

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I



www.sciencedirect.com

Numerical Analysis

A compact cell-centered Galerkin method with subgrid stabilization

Une méthode de Galerkine centrée aux mailles avec stencil compact et stabilisation de sous-grille

Daniele A. Di Pietro

IFP Energies nouvelles, 1 & 4, avenue de bois Préau, 92852 Rueil-Malmaison cedex, France

ARTICLE INFO	ABSTRACT							
Article history: Received 27 April 2010 Accepted after revision 16 November 2010	In this work we propose a compact cell-centered Galerkin method with subgrid stabilization for anisotropic heterogeneous diffusion problems on general meshes. Both essential theoretical results and numerical validation are provided.							
Available online 3 December 2010	© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.							

Presented by the Editorial Board

RÉSUMÉ

On propose une méthode de Galerkine centrée aux mailles avec stencil compact et stabilisation de sous-grille pour des problèmes de diffusion anisotrope et hétérogène. On présente à la fois les résultats théoriques essentiels et une validation numérique. © 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Version française abrégée

Soit $\Omega \subset \mathbb{R}^d$, $d \ge 2$, un ouvert polyhédrique borné. On considère le problème modèle suivant :

$$-\nabla \cdot (\boldsymbol{\kappa} \nabla \boldsymbol{u}) = f \quad \text{dans} \ \Omega, \qquad \boldsymbol{u} = \boldsymbol{0} \quad \text{sur} \ \partial \Omega, \tag{1}$$

où $f \in L^2(\Omega)$ et κ est un champ de diffusion tensoriel symétrique uniformement défini positif et constant par morceaux sur Ω . En posant $V \stackrel{\text{def}}{=} H_0^1(\Omega)$, la formulation faible de (1) s'écrit

Trouver
$$u \in V$$
 t.q. $\int_{\Omega} \kappa \nabla u \cdot \nabla v = \int_{\Omega} f v$ pour tout $v \in V$. (2)

La discrétisation de (2) par des méthodes de bas ordre sur des maillages généraux qui soient robustes par rapport à l'hétérogeneité et à l'anisotropie du tenseur de diffusion κ , et qui aient en même temps des bonnes proprietés de stabilité est un sujet actif de recherche. Les méthodes de volumes finis classiques réposent sur des bilans par maille faisant intervenir des flux numériques. L'idée d'obtenir des flux numériques à l'aide une reconstruction locale peut être attribuée à Aavatsmark et al. et à Edwards et al. (voir, respectivement, [1] et [6] pour une bibliographie). Le problème principal de ces méthodes est qu'il n'existe souvent pas de conditions simples à vérifier qui en garantissent la stabilité. Plus récemment, des approches variationnelles ont été proposées dans [7]. Les méthodes de ce type sont inconditionnellement coercives au prix d'un stencil

E-mail address: dipietrd@ifpenergiesnouvelles.fr.

¹⁶³¹⁻⁰⁷³X/\$ – see front matter © 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved. doi:10.1016/j.crma.2010.11.017



(a) Notation for the submesh

(b) Notation for an L-group g (in thick line)

(c) Basis function φ_T

Fig. 1. (a) and (b) Notation. (c) An example of actual basis function $\varphi_T = \mathcal{I}_h(\mathbf{e}_T)$ for the space V_h of (3) on a 3 × 3 quadrangular mesh. Observe that $\operatorname{supp}(\varphi_T)$ (thick line) is comprised between the maximal and minimal supports depicted on the left of Fig. 2(a). It can also be appreciated that φ_T is continuous across the faces of T.

élargi. De plus, elles ne sont définies qu'au niveau discret, ce qui ne permet pas d'appliquer les techniques d'analyse typiques des méthodes de Galerkine/éléments finis. L'analyse de convergence repose alors sur un argument de compacité, qui ne fournit pas une estimation de l'ordre de convergence [7,2]. Une alternative a été suggéré dans [4], où on propose une famille de méthodes variationnelles centrées aux mailles dites *cell-centered Galerkin* (ccG), et basées sur des espaces polynomiaux par morceaux incomplets. Ces méthodes sont inconditionnellement coercives, et il est possible d'étendre la forme bilinéaire à un espace contenant la solution exacte. L'analyse de convergence se fait alors dans l'esprit du Lemme de Céa. Le desavantage principal de la famille de méthodes ccG de [4] consiste à avoir un stencil élargi, comparable à celui de [7]. Dans cette note, on propose une variation qui réduit significativement le stencil tout en gardant les bonnes propriétés de la méthode originale (coercivité inconditionnnelle, robustesse par rapport à l'hétérogeneité et à l'anisotropie de κ). L'idée est de réduire le support des fonctions de base de l'espace discret en introduisant une reconstruction opportune sur une sousgrille. La consistance et la symétrie sont garanties par des termes de bord controlés par une stabilisation aux sous-faces. Dans les codes industriels (parallèles, d = 3), un stencil compact est crucial non seulement pour réduire l'occupation en mémoire, mais aussi pour minimiser les échanges de messages entre processeurs. En outre, vitesse de calcul et robustesse sont souvent faites passer avant la précision.

1. Discrete setting

1.1. Notation

The discretization of problem (2) is based on general polygonal meshes. In the context of the *h*-convergence analysis, we denote by $\mathcal{H} \subset \mathbb{R}^*_+$ a countable set having 0 as its unique accumulation point, and we adopt the concept of *admissible mesh sequence* $(\mathcal{T}_h)_{h \in \mathcal{H}}$ introduced in [2, Definition 2.1] (which may be consulted for further details on the regularity assumptions). We briefly recall here the relevant notations. For a given *meshsize* $h \in \mathcal{H}$, $\mathcal{T}_h = \{T\}$ denotes a partition of Ω into polygonal elements whose faces are portions of hyperplanes and such that κ is constant inside each element. The set of mesh vertices is denoted by \mathcal{N}_h . An *interface* is a portion of hyperplane $F \subset \partial T_1 \cap \partial T_2$, $T_1, T_2 \in \mathcal{T}_h$, with unit normal \mathbf{n}_F pointing out of T_1 , whereas a *boundary face* is a portion of hyperplane $F \subset \partial T \cap \partial \Omega$ with $T \in \mathcal{T}_h$ and outward normal \mathbf{n}_F . Interfaces are collected in the set \mathcal{F}_h^i , boundary faces in \mathcal{F}_h^b . We additionally set $\mathcal{F}_h \stackrel{\text{def}}{=} \mathcal{F}_h^i \cup \mathcal{F}_h^b$ and, for all $T \in \mathcal{T}_h$, $\mathcal{F}_T \stackrel{\text{def}}{=} \{F \in \mathcal{F}_h \mid F \subset \partial T\}$; conversely, we let $\mathcal{T}_F \stackrel{\text{def}}{=} \{T \in \mathcal{T}_h \mid F \subset \partial T\}$. The set of cells and faces sharing one vertex $N \in \mathcal{N}_h$ are respectively denoted by \mathcal{T}_N and \mathcal{F}_N . For each $T \in \mathcal{T}_h$ we assume that there exists $\mathbf{x}_T \in T$ (the *cell-center*) such that T is star-shaped with respect to \mathbf{x}_T . For all $F \in \mathcal{F}_T$ we let $d_{T,F} \stackrel{\text{def}}{=} \operatorname{dist}(\mathbf{x}_T, F)$, and we assume that $d_{T,F}$ is comparable to the face diameter h_F . For all $F \in \mathcal{F}_T$, $\mathcal{P}_{T,F}$ denotes the F-based pyramid of apex \mathbf{x}_T , i.e.

$$\forall T \in \mathcal{T}_h, \ \forall F \in \mathcal{F}_T, \quad \mathcal{P}_{T,F} \stackrel{\text{der}}{=} \big\{ \mathbf{x} \in T \mid \exists \mathbf{y} \in F, \ \exists \theta \in [0, 1] \text{ such that } \mathbf{x} = \theta \mathbf{x}_T + (1 - \theta) \mathbf{y} \big\}.$$

The subgrid obtained from the decomposition of the elements of \mathcal{T}_h into pyramids is denoted by \mathcal{S}_h , i.e. $\mathcal{S}_h = \{\mathcal{P}_{T,F}\}_{T \in \mathcal{T}_h, F \in \mathcal{F}_T}$ (notation exemplified in Fig. 1(a)). The set of lateral pyramid faces (*subfaces*) included in *T* is denoted by \mathcal{S}_T . For all $S \in \mathcal{S}_T$, $T \in \mathcal{T}_h$, we fix a unit normal \mathbf{n}_S and let, for all $\varphi \in H^1(\mathcal{S}_h) \stackrel{\text{def}}{=} \{v \in L^2(\Omega) \mid \forall T \in \mathcal{T}_h, \forall F \in \mathcal{F}_T, v \mid_{\mathcal{P}_{T,F}} \in H^1(\mathcal{P}_{T,F})\}$,

$$\llbracket \varphi \rrbracket \stackrel{\text{def}}{=} \varphi^{-} - \varphi^{+}, \qquad \{\varphi\} \stackrel{\text{def}}{=} \frac{\varphi^{-} + \varphi^{+}}{2}, \qquad \varphi^{\pm}(\mathbf{x}) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0^{+}} \varphi(\mathbf{x} \pm \epsilon \mathbf{n}_{S}) \quad \text{for a.e. } \mathbf{x} \in S.$$

The diameter of a subface $S \in S_T$, $T \in T_h$, is denoted by h_S . We introduce the following notation for the diffusion in the normal direction on faces and subfaces: For all $T \in T_h$,

for all $F \in \mathcal{F}_T$, $\lambda_F^T \stackrel{\text{def}}{=} \kappa|_T \mathbf{n}_F \cdot \mathbf{n}_F$ and for all $S \in \mathcal{S}_T$, $\lambda_S \stackrel{\text{def}}{=} \kappa|_T \mathbf{n}_S \cdot \mathbf{n}_S$.

For boundary faces we simply write λ_F to alleviate the notation. Finally, the space of the cell-centered unknowns is denoted by $\mathbb{R}^{\mathcal{T}_h}$, and we use the notation $\mathbf{v}_h = (v_T)_{T \in \mathcal{T}_h} \in \mathbb{R}^{\mathcal{T}_h}$.

1.2. The L-reconstruction

The L-reconstruction of [1] is the main ingredient to design the incomplete polynomial space. We recall here the basics and refer to $[2, \S3.1]$ for a comprehensive presentation. Define the set of *L*-groups as in [2, Definition 3.1],

$$\mathcal{G} \stackrel{\text{def}}{=} \{ \mathfrak{g} \subset \mathcal{F}_T \cap \mathcal{F}_N \mid T \in \mathcal{T}_N, \ N \in \mathcal{N}_h, \ \text{card}(\mathfrak{g}) = d \}.$$

For every $\mathfrak{g} \in \mathcal{G}$, we define the set of L-group cells $\mathcal{T}_{\mathfrak{g}} \stackrel{\text{def}}{=} \bigcup_{F \in \mathfrak{g}} \mathcal{T}_F$, among which we identify a cell $T_{\mathfrak{g}}$ such that $\mathfrak{g} \subset \mathcal{F}_{T_{\mathfrak{g}}}$ ($T_{\mathfrak{g}}$ may not be unique when concave elements are present); for all $F \in \mathfrak{g} \cap \mathcal{F}_h^i$, we denote by T_F the unique cell such that $F \subset \partial T_{\mathfrak{g}} \cap \partial T_F$ (the notation is exemplified in Fig. 1(b)). For a given $\mathbf{v}_h \in \mathbb{R}^{\mathcal{T}_h}$ we construct the function $\xi_{\mathbf{v}_h}^{\mathfrak{g}}$ such that

- (i) $\xi_{\mathbf{v}_h}^{\mathfrak{g}}$ is affine inside every $\{\mathcal{P}_{T,F}\}_{F \in \mathfrak{q}, T \in \mathcal{T}_F}$ and $\xi_{\mathbf{v}_h}^{\mathfrak{g}}(\mathbf{x}_T) = v_T$ for all $T \in \mathcal{T}_{\mathfrak{g}}$;
- (ii) both $\xi_{\mathbf{v}_h}^{\mathfrak{g}}$ and its diffusive flux are continuous across group interfaces, i.e., for all $F \in \mathfrak{g} \cap \mathcal{F}_h^i$ and every $\mathbf{x} \in F$,

$$\xi_{\mathbf{v}_{h}}^{\mathfrak{g}}\big|_{\mathcal{P}_{T_{\mathfrak{g},F}}}(\mathbf{x}) = \xi_{\mathbf{v}_{h}}^{\mathfrak{g}}\big|_{\mathcal{P}_{T_{F},F}}(\mathbf{x}) \quad \text{and} \quad \left(\boldsymbol{\kappa} \nabla \xi_{\mathbf{v}_{h}}^{\mathfrak{g}}\right)\big|_{\mathcal{P}_{T_{\mathfrak{g},F}}}(\mathbf{x}) \cdot \mathbf{n}_{F} = \left(\boldsymbol{\kappa} \nabla \xi_{\mathbf{v}_{h}}^{\mathfrak{g}}\right)\big|_{\mathcal{P}_{T_{F},F}}(\mathbf{x}) \cdot \mathbf{n}_{F}$$

(iii) for all $F \in \mathfrak{g} \cap \mathcal{F}_{h}^{b}$, $\xi_{\mathbf{v}_{h}}^{\mathfrak{g}}(\bar{\mathbf{x}}_{F}) = 0$ with $\bar{\mathbf{x}}_{F} \stackrel{\text{def}}{=} \int_{F} \mathbf{x}/|F|_{d-1}$.

It has been proved in [1] that $\xi_{\mathbf{v}_h}^{\mathfrak{g}}$ only depends on the values $(\mathbf{v}_T)_{T \in \mathcal{T}_{\mathfrak{g}}}$. In particular, it is inferred from [2, Lemma 3.1] that, for all $F \in \mathfrak{g}$ there holds $\xi_{\mathbf{v}_h}^{\mathfrak{g}}|_{\mathcal{P}_{T_{\mathfrak{g}},F}}(\mathbf{x}) = \mathbf{v}_{T_{\mathfrak{g}}} + \nabla \xi_{\mathbf{v}_h}^{\mathfrak{g}}|_{\mathcal{P}_{T_{\mathfrak{g}},F}} \cdot (\mathbf{x} - \mathbf{x}_{T_{\mathfrak{g}}})$ where $\nabla \xi_{\mathbf{v}_h}^{\mathfrak{g}}|_{\mathcal{P}_{T_{\mathfrak{g}},F}} = \mathbf{A}_{\mathfrak{g}}^{-1}\mathbf{b}(\mathbf{v}_h)$ with matrix $\mathbf{A}_{\mathfrak{g}} \in \mathbb{R}^{d,d}$ and linear *d*-valued application $\mathbf{b}_{\mathfrak{g}} : \mathbb{R}^{T_h} \to \mathbb{R}^d$ given by

$$\mathbf{A}_{\mathfrak{g}} \stackrel{\text{def}}{=} \begin{bmatrix} (\lambda_{F}^{T_{F}} d_{T_{F},F}^{-1} (\mathbf{x}_{T_{F}} - \mathbf{x}_{T_{\mathfrak{g}}}) + (\kappa_{T_{\mathfrak{g}}} - \kappa_{T_{F}}) \mathbf{n}_{T_{\mathfrak{g}},F})_{F \in \mathfrak{g} \cap \mathcal{F}_{h}^{i}}^{t} \\ (\lambda_{F} d_{T_{\mathfrak{g}},F}^{-1} (\mathbf{x}_{F} - \mathbf{x}_{T_{\mathfrak{g}}}))_{F \in \mathfrak{g} \cap \mathcal{F}_{h}^{i}}^{t} \end{bmatrix}, \qquad \mathbf{b}_{\mathfrak{g}} (\mathbf{v}_{h}) \stackrel{\text{def}}{=} \begin{bmatrix} (\lambda_{F}^{T_{F}} d_{T_{F},F}^{-1} (\mathbf{v}_{T_{F}} - \mathbf{v}_{T_{\mathfrak{g}}}))_{F \in \mathfrak{g} \cap \mathcal{F}_{h}^{i}}^{t} \\ (-\lambda_{F} d_{T_{\mathfrak{g}},F}^{-1} \mathbf{v}_{T_{\mathfrak{g}}})_{F \in \mathfrak{g} \cap \mathcal{F}_{h}^{i}}^{t} \end{bmatrix}.$$

(The cell and face centers and the normals are intended as column vectors, and $\mathbf{A}_{\mathfrak{g}}$ is defined by row blocks; $\mathbf{n}_{T_{\mathfrak{g}},F}$ denotes the normal to *F* pointing out of $T_{\mathfrak{g}}$.) When d = 1, $\nabla \xi_{\mathbf{v}_h}^{\mathfrak{g}}|_{\mathcal{P}_{T_{\mathfrak{g}},F}}$ coincides with the classical two-point gradient. Sufficient conditions for the inversibility of $\mathbf{A}_{\mathfrak{g}}$ are proposed in [2]; in particular, whenever $\boldsymbol{\kappa}$ is homogeneous, a sufficient condition is that the cell-centers $(\mathbf{x}_T)_{T \in \mathcal{T}_{\mathfrak{g}}}$ form a non-degenerate simplex. Finally, for all $F \in \mathfrak{g} \cap \mathcal{F}_h^i$, an explicit expression for $\xi_{\mathbf{v}_h}^{\mathfrak{g}}|_{\mathcal{P}_{T_F,F}}$ can also be derived (see again [2, Lemma 3.1]).

1.3. The discrete functional space

1 0

The procedure outlined in the previous section allows to reconstruct a piecewise affine field on S_h that is continuous across interfaces and has continuous diffusive fluxes. Let $\mathbb{P}_d^k(S_h)$ denote the space of broken polynomial functions on S_h of total degree at most k. We *assume* that the sets $\mathcal{G}_F \stackrel{\text{def}}{=} \{ \mathfrak{g} \in \mathcal{G} \mid F \subset \mathfrak{g} \text{ and } \mathbf{A}_\mathfrak{g} \text{ is invertible} \}$ are non-empty for all $F \in \mathcal{F}_h$. For each $F \in \mathcal{F}_h$, we select the L-group $\mathfrak{g}_F \in \mathcal{G}_F$ for which $\|\mathbf{A}_{\mathfrak{g}_F}^{-1}\|_{\infty}$ is the smallest (see [2, Eq. (B.9)]), and let $\mathcal{I}_h \in \mathcal{L}(\mathbb{R}^{\mathcal{T}_h}, \mathbb{P}_d^1(\mathcal{S}_h))$ realize the mapping $\mathbb{R}^{\mathcal{T}_h} \ni \mathbf{v}_h \mapsto \mathcal{I}_h(\mathbf{v}_h) \in \mathbb{P}_d^1(\mathcal{S}_h)$ with

$$\forall T \in \mathcal{T}_h, \ \forall F \in \mathcal{F}_T, \quad \mathcal{I}_h(\mathbf{v}_h)|_{\mathcal{P}_{T,F}} = \xi_{\mathbf{v}_h}^{\mathfrak{g}_F}|_{\mathcal{P}_{T,F}}$$

The injectivity of \mathcal{I}_h follows from the non-emptiness of the sets $\{\mathcal{G}_F\}_{F \in \mathcal{F}_h}$ together with the fact that every $T \in \mathcal{T}_h$ appears in at least one local reconstruction. Drawing on the ideas of [4], we set

$$V_h \stackrel{\text{def}}{=} \mathcal{I}_h(\mathbb{R}^{\mathcal{T}_h}) \subset \mathbb{P}^1_d(\mathcal{S}_h).$$
(3)

A bijective operator can be obtained restricting the co-domain of \mathcal{I}_h to V_h . Observe that the functions in V_h are continuous across mesh interfaces but they can jump across subfaces, as depicted in Fig. 1(c). Choose $\{\varphi_T \stackrel{\text{def}}{=} \mathcal{I}_h(\mathbf{e}_T)\}_{T \in \mathcal{I}_h}$ as basis for V_h (\mathbf{e}_T being the *T*th vector of the canonical basis of $\mathbb{R}^{\mathcal{I}_h}$). The reduced stencil with respect to the original ccG method of [4] follows from the more compact support of the basis functions, which reduces cell interactions. The actual support of each basis function φ_T varies according to the specific choices for $\{\mathfrak{g}_F\}_{F \in \mathcal{F}_T}$, but we can identify minimal and maximal supports (see Fig. 1(c)). Let

$$V_{\dagger} \stackrel{\text{def}}{=} \left\{ \boldsymbol{\nu} \in \boldsymbol{V} \cap H^{2}(\mathcal{S}_{h}) \mid \forall F \subset \partial T_{1} \cap \partial T_{2}, \ (\boldsymbol{\kappa} \nabla \boldsymbol{\nu}) \Big|_{T_{1}} \cdot \mathbf{n}_{F} = (\boldsymbol{\kappa} \nabla \boldsymbol{\nu}) \Big|_{T_{2}} \cdot \mathbf{n}_{F} \right\}$$



(a) Minimal (dark) and maximal (light+dark) support for $\varphi = \mathcal{I}_h(\mathbf{e}_T)$ (cell *T* in bold line). *Left.* Proposed method *Right*. Method of [4]





(b) Worst-case interacting cells for maximally supported φ_{T_1} and φ_{T_2} . Left. Proposed method Right. Method of [4]

Fig. 2. Stencil reduction related to the more compact support of basis functions with respect to the method of [4]. (a) T_h = quadrangular 3 × 3 mesh. (b) T_h = quadrangular 7 × 7 mesh, supports in shades of gray.

and set $V_{\dagger h} \stackrel{\text{def}}{=} V_{\dagger} + V_h$ (the additional local regularity in V_{\dagger} is required to *extend* the discrete bilinear form on a space containing the exact solution). As common in dG methods, we denote by ∇_h the broken gradient operator on S_h . The discrete bilinear form is defined on $V_{\dagger h} \times V_h$ as follows:

$$a_{h}(w, v_{h}) \stackrel{\text{def}}{=} \int_{\Omega} \kappa \nabla_{h} w \cdot \nabla_{h} v_{h} - \sum_{F \in \mathcal{F}_{h}^{b}} \int_{F} \left[(\kappa \nabla w \cdot \mathbf{n}_{F}) v_{h} + (\kappa \nabla v_{h} \cdot \mathbf{n}_{F}) w \right] + \sum_{F \in \mathcal{F}_{h}^{b}} \int_{F} \frac{\eta \lambda_{F}}{h_{F}} w v_{h} \\ - \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} \int_{S} \left(\kappa \{\nabla_{h} w\} \cdot \mathbf{n}_{S} \llbracket v_{h} \rrbracket + \kappa \{\nabla_{h} v_{h}\} \cdot \mathbf{n}_{S} \llbracket w \rrbracket \right) + \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} \int_{S} \frac{\eta \lambda_{S}}{h_{S}} \llbracket w \rrbracket \llbracket v_{h} \rrbracket.$$

$$(4)$$

The last term penalizes the jumps across subfaces; it is unnecessary to penalize the jumps across the faces of T_h since the functions in V_h are continuous across every interface (see Fig. 1(c)). Owing to the subface and boundary terms, two basis functions whose support shares at least one subface interact, thereby contributing to the stencil of the method (see Fig. 2(b)). All integrals in the expression of a_h can be evaluated using the center of mass as a quadrature node. The discrete problem reads

Find
$$u_h \in V_h$$
 such that $a_h(u_h, v_h) = \int_{\Omega} f v_h$ for all $v_h \in V_h$. (5)

1.4. Basic error estimate

We now identify some relevant properties of the bilinear form a_h . To prove coercivity, we equip the space V_h with the energy norm

$$\|\|\boldsymbol{v}_{h}\|\|^{2} \stackrel{\text{def}}{=} \|\boldsymbol{\kappa}^{1/2} \nabla_{h} \boldsymbol{v}_{h}\|_{[L^{2}(\Omega)]^{d}}^{2} + \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} \frac{\lambda_{S}}{h_{S}} \|[\![\boldsymbol{v}_{h}]\!]\|_{L^{2}(S)}^{2} + \sum_{F \in \mathcal{F}_{h}^{h}} \frac{\lambda_{F}}{h_{F}} \|\boldsymbol{v}_{h}\|_{L^{2}(F)}^{2}$$

To prove the boundedness of a_h , we introduce the following extended norm on $V_{\dagger h}$: For all $v \in V_{\dagger h}$,

$$|||v|||_{*}^{2} \stackrel{\text{def}}{=} |||v|||^{2} + \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} h_{S} \|\boldsymbol{\kappa}^{1/2} \{\nabla_{h} v\} \cdot \mathbf{n}_{S}\|_{L^{2}(S)}^{2} + \sum_{F \in \mathcal{F}_{h}^{b}} h_{F} \|\boldsymbol{\kappa}^{1/2} \nabla v \cdot \mathbf{n}_{F}\|_{L^{2}(F)}^{2}$$

The $\|\cdot\|$ -norm and the $\|\cdot\|_*$ -norm are uniformly equivalent on V_h . The bilinear form a_h enjoys the following properties:

- (i) Consistency. Let u solve (2) and additionally assume that $u \in V_{\dagger}$. Then, for all $v_h \in V_h$, $a_h(u, v_h) = \int_{\Omega} f v_h$.
- (ii) *Coercivity*. For all $\eta > \underline{\eta}$ (with $\underline{\eta}$ only depending on the mesh regularity), there exists $C_{\text{sta}} > 0$ independent of both the meshsize *h* and of the diffusion field κ such that for all $v_h \in V_h$, $a_h(v_h, v_h) \ge C_{\text{sta}} |||v_h|||^2$.
- (iii) Boundedness. There exists C_{bnd} independent of both the meshsize *h* and of the diffusion field κ such that, for all $(w, v_h) \in V_{\dagger h} \times V_h$, $a_h(w, v_h) \leq C_{\text{bnd}} |||w|||_* |||v_h|||$.

Using the above properties, the following error is classically inferred proceeding in the spirit of Céa's Lemma:

Theorem 1 (Error estimate). Let u solve (2) and additionally assume that $u \in V_{\dagger}$. Then,

$$|||u - u_h||| \leq (1 + C_{\text{bnd}}C_{\text{sta}}^{-1}) \inf_{w_h \in V_h} |||u - w_h|||_*.$$

Table 1

			U								
$card(T_h)$	$\ \ u-u_h\ \ _{L^2(\Omega)}$	Order	$\ \ u-u_h\ \ $	Order	Stencil	$card(T_h)$	$\ \ u-u_h\ \ _{L^2(\Omega)}$	Order	$ u-u_h $	Order	Stencil
Proposed method				Proposed	method						
3584	1.3027e-03	-	4.5746e-02	-	15.03	806	6.2131e-03	-	9.0381e-02	-	14.69
14336	3.1039e-04	2.07	2.2608e-02	1.02	15.16	3162	1.7417e-03	1.83	4.5942e-02	0.98	15.25
57 344	8.6435e-05	1.84	1.1581e-02	0.97	15.22	12632	5.7562e-04	1.60	2.2897e-02	1.00	15.38
229376	1.9971e-05	2.11	5.6241e-03	1.04	15.25	50 548	1.4492e-04	1.99	1.1341e-02	1.01	15.52
ccG method of [4]					ccG meth	nod of [4]					
3584	5.2227e-04	-	2.8501e-02	-	25.26	806	4.0093e-03	-	4.8250e-02	-	25.33
14336	1.2926e-04	2.01	1.4194e-02	1.01	25.73	3162	1.7803e-03	1.17	2.7231e-02	0.83	26.35
57 344	3.2545e-05	1.99	7.1126e-03	1.00	26.11	12632	4.9217e-04	1.85	1.3674e-02	0.99	26.85
229376	7.9803e-06	2.03	3.5293e-03	1.01	26.10	50 548	1.3945e-04	1.82	6.9776e-03	0.97	27.12

Comparison of the proposed method (top) with the original ccG method of [4] (bottom) on the heterogeneous test case (6) (left) and on the anisotropic test case (7) (right). The stencil is defined as the average number of nonzero element in a row.

The multiplicative constant in the right-hand side does not depend on the diffusion tensor, which renders the method robust with respect to both heterogeneity and anisotropy of κ . First order convergence rate can be inferred if u belongs to the space of [2, Lemma 3.2] or $u \in H^2(\Omega)$ in the homogeneous case.

1.5. Convergence to minimal regularity solutions

Convergence to solutions that are barely in *V* can also be proved. The $\|\|\cdot\|\|$ -norm satisfies a discrete Poincaré inequality on V_h (proceed as in [5, Theorem 6.1]). As a consequence, we have the uniform *a priori* estimate $\|\|u_h\|\| \leq \sigma_2 C_{\text{sta}}^{-1} \|f\|_{L^2(\Omega)}$ with σ_2 discrete Poincaré constant. In the spirit of [3], we define a discrete gradient that is weakly asymptotically consistent by introducing lifting operators. In particular, for all $T \in \mathcal{T}_h$, all $S \in \mathcal{S}_T$, $r_S : L^2(S) \to [\mathbb{P}^0_d(\mathcal{S}_h)]^d$ is such that, for all $\varphi \in L^2(S)$, $\int_{\Omega} r_S(\varphi) \cdot \tau_h = \int_S \varphi \{\tau_h\} \cdot \mathbf{n}_S$ for all $\tau_h \in [\mathbb{P}^0_d(\mathcal{S}_h)]^d$. Similarly, for all $F \in \mathcal{F}_h^b$ and all $\varphi \in L^2(F)$, $r_F : L^2(F) \to [\mathbb{P}^0_d(\mathcal{S}_h)]^d$ is such that $\int_{\Omega} r_F(\varphi) \cdot \tau_h = \int_F \varphi \tau_h \cdot \mathbf{n}_F$ for all $\tau_h \in [\mathbb{P}^0_d(\mathcal{S}_h)]^d$.

Lemma 2 (Discrete Rellich theorem). Let $(v_h)_{h \in \mathcal{H}}$ be a sequence in $(V_h)_{h \in \mathcal{H}}$ uniformly bounded in the $\|\|\cdot\|\|$ -norm and set, for all $v_h \in V_h$,

$$\mathfrak{G}_{h}(\nu_{h}) \stackrel{\text{def}}{=} \nabla_{h}\nu_{h} - \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} r_{S}(\llbracket \nu_{h} \rrbracket) - \sum_{F \in \mathcal{F}_{h}^{b}} r_{F}(\nu_{h}) \stackrel{\text{def}}{=} \nabla_{h}\nu_{h} - R_{h}(\nu_{h}).$$

Then, there exists $v_h \in H_0^1(\Omega)$ such that, up to a subsequence, $v_h \to v$ strongly in $L^2(\Omega)$ and $\mathfrak{G}_h(v_h) \rightharpoonup \nabla v$ weakly in $[L^2(\Omega)]^d$ (that is to say, \mathfrak{G}_h is weakly asymptotically consistent).

To gradient \mathfrak{G}_h is also strongly consistent for piecewise smooth functions with continuous diffusive fluxes belonging to the space of [2, Lemma 3.2]. The proof of Theorem 3 relies on the following equivalent form for a_h in $V_h \times V_h$:

$$a_{h}(u_{h}, v_{h}) = \int_{\Omega} \kappa \mathfrak{G}_{h}(u_{h}) \cdot \mathfrak{G}_{h}(v_{h}) + j_{h}(u_{h}, v_{h}),$$

where $j_{h}(u_{h}, v_{h}) = -\int_{\Omega} \kappa R_{h}(u_{h}) \cdot R_{h}(v_{h}) + \sum_{T \in \mathcal{T}_{h}} \sum_{S \in \mathcal{S}_{T}} \int_{S} \frac{\eta \lambda_{S}}{h_{S}} \llbracket u_{h} \rrbracket \llbracket v_{h} \rrbracket + \sum_{F \in \mathcal{F}_{h}^{b}} \int_{F} \frac{\eta \lambda_{F}}{h_{F}} u_{h} v_{h}.$

Theorem 3 (Convergence to minimal regularity solutions). Let $(u_h)_{h \in \mathcal{H}}$ denote the sequence of discrete solutions to (5) on the admissible mesh sequence $(\mathcal{T}_h)_{h \in \mathcal{H}}$, and let $u \in V$ solve (2). Then, $u_h \to u$ strongly in $L^2(\Omega)$ and $\nabla_h u_h \to \nabla u$ strongly in $[L^2(\Omega)]^d$.

2. Numerical validation

To assess the robustness with respect to the heterogeneity of κ , consider the following pseudo two-dimensional exact solution to (2) on the unit square domain $\Omega = (0, 1)^2$:

$$u = \begin{cases} -\frac{1}{2}x^2 + \frac{3+\epsilon}{4(1+\epsilon)}x & \text{if } x \leq \frac{1}{2}, \\ -\frac{1}{2\epsilon}x^2 + \frac{3+\epsilon}{4\epsilon(1+\epsilon)}x + \frac{\epsilon-1}{4\epsilon(1+\epsilon)} & \text{if } x > \frac{1}{2}. \end{cases} \qquad \kappa = \begin{cases} 1 & \text{if } x < \frac{1}{2}, \\ \epsilon & \text{if } x > \frac{1}{2}, \end{cases} \qquad f = 1, \end{cases}$$
(6)

with heterogeneity ratio $\epsilon = 10^{-3}$. To evaluate the method with respect to the anisotropy of κ , we consider the following exact solution on $\Omega = (0, 1)^2$:

$$\mu = \sin(\pi x)\sin(\pi y), \quad \kappa_{xx} = 1, \quad \kappa_{xy} = \kappa_{yx} = 0, \quad \kappa_{yy} = \epsilon,$$
(7)

with adapted right-hand side f and anisotropy ratio $\epsilon = 10^{-3}$. The results on non-structured triangular meshes are collected in Table 1. An inspection of the rightmost columns shows a significant stencil reduction in both cases for the proposed method. However, although both methods show the same formal order of convergence, a slight degradation of the precision is observed for the proposed method.

References

- I. Aavatsmark, G.T. Eigestad, B.T. Mallison, J.M. Nordbotten, A compact multipoint flux approximation method with improved robustness, Numer. Methods Partial Differential Equations 24 (5) (2008) 1329–1360, doi:10.1002/num.20320.
- [2] L. Agélas, D.A. Di Pietro, J. Droniou, The G method for heterogeneous anisotropic diffusion on general meshes, M2AN Math. Model. Numer. Anal. 44 (4) (2010) 597–625, doi:10.1051/m2an/2010021.
- [3] L. Agélas, D.A. Di Pietro, R. Eymard, R. Masson, An abstract analysis framework for nonconforming approximations of diffusion problems on general meshes, IJFV 7 (1) (2010) 1–29.
- [4] D.A. Di Pietro, Cell centered Galerkin methods, C. R. Math. Acad. Sci. Paris, Ser. I 348 (2010) 31-34, doi:10.1016/j.crma.2009.11.012.
- [5] D.A. Di Pietro, A. Ern, J.-L. Guermond, Discontinuous Galerkin methods for anisotropic semi-definite diffusion with advection, SIAM J. Numer. Anal. 46 (2) (2008) 805–831, doi:10.1137/060676106.
- [6] M.G. Edwards, Unstructured control-volume distributed full tensor finite volume schemes with flow based grids, Comput. Geosci. 6 (10) (2002) 433–452, doi:10.1023/A:1021243231313.
- [7] R. Eymard, Th. Gallouët, R. Herbin, Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces, IMA J. Numer. Anal. 30 (4) (2010), doi:10.1093/imanum/drn084.