofs of $v \circ \Phi_T$ are the dofs of the enhanced cubic Lagrange element if Φ_T is an affine map.
- The dofs of $v \circ \Phi_T$ are the dofs of the enhanced modified cubic Lagrange element if Φ_T is a cubic polynomial map.

An Isoparametric Finite Element Space V_h Associated with $\mathcal{T}_h$

For $T \in \mathcal{T}_h$,

$$\Phi_T : \hat{T} \longrightarrow T$$

is an affine map if T is straightedged and a cubic polynomial map if T has a curved edge.

Remark

Recall we can treat the solution u of the boundary value problem of the Monge-Ampère equation as a function in $H^4(\mathbb{R}^2)$ such that u is strictly convex in a neighborhood of $\bar{\Omega}$.

For $h \ll 1$,

$$\Delta(u \circ \Phi_T) \quad \text{is positive on } \hat{T}_\dagger$$

- Obvious if Φ_T is an affine map.
- Based on the approximation properties of Φ_T if it is a cubic polynomial map.

Interpolation Operator Π_h

The operator $\Pi_h : H^4(\mathbb{R}^2) \longrightarrow V_h$ is defined by the condition that

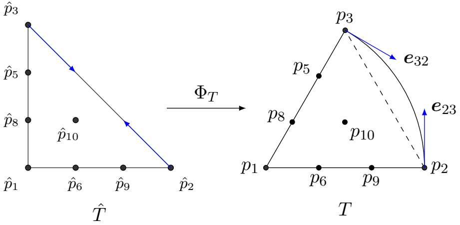
$$\zeta \circ \Phi_T \text{ and } (\Pi_h \zeta) \circ \Phi_T \text{ have identical dofs for all } T \in \mathcal{T}_h$$

Interpolation Operator Π_h

The operator $\Pi_h : H^4(\mathbb{R}^2) \longrightarrow V_h$ is defined by the condition that

$$\zeta \circ \Phi_T \text{ and } (\Pi_h \zeta) \circ \Phi_T \text{ have identical dofs for all } T \in \mathcal{T}_h$$

By construction, the restriction of $(\Pi_h \zeta) \circ \Phi_T$ on the edge of $\hat{T}$ corresponding to the curved edge of T is the one-dimensional cubic Hermite interpolant of $\zeta \circ \Phi_T$.

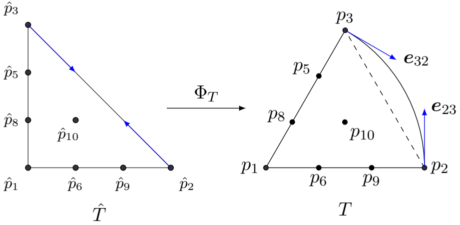


Interpolation Operator Π_h

The operator $\Pi_h : H^4(\mathbb{R}^2) \longrightarrow V_h$ is defined by the condition that

$$\zeta \circ \Phi_T \text{ and } (\Pi_h \zeta) \circ \Phi_T \text{ have identical dofs for all } T \in \mathcal{T}_h$$

By construction, the restriction of $(\Pi_h \zeta) \circ \Phi_T$ on the edge of $\hat{T}$ corresponding to the curved edge of T is the one-dimensional cubic Hermite interpolant of $\zeta \circ \Phi_T$.



Since the curved edge of T is tangential to $\partial\Omega$ at its two vertices on $\partial\Omega$, the restriction of $\Pi_h \zeta$ to $\partial\Omega_h$ is determined by the restriction of ζ to $\partial\Omega$.

The Constraint Set L_h

$$L_h = \{v \in V_h : v = \Pi_h \phi \text{ on } \partial\Omega_h \text{ and } \Delta(v \circ \Phi_T) \geq 0$$
$$\text{at the vertices of } \hat{T}_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

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

The Constraint Set L_h

$$L_h = \{v \in V_h : v = \Pi_h \phi \text{ on } \partial\Omega_h \text{ and } \Delta(v \circ \Phi_T) \geq 0$$
$$\text{at the vertices of } \hat{T}_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

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

Since $\Pi_h \phi$ is determined by $\phi = u$ on $\partial\Omega$, we have

$$\Pi_h \phi = \Pi_h u \quad \text{on} \quad \partial\Omega_h$$

The Constraint Set L_h

$$L_h = \{v \in V_h : v = \Pi_h \phi \text{ on } \partial\Omega_h \text{ and } \Delta(v \circ \Phi_T) \geq 0$$
$$\text{at the vertices of } \hat{T}_\dagger \text{ for every } T \in \mathcal{T}_h\}$$

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

Since $\Pi_h \phi$ is determined by $\phi = u$ on $\partial\Omega$, we have

$$\Pi_h \phi = \Pi_h u \quad \text{on} \quad \partial\Omega_h$$

Key Observation $\Pi_h u$ belongs to L_h.

- $\Pi_h u = \Pi_h \phi$ on $\partial\Omega_h$
- $\Delta\big((\Pi_h \phi) \circ \Phi_T\big) = \Delta(u \circ \Phi_T)$

The Discrete Problem

Find

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

where

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

The Discrete Problem

Find

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

where

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

Remark For an affine map Φ_T, we have

$$\frac{h^{-2}}{2}|\Delta(v\circ\Phi_T)|_{H^2(\hat{T})} \approx \frac{1}{2}\mathrm{tr}(J_T^t D_h^2 v J_T)|_{H^2(T)}^2$$

which is the term we used in the case where Ω is a convex polygon.

Error Analysis

Again we have *a priori* bounds because $\Pi_h u \in L_h$, and we can analyze the error by using the fundamental theorem of calculus and techniques for discontinuous Galerkin methods for elliptic problems in nondivergence form.

But we need to develop several new estimates for functions in the isoparametric finite element space V_h.

Error Analysis

Again we have *a priori* bounds because $\Pi_h u \in L_h$, and we can analyze the error by using the fundamental theorem of calculus and techniques for discontinuous Galerkin methods for elliptic problems in nondivergence form.

But we need to develop several new estimates for functions in the isoparametric finite element space V_h.

An Inverse Estimate

$$\|D^2 v\|_{L_\infty(T)} \le C\big(h^{-1}\|D^2 v\|_{L_2(T)} + h^4\|\nabla v\|_{L_\infty(T)}\big)$$

for all $T \in \mathcal{T}_h$.

Error Analysis

Again we have *a priori* bounds because $\Pi_h u \in L_h$, and we can analyze the error by using the fundamental theorem of calculus and techniques for discontinuous Galerkin methods for elliptic problems in nondivergence form.

But we need to develop several new estimates for functions in the isoparametric finite element space V_h.

An Inverse Estimate

$$\|D^2 v\|_{L_\infty(T)} \leq C\big(h^{-1}\|D^2 v\|_{L_2(T)} + h^4\|\nabla v\|_{L_\infty(T)}\big)$$

for all $T \in \mathcal{T}_h$.

A Discrete Sobolev Inequality

$$\|\nabla v\|^2_{L^\infty(\Omega_h)} \leq C\bigg\{(1+|\ln h|)^2\Big(\|D_h^2 v\|^2_{L^2(\Omega_h)} + \sum_{e\in\mathcal{E}_h^i} |e|^{-1}\|[\![\partial v/\partial n]\!]\|^2_{L^2(e)}\Big) + (1+|\ln h|)\max_{p\in\partial\Omega_h}\Big[\frac{\partial v}{\partial s}(p)\Big]^2\bigg\}$$

for all $v \in V_h$.

Error Analysis

Again we have *a priori* bounds because $\Pi_h u \in L_h$, and we can analyze the error by using the fundamental theorem of calculus and techniques for discontinuous Galerkin methods for elliptic problems in nondivergence form.

But we need to develop several new estimates for functions in the isoparametric finite element space V_h.

A Discrete Miranda-Talenti Estimate

$$\|D_h^2 v\|_{L^2(\Omega_h)} \le \|\Delta_h v\|_{L^2(\Omega_h)} + C_\dagger \Big[\Big(\sum_{e\in\mathcal{E}_h^i} |e|^{-1} \|[\![\partial v/\partial n]\!]\|_{L^2(e)}^2\Big)^{\frac{1}{2}} + h^3 \|\nabla v\|_{L^\infty(\Omega_h)}\Big]$$

for all $v \in V_h \cap H_0^1(\Omega_h)$. ($\Omega_h$ is convex for $h \ll 1$.)

Theorem

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

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

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

Theorem

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

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

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

Corollary

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

Theorem

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

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

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

Corollary

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

We can use the error estimator

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

to monitor the convergence of an approximate solution $\tilde{u}_h$ to u.

Numerical Results

We present the results of two examples for the domain Ω bounded by the ellipse

$$x_1^2 + 4x_2^2 = 1$$

We use quasi-uniform meshes in the computation.

Example 1

The exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

We compute the relative errors

$$e^r_{2,h} = \frac{\|u - \tilde{u}_h\|_h}{\|u\|_{H^2(\Omega_h)}} \qquad e^r_{1,h} = \frac{\|u - \tilde{u}_h\|_{H^1(\Omega_h)}}{\|u\|_{H^1(\Omega_h)}}$$

$$e^r_{0,h} = \frac{\|u - \tilde{u}_h\|_{L_2(\Omega_h)}}{\|u\|_{L_2(\Omega_h)}} \qquad e^r_{\infty,h} = \frac{\max_{p \in \mathcal{V}_h} |u(p) - \tilde{u}_h(p)|}{\|u\|_{L_\infty(\Omega_h)}}$$

($\mathcal{V}_h$ is the set of the vertices of $\mathcal{T}_h$)

and the error estimator

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

Example 1

The exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
1.12e0	4.8e-1	–	2.4e-1	–	5.4e-2	–	1.6e-2	–
7.22e-1	1.6e-1	2.5	5.1e-2	3.6	5.7e-3	5.1	8.0e-3	1.3
3.92e-1	3.9e-2	2.3	9.3e-3	2.8	1.1e-3	2.7	1.8e-3	2.7
2.02e-1	8.7e-3	2.3	1.1e-3	3.2	9.6e-5	3.7	1.4e-4	3.8
1.02e-1	1.8e-3	2.3	1.2e-4	3.3	7.2e-6	3.8	1.2e-5	3.6
5.13e-2	3.5e-4	2.4	1.2e-5	3.4	5.2e-7	3.8	1.0e-6	3.6

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	2.29e0	6.08e-1	1.32e-1	2.85e-2	6.11e-3	9.31e-4
Order	–	3.04	2.50	2.32	2.25	2.73
CPU time (s)	5.31e1	3.50e0	6.57e0	3.19e1	6.31e1	5.30e2

Example 1

The exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
1.12e0	4.8e-1	–	2.4e-1	–	5.4e-2	–	1.6e-2	–
7.22e-1	1.6e-1	2.5	5.1e-2	3.6	5.7e-3	5.1	8.0e-3	1.3
3.92e-1	3.9e-2	2.3	9.3e-3	2.8	1.1e-3	2.7	1.8e-3	2.7
2.02e-1	8.7e-3	2.3	1.1e-3	3.2	9.6e-5	3.7	1.4e-4	3.8
1.02e-1	1.8e-3	2.3	1.2e-4	3.3	7.2e-6	3.8	1.2e-5	3.6
5.13e-2	3.5e-4	2.4	1.2e-5	3.4	5.2e-7	3.8	1.0e-6	3.6

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	2.29e0	6.08e-1	1.32e-1	2.85e-2	6.11e-3	9.31e-4
Order	–	3.04	2.50	2.32	2.25	2.73
CPU time (s)	5.31e1	3.50e0	6.57e0	3.19e1	6.31e1	5.30e2

Example 1

The exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
1.12e0	4.8e-1	–	2.4e-1	–	5.4e-2	–	1.6e-2	–
7.22e-1	1.6e-1	2.5	5.1e-2	3.6	5.7e-3	5.1	8.0e-3	1.3
3.92e-1	3.9e-2	2.3	9.3e-3	2.8	1.1e-3	2.7	1.8e-3	2.7
2.02e-1	8.7e-3	2.3	1.1e-3	3.2	9.6e-5	3.7	1.4e-4	3.8
1.02e-1	1.8e-3	2.3	1.2e-4	3.3	7.2e-6	3.8	1.2e-5	3.6
5.13e-2	3.5e-4	2.4	1.2e-5	3.4	5.2e-7	3.8	1.0e-6	3.6

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	2.29e0	6.08e-1	1.32e-1	2.85e-2	6.11e-3	9.31e-4
Order	–	3.04	2.50	2.32	2.25	2.73
CPU time (s)	5.31e1	3.50e0	6.57e0	3.19e1	6.31e1	5.30e2

Example 1

The exact solution is given by

$$u(x) = e^{|x|^2/2}$$

Dean-Glowinski (2003)

h	$e^r_{2,h}$	order	$e^r_{1,h}$	order	$e^r_{0,h}$	order	$e^r_{\infty,h}$	order
1.12e0	4.8e-1	–	2.4e-1	–	5.4e-2	–	1.6e-2	–
7.22e-1	1.6e-1	2.5	5.1e-2	3.6	5.7e-3	5.1	8.0e-3	1.3
3.92e-1	3.9e-2	2.3	9.3e-3	2.8	1.1e-3	2.7	1.8e-3	2.7
2.02e-1	8.7e-3	2.3	1.1e-3	3.2	9.6e-5	3.7	1.4e-4	3.8
1.02e-1	1.8e-3	2.3	1.2e-4	3.3	7.2e-6	3.8	1.2e-5	3.6
5.13e-2	3.5e-4	2.4	1.2e-5	3.4	5.2e-7	3.8	1.0e-6	3.6

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	2.29e0	6.08e-1	1.32e-1	2.85e-2	6.11e-3	9.31e-4
Order	–	3.04	2.50	2.32	2.25	2.73
CPU time (s)	5.31e1	3.50e0	6.57e0	3.19e1	6.31e1	5.30e2

Example 2

$\psi = 1$ and $\phi = e^{|x|^2/2}$

The exact solution is unknown.

We use $\eta_h(\tilde{u}_h)$ to monitor the convergence of the approximate solution $\tilde{u}_h$ and also check the value of the cost $\mathcal{J}_h(\tilde{u}_h)$.

Example 2

$\psi = 1$ and $\phi = e^{|x|^2/2}$

The exact solution is unknown.

We use $\eta_h(\tilde{u}_h)$ to monitor the convergence of the approximate solution $\tilde{u}_h$ and also check the value of the cost $\mathcal{J}_h(\tilde{u}_h)$.

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	1.49e0	4.30e-1	6.31e-2	1.04e-2	2.04e-3	2.91e-4
Order	–	2.84	3.15	2.72	2.39	2.83
$\mathcal{J}_h(\tilde{u}_h)$	1.63e0	4.59e-1	4.37e-2	3.08e-3	2.01e-4	1.35e-5
Order	–	2.89	3.86	4.00	4.00	3.92
CPU time (s)	2.05e1	2.24e0	5.93e0	1.51e1	7.70e1	6.25e2

Example 2

$\psi = 1$ and $\phi = e^{|x|^2/2}$

The exact solution is unknown.

We use $\eta_h(\tilde{u}_h)$ to monitor the convergence of the approximate solution $\tilde{u}_h$ and also check the value of the cost $\mathcal{J}_h(\tilde{u}_h)$.

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	1.49e0	4.30e-1	6.31e-2	1.04e-2	2.04e-3	2.91e-4
Order	–	2.84	3.15	2.72	2.39	2.83
$\mathcal{J}_h(\tilde{u}_h)$	1.63e0	4.59e-1	4.37e-2	3.08e-3	2.01e-4	1.35e-5
Order	–	2.89	3.86	4.00	4.00	3.92
CPU time (s)	2.05e1	2.24e0	5.93e0	1.51e1	7.70e1	6.25e2

Example 2

$\psi = 1$ and $\phi = e^{|x|^2/2}$

The exact solution is unknown.

We use $\eta_h(\tilde{u}_h)$ to monitor the convergence of the approximate solution $\tilde{u}_h$ and also check the value of the cost $\mathcal{J}_h(\tilde{u}_h)$.

h	1.12e0	7.22e-1	3.92e-1	2.02e-1	1.02e-1	5.13e-2
$\eta_h(\tilde{u}_h)$	1.49e0	4.30e-1	6.31e-2	1.04e-2	2.04e-3	2.91e-4
Order	–	2.84	3.15	2.72	2.39	2.83
$\mathcal{J}_h(\tilde{u}_h)$	1.63e0	4.59e-1	4.37e-2	3.08e-3	2.01e-4	1.35e-5
Order	–	2.89	3.86	4.00	4.00	3.92
CPU time (s)	2.05e1	2.24e0	5.93e0	1.51e1	7.70e1	6.25e2

Concluding Remarks

Concluding Remarks

Using nonlinear least-squares methods to solve finite dimensional nonlinear equations is a classical approach. We have adopted this idea to tackle the Monge-Ampère equation.

Concluding Remarks

Using nonlinear least-squares methods to solve finite dimensional nonlinear equations is a classical approach. We have adopted this idea to tackle the Monge-Ampère equation.

Ingredients

- Design novel finite elements that can enforce the elementwise convexity of the discrete solutions as box constraints.
- Formulate discrete constrained nonlinear least-squares problems with regularization and penalty terms that lead to *a priori* bounds.
- Solve the discrete problems by globally convergent optimization algorithms that will terminate near stationary points of the cost function with an arbitrary initial guess.
- Derive error estimates by using the fundamental theorem of calculus to reduce the error analysis to a linear elliptic problem.
- Construct an *a posteriori* error estimator that can monitor the convergence of the computed solutions.