Monotone corrections for generic cell-centered Finite Volume approximations of anisotropic diffusion equations

Clément Cancès, Mathieu Cathala, Christophe Le Potier

Monotone corrections for generic cell-centered Finite Volume approximations of anisotropic diffusion equations *

Clément Cancès† Mathieu Cathala‡ Christophe Le Potier§

05/02/2013

Abstract

We present a nonlinear technique to correct a general Finite Volume scheme for anisotropic diffusion problems, which provides a discrete maximum principle. We point out general properties satisfied by many Finite Volume schemes and prove the proposed corrections also preserve these properties. We then study two specific corrections proving, under numerical assumptions, that the corresponding approximate solutions converge to the continuous one as the size of the mesh tends to 0. Finally we present numerical results showing that these corrections suppress local minima produced by the original Finite Volume scheme.

Keywords. Finite Volume scheme, Diffusion equation, Anisotropy, Maximum principle, Nonlinear corrections, Convergence.

AMS subject classification. 65N08, 65N12, 35J05

1 Statement of the problem

Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$. We consider the following elliptic problem:

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

with:

- $f \in L^2(\Omega)$, the source term;
- $\bar{u}$ the radioactive element concentration;

*This work was supported by the ANR project VFSitCom and by the GNR MoMaS.

†LJLL - UPMC Paris 06, 4 place Jussieu, F–75005 Paris. email: cances@ann.jussieu.fr

‡Université Montpellier II, Institut de Mathématiques et de Modélisation de Montpellier, CC 051, Place Eugène Bataillon, F–34095 Montpellier. email: mathieu.cathala@math.univ-montp2.fr

§CEA-Saclay, DEN, DM2S, STMF, LMEC, F–91191 Gif-sur-Yvette. email: clepotier@cea.fr

- $D : \Omega \to \mathcal{M}_d(\mathbb{R})$, the permeability, a bounded measurable function such that $D(x)$ is symmetric for a.e. $x \in \Omega$ and that there exists $\lambda > 0$ satisfying $D(x)\xi \cdot \xi \geq \lambda |\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^d$.

The elliptic operator from this simple problem occurs in more complex models of flows in porous media for instance related to underground nuclear waste repository or petroleum engineering. These particular applications require to design robust approximation methods to solve (1), one criterion consisting in the respect of the physical bounds. This is crucial, for example, for diffusion terms in modeling two-phase flows in porous media [19] and for coupling transport equation with a chemical model.

However, it is well known that classical linear methods discretizing diffusion operators do not always satisfy maximum principle for distorted meshes or with high anisotropy ratio [12, 19]. That is the reason why the question of constructing numerical methods for (1) ensuring the approximate solution satisfies a discrete maximum principle has been investigated. In [5], a non-linear stabilization term is introduced to design a Galerkin approximation of the Laplacian, but heterogeneous anisotropic tensors are not considered. More recently, a few non-linear finite volume schemes have been proposed to discretize elliptic problems [7, 11, 13, 15, 18, 21, 20]. For theses methods, the authors obtained the desired properties and accurate results which are generally second order in space. Unfortunately, none of these methods can ensure that they are coercive without conditions on the geometry or on the anisotropy ratio.

Starting from any given cell-centered finite volume scheme, our goal, in the present work, is to elaborate, in the spirit of methods described in [16], a general approach to construct non-linear corrections providing a discrete maximum principle while retaining some main properties of the scheme, in particular coercivity and convergence toward the solution of (1) as the size of the mesh tends to zero. To do so, we proceed step by step, beginning with a general correction and then refining it by considering successively the required properties. The corrections we obtain give nonoscillating solutions and can be applied, for example, to the cell-centered finite volume schemes developed in [1, 4, 2, 8, 14, 17]. Let us notice that these new corrections are quite easy to implement since they conserve the data structure used for the original linear scheme that has been corrected.

It is also worth mentioning the recent contribution [6] where a closely related question is investigated, i.e. the convergence of nonlinearly corrected Finite Volume schemes towards the solution of the unidimensional heat equation.

The paper is organized as follows. In section 2 we state the abstract framework about numerical schemes focusing on both discrete maximum principle and convergence of the solution to the scheme. Section 2.1 defines a specific structure of schemes (the so-called *LMP structure*, cf. Definition 2.2), which yields a discrete version of the local maximum principle. Section 2.2 specifies some basic properties of a numerical scheme, namely conservation property, coercivity and consistency. Using this abstract framework, we address in section 3 the problem of correcting a generic convergent cell centered finite volume scheme in order to enforce the LMP structure. In section 3.1 we state the main assumptions that can be made on the generic original scheme to be corrected, and that we expect to keep on its corrected version. Section 3.2 then establishes sufficient conditions for the corrections to bring the desired structure

while retaining conservation property and coercivity. Section 3.3 is devoted to the convergence of the corrected scheme. In section 4, we detail two examples of non-linear corrections and we perform for both a theoretical study of the corrected scheme. The convergence proofs rely in both cases on numerical assumptions on the approximate solutions to these schemes. The numerical results we present in section 5 confirm these assumptions seems to be actually fulfilled even for strongly anisotropic permeabilities.

2 Basics for numerical schemes

We first present the assumptions on the discretization of Ω.

Definition 2.1. *An admissible mesh of Ω is given by $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ where:*

- *$\mathcal{M}$ is a family of non-empty open polygonal connected disjoint subsets of Ω (the* control volumes*) such that $\overline{\Omega} = \cup_{K\in\mathcal{M}} \overline{K}$.*
- *$\mathcal{E}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the* edges *of the mesh) such that, for all $\sigma \in \mathcal{E}$, there exists an affine hyperplane E of $\mathbb{R}^d$ and $K \in \mathcal{M}$ verifying: : $\sigma \subset \partial K \cap E$ and σ is a non-empty open convex subset of E. We assume that, for all $K \in \mathcal{M}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \cup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. We also assume that, for all $\sigma \in \mathcal{E}$, either $\sigma \subset \partial\Omega$ or $\sigma \subset \overline{K} \cap \overline{L}$ for some $(K, L) \in \mathcal{M} \times \mathcal{M}$.*
- *$\mathcal{P} = (x_K)_{K\in\mathcal{M}}$ is a family of points of Ω (the* cell centers*) such that, for all $K \in \mathcal{M}$, $x_K \in K$ and K is star-shaped with respect to x_K.*

Remark 2.1. *Notice that the elements of $\mathcal{E}_K$ may not be the real edges of the control volume K (a full edge can be cut into several edges of the discretization). Notice also that no hypothesis is made on the convexity of the control volumes, so that two neighboring control volumes can share multiple edges.*

We use the following notations. The measure of a control volume K is denoted by $|K|$ and the $(d-1)$-dimensional measure of an edge σ is denoted by $|\sigma|$. For all $K, L \in \mathcal{M}$, we let $K|L = \mathcal{E}_K \cap \mathcal{E}_L$ be the (possibly empty) set of common edges to K and L and we set $|K|L| = \sum_{\sigma\in K|L} |\sigma|$. We define the set of interior (resp. boundary) edges as $\mathcal{E}_{\rm int} = \{\sigma \in \mathcal{E} \ ; \ \sigma \not\subset \partial\Omega\}$ (resp. $\mathcal{E}_{\rm ext} = \{\sigma \in \mathcal{E} \ ; \ \sigma \subset \partial\Omega\}$). For all $K \in \mathcal{M}$, we denote by $\mathcal{N}_K$ the subset of $\mathcal{M}$ of the neighboring control volumes, i.e.

$$\mathcal{N}_K = \{L \in \mathcal{M} \setminus \{K\} \text{ s.t. } K|L \neq \emptyset\}.$$

For all $x_K \in \mathcal{P}$, if $\sigma \in \mathcal{E}_K$, we denote by $d_{K,\sigma}$ the orthogonal distance between x_K and the hyperplane containing σ. For $\sigma \in \mathcal{E}$, we set $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$ if $\sigma \in K|L$ and $d_\sigma = d_{K,\sigma}$ if $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext}$.

To study the convergence of the schemes, we need the following two quantities: the size of the mesh

$$\text{size}(\mathcal{D}) = \sup_{K\in\mathcal{M}} \text{diam}(K)$$

and the regularity of the mesh

$$\mathrm{regul}(\mathcal{D}) = \sup_{\substack{K\in\mathcal{M}\\ \sigma\in\mathcal{E}_K}} \left\{ \frac{\mathrm{diam}(K)}{d_{K,\sigma}} \right\} + \sup_{\substack{K,L\in\mathcal{M}\\ \sigma\in\mathcal{E}_K\cap\mathcal{E}_L}} \left\{ \frac{d_{L,\sigma}}{d_{K,\sigma}} \right\}.$$

A cell-centered numerical scheme for (1) consists in a system of equations on some unknowns $(u_K)_{K\in\mathcal{M}}$ intended to approximate the values $(\bar{u}(x_K))_{K\in\mathcal{M}}$. More precisely it is given by a function

$$\begin{aligned} \mathcal{S}^{\mathcal{D}} : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} &\longrightarrow \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \\ u &\longmapsto (\mathcal{S}_K(u))_{K\in\mathcal{M}}, \end{aligned}$$

and consists in finding $u = (u_K)_{K\in\mathcal{M}}$ such that:

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = |K|\, f_K, \tag{2}$$

where f_K denotes the mean value of f on the cell K.

2.1 Local Maximum Principle structure

The main problem we address is to modify a cell-centered scheme in order to enforce the preservation of the maximum-principle. More precisely, using the terminology of [7], we focus on the following class of schemes.

Definition 2.2 (LMP structure)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω. A scheme $\mathcal{S}^{\mathcal{D}}$ for* (1) *has the Local Maximum Principle structure (LMP structure for short) if it can be written*

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = \sum_{L\in\mathcal{M}} \tau_{K,L}(u)(u_K - u_L) + \sum_{\sigma\in\mathcal{E}_{ext}} \tau_{K,\sigma}(u) u_K, \tag{3}$$

with functions $\tau_{K,L} : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \to \mathbb{R}_+$ (for $K, L \in \mathcal{M}$) and $\tau_{K,\sigma} : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \to \mathbb{R}_+$ (for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_{\mathrm{ext}}$) satisfying, for all $u \in \mathbb{R}^{\mathrm{Card}M}$,

$$\forall K \in \mathcal{M},\ \forall L \in \mathcal{N}_K, \quad \tau_{K,L}(u) > 0, \tag{4a}$$

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}, \quad \tau_{K,\sigma}(u) > 0. \tag{4b}$$

The schemes having the LMP structure meet a discrete version of the maximum principle as stated by the following property, whose proof is given in [7].

Proposition 2.1 (Discrete Maximum Principle)**.** *Assume that $f \geq 0$ on Ω. If $u = (u_K)_{K\in\mathcal{M}}$ is a solution to a scheme having the LMP structure, then* $\min_{K\in\mathcal{M}} u_K \geq 0$.

2.2 Convergent finite volume schemes

While correcting a scheme to provide it the LMP structure, we also want to preserve its main properties namely, on the one hand the Finite Volume structure and on the other hand the properties that lead to the convergence of its solution to the solution of the PDE (1). These properties are described in the following definitions.

2.2.1 Conservation property

Recall that a scheme for (1) is given, through (2), by a family $\mathcal{S}^{\mathcal{D}} = (\mathcal{S}_K)_{K\in\mathcal{M}}$ of functions $\mathcal{S}_K : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \to \mathbb{R}$ in the sense that, for $K \in \mathcal{M}$, the equation on the control volume K writes $\mathcal{S}_K(u) = |K|\, f_K$. We call conservative such a scheme if these equations can be written as a balance of approximate fluxes of the operator $\bar{u} \mapsto D\nabla\bar{u}$ in (1).

Definition 2.3 (Conservative scheme)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\mathcal{S}^{\mathcal{D}}$ define a scheme for* (1)*. $\mathcal{S}^{\mathcal{D}}$ is said to be conservative if there exists a family $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ of functions $F_{K,\sigma} : \mathbb{R}^{\mathrm{Card}\mathcal{M}} \to \mathbb{R}$ (the* numerical fluxes*) such that:*

$$\forall K \in \mathcal{M}, \forall L \in \mathcal{N}_K, \forall \sigma \in K|L, \quad F_{K,\sigma} + F_{L,\sigma} = 0, \tag{5}$$

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K = -\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}. \tag{6}$$

2.2.2 Coercivity

In order to estimate the solution of a scheme in a discrete version of the H_0^1 norm, it suffices for this scheme to fulfill some coercivity property, discrete analogue of the classical coercivity of the bilinear form that defines the variational formulation of (1).

To state this property we need to introduce some useful quantities. First we identify any element $u = (u_K)_{K\in\mathcal{M}}$ of $\mathbb{R}^{\mathrm{Card}\mathcal{M}}$ with the function u defined on Ω which is constant on each control volume of $\mathcal{M}$ and takes the value u_K on the cell $K \in \mathcal{M}$; we denote by $\mathcal{H}_{\mathcal{M}}$ the set of these functions. The space $\mathcal{H}_{\mathcal{M}}$ is then equipped with the discrete H_0^1 norm defined by:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad \|u\|_{\mathcal{D}}^2 = \sum_{\sigma\in\mathcal{E}} |\sigma| \frac{|u_K - u_L|^2}{d_\sigma}.$$

where, if $\sigma \in \mathcal{E}_{\mathrm{int}}$, K and L denote the cells on each side of σ and, if $\sigma \in \mathcal{E}_{\mathrm{ext}}$, K is the cell such that $\sigma \in \mathcal{E}_K$ and $u_L = 0$.

Definition 2.4 (Coercivity)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω. A scheme for* (1) *is coercive if there exists $\zeta > 0$ such that*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad \sum_{K\in\mathcal{M}} \mathcal{S}_K(u) u_K \geq \zeta \|u\|_{\mathcal{D}}^2. \tag{7}$$

The coercivity assumption allows to estimate a solution to a scheme in the discrete H_0^1 norm.

Proposition 2.2 (*a priori* estimate)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\mathcal{S}^{\mathcal{D}}$ define a coercive scheme for* (1) *with constant ζ in* (7)*. If $\theta \geq \mathrm{regul}(\mathcal{D})$, then there exists C_1 only depending on Ω, ζ and θ such that for any solution u to the scheme $\mathcal{S}^{\mathcal{D}}$*

$$\|u\|_{\mathcal{D}} \leq C_1 \|f\|_{L^2(\Omega)}. \tag{8}$$

Proof. For all $K \in \mathcal{M}$ we have $\mathcal{S}_K(u) = |K|\, f_K$. Multiplying this equality by u_K, summing over all the control volumes and using (7), we get

$$\zeta \|u\|_{\mathcal{D}}^2 \leq \int_\Omega fu. \tag{9}$$

Discrete Poincaré inequality (which can be deduced for instance from Lemma 5.3 of [9]) states that there exists C_2 only depending on Ω and θ such that

$$\|u\|_{L^2(\Omega)} \leq C_2 \, \|u\|_{\mathcal{D}}. \tag{10}$$

Inequality (8) then follows from (9) thanks to Cauchy-Schwarz inequality. □

2.2.3 Consistency

The discrete H_0^1 estimate that comes with a coercive scheme usually confers some compactness to its numerical solution, ensuring this solution converges to an element of $H_0^1(\Omega)$. In order to prove the latter is a weak solution to the problem (1), it then remains to ensure we can pass to the limit in the scheme.

Definition 2.5 (Consistency)**.** *Let $(\mathcal{D}^n)_{n\geq 1}$ be admissible meshes of Ω such that $\mathrm{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$. Let $(\mathcal{S}^n)_{n\geq 1}$ be such that, for all $n \geq 1$, $\mathcal{S}^n = (\mathcal{S}^n_K)_{K\in\mathcal{M}^n}$ is a scheme for (1) associated with discretization $\mathcal{D}^n = (\mathcal{M}^n, \mathcal{E}^n, \mathcal{P}^n)$. The family of schemes $(\mathcal{S}^n)_{n\geq 1}$ is consistent with (1) if, for any family $(u^n)_{n\geq 1}$ of discrete functions satisfying:*

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$,*
- *there exists $C_3 > 0$ such that, for all $n \geq 1$, $\|u^n\|_{\mathcal{D}^n} \leq C_3$,*
- *there exists $\bar{u} \in H_0^1(\Omega)$ such that $u^n \to \bar{u}$ in $L^2(\Omega)$ as $n \to \infty$,*

then

$$\forall \varphi \in \mathcal{C}_c^\infty(\Omega), \quad \lim_{n\to\infty} \sum_{K\in\mathcal{M}^n} \mathcal{S}^n_K(u^n)\varphi(x_K) = \int_\Omega D\nabla\bar{u}\cdot\nabla\varphi. \tag{11}$$

2.2.4 Cell-centered finite volume schemes for diffusion equations: a reader's digest

We present in the following table a brief review of which properties conservativity, coercivity and consistency are ensured by the schemes mentioned in the introduction.

Table 1: Cell-centered discretization schemes for anisotropic diffusion operators

Scheme	Conservativity	Coercivity	Consistency
MPFA O ([1, 3])	✓	✓[a]	✓[a]
Dioptre Scheme[b] ([2])	✓	✓	✓
Scheme from [8][c]	✓	✓	✓
SUSHI (barycentric form) ([9, 10])	not in the usual sense	✓	✓
VFSYM[d] ([14])	✓	✓	✓
Scheme from [17][d]	✓	✓	✓

[a] on parallelogram/parallelepiped
[b] with conditions on the meshes and the ratio anisotropy
[c] on admissible meshes
[d] on simplexes or parallelogram/parallelepiped

3 Non-linear corrections of a generic cell-centered finite volume scheme

Starting from a cell-centered scheme (for instance one from Table 1), we describe in this section how to construct a non-linear correction which gives the LMP structure while paying attention not to lose the main properties of the original scheme, namely conservativity, coercivity or consistency. We first state the main assumptions on the original scheme. Then we detail some general guidelines about the construction of such corrections.

3.1 The original scheme

Let us denote by A the continuous operator from problem (1) defined by $A(\bar{u}) = \mathrm{div}(D\nabla\bar{u})$.

In the following, we consider a generic discrete approximation $\mathcal{A}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ of the operator A. $\mathcal{A}^{\mathcal{D}}$ defines a scheme for (1) that writes

$$-\mathcal{A}^{\mathcal{D}}(u) = f_{\mathcal{D}}, \tag{12}$$

where we let $f_{\mathcal{D}} = (|K|\, f_K)_{K\in\mathcal{M}} \in \mathcal{H}_{\mathcal{M}}$. We assume that $\mathcal{A}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ is linear and invertible so that the original scheme (12) has a unique solution.

For the sake of clarity, it is convenient to introduce, for any $u \in \mathcal{H}_{\mathcal{M}}$, additional (trivial) values $(u_\sigma)_{\sigma\in\mathcal{E}_{\mathrm{ext}}}$ which we all take equal to zero. We denote by $V(K) \subset \mathcal{M} \cup \mathcal{E}_{\mathrm{ext}}$ the sets corresponding to the stencil of this scheme and we suppose the discrete linear operator $\mathcal{A}^{\mathcal{D}}$ writes in the following form[1]:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{A}_K(u) = \sum_{Z\in V(K)} \alpha_{K,Z}(u_Z - u_K) \tag{13}$$

(where with the previous convention $u_Z = 0$ if $Z = \sigma \in \mathcal{E}_{\mathrm{ext}}$). If need be by adding some null coefficients, we further suppose the stencil contains the neighboring cells (that is $V(K) \supset \mathcal{N}_K$) and that it is symmetric in the following sense:

$$\forall (K, L) \in \mathcal{M}^2, \quad L \in V(K) \implies K \in V(L). \tag{14}$$

In the following we address the problem of correcting this original scheme in order to provide it the LMP structure. Except from this property we want to reach, we focus on preserving any of the following additional properties:

(A1) The scheme defined by $\mathcal{A}^{\mathcal{D}}$ is conservative with numerical fluxes denoted by $F^{\mathcal{D}} = (F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$:

$$\forall K \in \mathcal{M}, \quad \mathcal{A}_K = \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}.$$

(A2) There exists $\zeta > 0$, independent of the mesh $\mathcal{D}$, such that the scheme defined by $\mathcal{A}^{\mathcal{D}}$ is coercive with constant ζ:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad -\sum_{K\in\mathcal{M}} \mathcal{A}_K(u)u_K \geq \zeta \,\|u\|_{\mathcal{D}}^2$$

[1] Using additionnal unknowns u_σ playing the role of approximation of $\bar{u}$ on the boundary edges, assuming (13) is nothing but assuming that the scheme is exact when applied to constant families: $\mathcal{A}^{\mathcal{D}}(u) = 0$ if $u = ((u_K)_{K\in\mathcal{M}}, (u_\sigma)_{\sigma\in\mathcal{E}_{\mathrm{ext}}}) = \text{constant}$.

(A3) Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes such that $\text{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$. Assume that $(\text{regul}(\mathcal{D}^n))_{n\geq 1}$ and $(\max_{K\in\mathcal{M}^n} \text{Card}\, V(K))_{n\geq 1}$ are bounded. Then the family of schemes defined by $(\mathcal{A}^{\mathcal{D}^n})_{n\geq 1}$ is consistent with problem (1).

3.2 General construction of non-linear corrections

Driven by the LMP structure, we consider corrections having the following form.

Definition 3.1 (Correction)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω. A correction for the scheme* (12) *defined by $\mathcal{A}^{\mathcal{D}}$ is a family $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ of functions $\beta_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}$. Given a correction β:*

- *the corrected scheme $\mathcal{S}^{\mathcal{D}}$ (from* (12)*) is defined by*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = -\mathcal{A}_K(u) + \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z), \tag{15}$$

- *the corrective term is the function $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ defined by*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{R}_K(u) = \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z). \tag{16}$$

3.2.1 Monotone corrections

The corrections defined above lead to a scheme having the LMP structure if they match the following conditions.

Proposition 3.1 (Monotone correction)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a correction for* (12)*. Let $(\gamma_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a family of functions $\gamma_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}_+$ such that, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,*

$$\text{if} \sum_{Z\in V(K)} |u_K - u_Z| \neq 0 \text{ then} \sum_{Z\in V(K)} \gamma_{K,Z}(u)\,|u_K - u_Z| = 1. \tag{17}$$

Assume that $\beta^{\mathcal{D}}$ satisfies, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,

$$\forall Z \in V(K), \quad \beta_{K,Z}(u) \geq \gamma_{K,Z}(u)\,|\mathcal{A}_K(u)|\,, \tag{18a}$$

$$\forall L \in \mathcal{N}_K, \quad \beta_{K,L}(u) > \gamma_{K,L}(u)\,|\mathcal{A}_K(u)|\,, \tag{18b}$$

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}, \quad \beta_{K,\sigma}(u) > \gamma_{K,\sigma}(u)\,|\mathcal{A}_K(u)|\,. \tag{18c}$$

Then the corrected scheme has the LMP structure.

Proof. Let $u \in \mathcal{H}_{\mathcal{M}}$. Using condition (17), the coordinate K of the original scheme (12) can be written

$$-\mathcal{A}_K(u) = -\sum_{Z\in V(K)} \gamma_{K,Z}(u)\,|u_K - u_Z|\,\mathcal{A}_K(u),$$

that is

$$-\mathcal{A}_K(u) = \sum_{Z\in V(K)} \{\gamma_{K,Z}(u)\text{sgn}(u_Z - u_K)\mathcal{A}_K(u)\}\,(u_K - u_Z). \tag{19}$$

Thus the coordinate K of the corrected scheme reads

$$\mathcal{S}_K(u) = \sum_{Z\in V(K)} \{\gamma_{K,Z}(u)\mathrm{sgn}(u_Z - u_K)\mathcal{A}_K(u) + \beta_{K,Z}(u)\}\,(u_K - u_Z). \quad (20)$$

Letting, for $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\tau_{K,Z}(u) = \gamma_{K,Z}(u)\mathrm{sgn}(u_Z - u_K)\mathcal{A}_K(u) + \beta_{K,Z}(u),$$

the corrected scheme takes the form of (3)

$$\mathcal{S}_K(u) = \sum_{Z\in V(K)} \tau_{K,Z}(u)(u_K - u_Z),$$

with $\tau_{K,Z} \geq 0$ according to (18a). To verify the corrected scheme has the LMP structure, it remains to check that these coefficients meet conditions (4), that is, they are positive whenever $Z \in \mathcal{N}_K$ or $Z = \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$. This last condition is precisely ensured by the strict inequalities (18b) and (18c). □

Remark 3.1. *Actually, the main condition we have to focus on when building a correction is condition* (18a)*. Indeed, assume a correction $\tilde{\beta}^{\mathcal{D}}$ matches condition* (18a)*, then, following the above calculus, we can see that the corresponding corrected scheme has the form of* (3) *with the non negative coefficients $\tau_{K,Z}$ given by*

$$\tau_{K,Z}(u) = \gamma_{K,Z}(u)\mathrm{sgn}(u_K - u_Z)\mathcal{A}_K(u) + \tilde{\beta}_{K,Z}(u).$$

Now, from $\tilde{\beta}^{\mathcal{D}}$, take positive numbers $(\nu_K)_{K\in\mathcal{M}}$ and define a new correction $\beta^{\mathcal{D}}$ by setting, for $u \in \mathcal{H}_{\mathcal{M}}$, $K \in \mathcal{M}$ and $Z \in V(K)$

$$\beta_{K,Z}(u) = \tilde{\beta}_{K,Z}(u) + \nu_K \frac{|K|Z|}{\mathrm{diam}(K)},$$

where we have extended the notation $K|Z$ to the edges $Z = \sigma \in V(K) \cap \mathcal{E}_{\mathrm{ext}}$ by setting $K|Z = \{\sigma\}$. Then the correction $(\beta^{\mathcal{D}})$ matches all the conditions of (18) *so that the scheme corrected with $\beta^{\mathcal{D}}$ has the LMP structure. Note that if we define a discrete Laplacian operator $\Delta^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ by*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \Delta_K(u) = \sum_{\sigma\in\mathcal{E}_K} \frac{|\sigma|}{\mathrm{diam}(K)}(u_L - u_K), \quad (21)$$

then using the correction $\beta^{\mathcal{D}}$ amounts to adding some numerical diffusion to the scheme corrected by $\tilde{\beta}^{\mathcal{D}}$. Indeed the scheme corrected with $\beta^{\mathcal{D}}$ writes, in terms of the correction $\tilde{\beta}^{\mathcal{D}}$, for $u \in \mathcal{H}_{\mathcal{M}}$ and $K \in \mathcal{M}$,

$$\mathcal{S}_K(u) = -\mathcal{A}_K(u) + \sum_{Z\in V(K)} \tilde{\beta}_{K,Z}(u)(u_K - u_Z) - \nu_K\Delta_K(u). \quad (22)$$

Then, provided that $\tilde{\beta}^{\mathcal{D}}$ has been chosen so that the corresponding corrected scheme is still consistent with the continuous operator A, we see that $\mathcal{S}_K(u)$ formally approximate the integral on K of $-A(u) - \nu_K\Delta u$. Taking ν_K to be of order $\mathrm{size}(\mathcal{D})$*, the effect of the last term can be expected*[2] *to vanish as* $\mathrm{size}(\mathcal{D}) \to 0$*. For instance, if $\nu_K = \mathrm{diam}(K)$, the correction turns to*

$$\beta_{K,Z}(u) = \tilde{\beta}_{K,Z}(u) + |K|Z|.$$

[2]This expectation is rigorously demonstrated in Remark 3.5.

Remark 3.2. *Conditions* (18) *ensures that the terms* $\beta_{K,Z}$ *are large enough to compensate the discrete maximum principle weakening contributions of* $-\mathcal{A}^{\mathcal{D}}$*, namely the coefficients in the right-hand side sum in* (19) *which correspond to elements* $Z \in V(K)$ *such that* $\mathcal{A}_K(u)(u_Z - u_K) < 0$*. Actually this condition entails that the compensation does not only happen on these weakening contribution but on all contributions. Thus the results remains true if we compensate only the weakening contributions. More precisely, setting*

$$V(K,u)^+ = \{Z \in V(K)\ ;\ \mathcal{A}_K(u)(u_Z - u_K) > 0\}$$

and

$$V(K,u)^- = \{Z \in V(K)\ ;\ \mathcal{A}_K(u)(u_Z - u_K) < 0\}\,,$$

we can take $\beta_{K,Z}(u) = 0$ *if* $Z \in V(K,u)^+$ *and change* (18) *into*

$$\begin{aligned}
&\forall Z \in V(K,u)^-, \quad \beta_{K,Z}(u) \geq \gamma_{K,Z}(u)\,|\mathcal{A}_K(u)|\,,\\
&\forall L \in \mathcal{N}_K \cap V(K,u)^-, \quad \beta_{K,L}(u) > \gamma_{K,L}(u)\,|\mathcal{A}_K(u)|\,,\\
&\forall \sigma \in \mathcal{E}_{\text{ext}} \cap V(K,u)^-, \quad \beta_{K,\sigma}(u) > \gamma_{K,\sigma}(u)\,|\mathcal{A}_K(u)|\,.
\end{aligned}$$

One first drawback of this choice is that it could lead to consider corrections that are not continuous functions of $u \in \mathcal{H}_{\mathcal{M}}$*, which is due to the fact the partition* $V(K) = V(K,u)^+ \cup V(K,u)^-$ *depends on* u*. Moreover, this would also break the symmetry of the correction that plays an important role as will be shown in Sections 3.2.2 and 3.2.3*

There are various ways to choose functions $\gamma_{K,Z}$ satisfying condition (17):

i) Taking, for $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\gamma_{K,Z}(u) = \frac{1}{\sum_{Y \in V(K)} |u_K - u_Y|} \tag{23}$$

if $\sum_{Y \in V(K)} |u_K - u_Y| \neq 0$ and $\gamma_{K,Z}(u) = 0$ else, condition (18a) writes

$$\beta_{K,Z}(u) \geq \frac{|\mathcal{A}_K(u)|}{\sum_{Y \in V(K)} |u_K - u_Y|}. \tag{24}$$

ii) For $u \in \mathcal{H}_{\mathcal{M}}$, let us define $V(K,u)^* = \{Z \in V(K)\ ;\ u_Z - u_K \neq 0\}$. Taking, for $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\gamma_{K,Z}(u) = \frac{1}{\text{Card}V(K,u)^*\,|u_Z - u_K|} \tag{25}$$

if $u_Z - u_K \neq 0$ and $\gamma_{K,Z}(u) = 0$ else, condition (18a) writes

$$\beta_{K,Z}(u) \geq \frac{|\mathcal{A}_K(u)|}{\text{Card}V(K,u)^*\,|u_Z - u_K|}. \tag{26}$$

3.2.2 Conservation preserving corrections

Even if the original scheme is a Finite Volume scheme in the sense that it matches assumption *(A1)*, this is not automatically the case of the corrected scheme. However a simple symmetry assumption on the correction ensures that the conservative structure is preserved.

The statement of this condition needs to introduce polygonal paths in the mesh as in [16]. Given an admissible mesh $\mathcal{D}$ of Ω we fix, for any pair $(I, J) \in \mathcal{M}^2$ such that $I \in V(J)$ (or equivalently $J \in V(I)$) a polygonal path IJ that does not include any edge or vertex of the mesh and that crosses any edge at most one time. Then, assuming the control volumes are sorted out, we denote by $\mathcal{C}$ the set $\mathcal{C} = \{IJ\ ;\ I \leq J\}$ and we let, for any edge $\sigma \in \mathcal{E}$, $\mathrm{ch}(\sigma)$ be the set of the polygonal paths IJ with $I \leq J$ and such that IJ crosses σ (see Figure 1). Finally, given a path $IJ \in \mathrm{ch}(\sigma)$ with $\sigma \in \mathcal{E}_K$, we set $\varepsilon_{K,\sigma,IJ} = 1$ if, from I

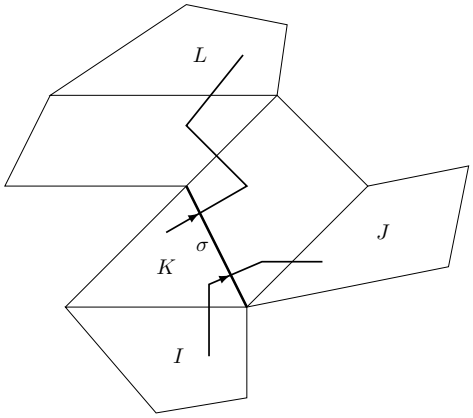

Figure 1: Paths IJ and KL in $\mathrm{ch}(\sigma)$, $J \in V(I)$, $L \in V(K)$.

to J, the path IJ enters the cell K through σ and $\varepsilon_{K,\sigma,IJ} = -1$ if it leaves K through σ.

Proposition 3.2 (Conservative corrections)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a correction for* (12)*. Assume the family $\beta^{\mathcal{D}}$ is symmetric:*

$$\forall K \in \mathcal{M}, \forall L \in V(K) \cap \mathcal{M}, \quad \beta_{K,L} = \beta_{L,K}. \tag{27}$$

If the original scheme is conservative, then so is the corrected one, with numerical fluxes $F'_{K,\sigma}$ given, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$, by

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{int}}, \quad F'_{K,\sigma}(u) = F_{K,\sigma}(u) - \sum_{IJ\in\mathrm{ch}(\sigma)} \varepsilon_{K,\sigma,IJ}\beta_{I,J}(u)(u_J - u_I) \tag{28a}$$

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}, \quad F'_{K,\sigma}(u) = F_{K,\sigma}(u) - \beta_{K,\sigma}(u)u_K \tag{28b}$$

Remark 3.3. *In case the correction $\beta^{\mathcal{D}}$ is symmetric (in the sense of* (27)*) the previous proposition states that correcting the original scheme with $\beta^{\mathcal{D}}$ amounts*

to correct the original fluxes $F_{K,\sigma}$ with the corrective fluxes $R_{K,\sigma}$ defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all interior edge $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm int}$ by

$$R_{K,\sigma}(u) = -\sum_{IJ \in {\rm ch}(\sigma)} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I), \tag{29}$$

and for all boundary edge $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext}$ by

$$R_{K,\sigma}(u) = -\beta_{K,\sigma}(u) u_K. \tag{30}$$

Proof of Proposition 3.2. We proceed as in the proof of Proposition 4.1 from [16]. Let us first remark that the corrective fluxes defined by (29) satisfy the conservativity condition (5) (this follows from the fact that, by definition, the quantity $\varepsilon_{K,\sigma,IJ}$ itself is conservative). Consequently the fluxes $F'_{K,\sigma}$ also satisfy this condition.

It remains to check that the corrective term $\mathcal{R}_K$ in (15) matches with the balance $-\sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma}$ of the corrective fluxes. On that account note that, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, if $IJ \in {\rm ch}(\sigma)$ is such that $K \notin \{I, J\}$ (*i.e.* the path crosses the cell K) and if IJ enters (resp. leaves) K across σ, then there exists $\sigma' \in \mathcal{E}_K$ such that IJ leaves (resp. enters) K across, this means $\varepsilon_{K,\sigma,IJ} = -\varepsilon_{K,\sigma',IJ}$. Thus, in the sum below, the terms corresponding to σ and σ' cancel so that we can state:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \sum_{\sigma \in \mathcal{E}_K} \sum_{\substack{IJ \in {\rm ch}(\sigma) \\ K \notin \{I,J\}}} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I) = 0.$$

Consequently, for any $u \in \mathcal{H}_{\mathcal{M}}$ and any $K \in \mathcal{M}$, the balance reduces to

$$-\sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{int}} R_{K,\sigma}(u) = \sum_{\sigma \in \mathcal{E}_K} \sum_{\substack{IJ \in {\rm ch}(\sigma) \\ K \in \{I,J\}}} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I)$$

which writes, in view of the definition of ${\rm ch}(\sigma)$ and $\varepsilon_{K,\sigma,IJ}$,

$$-\sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{int}} R_{K,\sigma}(u) = \sum_{L \in V(K) \cap \mathcal{M}} \beta_{K,L}(u)(u_K - u_L)$$

and then

$$-\sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma}(u) = \sum_{Z \in V(K)} \beta_{K,Z}(u)(u_K - u_Z) = \mathcal{R}_K(u).$$

□

3.2.3 Coercivity preserving corrections

If the correction is symmetric (in the sense of Proposition 3.2) it further suffices for the corrective functions to be non-negative to preserve the coercivity of the original scheme.

Proposition 3.3 (Coercivity preserving corrections)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K \in \mathcal{M}, Z \in V(K)}$ be a symmetric correction for* (12)*. Assume the family $\beta^{\mathcal{D}}$ is non-negative:*

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad \beta_{K,Z} \geq 0. \tag{31}$$

If the original scheme is coercive, then so is the corrected one, with the same constant.

Proof. Let $u \in \mathcal{H}_{\mathcal{M}}$. Assume the original scheme is coercive with constant ζ. Then

$$\sum_{K\in\mathcal{M}} \mathcal{S}_K(u)u_K \geq \zeta \left\| u \right\|_{\mathcal{D}}^2 + \sum_{K\in\mathcal{M}} u_K \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z).$$

Let us denote by $\mathcal{T}$ the last term of the inequality and remark that provided $\mathcal{T} \geq 0$, the coercivity of the original scheme is preserved. Now gathering by polygonal paths and using symmetry assumption (27) on $\beta^{\mathcal{D}}$ and assumption (14) on the stencil yields

$$\mathcal{T} = \sum_{IJ\in\mathcal{C}} \beta_{I,J}(u)(u_I - u_J)^2 + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}} \beta_{K,\sigma}(u)u_K^2$$

which proves, with (31), that $\mathcal{T} \geq 0$. □

Provided coefficients $\mathcal{R}_K$ of the corrective term are continuous functions of the unknown u, coercivity assumption also guaranties that there exists at least one solution to the corrected scheme.

Proposition 3.4 (Existence of a solution)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\beta^{\mathcal{D}}$ be a correction for* (12) *satisfying* (27) *and* (31)*. Assume that the original scheme is coercive and that the corrective term $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ is continuous. Then there exists one solution to the corrected scheme.*

Proof. The proof relies on Brower's topological degree. According to the hypothesis made on $\mathcal{R}^{\mathcal{D}}$, the application $h_t = -\mathcal{A}^{\mathcal{D}} + t\mathcal{R}^{\mathcal{D}}$ is continuous for all $t \in [0, 1]$. Then it is sufficient to show that, for R large enough, any solution to $h_t(u) = f_{\mathcal{D}}$ is bounded by R in $\mathcal{H}_{\mathcal{M}}$ to ensure that the degree of $h_1 = \mathcal{S}^{\mathcal{D}}$ on the ball of radius R at the point $f_{\mathcal{D}}$ is the same as the degree of $h_0 = -\mathcal{A}^{\mathcal{D}}$ which is not zero (since $\mathcal{A}^{\mathcal{D}}$ is invertible), and consequently to prove the existence of one solution to the corrected scheme $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$. The expected *a priori* estimate on the solution to $h_t(u) = f_{\mathcal{D}}$ is based on the coercivity of $-\mathcal{A}^{\mathcal{D}}$ and $\mathcal{S}^{\mathcal{D}}$. Indeed noting that $h_t = -(1-t)\mathcal{A}^{\mathcal{D}} + t\mathcal{S}^{\mathcal{D}}$ and denoting by ζ the coercivity constant of $-\mathcal{A}^{\mathcal{D}}$, Proposition 3.3 guarantees that the scheme defined by h_t is coercive with constant ζ. From Proposition 2.2 and the discrete Poincaré inequality (10) we get that any solution to $h_t(u) = f_{\mathcal{D}}$ is bounded in L^2 norm by $R = C_1C_2 \left\| f \right\|_{L^2(\Omega)}$. □

3.2.4 How to build monotone, conservative and coercive corrections

Assume the original scheme to be conservative and coercive. A simple way to construct corrections that match all the previous conditions ensuring the corrected scheme has the LMP structure and is still conservative and coercive is to take the following steps:

1. Choose a family $\gamma^{\mathcal{D}}$ such that (17) holds (for instance take $\gamma^{\mathcal{D}}$ as in (23) or (25));

2. Define the correction $b^{\mathcal{D}}$ by

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad b_{K,Z} = \gamma_{K,Z} \left|\mathcal{A}_K\right|. \tag{32}$$

This correction matches condition (18a)

3. (a) For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$, define $\tilde{\beta}_{K,\sigma} = b_{K,\sigma}$,
   (b) For $(K, L) \in \mathcal{M}^2$ such that $L \in V(K)$, define $\tilde{\beta}_{K,L}$ as a symmetric combination of $b_{K,L}$ and $b_{L,K}$ such that $\tilde{\beta}_{K,L} \geq b_{K,L}$. For instance one can take $\tilde{\beta}_{K,L} = b_{K,L} + b_{L,K}$ or $\tilde{\beta}_{K,L} = \max(b_{K,L}, b_{L,K})$.

   The correction $\tilde{\beta}^{\mathcal{D}}$ is thus symmetric, non-negative and satisfies condition (18a).

4. Augment $\tilde{\beta}^{\mathcal{D}}$ to match conditions (18b) and (18c): for instance define (see remark 3.1) $\beta^{\mathcal{D}}$ by

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad \beta_{K,Z} = \tilde{\beta}_{K,Z} + |K|Z|.$$

The correction $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K \in \mathcal{M}, Z \in V(K)}$ we obtain from these guidelines is thus symmetric, non-negative, hence it yields a scheme having the LMP structure.

As an example let us consider the following correction $\beta^{\mathcal{D}}$, similar to the non-linear correction proposed in [16], and defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K)$, by:

- If $Z = \sigma \in \mathcal{E}_{\text{ext}}$, then

$$\beta_{K,\sigma}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y \in V(K)} |u_Y - u_K|} + |\sigma|. \tag{33}$$

- If $Z = L \in \mathcal{M}$, then

$$\beta_{K,L}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y \in V(K)} |u_Y - u_K|} + \frac{|\mathcal{A}_L(u)|}{\sum_{Y \in V(L)} |u_Y - u_L|} + |K|L|. \tag{34}$$

If one of the quantities $\sum_{Y \in V(K)} |u_Y - u_K|$ or $\sum_{Y \in V(L)} |u_Y - u_L|$ is zero, we define $\beta_{K,Z}(u)$ in that case by dropping the corresponding term in (33) or (34). Note that each function $\beta_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}$ is continuous outside the set $\{u \in \mathcal{H}_{\mathcal{M}} \,;\, u_K - u_Z \neq 0\}$ and bounded on $\mathcal{H}_{\mathcal{M}}$ according to assumption (13) on the structure of the original scheme. Hence the corrective term $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ defined through (16) is continuous so that Proposition 3.4 guarantees the corresponding corrected scheme $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$ has at least one solution.

It has been proved in [16] that this correction gives a conservative and coercive scheme which has the LMP structure. This can also be shown by verifying this correction can be built following the guidelines 1–4 above. First, we consider the family $\gamma^{\mathcal{D}}$ given by (23) and then define, according to (32), correction $b^{\mathcal{D}}$ by:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \forall Z \in V(K), \quad b_{K,Z}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y \in V(K)} |u_Y - u_K|}.$$

We then follow steps 2 and 3 taking $\tilde{\beta}_{K,L} = b_{K,L} + b_{L,K}$ in 3b and we augment $\tilde{\beta}^{\mathcal{D}}$ according to step 4. Equation (34) finally writes $\beta_{K,L} = \tilde{\beta}_{K,L} + |K|L|$.

Starting from a different choice for the family $\gamma^{\mathcal{D}}$, namely the one previously defined by (25), the steps 1–4 can lead to the correction defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K,u)^*$, by

$$\beta_{K,Z}(u) = \Big\{ \max\left(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card} V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card} V(Z,u)^*} \right) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \Big\} \frac{1}{|u_K - u_Z|} \tag{35}$$

where we set $\frac{|\mathcal{A}_Z(u)|}{\mathrm{Card} V(Z,u)^*} = 0$ if $Z = \sigma \in \mathcal{E}_{\mathrm{ext}}$. The corresponding conservative and coercive corrected scheme $\mathcal{S}^{\mathcal{D}}$ which has the LMP structure writes, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,

$$\mathcal{S}_K(u) = -\mathcal{A}_K(u) + \sum_{Z \in V(K,u)^*} \Big\{ \max\Big(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card} V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card} V(Z,u)^*} \Big) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \Big\} \mathrm{sgn}(u_K - u_Z). \tag{36}$$

Note that the use of the terms $\mathrm{sgn}(u_K - u_Z)$ in this last correction is reminiscent of the form of the non-linear stabilization term proposed in [5] to design a Galerkin approximation of the Laplacian operator guaranteeing a discrete maximum principle on arbitrary meshes. The main drawback of the scheme (36) is that the corrective term is not continuous so that the existence of solutions to the non-linear system $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$ is not ensured. To obtain continuity we present in section 4.2 a regularized version of this scheme.

3.3 Convergence preserving corrections

Let us consider a coercive and consistent original scheme in the sense of assumptions *(A2)-(A3)*. From section 3.2, we know how to correct it in order to obtain a scheme which has the LMP structure and is still coercive. This last property ensures that the solution of such a corrected scheme still converges, up to a subsequence, to a function $\bar{u} \in H_0^1(\Omega)$. Moreover, from the consistency of the original scheme with problem (1), the behavior of the original part of the corrected scheme is known. Therefore, a simple way to prove that the limit $\bar{u}$ is a weak solution to the problem (1) is to make sure the corrective term vanishes as the size of the mesh tends to 0.

In addition to the geometrical regularity of the mesh, measured by the quantity $\mathrm{regul}(\mathcal{D})$, we want to take into account its compatibility with the original discretized operator $\mathcal{A}^{\mathcal{D}}$. To this end we first define the sets $\tilde{V}(K)$ by adding to $V(K)$ all the cells crossed by some polygonal path coming from K *i.e.* of the form KL ($L \in V(K)$). The sets $\tilde{V}(K)$ are then completed so that they are still symmetric that is:

$$\forall (K,L) \in \mathcal{M}^2, \quad L \in \tilde{V}(K) \implies K \in \tilde{V}(L).$$

Then we define the following quantity

$$\mathrm{reg}_{\mathcal{A}}(\mathcal{D}) = \mathrm{regul}(\mathcal{D}) + \max_{K \in \mathcal{M}, L \in \tilde{V}(K)} \frac{\mathrm{diam}(L)}{\mathrm{diam}(K)} + \max_{K \in \mathcal{M}} \mathrm{Card}(\tilde{V}(K)).$$

Proposition 3.5 (Convergence of the corrected scheme)**.** *Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that,* $\mathrm{size}(\mathcal{D}^n) \to 0$ *as* $n \to \infty$ *and* $(\mathrm{reg}_{\mathcal{A}}(\mathcal{D}^n))_{n\geq 1}$ *is bounded. Let $(\beta^n)_{n\geq 1}$ be a family of corrections associated with $(\mathcal{D}^n)_{n\geq 1}$ such that for all $n \geq 1$, β^n is symmetric and non-negative. For $n \geq 1$ we denote by $\mathcal{S}^n$ the corresponding corrected scheme.*

Assume that a family $(u^n)_{n\geq 1}$ satisfies:

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}}$ is a solution to $\mathcal{S}^n$;*
- *As $n \to \infty$,*

$$\sum_{K\in\mathcal{M}^n} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^n_{K,Z}(u^n)\,|u^n_K - u^n_Z| \to 0. \tag{37}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (1)*.*

Remark 3.4. *In the case where $V(K) \cap \mathcal{M}$ reduces to the neighboring cells $\mathcal{N}_K$ and where the paths in $\mathcal{C}$ cross only one edge, the family of corrective fluxes $R = (R_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ defined through* (29) *and* (30) *simply writes, for $\sigma \in K|L$,*

$$R_{K,\sigma}(u) = \beta_{K,\sigma}(u)(u_L - u_K).$$

Let us define, for a family of fluxes $F = (F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ and for a finite $p \geq 1$, discrete norms $N_{p,\mathcal{D}}(F)$ by

$$N_{p,\mathcal{D}}(F)^p = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} |\sigma|\,\mathrm{diam}(K) \left|\frac{F_{K,\sigma}}{|\sigma|}\right|^p.$$

We also define

$$N_{\infty,\mathcal{D}}(F) = \max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K} \left|\frac{F_{K,\sigma}}{|\sigma|}\right|.$$

Then condition (37) *reads*

$$N_{1,\mathcal{D}^n}(R(u^n)) \to 0 \text{ as } n \to \infty. \tag{38}$$

Notice that as a consequence of Hölder inequality, the following bound holds for any family of fluxes F and any $p \in [1,\infty]$:

$$N_{1,\mathcal{D}}(F) \leq (d\,|\Omega|\,\mathrm{regul}(\mathcal{D}))^{1-\frac{1}{p}}\, N_{p,\mathcal{D}}(F). \tag{39}$$

Thus, as $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ is bounded, condition (37) *holds if, for any $p > 1$, $N_{p,\mathcal{D}^n}(R(u^n)) \to 0$ as $n \to \infty$.*

Remark 3.5. *When choosing the augmentation in step 4 of the construction from section 3.2.4, one has to make sure this augmentation vanishes as* $\mathrm{size}(\mathcal{D}) \to 0$ *if not at the risk of jeopardizing the consistency of the scheme. In case this augmentation is conservative one may ensure that condition* (38) *holds. Note that this is the case for the above two corrections:*

- *Concerning* (33)-(34) *we know from Remark 3.1 that the additional numerical diffusion term writes, for all* $u \in \mathcal{H}_{\mathcal{M}}$ *and all* $K \in \mathcal{M}$,

$$\mathrm{diam}(K)\Delta_K(u) = \sum_{\sigma \in \mathcal{E}_K} r_{K,\sigma}(u),$$

with $r_{K,\sigma}(u) = |\sigma|(u_Z - u_K)$. *Now, remark that if* $\theta \geq \mathrm{regul}(\mathcal{D})$ *then we have*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad N_{2,\mathcal{D}}(r(u)) \leq C_4 \,\mathrm{size}(\mathcal{D})\, \|u\|_{\mathcal{D}},$$

with some $C_4 \in \mathbb{R}_+$ *only depending on* θ. *Provided both* $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ *and* $(\|u^n\|_{\mathcal{D}^n})_{n\geq 1}$ *are bounded, this entails that* $N_{2,\mathcal{D}^n}(r(u^n)) \to 0$ *as* $n \to \infty$.

Replacing $|K|L|$ *in* (34) *by the smaller quantity (used in section 4.1)*

$$\min\left(|K|L|\,, \frac{|K|}{\sum_{Y\in V(K)}|u_Y - u_K|} + \frac{|L|}{\sum_{Y\in V(L)}|u_Y - u_L|}\right)$$

and denoting by $\tilde{r}_{K,\sigma}$ *the corresponding fluxes, we have*

$$N_{\infty,\mathcal{D}}(\tilde{r}) \leq 2 \max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K} \frac{|K|}{|\sigma|}.$$

Assuming some reasonable regularity assumptions on the mesh, we can see that this last quantity scales as $\mathrm{size}(\mathcal{D})$ *so that this augmentation vanishes more strongly than the previous one.*

- *The augmentation chosen in* (35) *is conservative with fluxes* $\rho_{K,\sigma}$ *defined by* $\rho_{K,\sigma}(u) = |\sigma|\, d_\sigma \mathrm{sgn}(u_K - u_Z)$. *In that case,* (38) *follows from the following estimate:*

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad N_{\infty,\mathcal{D}}(\rho(u)) \leq 2\,\mathrm{size}(\mathcal{D}).$$

Proof. We proceed as mentioned above: we use first coercivity to extract a convergent subsequence of $(u^n)_{n\geq 1}$, then the consistency of the original scheme together with assumption (37) allow to pass to the limit in the corrected scheme.

Given $n \geq 1$, Proposition 3.3 shows that $\mathcal{S}^n$ is coercive with constant ζ and thus that the *a priori* estimate (8) holds for u^n. Since $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ is bounded and since ζ does not depend on n, this estimate proves that the sequence $(\|u^n\|_{\mathcal{D}^n})_{n\geq 1}$ is bounded. Thus, according to the discrete compactness results for bounded families in the discrete H_0^1 norm (see [9] lemmas 5.6 and 5.7 with $p = 2$), there exists $\bar{u} \in H_0^1(\Omega)$ such that, up to a subsequence, $u^n \to \bar{u}$ in $L^2(\Omega)$. Since (1) has a unique solution, if we prove that $\bar{u}$ is indeed this solution, then we get that the whole family $(u^n)_{n\geq 1}$ converges to $\bar{u}$ as $n \to \infty$.

To simplify the notations, we drop the index n and assume that $u = u^n$ converges to $\bar{u}$ as $\mathrm{size}(\mathcal{D}) \to 0$. Given $\varphi \in \mathcal{C}_c^\infty(\Omega)$ we set $\varphi_{\mathcal{D}} = (\varphi_K)_{K\in\mathcal{M}} \in \mathcal{H}_{\mathcal{M}}$ with $\varphi_K = \varphi(x_K)$. Multiplying the equation on K (15) by φ_K and summing over $K \in \mathcal{M}$ we get

$$-\sum_{K\in\mathcal{M}} \mathcal{A}_K(u)\varphi_K + \sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K = \int_\Omega f\varphi_{\mathcal{D}}. \tag{40}$$

The right-hand side tends to $\int_\Omega f\varphi$ as size$(\mathcal{D}) \to 0$. Besides, since $\mathrm{reg}_{\mathcal{A}}(\mathcal{D})$ is bounded, assumption *(A3)* on the consistency of the original scheme ensures that, along the extracted subfamily, we have

$$-\sum_{K\in\mathcal{M}} \mathcal{A}_K(u)\varphi_K \to \int_\Omega D\nabla\bar{u}\nabla\varphi,$$

as size$(\mathcal{D}) \to 0$.

Let us prove the corrected term in the left-hand side of (40) vanishes as size$(\mathcal{D}) \to 0$. Gathering by polygonal paths, we can write

$$\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K = \sum_{IJ\in\mathcal{C}} \beta_{I,J}(u)(u_I - u_J)(\varphi_I - \varphi_J) + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}} \beta_{K,\sigma}(u)u_K\varphi_K.$$

Hence

$$\left|\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K\right| \le \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)} \beta_{K,Z}(u)\,|u_K - u_Z|\,|\varphi_K - \varphi_Z|\,. \tag{41}$$

Now note that since φ is regular, compactly supported in Ω, and since $\mathrm{reg}_{\mathcal{A}}(\mathcal{D})$ is bounded, there exists C_5 not depending on $\mathcal{D}$ such that

$$|\varphi_K - \varphi_Z| \le C_5\,\mathrm{diam}(K)$$

for all $K \in \mathcal{M}$ and all $Z \in V(K)$. Using this last inequality in (41) proves, according to (37), that

$$\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K \to 0$$

as size$(\mathcal{D})$ goes to 0.

Sending size$(\mathcal{D}) \to 0$ in (40) (along the extracted subfamily) we finally get, for any $\varphi \in \mathcal{C}_c^\infty(\Omega)$,

$$\int_\Omega D\nabla\bar{u}\nabla\varphi = \int_\Omega f\varphi,$$

which proves, as announced, that $\bar{u}$ is the weak solution to (1). □

4 Examples of corrections

In this section, we assume that the original scheme is coercive and consistent in the sense of *(A2)-(A3)*. Using the tools from the previous section we study two actual examples of corrections. In both cases, we provide a numerical condition under which the convergence of the scheme is ensured.

4.1 A first correction

Given some parameter $\eta > 0$, we consider first the following correction $\beta^{\mathcal{D}}$ defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K)$, by:

- If $Z = \sigma \in \mathcal{E}_{\text{ext}}$, then

$$\beta_{K,\sigma}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)} |u_Y - u_K|} + \eta \min\left(|\sigma|, \frac{|K|}{\sum_{Y\in V(K)} |u_K - u_Y|} \right). \tag{42}$$

- If $Z = L \in \mathcal{M}$, then

$$\beta_{K,L}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)} |u_Y - u_K|} + \frac{|\mathcal{A}_L(u)|}{\sum_{Y\in V(L)} |u_Y - u_L|} + \eta \min\left(|K|L|, \frac{|K|}{\sum_{Y\in V(K)} |u_Y - u_K|} + \frac{|L|}{\sum_{Y\in V(L)} |u_Y - u_L|} \right). \tag{43}$$

This correction is slightly different from the one previously defined by (33)-(34). More precisely the difference lies in the last term that is in the augmentation chosen in step 4 of the guidelines from section 3.2.4. The modified augmentation chosen above still brings the LMP structure and takes better care of the convergence of the scheme since the stabilization term is smaller (see Remark 3.5).

Proposition 4.1. *Let $\eta > 0$ and let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that $\text{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$ and $(\text{reg}_{\mathcal{A}}(\mathcal{D}^n))_{n\geq 1}$ is bounded. For all $n \geq 1$ we denote by $\mathcal{S}^n : \mathcal{H}_{\mathcal{M}^n} \to \mathcal{H}_{\mathcal{M}^n}$ the corrected scheme defined through (42)–(43). Let $(u^n)_{n\geq 1}$ be a sequence of discrete functions satisfying:*

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$ is a solution to $\mathcal{S}^n$;*
- *As $n \to \infty$,*

$$\sup_{K\in\mathcal{M}^n} \left\{ \left| \mathcal{A}_K^{\mathcal{D}^n}(u^n) \right| \frac{\text{diam}(K)}{|K|} \right\} \to 0. \tag{44}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (1).

Proof. We show that the family of solutions $(u^n)_{n\geq 1}$ matches condition (37). For simplicity, we drop the index n. For all $K \in \mathcal{M}$ and all $Z \in V(K)$ we have

$$\beta_{K,Z}(u) |u_K - u_Z| \leq |\mathcal{A}_K(u)| + |\mathcal{A}_Z(u)| + \eta |K|Z| |u_K - u_Z|.$$

Thus, $\text{reg}_{\mathcal{A}}(\mathcal{D})$ being bounded, there exists C_6 independent of $\mathcal{D}$ such that

$$\sum_{K\in\mathcal{M}} \text{diam}(K) \sum_{Z\in V(K)} \beta_{K,Z}(u) |u_K - u_Z| \leq C_6 \sum_{K\in\mathcal{M}} \text{diam}(K) |\mathcal{A}_K(u)| + \eta N_{1,\mathcal{D}}(r(u)), \tag{45}$$

with $N_{1,\mathcal{D}}(r(u)) = \sum_{K\in\mathcal{M}} \text{diam}(K) \sum_{\sigma\in\mathcal{E}_K} |\sigma| |u_K - u_L|$. Remark 3.5 together with inequality (39) yield

$$N_{1,\mathcal{D}}(r(u)) \xrightarrow[\text{size}(\mathcal{D})\to 0]{} 0. \tag{46}$$

Besides, the first term of the right hand side in (45) can be bounded above as follows

$$\sum_{K\in\mathcal{M}} \text{diam}(K) |\mathcal{A}_K(u)| \leq |\Omega| \sup_{K\in\mathcal{M}} \left\{ |\mathcal{A}_K(u)| \frac{\text{diam}(K)}{|K|} \right\},$$

which, thanks to (44), implies

$$\sum_{K\in\mathcal{M}} \operatorname{diam}(K)\,|\mathcal{A}_K(u)| \xrightarrow[\operatorname{size}(\mathcal{D})\to 0]{} 0. \tag{47}$$

Substituting estimates (46) and (47) into (45) proves that, as $\operatorname{size}(\mathcal{D}) \to 0$,

$$\sum_{K\in\mathcal{M}} \operatorname{diam}(K) \sum_{Z\in V(K)} \beta_{K,Z}(u)\,|u_K - u_Z| \to 0,$$

which, according to Proposition 3.5, gives the desired result. □

4.2 A regularized correction

As we pointed out above, the main drawback of the correction defined by (35) is that the resulting scheme is not a continuous function of $u \in \mathcal{H}_\mathcal{M}$. Actually, discontinuity mainly comes from the family $\gamma^\mathcal{D}$ given by (25) which has been used to build the correction following the steps 1–4 from section 3.2.4. Given a positive parameter ε, let us replace $\gamma^\mathcal{D}$ by a smoothed family γ^ε which writes, for $u \in \mathcal{H}_\mathcal{M}$, $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\gamma^\varepsilon_{K,Z}(u) = \frac{1}{\operatorname{Card}_\varepsilon V(K,u)^*(|u_K - u_Z| + \varepsilon)}, \tag{48}$$

in which the smoothed version $\operatorname{Card}_\varepsilon V(K,u)^*$ of $\operatorname{Card} V(K,u)^*$ is defined, for $u \in \mathcal{H}_\mathcal{M}$ and $K \in \mathcal{M}$, by

$$\operatorname{Card}_\varepsilon V(K,u)^* = \sum_{Z\in V(K)} \frac{|u_K - u_Z|}{|u_K - u_Z| + \varepsilon}.$$

Note that this smoothed version of $\gamma^\mathcal{D}$ still matches the condition (17) of Propostion 3.1 so that, following the steps given in section 3.2.4, we can start from γ^ε to build a smoothed correction β^ε defined, for $u \in \mathcal{H}_\mathcal{M}$, $K \in \mathcal{M}$ and $Z \in V(K)$, by

$$\beta^\varepsilon_{K,Z}(u) = \max\left(\frac{|\mathcal{A}_K(u)|}{\operatorname{Card}_\varepsilon V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\operatorname{Card}_\varepsilon V(Z,u)^*}\right) \frac{1}{|u_K - u_Z| + \varepsilon} + \frac{\sum_{\sigma\in K|Z} |\sigma|\, d_\sigma}{|u_K - u_Z| + \varepsilon} \tag{49}$$

with the convention $\frac{|\mathcal{A}_Z(u)|}{\operatorname{Card}_\varepsilon V(Z,u)^*} = 0$ if $Z = \sigma \in \mathcal{E}_{\text{ext}}$.

The corresponding corrected scheme $\mathcal{S}^\varepsilon$ thus writes, for all $u \in \mathcal{H}_\mathcal{M}$ and all $K \in \mathcal{M}$,

$$\mathcal{S}^\varepsilon_K(u) = -\mathcal{A}_K(u) + \sum_{Z\in V(K)} \max\Big(\frac{|\mathcal{A}_K(u)|}{\operatorname{Card}_\varepsilon V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\operatorname{Card}_\varepsilon V(Z,u)^*}\Big)\operatorname{sgn}_\varepsilon(u_K - u_Z) + \sum_{\sigma\in\mathcal{E}_K} \rho^\varepsilon_{K,\sigma}(u), \tag{50}$$

where the real function $\mathrm{sgn}_\varepsilon : x \in \mathbb{R} \mapsto x/(|x| + \varepsilon)$ regularizes the function sgn and the additional corrective fluxes $\rho^\varepsilon_{K,\sigma}$ are defined, for $u \in \mathcal{H}_\mathcal{M}$, $K \in \mathcal{M}$, and $\sigma \in K|Z$ by $\rho^\varepsilon_{K,\sigma}(u) = |\sigma|\, d_\sigma \mathrm{sgn}_\varepsilon(u_K - u_Z)$. According to section 3.2.4 this scheme has the LMP structure, is coercive and Proposition 3.4 ensures it admits at least one solution. Moreover, if the original scheme is conservative, then this scheme is also.

Remark 4.1. *Considering a sequence (u^ε) of solutions to the regularized schemfffffes* (50) *and sending $\varepsilon \to 0$, one can expect to obtain a solution to the unregularized scheme defined by* (36)*. Indeed, thanks to the* a priori *estimate* (8)*, the sequence (u^ε) is bounded in the finite-dimensional space $\mathcal{H}_\mathcal{M}$ and then converges, up to a subsequence, to a discrete function $u \in \mathcal{H}_\mathcal{M}$. However, passing to the limit in* (50) *does not prove that u satisfies* (36)*. Actually, since the function* sgn *is not continuous at the origin, we can only conclude that, up to a subsequence, as $\varepsilon \to 0$*

$$\mathrm{sgn}_\varepsilon(u^\varepsilon_K - u^\varepsilon_Z) \to \begin{cases} \mathrm{sgn}(u_K - u_Z) & \text{if } u_Z \neq u_K \\ s_{K,Z} & \text{if } u_Z = u_K, \end{cases}$$

for some $s_{K,Z} \in [-1, 1]$. Then, as $\varepsilon \to 0$, $\mathrm{Card} V(K, u^\varepsilon)^* \to \Sigma(K)$ with

$$\Sigma(K) = \mathrm{Card} V(K, u)^* + \sum_{\substack{Z \in V(K) \\ u_Z = u_K}} |s_{K,Z}|\,.$$

Thus we can only conclude that u satisfies the limit scheme

$$-\mathcal{A}_K(u) + \sum_{Z \in V(K,u)^*} \Big\{ \max\left(\frac{|\mathcal{A}_K(u)|}{\Sigma(K)}, \frac{|\mathcal{A}_Z(u)|}{\Sigma(Z)} \right) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \Big\} \mathrm{sgn}(u_K - u_Z)$$
$$+ \sum_{\substack{Z \in V(K) \\ u_Z = u_K}} \Big\{ \max\left(\frac{|\mathcal{A}_K(u)|}{\Sigma(K)}, \frac{|\mathcal{A}_Z(u)|}{\Sigma(K)} \right) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \Big\} s_{K,Z} = |K|\, f_K,$$

which does not coincide with (36)*.*

In order to address the question of convergence for the scheme $\mathcal{S}^\varepsilon$, the proposition bellow gives an estimate on $\mathcal{A}^\mathcal{D}(u)$ if u is a solution to (50).

The statement of this proposition uses the sets $V(K,u)^+$ and $V(K,u)^-$ defined, as said before, by:

$$V(K,u)^+ = \{ Z \in V(K)\ ;\ \mathcal{A}_K(u)(u_Z - u_K) > 0 \}\,,$$
$$V(K,u)^- = \{ Z \in V(K)\ ;\ \mathcal{A}_K(u)(u_Z - u_K) < 0 \}\,.$$

Proposition 4.2. *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\theta \geq \mathrm{regul}(\mathcal{D})$ and $\varepsilon > 0$. Let u be a solution to $\mathcal{S}^\varepsilon$ and let $K_0 \in \mathcal{M}$ be such that*

$$\frac{|\mathcal{A}_{K_0}(u)|}{\mathrm{Card}_\varepsilon V(K_0, u)^*} = \max_{K \in \mathcal{M}} \frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K, u)^*}. \tag{51}$$

Assume that u satisfies:

$$\text{there exists } Z \in V(K_0, u)^+ \text{ such that } |u_{K_0} - u_Z| \geq \varepsilon. \tag{52}$$

Then there exists C_7 only depending on d and θ such that, for all $K \in \mathcal{M}$,

$$\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K, u)^*} \leq |K_0|\,|f_{K_0}| + C_7\,|K_0|\,. \tag{53}$$

Proof. It is sufficient to prove estimate (53) for $K = K_0$. Now the K_0 component of $\mathcal{S}^\varepsilon(u)$ reduces to

$$-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} \frac{|\mathcal{A}_{K_0}(u)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z) + \sum_{\sigma\in\mathcal{E}_{K_0}} \rho^\varepsilon_{K_0,\sigma}(u) = |K_0|\, f_{K_0}. \quad (54)$$

Summing separately on $V(K_0,u)^-$ and $V(K_0,u)^+$, we get

$$\begin{aligned}-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} \frac{|\mathcal{A}_{K_0}(u)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z) \\ = -\mathcal{A}_{K_0}(u)\Bigg(1 - \sum_{Z\in V(K_0,u)^-} \frac{|\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} \\ + \sum_{Z\in V(K_0,u)^+} \frac{|\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\Bigg).\end{aligned}$$

Since condition (17) for the family γ^ε can be written

$$\sum_{Z\in V(K_0,u)^-} \frac{|\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} + \sum_{Z\in V(K_0,u)^+} \frac{|\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} = 1,$$

we then have

$$\begin{aligned}-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} \frac{|\mathcal{A}_{K_0}(u)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z) \\ = \frac{-2\mathcal{A}_{K_0}(u)}{\mathrm{Card}_\varepsilon V(K_0,u)^*} \sum_{Z\in V(K_0,u)^+} |\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)|, \quad (55)\end{aligned}$$

Now since $|\mathrm{sgn}_\varepsilon(x)| \geq 1/2$ when $|x| \geq \varepsilon$, assumption (52) ensures that

$$\sum_{Z\in V(K_0,u)^+} |\mathrm{sgn}_\varepsilon(u_{K_0}-u_Z)| \geq 1/2.$$

Substituting (55) in (54), applying the triangle inequality and using this last bound lead to

$$\frac{|\mathcal{A}_{K_0}(u)|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} \leq |K_0|\,|f_{K_0}| + \sum_{\sigma\in\mathcal{E}_{K_0}} \left|\rho^\varepsilon_{K_0,\sigma}(u)\right|. \quad (56)$$

Finally remark that, for all $K \in \mathcal{M}$,

$$\sum_{\sigma\in\mathcal{E}_K} \left|\rho^\varepsilon_{K,\sigma}(u)\right| \leq \sum_{\sigma\in\mathcal{E}_K} |\sigma|\, d_\sigma \leq d(1+\theta)\,|K|. \quad (57)$$

Plugging this last inequality with $K = K_0$ into (56) gives the desired estimates. □

Adding some regularity assumption on the mesh, the following result states the convergence of the solution to the scheme $\mathcal{S}^\varepsilon$ provided this solution fulfills condition (52) above. In the following, for $u \in \mathcal{H}_\mathcal{M}$, we say that $K \in \mathcal{M}$ is a maximal cell for u if

$$\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K,u)^*} = \max_{L\in\mathcal{M}} \frac{|\mathcal{A}_L(u)|}{\mathrm{Card}_\varepsilon V(L,u)^*}. \tag{58}$$

Proposition 4.3. *Assume $f \in L^d(\Omega)$. Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that $\mathrm{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$ and $(\mathrm{reg}_\mathcal{A}(\mathcal{D}^n))_{n\geq 1}$ is bounded; assume that there exists $C_8 > 0$ verifying*

$$\forall n \geq 1, \forall K, L \in \mathcal{M}^n, \quad |K| \leq C_8 |L|, \tag{59}$$

$$\forall n \geq 1, \forall K \in \mathcal{M}^n, \quad \mathrm{diam}(K)^d \leq C_8 |K|. \tag{60}$$

Let $(\varepsilon_n)_{n\geq 1}$ be a sequence of positive real numbers and let $(u^n)_{n\geq 1}$ be a sequence of discrete functions satisfying:

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$ is a solution to the scheme $\mathcal{S}^{\varepsilon_n}$.*
- *For all $n \geq 1$, there exists a maximal cell $K_0^n \in \mathcal{M}^n$ for u^n for which*

$$\text{there exists } Z \in V(K_0^n, u^n)^+ \text{ such that } \left|u^n_{K_0^n} - u^n_Z\right| \geq \varepsilon_n. \tag{61}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (1).

Proof. We show that, thanks to assumption (61) made on $(u^n)_{n\geq 1}$, condition (37) of Proposition 3.5 is satisfied. For simplicity we drop the index n. From Proposition 4.2 and the triangle inequality, we know since $\mathrm{reg}_\mathcal{A}(\mathcal{D})$ is bounded that there exists a constant C_9 independent of $\mathcal{D}$ and ε such that, for all $K \in \mathcal{M}$,

$$\sum_{Z\in V(K)} \beta^\varepsilon_{K,Z}(u) |u_K - u_Z| \leq C_9 \int_{K_0} (|f|+1) + \sum_{\sigma\in\mathcal{E}_K} \left|\rho^\varepsilon_{K,\sigma}(u)\right|.$$

From Hölder inequality and assumption (59) we get

$$\sum_{Z\in V(K)} \beta^\varepsilon_{K,Z}(u) |u_K - u_Z| \leq C_{10} |K|^{\frac{d-1}{d}} \left(\int_{K_0} (|f|+1)^d\right)^{\frac{1}{d}} + \sum_{\sigma\in\mathcal{E}_K} \left|\rho^\varepsilon_{K,\sigma}(u)\right|,$$

with $C_{10} = \max\left(C_8^{\frac{d-1}{d}} C_9, C_9\right)$. Then, bounding $\mathrm{diam}(K)$ by $C_8^{\frac{1}{d}} |K|^{\frac{1}{d}}$, we get C_{11} that does not depend on $\mathcal{D}$ or ε such that

$$\sum_{K\in\mathcal{M}} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^\varepsilon_{K,Z}(u) |u_K - u_Z| \leq C_{11}\Big(\int_{K_0} (|f|+1)^d\Big)^{\frac{1}{d}} + N_{1,\mathcal{D}}(\rho^\varepsilon(u)). \tag{62}$$

Since $|f| + 1 \in L^d(\Omega)$, the first term of this right-hand side tends to 0 as $\mathrm{size}(\mathcal{D}) \to 0$. The norm comparison (39) shows that

$$N_{1,\mathcal{D}}(\rho^\varepsilon(u)) \leq C_{12} N_{\infty,\mathcal{D}}(\rho^\varepsilon(u)) \leq C_{13}\,\mathrm{size}(\mathcal{D})$$

(with constants depending neither on size($\mathcal{D}$) nor on ε). Therefore, as size($\mathcal{D}$) tends to 0,

$$\sum_{K\in\mathcal{M}} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^{\varepsilon}_{K,Z}(u)\,|u_K - u_Z| \to 0.$$

This guarantees we can apply Proposition 3.5 and conclude that $u \to \bar{u}$ in $L^2(\Omega)$ as size($\mathcal{D}$) $\to 0$. □

5 Numerical results

To deal with the nonlinear terms, we perform an iterative algorithm. Let us denote u^i the value of the solution where i is a fixed point iteration. We fix $u = u^i$ in $\beta_{K,Z}(u)$ in (15) and the iterative scheme can be written :

$$\forall K \in \mathcal{M}, \quad -\mathcal{A}_K(u^{i+1}) + \sum_{Z\in V(K)} \beta_{K,Z}(u^i)(u_K^{i+1} - u_Z^{i+1}) = |K| f_K.$$

We stop the algorithm when the criterion $\dfrac{\|u^{i+1} - u^i\|}{\|u^i\|} \leq 10^{-4}$ is satisfied. We start from the conservative and consistent original operator $\mathcal{A}^{\mathcal{D}}$ developed in [1]. Moreover, we use grids of squares of surface h^2 (h changing from $\frac{1}{8}$ to $\frac{1}{128}$) so that this scheme is also coercive (see Table 1). Some notations used to present the numerical results are given in Table 2.

Table 2: Notations.

h	size of the discretization
L^2 error	L^2 error of the computed solution with respect to the analytical solution
ratiol2	order of convergence, in L^2 norm, of the method
nit	number of iterations needed to compute the approximate solution of $\mathcal{S}$
Min. Val.	$\min\{u_K\ ;\ K \in \mathcal{M}\}$
Max. Val.	$\max\{u_K\ ;\ K \in \mathcal{M}\}$
$\lvert u_{K_0} - u_{Z^*}\rvert$	$\max\{\lvert u_{K_0} - u_Z\rvert\ ;\ Z \in V(K_0,u)^+\}$
$\frac{\lvert\mathcal{A}_{K^*}\rvert}{\lvert K^*\rvert}$	$\max\left\{\frac{\lvert\mathcal{A}_K\rvert}{\lvert K\rvert}\ ;\ K\ \in \mathcal{M}\right\}$

5.1 Stationary analytical solution

In order to numerically estimate the convergence of the scheme, let us consider the following elliptic problem:

$$\begin{cases} -\mathrm{div}(D\nabla\bar{u}) = f \text{ in } \Omega =]0,0.5[\times]0,0.5[\\ \bar{u}(x,y) = \sin(\pi x)\sin(\pi y) \text{ for } (x,y) \in \partial\Omega \end{cases} \tag{63}$$

with

$$D = \frac{1}{x^2+y^2}\begin{pmatrix} y^2+\alpha x^2 & -(1-\alpha)xy \\ -(1-\alpha)xy & x^2+\alpha y^2 \end{pmatrix}$$

and

$$\begin{cases} u_{\mathrm{ana}} = \sin(\pi x)\sin(\pi y), \\ f = -\mathrm{div}\, D\nabla u_{\mathrm{ana}}. \end{cases} \tag{64}$$

The parameter α is equal to 10^{-6} and the anisotropy ratio is equal to 10^6. We check that $f \geq 0$.
We show the results obtained in Table 3 with the scheme developed in [1] (S. 1), with the first correction (S. 2) and with the regularized correction (S. 3). For the scheme 2, we choose $\eta = 2$. For the scheme 3, we choose $\varepsilon = 4h^2$.

Table 3: Numerical results for (63) with the original scheme, the first correction and the regularized correction as a function of the discretization step.

h	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
L^2 error (S. 1)	5.21×10^{-1}	1.96×10^{-1}	7.14×10^{-2}	1.65×10^{-2}	2.14×10^{-3}
ratiol2 (S. 1)		1.41	1.46	2.11	2.95
Undershoots (S. 1)	12.5 %	10 %	5 %	2 %	1 %
Min. Val. (S. 1)	-2.9×10^{-1}	-2.4×10^{-1}	-1.4×10^{-1}	-5.26×10^{-2}	-1.33×10^{-2}
L^2 error (S. 2)	1.59×10^{-1}	8.98×10^{-2}	4.73×10^{-2}	2.47×10^{-3}	1.30×10^{-2}
ratiol2 (S. 2)		0.82	0.93	0.94	0.93
nit	7	11	13	13	13
$\frac{\lvert\mathcal{A}_{K^*}\rvert}{\lvert K^*\rvert}$	13.26	15.80	16.60	17.25	18.09
L^2 error (S. 3)	9.03×10^{-2}	4.27×10^{-2}	2.12×10^{-2}	1.00×10^{-2}	4.75×10^{-3}
ratiol2 (S. 3)		1.08	1.01	1.07	1.08
nit	15	17	18	18	15
$\lvert u_{K_0} - u_{Z^*}\rvert$	1.43×10^{-1}	3.62×10^{-2}	9.10×10^{-3}	2.28×10^{-3}	5.70×10^{-4}
ε	6.25×10^{-2}	1.56×10^{-2}	3.90×10^{-3}	9.77×10^{-4}	2.44×10^{-4}

It is clear that the original scheme is at least second order in space but we observe large oscillations. Concerning the scheme 2 and 3, they become first order in space but all oscillations disappear.
For the scheme 2, looking at the terms $\frac{|\mathcal{A}_{K^*}|}{|K^*|}$, the assumptions of Proposition 4.1 seem to hold in this case.
For the scheme 3, we also check the assumptions of Proposition 4.3. As we use squares, the grids satisfy clearly the inequalities (59)-(60). Moreover, looking at the terms $|u_{K_0} - u_{Z^*}|$, the inequalities (61) are verified for all the grids considered so that we may expect this inequality to hold with further refinement of the grid.

5.2 Stationary non analytical solution

In order to evaluate the respect of the discrete maximum principle, we now consider the problem:

$$\begin{cases} -\operatorname{div}(D\nabla\bar{u}) = f \text{ in } \Omega =]0, 0.5[\times]0, 0.5[\\ \bar{u} = 0 \text{ on } \partial\Omega \end{cases} \tag{65}$$

and

$$f(x, y) = \begin{cases} 10 \text{ if } (x, y) \in]0.25, 0.5[\times]0.25, 0.5[\\ 0 \text{ otherwise,} \end{cases} \tag{66}$$

where D is as before (see (63)). We also choose $\eta = 2$ and $\varepsilon = 4h^2$.
The Table 4 shows the minimum and the maximum values for the original scheme, the first correction and the regularized correction. It is interesting to observe that the oscillations can be quite large unless the grid is thin. Figure 3 shows that they can be numerous even on the thin grid. On the other hand, as

expected, no such oscillations appear with the modified schemes (Figure 2). For the two corrected schemes, the number of iterations seems to be bounded as a function of the discretization step when we refine the grid. Moreover, looking at the terms $\frac{|\mathcal{A}_{K^*}|}{|K^*|}$ and $|u_{K_0} - u_{Z^*}|$, the inequalities (44) and (61) are also satisfied for all the grids which signals a promising outlook for the convergence of the corrected schemes.

Table 4: Numerical results for (65) with the original scheme, the first correction and the regularized correction as a function of the discretization step.

h	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
Undershoots (S. 1)	37 %	28%	21 %	19 %	20%
Min. Val. (S. 1)	-4.62×10^{-2}	-3.91×10^{-2}	-1.08×10^{-2}	-1.09×10^{-2}	-4.71×10^{-3}
Max. Val. (S. 1)	2.97×10^{-1}	3.3×10^{-1}	3.5×10^{-1}	3.8×10^{-1}	4.1×10^{-1}
Min. Val. (S. 2)	2.38×10^{-3}	1.16×10^{-4}	8.75×10^{-7}	3.30×10^{-10}	1.82×10^{-15}
Max. Val. (S. 2)	9.41×10^{-2}	1.13×10^{-1}	1.16×10^{-1}	2.12×10^{-1}	2.62×10^{-1}
nit	8	11	13	19	20
$\frac{\lvert\mathcal{A}_{K^*}\rvert}{\lvert K^*\rvert}$	7.06	11.81	14.43	16.94	17.81
Min. Val. (S. 3)	1.12×10^{-3}	5.90×10^{-5}	1.55×10^{-6}	3.53×10^{-8}	7.95×10^{-10}
Max. Val. (S. 3)	1.21×10^{-1}	1.41×10^{-1}	1.95×10^{-1}	2.48×10^{-1}	2.92×10^{-1}
nit	8	13	16	20	21
$\lvert u_{K_0} - u_{Z^*}\rvert$	6.88×10^{-2}	2.17×10^{-2}	5.14×10^{-3}	1.25×10^{-3}	3.07×10^{-4}
ε	6.25×10^{-2}	1.56×10^{-2}	3.90×10^{-3}	9.77×10^{-4}	2.44×10^{-4}

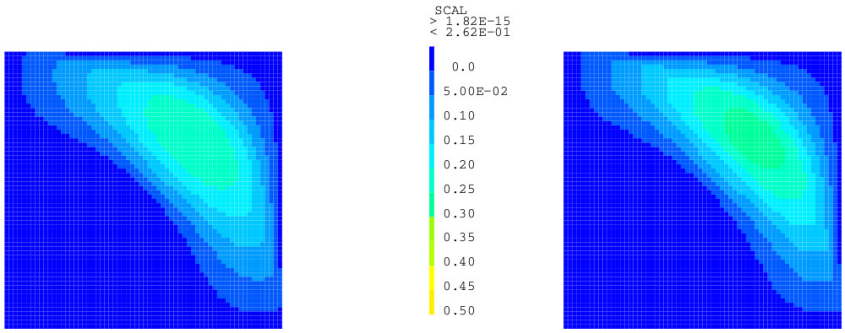

Figure 2: Concentration on a grid made of 4096 squares for the first correction (maximum value 0.26, minimum value 1.82×10^{-15}) and the regularized correction (maximum value 0.29, minimum value 7.95×10^{-10}).

Acknowledgments. The authors would like to thank Jérôme Droniou for precious advices and discussions.

References

[1] I. Aavatsmark I., T. Barkve T., O. Boe, T. Mannseth, *Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods*, Siam J. Sci. Comput., **19** (1998), no. 5, 1700–1716.

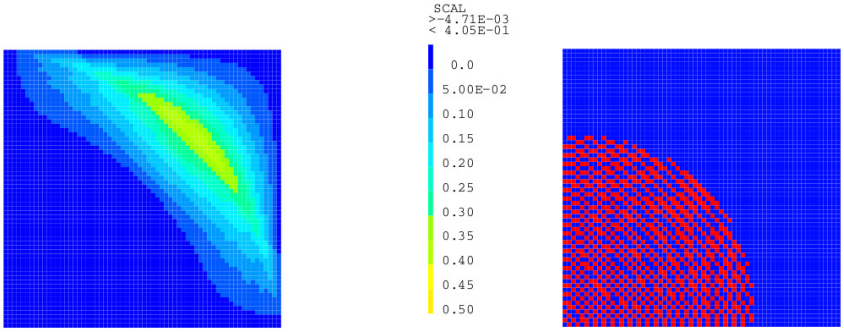

Figure 3: Concentration and position of the undershoots (in red) for the original scheme on a grid made of 4096 cells (maximum value 0.41, minimum value -4.71×10^{-3}).

[2] L. Agelas, R. Eymard, R. Herbin. *A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media*, C. R. Acad. Sci. Paris Ser. I, **347** (2009), no. 11-12, 673–676.

[3] L. Agelas, C. Guichard, R. Masson. *Convergence of Finite Volume MPFA O type Schemes for Heterogeneous Anisotropic Diffusion Problems on general meshes*, Int. J. of Finite Vol. **7** (2010), no.2.

[4] L. Agelas, R. Masson. *Convergence of the finite volume MPFA O scheme for heterogeneous anisotropic diffusion problems on general meshes*, C. R. Acad. Sci. Paris Ser. I, **346** (2008), no. 17-18, 1007–1012.

[5] E. Burman, A. Ern. *Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes*, C. R. Acad. Sci. Paris Ser. I, **338** (2004), no. 8, 641–646.

[6] B. Després. *Non linear finite volume schemes for the heat equation in 1D*, HAL: hal-00714781.

[7] J. Droniou, C. Le Potier. *Construction and Convergence Study of Schemes Preserving the Elliptic Local Maximum Principle*, SIAM Journal on Numerical Analysis, **49** (2011), no. 2, 459–490.

[8] R. Eymard, T. Gallouët, R. Herbin. *A cell-centred finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension*, IMA J. Numer. Anal., **26** (2006), no. 2, 326–353.

[9] R. Eymard, T. Gallouët, R. Herbin. *Discretisation of heterogeneous and anisotropic diffusion problems on general non-conforming meshes SUSHI: a scheme using stabilisation and hybrid interfaces*, IMA J. Numer. Anal., **30** (2010), no. 4, 1009–1043.

[10] R. Eymard, R. Herbin. *A new colocated finite volume scheme for the incompressible Navier-Stokes equations on general non matching grids*, C. R. Math. Acad. Sci. Paris, **344** (2007), no. 10, 659–662.

[11] A. Genty, C. Le Potier. *Maximum and Minimum Principles for Radionuclide Transport Calculations in Geological Radioactive Waste Repository : Comparisons Between a Mixed Hybrid Finite Element Method and Finite Volume Element Discretizations,* Transp. Porous Media, **88** (2011) 65–85.

[12] R. Herbin, F. Hubert. *Benchmark on discretization schemes for anisotropic diffusion problems on general grids, 5th International Symposium on Finite Volumes for Complex Applications*, R. Eymard and J.-M. Hérard, eds, ISTE, London; John Wiley, Inc., Hoboken, NJ, (2008), 659-692.

[13] I. Kapyrin. *A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes*, Dokl. Math., **76** (2007), no. 2, 734–738.

[14] C. Le Potier. *Schéma volumes finis pour des opérateurs de diffusion fortement anisotropes sur des maillages non structurés*, C. R. Acad. Sci. Paris Ser. I, **340** (2005), no. 12, 921–926.

[15] C. Le Potier. *A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators*, Int. J. Finite Vol., **6** (2009).

[16] C. Le Potier. *Correction non linéaire et principe du maximum pour la discrétisation d'opérateurs de diffusion avec des schémas volumes finis centrés sur les mailles.*, C. R. Acad. Sci. Paris, **348** (2010), no. 11-12, 691–695.

[17] K. Lipnikov, M Shashkov, I. Yotov, *Local flux mimetic finite difference methods*, Numer. Math., **112** (2009), no. 1, 115–152.

[18] K. Lipnikov, D. Svyatskiy, Yu. Vassilevski. *Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes*, J. Comput. Physics **228** (2009), no. 3, 703–716.

[19] J.M. Nordbotten, I. Aavastsmark, G.T. Eigestad, *Monotonicity of control volume methods*, Numer. Math. **106** (2007), no. 2, 255–288.

[20] Z. Sheng, G. Yuan, *The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes,* J. Comput. Physics **230** (2011), no. 7, 2588–2604.

[21] G. Yuan, Z. Sheng , *Monotone finite volume schemes for diffusion equations on polygonal meshes*, J. Comput. Physics **227** (2008), no. 12, 6288–6312.