Recent Results About the Discrete Maximum Principle for Higher-Order Finite Elements

Tomáš Vejchodský (vejchod@math.cas.cz)

Institute of Mathematics, Academy of Sciences
Žitná 25, 115 67 Prague 1
Czech Republic

Outline

Goal

- $-u'' + \kappa^2 u = f$ in $\Omega = (a_\Omega, b_\Omega) \quad u(a_\Omega) = u(b_\Omega) = 0$
- $f \geq 0 \quad \Rightarrow \quad u_{hp} \geq 0$

Outline

- Maximum principle
- *hp*-FEM
- Definition of the discrete maximum principle (DMP)
- Proof in 1D for $p = 1$
- Discrete Green's function (DGF)
- Scheme of the proof in 1D for arbitrary p
- Final result

(Continuous) maximum principle

$$-\Delta u + \kappa^2 u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega$$

$$\text{MaxP}: \quad f \le 0 \quad \Rightarrow \quad \max_{\overline{\Omega}} u \le \max\{0, \max_{\partial\Omega} u\} = 0$$

$$\Updownarrow$$

$$\text{MinP}: \quad f \ge 0 \quad \Rightarrow \quad \min_{\overline{\Omega}} u \ge \min\{0, \min_{\partial\Omega} u\} = 0$$

$$\Updownarrow$$

$$\text{ComP}: \quad f \ge 0 \quad \Rightarrow \quad u \ge 0$$

$$\Updownarrow$$

$$G(x,y) \ge 0 \text{ in } \Omega^2$$

$$u(y) = \int_\Omega G(x,y) f(x) \,\mathrm{d}x$$

$$-\Delta G_y + \kappa^2 G_y = \delta_y \quad \text{in } \Omega$$

$$G_y = 0 \quad \text{on } \partial\Omega$$

$$G(x,y) = G_y(x)$$

Proof of MaxP

Theorem

If $u \in H^1(\Omega)$*:* $\int_\Omega (\mathcal{A}\nabla u \cdot \nabla v + \kappa^2 uv)\,\mathrm{d}x = \int_\Omega fv\,\mathrm{d}x \quad \forall v \in H^1_0(\Omega)$ *and* $f \le 0$ *then* $\max_{\overline{\Omega}} u \le \max\{0, \max_{\partial\Omega} u\}$*.*

Proof.

- Set $M = \max\{0, \max_{\partial\Omega} u\}$.
- Define $v = \max\{u - M, 0\}$.
- $v \ge 0$ in Ω
- $v|_{\partial\Omega} = 0$
- $u(x) = v(x) + M$ for all $x \in \Omega$ such that $v(x) \ne 0$.
- $v \in H^1_0(\Omega)$ (We can use the Green's theorem.)
- $$0 \ge \int_\Omega fv\,\mathrm{d}x = \int_\Omega \left(\mathcal{A}\nabla u \cdot \nabla v + \kappa^2 uv\right)\,\mathrm{d}x = \int_\Omega \left(\mathcal{A}\nabla v \cdot \nabla v + \kappa^2 (v+M)v\right)\,\mathrm{d}x \ge 0$$
- Thus $v = 0$ and $u \le M$ in $\overline{\Omega}$.

Proof of (ComP $\Rightarrow$ MinP)

Theorem

ComP $\Rightarrow$ *MinP*

Proof.

- Set $m = \min_{\partial\Omega} u$.
- Let $m \le 0$.
  - $w = u - m$
  - $-\operatorname{div}(\mathcal{A}\nabla w) + \kappa^2 w = f - \kappa^2 m \ge 0$ in Ω
  - $w \ge 0$ on $\partial\Omega$.
  - Hence, by ComP $w \ge 0$ in $\overline{\Omega}$, i.e. $u \ge m$ and $\min_{\overline{\Omega}} u \ge m = \min\{0, m\} = \min\{0, \min_{\partial\Omega} u\}$.
- If $m > 0$ then
  - $u \ge 0$ on $\partial\Omega$
  - ComP for u implies $u \ge 0$ in $\overline{\Omega}$.
  - Thus, $\min_{\overline{\Omega}} u \ge 0 = \min\{0, m\}$.

$\square$

Discretization

- Weak

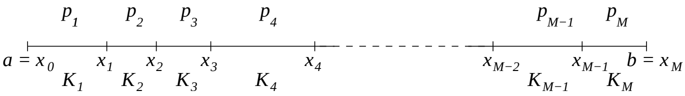
  $$u \in V = H_0^1(\Omega): \quad \underbrace{a(u,v)}_{\int_\Omega \nabla u \cdot \nabla v + \kappa^2 uv \,\mathrm{d}x} = \int_\Omega fv \,\mathrm{d}x \quad \forall v \in V$$

- *hp*-FEM

  $$u_{hp} \in V_{hp} \subset V: \quad a(u_{hp}, v_{hp}) = \int_\Omega f v_{hp} \,\mathrm{d}x \quad \forall v_{hp} \in V_{hp}$$

- $V_{hp} = \{v_{hp} \in V : v_{hp}|_{K_i} \in P^{p_i}(K_i),\ K_i \in \mathcal{T}_{hp}\}$

  Triangulation $\mathcal{T}_{hp}$ of Ω



Two definitions of the DMP

Notation

- For simplicity: $-\Delta u = f$
- $L^{2+}(\Omega) = \left\{ f \in L^2(\Omega) : f \geq 0 \text{ a.e. in } \Omega \right\}$
- $\mathcal{F} = \{\mathcal{T}_{hp} : \mathcal{T}_{hp} \text{ is a triangulation of } \Omega\}$
- $u_{hp}(x) = u_{\mathcal{T}_{hp},f}(x)$

Definitions

- DMP – not valid
  $\forall \mathcal{T}_{hp} \in \mathcal{F} \quad \forall f \in L^{2+}(\Omega) \quad u_{\mathcal{T}_{hp},f} \geq 0 \text{ in } \Omega$
- DMP (a) $\quad \exists \mathcal{F}_{DMP} \subset \mathcal{F}$:
  $\forall \mathcal{T}_{hp} \in \mathcal{F}_{DMP} \quad \forall f \in L^{2+}(\Omega) \quad u_{\mathcal{T}_{hp},f} \geq 0 \text{ in } \Omega$
- DMP (b)
  $\forall f \in L^{2+}(\Omega) \quad \exists \mathcal{T}_{hp} \in \mathcal{F} \quad u_{\mathcal{T}_{hp},f} \geq 0 \text{ in } \Omega$

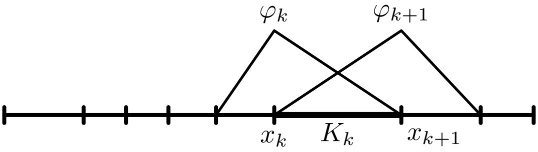
DMP in 1D for $p = 1$

$$-u'' + \kappa^2 u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega$$

Lemma

DMP $\Leftrightarrow$ $\kappa^2 h_K^2 \leq 6$ *for all* $K \in \mathcal{T}_{hp}$

Proof.



- $u_h(x) = \sum_{j=1}^{N} z_j \varphi_j(x)$ $\qquad A_{ij} = a(\varphi_j, \varphi_i) \quad b_i = \int_\Omega f \varphi_i$
- $Az = b \quad \Leftrightarrow \quad z = A^{-1} b$
- Assume $A^{-1} \geq 0$
- $f(x) \geq 0 \quad \Rightarrow \quad b \geq 0 \quad \Rightarrow \quad z \geq 0 \quad \Rightarrow \quad u_h(x) \geq 0$

$\square$

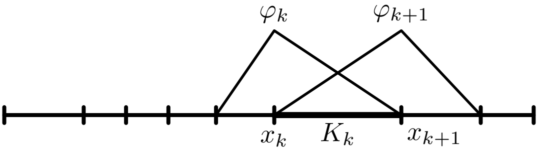
DMP in 1D for $p = 1$

$$-u'' + \kappa^2 u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega$$

Lemma

DMP $\quad \Leftrightarrow \quad \kappa^2 h_K^2 \leq 6$ *for all* $K \in \mathcal{T}_{hp}$

Proof.



- If A s.p.d. tridiagonal then
  $A^{-1} \geq 0 \quad \Leftrightarrow \quad A_{ij} \leq 0$ for $i \neq j \quad \Leftrightarrow \quad \kappa^2 h_K^2 \leq 6$
- $a(\varphi_k, \varphi_{k+1}) = \int_{x_k}^{x_{k+1}} (\varphi'_k \varphi'_{k+1} + \kappa^2 \varphi_k \varphi_{k+1}) \, \mathrm{d}x = -\frac{1}{h_K} + \kappa^2 \frac{h_K}{6}$
- $A_{k,k+1} \leq 0 \quad \Leftrightarrow \quad \kappa^2 h_K^2 \leq 6$

$\square$

DMP in 1D for $p = 1$

$$-u'' + \kappa^2 u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega$$

Lemma

DMP $\Leftrightarrow$ $\kappa^2 h_K^2 \leq 6$ *for all* $K \in \mathcal{T}_{hp}$

Counterintuitive:

- $\Omega = (0,1)$, $\kappa = 10$, $N_{elem} = 10$ $\Rightarrow$ DMP O.K.
- $\Omega = (0,10)$, $\kappa = 10$, $N_{elem} = 10$ $\Rightarrow$ NO DMP

Discrete Maximum Principle (DMP)

Definition (DMP)

Characterize such triangulations $\mathcal{T}_{hp}$ that
any $f \geq 0 \quad \Rightarrow \quad u_{hp} \geq 0$ in Ω.

Theorem

DMP $\quad \Leftrightarrow \quad G_{hp} \geq 0 \textit{ in } \Omega^2$

Proof.

$$G_{hp,y} \in V_{hp} \ : \ a(v_{hp}, G_{hp,y}) = \underbrace{\delta_y(v_{hp})}_{v_{hp}(y)} \quad \forall v_{hp} \in V_{hp}, \ \ y \in \Omega$$

$$u_{hp}(y) = a(u_{hp}, G_{hp,y}) = \int_\Omega G_{hp}(x,y) f(x) \, \mathrm{d}x$$

$G_{hp}(x,y) = G_{hp,y}(x)$ □

Discrete Maximum Principle (DMP)

Definition (DMP)

Characterize such triangulations $\mathcal{T}_{hp}$ that
any $f \geq 0 \quad \Rightarrow \quad u_{hp} \geq 0$ in Ω.

Theorem

DMP $\quad \Leftrightarrow \quad G_{hp} \geq 0$ *in* Ω^2

Theorem

Let $\varphi_1, \varphi_2, \dots, \varphi_N$ *be a basis of* V_{hp} *then*

$$G_{hp}(x,y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{A}_{ij}^{-1} \varphi_i(x) \varphi_j(y), \quad \text{where } \mathbb{A}_{ij} = a(\varphi_i, \varphi_j).$$

Remark: $\Omega \subset \mathbb{R}^1$, $\kappa = 0$, $\quad -u'' = f$

$h \leq 0.9|\Omega| \quad \Rightarrow \quad$ DMP (Vejchodský, Šolín, Math. Comp. 2007)

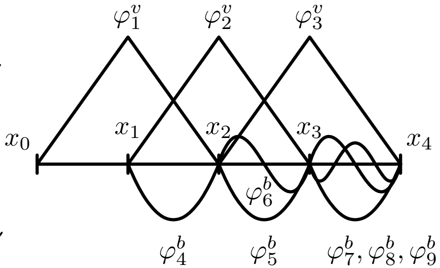
Scheme of the proof in 1D for arbitrary p

- Standard basis

$$\underbrace{\varphi_1^v, \varphi_2^v, \dots, \varphi_M^v}_{\text{vertex part}}, \underbrace{\varphi_{M+1}^b, \dots, \varphi_N^b}_{\text{bubble part}}$$

- New basis

$$\underbrace{\psi_1^v, \psi_2^v, \dots, \psi_M^v}_{\text{vertex part}}, \underbrace{\psi_{M+1}^b, \dots, \psi_N^b}_{\text{bubble part}}$$



$$\psi_i^b = \varphi_i^b$$

$$\psi_i^v = \varphi_i^v - \sum_{j=1}^{M} c_{ij}\varphi_{M+j}^b \quad \text{such that} \quad a(\psi_i^v, \varphi_j^b) = 0 \quad \forall j$$

Scheme of the proof in 1D for arbitrary p

- Standard basis

$$\underbrace{\varphi_1^v, \varphi_2^v, \dots, \varphi_M^v}_{\text{vertex part}}, \underbrace{\varphi_{M+1}^b, \dots, \varphi_N^b}_{\text{bubble part}}$$

- New basis

$$\underbrace{\psi_1^v, \psi_2^v, \dots, \psi_M^v}_{\text{vertex part}}, \underbrace{\psi_{M+1}^b, \dots, \psi_N^b}_{\text{bubble part}}$$

Stiffness matrices

$$\mathbb{A} = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix}$$

$$\widetilde{\mathbb{A}} = \begin{pmatrix} S & 0 \\ 0 & D \end{pmatrix}$$

$$S = A - BD^{-1}B^T$$

$$\psi_i^b = \varphi_i^b$$

$$\psi_i^v = \varphi_i^v - \sum_{j=1}^{M} c_{ij}\varphi_{M+j}^b \quad \text{such that} \quad a(\psi_i^v, \varphi_j^b) = 0 \quad \forall j$$

Scheme of the proof in 1D for arbitrary p

- Standard basis

$$\underbrace{\varphi_1^v, \varphi_2^v, \dots, \varphi_M^v}_{\text{vertex part}}, \underbrace{\varphi_{M+1}^b, \dots, \varphi_N^b}_{\text{bubble part}}$$

- New basis

$$\underbrace{\psi_1^v, \psi_2^v, \dots, \psi_M^v}_{\text{vertex part}}, \underbrace{\psi_{M+1}^b, \dots, \psi_N^b}_{\text{bubble part}}$$

Stiffness matrices

$$\mathbb{A} = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix}$$

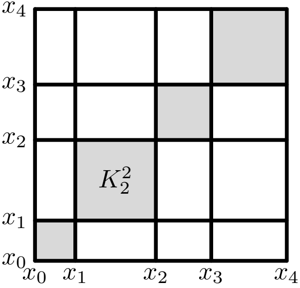
$$\widetilde{\mathbb{A}} = \begin{pmatrix} S & 0 \\ 0 & D \end{pmatrix}$$

$$S = A - BD^{-1}B^T$$

$$G_{hp}(x,y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \widetilde{\mathbb{A}}_{ij}^{-1} \psi_i(x) \psi_j(y)$$

$$= \underbrace{\sum_{i=1}^{M} \sum_{j=1}^{M} S_{ij}^{-1} \psi_i^v(x) \psi_j^v(y)}_{G_{hp}^v(x,y)} + \underbrace{\sum_{i=1}^{N-M} \sum_{j=1}^{N-M} D_{ij}^{-1} \psi_{M+i}^b(x) \psi_{M+j}^b(y)}_{G_{hp}^b(x,y)}$$

Scheme of the proof in 1D for arbitrary p



Stiffness matrices

$$\mathbb{A} = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix}$$

$$\widetilde{\mathbb{A}} = \begin{pmatrix} S & 0 \\ 0 & D \end{pmatrix}$$

$$S = A - BD^{-1}B^T$$

$$G_{hp}(x,y) = \sum_{i=1}^{N}\sum_{j=1}^{N} \widetilde{\mathbb{A}}_{ij}^{-1}\psi_i(x)\psi_j(y)$$

$$= \underbrace{\sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}^{-1}\psi_i^v(x)\psi_j^v(y)}_{G_{hp}^v(x,y)} + \underbrace{\sum_{i=1}^{N-M}\sum_{j=1}^{N-M} D_{ij}^{-1}\psi_{M+i}^b(x)\psi_{M+j}^b(y)}_{G_{hp}^b(x,y)}$$

Scheme of the proof in 1D for arbitrary p

To prove $G_{hp}(x,y) \geq 0$ we require

(a) $\psi_i^v \geq 0$

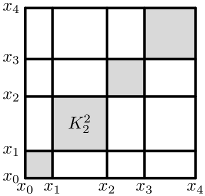
(b) $a(\psi_i^v, \psi_j^v) \leq 0$ for $i \neq j$

(c) $G_{hp}(x,y)|_{K_k^2} = G_{hp}^v(x,y)|_{K_k^2} + G_{hp}^b(x,y)|_{K_k^2} \geq 0$

Stiffness matrices

$$\mathbb{A} = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix}$$

$$\widetilde{\mathbb{A}} = \begin{pmatrix} S & 0 \\ 0 & D \end{pmatrix}$$



$$G_{hp}(x,y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \widetilde{\mathbb{A}}_{ij}^{-1} \psi_i(x)\psi_j(y)$$

$$= \underbrace{\sum_{i=1}^{M} \sum_{j=1}^{M} S_{ij}^{-1} \psi_i^v(x)\psi_j^v(y)}_{G_{hp}^v(x,y)} + \underbrace{\sum_{i=1}^{N-M} \sum_{j=1}^{N-M} D_{ij}^{-1} \psi_{M+i}^b(x)\psi_{M+j}^b(y)}_{G_{hp}^b(x,y)}$$

Verification of (a), (b), (c)

(a) $\kappa^2 h_K^2 \le \alpha^{p_K}$ $\Rightarrow$ $\psi_i^v|_K \ge 0$

(b) $\kappa^2 h_K^2 \le \beta^{p_K}$ $\Rightarrow$ $a(\psi_i^v, \psi_j^v) \le 0$ for $i \ne j$

(c) $\kappa^2 h_K^2 \le \gamma^{p_K} \dfrac{H_{\mathrm{rel}}^K}{1 - H_{\mathrm{rel}}^K} + \delta^{p_K}$ $\Rightarrow$ $G_{hp}(x,y)|_{K_k^2} \ge 0$

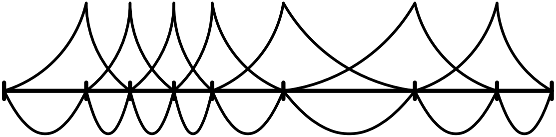
p	α^p	β^p	γ^p	δ^p
1	∞	6	0	∞
2	20/3	∞	0	∞
3	38.61	25.89	5.608	0
4	18.91	∞	2.936	3.614
5	49.44	59.82	7.799	0
6	37.56	∞	7.247	0.887
7	72.82	107.81	9.791	0
8	62.62	∞	9.709	0
9	104.09	169.85	11.510	0
10	94.10	∞	10.644	0

$$H_{\mathrm{rel}}^K = \frac{h_K}{|\Omega|}$$

$$H_{\mathrm{rel}}^K \le 1/3$$

Verification of (c)

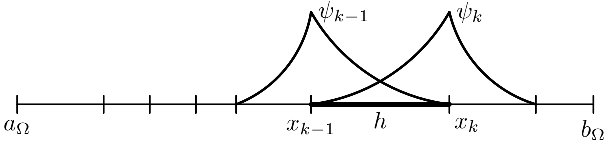
$$V_{hp} = \text{span}\{\psi_1^v, \dots, \psi_M^v, \psi_{M+1}^b, \dots, \psi_N^b\}$$
$$G_{hp} = G_{hp}^v + G_{hp}^b$$



$$G_{hp}^v(x,y) = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}^{-1}\psi_i(x)\psi_j(y) \qquad (x,y) \in K^2$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2} S_{k-2+i,k-2+j}^{-1}\psi_{k-2+i}(x)\psi_{k-2+j}(y)$$
$$\geq \sum_{i=1}^{2}\sum_{j=1}^{2} \widetilde{S}_{k-2+i,k-2+j}^{-1}\widetilde{\psi}_{k-2+i}(x)\widetilde{\psi}_{k-2+j}(y)$$
$$\geq \frac{1}{a(\widehat{\psi},\widehat{\psi})}\widehat{\psi}(\hat{x})\widehat{\psi}(\hat{y})$$

Verification of (c)

$$V_{hp} = \text{span}\{\psi_1^v, \dots, \psi_M^v, \psi_{M+1}^b, \dots, \psi_N^b\}$$
$$G_{hp} = G_{hp}^v + G_{hp}^b$$



$$G_{hp}^v(x,y) = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}^{-1}\psi_i(x)\psi_j(y) \qquad (x,y) \in K^2$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} S_{k-2+i,k-2+j}^{-1}\psi_{k-2+i}(x)\psi_{k-2+j}(y)$$

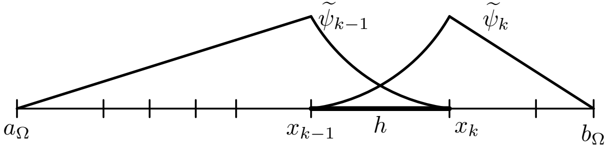
$$\geq \sum_{i=1}^{2}\sum_{j=1}^{2} \widetilde{S}_{k-2+i,k-2+j}^{-1}\widetilde{\psi}_{k-2+i}(x)\widetilde{\psi}_{k-2+j}(y)$$

$$\geq \frac{1}{a(\widehat{\psi},\widehat{\psi})}\widehat{\psi}(\hat{x})\widehat{\psi}(\hat{y})$$

Verification of (c)

$$V_{hp} = \text{span}\{\psi_1^v, \ldots, \psi_M^v, \psi_{M+1}^b, \ldots, \psi_N^b\}$$

$$G_{hp} = G_{hp}^v + G_{hp}^b$$



$$G_{hp}^v(x,y) = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}^{-1}\psi_i(x)\psi_j(y) \qquad (x,y) \in K^2$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} S_{k-2+i,k-2+j}^{-1}\psi_{k-2+i}(x)\psi_{k-2+j}(y)$$

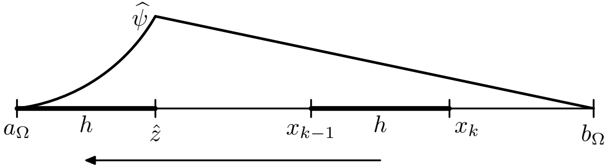
$$\geq \sum_{i=1}^{2}\sum_{j=1}^{2} \widetilde{S}_{k-2+i,k-2+j}^{-1}\widetilde{\psi}_{k-2+i}(x)\widetilde{\psi}_{k-2+j}(y)$$

$$\geq \frac{1}{a(\widehat{\psi},\widehat{\psi})}\widehat{\psi}(\hat{x})\widehat{\psi}(\hat{y})$$

Verification of (c)

$V_{hp} = \text{span}\{\psi_1^v, \dots, \psi_M^v, \psi_{M+1}^b, \dots, \psi_N^b\}$

$G_{hp} = G_{hp}^v + G_{hp}^b$



$$
\begin{aligned}
G_{hp}^v(x,y) &= \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}^{-1}\psi_i(x)\psi_j(y) \qquad\qquad (x,y)\in K^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} S_{k-2+i,k-2+j}^{-1}\psi_{k-2+i}(x)\psi_{k-2+j}(y) \\
&\geq \sum_{i=1}^{2}\sum_{j=1}^{2} \widetilde{S}_{k-2+i,k-2+j}^{-1}\widetilde{\psi}_{k-2+i}(x)\widetilde{\psi}_{k-2+j}(y) \\
&\geq \frac{1}{a(\widehat{\psi},\widehat{\psi})}\widehat{\psi}(\hat{x})\widehat{\psi}(\hat{y})
\end{aligned}
$$

Conclusion

$$-\Delta u + \kappa^2 u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega$$

$$u_{hp} \in V_{hp}: \qquad a(u_{hp}, v_{hp}) = \int_\Omega f v_{hp}\,\mathrm{d}x \quad \forall v_{hp} \in V_{hp}$$

Theorem

Let $\mathcal{T}_{hp}$ be a finite element mesh in an interval $\Omega = (a_\Omega, b_\Omega)$. Let us consider arbitrary polynomial degrees up to 10. Denote by h_K and $H^K_{\mathrm{rel}} = h_K/(b_\Omega - a_\Omega)$ the length and the relative length of the element $K \in \mathcal{T}_{hp}$, respectively. If

$$\frac{\kappa^2 h_K^2}{\kappa^2 h_K^2 + \gamma^3} \le H^K_{\mathrm{rel}} \le 1/3 \quad \text{for all } K \in \mathcal{T}_{hp},$$

where $\gamma^3 \approx 5.608797$, then the approximate problem satisfies the DMP.

Thank you for your attention

Tomáš Vejchodský (vejchod@math.cas.cz)

Institute of Mathematics, Academy of Sciences
Žitná 25, 115 67 Prague 1
Czech Republic