Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/180420

Discontinuous optimal operator for the $\mathbb{P}_1$ Galerkin method approximation of elliptic PDEs with a weakly regular source term

Rachid Messaoudi[1,*], *Abdeluaab Lidouh*[2], *Mohamed A. Hafez*[3] *and Shawkat Alkhazaleh* [4]

[1] Mathematics Department, LaMAO, Multidisciplinary Faculty of Nador, Mohammed First University, Oujda, Morocco
[2] Mathematics Department, LACSA, Faculty of Science of Oujda, Mohammed First University, Oujda, Morocco
[3] Department of Civil Engineering, Faculty of Engineering, FEQS, INTI-IU UNiversity, Nilai, Malaysia
[4] Department of Mathematics, Faculty of sciences and information technology, Jadara University, Irbid, Jordan

Received: 2 Feb. 2024, Revised: 23 Mar. 2023, Accepted: 24 May 2024
Published online: 1 Jul. 2024

Abstract: In this paper, the standard $\mathbb{P}_1$-discontinuous Galerkin method approximation for elliptic PDEs with a weakly regular source term and $\mathbb{L}^\infty$ -coefficients is considered. We propose introducing a new truncated interpolation operator $I^d_{h,k}$ to replace the operator I^k_h used in [1,6]. We prove that it is possible to eliminate a principal constraint imposed on the $N\times N$ stiffness matrix Q.
The statements and proofs of [1,6] remain valid according to the new operator.

Keywords: $\mathbb{P}_1$-discontinuous Galerkin method, truncated interpolation operator, diagonally dominant matrix, stiffness matrix, piecewise affine variant.

1 Introduction and Preliminaries

We consider the Dirichlet problem in $2D$ or $3D$:

$$\begin{cases} -\operatorname{div}(A\nabla u) = f \ \text{ in } \ \Omega, \\ \qquad\qquad u = 0 \ \text{ on } \ \partial\Omega, \end{cases} \tag{1}$$

on Ω (open bounded set of $\mathbb{R}^d$) where $f\in\mathbb{L}^1(\Omega)$, and A is a coercive matrix such that $A\in\mathbb{L}^\infty(\Omega)^{d\times d}$.

The discrete problem considered is

$$\begin{cases} u_h\in\mathbb{V}_h, \\ \forall v_h\in\mathbb{V}_h, \quad a_h^{swip}(u_h,v_h) = \int_\Omega f v_h\, dx. \end{cases} \tag{2}$$

with, $\forall T\in\mathbb{T}_h$ and $\forall F\in\mathbb{F}_h$,

$$\mathbb{V}_h = \{v_h\in\mathbb{L}^2(\Omega):\ v_h|_T\in\mathbb{P}_1[T], \int_F [\![v_h]\!] = 0\}, \tag{3}$$

where the symmetric weighted interior penalty (SWIP) bilinear form a_h^{swip} is defined as in [1,3].

The goal of this paper is to solve problem (2) using the $\mathbb{P}_1$-discontinuous Galerkin method (cf. [1]) and the renormalized solution class (cf. [2,4]), without a diagonal dominance of the stiffness matrix Q as condition (4).

For this purpose, we insert a new truncated interpolation operator $I^d_{h,k}$, and prove the following similar convergence results.

Theorem 1. *The unique renormalized solution u_h of (2), satisfies*

$$\forall k>0;\ \forall q;\ s.t.;\ 1\le q<1+\frac{1}{d-1}:$$

$$u_h\longrightarrow u \quad \textit{strongly in } \ \mathbb{L}^q(\Omega),$$

$$\nabla_h u_h\longrightarrow\nabla u \quad \textit{strongly in } \ [\mathbb{L}^q(\Omega)]^d,$$

$$|u_h|_{J,A,q}\longrightarrow 0,$$

$$I^d_{h,k}(u_h)\longrightarrow T_k(u) \quad \textit{strongly in } \ \mathbb{L}^2(\Omega),$$

$$\nabla_h(I^d_{h,k}(u_h))\longrightarrow\nabla T_k(u) \quad \textit{strongly in } \ \left[\mathbb{L}^2(\Omega)\right]^d,$$

$$\left|I^d_{h,k}(u_h)\right|_{J,A}\longrightarrow 0,$$

when $h\longrightarrow 0$.

* Corresponding author e-mail: m_rachid_ens@yahoo.fr

© 2024 NSP
Natural Sciences Publishing Cor.

The idea is based on the new operator $I^d_{h,k}$ by returning all the data associated with the vertices $s_{i,T}$, at only two points that represent local extrema of $v_h \in \mathbb{L}^2(\overline{\Omega})$. Thus, (depending on each v_h) we find a new 2×2 matrix $\widetilde{Q}$ that easily replaces the following condition (see [1],[5],[6])

$$\forall i \in \{1,2,...,N\} : Q_{ii} - \sum_{\substack{j=1\\ j\neq i}}^{N} |Q_{ij}| \geq 0. \tag{4}$$

Remark. Note that the condition

$$\forall i \in \{1,2,...,N\}, \forall j \neq i : \; Q_{ij} \leq 0, \tag{5}$$

is equivalent to (4), if $s_{i,T}$ is a strictly interior vertex (cf. Remark 6.2 in [6]), which is difficult to achieve if the degree of polynomial approximation exceeds 2. (cf. example in [5]).

Notations

N represents the number of all interior centers c_i of faces F_i in any triangulation $\mathbb{T}_h$.

For every d-simplex $T \in \mathbb{T}_h$, every $v_h \in \mathbb{V}_h$ and every center $c_{i,T}$ of faces $F_{i,T} \in T$, we successively set

$$v_{i,T} = v(c_{i,T}),$$

$$v_{m,T} = \min_{0\leq i\leq d} v_{i,T}, v_{M,T} = \max_{0\leq i\leq d} v_{i,T},$$

$$\varphi_{i,T} := 1 - d\lambda_{i,T}, \; i = 0,...,d, \tag{6}$$

$$\begin{cases} \psi_{m,T} := \displaystyle\sum_{\alpha^{v_h}_{i,T}\neq 1} (1-\alpha^{v_h}_{i,T})\varphi_{i,T}, \\ \psi_{M,T} := \displaystyle\sum_{\alpha^{v_h}_{i,T}\neq 0} \alpha^{v_h}_{i,T}\varphi_{i,T}, \end{cases} \tag{7}$$

where

$$\begin{cases} \alpha^{v_h}_{i,T} := \dfrac{v_{i,T}-v_{m,T}}{v_{M,T}-v_{m,T}}, \; if \; v_{M,T} \neq v_{m,T}, \\ \alpha^{v_h}_{i,T} := 0, \;\; \text{else}. \end{cases}$$

The truncated interpolation operator $I^d_{h,k}$ is defined by

$$\begin{cases} \forall v \in \mathbb{L}^2(\overline{\Omega}) \; s.t. \; \displaystyle\int_F [\![v_h]\!] = 0, \\ I^d_{h,k}(v) := T_k(v_{m,T})\psi_{m,T} + T_k(v_{M,T})\psi_{M,T}. \end{cases} \tag{8}$$

One can easily check that

$$I^d_{h,k}(v_h) = \sum_{i=0}^{d}((1-\alpha^{v_h}_{i,T})T_k(v_{m,T}) + \alpha^{v_h}_{i,T}T_k(v_{M,T}))\varphi_{i,T}, \tag{9}$$

$$v_h(x) = v_{m,T}\psi_{m,T} + v_{M,T}\psi_{M,T}, \tag{10}$$

where

$$|(1-\alpha^{v_h}_{i,T})T_k(v_{m,T}) + \alpha^{v_h}_{i,T}T_k(v_{M,T})| \leq k, \tag{11}$$

$$\psi_{m,T} + \psi_{M,T} = 1, \tag{12}$$

and

$$\psi_{a,T}(c_{j,T}) = \mathbb{1}_{\{i/\, v_{i,T}=v_{a,T}\}}(j), \; a \in \{m,M\}. \tag{13}$$

2 Main results

In this section, we will prove that the main results are similar to those of [1,6] associated with the new operator $I^d_{h,k}$. Our goal is to prove that all the convergence results remain valid but without needing condition (4).

Proposition 1. *If for some $v_h \in \mathbb{V}_h$ and $k > 0$, there exists $z \in T$ s.t. $|v_h(z)| \geq k$; then,*

there exists a d-simplex $T^* \subset T$ and $y \in T$ s.t.

$$\left|I^d_{h,k}(v_h)\right| \geq \frac{k}{2}, \; on \; T^*,$$

where

$$T^* = y - s_{i_0,T} + \left\{x^* \in T, \; \lambda_{i_0,T}(x^*) \geq \frac{11}{12}\right\}.$$

Proof. Let $v_h \in \mathbb{V}_h$, $k > 0$ and $T \in \mathbb{T}_h$.

There exists an element y $\in$ T s.t. $|\, I^d_{h,k}(v_h)(y) \,| \geq k$; indeed,

- if $|\, v_{M,T} \,| < k$, then, by (10)

$$I^d_{h,k}(v_h) = v_{m,T}\psi_{m,T} + v_{M,T}\psi_{M,T} = v_h,$$

so,

$$y = z,$$

- if $|\, v_{M,T} \,| \geq k$, it follows that

$$I^d_{h,k}(v_h)(c_{M,T}) = k,$$

and one can take

$$y = c_{M,T}.$$

On the other hand, it is possible to find $i_0 \in \{0,1,...d\}$ s.t.

$$\lambda_{i_0,T}(y) \geq \frac{1}{12}. \tag{14}$$

Therefore, if we consider the d-simplex contained in T defined as follows:

$$T^* = \left\{x^* \in T, \lambda_{i_0,T}(x^*) \geq \frac{11}{12}\right\}, \tag{15}$$

© 2024 NSP
Natural Sciences Publishing Cor.

this allows one to see that

$$\forall x^* \in T^*,\ \forall x = y - s_{i_0,T} + x^* : x \in T,$$

thanks to

$$y - s_{i_0,T} + x^* = \sum_{\substack{i=0\\ i\neq i_0}}^{d} (\lambda_{i,T}(y) + \lambda_{i,T}(x^*))s_{i,T} + \\ + (\lambda_{i_0,T}(y) - 1\lambda_{i_0,T}(x^*))s_{i_0,T},$$

and to

$$\lambda_{i_0,T}(y) - 1 + \lambda_{i_0,T}(x^*) \geq 0,$$

deducted from (14) and (15).

Then, one can argue that

$$T^* \subset T.$$

Using (9), we can establish the following identity

$$\nabla(I^d_{h,k}(v_h)) = -d\sum_{i=0}^{d}\Big((1-\alpha^{v_h}_{i,T})T_k(v_{m,T}) + \\ + \alpha^{v_h}_{i,T}T_k(v_{M,T})\Big)\nabla\lambda_{i,T}$$

which, together with (11) and the identity

$$\nabla\lambda_{i,T}(s_{j,T} - s_{i_0,T}) = \delta_{i,j}(1-\delta_{i_0,j}),$$

yields

$$\left|\nabla(I^d_{h,k}(v_h))(s_{j,T} - s_{i_0,T})\right| \leq 2kd \leq 6k.$$

Recalling that

$$x - y = \sum_{\substack{j=0\\ j\neq i_0}}^{d} (\lambda_{j,T}(x) - \lambda_{j,T}(y))\,(s_{j,T} - s_{i_0,T}),$$

so, we observe that, $\forall x \in T^*$

$$\begin{aligned}\left|I^d_{h,k}(v_h)(x) - I^d_{h,k}(v_h)(y)\right| &= \left|\nabla(I^d_{h,k}(v_h))(x-y)\right| \\ &\leq 6k\sum_{\substack{j=0\\ j\neq i_0}}^{d}\left|\lambda_{j,T}(x) - \lambda_{j,T}(y)\right| \\ &\leq 6k\sum_{\substack{j=0\\ j\neq i_0}}^{d}\left|\nabla\lambda_{j,T}(x-y)\right| \\ &\leq 6k\sum_{\substack{j=0\\ j\neq i_0}}^{d}\left|\nabla\lambda_{j,T}(x^* - s_{i_0,T})\right| \\ &\leq 6k\sum_{\substack{j=0\\ j\neq i_0}}^{d}\left|\lambda_{j,T}(x^*) - \lambda_{i,T}(s_{i_0,T})\right| \\ &\leq 6k\sum_{\substack{j=0\\ j\neq i_0}}^{d}\lambda_{j,T}(x^*) \leq \frac{k}{2}.\end{aligned}$$

This completes the proof. □

Now, we will prove the following proposition without using condition (4) which is imposed in [1,5,6].

Proposition 2. *For every $v_h \in \mathbb{V}_h$ and every $k > 0$,*

$$a^{swip}_h(v_h - I^d_{h,k}(v_h), I^d_{h,k}(v_h)) \geq 0. \tag{16}$$

Proof. Since

$$v_h(x) = v_{m,T}\psi_{m,T}(x) + v_{M,T}\psi_{M,T}(x),$$

and

$$I^d_{h,k}(v_h)(x) = T_k(v_{m,T})\psi_{m,T}(x) + T_k(v_{M,T})\psi_{M,T}(x),$$

it follows that

$$a^{swip}_h(v_h - I^d_{h,k}(v_h), I^d_{h,k}(v_h)\ \mathbb{1}_T) = \sum_{i\in\{m,M\}} Z^{v_h}_i$$

where

$$Z^{v_h}_i := (v_{i,T} - T_k(v_{i,T}))\left(T_k(v_{i,T})\widetilde{Q}_{ii,T} + T_k(v_{j,T})\widetilde{Q}_{ij,T}\right)\ (j \neq i),$$

and

$$\widetilde{Q}_{ij,T} := a^{swip}_h(\psi_{i,T}, \psi_{j,T}\ \mathbb{1}_T).$$

Fixing $i \in \{m, M\}$, there are two possibilities:

- if $|v_{i,T}| \leq k$, then $v_{i,T} - T_k(v_{i,T}) = 0$ and,

$$Z^{v_h}_i = 0,$$

- if $|v_{i,T}| > k$, note that $v_{i,T} - T_k(v_{i,T})$ has the same sign as $T_k(v_{i,T})$; therefore,

$$(v_{i,T} - T_k(v_{i,T}))T_k(v_{i,T}) = k|v_{i,T} - T_k(v_{i,T})|.$$

© 2024 NSP
Natural Sciences Publishing Cor.

The above identity combined with

$$\widetilde{Q}_{ii,T} - |\widetilde{Q}_{ij,T}| = \widetilde{Q}_{ii,T} + \widetilde{Q}_{ij,T} = 0, i,j \in \{m,M\}\ (i \neq j).$$

leads to

$$Z_i^{v_h} \geq k|v_{i,T} - T_k(v_{i,T})|(\widetilde{Q}_{ii,T} - |\widetilde{Q}_{ij,T}|) \geq 0.$$

Hence, in both cases, we claim that

$$Z_i^{v_h} \geq 0, i \in \{m,M\}.$$

Therefore, the inequality (16) is deduced. □

Proposition 3. *Let $k > 0$; the following bound holds for any $v_h \in \mathbb{V}_h$*

$$||I^d_{h,k}(v_h)||_\infty \leq k(d^2-1). \tag{17}$$

Proof. Let $v_h \in \mathbb{V}_h$ and $k > 0$, so from (9)

$$\begin{aligned} ||I^d_{h,k}(v_h)||_\infty &\leq \sum_{i=0}^{d} |(1-\alpha^{v_h}_{i,T})T_k(v_{m,T}) + \alpha^{v_h}_{i,T}T_k(v_{M,T})| \\ &\leq k(d^2-1), \end{aligned}$$

since

$$\max_{0\leq i\leq d} |\varphi_{i,T}| = d-1.$$

Therefore, (17) is obtained. □

We now prove the following main result, which is a piecewise affine variant according to the result of L. Boccardo & T. Gallouët [4,8].

Theorem 2. *For every $k > 0$ and every $h > 0$, the unique renormalized solution u_h of (2) satisfies*

$$\int_\Omega |\nabla_h I^d_{h,k}(u_h)|^2\, dx \ \leq\ kC_1\ \ ||\, f\, ||_{\mathbb{L}^1(\Omega)}\,. \tag{18}$$

where the constant C_1 is independent of h.

Proof. The use of $I^d_{h,k}(u_h)$ as a test function in (2) combined with (16) leads us to

$$a_h^{swip}(I^d_{h,k}(u_h), I^d_{h,k}(u_h))\ \leq \int_\Omega f I^d_{h,k}(u_h)dx. \tag{19}$$

Based on the coercivity hypothesis of A (Theorem 2.2 in [1]), one can write

$$\begin{aligned} &\int_\Omega |\nabla_h(I^d_{h,k}(u_h))|^2\, dx \leq \\ &\leq \alpha^{-1}\int_\Omega |A\nabla_h(I^d_{h,k}(u_h))\nabla_h(I^d_{h,k}(u_h))|\, dx \\ &\leq \alpha^{-1}\, ||\, I^d_{h,k}(u_h)\, ||^2_{swip}, \end{aligned}$$

together with (17), (19) and the discrete coercivity of the SWIP bilinear form a_h^{swip} (Lemma 4.51 in[3]), we see that

$$\begin{aligned} ||\, I^d_{h,k}(u_h)\, ||^2_{swip} &\leq C^* a_h^{swip}(I^d_{h,k}(u_h), I^d_{h,k}(u_h)) \\ &\leq C^*\, ||\, I^d_{h,k}(u_h)\, ||_\infty ||\, f\, ||_{\mathbb{L}^1(\Omega)}, \end{aligned}$$

allows us to deduce the estimate (18),
where $C^* := \dfrac{\eta+1}{\eta-(d+1)C_{tr}^2}$ and $C_1 := \alpha^{-1}C^*$. □

Theorem 3. *Let $v_h \in \mathbb{V}_h$. For every q s.t. $1 \leq q < 1 + \dfrac{1}{d-1}$ it holds that*

$$\|u_h\|_{swip,q} \leq C_2\ ||\, f\, ||_{\mathbb{L}^1(\Omega)}\,. \tag{20}$$

where the constant C_2 is independent of h.

Proof. (cf. [1,6])
Let $\lambda > 0$ and $k > 0$. If $\max |u_{h|_T}| < k$, then

$$I_h^k(u_h)_{|_T} = u_{h|_T}.$$

Combined with (18), this implies

$$\begin{aligned} &|\bigcup_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|<k}} \{x \in T : |\nabla_h u_h| \geq \lambda\}| \leq \\ &\leq |\bigcup_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|<k}} \left\{x \in T : |\nabla_h(I_h^k(u_h))| \geq \lambda\right\}| \\ &\leq \frac{1}{\lambda^2}\int_\Omega |\nabla_h(I_h^k(u_h))|^2\, dx \\ &\leq \frac{C_1 k}{\lambda^2}||f||_{\mathbb{L}^1(\Omega)} \\ &\leq \left(\frac{||f||_{\mathbb{L}^1(\Omega)}}{\lambda}\right)^{\frac{22^*}{2+2^*}}, \end{aligned}$$

for

$$k = \frac{1}{C_1}\sqrt[2^*+2]{\lambda^4 ||f||^{2^*-2}_{\mathbb{L}^1(\Omega)}}. \tag{21}$$

Hence,

$$\begin{aligned} \sum_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|\geq k}} |T| &= \sum_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|\geq k}} \frac{1}{C_0}|T^*| \\ &\leq \sum_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|\geq k}} \frac{1}{C_0}\left(\frac{2}{k}\right)^{2^*}\int_{T^*}\left|I_h^k(u_h(x))\right|^{2^*} \mathrm{d}x \\ &\leq \frac{1}{C_0}\left(\frac{2}{k}\right)^{2^*}\int_\Omega \left|I_h^k(u_h(x))\right|^{2^*} \mathrm{d}x \\ &\leq \frac{1}{C_0}\left[\frac{2\sigma_{2,2^*}\sqrt{1+C_3}}{k}\right]^{2^*} |||\ \nabla_h(I_h^k(u_h(x)))\,|||^{2^*}_{\mathbb{L}^2(\Omega)}, \end{aligned}$$

where $C_3 := C(\sigma,d)$ (see Lemma 3.2 [1]).

© 2024 NSP
Natural Sciences Publishing Cor.

So, by (18), one can see that

$$\sum_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|\geq k}} |T| \leq \frac{\left[2\sigma_{2,2^*}\sqrt{C_1(1+C_3)}\right]^{2^*}}{C_0}\left[\frac{\|f\|_{\mathbb{L}^1(\Omega)}}{k}\right]^{\frac{2^*}{2}}. \tag{22}$$

Combining the above result with (21) yields

$$\sum_{\substack{T\in\mathbb{T}_h\\ \max|u_{h|_T}|\geq k}} |T| \leq \frac{1}{C_0}\left[2\sigma_{2,2^*}C_1\sqrt{1+C_3}\right]^{2^*}\left[\frac{\|f\|_{\mathbb{L}^1(\Omega)}}{\lambda}\right]^{\frac{22^*}{2^*+2}},$$

it follows that

$$\||\nabla u_h|\|_{\mathbb{L}^{\frac{22^*}{2^*+2},\infty}(\Omega)} \leq C_4\|f\|_{\mathbb{L}^1(\Omega)},$$

where

$$C_4=\left[\frac{\left(2\sigma_{2,2^*}C_1\sqrt{1+C_3}\right)^{2^*}}{C_0}+1\right]^{\frac{1}{2}+\frac{1}{2^*}}.$$

using Lemma 3.2 (in [1]) and the embedding inequality

$$\||\nabla_h v_h|\|_{\mathbb{L}^q(\Omega)} \leq C(q,r,|\Omega|)\,\||\nabla v_h|\|_{\mathbb{L}^{r,\infty}(\Omega)}.$$

We infer

$$\|u_h\|_{swip,q} \leq C\big(q,\sigma,d,\|A\|_{\mathbb{L}^\infty(\Omega)^{d\times d}}\big)\|\nabla_h u_h\|_{\mathbb{L}^{\frac{22^*}{2^*+2}}(\Omega)},$$

whence the assertion (20). □

Theorem 4. *Under the assumptions of (Theorem 2.2 in [1]), the solution u_h of (2) satisfies for every q with* $1\leq q<1+\dfrac{1}{d-1}$

$$u_h \longrightarrow u \quad \textit{strongly in } \mathbb{L}^q(\Omega),$$

$$\nabla_h u_h \longrightarrow \nabla u \quad \textit{strongly in } [\mathbb{L}^q(\Omega)]^d,$$

$$|u_h|_{J,A,q} \longrightarrow 0,$$

when h tends to zero, where u is the unique renormalized solution of (1).

Proof. Let $n\in\mathbb{N}$, $\rho>0$ and $f_\rho=T_{\frac{1}{\rho}}(f)$. If $u_{h,\rho}$ denotes the unique solution of the problem

$$\begin{cases} u_{h,\rho}\in\mathbb{V}_h,\\ \forall v_h\in\mathbb{V}_h,\quad a_h^{swip}(u_{h,\rho},v_h)=\displaystyle\int_\Omega f_\rho v_h\,dx. \end{cases} \tag{23}$$

So, one can see that

$$\forall v_h\in\mathbb{V}_h,\quad a_h^{swip}(u_h-u_{h,\rho},v_h)=\int_\Omega (f-f_\rho)v_h\,dx.$$

It is known (see [9]) that

$$u_{h,\rho}\longrightarrow u_\rho \quad \text{strongly in } \mathbb{L}^2(\Omega), \tag{24}$$

$$\nabla_h u_{h,\rho}\longrightarrow \nabla u_\rho \quad \text{strongly in } \left[\mathbb{L}^2(\Omega)\right]^d, \tag{25}$$

$$|u_{h,\rho}|_{J,A}\longrightarrow 0, \tag{26}$$

when h tends to zero, where u_ρ is the unique renormalized solution of the problem

$$\begin{cases} -\mathrm{div}(A\nabla u_\rho)=f_\rho \text{ in } \Omega,\\ u_\rho=0 \quad \text{on } \partial\Omega. \end{cases} \tag{27}$$

The estimate (see Theorem 2.1 in [1])

$$\alpha\|u_\rho-u\|_{\mathbb{W}_0^{1,q}(\Omega)} \leq C(d,|\Omega|,q)\,\|f_\rho-f\|_{\mathbb{L}^1(\Omega)},$$

yields

$$\|u_\rho-u\|_{swip,q}\leq \frac{\|A\|_{\mathbb{L}^\infty(\Omega)^{d\times d}}\,C(d,|\Omega|,q)}{\alpha}\|f_\rho-f\|_{\mathbb{L}^1(\Omega)}. \tag{28}$$

Therefore, from the inequality (20) together with (25), (26) and (28), we see that

$$\limsup_{h\to 0}\|u_h-u\|_{swip,q}\leq C_5\|f-f_\rho\|_{\mathbb{L}^1(\Omega)},$$

for every $\rho>0$ and every q s.t. $1\leq q<1+\dfrac{1}{d-1}$, with $C_5:=\alpha^{-1}C^*C\big(d,|\Omega|,q,\sigma,\|A\|_{\mathbb{L}^\infty(\Omega)^{d\times d}}\big)$.
Again, from Theorem 2.1 in [1],

$$\lim_{\rho\to 0}\|f-f_\rho\|_{\mathbb{L}^1(\Omega)}=0.$$

This completes the proof. □

Proposition 4. *Under the assumptions of (Theorem 2.2 in [1]), the solution u_h of (2) satisfies*

$$|I_h^k(u_h)-T_k(u_h)|=O(h). \tag{29}$$

Proof. Let us consider the set $\mathscr{B}_{k,s}(v)$ defined by

$$\mathscr{B}_{k,s}(v)=\bigcup\{T\in\mathbb{T}_h:\min_T|v|\leq s,\max_T|v|\geq k\}. \tag{30}$$

Using Lemma 3.5 in [1], one can write

$$|\mathscr{B}_{k,s}(v)|\leq\frac{h^2}{(k-s)^2}\int_\Omega|\nabla_h(v)|^2\,\mathrm{d}x, \tag{31}$$

combining (31) with (18), yields

$$|\mathscr{B}_{k,s}(I^d_{h,\frac{1}{h}}(u_h))|=O(h). \tag{32}$$

© 2024 NSP
Natural Sciences Publishing Cor.

Furthermore, from (22), we observe that

$$\sum_{\substack{T\in\mathbb{T}_h \\ \max|u_{h|T}|\geq\frac{1}{h}}} |T| = O(h^{\frac{2^*}{2}}). \quad (33)$$

Finally, since

$$\bigcup\{T\in\mathscr{B}_{k,s}(u_h),\ \max|u_{h|T}|\geq\frac{1}{h}\}\subset\mathscr{B}_{k,s}(I^d_{h,\frac{1}{h}}(u_h)),$$

it follows that

$$|\mathscr{B}_{k,s}(u_h)|\leq|\mathscr{B}_{k,s}(I^d_{h,\frac{1}{h}}(u_h))|+\sum_{\substack{T\in\mathscr{B}_{k,s}(u_h) \\ \max|u_{h|T}|<\frac{1}{h}}}|T|.$$

Therefore, through (32) and (33), we can see that

$$|\mathscr{B}_{k,s}(u_h)|=O(h),$$

where $\frac{2^*}{2}>2$.

This allows us to deduce that

$$\left|\mathscr{B}_{k,s}(u_h)\right|\underset{h\to 0}{\longrightarrow}0. \quad (34)$$

Fix $k>0$ and $\varepsilon>0$ such that $\varepsilon<k$ and let us consider

$$\mathscr{E}_\varepsilon=\left\{x\in\Omega:\left|I^d_{h,k}(u_h(x))-T_k(u_h(x))\right|\geq\varepsilon\right\}.$$

Let $x\in\mathscr{I}_\varepsilon$ and $T\in\mathbb{T}_h$ with $x\in T$. It is easily checked that

$$I^d_{h,k}(u_h)_{|T}\neq T_k(u_h)_{|T},$$

what implies that $\max_T|v_h|>k$. So there are four possible cases.

i) $\min(u_{h|T})=0$ or $\min(I^d_{h,k}(u_h)_{|T})=0$: then,

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k,s}(u_h)\cup\mathscr{B}_{k,s}\left(I^d_{h,k}(u_h)\right),$$

for every $s\in]0,k[$.

(ii) $0\leq u_{m,T}\leq u_{M,T}\leq k$ or $-k\leq u_{m,T}\leq u_{M,T}\leq 0$, so

$$I^d_{h,k}(u_h)_{|T}=u_{h|T},$$

then

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k+\varepsilon,k}(I^d_{h,k}(u_h)).$$

(iii) $0\leq u_{m,T}<k<u_{M,T}$ or $u_{m,T}<-k<u_{M,T}\leq 0$, so

- If $\left|I^d_{h,k}(u_h(x))\right|-|T_k(u_h(x))|\geq\varepsilon$, we distinguish 3 cases:

1^{st} case : $|u_h(x)|\geq k$ so,

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k+\varepsilon,k}\left(I^d_{h,k}(u_h)\right),$$

2^{nd} case : $|u_h(x)|<k-\frac{\varepsilon}{2}$, then

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k,k-\frac{\varepsilon}{2}}(u_h),$$

3^{th} case : $k-\frac{\varepsilon}{2}\leq|u_h(x)|<k$, so

$$\left|I^d_{h,k}(u_h(x))\right|\geq k+\frac{\varepsilon}{2},$$

then,

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k+\frac{\varepsilon}{2},k}(u_h),$$

- If $|T_k(u_h(x))|-\left|I^d_{h,k}(u_h(x))\right|\geq\varepsilon$, so

$$\left|I^d_{h,k}(u_h(x))\right|\leq k-\varepsilon,$$

therefore,

$$\mathscr{E}_\varepsilon\subset\mathscr{B}_{k,k-\varepsilon}\left(I^d_{h,k}(u_h)\right).$$

(29) is then a consequence of (18), (31) and (34). □

Theorem 5. *Under the assumptions of (Theorem 2.2 in [1]), the solution u_h of (2) satisfies*

$$I^d_{h,k}(u_h)\underset{h\to 0}{\longrightarrow}T_k(u)\quad \textit{strongly in } \mathbb{L}^2(\Omega), \quad (35)$$

$$\nabla_h(I^d_{h,k}(u_h))\underset{h\to 0}{\longrightarrow}\nabla T_k(u)\quad \textit{strongly in } \left[\mathbb{L}^2(\Omega)\right]^d, \quad (36)$$

$$\left|I^d_{h,k}(u_h)\right|_{J,A}\underset{h\to 0}{\longrightarrow}0, \quad (37)$$

for every $k>0$.

Proof. From the assertion (18 together with (29) and (Theorem 5.7 in [3]) it follows that

$$I^d_{h,k}(u_h)\underset{h\to 0}{\longrightarrow}T_k(u)\quad \text{strongly in } \mathbb{L}^2(\Omega), \quad (38)$$

$$G^l_h(I^d_{h,k}(u_h))\underset{h\to 0}{\rightharpoonup}\nabla T_k(u)\quad \text{Weakly in } \left[\mathbb{L}^2(\Omega)\right]^d, \quad (39)$$

for every $k>0$.

On the other hand, following (17), discrete Rellich-Kondrachov's compactness theorem (theorem 5.6 in [3]) and Lebesgue's dominated convergence theorem, we observe that,

$$\int_\Omega fI^d_{h,k}(u_h)\,\mathrm{d}x\underset{h\to 0}{\longrightarrow}\int_\Omega fT_k(u)\,\mathrm{d}x.$$

Combining the above result with (19) yields

$$\limsup_{h\to 0}a^{swip}_h(I^d_{h,k}(u_h),I^d_{h,k}(u_h))\ \leq\int_\Omega fT_k(u)dx.$$

Furthermore, according to Proposition 4.36 in [3] (take $v_h=I^d_{h,k}(u_h)$), it follows that

$$\int_\Omega A\ G^l_h(I^d_{h,k}(u_h))\ G^l_h(I^d_{h,k}(u_h))\ \leq a^{swip}_h(I^d_{h,k}(u_h),I^d_{h,k}(u_h)).$$

© 2024 NSP
Natural Sciences Publishing Cor.

Hence,

$$\limsup_{h\to 0}\int_{\Omega} A\, G_h^l(I_h^k(u_h))\, G_h^l(I_h^k(u_h)) \;\leq \int_{\Omega} fT_k(u)dx. \tag{40}$$

Therefore, owing to Definition 1.1 in [6] for the renormalized solution u, of (1), we see that

$$\int_{\Omega} A\, \nabla T_k(u)\nabla T_k(u) = \int_{\Omega} fT_k(u)dx, \tag{41}$$

which, combined with (40), leads us to claim that

$$\limsup_{h\to 0}\int_{\Omega} A\, G_h^l(I_{h,k}^d(u_h))\, G_h^l(I_{h,k}^d(u_h)) \leq \int_{\Omega} A\, \nabla T_k(u)\nabla T_k(u),$$

and using (39)

$$G_h^l(I_{h,k}^d(u_h)) \underset{h\to 0}{\longrightarrow} \nabla T_k(u) \quad \text{strongly in} \quad \left[\mathbb{L}^2(\Omega)\right]^d. \tag{42}$$

We also claim by (Proposition 4.36 in [3]), that for all $v_h \in \mathbb{V}_h$ and all $\eta > (d+1)C_{tr}^2$:

$$\left|I_{h,k}^d(u_h)\right|_{J,A}^2 \leq \frac{1}{\eta-(d+1)C_{tr}^2}\left[a_h^{swip}(I_{h,k}^d(u_h),I_{h,k}^d(u_h))+ - \left\|A^{\frac{1}{2}}\, G_h^l(I_{h,k}^d(u_h))\right\|_{[\mathbb{L}^2(\Omega)]^d}^2\right]. \tag{43}$$

Since the right-hand side tends to zero, the assertion (37) follows.

Finally, combining (Proposition 4.34 in [3]) with the triangle inequality yields

$$\left\|\nabla_h I_{h,k}^d(u_h) - \nabla T_k(u)\right\|_{[\mathbb{L}^2(\Omega)]^d} \leq \sqrt{d+1}\, C_{tr}\left|I_{h,k}^d(u_h)\right|_{J,A} + +\left\|G_h^l(I_{h,k}^d(u_h)) - \nabla T_k(u)\right\|_{[\mathbb{L}^2(\Omega)]^d}.$$

The proof of (36) is then completed. □

3 Example

Let us consider Poisson's equation:

$$-\Delta u = f,$$

with $\Omega = [-1,1]^d$ (d=2 or 3), where f is a point source according to the Dirac delta distribution.

From Proposition 4 and (Proposition 3.2 in [1]), we observe that

$$|I_{h,k}^d(u_h) - I_h^k(u_h)| = O(h)$$

where $I_h^k(u_h)$ is the usual operator used in [1].

In particular, the solution u_h in this example, is symmetrical with respect to the zero center of Ω.

Indeed, this symmetry results from the following factors:

- a uniform triangulation which is symmetrical with respect to the center o,
- the symmetry of the Dirac distribution. It can be approximated for example by using Gaussian functions,
- the stiffness matrix Q is symmetrical by construction,
- the set Ω is symmetrical with respect to its center o,
- the Dirichlet conditions are also symmetrical with respect to the point o,
- and the matrix A is symmetrical since $A = id$ for a Poisson's equation.

Then, one can easily see that on the 2^{d-1} diagonals

$$I_{h,k}^d(u_h) = I_h^k(u_h).$$

4 Conclusion and perspectives

The advantage of this work is that it approaches the solution of linear elliptic problems with $\mathbb{L}^1$ data by searching for the optimal unconstrained triangulation in our discontinuous affine case [1]. We try to increase the degree of approximation in this same case. However, the constraint (4) for the conformal quadratic approximation [5] is difficult, and it is necessary to look for an alternative in parallel to make examples for an analysis of the convergence rate and error.

References

[1] L. Abdeluaab and M. Rachid, Affine Discontinuous Galerkin Method Approximation of Second-Order Linear Elliptic Equations in Divergence Form with Right-Hand Side in $\mathbb{L}^1$, *International Journal of Differential Equation*, **vol. 2018**, 4650512.

[2] G. Dal Maso, F. Murat, L. Orsina and A. Prignet, Renormalized solutions of elliptic equations with general measure data, *Ann. Scuola Norm. Sup. Pisa,* **28**, 741–808, (1999).

[3] D. A. Di Pietro, A. Ern, and J.-L. Guermond. Discontinuous Galerkin methods for anisotropic semi definite diffusion with advection, *SIAM J. Numer. Anal.,* **46(2)**, 805-831, (2008).

[4] P. Bénilan, L. Boccardo, T. Gallouët, R. Gariepy, M. Pierre and J.L. Vázquez, An $\mathbb{L}^1$-theory of existence and uniqueness of solutions of nonlinear elliptic equations, *Ann. Scuola Norm. Sup. Pisa,* **22(4)**, (1995), 241–273

[5] L. Abdeluaab, M. Rachid and S. Belkassem, Analysis of a quadratic Finite Element Method for Second-Order Linear Elliptic PDE, with low regularity data, *Numerical Functional Analysis and Optimization,* **41(5)**, 507–541.

[6] J. Casado-Díaz, T. Chacón Rebollo, V. Girault, M. Gómez Mármol, and F. Murat, Finite elements approximation of second order linear elliptic equations in divergence form with right-hand side in $\mathbb{L}^1$, *Numer. Math.,* **105(3)**, 337–374, (2007).

© 2024 NSP
Natural Sciences Publishing Cor.

[7] T. Gallouët, R.Herbin, and J.-C. Latché, A convergent finite element-finite volume scheme for the compressible stokes problem. Part 1: The isothermal case, *Mathematics of computation*, 1333–1352, (2009).
[8] L. Boccardo and T. Gallouët, Nonlinear elliptic and parabolic equations involving measure data, *J. Funct. Anal.,* **87**, 149–169, (1989).
[9] P.G. Ciarlet., The finite element method for elliptic problems, *North- Holland, Amsterdam,* (1978).
[10] A. Ern ,J-L Guermond., Theory and Practice of Finite Elements, *Applied Mathematical series,* **159**, springer, New York, (2004).

© 2024 NSP
Natural Sciences Publishing Cor.