Finite volume schemes for anisotropic and heterogeneous diffusion operators on non-conforming meshes

Thanh Hai Ong

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

École doctorale : **MSTIC**

Discipline : **MATHÉMATIQUES**

présentée et soutenue publiquement par

Thanh Hai Ong

le 13/11/2012

Cell-Centered Scheme For Heterogeneous Anisotropic Diffusion Problems On General Meshes

---

devant le jury composé de

ANDREIANOV Boris	-	Rapporteur
BOUCHUT François	-	Examinateur
DESPRÉS Bruno	-	Rapporteur
DI PIETRO Daniele	-	Rapporteur
EYMARD Robert	-	Directeur de thèse
HERBIN Raphaèle	-	Examinateur
LE POTIER Christophe	-	Encadrant CEA

Remerciements

The thesis was completed with the guidance and the help of several individuals who contributed and extended their valuable assistance in the preparation and completion.

Foremost, I am grateful to my advisors Robert Eymard, Christophe Le Potier and Jérôme Dronjou whose sincerity and encouragement I will never forget. Moreover, my advisors has been my inspiration as I overcome all the obstacles in the completion this research study.

Besides my advisors, I would like to thank the other Jury members: Boris Andreianov, Daniele Di Pietro, Bruno Després, Raphaèle Herbin and François Bouchut, for their encouragement, insightful comments and interesting questions.

I am thankful to Mrs. Sylvie Cach, the administrative officer of the University Paris Est, who helped and gave me advices, informations about the necessary documents for my PhD procedure.

I thank my colleagues in Cea-Saclay: Pascal Omnes, François, Kun Kun, Huyen, Lan, Hy, Trinh, Minh...for the discussions and for the fun we have had in the three years. I also thank my friends in École Polytechnique, University of Orleans and University of Tours: Du, Hao, Trung, Bao, Phuc, Liem, Binh, Trang, Diem, Nhan, Phuong, Thuy, Chi, Ky, Nhat, Nguyen Anh, Tai, Viet, Dien,...for playing football, discussions and helping me.

Finally, I deep thank to my family for encouraging, supporting me spiritually throughout my life.

Contents

List of Figures

Part I

Introduction and Presentation of manuscript

Chapter 1

Introduction

Contents

1.1 Problems and numerical simulations of nuclear waste disposal.

Nuclear waste includes leftovers and products of radioactive materials used to generate electricity, carry out certain health-care procedures, perform a variety of commercial processes and conduct university researches. Because different types of nuclear waste have been varying levels of radioactivity, according to [12], waste is divided into the two main categories: high-level waste and low-level waste. The goal of long-term radioactive waste management is to protect humankind and its environment from the effects of the materials comprised in this waste, most importantly from radiological hazards, because the possible returns to the biosphere of minute amounts of radionuclide can have no unacceptable health or environmental impact. The object of nuclear waste disposal which is seen as the reference solution help to ensure the confinement of radioactivity within a definite space, to be segregated from humankind and the environment in over several tens of thousand years, in the case of long-lived waste, or even longer. Disposal in deep geological strata (typically, 500 m down) inherently makes for deployment of a more passive technical solution, with the ability to stand, with no increased risk, an absence of surveillance. The geological environment of such a disposal facility thus forms a further, essential barrier. Therefore, it is necessary to research into nuclear waste disposal simulation. In the framework of nuclear waste disposal simulation, we are interested in the two kinds of equations: the familiar ground-water flow equations and the transport equations having a term of radioactive decay. This term of radioactive decay allows to study the migration of a radionuclide (for example Iodine 129 or Plutonium 242) in highly anisotropic heterogeneous geological layers.

In [36] and [37], the familiar ground-water flow models are governed by the relations expressed in Darcy's law and the conservation of mass. The Darcy's law which represents the rate of flow water through a porous media related to the properties of the water, the properties of the porous media, and the gradient of the hydraulic head, is written by:

$$q_i = -K_{ij}\frac{\partial h}{\partial x_j},$$

where q_i is the specific discharge, K_{ij} is the hydraulic conductivity of the porous medium, and h is the hydraulic head.
A general groundwater flow equation which may be derived by combining Darcy's law with the continuity equation, are written in Cartesian tensor, follows as:

$$\frac{\partial}{\partial x_i}\left(K_{ij}\frac{\partial h}{\partial x_i}\right) = S_S\frac{\partial h}{\partial t} + W^*,$$

where S_S is the specific storage, t is time, W^* which is the volumetric flux per unit volume (positive for outflow and negative for inflow), and x_i are the Cartesian coordinates. Fluid properties such as density and viscosity may vary significantly in space or time. This may occur where water temperature or dissolved-solids concentration changes significantly. When the water properties are heterogeneous and (or) transient, the relations among water levels, hydraulic heads, fluid pressures, and flow velocities are neither simple nor straightforward. In such cases, the flow equation is written and solved in terms of fluid pressures,

fluid densities, and the intrinsic permeability of the porous media.

In [10], the transport model of radionuclide (RN) in porous media is described by convection-diffusion-dispersion equations . These equations are to particularly consider the two concentrations of Iodine 129 (C_1) and Plutonium 242 (C_2) escaping from the repository cave into the water. They are written by:

$$R_i w \left(\frac{\partial C_i}{\partial t} + \lambda_i C_i \right) - \nabla.\left(D_i \nabla C_i\right) + u.\nabla C_i = f_i \quad in\ \Omega \times (0,T) \quad i = 1,2,$$

where

- R_i is the latency Retardation factor, with value 1 for ^{129}I, 10^5 for ^{242}Pu in the clay layer and 1 elsewhere for both Iodine and Plutonium,
- the effective porosity w, is equal to 0.001 for ^{129}I, 0.2 for ^{242}Pu in the clay player and 0.1 elsewhere for both,
- $\lambda_i = log2/T_i$ with T_i being the half life time of the element: 1.57 10^7 for ^{129}I, 3.7610^5 for ^{242}Pu (in year),
- u is the velocity of the flow,
- f_i (i = 1, 2) are the source term,
- the effective diffusion/dispersion tensors D_i for any $i = 1, 2$ depend on the Darcy velocity, as follows:

$$D_i = d_{ei} I + |u| \left[\alpha_{li} E(u) + \alpha_{ti} \left(I - E(u)\right)\right]$$

with

$$E_{kj}(u) = \frac{u_k u_j}{|u|^2}.$$

We have a sample table of diffusion dispersion coefficients for the radioactive elements in the 4 layers:

	Iodien 129			Plutonium 242		
	$d_{e_1}(m^2/year)$	α_{l_1}	α_{t_1}	$d_{e_2}(m^2/year)$	α_{l_2}	α_{t_2}
Dogger	5.04E-4	50	1	5.04E-4	50	1
Clay	9.48E-7	0	0	4.42E-4	0	0
Limestone	5.0E-4	50	1	5.04E-4	50	1
Marl	5.0E-4	0	0	5.0E-4	0	0

One of the important ingredients for the numerical solutions of these ground-water flow equations and transport equations is the discretization of anisotropic heterogeneous diffusion terms on general meshes. For this complex task, there are many proposed discretization methods which have been classified by the list of well-known discretization methods. This

list involves the finite difference method, the finite element method and the finite volume method. Each of these methods has its own advantages and disadvantages.

Firstly, the finite difference method is simple. Nevertheless, it is not widely used in practical applications, since we need a smoothness assumption of the solution and it is not applicable for the domains with a complex geometry.

Secondly, the standard finite element method has the following advantages:

- It can be applied in domains with complex shapes. These domains can be discretized by triangular meshes.
- It uses the spaces of piecewise polynomials of degree 1 to approximate the solution function. The basic functions of these spaces have small supports, so the computation of this method is simple.

Unfortunately, in discontinuous diffusion problems coupled with convective transport models, the approximate solutions can be inaccurate [56], when they are computed by the standard finite element method.

Thirdly, the finite volume method is known as a accuracy and cheap method for the discretization of conservation laws to approximate the solutions of anisotropic heterogeneous diffusion problems. Furthermore, it exists local conservativity of the fluxes that is significant in physics. Now, this method is classified into the two main categories:

- “Cell-centered schemes” compute approximate values of the solution function at the centers of the cells of the primary meshes.
- Other schemes use not only usual cell unknowns but also interface unknowns to compute approximation values of the solution function. In [20], HFV, MFD and MFV involve the cell and edge unknowns. These schemes belong to the same family. In addition, the DDFV schemes in [18], [31] which give precise solutions use techniques of dual mesh. They depend on the cell and vertex unknowns. Obviously, all schemes of this category are more expensive than cell-centered schemes, because they use more unknowns.

Therefore, we pay attention to “Cell-centered schemes” with small stencils and only using cell unknowns. By these two advantages, they are often used in industrial codes. "Cell-centered schemes" still have disadvantages. The so-called Multipoint Flux Approximation (MPFA) [2], [3] involves the reconstruction of the gradient in order to evaluate the fluxes. But these methods only satisfy coercivity under suitable conditions on both the mesh and the permeability tensor Λ. In [7], the authors need a coercivity assumption linking the mesh and the tensor. Additionally, some proposed schemes only need either conditions on meshes or conditions on the permeability tensor. For examples, in [5], the condition is that the meshes are not too distorted. In [16], the authors need a sufficient coercivity condition. In this kind of schemes, let us also cite methods [21, 46, 50, 51, 42, 43], the methods allow to obtain maximum and minimum principles for diffusion problems on distorted meshes.

1.2 Content and objective of the thesis.

In my work, the proposed scheme is designed on general meshes for heterogeneous and anisotropic permeability tensors with the main following characteristics:

- The main idea of the scheme is based on that of the standard finite element method and a technique of dual mesh. The dual mesh is chosen to be easily recovered a cell centered scheme, i.e, the dual mesh unknowns are computed by linear combinations of cell unknowns. This is different from other schemes which use techniques of the dual mesh such as DDFV schemes, because we can not recover a cell centered scheme with these dual meshes.
- It is a cell-centered scheme and its stencil is equal or less than nine on quadrangular meshes and twenty seven on hexahedral meshes.
- In heterogeneous and homogeneous anisotropic cases, it is locally conservative.
- In general cases, using a light assumption (hypothesis 3.1), the matrix which is associated to our scheme is symmetric and positive definite on general meshes.
- It is exact on cell-wise affine solutions for cell-wise constant diffusion tensors.

Its name is the finite element cell-centered scheme (the FECC scheme).

On the other hand, we focus on studying the existence of a maximum principle. This existence is one of the fundamental properties of approximate solutions for the groundwater flow and the transport equations, because violation of the maximum principle can lead to non-physical solutions. In these equations, the convective term is discretized using a classical upwind scheme that satisfies the minimum and the maximum principles [49], [39]. Hence, we only pay attention to the existence of the principles for the diffusion-dispersive term. However, classical finite volume and finite element schemes do not satisfy these principles for distorted meshes or for high anisotropy ratio of diffusion tensors [19], [45], [34], [41]. In [45], the authors proved that it is impossible to construct nine-point methods which unconditionally satisfy the monotonicity criteria when the discretization satisfies local conservation and exact reproduction of linear potential fields. The proposed non-linear finite volume schemes [21], [29], [35], [51], [43], [55], [30] satisfy the desired properties and the accurate results, but they are coercive with conditions on the meshes and on the anisotropy ratio. The FECC scheme also violates the maximum principle (see Test 2 in the section 7.3 of chapter 7). That is why we will study three non-linear corrections for the FECC schemes. Their constructions are based in the spirit of methods developed in [53] and [13].

Finally, in this work, the efficiency of the FECC scheme is demonstrated though numerical tests of the 5th & 6th International Symposium on Finite Volumes for Complex Applications - FVCA 5 & 6. In addition, the comparison with classical finite volume schemes emphasizes the precision of the method. We show the good behaviour of algorithms for nonconforming meshes. We also present 3D numerical results for the linear and the non-linear schemes.

1.3 Structure of the thesis.

The thesis started with a discussion about problems and numerical simulations of nuclear waste disposal. From these problems, we focused on analysis of advantages and disadvantages of classical numerical schemes used in the discretization of anisotropic heterogeneous diffusion equation on general meshes. The thesis is then organized, as follows:

In Chapter 2, we concentrate on the constructions of the dual, the third meshes and the definition of the FECC scheme in isotropic homogeneous cases and in anisotropic heterogeneous cases in two dimensions.

In Chapter 3, we present the 3D extension of the constructions of the dual, the third meshes and the definition of the FECC scheme in isotropic homogeneous cases and in anisotropic heterogeneous cases.

In Chapter 4, we show the mathematical properties of the FECC scheme including the symmetric positive definite matrix associated to the scheme, the small stencil, the relationship between the FECC scheme and the scheme of [5], the accuracy on cell-wise affine solutions for cell-wise constant diffusion tensors.

In Chapter 5, the convergence of the FECC scheme with the piecewise Lipschitz-continuous tensor Λ is proved. The key point of the proof consists in showing both the strong and the weak consistency of the method.

In Chapter 6, we present non-linear corrections for the FECC schemes and the properties of the modified schemes.

In Chapter 7, we show 2D and 3D numerical results of FVCA 5 & 6 and comparisons with classical finite volume schemes. Additionally, we give numerical results to test the existence, violation of the maximum principle for the FECC schemes and the non-linear FECC schemes (NLFECC1 and NLFECC2).

In Chapter 8, the conclusion and the perspectives of the schemes are discussed.

Part II

A new cell-centered scheme for heterogeneous anisotropic diffusion problem on general meshes

Chapter 2

The FECC scheme in two dimensions

Contents

2.1 Motivation.

In this chapter, we construct a new cell centered scheme for approximating heterogeneous anisotropic diffusion operators on general meshes. For this work, we recall that the Multipoint Flux Approximate (MPFA) [2], [3] involves the reconstruction of the gradient in order to evaluate the fluxes. Nevertheless, these methods only satisfy coercivity under suitable conditions on both the mesh and the permeability tensor Λ. In [7], the authors need a coercivity assumption which links the mesh and the tensor. There are also some schemes which need either conditions on meshes or conditions on the permeability tensor. For example, in [5], the condition is that the meshes are not too distorted. In addition, the Sushi scheme [23] is unconditionally coercive, but its stencil includes the neighbours of the neighbours of a given control volume.

The proposed scheme [52] has the main characteristics:

- Its main idea is based on the P1 standard finite element method on the third triangular mesh built from a particular dual mesh.
- It is a cell-centered scheme and its stencil is equal or less than nine on quadrangular meshes.
- In heterogeneous and homogeneous anisotropic cases, it is locally conservative.
- In general cases, using a light assumption (hypothesis 2.4.1), the matrix which is associated to our scheme is symmetric and positive definite on general meshes.

2.2 The continuous problem.

The approximation of the solutions of the anisotropic heterogeneous diffusion problems is an important issue in several engineering fields. We mention a kind of these problems, as follows:

$$\begin{cases} -\mathrm{div}\left(\Lambda(x)\nabla u\right) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

where the following assumptions hold:

1. Ω is an open bounded connected polygonal subset of $\mathbb{R}^d$ with $d = 2, 3$.
2. The diffusion (or permeability) tensor $\Lambda : \Omega \to \mathbb{R}^{d\times d}$ is symmetric, uniformly positive-definite and such that the set of its eigenvalues is included in $[\underline{\lambda}, \overline{\lambda}]$, with $\underline{\lambda}$ and $\overline{\lambda}$ $\in \mathbb{R}$ satisfying $0 < \underline{\lambda} \leq \overline{\lambda}$.
3. The function f is the source term and belongs to $L^2(\Omega)$.

With assumptions (1)-(3), u is called the weak solution of (2.1) if u satisfies

$$u \in H_0^1(\Omega) \text{ and } \forall v \in H_0^1(\Omega), \int_\Omega \left(\Lambda(x)\nabla u(x)\right).\nabla v(x)dx = \int_\Omega f(x)v(x)dx. \tag{2.2}$$

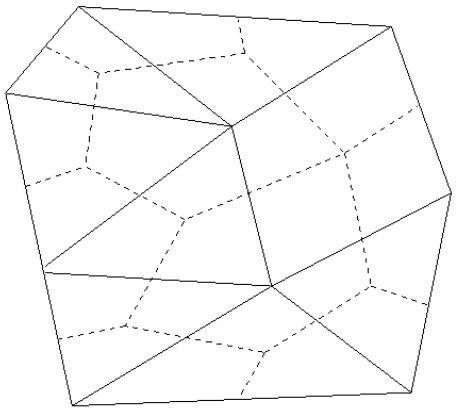


Figure 2.1: A sample primary mesh (solid lines) and its dual mesh (dashed lines).

2.3 Notations.

Let Ω be an open bounded polygonal set of $\mathbb{R}^2$ with the boundary $\partial\Omega$. We denote the three discretization families of Ω by $\mathcal{D}$, $\mathcal{D}^*$ and $\mathcal{D}^{**}$, which are given by

2.3.1 The 2D primary discretization family $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$.

The primary discretization family is denoted by $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, such that:

- $\mathcal{M}$ is a finite family of non empty connected open disjoint subsets of Ω such that $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$. $m_K > 0$ denotes the measure of K (the "primary control volume").
- $\mathcal{E}$ (the set of "edges" of the primary grid) is a set of disjoint subsets of $\overline{\Omega}$ such that, for all $\sigma \in \mathcal{E}$, σ is a segment in $\mathbb{R}$, $m_\sigma > 0$ denotes the measure of σ. Let K be an element of $\mathcal{M}$, we assume that there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \bigcup_{\sigma\in\mathcal{E}_K} \sigma$ and $\mathcal{E} = \bigcup_{K\in\mathcal{M}} \mathcal{E}_K$. The set of interior edges is denoted by $\mathcal{E}_{\text{int}}$ (resp. $\mathcal{E}_{\text{ext}}$) with $\mathcal{E}_{\text{int}} = \{\sigma \in \mathcal{E} | \sigma \not\subset \partial\Omega\}$ (resp. $\mathcal{E}_{\text{ext}} = \{\sigma \in \mathcal{E} | \sigma \subset \partial\Omega\}$).
- $\mathcal{P} = (x_K)_{K\in\mathcal{M}}$ is a set of all mesh points of the primary grid. For all $K \in \mathcal{M}$, $x_K \in K$ and K is assumed to be x_K-star-shaped, i.e for all $x \in K$, $[x_K, x] \in K$.

2.3.2 The 2D dual discretization family $\mathcal{D}^* = (\mathcal{M}^*, \mathcal{E}^*, \mathcal{P}^*, \mathcal{V}^*)$.

The dual discretization family $\mathcal{D}^* = (\mathcal{M}^*, \mathcal{E}^*, \mathcal{P}^*, \mathcal{V}^*)$ is defined, as follows:

- The dual control volumes K^* are defined by connecting mesh points of the primary control volumes and the midpoints of the edges belonging to $\partial\Omega$. Moreover, we assume that the lines joining their mesh points are inside Ω. In this case, $\mathcal{M}^*$ which is

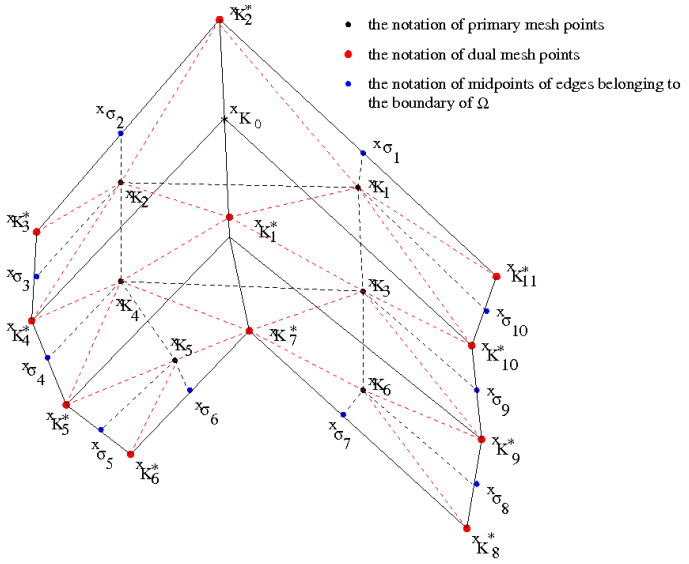


Figure 2.2: An example for the remark 2.3.2.

a set of all dual control volumes such that $\overline{\Omega} = \bigcup_{K^* \in \mathcal{M}^*} \overline{K^*}$, is defined and we assume it fits the initial domain Ω. We denote by $m_{K^*} > 0$ the measure of K^*.

Remark 2.3.1: A sufficient condition to define the mesh $\mathcal{M}^*$ is that, for neighbouring control volumes, the line joining their centers intersects their common edge. This condition is not necessary.

- $\mathcal{E}^*$ (the "edges" of the dual grid) is a set of disjoint subsets of $\overline{\Omega}$ such that, for all $\sigma^* \in \mathcal{E}^*$, σ^* is a segment in $\mathbb{R}$, the measure of σ^* is denoted by $m_{\sigma^*} > 0$. Let K^* be an element of $\mathcal{M}^*$, we assume that there exists a subset $\mathcal{E}_{K^*}$ of $\mathcal{E}^*$ such that $\partial K^* = \bigcup_{\sigma \in \mathcal{E}_{K^*}} \sigma$ and $\mathcal{E}^* = \bigcup_{K^* \in \mathcal{M}^*} \mathcal{E}_{K^*}$.
- $\mathcal{P}^* = (x_{K^*})_{K^* \in \mathcal{M}^*}$ is the set of all mesh points of the dual grid.
- $\mathcal{V}^*$ is the set of all vertices of the dual meshes which includes the primary mesh points, the midpoints of the edges belonging to $\partial\Omega$ and boundary vertices of Ω.

Remark 2.3.2: We do not always use the vertices of the primary mesh as dual mesh points $\{x_{K^*}\}_{K^* \in \mathcal{M}^*}$. For example, we consider the following polygon Ω: we can not choose the vertex x_{K_0} of the primary mesh to define the dual mesh of $(x_{K_1}, x_{K_2}, x_{K_3}, x_{K_4}) \in \mathcal{M}^*$, because it is outside $(x_{K_1}, x_{K_2}, x_{K_3}, x_{K_4})$.

2.3.3 The 2D third discretization family $\mathcal{D}^{**} = (\mathcal{M}^{**}, \mathcal{V}^{**})$.

The notation $\mathcal{D}^{**} = (\mathcal{M}^{**}, \mathcal{V}^{**})$ which is the third discretization family, is constructed by:

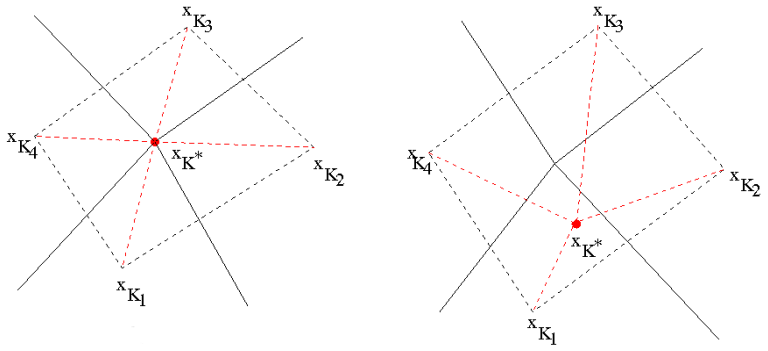


Figure 2.3: the two examples of the third meshes.

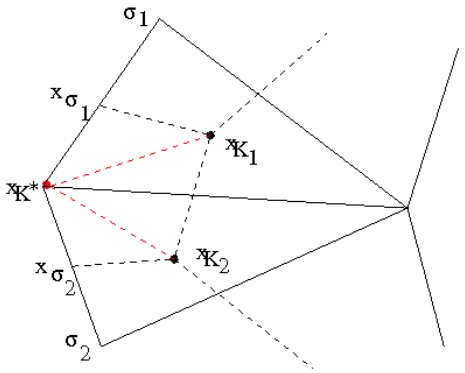


Figure 2.4: a dual mesh point x_{K^*} is on the boundary $\partial\Omega$.

Let be $K^* \in \mathcal{M}^*$, if all edges of $\mathcal{E}_{K^*}$ do not belong to the boundary $\partial\Omega$, the set of vertices of K^* only contains mesh points of the primary control volumes. A point x_{K^*} is chosen inside K^* and connected to all vertices of K^*, for examples Figure 2.3.

If K^* has a vertex x_{K^*} belonging to the boundary of Ω, its dual mesh point is equal to the vertex x_{K^*} (see Figure 2.4). We connect x_{K^*} to the other vertices of K^*.

In Figures (2.2)-(2.4), we denote that

- The primary mesh is represented by solid black lines.
- The dual mesh is represented by dashed black lines.
- The third mesh is represented by dashed red and black lines.
- The primary mesh points $x_{K_1}, x_{K_2}, x_{K_3}, x_{K_4}$ are elements of $\mathcal{P}$.
- The dual mesh points x_{K^*}, x_{L^*} are elements of $\mathcal{P}^*$.
- The edges σ_1, σ_2 are edges of the boundary of Ω.

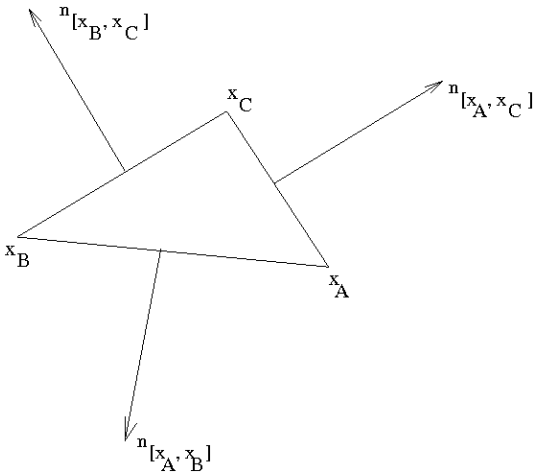


Figure 2.5: the triangle (x_A, x_B, x_C).

- The points x_{σ_1}, x_{σ_2} are midpoints of the edges σ_1, σ_2.

From the construction of the third grid, this implies that it is a sub-grid of the dual grid.

- $\mathcal{M}^{**}$ is a finite family of sub-triangles such that $\overline{\Omega} = \underset{T\in\mathcal{M}^{**}}{\cup} \overline{T}$.
- $\mathcal{V}^{**}$ is a finite set of all vertices of the third grid such that, for all $T \in \mathcal{M}^{**}, \mathcal{V}_T^{**}$ is a set of three vertices of triangle T and $\mathcal{V}^{**} = \underset{T\in\mathcal{M}^{**}}{\bigcup} \mathcal{V}_T^{**}$. The set of interior vertices is denoted by $\mathcal{V}_{\text{int}}^{**} = \mathcal{P} \cup \mathcal{P}^*$. Moreover, the functions p_K and p_{K^*} with $K \in \mathcal{M}$, $K^* \in \mathcal{M}^*$ are piecewise linear continuous functions defined by

$$p_K(x) = \begin{cases} 1 & \text{at } x = x_K,\ x_K \in \mathcal{P}, \\ 0 & \text{at } x \in \mathcal{V}_{\text{int}}^{**} \setminus \{x_K\}, \\ & 0 \quad \text{on } \partial\Omega. \end{cases}$$

$$p_{K^*}(x) = \begin{cases} 1 & \text{at } x = x_{K^*},\ x_{K^*} \in \mathcal{P}^*, \\ 0 & \text{at } x \in \mathcal{V}_{\text{int}}^{**} \setminus \{x_{K^*}\}, \\ & 0 \quad \text{on } \partial\Omega. \end{cases}$$

  Additionally, we introduce some notations $n_{[x_A,x_B]}$, $n_{[x_A,x_C]}$, $n_{[x_B,x_C]}$ which are outward normal vectors of the triangle (x_A, x_B, x_C) (Figure 2.5). The lengths of these vectors are equal to the segments $[x_A, x_B]$, $[x_A, x_C]$, $[x_B, x_C]$ and $m_{(x_A,x_B,x_C)}$ is the measure of the triangle (x_A, x_B, x_C).

2.4 Presentation of the 2D FECC scheme.

Now, we introduce our scheme *A Cell-Centered Scheme For Heterogeneous Anisotropic Diffusion Problems On General Meshes*. We name it FECC for the Finite Element Cell-Centered scheme.

Definition 2.4.1: Let us define the discrete function space $\mathcal{H}_\mathcal{D}$ as the set of all $((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*})$, $u_K \in \mathbb{R}$, $K \in \mathcal{M}$, $u_{K^*} \in \mathbb{R}$, $K^* \in \mathcal{M}^*$ and $u_{K^*} = 0$ if x_{K^*} belongs to the boundary of Ω.

$P(v)$ which is a function on Ω, is constructed from the value $v = ((v_K)_{K\in\mathcal{M}}, (v_{K^*})_{K^*\in\mathcal{M}^*})$. The function $\nabla_{\mathcal{D},\Lambda} u$ is intended to be a discrete gradient of ∇u, taking into account $u = ((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*})$. As a result, equation (2.2) is discretized by the following discrete variational formulation

$$\int_\Omega (\Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x)) . \nabla_{\mathcal{D},\Lambda} v(x)\, dx = \int_\Omega f(x)P(v)(x)dx \quad \text{for all } v \in \mathcal{H}_\mathcal{D}. \tag{2.3}$$

From equation (2.3), we describe the FECC scheme in each of the following cases of Λ.

2.4.1 Isotropic homogeneous cases in 2D.

The main idea of the FECC scheme is the same as that of the standard finite element method ($P1$) on the third triangular mesh. The domain Ω is partitioned by this third mesh.
For any $u = ((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*}) \in \mathcal{H}_\mathcal{D}$, $P(u)$ is defined by

$$P(u)(x) = \sum_{K\in\mathcal{M}} u_K . p_K(x) + \sum_{K^*\in\mathcal{M}^*} u_{K^*} . p_{K^*}(x), \tag{2.4}$$

and $\nabla_{\mathcal{D},\Lambda} u$ is defined by

$$\nabla_{\mathcal{D},Id} u(x) = \nabla_{\mathcal{D},Id} P(u)(x) = \sum_{K\in\mathcal{M}} u_K \nabla p_K(x) + \sum_{K^*\in\mathcal{M}^*} u_{K^*} \nabla p_{K^*}(x). \tag{2.5}$$

Substituting the definitions (2.4) and (2.5) into the equation (2.3), for each $L \in \mathcal{M} \cup \mathcal{M}^*$, we choose $v = ((v_K)_{K\in\mathcal{M}}, (v_{K^*})_{K^*\in\mathcal{M}^*}) \in \mathcal{H}_\mathcal{D}$ such that $v_K = 0$ if $K \neq L$, $v_{K^*} = 0$ if $K^* \neq L$, $v_L = 1$ and $P(v) = p_L$. The resulting equation can be re-written in the following form

$$\int_\Omega \left(\sum_{K\in\mathcal{M}} u_K \nabla p_K + \sum_{K^*\in\mathcal{M}^*} u_{K^*} \nabla p_{K^*} \right) . \nabla p_L \; dx = \int_\Omega f . p_L \; dx. \tag{2.6}$$

We present the construction of the FECC scheme in an isotropic homogeneous case.

Step 1 *Recover all u_{K^*} with $K^* \in \mathcal{M}^*$ by linear functions of $(u_K)_{K\in\mathcal{M}}$ and constants depending on function f.*

We choose p_L equal to p_{K^*}. We have

$$\int_\Omega \left(\left(\sum_{K\in\mathcal{M}} u_K \nabla p_K \right) . \nabla p_{K^*} + u_{K^*} \nabla p_{K^*} . \nabla p_{K^*} \right) \; dx = \int_\Omega f . p_{K^*} \; dx,$$

because $supp\{p_{K^*}\}$ is a subset of K^* for all $K^* \in \mathcal{M}^*$.
Thus, u_{K^*} is equal to $\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)$ defined by

$$\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right) = -\sum_{K\in\mathcal{M}} u_K \int_\Omega \nabla p_K(x) \frac{\nabla p_{K^*}(x)}{\|\nabla p_{K^*}\|^2_{L^2(\Omega)}}\, dx + \int_\Omega f(x) \frac{p_{K^*}(x)}{\|\nabla p_{K^*}\|^2_{L^2(\Omega)}}\, dx.$$

Step 2 *Transform the variables in formula (2.4).*

$$P(u)(x) = \sum_{K\in\mathcal{M}} u_K.p_K(x) + \sum_{K^*\in\mathcal{M}^*} \Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right).p_{K^*}(x). \tag{2.7}$$

Step 3 *Construct a system of linear equations.*

Substituting (2.7) into (2.6), for each p_L belonging to $\{p_K\}_{K\in\mathcal{M}}$, we get

$$\int_\Omega \left(\sum_{K\in\mathcal{M}} u_K \left(\nabla p_K.\nabla p_L\right) + \sum_{K^*\in\mathcal{M}^*} \Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{P}}, f\right)\left(\nabla p_{K^*}.\nabla p_L\right) \right) dx = \int_\Omega f.p_L\, dx. \tag{2.8}$$

This is a linear equation which only involves the cell unknowns $\{u_K\}_{K\in\mathcal{M}}$. Hence, we construct a system of linear equations

$$A.U = B, \tag{2.9}$$

where U is a vector $(u_K)_{K\in\mathcal{M}}$ and A is a square matrix in $\mathbb{R}^{card(\mathcal{M})\times card(\mathcal{M})}$. All unknowns $(u_{K^*})_{K^*\in\mathcal{P}^*}$ have been eliminated, the scheme is thus indeed cell-centered.

2.4.2 Anisotropic heterogeneous cases in 2D.

To simplify the description of the FECC scheme, we assume that, for neighbouring control volumes, the line joining their primary mesh points intersects their common edge. Hence, the dual mesh is centred around the vertices of the primary mesh and the dual mesh points are the vertices of the primary mesh. Taking any $\sigma \in \mathcal{E}_{\text{int}}$ such that $\mathcal{M}_\sigma = \{K, L\}$, $x_K,\ x_L \in \mathcal{P}$ and $x_{K^*} \in \mathcal{P}^*$, (x_K, x_L, x_{K^*}) is a notation of a triangle of $\mathcal{M}^{**}$. We take the values $u_{K^*},\ u_K,\ u_L$ of u at $x_{K^*},\ x_K,\ x_L$. From values $u_{K^*},\ u_K,\ u_L$, we want to construct a discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ on the triangle (x_{K^*}, x_K, x_L) taking into account the heterogeneity of Λ. We consider the function

$$P_{(K^*,K,L)}(u) \ : \quad (x_{K^*}, x_K, x_L) \to \mathbb{R},$$

where it is continuous, linear on (x_{K^*}, x_K, x_σ) and (x_{K^*}, x_L, x_σ) (two half triangles of (x_K, x_L, x_{K^*})). We introduce a value $u_\sigma^{K^*}$ (a temporary unknown) at x_σ. The point x_σ is an intersecting point between the line joining two mesh points x_K, x_L and the internal edge σ. The discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ is then defined by

- on the triangle (x_{K^*}, x_K, x_σ)

$$P_{(K^*,K,L)}(u)(x) = \begin{cases} u_K & x = x_K, \\ u_{K^*} & x = x_{K^*}, \\ u_\sigma^{K^*} & x = x_\sigma. \end{cases}$$

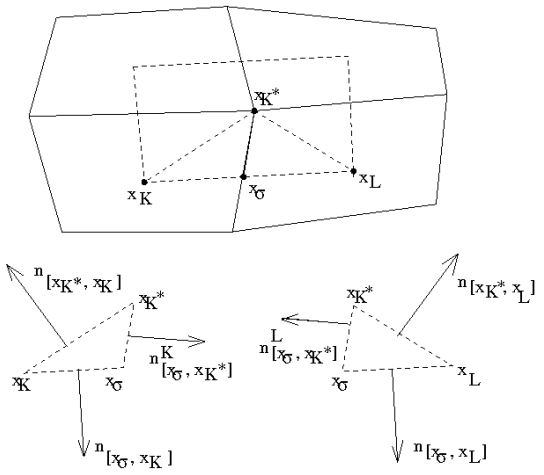


Figure 2.6: the two sub-triangle (x_{K^*}, x_K, x_σ) and (x_{K^*}, x_L, x_σ) and their outward normal vectors.

$$\begin{aligned}
&\nabla_{\mathcal{D},\Lambda} u = \nabla_{\mathcal{D},\Lambda} P_{(K^*,K,L)}(u) \\
= \quad & \frac{-P_{(K^*,K,L)}(u)(x_\sigma) n_{[x_{K^*},x_K]} - P_{(K^*,K,L)}(u)(x_K) n^K_{[x_\sigma,x_{K^*}]} - P_{(K^*,K,L)}(u)(x_{K^*}) n_{[x_\sigma,x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}} \\
= \quad & \frac{-u_\sigma^{K^*} n_{[x_{K^*},x_K]} - u_K n^K_{[x_\sigma,x_{K^*}]} - u_{K^*} n_{[x_\sigma,x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}},
\end{aligned}$$

where $n^K_{[x_\sigma,x_{K^*}]}$ is outer normal vector to the triangle (x_{K^*}, x_K, x_σ). The length of the vector $n^K_{[x_\sigma,x_{K^*}]}$ is equal to the length of the segment $[x_\sigma, x_{K^*}]$. If x_σ belongs to the boundary of Ω then $u_\sigma^{K^*} = 0$.

- on the triangle (x_{K^*}, x_L, x_σ)

$$P_{(K^*,K,L)}(u)(x) = \begin{cases} u_L & x = x_K, \\ u_{K^*} & x = x_{K^*}, \\ u_\sigma^{K^*} & x = x_\sigma. \end{cases}$$

$$\begin{aligned}
&\nabla_{\mathcal{D},\Lambda} u = \nabla_{\mathcal{D},\Lambda} P_{(K^*,K,L)}(u) \\
= \quad & \frac{-P_{(K^*,K,L)}(u)(x_\sigma) n_{[x_{K^*},x_L]} - P_{(K^*,K,L)}(u)(x_L) n^L_{[x_\sigma,x_{K^*}]} - P_{(K^*,K,L)}(u)(x_{K^*}) n_{[x_\sigma,x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}} \\
= \quad & \frac{-u_\sigma^{K^*} n_{[x_{K^*},x_L]} - u_L n^L_{[x_\sigma,x_{K^*}]} - u_{K^*} n_{[x_\sigma,x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}},
\end{aligned}$$

where $n^L_{[x_\sigma,x_{K^*}]}$ is outer normal vector to the triangle (x_{K^*}, x_L, x_σ). The length of the vector $n^L_{[x_\sigma,x_{K^*}]}$ is equal to the length of segment $[x_\sigma, x_{K^*}]$. If x_σ belongs to boundary of Ω then $u_\sigma^{K^*} = 0$.

These definitions depend on $u_\sigma^{K^*}$ but we fix $u_\sigma^{K^*}$ by imposing the Local Conservativity of the Fluxes condition, i.e

$$\Lambda_K \left(\nabla_{\mathcal{D},\Lambda} u\right)_{|(x_{K^*},x_K,x_\sigma)} .n^K_{[x_\sigma,x_{K^*}]} + \Lambda_L \left(\nabla_{D,\Lambda} u\right)_{|(x_{K^*},x_L,x_\sigma)} .n^L_{[x_\sigma,x_{K^*}]} = 0, \tag{2.10}$$

where Λ_K, Λ_L are values of Λ on K and L.

Equation (2.10) corresponds to the following equation

$$u_\sigma^{K^*} = \beta_K^{K^*,\sigma} u_K + \beta_L^{K^*,\sigma} u_L + \beta_{K^*}^{K^*,\sigma} u_{K^*}, \tag{2.11}$$

with

$$\beta_K^{K^*,\sigma} = \left(\frac{(n^K_{[x_\sigma,x_{K^*}]})^T \Lambda_K n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}}\right) \Bigg/ \left(-\frac{(n^K_{[x_\sigma,x_{K^*}]})^T \Lambda_K n_{[x_{K^*},x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}} - \frac{(n^L_{[x_\sigma,x_{K^*}]})^T \Lambda_L n_{[x_{K^*},x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\right),$$

$$\beta_L^{K^*,\sigma} = \left(\frac{(n^L_{[x_\sigma,x_{K^*}]})^T \Lambda_L n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\right) \Bigg/ \left(-\frac{(n^K_{[x_\sigma,x_{K^*}]})^T \Lambda_K n_{[x_{K^*},x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}} - \frac{(n^L_{[x_\sigma,x_{K^*}]})^T \Lambda_L n_{[x_{K^*},x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\right),$$

$$\beta_{K^*}^{K^*,\sigma} = 1 - \beta_K^{K^*,\sigma} - \beta_L^{K^*,\sigma}.$$

From equation (2.11), the unknown $u_\sigma^{K^*}$ is computed by u_K, u_{K^*} and u_L. Thus, the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ on (x_K, x_L, x_{K^*}) only depends on these three values.

Hypothesis 2.1: we assume

$$\left(-\frac{(n^K_{[x_\sigma,x_{K^*}]})^T \Lambda_K n_{[x_{K^*},x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}} - \frac{(n^L_{[x_\sigma,x_{K^*}]})^T \Lambda_L n_{[x_{K^*},x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\right) \neq 0.$$

Remark 2.4.1:

- In isotropic heterogeneous cases, if the primary mesh is an admissible mesh (see definition 3.1, paper $37-39$ in [22]), the unknown $u_\sigma^{K^*}$ is computed by

$$u_\sigma^{K^*} = \underbrace{\beta_K^{K^*,\sigma}}_{>0} u_K + \underbrace{\beta_L^{K^*,\sigma}}_{>0} u_L,$$

  because $n^K_{[x_\sigma,x_{K^*}]}.\, n_{[x_\sigma,x_K]} = 0$ and $n^L_{[x_\sigma,x_{K^*}]}.\, n_{[x_\sigma,x_L]} = 0$.

- In isotropic homogeneous cases, we do not need the hypothesis 3.1, since the coefficients are different from 0.

We present the construction of the FECC scheme in anisotropic heterogeneous cases.

Step 1 *Recover all u_{K^*} with $K^* \in \mathcal{M}^*$ by linear functions of $(u_K)_{K\in\mathcal{M}}$ and constants depending on function f.*

For each $K^* \in \mathcal{M}^*$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*})$ such that $v_L = 0$ for all $L \in \mathcal{M}$, $v_{L^*} = 0$ if $L^* \neq K^*$ and $v_{K^*} = 1$ in equation (2.3)

$$\int_\Omega (\Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x)) .\nabla_{\mathcal{D},\Lambda}v(x)\, dx = \int_\Omega f(x)P(v)(x)dx.$$

The discrete gradient $\nabla_{\mathcal{D},\Lambda}v$ is equal to 0 on L^* which is different from K^*. It implies that $\int_\Omega (\Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x)) .\nabla_{\mathcal{D},\Lambda}v(x)\ dx$ is presented by the linear function of $(u_K)_{K\in\mathcal{M}}$, u_{K^*} and a constant depending on function f. Therefore, the unknown u_{K^*} is computed by a linear function of $\{u_K\}_{K\in\mathcal{M}}$ and a constant depending on function f. This linear function is also denoted by $\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)$.

Step 2 *Reconstruct the discrete gradient* $\nabla_{\mathcal{D},\Lambda}u$.

In the definition of the discrete gradient $\nabla_{\mathcal{D},\Lambda}u$, we transform all the unknowns $\{u_{K^*}\}_{K^*\in\mathcal{M}^*}$ by $\left\{\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)\right\}_{K^*\in\mathcal{M}^*}$. Hence, $\nabla_{\mathcal{D},\Lambda}u$ does not depend on unknowns $\{u_{K^*}\}_{K^*\in\mathcal{M}^*}$.

Step 3 *Construct a system of linear equations.*

In equation (2.3), for each $K \in \mathcal{M}$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*}) \in \mathcal{H}_\mathcal{D}$ such that $v_{L^*} = 0$ for all $L^* \in \mathcal{M}^*$, $v_L = 0$ if $L \neq K$ and $v_K = 1$. This resulting equation is a linear equation which only involves unknowns $\{u_K\}_{K\in\mathcal{M}}$. Thus, we construct a system of linear equations

$$A.U = B, \tag{2.12}$$

where U is the vector $(u_K)_{K\in\mathcal{M}}$ and A is a square matrix in $\mathbb{R}^{card(\mathcal{M})\times card(\mathcal{M})}$.

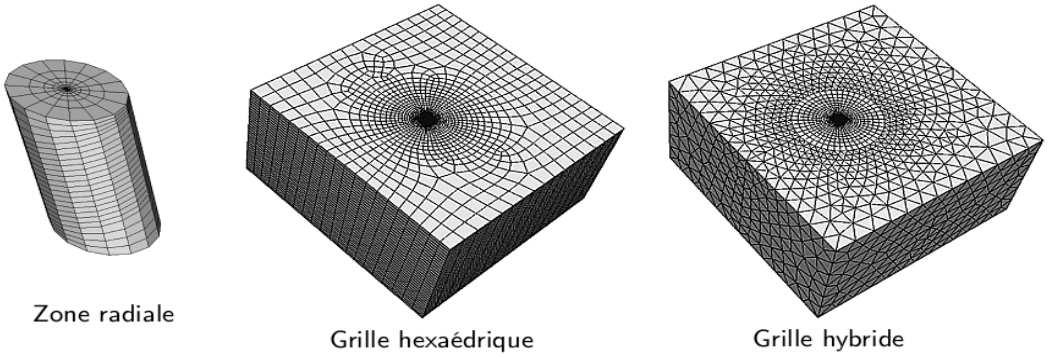
Chapter 3

The FECC scheme in three dimensions

Contents

3.1 Motivation.

In industrial engineering codes, the meshes are generated from the modeling of the underground geological layers. These layers produce complex shapes, such as: faults, inclined wells, highly heterogeneous permeability fields...By this issue, 2D meshes need to be extended to 3D meshes, for examples: tetrahedral meshes, prism meshes, hexahedral meshes, pyramidal meshes...



To take into account these meshes, we present the FECC scheme and the constructions of the dual, the third meshes in 3D. We recover the same properties as in two dimensions.

3.2 Notations.

Let Ω be a bounded open domain of $\mathbb{R}^3$ with the boundary $\partial\Omega$. Before we introduce the three discretization families $\mathcal{D}$, $\mathcal{D}^*$ and $\mathcal{D}^{**}$ of Ω, we explain some notations in Figures (3.1)-(3.10):

- The edges of the primary mesh are represented by the black color.
- The edges of the dual mesh are only represented by the blue color.
- The edges of the third mesh are represented by the blue and red colors.
- The mesh points of the primary mesh are represented by the black dots.
- The mesh points of the dual mesh are represented by the blue dots.
- The mesh points of the third mesh are represented by the red dots.
- The edge points are represented by the pink dots.

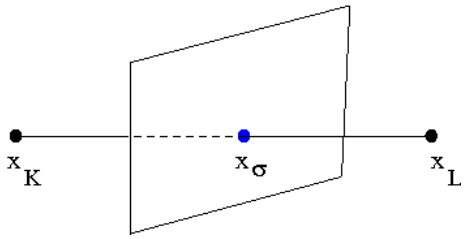


Figure 3.1: an intersecting point x_σ on the non distorted primary mesh $\mathcal{M}$.

3.2.1 The 3D primary discretization family $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P}, \mathcal{F}, \mathcal{V})$

The domain Ω is discretized by the primary mesh $\mathcal{M}$. The primary mesh $\mathcal{M}$ is the set of open disjoint polyhedral control volumes $K \in \Omega$ such that $\bigcup_{K \in \mathcal{M}} \overline{K} = \overline{\Omega}$.

The set of interfaces of the two control volumes K and L is denoted by $\mathcal{F}_{K,L} = \overline{K} \cap \overline{L}$. The set $\mathcal{F}_{K,L}$ is a subset of $\mathcal{F}_{int}$ with $\mathcal{F}_{int} = \bigcup_{\substack{K,L \in \mathcal{M} \\ K \cap L \neq \emptyset}} \mathcal{F}_{K,L}$. In addition, the set of boundary interfaces of $\partial K \cap \partial\Omega$ is denoted by $\mathcal{F}_{K,\partial\Omega}$. The set $\mathcal{F}_{K,\partial\Omega}$ belongs to $\mathcal{F}_{ext}$, where $\mathcal{F}_{ext} = \bigcup_{\substack{K \in \mathcal{M} \\ \partial K \cap \partial\Omega \neq \emptyset}} \mathcal{F}_{K,\partial\Omega}$. For each edge e on the boundary $\partial\Omega$, a set $\mathcal{F}^e_{ext}$ of which all elements contain the edge e is a subset of $\mathcal{F}_{ext}$. We then denote the union of $\mathcal{F}_{int}$ and $\mathcal{F}_{ext}$ by $\mathcal{F}$.

Vertices and edges of all faces in $\mathcal{F}$ are vertices and edges of $\mathcal{M}$. The two sets of these vertices and edges are respectively denoted by $\mathcal{V}$ and $\mathcal{E}$. Moreover, the set $\mathcal{E}$ has the two subsets $\mathcal{E}_{ext}$ and $\mathcal{E}_{int}$ such that $\mathcal{E} = \mathcal{E}_{int} \cup \mathcal{E}_{ext}$. The set $\mathcal{E}_{int}$ (resp. $\mathcal{E}_{ext}$) is the set of all edges inside Ω (resp. on the boundary $\partial\Omega$). For each edge $e \in \mathcal{E}$, the set $\mathcal{M}_e$ is a subset of $\mathcal{M}$, where the edge e is a common edge of all control volumes in $\mathcal{M}_e$.

For each $K \in \mathcal{M}$, the mesh point x_K is an inside point of the control volume K, and the set of mesh points of all control volumes in $\mathcal{M}$ is denoted by $\mathcal{P} = \{x_K | K \in \mathcal{M}\}$. Besides, we define a face point, as follows:

- Consider the two neighbouring control volumes K and L, $\sigma \in \mathcal{F}_{int}$, the face point x_σ is defined by an intersecting point between the segment $[x_K, x_L]$ and the face σ.

  Remark 3.2.1: if the primary mesh $\mathcal{M}$ is not distorted (Figure 3.1), there exists an unique face point x_σ inside the face $\sigma \in \mathcal{F}_{K,L}$. On the contrary, if the primary mesh $\mathcal{M}$ is distorted (Figure 3.2), then there are some face points outside the face $\sigma \in \mathcal{F}_{K,L}$.

- If a face σ belongs to $\mathcal{F}_{K,\partial\Omega} \subset \mathcal{F}_{ext}$, then the face point x_σ is an isobarycenter of σ.

Furthermore, let be an edge e belonging to $\mathcal{E}_{ext}$, we define the boundary edge point x_e, as follows:

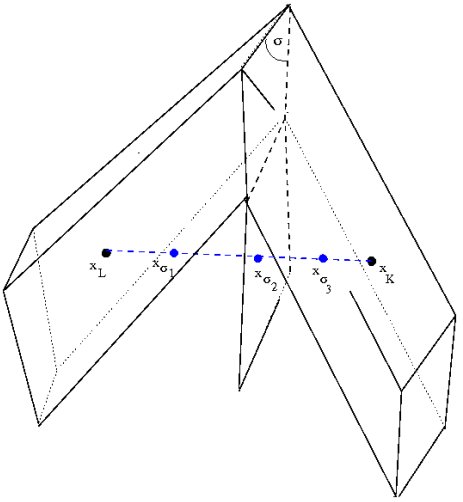


Figure 3.2: the three intersecting points x_{σ_1}, x_{σ_2} and x_{σ_3} on the distorted primary mesh $\mathcal{M}$.

- On the planar boundary $\partial\Omega$, the boundary edge point x_e is an intersecting point between a segment $[x_{\sigma_1}, x_{\sigma_2}]$ and an edge e (Figure 3.4), where σ_1 and σ_2 are the two neighbouring elements of $\mathcal{F}_{ext}$ and the edge e also belongs to the boundary $\partial\Omega$.
- On the non planar boundary $\partial\Omega$, the boundary edge point x_e is the midpoint of the edge e (Figure 3.5).

3.2.2 The 3D dual discretization family $\mathcal{D}^* = (\mathcal{M}^*, \mathcal{E}^*, \mathcal{P}^*, \mathcal{F}^*, \mathcal{V}^*)$

The set of dual vertices $\mathcal{V}^*$ is defined in the two following cases:

- With the planar boundary $\partial\Omega$, the set $\mathcal{V}^*$ is equal to $\mathcal{P} \cup \{x_\sigma | \sigma \in \mathcal{F}_{ext}\}$.
- With the non planar boundary $\partial\Omega$, the set $\mathcal{V}^*$ contains $\mathcal{P} \cup \{x_\sigma | \sigma \in \mathcal{F}_{ext}\}$ and the boundary edge points.

To construct the dual faces, we connect all points of $\mathcal{V}^*$, as follows:

- For each edge $e \in \mathcal{E}_{int}$, the dual face σ_e^* is defined by connecting mesh points of all elements in $\mathcal{M}_e$ (Figure 3.3). The point x_e is an intersecting point between e and σ_e^*.
- For each edge $e \in \mathcal{E}_{ext}$, the definition of the dual face σ_e^* bases on the two types of the boundary $\partial\Omega$: If the edge e belongs to the planar boundary $\partial\Omega$, the dual face σ_e^* is built by connecting all points in $\{x_K | K \in \mathcal{M}_e\} \cup \{x_\sigma | \sigma \in \mathcal{F}_{ext}^e\}$ (Figure 3.4). If the edge e belongs to the non planar boundary $\partial\Omega$, the dual face σ_e^* is built by connecting all points in $\{x_K | K \in \mathcal{M}_e\} \cup \{x_\sigma | \sigma \in \mathcal{F}_{ext}^e\} \cup \{x_e\}$ (Figure 3.5).

The set $\mathcal{F}^*$ is the set of all dual faces. We assume that all dual faces in $\mathcal{F}^*$ are inside $\overline{\Omega}$, and each dual face is star shaped. Therefore, the dual mesh $\mathcal{M}^*$ which can be defined by

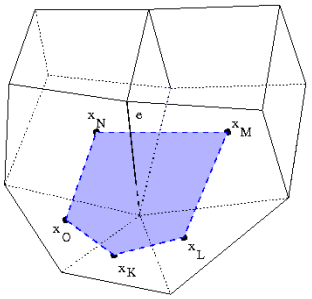


Figure 3.3: the dual face $(x_K, x_L, x_M, x_N, x_O)$ of the dual mesh $\mathcal{M}^*$.

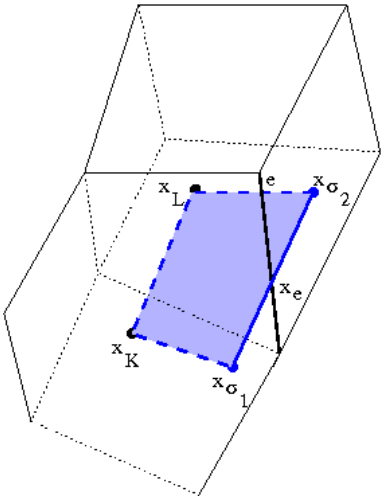


Figure 3.4: an element $(x_{\sigma_1}, x_K, x_L, x_{\sigma_2})$ of $\mathcal{F}^*$, where e is on the planar boundary $\partial\Omega$.

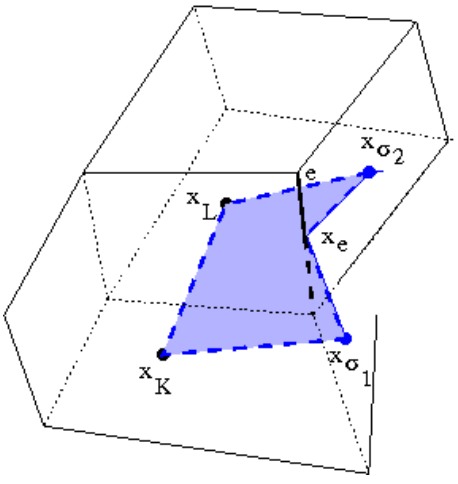


Figure 3.5: an element $(x_{\sigma_1}, x_K, x_L, x_{\sigma_2})$ of $\mathcal{F}^*$, where e is on the non planar boundary $\partial\Omega$.

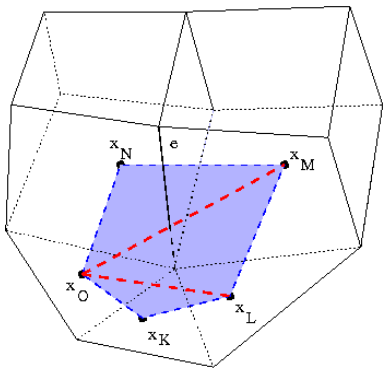


Figure 3.6: The triangles (x_O, x_K, x_L), (x_O, x_L, x_M) and (x_O, x_M, x_N) are three elements of $\mathcal{F}^{**}$, with $e \in \mathcal{E}_{\text{int}}$.

the set of all dual faces $\mathcal{F}^*$, is the set of dual control volumes K^* such that $\bigcup_{K^* \in \mathcal{M}^*} \overline{K}^* = \overline{\Omega}$. We associate to each control volume K^* a dual mesh point x_{K^*}. The dual mesh point x_{K^*} which is defined by either an inside point of K^* or a vertex of K^* can be connected to the other vertices of K^*. Additionally, the set of all dual mesh points is denoted by $\mathcal{P}^*$. The two sets $\mathcal{E}^*$ and $\mathcal{V}^*$ are respectively the sets of edges and vertices of all dual faces in $\mathcal{F}^*$.

3.2.3 The 3D third discretization family $\mathcal{D}^{**} = (\mathcal{M}^{**}, \mathcal{E}^{**}, \mathcal{P}^{**}, \mathcal{F}^{**}, \mathcal{V}^{**})$

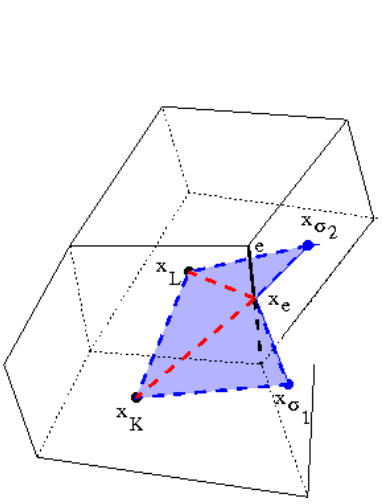
The third mesh $\mathcal{M}^{**}$ which is the tetrahedral mesh is constructed, as follows:
For each dual control volume K^* and each dual face σ^* of K^*, because the dual face σ^* is star shaped, there exists a point used to connected to the other vertices of the face σ^*. This implies that the dual face σ^* is partitioned by triangles, and the set of these triangles is denoted by $\mathcal{F}^{**}$. The two sets of vertices and edges of all triangles in $\mathcal{F}^{**}$ are respectively associated to the notations $\mathcal{V}^{**}$ and $\mathcal{E}^{**}$.

Remark 3.2.3: The set $\mathcal{E}^*$ is a subset of the set $\mathcal{E}^{**}$.

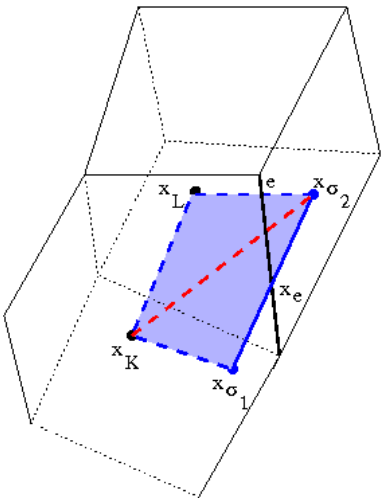
Furthermore, we have Figure-examples (3.6)-(3.7) for some elements of $\mathcal{M}^{**}$, $\mathcal{E}^{**}$, $\mathcal{P}^{**}$, $\mathcal{F}^{**}$, $\mathcal{V}^{**}$. These figures correspond to each location of the edge e in $\overline{\Omega}$:

- the edge e is inside the domain Ω Figure 3.6,
- the edge e is on the boundary $\partial\Omega$ Figure 3.7(a) and 3.7(b).

We then connect the dual mesh point x_{K^*} of K^* to all vertices of σ^*, so tetrahedra are constructed from the dual mesh point x_{K^*} and triangles on the dual face σ^*: these tetrahedra are elements of the third mesh $\mathcal{M}^{**}$ (see Figure 3.8).



(a) The triangles (x_K, x_{σ_1}, x_e), (x_K, x_L, x_e) and (x_L, x_{σ_1}, x_e) are three elements of $\mathcal{F}^{**}$, where e is on the non planar boundary $\partial\Omega$.



(b) the two elements $(x_K, x_{\sigma_1}, x_{\sigma_2})$ and (x_K, x_L, x_{σ_2}) belong to $\mathcal{F}^*$, where e is on the planar boundary $\partial\Omega$.

Figure 3.7:

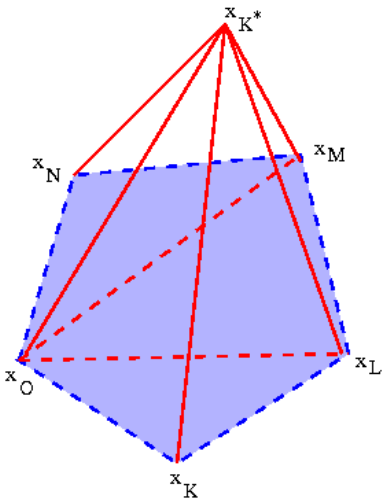


Figure 3.8: the three tetrahedra (x_{K^*}, x_K, x_L, x_O), (x_{K^*}, x_L, x_M, x_O) and (x_{K^*}, x_M, x_N, x_O) of $\mathcal{M}^{**}$.

3.3 Presentation of the 3D FECC scheme.

Firstly, we recall the following definitions:

- The discrete function space $\mathcal{H}_{\mathcal{D}}$ is the set of all vectors $\left((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*}\right)$, $u_K \in \mathbb{R}$, $K \in \mathcal{M}$, $u_{K^*} \in \mathbb{R}$, $K^* \in \mathcal{M}^*$ and $u_{K^*} = 0$ if x_{K^*} is on the boundary $\partial\Omega$.
- The function $P(v)$ is defined from the value $v = \left((v_K)_{K\in\mathcal{M}}, (v_{K^*})_{K^*\in\mathcal{M}^*}\right)$.
- The gradient operator ∇u is discretized by $\nabla_{\mathcal{D},\Lambda} u$. The discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ takes into account $\left((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*}\right) \in \mathcal{H}_{\mathcal{D}}$.

3.3.1 Isotropic homogeneous cases in 3D.

The function $P(u)$ and the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ are defined by:

$$P(u)(x) = \sum_{K\in\mathcal{M}} u_K p_K(x) + \sum_{K^*\in\mathcal{M}^*} u_{K^*} p_{K^*}(x), \tag{3.1}$$

$$\nabla_{\mathcal{D},\Lambda} u(x) = \sum_{K\in\mathcal{M}} u_K . \nabla p_K(x) + \sum_{K^*\in\mathcal{M}^*} u_{K^*} . \nabla p_{K^*}(x), \tag{3.2}$$

the function p_K (resp. p_{K^*}) which is a piecewise linear continuous function, is equal to 1 at x_K (resp. x_{K^*}) and 0 at the other primary mesh and dual mesh points.

3.3.2 Anisotropic heterogeneous cases in 3D.

To simplify the description of the 3D FECC scheme, we assume that, for neighbouring control volumes, the line joining their primary mesh points intersects their common face.

Without loss of generality, we use Figure 3.9 to present the definitions of the discrete gradient $\nabla_{D,\Lambda} u$: The discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ must be taken into account of the heterogeneity of Λ and considered on the two neighbouring tetrahedra (x_{K^*}, x_A, x_B, x_D) and (x_{K^*}, x_B, x_C, x_D).

Let us consider the tetrahedron (x_{K^*}, x_B, x_C, x_D). It is partitioned by the four polyhedra $(x_{K^*}, x_B, x_{\sigma_2}, x_e, x_{\sigma_6})$, $(x_{K^*}, x_C, x_{\sigma_1}, x_e, x_{\sigma_2})$, $(x_{K^*}, x_D, x_{\sigma_1}, x_e, x_{\sigma_5})$, $(x_e, x_{\sigma_5}, x_{\sigma_6}, x_{K^*})$. The discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ is defined for each sub-polyhedron belonging to the tetrahedron (x_{K^*}, x_B, x_C, x_D) :

$$\begin{aligned}
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_B,x_{\sigma_2},x_e,x_{\sigma_6})} &= (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_B,x_{\sigma_e},x_{\sigma_2})} = (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_B,x_{\sigma_e},x_{\sigma_6})}, \\
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_{\sigma_1},x_e,x_{\sigma_2})} &= (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_{\sigma_e},x_{\sigma_1})} = (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_{\sigma_e},x_{\sigma_2})}, \\
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_D,x_{\sigma_1},x_e,x_{\sigma_5})} &= (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_D,x_{\sigma_e},x_{\sigma_1})} = (\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_D,x_{\sigma_e},x_{\sigma_5})}.
\end{aligned}$$

We obtain the formulas of the discrete gradients $(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_B,x_e,x_{\sigma_2})}$, $(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_e,x_{\sigma_1})}$, $(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_e,x_{\sigma_2})}$ and $(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_D,x_e,x_{\sigma_1})}$, as follows:

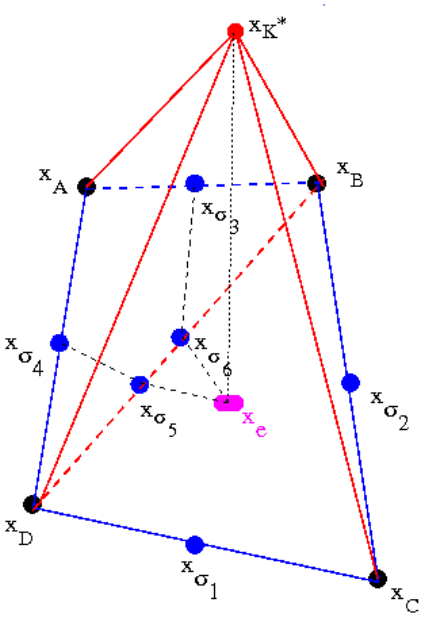


Figure 3.9: the two neighbouring tetrahedra (x_{K^*}, x_A, x_B, x_D) and (x_{K^*}, x_B, x_C, x_D).

$$
\begin{aligned}
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_B,x_e,x_{\sigma_2})} &= \frac{-u_B n^B_{(x_{K^*},x_e,x_{\sigma_2})} - u_e^{K^*} n^B_{(x_{K^*},x_B,x_{\sigma_2})} - u_{\sigma_2}^{K^*} n^B_{(x_{K^*},x_B,x_e)} - u_{K^*} n^B_{(x_B,x_e,x_{\sigma_2})}}{3m_{(x_{K^*},x_B,x_e,x_{\sigma_2})}},\\
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_e,x_{\sigma_1})} &= \frac{-u_C n^C_{(x_{K^*},x_e,x_{\sigma_1})} - u_e^{K^*} n^C_{(x_{K^*},x_C,x_{\sigma_1})} - u_{\sigma_1}^{K^*} n^{C,\sigma_1}_{(x_{K^*},x_C,x_e)} - u_{K^*} n^C_{(x_C,x_e,x_{\sigma_1})}}{3m_{(x_{K^*},x_C,x_e,x_{\sigma_1})}},\\
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_C,x_e,x_{\sigma_2})} &= \frac{-u_C n^C_{(x_{K^*},x_e,x_{\sigma_2})} - u_e^{K^*} n^C_{(x_{K^*},x_C,x_{\sigma_2})} - u_{\sigma_2}^{K^*} n^{C,\sigma_2}_{(x_{K^*},x_C,x_e)} - u_{K^*} n^C_{(x_C,x_e,x_{\sigma_2})}}{3m_{(x_{K^*},x_C,x_e,x_{\sigma_2})}},\\
(\nabla_{\mathcal{D},\Lambda} u)|_{(x_{K^*},x_D,x_e,x_{\sigma_1})} &= \frac{-u_D n^D_{(x_{K^*},x_e,x_{\sigma_1})} - u_e^{K^*} n^D_{(x_{K^*},x_D,x_{\sigma_1})} - u_{\sigma_1}^{K^*} n^D_{(x_{K^*},x_D,x_e)} - u_{K^*} n^D_{(x_D,x_e,x_{\sigma_1})}}{3m_{(x_{K^*},x_D,x_e,x_{\sigma_1})}}.
\end{aligned}
$$

We remark that if the edges σ_1, σ_2, e belong to the boundary $\partial\Omega$, $u_{\sigma_1}^{K^*}$, $u_{\sigma_2}^{K^*}$, $u_e^{K^*}$ are equal to 0.

The notations which are used in the above formulas are defined in Figure 3.10, where the vectors $n_{(x_A,x_B,x_C)} = \frac{1}{2}(x_B - x_A) \times (x_C - x_A)$, $n_{(x_A,x_B,x_D)} = \frac{1}{2}(x_B - x_A) \times (x_D - x_A)$, $n_{(x_A,x_C,x_D)} = \frac{1}{2}(x_C - x_A) \times (x_D - x_A)$ and $n_{(x_B,x_C,x_D)} = \frac{1}{2}(x_C - x_B) \times (x_D - x_B)$. The measure $m_{(x_A,x_B,x_C,x_D)} > 0$ is the volume of (x_A, x_B, x_C, x_D).

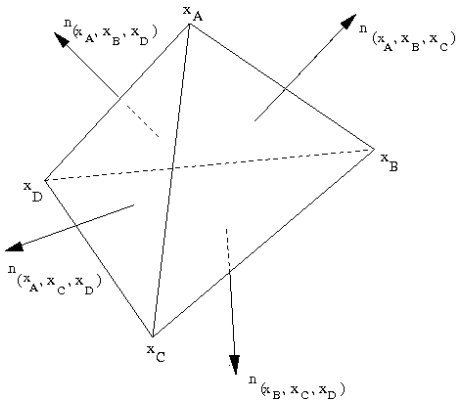


Figure 3.10: A tetrahedron and its outward normal vectors.

Additionally, we introduce a new auxiliary edge unknowns $u_e^{K^*}$ with $e \in \mathcal{E}_{int}$, computed by:

$$u_e^{K^*} = \alpha_{\sigma_1}^e . u_{\sigma_1}^{K^*} + \alpha_{\sigma_2}^e . u_{\sigma_2}^{K^*} + (1 - \alpha_{\sigma_1}^e - \alpha_{\sigma_2}^e) . u_C,$$

$$\Leftrightarrow \qquad u_e^{K^*} - \alpha_{\sigma_1}^e . u_{\sigma_1}^{K^*} - \alpha_{\sigma_2}^e . u_{\sigma_2}^{K^*} = (1 - \alpha_{\sigma_1}^e - \alpha_{\sigma_2}^e) . u_C, \tag{3.3}$$

where the two coefficients $\alpha_{\sigma_1}^e$ and $\alpha_{\sigma_2}^e$ satisfy

$$x_e - x_B = \alpha_{\sigma_1}^e (x_{\sigma_1} - x_B) + \alpha_{\sigma_2}^e (x_{\sigma_2} - x_B).$$

We then apply the Local Conservativity of the Fluxes:

- on the tetrahedron (x_{K^*}, x_B, x_C, x_e)

$$\left[\Lambda_B \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*}, x_B, x_e, x_{\sigma_2})}\right] . n^B_{(x_{K^*}, x_e, x_{\sigma_2})} + \left[\Lambda_C \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*}, x_C, x_e, x_{\sigma_2})}\right] . n^C_{(x_{K^*}, x_e, x_{\sigma_2})} = 0.$$

$$\Leftrightarrow u_{\sigma_2}^{K^*} - \beta_e^{K^*,\sigma_2} u_e = \beta_B^{K^*,\sigma_2} u_B + \beta_C^{K^*,\sigma_2} u_C + \beta_{K^*}^{K^*,\sigma_2} u_{K^*}. \tag{3.4}$$

- on the tetrahedron (x_{K^*}, x_C, x_D, x_e)

$$\left[\Lambda_C \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*}, x_C, x_e, x_{\sigma_1})}\right] . n^C_{(x_{K^*}, x_e, x_{\sigma_1})} + \left[\Lambda_D \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*}, x_D, x_e, x_{\sigma_1})}\right] . n^D_{(x_{K^*}, x_e, x_{\sigma_1})} = 0.$$

$$\Leftrightarrow u_{\sigma_1}^{K^*} - \beta_e^{K^*,\sigma_1} u_e = \beta_C^{K^*,\sigma_1} u_C + \beta_D^{K^*,\sigma_1} u_D + \beta_{K^*}^{K^*,\sigma_1} u_{K^*}. \tag{3.5}$$

From (3.3) – (3.5), we get the following system of linear equations:

$$\begin{pmatrix} 1 & -\alpha_{\sigma_1}^e & -\alpha_{\sigma_2}^e \\ -\beta_e^{K^*,\sigma_1} & 1 & 0 \\ -\beta_e^{K^*,\sigma_2} & 0 & 1 \end{pmatrix} \begin{pmatrix} u_e \\ u_{\sigma_1}^{K^*} \\ u_{\sigma_2}^{K^*} \end{pmatrix} = \begin{pmatrix} 0 & 1 - \alpha_{\sigma_1}^e - \alpha_{\sigma_2}^e & 0 & 0 \\ 0 & \beta_C^{K^*,\sigma_1} & \beta_D^{K^*,\sigma_1} & \beta_{K^*}^{K^*,\sigma_1} \\ \beta_B^{K^*,\sigma_2} & \beta_C^{K^*,\sigma_2} & 0 & \beta_{K^*}^{K^*,\sigma_2} \end{pmatrix} \begin{pmatrix} u_B \\ u_C \\ u_D \\ u_{K^*} \end{pmatrix}.$$

Hypothesis 3.1: We assume that

$$\det \begin{pmatrix} 1 & -\alpha_{\sigma_1}^e & -\alpha_{\sigma_2}^e \\ -\beta_e^{K^*,\sigma_1} & 1 & 0 \\ -\beta_e^{K^*,\sigma_2} & 0 & 1 \end{pmatrix} \neq 0.$$

Hence, the auxiliary unknowns $u_e^{K^*}$, $u_{\sigma_1}^{K*}$ and $u_{\sigma_2}^{K*}$ are eliminated by linear combinations depending on u_B, u_C, u_D and u_{K^*}.

Using again the Local Conservativity of the Fluxes:

$$\begin{cases} \left[\Lambda_D \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_D,x_e,x_{\sigma_5})}\right].n^D_{(x_{K^*},x_e,x_{\sigma_5})} + \left[\Lambda_A \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_e,x_{\sigma_5},x_{\sigma_6})}\right].n^A_{(x_{K^*},x_e,x_{\sigma_5})} = 0, \\ \left[\Lambda_B \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_B,x_e,x_{\sigma_6})}\right].n^B_{(x_{K^*},x_e,x_{\sigma_6})} + \left[\Lambda_A \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_e,x_{\sigma_5},x_{\sigma_6})}\right].n^A_{(x_{K^*},x_e,x_{\sigma_6})} = 0, \end{cases}$$

and the linear combination for eliminating $u_e^{K^*}$, we can compute $u_{\sigma_5}^{K*}$ and $u_{\sigma_6}^{K*}$ with linear combinations depending on u_B, u_C, u_D and u_{K^*}.

Thanks to the above linear combinations and the Local Conservativity of the Fluxes:

$$\begin{aligned} \left[\Lambda_B \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_B,x_{\sigma_3},x_{\sigma_6})}\right].n^B_{(x_{K^*},x_{\sigma_3},x_{\sigma_6})} + \left[\Lambda_A \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_A,x_{\sigma_3},x_{\sigma_6})}\right].n^A_{(x_{K^*},x_{\sigma_3},x_{\sigma_6})} &= 0, \\ \left[\Lambda_D \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_D,x_{\sigma_4},x_{\sigma_5})}\right].n^D_{(x_{K^*},x_{\sigma_4},x_{\sigma_5})} + \left[\Lambda_A \left(\nabla_{\mathcal{D},\Lambda} u\right)\Big|_{(x_{K^*},x_A,x_{\sigma_4},x_{\sigma_5})}\right].n^A_{(x_{K^*},x_{\sigma_4},x_{\sigma_5})} &= 0, \end{aligned}$$

we can transform $u_{\sigma_3}^{K*}$ and $u_{\sigma_4}^{K*}$ into linear combinations depending on u_A, u_B, u_C, u_D and u_{K^*}.

Therefore, we can reconstruct the definition of the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ which only depends on $(u_K)_{K\in\mathcal{M}}$ and $(u_{K^*})_{K^*\in\mathcal{M}^*}$.

3.4 Construction in three dimensions.

The main idea of the 3D FECC scheme which is also based on the standard finite element method (P_1) on the third tetrahedral meshes is presented by the three following steps:

Step 1 *Recover all u_{K^*} with $K^* \in \mathcal{M}^*$ by linear functions of $(u_K)_{K\in\mathcal{M}}$ and constants depending on the function f.*

For each $K^* \in \mathcal{M}^*$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*})$ such that $v_L = 0$ for all $L \in \mathcal{M}$, $v_{L^*} = 0$ if $L^* \neq K^*$ and $v_{K^*} = 1$ in equation (3.3)

$$\int_\Omega \left(\Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x)\right).\nabla_{\mathcal{D},\Lambda} v(x)\, dx = \int_\Omega f(x) P(v)(x) dx.$$

The discrete gradient $\nabla_{\mathcal{D},\Lambda} v$ is equal to 0 on L^* which is different from K^*. We deduce that $\int_\Omega \left(\Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x)\right).\nabla_{\mathcal{D},\Lambda} v(x)\, dx$ is presented by the linear function of $(u_K)_{K\in\mathcal{M}}$, u_{K^*}

and a constant depending on function f. Therefore, the unknown u_{K^*} is computed by a linear function of $\{u_K\}_{K\in\mathcal{M}}$ and a constant depending on function f. This linear function is denoted by $\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)$. In isotropic homogeneous cases, $\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)$ is formulated by:

$$\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right) = -\sum_{K\in\mathcal{M}} u_K \int_\Omega \nabla p_K(x) \frac{\nabla p_{K^*}(x)}{\|\nabla p_{K^*}\|^2_{L^2(\Omega)}}\, dx + \int_\Omega f(x) \frac{p_{K^*}(x)}{\|\nabla p_{K^*}\|^2_{L^2(\Omega)}}\, dx.$$

Step 2 *Reconstruct the discrete gradient* $\nabla_{\mathcal{D},\Lambda} u$.

In the definition of the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$, we transform all the unknowns $\{u_{K^*}\}_{K^*\in\mathcal{M}^*}$ into $\left\{\Pi_{K^*}\left(\{u_K\}_{K\in\mathcal{M}}, f\right)\right\}_{K^*\in\mathcal{M}^*}$. Hence, $\nabla_{\mathcal{D},\Lambda} u$ does not depend on unknowns $\{u_{K^*}\}_{K^*\in\mathcal{M}^*}$.

Step 3 *Construct a system of linear equations.*

In equation (3.3), for each $K \in \mathcal{M}$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*}) \in \mathcal{H}_{\mathcal{D}}$ such that $v_{L^*} = 0$ for all $L^* \in \mathcal{M}^*$, $v_L = 0$ if $L \neq K$ and $v_K = 1$. This resulting equation is a linear equation that only involves unknowns $\{u_K\}_{K\in\mathcal{M}}$. Thus, we construct a system of linear equations

$$A.U = B, \tag{3.6}$$

where U is the vector $(u_K)_{K\in\mathcal{M}}$ and A is a square matrix in $\mathbb{R}^{card(\mathcal{M})\times card(\mathcal{M})}$.

Part III

Mathematical properties

Chapter 4

Properties of the FECC scheme

Contents

4.1 Motivation.

In the fourth chapter, we show the main properties of the FECC scheme in isotropic homogeneous cases and discontinuous anisotropic cases.

- Symmetry and positive definiteness: these properties are interested in building discretization schemes for diffusing flows in heterogeneous anisotropic porous media. Multipoint schemes which are studied in [1], [2], [3], [16] and [27] do not satisfy these properties. Moreover, the two properties allow us to use efficient methods to solve the system of linear equations.

- Small stencil: the FECC scheme is cell-centered scheme. Its stencil is compact, equal or less than nine on quadrangular meshes and twenty seven on hexahedral meshes.

- We recover the exact solution if Λ is piecewise constant in polygonal sub-domains and u (the solution of the diffusion problem (2.1)) is affine in each of these sub-domains.

- The scheme is convergent in discontinuous anisotropic heterogeneous cases, where the tensor Λ is piecewise Lipschitz-continuous.

4.2 Symmetry and positive definiteness of the FECC scheme.

Lemma 4.1 *With hypothesis (2.1), for anisotropic heterogeneous cases, the matrix A of the system (2.12) is symmetric and positive definite on general meshes.*

Proof of lemma 4.1

By definition, the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ depends on elements of sets $\{u_K\}_{K\in\mathcal{M}}$, $\{u_{K^*}\}_{K^*\in\mathcal{M}^*}$ and $\left\{u_\sigma^{K^*}\right\}_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}}$. The set $\left\{u_\sigma^{K^*}\right\}_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}}$ is only considered in anisotropic heterogeneous cases. Hence, we can present

$$\int_\Omega \left(\Lambda\nabla_{\mathcal{D},\Lambda} u\right).\nabla_{\mathcal{D},\Lambda} v dx = \mathcal{U}^T A_\Lambda \mathcal{V},$$

where $\mathcal{U}$, $\mathcal{V}$ are defined by

$$\mathcal{U} = \begin{pmatrix} (u_{K^*})_{K^*\in\mathcal{M}^*} \\ (u_K)_{K\in\mathcal{M}} \\ \left(u_\sigma^{K^*}\right)_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}} \end{pmatrix}, \qquad \mathcal{V} = \begin{pmatrix} (v_{K^*})_{K^*\in\mathcal{M}^*} \\ (v_K)_{K\in\mathcal{M}} \\ \left(v_\sigma^{K^*}\right)_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}} \end{pmatrix},$$

$$\mathcal{M}_\sigma^* = \{K^*\in\mathcal{M}^* \text{ such that } \sigma\cap K^*\neq\emptyset\},$$

with

$$\begin{aligned}(u_K)_{K\in\mathcal{M}}\,,\ (v_K)_{K\in\mathcal{M}} &\in M^{card(\mathcal{M})\times 1},\\ (u_{K^*})_{K^*\in\mathcal{M}^*}\,,\ (v_{K^*})_{K^*\in\mathcal{M}^*} &\in M^{card(\mathcal{M}^*)\times 1},\end{aligned}$$

$$\left(u_\sigma^{K^*}\right)_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}}\,,\ \left(v_\sigma^{K^*}\right)_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ K^*\in\mathcal{M}_\sigma^*}} \in M^{\left\{\sum\limits_{\sigma\in\mathcal{E}_{\rm int}} card(\mathcal{M}_\sigma^*)\right\}\times 1}.$$

Setting $m = card(\mathcal{M}) + card(\mathcal{M}^*) + \sum\limits_{\sigma\in\mathcal{E}_{\rm int}} card(\mathcal{M}_\sigma^*)$ and $n = card(\mathcal{M}) + card(\mathcal{M}^*)$, we obtain

$$A_\Lambda = ((a_\Lambda)_{ij})_{i,j\in\overline{1,m}} \quad \text{and} \quad \mathcal{U},\mathcal{V}\in M^{m\times 1}.$$

Moreover, A_Λ is symmetric. This is implied from $\int\limits_\Omega (\Lambda\nabla_{D,\Lambda}u)\,.\nabla_{D,\Lambda}v dx = \int\limits_\Omega \nabla_{D,\Lambda}u.\,(\Lambda\nabla_{D,\Lambda}v)\,dx$, because the tensor Λ is symmetric.

In anisotropic heterogeneous cases, the discrete gradient $\nabla_{\mathcal{D},\Lambda}u$ can be re-written by

- on the triangle (x_{K^*}, x_K, x_σ)

$$\nabla_{D,\Lambda}u = -\frac{\left[\begin{array}{l}\left(\beta_K^{K^*,\sigma}n_{[x_{K^*},x_K]} + n^K_{[x_\sigma,x_{K^*}]}\right)u_K + \left(\beta_L^{K^*,\sigma}n_{[x_{K^*},x_K]}\right)u_L\\ +\left(\beta_{K^*}^{K^*,\sigma}n_{[x_{K^*},x_K]} + n_{[x_\sigma,x_K]}\right)u_{K^*}\end{array}\right]}{2m_{(x_{K^*},x_K,x_\sigma)}}, \tag{4.1}$$

- on the triangle (x_{K^*}, x_L, x_σ)

$$\nabla_{D,\Lambda}u = -\frac{\left[\begin{array}{l}\left(\beta_K^{K^*,\sigma}n_{[x_{K^*},x_L]}\right)u_K + \left(\beta_L^{K^*,\sigma}n_{[x_{K^*},x_L]} + n^L_{[x_\sigma,x_{K^*}]}\right)u_L\\ +\left(\beta_{K^*}^{K^*,\sigma}n_{[x_{K^*},x_L]} + n_{[x_\sigma,x_L]}\right)u_{K^*}\end{array}\right]}{2m_{(x_{K^*},x_L,x_\sigma)}}, \tag{4.2}$$

because $u_\sigma^{K^*} = \beta_K^{K^*,\sigma}u_K + \beta_L^{K^*,\sigma}u_L + \beta_{K^*}^{K^*,\sigma}u_{K^*}$.

Now, the discrete gradient $\nabla_{\mathcal{D},\Lambda}u$ depends on elements of set $\left\{(u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*}\right\}$ in general cases. Therefore, there exists a matrix $C^*\in M^{m\times n}$ such that $\mathcal{U} = C^*\mathcal{U}^*$ with $\mathcal{U}\neq 0$, which implies

$$\mathcal{U}^T A_\Lambda \mathcal{U} = (\mathcal{U}^*)^T\left(C^{*T}A_\Lambda C^*\right)\mathcal{U}^*,$$

where $C^*\in M^{m\times n}$ and $\mathcal{U}^* = \begin{pmatrix}(u_{K^*})_{K^*\in\mathcal{M}^*}\\ (u_K)_{K\in\mathcal{M}}\end{pmatrix}\in M^{n\times 1}$.

As the matrix A_Λ is symmetric, the matrix $G = C^{*T}A_\Lambda C^*$ is also symmetric.

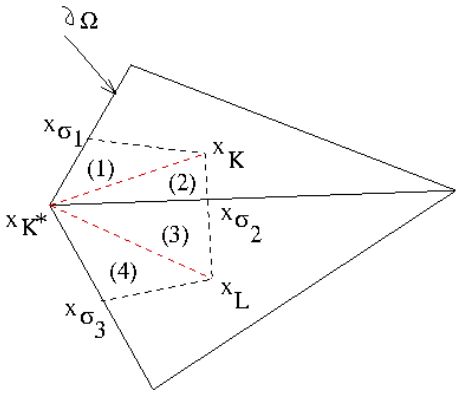


Figure 4.1: an example triangle $T_0 = (x_{K^*}, x_{\sigma_1}, x_K) \in \mathcal{M}^{**}$.

According to step 1 of the construction scheme, for each $K^* \in \mathcal{M}^*$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*})$ such that $v_L = 0$ for all $L \in \mathcal{M}$, $v_{L^*} = 0$ if $L^* \neq K^*$ and $v_{K^*} = 1$ in equation (2.3)

$$\int_\Omega (\Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x)) . \nabla_{\mathcal{D},\Lambda} v(x)\, dx = \int_\Omega f(x) P(v)(x) dx.$$

From these linear equations, the first system of linear equations is constructed by:

$$DU^* + EU = F^*, \tag{4.3}$$

where $F^* \in M^{card(\mathcal{M}^*)\times 1}$, $D \in M^{card(\mathcal{M}^*)\times card(\mathcal{M}^*)}$ and $E \in M^{card(\mathcal{M}^*)\times card(\mathcal{M})}$.

According to step 3 of the construction scheme, for each $K \in \mathcal{M}$, we choose $v = (\{v_L\}_{L\in\mathcal{M}}, \{v_{L^*}\}_{L^*\in\mathcal{M}^*}) \in \mathcal{H}_\mathcal{D}$ such that $v_{L^*} = 0$ for all $L^* \in \mathcal{M}^*$, $v_L = 0$ if $L \neq K$ and $v_K = 1$ in equation (2.3). We get the second system of linear equations, as follows:

$$MU^* + NU = F, \tag{4.4}$$

where $F \in M^{card(\mathcal{M})\times 1}$, $N \in M^{card(\mathcal{M})\times card(\mathcal{M})}$ and $M \in M^{card(\mathcal{M})\times card(\mathcal{M}^*)}$. Both F and F^* depend on function f.

From (4.3) and (4.4), it follows that $G = \begin{pmatrix} D & E \\ M & N \end{pmatrix}$ where $M = E^T$ and two square matrices D, N are symmetric, because the matrix G is symmetric.

Next, we prove that the matrix G is positive definite. Assume that $\mathcal{U}^* \neq 0$, there are two following cases:
In the first case where $u_K \neq 0$ for all $K \in \mathcal{M}$, we consider $T_0 = (x_{K^*}, x_{\sigma_1}, x_K) \in \mathcal{M}^{**}$. This triangle has an edge belonging to the boundary of Ω as Figure 4.1. On the triangle $T_0 = (x_{K^*}, x_{\sigma_1}, x_K)$, the discrete gradient $\nabla_{D,\Lambda} u$ is defined by

$$\nabla_{D,\Lambda} u = \frac{-u_K n_{[x_{K^*}, x_{\sigma_1}]}}{2m_{(x_{K^*}, x_{\sigma_1}, x_K)}},$$

because $u_{K^*} = u_{\sigma_1}^{K^*} = 0$.
All eigenvalues of tensor Λ are equal or greater than $\underline{\lambda} > 0$, thus

$$\begin{aligned}\mathcal{U}^{*T} G \mathcal{U}^* = \int_\Omega \left(\Lambda \nabla_{\mathcal{D},\Lambda} u\right).\nabla_{\mathcal{D},\Lambda} u dx &\geq \underline{\lambda} \int_\Omega \left(\nabla_{\mathcal{D},\Lambda} u\right)^2 dx \\ &\geq \underline{\lambda} \int_{T_0} \left(\nabla_{D,\Lambda} u\right)^2 dx = \underline{\lambda} \frac{\left(u_K n_{[x_{K^*}, x_{\sigma_1}]}\right)^2}{4 m_{(x_{K^*}, x_{\sigma_1}, x_K)}} > 0.\end{aligned}$$

In the second case, there exists $K \in \mathcal{M}$ such that $u_K = 0$. In this case, we have a triangle $T_0 = (x_{K^*}, x_K, x_L)$ such that $u_L \neq 0$ or $u_{K^*} \neq 0$ (see the figure 3.1).
The integral $\int\limits_{T_0} \left(\nabla_{D,\Lambda} u\right)^2 dx$ is computed by:

$$\int_{T_0} \left(\nabla_{D,\Lambda} u\right)^2 dx = \int_{(x_{K^*}, x_\sigma, x_K)} \left(\nabla_{D,\Lambda} u\right)^2 dx + \int_{(x_{K^*}, x_\sigma, x_L)} \left(\nabla_{D,\Lambda} u\right)^2 dx.$$

On the triangle (x_{K^*}, x_σ, x_K), the discrete gradient $\nabla_{\mathcal{D},\Lambda} u$ is defined by

$$\nabla_{D,\Lambda} u = \frac{-u_{K^*}\left(n_{[x_K, x_\sigma]} + \beta_{K^*}^{K^*,\sigma} n_{[x_K, x_{K^*}]}\right) - u_L \beta_L^{K^*,\sigma} n_{[x_K, x_{K^*}]}}{2 m_{(x_{K^*}, x_\sigma, x_K)}}.$$

It follows that

$$\int_{(x_{K^*}, x_\sigma, x_K)} \left(\nabla_{D,\Lambda} u\right)^2 dx = \frac{\left\{u_{K^*}\left(n_{[x_K, x_\sigma]} + \beta_{K^*}^{K^*,\sigma} n_{[x_K, x_{K^*}]}\right) + u_L \beta_L^{K^*,\sigma} n_{[x_K, x_{K^*}]}\right\}^2}{4 m_{(x_{K^*}, x_\sigma, x_K)}} > 0,$$

since the direction of the vector $\left(n_{[x_K, x_\sigma]} + \beta_{K^*}^{K^*,\sigma} n_{[x_K, x_{K^*}]}\right)$ is different from direction of the vector $n_{[x_K, x_{K^*}]}$, $\beta_L^{K^*,\sigma} \neq 0$ (use hypothesis 3.1) and $\left[\begin{array}{l} u_L \neq 0, \\ u_{K^*} \neq 0. \end{array}\right.$
Similarly to the first case, we get

$$\mathcal{U}^{*T} G \mathcal{U}^* = \int_\Omega \left(\Lambda \nabla_{\mathcal{D},\Lambda} u\right).\nabla_{\mathcal{D},\Lambda} u dx \geq \underline{\lambda} \int_\Omega \left(\nabla_{\mathcal{D},\Lambda} u\right)^2 dx \geq \underline{\lambda} \int_{T_0} \left(\nabla_{D,\Lambda} u\right)^2 dx > 0.$$

Therefore, the matrix G is positive definite.

From (4.3), $\mathcal{U}^*$ is computed by $\mathcal{U}^* = D^{-1}\left(F^* - EU\right)$. In this formula, the matrix D^{-1} exists, because we apply the property of Schur complement of D in G (see more theorem 1.20, paper 44 in [28]) and G is symmetric, positive definite. Thus, (4.4) is transformed as follows

$$\underbrace{\left(N - E^T D^{-1} E\right)}_{=A} U = \underbrace{F - E^T D^{-1} F^*}_{=B},$$

where the matrices A, B are defined in the systems of linear equations (2.9) and (2.12).

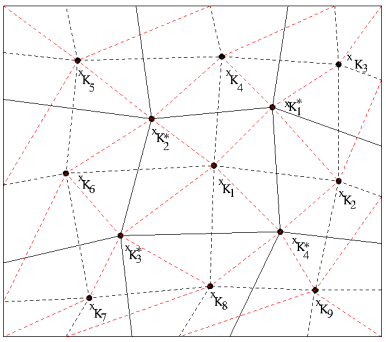


Figure 4.2: an example for the dual and the third meshes such that the stencil of the 2D FECC scheme is equal to 9.

Since G, D are symmetric, positive definite and using the property of Schur complement in G (see more theorem 1.12, paper 34 in [28]), we conclude that A is symmetric, positive definite. This allows us to use efficient solving methods to solve the systems of linear equations (2.9) and (2.12). $\square$

4.3 Isotropic homogeneous cases.

4.3.1 Small stencil.

Property 4.1 *The stencil of the FECC scheme is equal to* 9 *on quadrangular primary meshes.*

Proof of property 4.1

If the dual grid and the third grid are described by Figure 4.2, then the stencil is equal to 9. *Step 1*
The intersecting domains are not empty between $supp\{p_{K_1}\}$ and each of the following domains: $supp\{p_{K_2}\}$, $supp\{p_{K_4}\}$, $supp\{p_{K_6}\}$, $supp\{p_{K_8}\}$, $supp\{p_{K_1^*}\}$, $supp\{p_{K_2^*}\}$, $supp\{p_{K_3^*}\}$, $supp\{p^*_{K_4}\}$, and are empty between $supp\{p_{K_1}\}$ and the others: $supp\{p_{K_3}\}$, $supp\{p_{K_5}\}$, $supp\{p_{K_7}\}$, $supp\{p_{K_9}\}$. Therefore, equation (2.6) can be written as

$$\int_\Omega \left(\begin{array}{l} u_{K_1}\nabla p_{K_1} + u_{K_2}\nabla p_{K_2} + u_{K_4}\nabla p_{K_4} + u_{K_6}\nabla p_{K_6} + u_{K_8}\nabla p_{K_8} \\ +u_{K_1^*}\nabla p_{K_1^*} + u_{K_2^*}\nabla p_{K_2^*} + u_{K_3^*}\nabla p_{K_3^*} + u_{K_4^*}\nabla p_{K_4^*} \end{array} \right).\nabla p_{K_1} dx = \int_\Omega f.p_{K_1}\, dx.$$

Step 2

$$\begin{aligned}
u_{K_1^*} &= \alpha_1^{K_1^*} u_{K_1} + \alpha_2^{K_1^*} u_{K_2} + \alpha_3^{K_1^*} u_{K_3} + \alpha_4^{K_1^*} u_{K_4} + \alpha_{K_1^*}(f).\\
u_{K_2^*} &= \alpha_1^{K_2^*} u_{K_1} + \alpha_4^{K_2^*} u_{K_4} + \alpha_5^{K_2^*} u_{K_5} + \alpha_6^{K_2^*} u_{K_6} + \alpha_{K_2^*}(f).\\
u_{K_3^*} &= \alpha_1^{K_3^*} u_{K_1} + \alpha_6^{K_3^*} u_{K_6} + \alpha_7^{K_3^*} u_{K_7} + \alpha_8^{K_3^*} u_{K_8} + \alpha_{K_3^*}(f).\\
u_{K_4^*} &= \alpha_1^{K_4^*} u_{K_1} + \alpha_2^{K_4^*} u_{K_2} + \alpha_8^{K_4^*} u_{K_8} + \alpha_9^{K_4^*} u_{K_9} + \alpha_{K_4^*}(f).
\end{aligned}$$

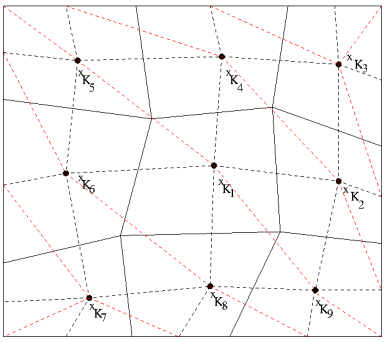


Figure 4.3: an example for the dual and the third meshes such that the stencil of the 2D FECC scheme is equal to 7.

Step 3

$$\int_\Omega \left(\begin{array}{l} u_{K_1}\nabla p_{K_1} + u_{K_2}\nabla p_{K_2} + u_{K_4}\nabla p_{K_4} + u_{K_6}\nabla p_{K_6} + u_{K_8}\nabla p_{K_8} \\ + \left(\alpha_1^{K_1^*} u_{K_1} + \alpha_2^{K_1^*} u_{K_2} + \alpha_3^{K_1^*} u_{K_3} + \alpha_4^{K_1^*} u_{K_4}\right) \nabla p_{K_1^*} \\ + \left(\alpha_1^{K_2^*} u_{K_1} + \alpha_4^{K_2^*} u_{K_4} + \alpha_5^{K_2^*} u_{K_5} + \alpha_6^{K_2^*} u_{K_6}\right) \nabla p_{K_2^*} \\ + \left(\alpha_1^{K_3^*} u_{K_1} + \alpha_6^{K_3^*} u_{K_6} + \alpha_7^{K_3^*} u_{K_7} + \alpha_8^{K_3^*} u_{K_8}\right) \nabla p_{K_3^*} \\ + \left(\alpha_1^{K_4^*} u_{K_1} + \alpha_2^{K_4^*} u_{K_2} + \alpha_8^{K_4^*} u_{K_8} + \alpha_9^{K_4^*} u_{K_9}\right) \nabla p_{K_4^*} \end{array} \right) .\nabla p_{K_1} dx =$$

$$\int_\Omega \left\{ f.p_{K_1} - \left[\alpha_{K_1^*}(f)\nabla p_{K_1^*} - \alpha_{K_2^*}(f)\nabla p_{K_2^*} - \alpha_{K_3^*}(f)\nabla p_{K_3^*} - \alpha_{K_4^*}(f)\nabla p_{K_4^*}\right].\nabla p_{K_1} \right\} dx.$$

This equation only depends on nine cell unknowns, so the stencil is equal to nine. □

Remark 4.1: In some particular cases, the stencil can be 7 or even 5. We show two following examples for these cases:

a) If the dual grid and the third grid are described by Figure 4.3, then the stencil is equal to 7. In Figure 4.2, the polygon $(x_{K_2}, x_{K_4}, x_{K_5}, x_{K_6}, x_{K_8}, x_{K_9})$ is an element of the dual grid where the mesh point x_{K^*} of this element coincides in x_{K_1}.
The intersecting domains are not empty between $supp\{p_{K_1}\}$ and each of the following domains: $supp\{p_{K_2}\}$, $supp\{p_{K_4}\}$, $supp\{p_{K_5}\}$, $supp\{p_{K_6}\}$, $supp\{p_{K_8}\}$, $supp\{p_{K_9}\}$, and are empty between $supp\{p_{K_1}\}$ and the others: $supp\{p_{K_3}\}$, $supp\{p_{K_7}\}$. Hence, equation (2.6) can be written as

$$\int_\Omega \left(\begin{array}{l} u_{K_1}\nabla p_{K_1} + u_{K_2}\nabla p_{K_2} + u_{K_4}\nabla p_{K_4} + u_{K_5}\nabla p_{K_5} \\ + u_{K_6}\nabla p_{K_6} + u_{K_8}\nabla p_{K_8} + u_{K_9}\nabla p_{K_9} \end{array} \right) .\nabla p_{K_1} dx = \int_\Omega f.p_{K_1}\, dx.$$

There are only seven main unknowns in the equation of step 1, thus the stencil is equal to seven.

b) We consider a particular case with a primary grid of squares. The dual grid is constructed as Figure 4.4. In Figure 4.4, the polygon $(x_{K_2}, x_{K_4}, x_{K_5}, x_{K_6}, x_{K_8}, x_{K_9})$

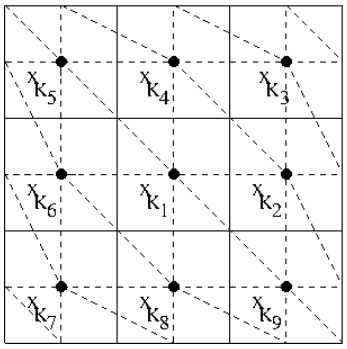


Figure 4.4: an example for the dual and the third meshes such that the stencil of the 2D FECC scheme is equal to 5.

is an element of the dual mesh where the mesh point x_{K^*} of this element coincides in x_{K_1}.
The intersecting domains are not empty between $supp\{p_{K_1}\}$ and each of the following domains: $supp\{p_{K_2}\}$, $supp\{p_{K_4}\}$, $supp\{p_{K_5}\}$, $supp\{p_{K_6}\}$, $supp\{p_{K_8}\}$, $supp\{p_{K_9}\}$, and are empty between $supp\{p_{K_1}\}$ and the others: $supp\{p_{K_3}\}$, $supp\{p_{K_7}\}$. Hence, equation (2.6) can be written as

$$\int_\Omega \left(\begin{array}{l} u_{K_1}\nabla p_{K_1} + u_{K_2}\nabla p_{K_2} + u_{K_4}\nabla p_{K_4} + u_{K_5}\nabla p_{K_5} \\ +u_{K_6}\nabla p_{K_6} + u_{K_8}\nabla p_{K_8} + u_{K_9}\nabla p_{K_9} \end{array} \right) . \nabla p_{K_1} dx = \int_\Omega f . p_{K_1}\, dx.$$

Moreover, we have that

- on the triangle $(x_{K_1}, x_{K_4}, x_{K_5})$

$$\nabla p_{K_5} .\ \nabla p_{K_1} = \frac{n_{[x_{K_1}, x_{K_4}]} . n_{[x_{K_4}, x_{K_5}]}}{\left(2m_{(x_{K_1}, x_{K_4}, x_{K_5})}\right)^2} = 0,$$

  because $n_{[x_{K_1}, x_{K_4}]} . n_{[x_{K_4}, x_{K_5}]}$ is equal to 0,
- on the triangle $(x_{K_1}, x_{K_5}, x_{K_6})$

$$\nabla p_{K_5} .\ \nabla p_{K_1} = \frac{n_{[x_{K_1}, x_{K_6}]} . n_{[x_{K_5}, x_{K_6}]}}{\left(2m_{(x_{K_1}, x_{K_5}, x_{K_6})}\right)^2} = 0,$$

  because $n_{[x_{K_1}, x_{K_6}]} . n_{[x_{K_5}, x_{K_6}]}$ is equal to 0,
- on the triangle $(x_{K_1}, x_{K_2}, x_{K_9})$

$$\nabla p_{K_9} . \nabla p_{K_1} = \frac{n_{[x_{K_1}, x_{K_2}]} . n_{[x_{K_2}, x_{K_9}]}}{\left(2m_{(x_{K_1}, x_{K_2}, x_{K_9})}\right)^2} = 0,$$

  because $n_{[x_{K_1}, x_{K_2}]} . n_{[x_{K_2}, x_{K_9}]}$ is equal to 0 ,

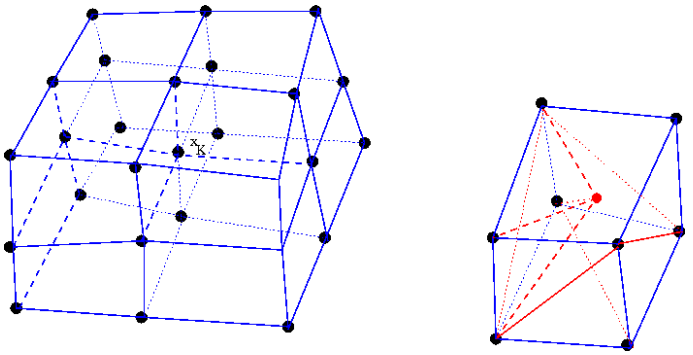


Figure 4.5: an example for the dual and the third meshes such that the stencil of the 3D FECC scheme is equal to 27.

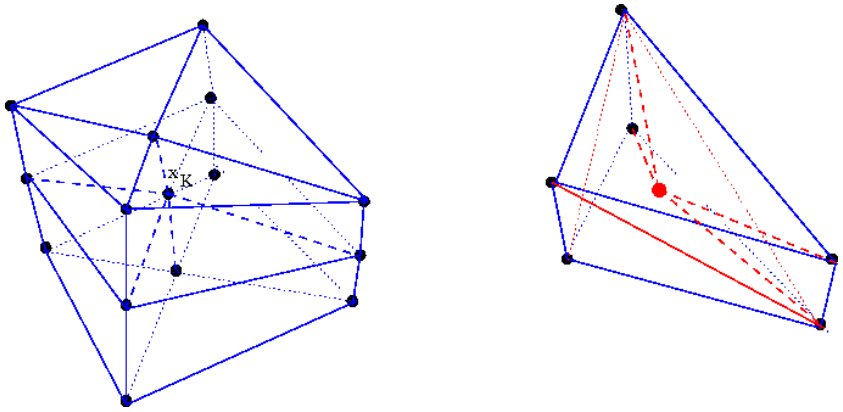


Figure 4.6: an example for the dual and the third meshes such that the stencil of the 3D FECC scheme is equal to 15.

- on the triangle $(x_{K_1}, x_{K_8}, x_{K_9})$

$$\nabla p_{K_9}.\nabla p_{K_1} = \frac{n_{[x_{K_1},x_{K_8}]}.\ n_{[x_{K_8},x_{K_9}]}}{\left(2m_{(x_{K_1},x_{K_8},x_{K_9})}\right)^2} = 0,$$

because $n_{[x_{K_1},x_{K_8}]}.n_{[x_{K_8},x_{K_9}]}$ is equal to 0 .

Hence, we get

$$\int_\Omega \left(\begin{array}{l} u_{K_1}\nabla p_{K_1} + u_{K_2}\nabla p_{K_2} + u_{K_4}\nabla p_{K_4} \\ +u_{K_6}\nabla p_{K_6} + u_{K_8}\nabla p_{K_8} \end{array} \right).\nabla p_{K_1} dx = \int_\Omega f.p_{K_1}\, dx,$$

which implies that the stencil is equal to 5.

The stencil of the 3D FECC scheme is equal or less than 27 on the hexahedral primary meshes, for examples: Figures 4.5 and 4.6. □

4.3.2 Relationship between the FECC scheme and the scheme in [5].

According to the construction in isotropic cases, the affine function u on (x_{K^*}, x_K, x_L) is also affine on (x_{K^*}, x_K, x_σ) and (x_{K^*}, x_L, x_σ). In addition, this function has continuous fluxes because Λ is continuous. Therefore, it corresponds to the function u constructed in the heterogeneous cases, as follows:

$$\begin{aligned}(\nabla_{D,Id}u)_{|(x_{K^*},x_K,x_\sigma)} &= \frac{-u_\sigma^{K^*} n_{[x_{K^*},x_K]} - u_K n^K_{[x_\sigma,x_{K^*}]} - u_{K^*} n_{[x_\sigma,x_K]}}{2m_{(x_{K^*},x_K,x_\sigma)}},\\ (\nabla_{D,Id}u)_{|(x_{K^*},x_L,x_\sigma)} &= \frac{-u_\sigma^{K^*} n_{[x_{K^*},x_L]} - u_L n^L_{[x_\sigma,x_{K^*}]} - u_{K^*} n_{[x_\sigma,x_L]}}{2m_{(x_{K^*},x_L,x_\sigma)}},\end{aligned}$$

where $u_\sigma^{K^*}$ is a temporary unknown. This unknown is eliminated by imposing the continuity of the fluxes:

$$(\nabla_{D,\Lambda}u)_{|(x_{K^*},x_K,x_\sigma)}\,.n^K_{[x_\sigma,x_{K^*}]} + (\nabla_{D,\Lambda}u)_{|(x_{K^*},x_L,x_\sigma)}\,.n^L_{[x_\sigma,x_{K^*}]} = 0.$$

$$\left(\alpha_\sigma^K + \alpha_\sigma^L\right) u_\sigma^{K^*} = \alpha_K u_K + \alpha_L u_L + \left(\alpha_{K^*}^K + \alpha_{K^*}^L\right) u_{K^*}, \tag{4.5}$$

where

$$\alpha_K = -\frac{n^K_{[x_\sigma,x_{K^*}]}.\, n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}}\ ,\quad \alpha^K_{K^*} = -\frac{n_{[x_\sigma,x_K]}.\, n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}}\ ,\quad \alpha^K_\sigma = \frac{n_{[x_{K^*},x_K]}.\, n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}},$$

$$\alpha_L = -\frac{n^L_{[x_\sigma,x_{K^*}]}.\, n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\ ,\quad \alpha^L_{K^*} = -\frac{n_{[x_\sigma,x_L]}.\, n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\ ,\quad \alpha^L_\sigma = \frac{n_{[x_{K^*},x_L]}.\, n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}}.$$

The following property presents a formula to compute the unknown u_σ in terms of the unknowns $u_K,\ u_L$.

Property 4.2 *The unknown $u_\sigma^{K^*}$ of (4.5) satisfies*

$$u_\sigma^{K^*} = \frac{d_{L,\sigma}}{d_{K,\sigma}+d_{L,\sigma}}u_K + \frac{d_{K,\sigma}}{d_{K,\sigma}+d_{L,\sigma}}u_L, \tag{4.6}$$

which is the value obtained by the scheme described in [5] using the harmonic averaging points.

In (4.6), $d_{K,\sigma}$, $d_{L,\sigma}$ are greater than 0 and denote the measures of the segments $[x_K, x_{K,\sigma}]$, $[x_L, x_{L,\sigma}]$, respectively. Two points $x_{K,\sigma}$, $x_{L,\sigma}$ are the two orthogonal projection points of x_K and x_L onto σ.

Proof of property 4.2

The coefficients of equation (4.5) which are computed in detail are based on Figure 4.7.

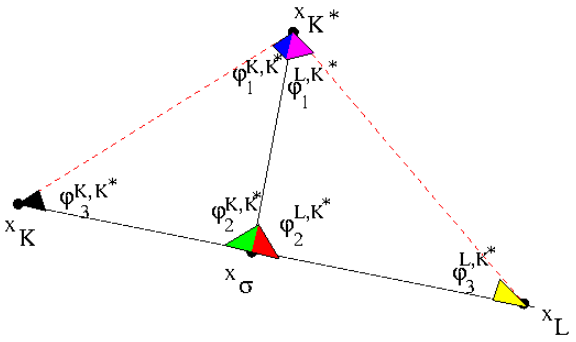


Figure 4.7: an element (x_{K^*}, x_K, x_L) of a third mesh and its angles.

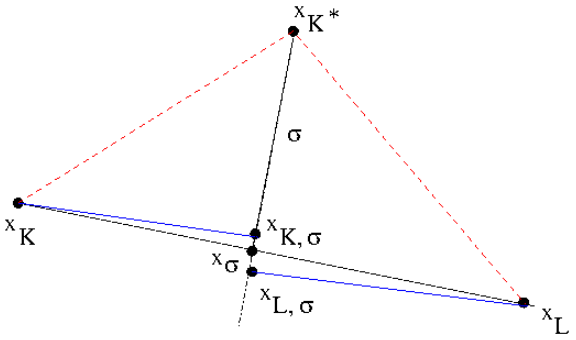


Figure 4.8: $x_{K,\sigma}$, $x_{L,\sigma}$ are the two orthogonal projection points of x_K, x_L on σ.

a) Calculation of the coefficient u_{K^*}:

$$\begin{aligned}
\alpha_{K^*}^K + \alpha_{K^*}^L &= -\frac{n_{[x_\sigma,x_K]}.\ n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}} - \frac{n_{[x_\sigma,x_L]}.n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}} \\
&= \frac{m_{[x_\sigma,x_K]}.\ m_\sigma \cos\left(\varphi_2^{K,K^*}\right)}{m_{[x_\sigma,x_K]}.\ m_\sigma \sin\left(\varphi_2^{K,K^*}\right)} + \frac{m_{[x_\sigma,x_L]}.\ m_\sigma \cos\left(\varphi_2^{L,K^*}\right)}{m_{[x_\sigma,x_L]}.\ m_\sigma \sin\left(\varphi_2^{L,K^*}\right)} \\
&= \frac{\cos\left(\varphi_2^{K,K^*}\right)}{\sin\left(\varphi_2^{K,K^*}\right)} + \frac{\cos\left(\varphi_2^{L,K^*}\right)}{\sin\left(\varphi_2^{L,K^*}\right)} \\
&= \frac{\cos\left(\Pi - \varphi_2^{L,K^*}\right)}{\sin\left(\Pi - \varphi_2^{L,K^*}\right)} + \frac{\cos\left(\varphi_2^{L,K^*}\right)}{\sin\left(\varphi_2^{L,K^*}\right)} = 0.
\end{aligned}$$

b) Calculation of the coefficient u_σ:

$$\begin{aligned}
\alpha_\sigma^K + \alpha_\sigma^L &= \frac{n_{[x_{K^*},x_K]}.\ n^K_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_K,x_\sigma)}} + \frac{n_{[x_{K^*},x_L]}.n^L_{[x_\sigma,x_{K^*}]}}{2m_{(x_{K^*},x_L,x_\sigma)}} \\
&= \frac{m_{[x_{K^*},x_K]}.\ m_\sigma.\ \cos\left(\varphi_1^{K,K^*}\right)}{m_\sigma.\ d_{K,\sigma}} + \frac{m_{[x_{K^*},x_L]}.m_\sigma.\ \cos\left(\varphi_1^{L,K^*}\right)}{m_\sigma.\ d_{L,\sigma}} \\
&= -\frac{m_{[x_{K^*},x_{K,\sigma}]}}{d_{K,\sigma}} - \frac{m_{[x_{K^*},x_{L,\sigma}]}}{d_{L,\sigma}} \\
&= \frac{m_\sigma.\,(d_{K,\sigma}+d_{L,\sigma}) - m_{[x_{K,\sigma},x_\sigma]}.d_{K,\sigma} + m_{[x_{L,\sigma},x_\sigma]}.d_{L,\sigma}}{d_{K,\sigma}.d_{L,\sigma}} \\
&= -\frac{m_\sigma\,(d_{K,\sigma}+d_{L,\sigma})}{d_{K,\sigma}\,d_{L,\sigma}},
\end{aligned}$$

because $m_{[x_{K^*},x_{K,\sigma}]} = m_{[x_{K^*},x_\sigma]} - m_{[x_{K,\sigma},x_\sigma]}$, $m_{[x_{K^*},x_{L,\sigma}]} = m_{[x_{K^*},x_\sigma]} + m_{[x_{L,\sigma},x_\sigma]}$, $m_{[x_{K,\sigma},x_\sigma]}.\ d_{L,\sigma} = m_{[x_{L,\sigma},x_\sigma]}.\ d_{K,\sigma}$.

c) Calculation of the coefficient u_K:

$$\alpha_K = -\frac{\left(n^K_{[x_\sigma,x_{K^*}]}\right)^2}{2m_{(x_{K^*},x_K,x_\sigma)}} = -\frac{(m_\sigma)^2}{m_\sigma.\ d_{K,\sigma}} = -\frac{m_\sigma}{d_{K,\sigma}}.$$

d) Calculation of the coefficient u_L:

$$\alpha_L = -\frac{\left(n^L_{[x_\sigma,x_{K^*}]}\right)^2}{2m_{(x_{K^*},x_L,x_\sigma)}} = -\frac{(m_\sigma)^2}{m_\sigma d_{L,\sigma}} = -\frac{m_\sigma}{d_{L,\sigma}}.$$

From the above calculations, we get:

$$\begin{aligned}
\left(-\frac{m_\sigma\,(d_{K,\sigma}+d_{L,\sigma})}{d_{K,\sigma}\,d_{L,\sigma}}\right) u_\sigma^{K^*} &= -\frac{m_\sigma}{d_{K,\sigma}}u_K - \frac{m_\sigma}{d_{L,\sigma}}u_L. \\
u_\sigma^{K^*} &= \frac{d_{L,\sigma}}{d_{K,\sigma}+d_{L,\sigma}}u_K + \frac{d_{K,\sigma}}{d_{K,\sigma}+d_{L,\sigma}}u_L. \qquad \square
\end{aligned}$$

Remark 4.2: In heterogeneous strongly anisotropic cases, the harmonic averaging point in [5] does not provide an "acceptable" interpolation u_σ. Even in this case, in Test 5 (see numerical results), we show that the FECC scheme can obtain precise results.

Remark 4.3: In property 4.1.2, we show a relationship between the FECC scheme and the scheme introduced in [5]. However, even in isotropic homogeneous cases, there is a difference between these two schemes.

For a grid of squares (the length of each edge is equal to a, see Figure 4.9), we consider two control volumes $K_{i,j}$ (mesh point $x_{i,j}$) and $K_{i+1,j}$ (mesh point $x_{i+1,j}$) of the primary grid such that their edges do not belong to the boundary $\partial\Omega$. The third grid is defined in the

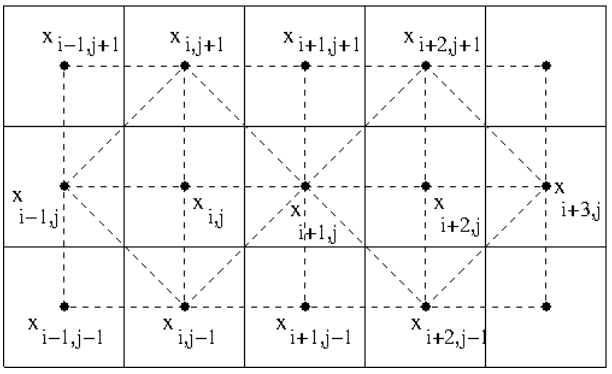


Figure 4.9: the dual and the third meshes are used in Remark 4.3.

following figure (dashed red and black lines): Using the scheme described in [5], for a given constant function f ($\neq 0$), we get the two following linear equations at $x_{i,j}$ and $x_{i+1,j}$:

$$\begin{aligned} 4u_{i,j} - u_{i,j+1} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} &= \int\limits_{K_{i,j}} f dx = f.a^2, \\ 4u_{i+1,j} - u_{i,j} - u_{i+1,j+1} - u_{i+2,j} - u_{i+1,j-1} &= \int\limits_{K_{i+1,j}} f dx = f.a^2. \end{aligned}$$

Using the FECC scheme, we get the two following linear equations at $x_{i,j}$ and $x_{i+1,j}$:

$$\begin{aligned} 4u_{i,j} - u_{i,j+1} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} &= \int\limits_{(x_{i,j+1};\, x_{i+1,j};\, x_{i,j-1};\, x_{i-1,j})} f.p_{i,j} dx = \frac{2.f.a^2}{3}, \\ 4u_{i+1,j} - u_{i,j} - u_{i+1,j+1} - u_{i+2,j} - u_{i+1,j-1} &= \int\limits_{(x_{i,j+1};\, x_{i+2,j+1};\, x_{i+2,j-1};\, x_{i,j-1})} f.p_{i+1,j} dx = \frac{4.f.a^2}{3}. \end{aligned}$$

At each point $x_{i,j}$ and $x_{i+1,j}$, we observe that the right hand sides of the two linear equations are different and that the left hand sides are the same. □

4.4 Discontinuous anisotropic heterogeneous cases.

4.4.1 Exact solution on piecewise affine functions.

In the property 4.3, we assume that, for neighbouring control volumes, the line joining their primary mesh points intersects their common edge. Hence, the dual meshes can be centred around the vertices of the primary meshes, where the dual mesh points are the vertices of the primary meshes.

Property 4.3 *Let the tensor Λ be piecewise constant in polyhedral sub-domains and the*

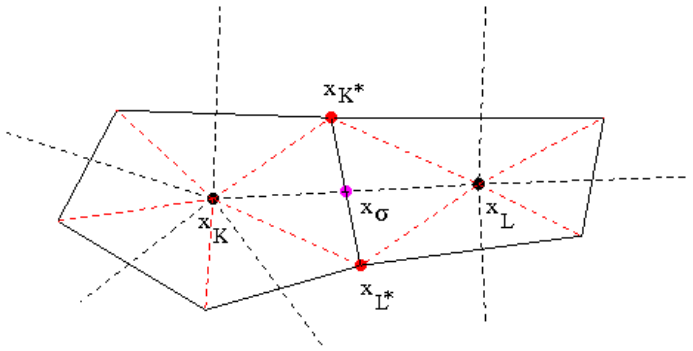


Figure 4.10: the two elements (x_{K^*}, x_K, x_L), (x_{L^*}, x_K, x_L) of the third mesh.

vector $u = \left((u_K)_{K\in M}, (u_{K^*})_{K^*\in M^*}\right) \in \mathcal{H}_{\mathcal{D}}$ *such that, for each* $K \in \mathcal{M}$, *there exists* $G_K \in \mathbb{R}^2$

$$u_{K^*} - u_K = G_K.(x_{K^*} - x_K), \tag{4.7}$$

for all $x_{K^*} \in \mathcal{V}_K$, $K^* \in \mathcal{M}^*$ *and such that for any* $\sigma \in \mathcal{E}_K \cup \mathcal{E}_L$ *which is the interface between the two neighbouring control volume* K *and* L,

$$\Lambda_K G_K.n_{K,L} + \Lambda_L G_L.n_{K,L} = 0, \tag{4.8}$$

where $n_{K,L}$ *is a outward normal unit vector of* K *at* σ. *Moreover, basing on Figure 4.10, we have*

$$(\nabla_{\mathcal{D},\Lambda} u)\big|_{(x_K,x_{K^*},x_\sigma)}.\ \frac{(x_\sigma - x_{K^*})}{m_{[x_\sigma,x_{K^*}]}} + (\nabla_{\mathcal{D},\Lambda} u)\big|_{(x_K,x_{L^*},x_\sigma)}.\ \frac{(x_\sigma - x_{L^*})}{m_{[x_\sigma,x_{L^*}]}} = 0. \tag{4.9}$$

Then u *is the unique solution of the discrete diffusion problem:*

$$\forall v \in \mathcal{H}_{\mathcal{D}}, \int_\Omega \Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x).\nabla_{\mathcal{D},\Lambda} v(x)\,dx = 0. \tag{4.10}$$

Proof of property 4.3

For each $K \in \mathcal{M}$, $\sigma \in \mathcal{E}_K \cap \mathcal{E}_L$ with $L \in \mathcal{M}$, the Local Conservativity of the Fluxes is imposed on σ, we obtain:

- on the triangle (x_K, x_L, x_{K^*}),

$$u_\sigma^{K^*}\left(\gamma_K^{K^*,\sigma} + \gamma_L^{K^*,\sigma} + \gamma_{K^*}^{K^*,\sigma}\right) = \gamma_K^{K^*,\sigma} u_K + \gamma_L^{K^*,\sigma} u_L + \gamma_{K^*}^{K^*,\sigma} u_{K^*}, \tag{4.11}$$

- on the triangle (x_K, x_L, x_{L^*}),

$$u_\sigma^{L^*}\left(\gamma_K^{L^*,\sigma} + \gamma_L^{L^*,\sigma} + \gamma_{L^*}^{L^*,\sigma}\right) = \gamma_K^{L^*,\sigma} u_K + \gamma_L^{L^*,\sigma} u_L + \gamma_{L^*}^{L^*,\sigma} u_{L^*}. \tag{4.12}$$

The coefficients are computed, as follows:

- in (4.11),

$$\begin{aligned}
\gamma_K^{K^*,\sigma} &= \frac{\left(\Lambda_K n^K_{[x_{K^*},x_\sigma]}\right).n^K_{[x_{K^*},x_\sigma]}}{2m_{(x_{K^*},x_K,x_\sigma)}} = \frac{\left(\Lambda_K \overline{n}^K_{[x_{K^*},x_\sigma]}\right).\overline{n}^K_{[x_{K^*},x_\sigma]}.m_{[x_{K^*},x_\sigma]}}{d_{K,\sigma}},\\
\gamma_L^{K^*,\sigma} &= \frac{\left(\Lambda_L n^L_{[x_{K^*},x_\sigma]}\right).n^L_{[x_{K^*},x_\sigma]}}{2m_{(x_{K^*},x_L,x_\sigma)}} = \frac{\left(\Lambda_L \overline{n}^L_{[x_{K^*},x_\sigma]}\right).\overline{n}^L_{[x_{K^*},x_\sigma]}.m_{[x_{K^*},x_\sigma]}}{d_{L,\sigma}},\\
\gamma_{K^*}^{K^*,\sigma} &= \frac{\left(\Lambda_K n^{K^*}_{[x_K,x_\sigma]}\right).n^K_{[x_{K^*},x_\sigma]}}{2m_{(x_{K^*},x_K,x_\sigma)}} + \frac{\left(\Lambda_L n^{K^*}_{[x_L,x_\sigma]}\right).n^L_{[x_{K^*},x_\sigma]}}{2m_{(x_{K^*},x_L,x_\sigma)}}\\
&= \frac{\left(\Lambda_K \overline{n}^{K^*}_{[x_K,x_\sigma]}\right).\overline{n}^K_{[x_{K^*},x_\sigma]}.m_{[x_{K^*},x_\sigma]}}{d_{K^*,[x_K,x_L]}} + \frac{\left(\Lambda_L \overline{n}^{K^*}_{[x_L,x_\sigma]}\right).\overline{n}^L_{[x_{K^*},x_\sigma]}.m_{[x_{K^*},x_\sigma]}}{d_{K^*,[x_K,x_L]}},
\end{aligned}$$

- in (4.12),

$$\begin{aligned}
\gamma_K^{L^*,\sigma} &= \frac{\left(\Lambda_K n^K_{[x_{L^*},x_\sigma]}\right).n^K_{[x_{L^*},x_\sigma]}}{2m_{(x_{L^*},x_K,x_\sigma)}} = \frac{\left(\Lambda_K \overline{n}^K_{[x_{L^*},x_\sigma]}\right).\overline{n}^K_{[x_{L^*},x_\sigma]}.m_{[x_{L^*},x_\sigma]}}{d_{K,\sigma}},\\
\gamma_L^{L^*,\sigma} &= \frac{\left(\Lambda_L n^L_{[x_{L^*},x_\sigma]}\right).n^L_{[x_{L^*},x_\sigma]}}{2m_{(x_{L^*},x_L,x_\sigma)}} = \frac{\left(\Lambda_L \overline{n}^L_{[x_{L^*},x_\sigma]}\right).\overline{n}^L_{[x_{L^*},x_\sigma]}.m_{[x_{L^*},x_\sigma]}}{d_{L,\sigma}},\\
\gamma_{L^*}^{L^*,\sigma} &= \frac{\left(\Lambda_K n^{L^*}_{[x_K,x_\sigma]}\right).n^K_{[x_{L^*},x_\sigma]}}{2m_{(x_{L^*},x_K,x_\sigma)}} + \frac{\left(\Lambda_L n^{L^*}_{[x_L,x_\sigma]}\right).n^L_{[x_{L^*},x_\sigma]}}{2m_{(x_{L^*},x_L,x_\sigma)}}\\
&= \frac{\left(\Lambda_K \overline{n}^{L^*}_{[x_K,x_\sigma]}\right).\overline{n}^K_{[x_{L^*},x_\sigma]}.m_{[x_{L^*},x_\sigma]}}{d_{L^*,[x_K,x_L]}} + \frac{\left(\Lambda_L \overline{n}^{L^*}_{[x_L,x_\sigma]}\right).\overline{n}^L_{[x_{L^*},x_\sigma]}.m_{[x_{L^*},x_\sigma]}}{d_{L^*,[x_K,x_L]}},
\end{aligned}$$

where the vectors $\overline{n}^K_{[x_{K^*},x_\sigma]}$, $\overline{n}^L_{[x_{K^*},x_\sigma]}$, $\overline{n}^{K^*}_{[x_K,x_\sigma]}$, $\overline{n}^{K^*}_{[x_L,x_\sigma]}$, $\overline{n}^K_{[x_{L^*},x_\sigma]}$, $\overline{n}^L_{[x_{L^*},x_\sigma]}$, $\overline{n}^{L^*}_{[x_K,x_\sigma]}$ and $\overline{n}^{L^*}_{[x_L,x_\sigma]}$ are respectively the unit vectors of $n^K_{[x_{K^*},x_\sigma]}$, $n^L_{[x_{K^*},x_\sigma]}$, $n^{K^*}_{[x_K,x_\sigma]}$, $n^{K^*}_{[x_L,x_\sigma]}$, $n^K_{[x_{L^*},x_\sigma]}$, $n^L_{[x_{L^*},x_\sigma]}$, $n^{L^*}_{[x_K,x_\sigma]}$ and $n^{L^*}_{[x_L,x_\sigma]}$. The notations $d_{K^*,[x_K,x_L]}$ and $d_{L^*,[x_K,x_L]}$ are the two distances from x_{K^*} to $[x_K,x_L]$ and from x_{L^*} to $[x_K,x_L]$.

By the above computations of the coefficients of (4.11) and (4.12), we get the relationships between their coefficients:

$$\begin{aligned}
\frac{\gamma_K^{K^*,\sigma}}{m_{[x_{K^*},x_\sigma]}} &= \frac{\gamma_K^{L^*,\sigma}}{m_{[x_{L^*},x_\sigma]}}; \quad \frac{\gamma_L^{K^*,\sigma}}{m_{[x_{K^*},x_\sigma]}} = \frac{\gamma_L^{L^*,\sigma}}{m_{[x_{L^*},x_\sigma]}};\\
\gamma_{K^*}^{K^*,\sigma} &= -\gamma_{L^*}^{L^*,\sigma} \text{ because of } \frac{m_{[x_{K^*},x_\sigma]}}{d_{K^*,[x_K,x_L]}} = \frac{m_{[x_{L^*},x_\sigma]}}{d_{x_{L^*},[x_K,x_L]}}.
\end{aligned}$$

Let $\frac{(4.11)}{m_{[x_{K^*},x_\sigma]}}$ subtracts $\frac{(4.12)}{m_{[x_{L^*},x_\sigma]}}$, we get

$$\left(\frac{\gamma_K^{K^*,\sigma}+\gamma_L^{K^*,\sigma}}{m_{[x_{K^*},x_\sigma]}}\right).\left(u_\sigma^{K^*}-u_\sigma^{L^*}\right) = -\frac{\gamma_{K^*}^{K^*,\sigma}}{m_{[x_{K^*},x_\sigma]}}.\left(u_\sigma^{K^*}-u_{K^*}\right)+\frac{\gamma_{L^*}^{L^*,\sigma}}{m_{[x_{L^*},x_\sigma]}}.\left(u_\sigma^{L^*}-u_{L^*}\right). \tag{4.13}$$

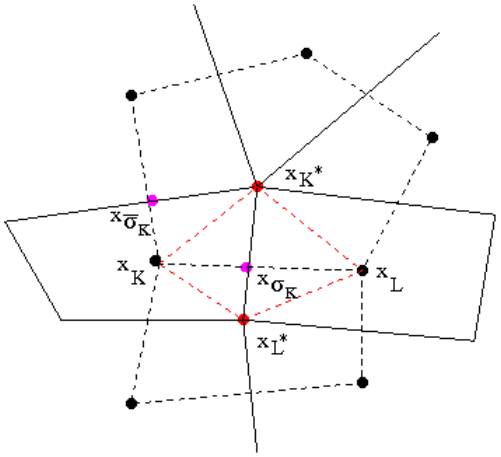


Figure 4.11: x_K, $x_L \in \mathcal{P}$; x_{K^*}, $x_{L^*} \in \mathcal{P}^*$; and the two intersecting points x_{σ_K}, $x_{\overline{\sigma}_K}$.

Thanks to (4.9) and definitions of $(\nabla_{D,\Lambda}u)\Big|_{(x_K,x_{K^*},x_\sigma)}$ and $(\nabla_{D,\Lambda}u)\Big|_{(x_K,x_{L^*},x_\sigma)}$, we deduce $u_\sigma^{K^*} = u_\sigma^{L^*} = u_\sigma$ from (4.13).

By (4.9) and $u_\sigma^{K^*} = u_\sigma^{L^*} = u_\sigma$, we obtain:

$$\frac{u_\sigma - u_{K^*}}{m_{[x_{K^*},x_\sigma]}} + \frac{u_\sigma - u_{L^*}}{m_{[x_{L^*},x_\sigma]}} = 0. \tag{4.14}$$

The left hand side of (4.14) is computed, as follows:

$$\begin{aligned}
&\frac{u_\sigma - u_{K^*}}{m_{[x_{K^*},x_\sigma]}} + \frac{u_\sigma - u_{L^*}}{m_{[x_{L^*},x_\sigma]}} = \frac{u_\sigma - u_K + u_K - u_{K^*}}{m_{[x_{K^*},x_\sigma]}} + \frac{u_\sigma - u_K + u_K - u_{L^*}}{m_{[x_{L^*},x_\sigma]}} \\
= \;& (u_\sigma - u_K).\left(\frac{1}{m_{[x_{K^*},x_\sigma]}} + \frac{1}{m_{[x_{L^*},x_\sigma]}}\right) + \frac{G_K.(x_K - x_{K^*})}{m_{[x_{K^*},x_\sigma]}} + \frac{G_K.(x_K - x_{L^*})}{m_{[x_{L^*},x_\sigma]}} \\
= \;& (u_\sigma - u_K).\left(\frac{1}{m_{[x_{K^*},x_\sigma]}} + \frac{1}{m_{[x_{L^*},x_\sigma]}}\right) + \left(\frac{1}{m_{[x_{K^*},x_\sigma]}} + \frac{1}{m_{[x_{L^*},x_\sigma]}}\right) G_K.(x_K - x_\sigma) \\
+ \;& \underbrace{\frac{G_K.(x_\sigma - x_{K^*})}{m_{[x_{K^*},x_\sigma]}} + \frac{G_K.(x_\sigma - x_{L^*})}{m_{[x_{L^*},x_\sigma]}}}_{=0},
\end{aligned}$$

which implies that $u_\sigma - u_K = G_K.(x_\sigma - x_K)$. Therefore, for each $K \in \mathcal{M}$, the discrete gradient $(\nabla_{D,\Lambda}u)\,|_K$ is equal to G_K.
Then we prove:

$$\forall v \in \mathcal{H}_\mathcal{D}, \int_\Omega \Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x).\nabla_{\mathcal{D},\Lambda}v(x)\,dx = 0 \tag{4.15}$$

with the following two steps and Figure 4.11:

a) For each $K^* \in \mathcal{M}^*$, let a basic function v_{K^*} be continuous piecewise linear on Ω, such that $v_{K^*}(x_{K^*}) = 1$, $v_{K^*}(x_L) = 0$ and $v_{K^*}(x_\sigma) = 0$, where each control volume L belongs

to $\{\mathcal{M}\cup\mathcal{M}^*\}\backslash\{K^*\}$, x_σ is an intersecting point between $[x_K, x_L]$ and $\sigma\in\mathcal{E}_K\cap\mathcal{E}_L\cap\mathcal{E}_{\rm int}$. Moreover, v_{K^*} is also equal to 0 on the boundary $\partial\Omega$.

Using $supp(v_{K^*})$ and the constructions of the dual and the third meshes, we obtain:

$$\int_\Omega \Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x).\nabla_{\mathcal{D},\Lambda}v_{K^*}(x)\,dx = \frac{1}{4}\sum_{\substack{x_K\in\mathcal{V}_{K^*}\\ K\in\mathcal{M}\\ \sigma_K,\overline{\sigma}_K\in\mathcal{E}_K}}\Big[(\Lambda_K G_K).n^{K^*}_{[x_K,x_{\sigma_K}]}+(\Lambda_K G_K).n^{K^*}_{[x_K,x_{\overline{\sigma}_K}]}\Big]$$

$$=\frac{1}{4}\sum_{\substack{x_K\in\mathcal{V}_{K^*}\\ K\in\mathcal{M}\\ \sigma_K,\overline{\sigma}_K\in\mathcal{E}_K}}\Big[-(\Lambda_K G_K).n^{K}_{[x_{K^*},x_{\sigma_K}]}-(\Lambda_K G_K).n^{K}_{[x_{K^*},x_{\overline{\sigma}_K}]}\Big]. \tag{4.16}$$

Applying (4.8) in (4.16), we get:

$$\forall K^*\in\mathcal{M}^*,\ \int_\Omega \Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x).\nabla_{\mathcal{D},\Lambda}v_{K^*}(x)\,dx=0. \tag{4.17}$$

b) For each $K\in\mathcal{M}$, let a basic function v_K be continuous piecewise linear on Ω, such that $v_K(x_K)=1$ and $v_K(x_L)=0$, where each control volume L belongs to $\{\mathcal{M}\cup\mathcal{M}^*\}\backslash\{K\}$. Besides, v_K is also equal to 0 on the boundary $\partial\Omega$.

We compute the integral $\int_\Omega \Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x).\nabla_{\mathcal{D},\Lambda}v_K(x)\,dx$

$$\begin{aligned}
&= \frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_K\cap\mathcal{E}_L\\ \sigma_K=[x_{K^*},x_{L^*}]\\ L\in\mathcal{M},\ K^*\text{ and }L^*\in\mathcal{M}^*}}\left[\begin{array}{l}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}\\ +v_K(x_{\sigma_K}).(\Lambda_K G_K).n^K_{[x_K,x_{K^*}]}+v_K(x_{\sigma_K}).(\Lambda_L G_L).n^L_{[x_L,x_{K^*}]}\\ +v_K(x_{\sigma_K}).(\Lambda_K G_K).n^K_{[x_K,x_{L^*}]}+v_K(x_{\sigma_K}).(\Lambda_L G_L).n^L_{[x_L,x_{L^*}]}\end{array}\right]\\
&+ \frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_{\rm ext}\\ \sigma_K=[x_{K^*},x_{L^*}]\\ K^*\text{ and }L^*\in\mathcal{M}^*}}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}\\
&= \frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_K\cap\mathcal{E}_L\\ \sigma_K=[x_{K^*},x_{L^*}]\\ L\in\mathcal{M},\ K^*\text{ and }L^*\in\mathcal{M}^*}}\left[\begin{array}{l}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}\\ -v_K(x_{\sigma_K}).(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}-v_K(x_{\sigma_K}).(\Lambda_L G_L).n^L_{[x_{K^*},x_{L^*}]}\end{array}\right]\\
&+ \frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_{\rm ext}\\ \sigma_K=[x_{K^*},x_{L^*}]\\ K^*\text{ and }L^*\in\mathcal{M}^*}}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}\\
&= \frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_K\cap\mathcal{E}_L\\ \sigma_K=[x_{K^*},x_{L^*}]\\ L\in\mathcal{M},\ K^*\text{ and }L^*\in\mathcal{M}^*}}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}+\frac{1}{4}\sum_{\substack{\sigma_K\in\mathcal{E}_{\rm ext}\\ \sigma_K=[x_{K^*},x_{L^*}]\\ K^*\text{ and }L^*\in\mathcal{M}^*}}(\Lambda_K G_K).n^K_{[x_{K^*},x_{L^*}]}.
\end{aligned}\tag{4.18}$$

Applying (4.8) in (4.18), it follows that

$$\forall K \in \mathcal{M}, \int_{\Omega} \Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x).\nabla_{\mathcal{D},\Lambda}v_K(x)\,dx = 0. \tag{4.19}$$

The uniqueness of u is implied from Lemma 4.1. $\square$

Chapter 5

Convergence of the FECC scheme.

Contents

5.1 Motivation.

We are interested in this chapter by the theoretical convergence of the FECC scheme in the general case where the tensor $\Lambda(x)$ can be discontinuous. The key point of the proof includes showing both the strong and the weak consistency of the FECC scheme.

5.2 Proof of the convergence.

We denote by $P_1(v)$ the traditional P_1 function on Ω, constructed on $\mathcal{M}^{**}$. We first show the convergence of a variant of the original scheme, which we call FECCB, which satisfies the following discrete variational formulation:

$$u \in \mathcal{H}_{\mathcal{D}} \int_{\Omega} (\Lambda(x)\nabla_{\mathcal{D},\Lambda}u(x)) . \nabla_{\mathcal{D},\Lambda}v(x)\, dx = \int_{\Omega} f(x)P_1(v)(x)dx \quad \text{for all } v \in \mathcal{H}_{\mathcal{D}}. \tag{5.1}$$

To simplify the presentation, we assume that, for neighbouring control volumes, the line joining their primary mesh points intersects their common edge. Let

$\mathcal{M}^{**}_{\Lambda} = \{K \in \mathcal{M}^{**} \mid \Lambda(x) \text{ is not continuous on } K\}$, $\mathcal{M}^{**}_{Const} = \{K \in \mathcal{M}^{**} \mid \Lambda(x) \text{ is constant on } K\}$.

If $\forall K \in \mathcal{M}^{**}\backslash\mathcal{M}^{**}_{Const}$, let us assume that $\Lambda(x) = \Lambda_1$ on K_1 (K_1 is the triangle (x_K, x_σ, x_{K^*}) and that $\Lambda(x) = \Lambda_2$ on K_2 (K_2 is the triangle (x_L, x_σ, x_{K^*})) (see Figure 5.1).

For $K \in \mathcal{M}^{**}_{\Lambda}$, we only choose two discontinuities to simplify the presentation but the method can be generalized to a greater number of discontinuities.

We denote by h_K the diameter of the triangle K and $\rho_K = sup\{diam(S) : \text{S is a ball contained in K}\}$.

As described in section 2, we recall that $\mathcal{V}^{**}_K$ is the set of the vertices of the triangle $K \in \mathcal{M}^{**}$. Moreover, the size of the discretization is defined by

$$h_{\mathcal{M}^{**}} = sup\{h_K, K \in \mathcal{M}^{**}\}$$

For all the triangular cells belonging to $\mathcal{M}^{**}$, we join the centers of gravity of the triangles $x_{bar,K}$ to the midpoints of the edges $(x_{K,K^*}, x_{K,L}, x_{K^*,L})$. The vectors $\vec{\tau}_{K,K^*}$, $\vec{\tau}_{K^*,L}$, $\vec{\tau}_{L,K}$ are orthogonal vectors (with the same length) to the sides $[x_{K,K^*}, x_{bar,K}]$, $[x_{K^*,L}, x_{bar,K}]$, $[x_{K,L}, x_{bar,K}]$.

We denote by Ai_K, Ai_L and Ai_{K^*}, the polygons $(x_K, x_{K,L}, x_{bar,K}, x_{K,K^*})$, $(x_L, x_{K,L}, x_{bar,K}, x_{K^*,L})$ and $(x_{K^*}, x_{K^*,L}, x_{bar,K}, x_{K,K^*})$. The vectors $\vec{n}_K$, $\vec{n}_{K^*}$, $\vec{n}_L$, $\vec{n}_\sigma$, $\vec{n}_{K^*,1}$ and $\vec{n}_{K^*,2}$ are orthogonal vectors (with the same length) to the sides $[x_{K^*}, x_L]$, $[x_K, x_L]$, $[x_{K^*}, x_K]$, $[x_{K^*}, x_\sigma]$, $[x_K, x_\sigma]$ and $[x_\sigma, x_L]$ (see Figure 5.1). We define $\Pi^0_{\mathcal{D}^{**}}u$ which is a piecewise constant reconstruction by:

$$\begin{aligned}
\Pi^0_{\mathcal{D}^{**}}u(x) &= \Pi^0_K u = u_K \quad \text{if } x \in Ai_K, \\
\Pi^0_{\mathcal{D}^{**}}u(x) &= \Pi^0_L u = u_L \quad \text{if } x \in Ai_L, \\
\Pi^0_{\mathcal{D}^{**}}u(x) &= \Pi^0_{K^*} u = u_{K^*} \quad \text{if } x \in Ai_{K^*}.
\end{aligned}$$

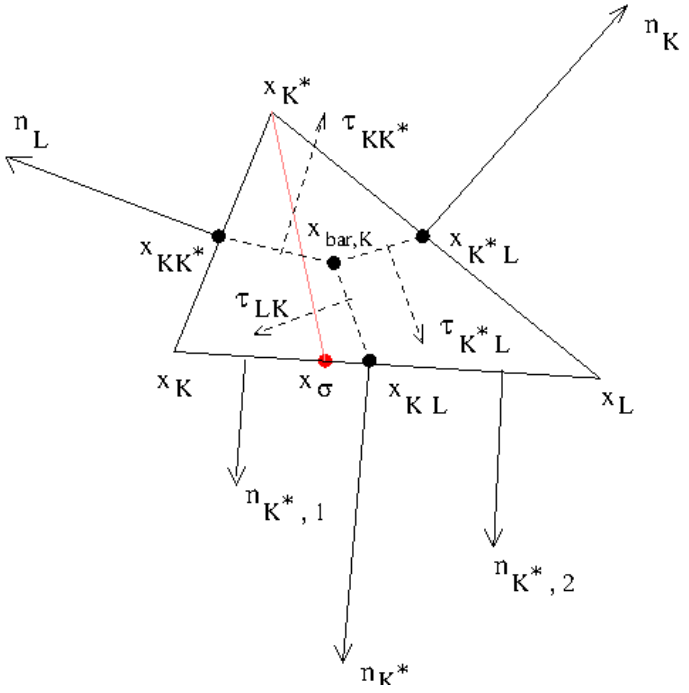


Figure 5.1: the triangle (x_{K^*}, x_K, x_L), its sub-polygons $Ai_K = (x_K, x_{K,L}, x_{bar,K}, x_{K,K^*})$, $Ai_L = (x_L, x_{K,L}, x_{bar,K}, x_{K^*,L})$, $Ai_{K^*} = (x_{K^*}, x_{K^*,L}, x_{bar,K}, x_{K,K^*})$ and their outward normal vectors.

We define the discrete H^1 norm of u by:

$$||u||^2_{1,\mathcal{D}^{**}} = \sum_{K\in\mathcal{M}^{**}} \frac{|\vec{\tau}_{K,L}|}{d(K,L)}(u_K - u_L)^2 + \frac{|\vec{\tau}_{K,K^*}|}{d(K,K^*)}(u_K - u_{K^*})^2 + \frac{|\vec{\tau}_{L,K^*}|}{d(L,K^*)}(u_L - u_{K^*})^2,$$

where $d(K,L)$ is the distance between x_K and $[x_{K,L}, x_{bar,K}]$ $(d(K,L) = d(L,K))$, $d(K,K^*)$ is the distance between x_{K^*} and $[x_{K,K^*}, x_{bar,K}]$ $(d(K,K^*) = d(K^*,K))$ and $d(L,K^*)$ is the distance between x_L and $[x_{K^*,L}, x_{bar,K}]$ $(d(L,K^*) = d(K^*,L))$.

Following the definition given in [26], we measure the strong consistency with the interpolation error function $S(\varphi) = \{||P_1(\varphi) - \varphi||^2_{L^2(\Omega)} + ||\nabla_{\mathcal{D},\Lambda}\varphi - \nabla\varphi||^2_{L^2(\Omega)^2}\}^{\frac{1}{2}}$,
$\varphi \in [C_c^\infty(\Omega)]$ and the dual consistency with the conformity error function

$$W_{\mathcal{D}^{**}}(\vec{\varphi}) = max_{u\in\mathcal{H}_\mathcal{D}}||\frac{1}{\nabla_{\mathcal{D},\Lambda}u(x)}|| \int_\Omega (\nabla_{\mathcal{D},\Lambda}u(x).\vec{\varphi} + P_1(u)(x)div\vec{\varphi}(x))dx, \forall\vec{\varphi} \in [C_c^\infty(\Omega)]^2. \ \square$$

Lemma 5.1 *With hypothesis 2.1, let $\mathcal{S}$ be a sequence of discretizations $\mathcal{D}^{**} = (\mathcal{H}_\mathcal{D}, h_{\mathcal{M}^{**}}, P_1(u), \nabla_{\mathcal{D},\Lambda})$ previously defined. We assume that there exists θ such that for all $\mathcal{D}^{**} \in \mathcal{S}$,*
for $K \in \mathcal{M}^{**}\backslash\{\mathcal{M}^{**}_\Lambda \cup \mathcal{M}^{**}_{Const}\}$,

$$\frac{min(|\vec{\tau}_{K,L}|, |\vec{\tau}_{K,K^*}|, |\vec{\tau}_{K,K^*}|)}{|\vec{n}_L|}\frac{|K_1|}{|K|} > \theta \ and \ \frac{min(|\vec{\tau}_{K,L}|, |\vec{\tau}_{K,K^*}|, |\vec{\tau}_{K,K^*}|)}{|\vec{n}_K|}\frac{|K_2|}{|K|} > \theta.$$

Then, for $K \in \mathcal{M}^{**}\backslash\mathcal{M}^{**}_{\Lambda}$ *the gradient* $\nabla_{\mathcal{D},\Lambda}u$ *satisfies*:

$|K|\nabla_{\mathcal{D},\Lambda}u = (u_{K^*}-u_K)(\vec{\tau}_{K,K^*}+\vec{\epsilon}_{K,K^*})+(u_L-u_{K^*})(\vec{\tau}_{K^*,L}+\vec{\epsilon}_{K^*,L})+(u_K-u_L)(\vec{\tau}_{L,K}+\vec{\epsilon}_{L,K})$

with $lim_{h_{\mathcal{M}^{**}}\to 0}\frac{|\vec{\epsilon}_{K,K^*}|}{|\vec{\tau}_{K,K^*}|}=0$, $lim_{h_{\mathcal{M}^{**}}\to 0}\frac{|\vec{\epsilon}_{K^*,L}|}{|\vec{\tau}_{K^*,L}|}=0$ *and* $lim_{h_{\mathcal{M}^{**}}\to 0}\frac{|\vec{\epsilon}_{L,K}|}{|\vec{\tau}_{L,K}|}=0$.

Proof of lemma 5.1

First case: $\Lambda_1 = \Lambda_2$**.**

Using the stokes formula, we obtain (see Figure 5.1):

$$2|K|\nabla_{\mathcal{D},\Lambda}u = -u_{K^*}\vec{n}_{K^*} - u_L\vec{n}_L - u_K\vec{n}_K$$

which becomes

$$|K|\nabla_{\mathcal{D},\Lambda} = \frac{1}{6}\{(u_{K^*}-u_K)(\vec{n}_K-\vec{n}_{K^*})+(u_K-u_L)(\vec{n}_L-\vec{n}_K)+(u_L-u_{K^*})(\vec{n}_{K^*}-\vec{n}_L)\}.$$

As the vectors $\vec{\tau}_{L,K}$, $\vec{\tau}_{K,K^*}$ and $\vec{\tau}_{K^*,L}$ satisfy $\vec{\tau}_{L,K}=\frac{1}{6}(\vec{n}_L-\vec{n}_K)$,
$\vec{\tau}_{K,K^*}=\frac{1}{6}(\vec{n}_K-\vec{n}_{K^*})$, $\vec{\tau}_{K^*,L}=\frac{1}{6}(-\vec{n}_L+\vec{n}_{K^*})$, we obtain:

$$|K|\nabla_{\mathcal{D},\Lambda}u = (u_{K^*}-u_K)\vec{\tau}_{K,K^*}+(u_L-u_{K^*})\vec{\tau}_{K^*,L}+(u_K-u_L)\vec{\tau}_{L,K}.$$

Second case: $\Lambda_1 \neq \Lambda_2$**.**

For $x \in K_1$, the gradient $\nabla_{\mathcal{D},\Lambda}u$ satisfies:

$$2|K_1|\nabla_{\mathcal{D},\Lambda}u = -u_{K^*}\vec{n}_{K^*,1} - u_\sigma\vec{n}_L - u_K\vec{n}_\sigma.$$

We can write

$$u_\sigma = \beta_K u_K + \beta_L u_L + \beta_{K^*}u_{K^*},$$

with $\beta_K+\beta_L+\beta_{K^*}=1$, $lim_{h_{\mathcal{M}^{**}}\to 0}\beta_{K^*}=0$, $lim_{h_{\mathcal{M}^{**}}\to 0}\beta_K=\frac{[x_L,x_\sigma]}{[x_L,x_K]}$,
and $lim_{h_{\mathcal{M}^{**}}\to 0}\beta_L=\frac{[x_K,x_\sigma]}{[x_L,x_K]}$ because, for $K\in\mathcal{M}^{**}\backslash\{\mathcal{M}^{**}_{\Lambda}\cup\mathcal{M}^{**}_{Const}\}$ ($\Lambda(x)$ is continuous),
$lim_{h_{\mathcal{M}^{**}}\to 0}||\Lambda_1-\Lambda_2||=0$.

The gradient becomes:

$$2|K_1|\nabla_{\mathcal{D},\Lambda}u = -u_{K^*}(\vec{n}_{K^*,1}+\epsilon_{K^*}\vec{n}_L)-u_L(\frac{[x_K,x_\sigma]}{[x_L,x_K]}+\epsilon_L)\vec{n}_L-u_K(\vec{n}_\sigma+\vec{n}_L(\epsilon_K+\frac{[x_L,x_\sigma]}{[x_L,x_K]})),$$

with $\epsilon_K+\epsilon_L+\epsilon_{K^*}=0$, $lim_{h_{\mathcal{M}^{**}}\to 0}\epsilon_{K^*}=0$, $lim_{h_{\mathcal{M}^{**}}\to 0}\epsilon_K=0$ and $lim_{h_{\mathcal{M}^{**}}\to 0}\epsilon_L=0$.

As $\frac{(\vec{n}_\sigma+\frac{[x_L,x_\sigma]}{[x_L,x_K]}\vec{n}_L)}{|K_1|}=\frac{\vec{n}_K}{|K|}$, $\frac{\frac{[x_K,x_\sigma]}{[x_L,x_K]}\vec{n}_L}{|K_1|}=\frac{\vec{n}_L}{|K|}$ and $\frac{\vec{n}_{K^*,1}}{|K_1|}=\frac{\vec{n}_{K^*}}{|K|}$, we deduce:

$$\begin{aligned}2|K|\nabla_{\mathcal{D},\Lambda}u &= -u_{K^*}(\vec{n}_{K^*}+\frac{|K|}{|K_1|}\epsilon_{K^*}\vec{n}_L)-u_L(1+\frac{|K|}{|K_1|}\epsilon_L)\vec{n}_L-u_K(\vec{n}_K+\frac{|K|}{|K_1|}\epsilon_K\vec{n}_L)\\ &= -u_{K^*}\vec{p}_{K^*}-u_L\vec{p}_L-u_K\vec{p}_K.\end{aligned}$$

Finally, we obtain:

$$|K|\nabla_{\mathcal{D},\Lambda}u = \frac{1}{6}\{(u_{K^*}-u_K)(\vec{p}_K-\vec{p}_{K^*})+(u_K-u_L)(\vec{p}_L-\vec{p}_K)+(u_L-u_{K^*})(\vec{p}_{K^*}-\vec{p}_L)\}.$$

Using the assumption on the grid ($\frac{min(|\vec{\tau}_{K,L}|,|\vec{\tau}_{K,K^*}|,|\vec{\tau}_{K,K^*}|)}{|\vec{n}_L|}\frac{|K_1|}{|K|} > \theta$), choosing

$$\begin{aligned}
\vec{\epsilon}_{K,K^*} &= -\frac{|K|}{|K_1|}\epsilon_{K^*}\vec{n}_L + \frac{|K|}{|K_1|}\epsilon_K\vec{n}_L,\\
\vec{\epsilon}_{K^*,L} &= \frac{|K|}{|K_1|}\epsilon_{K^*}\vec{n}_L - \frac{|K|}{|K_1|}\epsilon_L\vec{n}_L,\\
\vec{\epsilon}_{L,K} &= \frac{|K|}{|K_1|}\epsilon_L\vec{n}_L - \frac{|K|}{|K_1|}\epsilon_K\vec{n}_L
\end{aligned}$$

and using the values of $\vec{\tau}_{L,K}$,$\vec{\tau}_{K,K^*}$ and $\vec{\tau}_{K^*,L}$, we obtain the desired property for $x \in K_1$.

For $x \in K_2$, the computation of the gradient is similar. □

Lemma 5.2 *With hypothesis 2.1, let $\mathcal{S}$ be a sequence of discretization $\mathcal{D}^{**} = (\mathcal{H}_{\mathcal{D}}, h_{\mathcal{M}^{**}}, P_1(u), \nabla_{\mathcal{D},\Lambda})$ previously defined. We assume that there exists θ, such that for all $\mathcal{D}^{**} \in \mathcal{S}$, for $K \in \mathcal{M}^{**}_{\Lambda}$*

- $|\frac{\vec{n}_\sigma\Lambda_2\vec{n}_K}{|K_2|} - \frac{\vec{n}_\sigma\Lambda_1\vec{n}_L}{|K_1|}| \geq \theta(\frac{\vec{n}_\sigma\Lambda_1\vec{n}_\sigma}{|K_1|} + \frac{\vec{n}_\sigma\Lambda_2\vec{n}_\sigma}{|K_2|})$, *($H_1$)*
- $\frac{min(|\vec{\tau}_{K,L}|,|\vec{\tau}_{K,K^*}|,|\vec{\tau}_{K,K^*}|)}{max(|\vec{n}_\sigma|,|\vec{n}_{K^*,1}|,|\vec{n}_{K^*,2}|)} > \theta$, *($H_2$).*

Then, there exists a constant C_2 such that the gradients $\nabla_{K_1}u$ and $\nabla_{K_2}u$ satisfy:

$$|K_1|\nabla_{K_1}u = (u_{K^*}-u_K)\vec{\theta}_1(K,K^*) + (u_L-u_{K^*})\vec{\theta}_1(K^*,L) + (u_K-u_L)\vec{\theta}_1(L,K)$$

and

$$|K_2|\nabla_{K_2}u = (u_{K^*}-u_K)\vec{\theta}_2(K,K^*) + (u_L-u_{K^*})\vec{\theta}_2(K^*,L) + (u_K-u_L)\vec{\theta}_2(L,K).$$

with $|\vec{\theta}_1(K,K^*)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{K,K^*}|$, $|\vec{\theta}_1(K^*,L)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{K^*,L}|$, $|\vec{\theta}_1(L,K)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{L,K}|$, $|\vec{\theta}_2(K,K^*)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{K,K^*}|$, $|\vec{\theta}_2(K^*,L)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{K^*,L}|$ and $|\vec{\theta}_2(L,K)| \leq \frac{C_2}{\theta^2}|\vec{\tau}_{L,K}|$.

Remark 5.1: In order to satisfy assumption (H_1), it is sufficient to choose the primary mesh points x_K and x_L close enough from x_σ. In practice, this is a light hypothesis.

Proof of lemma 5.2

Let us denote $Det = -\frac{\vec{n}_\sigma\Lambda_1\vec{n}_L}{|K_1|} + \frac{\vec{n}_\sigma\Lambda_2\vec{n}_K}{|K_2|}$.
We can write:

$$u_\sigma = \beta_K u_K + \beta_L u_L + \beta_{K^*}u_{K^*} \tag{5.2}$$

with $\beta_K + \beta_L + \beta_{K^*} = 1$ where $\beta_K = \frac{\vec{n}_\sigma \Lambda_1 \vec{n}_\sigma}{Det|K_1|}$, $\beta_L = \frac{\vec{n}_\sigma \Lambda_2 \vec{n}_\sigma}{Det|K_2|}$ and $\beta_{K^*} = 1 - \beta_K - \beta_L$.

Using assumption (H_1), we obtain $|Det| \geq \theta(\frac{\vec{n}_\sigma \Lambda_1 \vec{n}_\sigma}{|K_1|} + \frac{\vec{n}_\sigma \Lambda_2 \vec{n}_\sigma}{|K_2|})$.

We deduce $|\beta_K| \leq \frac{1}{\theta}$, $|\beta_L| \leq \frac{1}{\theta}$ and $\beta_{K^*} \leq \frac{2}{\theta} + 1$. Using the formula:

$$2|K_1| \nabla_{K_1} u = -(u_{K^*} - u_\sigma)\vec{n}_{K^*,1} - (u_K - u_\sigma)\vec{n}_\sigma,$$

$$2|K_2| \nabla_{K_2} u = -(u_{K^*} - u_\sigma)\vec{n}_{K^*,2} - (u_L - u_\sigma)\vec{n}_\sigma$$

and equality (5.2), we obtain: $\vec{\theta}_1(K, K^*) = -\frac{\beta_K}{2}\vec{n}_{K^*,1} + \frac{\beta_{K^*}}{2}\vec{n}_\sigma$, $\vec{\theta}_1(K^*, L) = \frac{\beta_L}{2}\vec{n}_{K^*,1}$, $\vec{\theta}_1(L, K) = -\frac{\beta_L}{2}\vec{n}_\sigma$, $\vec{\theta}_2(K, K^*) = -\frac{\beta_K}{2}\vec{n}_{K^*,2}$, $\vec{\theta}_2(K^*, L) = \frac{\beta_L}{2}\vec{n}_{K^*,2} - \frac{\beta_{K^*}}{2}\vec{n}_\sigma$ and $\vec{\theta}_2(L, K) = \frac{\beta_L}{2}\vec{n}_\sigma$.
We conclude using assumption (H_2). $\square$

Proposition 5.1: *With hypothesis 2.1, let $\mathcal{S}$ be a sequence of discretizations $\mathcal{D}^{**} = (\mathcal{H}_\mathcal{D}, h_{\mathcal{M}^{**}}, P_1(u), \nabla_{\mathcal{D},\Lambda})$ previously defined. We assume that there exists θ, such that for all $\mathcal{D}^{**} \in \mathcal{S}$:*

- $\rho_K > \theta h_K \ \forall K \in \mathcal{M}^{**}_{Const}$ *(H_1),*
- $\rho_{K_1} > \theta h_{K_1} \ \forall K \in \mathcal{M}^{**} \backslash \mathcal{M}^{**}_{Const}$ *(H_2),*
- $\rho_{K_2} > \theta h_{K_2} \ \forall K \in \mathcal{M}^{**} \backslash \mathcal{M}^{**}_{Const}$ *(H_3),*
- $d(K, L) > \frac{1}{\theta}|\vec{\tau}_{K,L}|$, $d(K, K^*) > \frac{1}{\theta}|\vec{\tau}_{K,K^*}|$, $d(L, K^*) > \frac{1}{\theta}|\vec{\tau}_{L,K^*}|$, $\forall K \in \mathcal{M}^{**}$ *(H_4),*
- $|\frac{\vec{n}_\sigma \Lambda_2 \vec{n}_K}{|K_2|} - \frac{\vec{n}_\sigma \Lambda_1 \vec{n}_L}{|K_1|}| \geq \theta(\frac{\vec{n}_\sigma \Lambda_1 \vec{n}_\sigma}{|K_1|} + \frac{\vec{n}_\sigma \Lambda_2 \vec{n}_\sigma}{|K_2|})$ *for $K \in \mathcal{M}^{**}_\Lambda$ (H_5),*
- $\frac{min(|\vec{\tau}_{K,L}|, |\vec{\tau}_{K,K^*}|, |\vec{\tau}_{K,K^*}|)}{max(|\vec{n}_\sigma|, |\vec{n}_{K^*,1}|, |\vec{n}_{K^*,2}|)} > \theta$ *for $K \in \mathcal{M}^{**}_\Lambda$ (H_6),*
- $\frac{min(|\vec{\tau}_{K,L}|, |\vec{\tau}_{K,K^*}|, |\vec{\tau}_{K,K^*}|)}{|\vec{n}_L|} \frac{|K_1|}{|K|} > \theta$ *and*
  $\frac{min(|\vec{\tau}_{K,L}|, |\vec{\tau}_{K,K^*}|, |\vec{\tau}_{K,K^*}|)}{|\vec{n}_K|} \frac{|K_2|}{|K|} > \theta$ *for $K \in \mathcal{M}^{**} \backslash \{\mathcal{M}^{**}_\Lambda \cup \mathcal{M}^{**}_{Const}\}$ (H_7).*

Then, the FECCB scheme is coercive, that is to say there exists $C_{\mathcal{D}^{**}}$ such that

$$||P_1(u)||_{L^2(\Omega)} \leq C_{\mathcal{D}^{**}} ||\nabla_{\mathcal{D},\Lambda} u||, \ u \in \mathcal{H}_\mathcal{D}.$$

*Moreover $\forall \varphi \in [C_c^\infty(\Omega)]$, $lim_{h_{\mathcal{M}^{**}} \to 0} S(\varphi) = 0$ and $\forall \vec{\varphi} \in [C_c^\infty(\Omega)]^2$
$lim_{h_{\mathcal{M}^{**}} \to 0} W_{\mathcal{D}^{**}}(\vec{\varphi}) = 0$. With these three properties, we can apply the corollary 2.3 described in [26]. It means that the FECCB scheme is convergent, that is to say, $P_1(u)$*

converges to the exact solution u_{exa} of the problem and $\nabla_{\mathcal{D},\Lambda} u$ tends to ∇u_{exa} as $h_{\mathcal{M}^{**}} \to 0$.

Proof of proposition 5.1:

Following Lemma 5.3 in [23], there exists C_3 only depending on θ (Poincaré inequality) such that:

$$||\Pi^0_{\mathcal{D}^{**}} u||^2_{L^2(\Omega)} \leq C_3 ||u||^2_{1,\mathcal{D}^{**}}, \forall u \in \mathcal{H}_{\mathcal{D}}.$$

Let us show that there exists C_4 only depending on Ω and θ such that:

$$||u||^2_{1,\mathcal{D}^{**}} \leq C_4 ||\nabla_{\mathcal{D},\Lambda} u(x)||^2_{\{L^2(\Omega)\}^2}. \tag{5.3}$$

The proof is close to the one described in [26] (lemma 3.1). We denote by $\nabla_K u$ the value of $\nabla_{\mathcal{D},\Lambda} u(x)$ if $K \in \mathcal{M}^{**}_{Const}$, and $\nabla_{K_1} u$ (resp. $\nabla_{K_2} u$) the value of $\nabla_{\mathcal{D},\Lambda} u(x)$ on K_1 (resp. K_2) if $K \in \mathcal{M}^{**} \backslash \mathcal{M}^{**}_{Const}$. For all $K \in \mathcal{M}^{**}_{Const}$, for all $s \in \mathcal{V}^{**}_K$ and $r \in \mathcal{V}^{**}_K$, we can write:

$$u_s - u_r = \nabla_K u.(s - r).$$

We obtain:

$$||\nabla_K u||^2 \geq \frac{1}{3}\left(\frac{(u_K - u_L)^2}{[x_K, x_L]^2} + \frac{(u_K - u_{K^*})^2}{[x_K, x_{K^*}]^2} + \frac{(u_L - u_{K^*})^2}{[x_L, x_{K^*}]^2}\right)$$

and

$$|K|||\nabla_K u||^2 \geq \frac{1}{3}\left((u_K - u_L)^2 \frac{\rho_K}{[x_K, x_L]} + (u_K - u_{K^*})^2 \frac{\rho_K}{[x_K, x_{K^*}]} + (u_L - u_{K^*})^2 \frac{\rho_K}{[x_L, x_{K^*}]}\right). \tag{5.4}$$

For $K \in \mathcal{M}^{**} \backslash \mathcal{M}^{**}_{Const}$, we get:

$$||\nabla_{K_1} u||^2 \geq \frac{1}{2}\left(\frac{(u_K - u_\sigma)^2}{[x_K, x_\sigma]^2} + \frac{(u_K - u_{K^*})^2}{[x_K, x_{K^*}]^2}\right)$$

and

$$||\nabla_{K_2} u||^2 \geq \frac{1}{2}\left(\frac{(u_L - u_\sigma)^2}{[x_\sigma, x_L]^2} + \frac{(u_L - u_{K^*})^2}{[x_L, x_{K^*}]^2}\right).$$

We deduce that there exists C_5 such that:

$$|K_1|||\nabla_{K_1} u||^2 + |K_2|||\nabla_{K_2} u||^2 \geq C_5\{(min(\tfrac{\rho_{K_1}}{[x_K, x_\sigma]}, \tfrac{\rho_{K_2}}{[x_L, x_\sigma]})(u_K - u_\sigma)^2 + (u_L - u_\sigma)^2) +$$

$$\tfrac{\rho_{K_1}}{[x_K, x_{K^*}]}(u_K - u_{K^*})^2 + \tfrac{\rho_{K_2}}{[x_L, x_{K^*}]}(u_L - u_{K^*})^2\}. \tag{5.5}$$

Using the inequality $(u_K - u_\sigma)^2 + (u_L - u_\sigma)^2 \geq \frac{1}{2}(u_K - u_L)^2$, we obtain that there exists C_{21} such that:

$$\sum_{K \in \mathcal{M}^{**} \backslash \mathcal{M}^{**}_{Const}} |K_1|||\nabla_{K_1} u||^2 + |K_2|||\nabla_{K_2} u||^2 \geq$$

$C_{21}\sum_{K\in\mathcal{M}^{**}\backslash\mathcal{M}^{**}_{Const}} min(\frac{\rho_{K_1}}{[x_K,x_\sigma]},\frac{\rho_{K_2}}{[x_L,x_\sigma]})(u_K-u_L)^2+\frac{\rho_{K_1}}{[x_K,x_{K^*}]}(u_K-u_{K^*})^2+$

$\frac{\rho_{K_2}}{[x_L,x_{K^*}]}(u_L-u_{K^*})^2$.

Using assumptions (H_1), (H_2), (H_3) and inequality (5.4), there exists C_{23} such that:

$$||\nabla_{\mathcal{D},\Lambda}u(x)||^2 \geq C_{23}\sum_{K\in\mathcal{M}^{**}}\theta(u_K-u_L)^2+\theta(u_K-u_{K^*})^2+\theta(u_L-u_{K^*})^2.$$

Using assumption (H_4), we get:

$$||\nabla_{\mathcal{D},\Lambda}u(x)||^2 \geq C_{23}\sum_{K\in\mathcal{M}^{**}}\frac{|\vec{\tau}_{K,L}|}{d(K,L)}(u_K-u_L)^2+\frac{|\vec{\tau}_{K,K^*}|}{d(K,K^*)}(u_K-u_{K^*})^2+\frac{|\vec{\tau}_{L,K^*}|}{d(L,K^*)}(u_L-u_{K^*})^2,$$

which is the desired inequality.

Moreover, for $K\in\mathcal{M}^{**}$, we write:

$$\text{if } x\in Ai_K,\quad P_1(u)(x)=\Pi^0_{\mathcal{D}^{**}}u(x)+\nabla_{P_{1,K}}u.(x-x_K),$$

$$\text{if } x\in Ai_L,\quad P_1(u)(x)=\Pi^0_{\mathcal{D}^{**}}u(x)+\nabla_{P_{1,L}}u.(x-x_L),$$

$$\text{if } x\in Ai_{K^*},\quad P_1(u)(x)=\Pi^0_{\mathcal{D}^{**}}u(x)+\nabla_{P_{1,K^*}}u.(x-x_{K^*}).$$

We obtain that

$$||P_1(u)||_{L^2(\Omega)}\leq||\Pi^0_{\mathcal{D}^{**}}u||_{L^2(\Omega)}+h_{\mathcal{M}^{**}}||\nabla_{P_1}u(x)||_{\{L^2(\Omega)\}^2}. \tag{5.6}$$

With Lemma 5.1, we obtain the formulation:

$$|K|\nabla_{P_1}u=(u_K-u_{K^*})\vec{\tau}_{K,K^*}+(u_{K^*}-u_L)\vec{\tau}_{K^*,L}+(u_L-u_K)\vec{\tau}_{L,K}\ for\ x\in K,\ for\ K\in\mathcal{M}^{**}.$$

We define $S_K=((K,L),(K,K^*),(K^*,L))$. Using the Cauchy-Schwarz inequality and $\sum_{K\in\mathcal{M}^{**}}\sum_{(M,N)\in S_K}|\vec{\tau}_{M,N}|d(M,N)=2|\Omega|$, we get that there exists C_{20} such that:

$$||\nabla_{P_1}u||_{\{L^2(\Omega)\}^2}\leq C_{20}||u||_{1,\mathcal{D}^{**}}. \tag{5.7}$$

Using (5.3), (5.6) and (5.7), we obtain that there exists C_{22} such that:

$$||P_1(u)||_{L^2(\Omega)}\leq C_{22}||\nabla_{\mathcal{D},\Lambda}u(x)||_{\{L^2(\Omega)\}^2}. \tag{5.8}$$

We conclude that the scheme is coercive.

Let us estimate the strong consistency of the discretization. Let $\varphi\in[C^\infty_c(\Omega)]$. As we use the Stokes formula to approximate the gradient, with assumptions (H_1), (H_2) and (H_3), we obtain in the same way as lemma 4.3 described in [23] that there exists C_6 only depending on θ and φ such that

$$||\nabla_{\mathcal{D},\Lambda}\varphi-\nabla\varphi||_{L^2(\Omega)^2}\leq C_6h_{M^{**}}.$$

Moreover, using Lemma 3.1 in [54], we obtain that $||P_1(\varphi) - \varphi||_{L^2(\Omega)} \leq h_{M^{**}}||\nabla\varphi||_{L^2(\Omega)^2}$. We deduce that the interpolation error function $S(\varphi)$ tends toward zero if $h_{M^{**}}$ tends toward zero.

Let $\vec{\varphi} \in [C_c^\infty(\Omega)]^2$. Let us compute $T = \int_\Omega (\nabla_{\mathcal{D},\Lambda} u(x).\vec{\varphi} + P_1(u) div\vec{\varphi}(x))dx$.

We denote by $\vec{\varphi}_K$ the average value of $\vec{\varphi}(x)$ if $K \in \mathcal{M}^{**}\backslash\mathcal{M}_\Lambda^{**}$, $\vec{\varphi}_{K_1}$ (resp. $\vec{\varphi}_{K_2}$) the average value of $\vec{\varphi}(x)$ on K_1 (resp. K_2) if $K^{**} \in \mathcal{M}_\Lambda^{**}$, $\vec{\varphi}_{M,N(M,N)\in S_K}$ the average value of $\vec{\varphi}(x)$ on $\vec{\tau}_{M,N,(M,N)\in S_K}$.

We get $T = T_1 + T_2 + T_3$ with

$$\begin{aligned}
T_1 &= \sum_{K\in\mathcal{M}_{Const}^{**}} |K|\nabla_K u.\vec{\varphi}_K + \sum_{K\in\mathcal{M}^{**}\backslash\{\mathcal{M}_{Const}^{**}\cup\mathcal{M}_\Lambda^{**}\}} |K_1|\nabla_{K_1} u.\vec{\varphi}_{K_1} \\
&+ |K_2|\nabla_{K_2} u.\vec{\varphi}_{K_2^{**}} + \sum_{K\in\mathcal{M}_\Lambda^{**}} |K_1|\nabla_{K_1} u.\vec{\varphi}_{K_1} + |K_2|\nabla_{K_2} u.\vec{\varphi}_{K_2^{**}} \\
&= T_{11} + T_{12} + T_{13} + T_{14} + T_{15},
\end{aligned}$$

$$T_2 = \sum_{K\in\mathcal{M}^{**}} (u_K - u_{K^*})\vec{\tau}_{K,K^*}.\vec{\varphi}_{K,K^*} + (u_{K^*} - u_L)\vec{\tau}_{K^*,L}.\vec{\varphi}_{K^*,L} + (u_L - u_K)\vec{\tau}_{L,K}.\vec{\varphi}_{K,L},$$

$$\begin{aligned}
T_3 &= \sum_{K\in\mathcal{M}^{**}} \int_{Ai_K} \nabla_{P_{1,K}} u.(x - x_K) div\vec{\varphi}(x)dx + \int_{Ai_L} \nabla_{P_{1,L}} u.(x - x_L) div\vec{\varphi}(x)dx \\
&+ \int_{Ai_{K^*}} \nabla_{P_{1,K^*}} u.(x - x_{K^*}) div\vec{\varphi}(x)dx.
\end{aligned}$$

We obtain:

$$|T_3| \leq h_{\mathcal{M}}^{**}||\nabla_{P_1} u||_{\{L^2(\Omega)\}^2}||div\vec{\varphi}||_{L^2(\Omega)}. \tag{5.9}$$

With assumption (H_7) and Lemma 5.1, we get:

$$\begin{aligned}
T_{11} + T_{12} + T_{13} &= \sum_{K\in\mathcal{M}^{**}\backslash\mathcal{M}_\Lambda^{**}} \{(u_{K^*} - u_K)(\vec{\tau}_{K,K^*} + \vec{\epsilon}_{K,K^*}) + \\
&\quad (u_L - u_{K^*})(\vec{\tau}_{K^*,L} + \vec{\epsilon}_{K^*,L}) + (u_K - u_L)(\vec{\tau}_{L,K} + \vec{\epsilon}_{L,K})\}.\vec{\varphi}_K.
\end{aligned}$$

On the other hand, with Lemma 5.2, we write:

$$T_{14} = \sum_{K\in\mathcal{M}_\Lambda^{**}} ((u_{K^*} - u_K)\vec{\theta}_1(K,K^*) + (u_L - u_{K^*})\vec{\theta}_1(K^*,L) + (u_K - u_L)\vec{\theta}_1(L,K)).\vec{\varphi}_{K_1}$$

and

$$T_{15} = \sum_{K\in\mathcal{M}_\Lambda^{**}} ((u_{K^*} - u_K)\vec{\theta}_2(K,K^*) + (u_L - u_{K^*})\vec{\theta}_2(K^*,L) + (u_K - u_L)\vec{\theta}_2(L,K)).\vec{\varphi}_{K_2}.$$

Let us define

$$Term_1 = \sum_{K \in \mathcal{M}^{**} \setminus \mathcal{M}^{**}_\Lambda} \sum_{(M,N) \in S_K} (|\vec{\tau}_{M,N}| + |\vec{\epsilon}_{M,N})|) d(M,N)(|\vec{\varphi}_K - \vec{\varphi}_{M,N}|^2)$$

and

$$Term_2 = \sum_{K \in \mathcal{M}^{**}_\Lambda} \sum_{(M,N) \in S_K} |\vec{\tau}_{M,N}| d(M,N)(|\frac{\vec{\theta}_1(M,N).\vec{\varphi}_{K_1}}{|\vec{\tau}_{M,N}|} + \frac{\theta_2(M,N)\vec{\varphi}_{K_2}}{|\vec{\tau}_{M,N}|} - \vec{\varphi}_{M,N}|^2).$$

Using the Cauchy-Schwarz inequality, we obtain:

$$|T_1 + T_2|^2 \leq ||u||^2_{1,\mathcal{D}^{**}} (Term_1 + Term_2).$$

Besides, using the regularity of $\vec{\varphi}$, there exists a constant C_ϕ such that

$$\forall (M,N) \in S_K \quad |\varphi_K - \varphi_{M,N}|^2 \leq C_\phi h^2_{\mathcal{M}^{**}}.$$

Using $\sum_{K \in \mathcal{M}^{**}} \sum_{(M,N) \in S_K} |\vec{\tau}_{M,N}| d(M,N) = 2|\Omega|$, and Lemma 5.1, there exists C_{30} such that

$$Term_1 \leq C_{30}|\Omega|(h^2_{\mathcal{M}^{**}}(1 + \epsilon_1(h_{\mathcal{M}^{**}})))$$

with $lim_{h_{\mathcal{M}^{**}} \to 0} \epsilon_1(h_{\mathcal{M}^{**}}) = 0$.

Moreover, using the regularity of $\vec{\varphi}$, assumptions (H_5), (H_6) and Lemma 5.2, we get that there exists C_{10} depending on θ and $\vec{\varphi}$ such that

$$\forall (M,N) \in S_K \ (|\frac{\vec{\theta}_1(M,N).\vec{\varphi}_{K_1}}{|\vec{\tau}_{M,N}|} + \frac{\theta_2(M,N)\vec{\varphi}_{K_2}}{|\vec{\tau}_{M,N}|} - \vec{\varphi}_{M,N}|^2) \leq C_{10}.$$

Following the same arguments as those described in [21] (Theorem 3.8), as the tensor $\lambda(x)$ is piecewise Lipschitz-continuous, we deduce that $\sum_{K \in \mathcal{M}^{**}_\Lambda} \sum_{(M,N) \in S_K} |\vec{\tau}_{M,N}| d(M,N)$ tends toward zero if $h_{\mathcal{M}^{**}}$ tends toward zero. (It means that the dimension of the zones where the tensor $\lambda(x)$ is discontinuous is inferior to one.)

We finally obtain that there exists C_8 such that:

$$|T_1 + T_2| \leq C_8 ||u||_{1,\mathcal{D}^{**}} (h_{\mathcal{M}^{**}} + \epsilon_2(h_{\mathcal{M}^{**}}))$$

with $lim_{h_{\mathcal{M}^{**}} \to 0} \epsilon_2(h_{\mathcal{M}^{**}}) = 0$. Using (5.3), (5.9) and (5.7), we deduce that there exists C_9 only depending on θ and $\vec{\varphi}$ such that $W_{\mathcal{D}^{**}}(\vec{\varphi}) \leq C_9(h_{\mathcal{M}^{**}} + \epsilon(h_{\mathcal{M}^{**}}))$. This is the dual consistency described in [26]. □

Corollary 5.1 *With hypothesis 2.1, let $\mathcal{S}$ be a sequence of discretizations $\mathcal{D}^{**} = (\mathcal{H}_\mathcal{D}, h_{\mathcal{M}^{**}}, P(u), \nabla_{\mathcal{D},\Lambda})$ defined in (2.3). Under the assumptions of Proposition 5.1, the FECC scheme is convergent, that is to say, $P(u)$ converges to the exact solution u_{exa} of the problem and $\nabla_{\mathcal{D},\Lambda} u$ tends to ∇u_{exa} as $h_{\mathcal{M}^{**}} \to 0$.*

Proof of corollary 5.1

We measure the strong consistency of the FECC scheme with the interpolation error function $S_i(\varphi) = \{||P(\varphi) - \varphi||^2_{L^2(\Omega)} + ||\nabla_{\mathcal{D},\Lambda}\varphi - \nabla\varphi||^2_{L^2(\Omega)^2}\}^{\frac{1}{2}}$, $\varphi \in [C_c^\infty(\Omega)]$ and the dual consistency with the conformity error function

$$W_{i,\mathcal{D}^{**}}(\vec{\varphi}) = max_{u\in\mathcal{H}_\mathcal{D}}||\frac{1}{\nabla_{\mathcal{D},\Lambda}u(x)}||\int_\Omega(\nabla_{\mathcal{D},\Lambda}u(x).\vec{\varphi} + P(u)(x)div\vec{\varphi}(x))dx, \forall\vec{\varphi}\in[C_c^\infty(\Omega)]^2.$$

Using the definition of $P(u)$ and $P_1(u)$, we obtain that:

$$||P(u) - P_1(u)||_{L^2(\Omega)} \leq h_{\mathcal{M}^{**}}(||\nabla_{P_1}u||_{L^2(\Omega)^2} + ||\nabla_{\mathcal{D},\Lambda}u||_{L^2(\Omega)^2}).$$

With (5.3) and (5.7), there exists C_{40} such that:

$$||P(u) - P_1(u)||_{L^2(\Omega)} \leq C_{40}h_{\mathcal{M}^{**}}||\nabla_{\mathcal{D},\Lambda}u||_{L^2(\Omega)^2}. \tag{5.10}$$

In the same way, for $\varphi \in [C_c^\infty(\Omega)]$, we get that:

$$||P(\varphi) - P_1(\varphi)||_{L^2(\Omega)} = \epsilon_3(h_{\mathcal{M}^{**}}) \tag{5.11}$$

with $lim_{h_{\mathcal{M}^{**}}\to 0}\epsilon_3(h_{\mathcal{M}^{**}}) = 0$. Using (5.8), we deduce that there exists C_{41} such that

$$||P(u)||_{L^2(\Omega)} \leq C_{41}||\nabla_{\mathcal{D},\Lambda}u(x)||_{\{L^2(\Omega)\}^2}. \tag{5.12}$$

We deduce that the FECC scheme is coercive. Using proposition 5.1, (5.10) and (5.11), we obtain that $\forall\varphi\in[C_c^\infty(\Omega)]$, $lim_{h_{\mathcal{M}^{**}}\to 0}S_i(\varphi) = 0$ and $\forall\vec{\varphi}\in[C_c^\infty(\Omega)]^2$ $lim_{h_{\mathcal{M}^{**}}\to 0}W_{i,\mathcal{D}^{**}}(\vec{\varphi}) = 0$. We conclude applying corollary 2.3 described in [26]. □

Part IV

Discrete Maximum Principle

Chapter 6

Non-linear corrections of the FECC schemes

Contents

6.1 Motivation.

For diffusion terms in models of flows in porous media, for coupling transport equations with a chemical model and for heat flows from a cold material to a hot one, the existence of a maximum principle of their solutions plays very important role, because violation of the maximum principle might lead to non-physical solutions. Unfortunately, classical finite volume and finite element schemes fail to satisfy the maximum principle for distorted meshes or for high anisotropy ratio of diffusion tensors [19], [45], [34], [41]. In [11], to satisfy the maximum principle, almost the discretization schemes for the Laplace operator must comply with some type of geometrical constraint on the mesh, such as 2D & 3D acute type condition in [14], [38]. Additionally, in [45], the authors proved that it is impossible to construct nine-point methods which unconditionally satisfy the monotonicity criteria when the discretization satisfies local conservation and exact reproduction of linear potential fields. However, in practical, these conditions are difficult to satisfy, especially in three dimensions. The FECC scheme also violates the discrete maximum principle: examples are shown in the section 7.3 of chapter 7. Our objective is to construct non-linear corrections for the FECC schemes providing a discrete maximum principle. The non-linear FECC schemes still preserve the main properties of the scheme including coercivity, symmetry, existence of a solution. Moreover, we also have the condition for convergence toward the solution of (2.1) when the size of the meshes tends to 0. Besides, it is easy for us to implement this modification, because we can use the computed data of the linear systems associated to the FECC schemes. The spirit of these non-linear corrections is based on the methods presented in [53] and [13]. In fact, some authors also proposed non-linear finite volume schemes [21], [29], [35], [51], [43], [55], [30] to discretize elliptic problems. These schemes satisfy the desired properties and accurate results. However, they are coercive with conditions on the meshes and on the anisotropy ratio.

6.2 Notations.

Firstly, we introduce some notations and definitions used in this chapter:

- For the definition of $\mathcal{E}$, we further assume that, for all $\sigma \in \mathcal{E}$, either $\sigma \in \partial\Omega$ or $\overline{\sigma} = \overline{K} \cap \overline{L}$ for $(K, L) \in \mathcal{M} \times \mathcal{M}$.

- For $\sigma \in \mathcal{E}$, we set $d_\sigma = \begin{cases} d_{K,\sigma} + d_{L,\sigma} & \text{if } \sigma = K|L \in \mathcal{E}_{\text{int}} \\ d_{K,\sigma} & \text{if } \sigma = E_K \cap \mathcal{E}_{ext} \end{cases}$, where $d_{K,\sigma}$ is the orthogonal distance between x_K and the hyperplane containing σ.

- The size of the meshes is denoted by $\text{size}(\mathcal{D}) = \sup_{K \in \mathcal{M}} \text{diam}(K)$,

- The coefficient η which is a positive constant is considered in the section 7.3 of Chapter 7.

- The regularities of the primary and dual meshes are defined by:

$$\begin{aligned}
\text{regul}(\mathcal{D}) &= \sup_{K\in\mathcal{M}}\left\{\max\left(\frac{\text{diam}(K)^d}{\rho_K^d}, \text{Card}(\mathcal{E}_K)\right)\right\} + \sup_{\substack{K\in\mathcal{M}\\ \sigma\in\mathcal{E}_K}}\left\{\frac{\text{diam}(K)}{d_{K,\sigma}}\right\} + \sup_{\substack{K\in\mathcal{E}_{int}\\ \sigma=K|L}}\left\{\frac{d_{L,\sigma}}{d_{K,\sigma}}\right\},\\
\text{regul}(\mathcal{D}^*) &= \sup_{K^*\in\mathcal{M}^*}\{\text{Card}(\mathcal{E}^*_{K^*})\} + \sup_{\substack{K^*\in\mathcal{M}^*\\ \sigma^*\in\mathcal{E}^*_{K^*}}}\left\{\frac{\text{diam}(K^*)}{d_{K^*,\sigma^*}}\right\} + \sup_{\substack{K^*\in\mathcal{E}^*_{int}\\ \sigma=K^*|L^*}}\left\{\frac{d_{L^*,\sigma^*}}{d_{K^*,\sigma^*}}\right\},
\end{aligned}$$

  where for $K\in\mathcal{M}$, ρ_K is the supremum of the radius of the balls contained in K, and $d=2,3$.

- $\mathcal{H}_\mathcal{M}$ is the set of all functions u containing $u=(u_K)_{K\in\mathcal{M}}\in\mathbb{R}^{Card(\mathcal{M})}$ such that u is defined on Ω, is constant on each control volume of $\mathcal{M}$ and takes the value u_K in the cell $K\in\mathcal{M}$. The space $\mathcal{H}_\mathcal{M}$ is equipped with the discrete norm defined by:

$$\forall u\in\mathcal{H}_\mathcal{M}, ||u||^2_\mathcal{D} = \sum_{\substack{\sigma\in\mathcal{E}_{\text{int}}\\ \sigma\in\mathcal{E}_K\cap\mathcal{E}_L}} |\sigma|\frac{|u_K-u_L|^2}{d_\sigma} + \sum_{\sigma\in\mathcal{E}_{ext}} |\sigma|\frac{|u_K|^2}{d_\sigma}.$$

- The set V(K) is defined in the two following cases:

  To correct the initial FECC scheme: For each $K\in\mathcal{M}\cup\mathcal{M}^*$, $V(K)$ which belongs to $\{\mathcal{M}\cup\mathcal{M}^*\cup\mathcal{E}_{ext}\}$ corresponds to the stencil of the two systems (6.3) and (6.4). The stencil is symmetric, i.e.

$$\forall(K,L)\in(\mathcal{M}\cup\mathcal{M}^*,\mathcal{M}\cup\mathcal{M}^*),\quad L\in V(K)\Rightarrow K\in V(L). \tag{6.1}$$

  To correct the second FECC scheme: For each $K\in\mathcal{M}$, $V(K)$ is a subset of $\mathcal{M}\cup\mathcal{E}_{ext}$ corresponding to the stencil of the system (6.6). The stencil is also symmetric, i.e.

$$\forall(K,L)\in\mathcal{M}^2,\quad L\in V(K)\Rightarrow K\in V(L). \tag{6.2}$$

- **Definition 6.1** (Discrete Maximum Principle). The non-linear FECC schemes satisfy the following discrete maximum principle: if $f\geq 0$ on Ω, then the solution $u=(u_K)_{K\in\mathcal{M}}$ of the non-linear FECC schemes satisfy $min_{K\in\mathcal{M}}u_K\geq 0$.

Before we represent the non-linear corrections, we recall the initial FECC scheme and introduce another FECC scheme named "the second FECC scheme".

The initial FECC scheme: From (4.3) and (4.4) of chapter 2, we construct the two systems of linear equations

$$DU^* + EU = F^*, \tag{6.3}$$
$$MU^* + NU = F. \tag{6.4}$$

where $U = (U_K)_{K\in\mathcal{M}}, U^* = (U^*{}_{K^*})_{K^*\in\mathcal{M}^*}, D \in M^{Card(\mathcal{M}^*)\times Card(\mathcal{M}^*)}, E \in M^{Card(\mathcal{M}^*)\times Card(\mathcal{M})}$, $M \in M^{Card(\mathcal{M})\times Card(\mathcal{M}^*)}, N \in M^{Card(\mathcal{M})\times Card(\mathcal{M})}$. The two operators $F, F^* \in M^{Card(\mathcal{M}^*)\times 1}$ are defined by:

$$F = \left(\int_\Omega f(x).P(p_K)(x)dx \right)_{K\in M}, \quad F^* = \left(\int_\Omega f(x).P(p_{K^*})(x)dx \right)_{K^*\in M^*},$$

where let be K belonging to $\mathcal{M}\cup\mathcal{M}^*$, we define $P(p_K)(x_K) = 1$ and $P(p_K)(x_L) = 0$ with $L \neq K$ and $L \in \mathcal{M}\cup\mathcal{M}^*$. Moreover, for all $K \in \mathcal{M}\cup\mathcal{M}^*$, $P(p_K)(x)$ is equal to 0 on the boundary $\partial\Omega$.

Both left hand sides of (6.3) and (6.4) $\begin{pmatrix} D & E \\ M & N \end{pmatrix} . \begin{pmatrix} U^* \\ U \end{pmatrix}$ can be rewritten by $-\mathcal{A}^{\mathcal{D},\mathcal{D}^*}(u)$, where $u \in \mathcal{H}_\mathcal{D}$ and the discrete linear operator $\mathcal{A}^{\mathcal{D},\mathcal{D}^*} : \mathcal{H}_\mathcal{D} \to \mathcal{H}_\mathcal{D}$ is defined by:

$$\forall u \in \mathcal{H}_\mathcal{D}, \forall K \in \{\mathcal{M}\cup\mathcal{M}^*\}, \quad \mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u) = \sum_{Z\in V(K)} \alpha_{K,Z}(u_Z - u_K).$$

From the two systems (6.3) and (6.4), the system of linear equations (6.5) is built as (2.9) (in 2D homogeneous isotropic cases), or (2.12) (2D heterogeneous anisotropic cases), or (3.6) (in three dimensions)

$$\underbrace{A}_{N-E^TD^{-1}E} .U = \underbrace{B}_{F-E^TD^{-1}F^*}. \tag{6.5}$$

On the other hand, we can reconstruct the right hand side of (6.5) by:

$$\overline{F}^* = 0, \ \overline{F} = (|K|f_K)_{K\in\mathcal{M}}, \ \overline{B} = \overline{F},$$

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

The second FECC scheme is defined by:

$$A.u = \overline{B} \quad \text{with } u \in \mathcal{H}_\mathcal{M}, \tag{6.6}$$

where the matrix A is the same as the matrix A in (6.5). The left hand side of (6.6) can be also rewritten by $-\mathcal{A}^\mathcal{D}(u)$, where the discrete linear operator $\mathcal{A}^\mathcal{D} : \mathcal{H}_\mathcal{M} \to \mathcal{H}_\mathcal{M}$ is defined by:

$$\forall u \in \mathcal{H}_\mathcal{M}, \forall K \in \mathcal{M}, \quad \mathcal{A}_K^\mathcal{D}(u) = \sum_{Z\in V(K)} \overline{\alpha}_{K,Z}(u_Z - u_K).$$

6.3 Construction of non-linear corrections.

6.3.1 First non-linear FECC scheme.

For all $u = \left((u_K)_{K\in\mathcal{M}}, (u_{K^*})_{K^*\in\mathcal{M}^*}\right) \in \mathcal{H}_\mathcal{D}$, and all $Z \in \{V(K)\cup V(K^*)\} \subset \{\mathcal{M}\cup\mathcal{M}^*\cup\mathcal{E}_{ext}\}$, we define the first non-linear FECC scheme (**NLFECC1**), as follows: for each $K \in$

$\mathcal{M} \cup \mathcal{M}^*$,

$$\mathcal{S}_K^{\mathcal{D},\mathcal{D}^*}(u) = -\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u) + \sum_{Z \in V(K)} \beta_{K,Z}(u)(u_K - u_Z) = \int_\Omega f(x).P(p_K)(x)dx, \qquad (6.7)$$

we remark that if x_K, x_Z are on the boundary $\partial\Omega$, u_{K^*}, u_Z are equal to 0.

In (6.7), for any K belonging to $\mathcal{M} \cup \mathcal{M}^*$, $\beta_{K,Z}$ is defined by the following non-linear corrections:

The first correction $\beta^{\mathcal{D},\mathcal{D}^*}$:

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

$$\beta_{K,\sigma}^{\mathcal{D},\mathcal{D}^*}(u) = \eta.\frac{\left|\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(K)} |u_Y - u_K|}.$$

- If $Z = L \in \mathcal{M} \cup \mathcal{M}^*$, then:

$$\beta_{K,L}^{\mathcal{D},\mathcal{D}^*}(u) = \eta.\left(\frac{\left|\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(K)} |u_Y - u_K|} + \frac{\left|\mathcal{A}_L^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(L)} |u_Y - u_L|}\right).$$

  We note that $\sum\limits_{Y \in V(K)} |u_Y - u_K|$ or $\sum\limits_{Y \in V(L)} |u_Y - u_L|$ is equal to 0, then the term $\frac{\left|\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(K)} |u_Y - u_K|}$ or $\frac{\left|\mathcal{A}_L^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(L)} |u_Y - u_L|}$ disappears.

The second correction $\overline{\beta}^{\mathcal{D},\mathcal{D}^*}$:

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

$$\overline{\beta}_{K,\sigma}^{\mathcal{D},\mathcal{D}^*}(u) = \eta.\frac{\left|\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}.$$

- If $Z = L \in \mathcal{M} \cup \mathcal{M}^*$, then:

$$\overline{\beta}_{K,L}^{\mathcal{D},\mathcal{D}^*}(u) = \eta.\left\{\frac{\left|\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)} + \frac{\left|\mathcal{A}_L^{\mathcal{D},\mathcal{D}^*}(u)\right|}{\sum\limits_{Y \in V(L)} \frac{1}{2}\left(|u_Y| + |u_L|\right)}\right\}.$$

The third correction $\widetilde{\beta}^{\mathcal{D},\mathcal{D}^*}$:

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

$$\widetilde{\beta}^{\mathcal{D},\mathcal{D}^*}_{K,\sigma}(u) = \eta.\frac{\left|\mathcal{A}^{\mathcal{D},\mathcal{D}^*}_K(u)\right|}{\sum\limits_{Y\in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}.$$

- If $Z = L \in \mathcal{M} \cup \mathcal{M}^*$, then:

$$\widetilde{\beta}^{\mathcal{D},\mathcal{D}^*}_{K,L}(u) = \eta.\sup\left\{\frac{\left|\mathcal{A}^{\mathcal{D},\mathcal{D}^*}_K(u)\right|}{\sum\limits_{Y\in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}, \frac{\left|\mathcal{A}^{\mathcal{D},\mathcal{D}^*}_L(u)\right|}{\sum\limits_{Y\in V(L)} \frac{1}{2}\left(|u_Y| + |u_L|\right)}\right\}.$$

6.3.2 Second non-linear FECC scheme

For all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K)$, the second non-linear FECC scheme (**NLFECC2**) is defined by:

$$\mathcal{S}^{\mathcal{D}}_K(u) = -\mathcal{A}^{\mathcal{D}}_K(u) + \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z) = |K|.f_K. \tag{6.8}$$

For any K belonging to $\mathcal{M}$, we use the following corrections to define $\beta_{K,Z}$ in (6.8):

The first correction $\beta^{\mathcal{D}}$:

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

$$\beta^{\mathcal{D}}_{K,\sigma}(u) = \eta.\frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y\in V(K)} |u_Y - u_K|},$$

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

$$\beta^{\mathcal{D}}_{K,L}(u) = \eta.\left(\frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y\in V(K)} |u_Y - u_K|} + \frac{\left|\mathcal{A}^{\mathcal{D}}_L(u)\right|}{\sum\limits_{Y\in V(L)} |u_Y - u_L|}\right).$$

We remark that $\sum\limits_{Y\in V(K)} |u_Y - u_K|$ or $\sum\limits_{Y\in V(L)} |u_Y - u_L|$ is equal to 0, then the term $\frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y\in V(K)} |u_Y - u_K|}$ or $\frac{\left|\mathcal{A}^{\mathcal{D}}_L(u)\right|}{\sum\limits_{Y\in V(L)} |u_Y - u_L|}$ disappears.

The second correction $\overline{\beta}^{\mathcal{D}}$:

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

$$\overline{\beta}^{\mathcal{D}}_{K,\sigma}(u) = \eta.\frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y\in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}.$$

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

$$\overline{\beta}^{\mathcal{D}}_{K,L}(u) = \eta. \left\{ \frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y \in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)} + \frac{\left|\mathcal{A}^{\mathcal{D}}_L(u)\right|}{\sum\limits_{Y \in V(L)} \frac{1}{2}\left(|u_Y| + |u_L|\right)} \right\}.$$

The third correction $\widetilde{\beta}^{\mathcal{D}}$:

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

$$\widetilde{\beta}^{\mathcal{D}}_{K,\sigma}(u) = \eta. \frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y \in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}.$$

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

$$\widetilde{\beta}^{\mathcal{D}}_{K,L}(u) = \eta. \sup \left\{ \frac{\left|\mathcal{A}^{\mathcal{D}}_K(u)\right|}{\sum\limits_{Y \in V(K)} \frac{1}{2}\left(|u_Y| + |u_K|\right)}, \frac{\left|\mathcal{A}^{\mathcal{D}}_L(u)\right|}{\sum\limits_{Y \in V(L)} \frac{1}{2}\left(|u_Y| + |u_L|\right)} \right\}.$$

6.4 Properties of the non-linear FECC schemes.

Now, we consider properties of the NLFECC2 scheme using the first non-linear correction $\beta^{\mathcal{D}}$. For the other non-linear FECC schemes, we prove similarly.

The family $\beta = (\beta_{K,L})_{K \in \mathcal{M}, Z \in V(K)}$ satisfies the two following properties:

- **symmetry**: $\forall K \in \mathcal{M}, \forall Z \in V(K) \cap \mathcal{M}, \quad \beta_{K,Z} = \beta_{Z,K}$.
- **positivity**: $\forall K \in \mathcal{M}, \forall Z \in V(K), \quad \beta_{K,Z} \geq 0$.

These properties are useful to consider the following properties:

6.4.1 Coercivity.

Proposition 6.1 *Let $\theta \geq regul(\mathcal{D})$, $regul(\mathcal{D}^*)$ be given, then there exists a positive constant C_1 depending on θ, Ω, such that*

$$\sum_{K \in \mathcal{M}} \mathcal{S}^{\mathcal{D}}_K(u) u_K = -\mathcal{A}^{\mathcal{D}}(u) + \sum_{K \in \mathcal{M}} u_K \sum_{Z \in V(K)} \beta^{\mathcal{D}}_{K,L}(u)(u_K - u_Z) \geq C_1 ||u||^2_{\mathcal{D}}. \tag{6.9}$$

Proof of proposition 6.1

We firstly prove that there exists $C > 0$ only depending on θ, Ω such that:

$$\int_\Omega \left(\Lambda(x) \nabla_{\mathcal{D},\Lambda} u(x)\right) . \nabla_{\mathcal{D},\Lambda} u(x) \, dx \geq C. ||u||^2_{\mathcal{D}}. \tag{6.10}$$

Using the definition of $\mathcal{V}^*$, let each $\sigma^* \in \mathcal{E}^*_{int}$, there exist two points x_K and x_L belonging to $\mathcal{P}$ are vertices of σ^*. Besides, we define an intersecting point x_σ between σ^* and $\sigma = \mathcal{E}_K \cup \mathcal{E}_L$. This intersecting point is inside σ (see Figure 4.10). At a triangle $(x_K, x_L, x_{K^*}) \in \mathcal{M}^{**}$, we estimate the absolute value of $\nabla_{\mathcal{D},\Lambda} u$ on each a sub-triangle (x_{K^*}, x_K, x_σ), (x_{K^*}, x_L, x_σ) of (x_K, x_L, x_{K^*}), as follows:

$$\left\{\begin{array}{l} (\nabla_{\mathcal{D},\Lambda} u)_{(x_{K^*},x_K,x_\sigma)} .(x_\sigma - x_K) = u_\sigma^{K^*} - u_K, \\ (\nabla_{\mathcal{D},\Lambda} u)_{(x_{K^*},x_L,x_\sigma)} .(x_\sigma - x_L) = u_\sigma^{K^*} - u_L. \end{array}\right. \Rightarrow \left\{\begin{array}{l} \left|(\nabla_{\mathcal{D},\Lambda} u)_{(x_{K^*},x_K,x_\sigma)}\right| \geq \frac{\left|u_\sigma^{K^*} - u_K\right|}{m_{[x_K,x_\sigma]}}, \\ \left|(\nabla_{\mathcal{D},\Lambda} u)_{(x_{K^*},x_L,x_\sigma)}\right| \geq \frac{\left|u_\sigma^{K^*} - u_L\right|}{m_{[x_L,x_\sigma]}}. \end{array}\right. \tag{6.11}$$

Hence, we get that

$$\begin{aligned} \int\limits_{T_{K^*,\sigma^*}=(x_K,x_L,x_{K^*})} |\nabla_{\mathcal{D},\Lambda} u|^2 \, dx &= \int\limits_{(x_{K^*},x_K,x_\sigma)} |\nabla_{\mathcal{D},\Lambda} u|^2 \, dx + \int\limits_{(x_{K^*},x_L,x_\sigma)} |\nabla_{\mathcal{D},\Lambda} u|^2 \, dx \\ &\geq m_{(x_{K^*},x_K,x_\sigma)} \cdot \left(\frac{\left|u_\sigma^{K^*} - u_K\right|}{m_{[x_K,x_\sigma]}}\right)^2 + m_{(x_{K^*},x_L,x_\sigma)} \cdot \left(\frac{\left|u_\sigma^{K^*} - u_L\right|}{m_{[x_L,x_\sigma]}}\right)^2 \\ &\geq d_{K^*,\sigma^*} \frac{\left|u_\sigma^{K^*} - u_K\right|^2}{m_{[x_K,x_\sigma]}} + d_{K^*,\sigma^*} \frac{\left|u_\sigma^{K^*} - u_L\right|^2}{m_{[x_L,x_\sigma]}} \geq d_{K^*,\sigma^*} \frac{|u_K - u_L|^2}{|\sigma^*|}. \end{aligned}$$

Similarly,

$$\int\limits_{T_{L^*,\sigma^*}=(x_K,x_L,x_{L^*})} |\nabla_{\mathcal{D},\Lambda} u|^2 \, dx \geq d_{L^*,\sigma^*} \frac{|u_K - u_L|^2}{|\sigma^*|}.$$

For each $K^* \in \mathcal{M}^*$, $\sigma^* \in \mathcal{E}^*_{K^*}$, there exists $\sigma \in \mathcal{E}_K$ with $K \in \mathcal{M}$, such that $\sigma \cap \sigma^* \neq \emptyset$. We find a relationship between $\frac{|\sigma|}{d_\sigma}$ and $\frac{d_{\sigma^*}}{|\sigma^*|}$:

$$\left\{\begin{array}{l} \theta \geq \text{regul}(\mathcal{D}) \geq \frac{diam(K)}{d_{K,\sigma}} \geq \frac{|\sigma|}{d_{K,\sigma}} \geq \frac{|\sigma|}{d_\sigma} \\ \theta \geq \text{regul}(\mathcal{D}^*) \geq \frac{diam(K^*)}{d_{K^*,\sigma^*}} \geq \frac{|\sigma^*|}{d_{K^*,\sigma^*}} \geq \frac{|\sigma^*|}{d_{\sigma^*}} \\ |\sigma| \geq d_{\sigma^*} \quad \text{and} \quad d_\sigma \leq |\sigma^*| \end{array}\right. \Rightarrow \theta^2 . \frac{d_{\sigma^*}}{|\sigma^*|} \geq \frac{|\sigma|}{d_\sigma} \geq \frac{d_{\sigma^*}}{|\sigma^*|}, \tag{6.12}$$

where d_{σ^*} is equal to $\left\{\begin{array}{ll} d_{K^*,\sigma^*} + d_{L^*,\sigma^*} & \text{if } \sigma = K^*|L^* \in \mathcal{E}^*_{\text{int}} \\ d_{K^*,\sigma^*} & \text{if } \sigma = \mathcal{E}^*_{K^*} \cap \mathcal{E}^*_{ext} \end{array}\right.$ with the two orthogonal distances from x_{K^*} and x_{L^*} to σ^* denoted by d_{K^*,σ^*}, d_{L^*,σ^*} respectively.

By (6.11) and (6.12), it follows that

$$\begin{aligned}
\int_\Omega |\nabla_{\mathcal{D},\Lambda} u|^2\, dx &= \sum_{\substack{\sigma^*\in\mathcal{E}^*_{\rm int}\\ \sigma^*=\mathcal{E}^*_{K^*}\cap\mathcal{E}^*_{L^*}}} \left(\int_{T_{K^*,\sigma^*}\in\mathcal{M}^{**}} |\nabla_{\mathcal{D},\Lambda} u|^2\, dx + \int_{T_{L^*,\sigma^*}\in\mathcal{M}^{**}} |\nabla_{\mathcal{D},\Lambda} u|^2\, dx \right) \\
&+ \sum_{\substack{\sigma^*\in\mathcal{E}^*_{ext}\\ \sigma^*=\mathcal{E}^*_{K^*}\cap\mathcal{E}^*_{ext}}} \left(\int_{T_{K^*,\sigma^*}\in\mathcal{M}^{**}} |\nabla_{\mathcal{D},\Lambda} u|^2\, dx \right) \\
&\geq \sum_{\substack{\sigma^*\in\mathcal{E}^*_{\rm int}\\ \sigma\cap\sigma^*\neq\emptyset,\sigma=\mathcal{E}_K\cap\mathcal{E}_L\in\mathcal{E}_{int}}} \left(\frac{d_{\sigma^*}\,|u_K-u_L|^2}{|\sigma^*|} \right) + \sum_{\substack{\sigma^*\in\mathcal{E}^*_{\rm int}\\ \sigma\cap\sigma^*\neq\emptyset,\,\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}}} \left(\frac{d_{\sigma^*}\,|u_K|^2}{|\sigma^*|} \right) \\
&+ \sum_{\substack{\sigma^*\in\mathcal{E}^*_{ext}\\ \sigma\cap\sigma^*\neq\emptyset,\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}}} \left(\frac{d_{\sigma^*}\,|u_K|^2}{|\sigma^*|} \right) \\
&\geq \frac{1}{\theta^2} \left(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma\in\mathcal{E}_K\cap\mathcal{E}_L}} |\sigma| \frac{|u_K-u_L|^2}{d_\sigma} + \sum_{\sigma\in\mathcal{E}_{ext}} |\sigma| \frac{|u_K|^2}{d_\sigma} \right) = \theta^2 ||u||^2_{\mathcal{D}}.
\end{aligned}$$

Therefore, the left hand side of (6.10) is estimated by

$$\int_\Omega (\Lambda(x)\nabla_{\mathcal{D},\Lambda} u(x)) \,.\nabla_{\mathcal{D},\Lambda} u(x) dx \geq \underline{\lambda} \int_\Omega |\nabla_{\mathcal{D},\Lambda} u(x)|^2\, dx \geq \underbrace{\underline{\lambda}\frac{1}{\theta^2}}_{=C} ||u||_{\mathcal{D}}.$$

Using the symmetry and positivity properties of the family $\beta^{\mathcal{D}}$ and (6.2), it implies that

$$\begin{aligned}
\sum_{K\in\mathcal{M}} u_K \sum_{Z\in V(K)} \beta^{\mathcal{D}}_{K,L}(u)(u_K-u_Z) &= \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{M}} \beta^{\mathcal{D}}_{K,Z}(u)(u_K-u_Z)^2 \\
&+ \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{E}_{ext}} \beta^{\mathcal{D}}_{K,Z}(u)(u_K)^2 \geq 0.
\end{aligned}$$

Together (6.10), we get $\sum_{K\in\mathcal{M}} \mathcal{S}^{\mathcal{D}}_K(u) u_K \geq C_1 ||u||^2_{\mathcal{D}}$. □

6.4.2 A prior estimate.

Proposition 6.2 *If $\theta \geq regul(\mathcal{D}), regul(\mathcal{D}^*)$, there then exists C_2 only depending on Ω and θ such that for any solution u to the NLFECC2 scheme $\mathcal{S}^{\mathcal{D}}$:*

$$||u||_{\mathcal{D}} \leq C_2 ||f||_{L^2(\Omega)}. \tag{6.13}$$

Proof of proposition 6.2

From proposition 6.1, we get that

$$C_1||u||_{\mathcal{D}}^2 \leq \sum_{K\in\mathcal{M}} \mathcal{S}_K^{\mathcal{D}}(u)u_K = \sum_{K\in\mathcal{M}} |K|f_K u_K$$
$$= \int_\Omega f.u\ dx \underset{Holder}{\leq} ||f||_{L^2(\Omega)}||u||_{L^2(\Omega)} \leq C_3||f||_{L^2(\Omega)}||u||_{\mathcal{D}}.$$

The inequation $||u||_{L^2(\Omega)} \leq C_3||u||_{\mathcal{D}}$ is implied by the discrete Poincare inequality. This inequality can be deduced from Lemma 5.3 of [23]. It states that there exists C_3 only depending on Ω and θ. Similarly, we also have the following priori estimate of the second FECC scheme: for any solution $u \in \mathcal{H}_{\mathcal{M}}$ of the second FECC scheme, there exists C_4 such that

$$||u||_{\mathcal{D}} \leq C_4||f||_{L^2(\Omega)}. \qquad \square$$

6.4.3 Existence of a solution.

Proposition 6.3 *There exists one solution to the NLFECC2 scheme (6.8) with the correction $\beta^{\mathcal{D}}$.*

Proof of proposition 6.3

Due to the proof of Proposition 3.4 [13](http://hal.archives-ouvertes.fr/hal-00643838/), we put the function

$$H : [0,1] \times H_{\mathcal{M}} \to H_{\mathcal{M}}$$
$$(t,u) \qquad \to H(t,u) = -\mathcal{A}^{\mathcal{D}}(u) + t\left(\sum_{Z\in V(K)} \beta_{K,Z}^{\mathcal{D}}(u_K - u_Z)\right)_{K\in\mathcal{M}} = (1-t)\left[-\mathcal{A}^{\mathcal{D}}(u)\right] + t\mathcal{S}^{\mathcal{D}}(u),$$

we have $H(0,u) = -\mathcal{A}^{\mathcal{D}}(u)$ and $H(1,u) = \mathcal{S}^{\mathcal{D}}(u)$.

The operator $H(t,.)$ is continuous on $\mathcal{H}_{\mathcal{M}}$ for all $t \in [0,1]$, because for all $K \in \mathcal{M}$ and all $Z \in V(K)$, we have the following inequation

$$\beta_{K,Z}^{\mathcal{D}}(u)|u_K - u_Z| \leq \begin{cases} \eta.\left(|\mathcal{A}_K^{\mathcal{D}}(u)| + |\mathcal{A}_Z^{\mathcal{D}}(u)|\right) & \text{if } Z \in V(K)\cap\mathcal{M}, \\ \eta.|\mathcal{A}_K^{\mathcal{D}}(u)| & \text{if } Z \in V(K)\cap\mathcal{E}_{ext}. \end{cases}$$

Because of results of the two propositions 6.1 and 6.2, the set of all solutions of $H(t,u) = (|K|f_K)_{K\in\mathcal{M}}$ is bounded in $\mathcal{H}_{\mathcal{M}}$.

Therefore, H is a homotopy, the degree of H(0,.) is different from 0. We then deduce existence of a solution for $H(1,.) = (|K|f_K)_{K\in\mathcal{M}}$ by applying the Brouwer's theorem. $\square$

6.4.4 Convergence.

In this section, we will show conditions to get the convergence of the NLFECC2 scheme.

Proposition 6.4 *If we have the following two conditions:*

- *There exists $\theta > 0$ not depending on $\mathcal{D}$ such that*

$$\theta > regul(\mathcal{D}) + regul(\mathcal{D}^*) + \sup_{K\in\mathcal{M}, L\in V(K)} \frac{diam(L)}{diam(K)}, \tag{6.14}$$

- $$\sup_{K\in\mathcal{M}} \left\{ |\mathcal{A}_K^{\mathcal{D}}(u)| \frac{diam(K)}{|K|} \right\} \to 0 \quad as\ size(D) \to 0, \tag{6.15}$$

then for each $\varphi \in \mathcal{C}_c^\infty(\Omega)$, $\varphi_{\mathcal{D}} = (\varphi_K)_{K\in\mathcal{M}} \in \mathcal{H}_{\mathcal{M}}$ with $\varphi_K = \varphi(x_K)$, we obtain

$$\sum_{K\in\mathcal{M}} \varphi_K \sum_{Z\in V(K)} \beta_{K,Z}^{\mathcal{D}}(u_K - u_Z) \to 0 \quad as\ size(\mathcal{D}) \to 0,$$

where $u \in \mathcal{H}_{\mathcal{M}}$ is a solution of the NLFECC2 scheme $\mathcal{S}^{\mathcal{D}}$.

Proposition 6.4 relates to the convergence of the NLFECC2 scheme, as follows:
When the equation (6.8) is multiplied by φ_K on K and summed over $K \in \mathcal{M}$, it is transformed into:

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

$\sum_{K\in\mathcal{M}} |K| f_K \varphi_K$ tends to the integral $\int_\Omega f\varphi$. Moreover, by the existence of the consistency of the second FECC scheme, it ensures that

$$-\sum_{K\in\mathcal{M}} \mathcal{A}_K^{\mathcal{D}}(u)\varphi_K \to \int_\Omega \Lambda \nabla u \nabla \varphi \, dx \quad \text{as } size(\mathcal{D}) \to 0.$$

Proof of proposition 6.4

This proposition is the same proof as Proposition 3.6 [13]. From the symmetric family $\beta^{\mathcal{D}}$ and (6.2), we can write:

$$\begin{aligned}
\sum_{K\in\mathcal{M}} \varphi_K \sum_{Z\in V(K)} \beta_{K,Z}^{\mathcal{D}}(u_K - u_Z) &= \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{M}} \beta_{K,Z}^{\mathcal{D}}(u)(u_K - u_Z)(\varphi_K - \varphi_Z) \\
&+ \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{E}_{ext}} \beta_{K,Z}^{\mathcal{D}}(u)(u_K)(\varphi_K) \\
&\leq \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{M}} \beta_{K,Z}^{\mathcal{D}}(u) \left|u_K - u_Z\right| \left|\varphi_K - \varphi_Z\right| \\
&+ \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)\cap\mathcal{E}_{ext}} \beta_{K,Z}^{\mathcal{D}}(u) \left|u_K\right| \left|\varphi_K\right|.
\end{aligned}$$

Together $\frac{|u_K - u_Z|}{\sum_{Y\in V(K)} |u_K - u_Y|} \leq 1$ for all $K \in \mathcal{M}$, all $Z \in V(K)$, we then get

$$\sum_{K\in\mathcal{M}} \varphi_K \sum_{Z\in V(K)} \beta_{K,Z}^{\mathcal{D}}(u_K - u_Z) \leq \sum_{K\in\mathcal{M}} \left[\begin{array}{l} \sum\limits_{Z\in V(K)\cap\mathcal{M}} \left(\left|\mathcal{A}_K^{\mathcal{D}}(u)\right| \left|\varphi_K - \phi_Z\right| + \left|\mathcal{A}_Z^{\mathcal{D}}(u)\right| \left|\varphi_K - \varphi_Z\right| \right) \\ + \sum\limits_{\sigma\in V(K)\cap\mathcal{E}_{ext}} \left(\left|\mathcal{A}_K^{\mathcal{D}}(u)\right| \left|\varphi_K\right| \right) \end{array} \right].$$

Note that for each $K \in \mathcal{M}$, each $Z \in V(K)$, there exists $K^* \in \mathcal{M}^*$ satisfying $x_K, x_Z \in \mathcal{V}^*_{K^*}$. Thanks to (6.14), we find a positive constant C_5 not depending on $\mathcal{D}$ such that

$$|\varphi_K - \varphi_Z| \leq C_5 \text{diam}(K) \quad \text{for all } K \in \mathcal{M} \text{ and all } Z \in V(K)$$

with the function $\varphi \in \mathcal{C}_0^\infty(\Omega)$.
Therefore,

$$\sum_{K\in\mathcal{M}} \varphi_K \sum_{Z\in V(K)} \beta^{\mathcal{D}}_{K,Z}(u_K - u_Z) \leq C_7 \sum_{K\in\mathcal{M}} \text{diam}(K)|\mathcal{A}^{\mathcal{D}}_K(u)| \leq C_7|\Omega| \sup_{K\in\mathcal{M}} \left\{ |\mathcal{A}^{\mathcal{D}}_K(u)| \frac{\text{diam}(K)}{|K|} \right\}.$$

By the condition (6.15), we deduce that

$$\sum_{K\in\mathcal{M}} \varphi_K \sum_{Z\in V(K)\cap\mathcal{M}} \beta^{\mathcal{D}}_{K,Z}(u_K - u_Z) \to 0 \quad \text{size}(\mathcal{D}) \to 0. \qquad \square$$

6.4.5 Discrete maximum principle.

Proposition 6.5 *The NLFECC2 scheme with the first correction $\beta^{\mathcal{D}}$ satisfies the discrete maximum principle (see Definition 6.1).*

Proof of proposition 6.5

Proposition 6.5 is also proved as Proposition 2.1 [13]. We put $u_{K_0} = \min_{K\in\mathcal{M}} u_K$ with $K^0 \in \mathcal{M}$. From (6.8), we consider $\mathcal{S}^{\mathcal{D}}_{K_0}(u)$:

$$\begin{aligned}
\mathcal{S}^{\mathcal{D}}_{K_0}(u) &= \sum_{Z\in V(K_0)} \mathcal{A}^{\mathcal{D}}_{K_0}(u) \frac{|u_{K_0} - u_Z|}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} + \sum_{\sigma\in V(K_0)\cap\mathcal{E}_{ext}} \left(\frac{\left|\mathcal{A}^{\mathcal{D}}_{K_0}(u)\right|}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} \right) .u_{K_0} \\
&+ \sum_{Z\in V(K_0)\cap\mathcal{M}} \left(\frac{\left|\mathcal{A}^{\mathcal{D}}_{K_0}(u)\right|}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} + \frac{\left|\mathcal{A}^{\mathcal{D}}_{Z}(u)\right|}{\sum_{Y\in V(K_0)} |u_{Z} - u_Y|} \right) .(u_{K_0} - u_Z) \\
&= |K_0| f_{K_0} \geq 0.
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \mathcal{S}^{\mathcal{D}}_{K_0}(u) &= \sum_{Z\in V(K_0)\cap\mathcal{M}} \underbrace{\left(\begin{array}{l} \frac{\mathcal{A}^{\mathcal{D}}_{K_0}(u).\text{sgn}(u_{K_0} - u_Z)}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} + \frac{\left|\mathcal{A}^{\mathcal{D}}_{K_0}(u)\right|}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} \\ \\ + \frac{\left|\mathcal{A}^{\mathcal{D}}_{Z}(u)\right|}{\sum_{Y\in V(K_0)} |u_{Z} - u_Y|} \end{array} \right)}_{\geq 0} .(u_{K_0} - u_Z) \\
&+ \sum_{\sigma\in V(K_0)\cap\mathcal{E}_{ext}} \underbrace{\left(\frac{\mathcal{A}^{\mathcal{D}}_{K_0}(u).\text{sgn}(u_{K_0})}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} + \frac{\left|\mathcal{A}^{\mathcal{D}}_{K_0}(u)\right|}{\sum_{Y\in V(K_0)} |u_{K_0} - u_Y|} \right)}_{\geq 0} .u_{K_0} = |K_0|\, f_{K_0} \geq 0.
\end{aligned}$$

Obviously, u_{K_0} must be equal or great than 0. $\square$

Part V

Numerical tests in 2D and 3D

Chapter 7

Numerical results

Contents

7.1 Motivation.

In this chapter, the efficiency of the 2D & 3D FECC schemes is demonstrated through numerical tests of the 5th and 6th International Symposium on Finite Volumes for Complex Applications - FVCA 5 & 6. Moreover, the comparison with classical finite volume schemes emphasizes the precision of the method. We also show the good behaviour of the algorithm for nonconforming meshes. Besides, in the section (7.3), there are examples for the maximum principle violations of the FECC schemes and 3D numerical results with non-linear corrections.

7.2 2D numerical tests.

7.2.1 Notations in 2D numerical tests.

We introduce some notations for all the tests in the section 7.2:

- nunkw: number of unknowns.
- umin: value of the minimum of the approximate solution.
- umax: value of the maximum of the approximate solution.

Let us denote by u_{ana} the exact solution, $u_{\mathcal{M}} = (u_K)_{K\in\mathcal{M}}$ the piecewise constant approximate solution.

- erl2, the relative discrete L^2 norm of the error, as follows:

$$\text{erl2} = \left(\frac{\sum_{K\in\mathcal{M}} |K|(u_{ana}(x_K) - u_K)^2}{\sum_{K\in\mathcal{M}} |K| u_{ana}(x_K)^2} \right)^{\frac{1}{2}} .$$

- ergrad, the relative L^2 norm of the error on the gradient.
- ratiol2: for $i \geq 2$,

$$\text{ratiol2(i)} = -2 \frac{\ln(\text{erl2(i)}) - \ln(\text{erl2(i - 1)})}{\ln(\text{nunkw(i)}) - \ln(\text{nunkw(i - 1)})} .$$

- ratiograd, for $i \geq 2$, the same formula as above with ergrad instead of erl2.

7.2.2 2D numerical results.

We use numerical tests in the 2D benchmark on discretization schemes FVCA 5.

Test 2D.1*: *Mild anisotropy

We consider a homogeneous anisotropic tensor, as follows:

$$\Lambda = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}.$$

Test 2D.1.1: The exact solution u_{ana} and the source term f

$$\begin{cases} u_{\text{ana}}(x,y) = 16x(1-x)y(1-y) \quad \text{in} \quad (0,1)\times(0,1)\,, \\ u_{\text{ana}}(x,\text{y}) = 0 \quad \text{on the boundary of } [0,1]\times[0,1], \\ f(x,y) = -\nabla.\,(\Lambda\nabla u_{\text{ana}})\,. \end{cases}$$

nunkw	erl2	ratiol2	umin	umax	ergrad	ratiograd
56	9.74303E-03		9.12E-02	9.28E-01	1.46E-02	
224	2.44889E-03	1.99E+00	2.54E-02	9.28E-01	8.15E-03	0.848
896	6.08651E-04	2.00E+00	6.70E-03	9.95E-01	4.26E-03	0.936
3584	1.52175E-04	1.99E+00	1.73E-03	9.99E-01	2.17E-03	0.967
14336	3.81026E-05	1.99E+00	4.36E-04	1.00E+00	1.10E-03	0.983

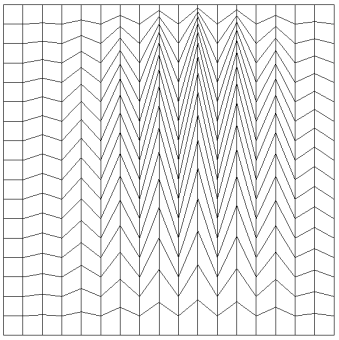
Mesh 1 - regular triangular mesh.

nunkw	erl2	umin	umax	ergrad
289	2.68581E-03	1.26800E-02	1.0020E+00	2.81E-02

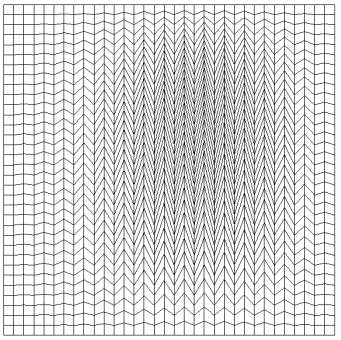
Mesh 4.1 - distorted quadrangular mesh.

nunkw	erl2	umin	umax	ergrad
1089	7.60982E-04	3.48999E-03	1.0007E+00	1.29E-02

Mesh 4.2 - distorted quadrangular mesh.



Mesh 4.1 - distorted quadrangular mesh.



Mesh 4.2 - distorted quadrangular mesh.

Test 2D.1.2: The exact solution and the source term f

$$\begin{cases} u_{\text{ana}}(x,y) = sin((1-x)(1-y)) + (1-x)^3(1-y)^2, \\ f(x,y) = -\nabla.(\Lambda \nabla u_{\text{ana}}). \end{cases}$$

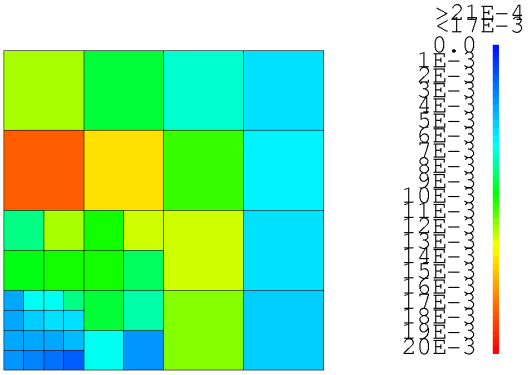
nunkw	erl2	ratiol2	umin	umax	ergrad	ratiograd
56	2.25334E-03		7.16810E-03	1.3786	1.56E-03	
224	6.03417E-04	1.90E+00	1.77495E-03	1.5973	9.52E-04	0.718
896	1.54969E-04	1.96E+00	4.42261E-04	1.7160	5.44E-04	0.808
3584	3.91813E-05	1.98E+00	1.10442E-04	1.7779	2.93E-04	0.890
14336	9.84396E-06	1.99E+00	2.75983E-05	1.8095	1.53E-04	0.938

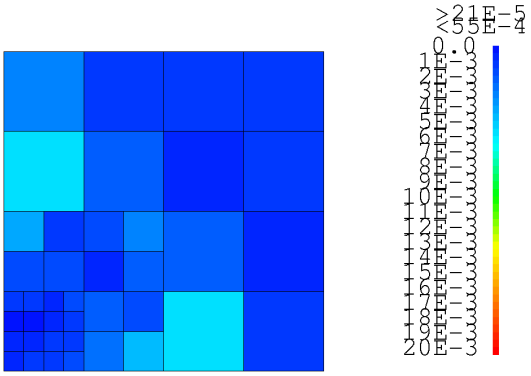
Mesh 1 - regular triangular mesh.

nunkw	erl2	ratiol2	ergrad	ratiograd
40	5.41026E-03		2.43e-02	
160	1.29132E-03	2.06E+00	1.35E-02	0.848
640	3.06998E-04	2.07E+00	7.12E-03	0.926
2560	7.43874E-05	2.04E+00	3.65E-03	0.964
10240	1.82906E-05	2.02E+00	1.84E-03	0.982

Mesh 3 - locally refined nonconforming rectangular mesh.

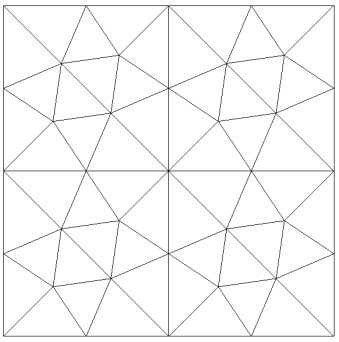
The error between the exact solution and the computed solution



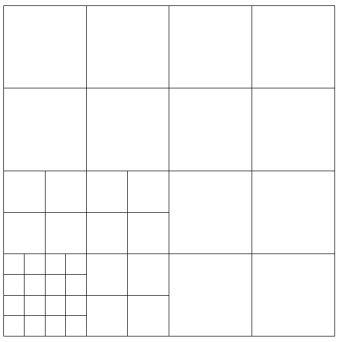


Left: Result of the MPFA scheme [2].

Right: Result of the FECC scheme.



Mesh 1 - regular triangular mesh.



Mesh 3 - locally refined nonconforming rectangular mesh.

Test 2D.2: Heterogeneous rotating anisotropy

The tensor Λ satisfies the equation:

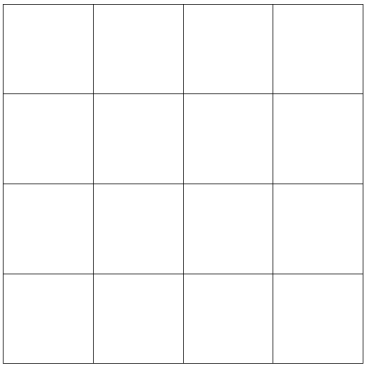
$$\Lambda = \frac{1}{(x^2+y^2)} \begin{pmatrix} 10^{-3}x^2+y^2 & \left(10^{-3}-1\right)xy \\ \left(10^{-3}-1\right)xy & x^2+10^{-3}y^2 \end{pmatrix}.$$

We define the exact solution and the source term f as:

$$\begin{cases} \mathrm{u_{ana}(x,y)} = \sin(\pi \mathrm{x}) \sin(\pi \mathrm{y}) \quad \text{in} \quad (0,1)\times(0,1)\,, \\ \mathrm{u_{ana}(x,y)} = 0 \text{ on the boundary of } [0,1]\times[0,1], \\ f(x,y) = -\nabla.\left(\Lambda \nabla u_{\mathrm{ana}}\right). \end{cases}$$

nunkw	erl2	ratiol2	umin	umax	ergrad	ratiograd
16	7.02265E-02		1.35E-01	9.34E-01	9.04E-02	
64	1.67141E-02	2.07E+00	3.68E-02	9.8E-01	5.03E-02	0.770
256	4.25124E-03	1.97E+00	9.5E-03	9.94E-01	2.80E-02	0.919
1024	1.09645E-03	1.95E+00	2.4E-03	9.98E-01	1.43E-02	0.970
4096	2.81843E-04	1.96E+00	6.02E-04	9.99E-01	7.21E-03	0.988

Mesh 2 - uniform rectangular mesh.



Mesh 2 - uniform rectangular mesh.

We consider the problem:

$$\begin{cases} div\,(\Lambda\nabla u) = div\,(\Lambda\nabla u_{\text{ana}}) & \text{in } \Omega = (0,1)\times(0,1)\,, \\ u(x,y) = u_{\text{ana}}(x,y) & \text{on } \partial\Omega, \end{cases}$$

for the following tests:

Test 2D.3: ***Discontinuous anisotropy*** (see for more detail in the section 4.2.1 of [21])

The analytical solution is

$$\begin{cases} u_{\text{ana}}(x,y) = \cos(\pi x)\sin(\pi y) & \text{if } x \le 0.5, \\ u_{\text{ana}}(x,y) = 0.01\cos(\pi x)\sin(\pi y) & \text{if } x > 0.5. \end{cases}$$

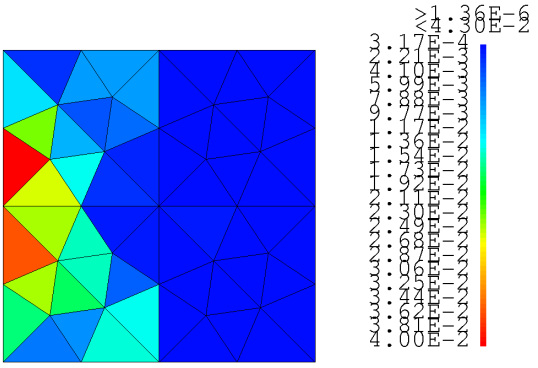
Consider the tensor

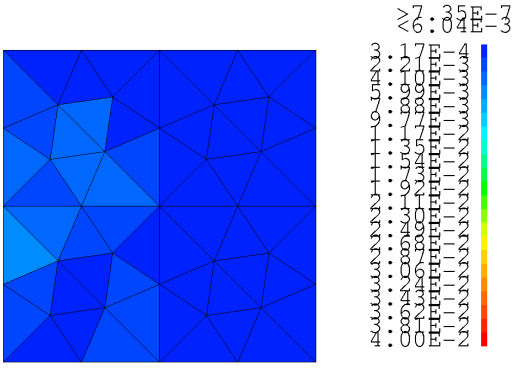
$$\Lambda(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } x \le 0.5, \quad \Lambda(x,y) = \begin{pmatrix} 100 & 0 \\ 0 & 0.01 \end{pmatrix} \text{ if } x > 0.5.$$

nunkw	erl2	ratiol2	umin	umax	ergrad	ratiograd
56	5.45056E-03		-9.07E-03	9.06E-01	8.131501E-03	
224	1.37517E-03	1.98E+00	-9.76E-03	9.75E-01	6.623965E-03	0.296
896	3.44881E-04	1.99E+00	-9.93E-03	9.93E-01	3.741116E-03	0.824
3584	8.65861E-05	1.99E+00	-9.98E-03	9.98E-01	1.972594E-03	0.923
14336	2.17672E-05	1.99E+00	-9.99E-03	9.99E-01	1.011825E-03	0.963

Mesh 1 - regular triangular mesh.

The error between the exact solution and the computed solution





Left: Result of the Diamond scheme [16]. Right: Result of the FECC scheme.

Test 2D.3.b*: *Discontinuous anisotropy

The analytical solution is

$$\begin{cases} u_{\text{ana}}(x,y) = \cos(\pi x)\sin(\pi y) & \text{if } x \leq 0.5, \\ u_{\text{ana}}(x,y) = 0.01\cos(\pi x)\sin(\pi y) & \text{if } x > 0.5. \end{cases}$$

The tensor is

$$\Lambda(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } x \leq 0.5, \quad \Lambda(x,y) = \begin{pmatrix} 100 & 0 \\ 0 & 0.01 \end{pmatrix} \text{ if } x > 0.5.$$

We use the harmonic averaging points y_σ introduced by [5] to define the dual grid.

$$y_\sigma = \frac{\lambda_L d_{K,\sigma} y_L + \lambda_K d_{L,\sigma} y_K}{\lambda_L d_{K,\sigma} + \lambda_K d_{L,\sigma}} + \frac{d_{K,\sigma} d_{L,\sigma}}{\lambda_L d_{K,\sigma} + \lambda_K d_{L,\sigma}} (\lambda_K^\sigma - \lambda_L^\sigma),$$

where notations are defined in Lemma 2.1 of [5], page 2.

We obtain the following numerical results with this modification:

nunkw	erl2	ratiol2	umin	umax
56	4.99875E-03		-9.07E-03	9.06E-01
224	1.28975E-03	1.95E+00	-9.76E-03	9.75E-01
896	3.32247E-04	1.95E+00	-9.93E-03	9.93E-01
3584	8.47105E-05	1.97E+00	-9.98E-03	9.98E-01
14336	2.14672E-05	1.98E+00	-9.99E-03	9.99E-01

The results are slightly more accurate than the previous results.

Test 2D.4*: *Discontinuous strong anisotropy

The analytical solution is

$$\begin{cases} u_{\text{ana}}(x,y) = \cos(\pi x)\sin(\pi y) & \text{if } x \leq 0.5, \\ u_{\text{ana}}(x,y) = 10^{-6}\cos(\pi x)\sin(\pi y) & \text{if } x > 0.5. \end{cases}$$

Consider the tensor

$$\Lambda(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } x \leq 0.5, \quad \Lambda(x,y) = \begin{pmatrix} 10^6 & 0 \\ 0 & 0.01 \end{pmatrix} \text{ if } x > 0.5.$$

nunkw	erl2	ratiol2	umin	umax	ergrad	ratiograd
56	5.45798E-03		-9.07E-07	9.06E-01	8.138566E-03	
224	1.37250E-03	1.99E+00	-9.76E-07	9.75E-01	6.624536E-03	0.296
896	3.43047E-04	2.00E+00	-9.93E-07	9.93E-01	3.741224E-03	0.824
3584	8.58622E-05	1.99E+00	-9.98E-07	9.98E-01	1.972620E-03	0.923
14336	2.14862E-05	1.99E+00	-9.99E-07	9.99E-01	1.011832E-03	0.963

Mesh 1 - regular triangular mesh.

Test 2D.4.b: Discontinuous strong anisotropy

We also use the harmonic averaging points y_σ introduced by [5] to define the dual grid. The analytical solution is

$$\begin{cases} u_{\text{ana}}(x,y) = \cos(\pi x)\sin(\pi y) & \text{if } x \leq 0.5, \\ u_{\text{ana}}(x,y) = 10^{-6}\cos(\pi x)\sin(\pi y) & \text{if } x > 0.5. \end{cases}$$

The tensor is

$$\Lambda(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } x \leq 0.5, \quad \Lambda(x,y) = \begin{pmatrix} 10^6 & 0 \\ 0 & 0.01 \end{pmatrix} \text{ if } x > 0.5.$$

We obtain the following numerical results:

nunkw	erl2	ratiol2	umin	umax
56	5.03257E-03		-9.07E-07	9.06E-01
224	1.29392E-03	1.95E+00	-9.76E-07	9.75E-01
896	3.32255E-04	1.96E+00	-9.93E-07	9.93E-01
3584	8.44700E-05	1.97E+00	-9.98E-07	9.98E-01
14336	2.13100E-05	1.98E+00	-9.99E-07	9.99E-01

The results are slightly more accurate than before.

Remark 7.1: In test $2D.3.b$ and test $2D.4.b$, the tensors are discontinuous on the line (d): $x = 0.5$. All the points y_σ belong to the edges σ which are common edges of the two adjacent control volumes, computed by

$$y_\sigma = \frac{\lambda_L d_{K,\sigma} y_L + \lambda_K d_{L,\sigma} y_K}{\lambda_L d_{K,\sigma} + \lambda_K d_{L,\sigma}},$$

because all the vectors λ_K^σ, λ_L^σ are equal to 0.

In test 5, we show that [5] does not provide an "acceptable" y_σ.

Test 2D.5: Discontinuous anisotropy (case where the harmonic averaging points are not always defined).

The analytical solution is

$$u_{\text{ana}}(x,y) = \sin(\pi x).$$

The tensor is

$$\Lambda(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } x \leq 0.5, \quad \Lambda(x,y) = \begin{pmatrix} 1 & 9 \\ 9 & 100 \end{pmatrix} \text{ if } x > 0.5.$$

We obtain the following numerical results with the FECC scheme:

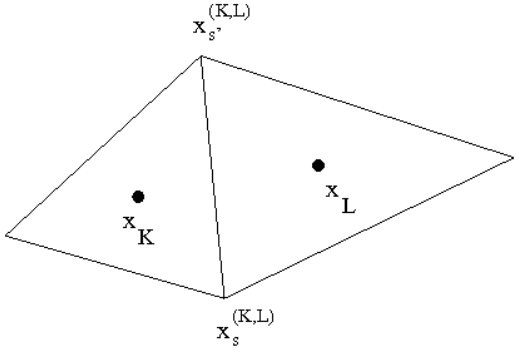
nunkw	erl2	ratiol2	umin	umax
56	2.81812E-01		1.51E-01	1.80E+00
224	2.59236E-02	3.44E+00	7.81E-02	1.10E+00
896	3.56133E-03	2.86E+00	3.92E-02	1.02E+00
3584	5.81922E-04	2.61E+00	1.96E-02	1.00E+00
14336	1.43293E-04	2.02E+00	9.81E-03	1.00E+00

Here, the FECC scheme is defined because the dual mesh is defined. Initially, the primary mesh points were located at the barycentre of each triangle cell. We chose to move them slightly such that the hypothesis 3.1 is satisfied for any edge of the primary grid.

We give here the coordinates of a few y_σ for the coarse grid.

nunkw	K	L	$x_s^{(K,L)}$	$x_{s'}^{(K,L)}$	y_σ
56	6	18	$(0.5, 0)$	$(0.5, 0.25)$	$(0.5, -0.105128205)$
	14	24	$(0.5, 0.25)$	$(0.5, 0.5)$	$(0.5, 0.120512821)$

The scheme of [5] is not defined, because there are some harmonic averaging points y_σ which are outside the edges σ.

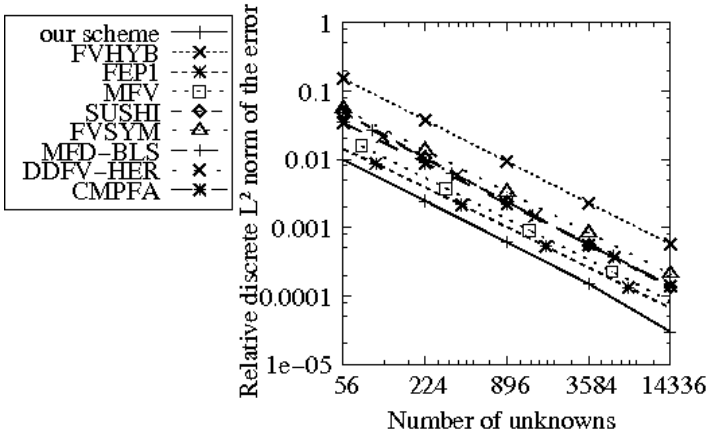


In this figure, the edge σ which is a common edge between K and L, has two vertices $x_s^{(K,L)}$ and $x_{s'}^{(K,L)}$.

By test 5, we see another difference between the FECC scheme and the scheme of [5].

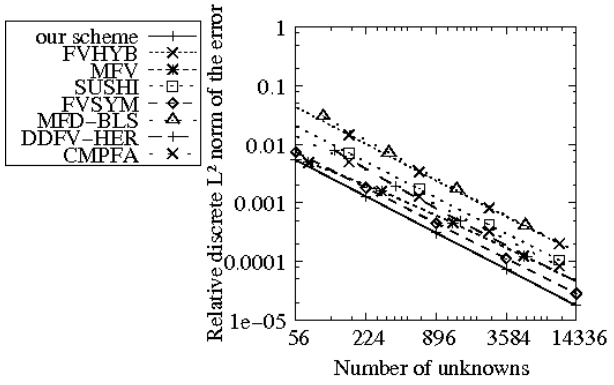
7.2.3 Comments about 2D numerical results.

In test 2D.1.1, for triangular cells, when the mesh and the solution are regular, we obtain a second order convergence in the L^2 norm and an order close to 1 for the gradient.

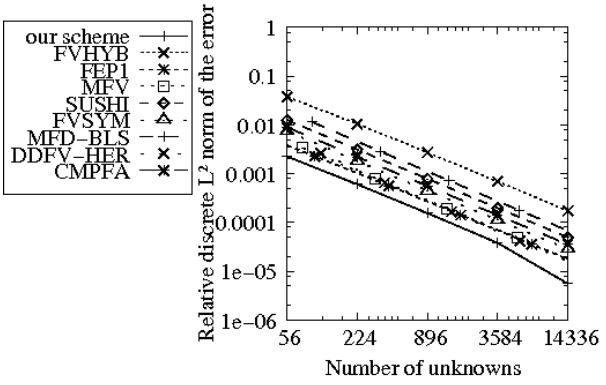


Mesh 1 – regular triangular mesh

In test 2D.1.2, we obtain an order of convergence in the L^2 norm close to 2 for regular triangular meshes and locally refined nonconforming rectangular meshes. The order of convergence of the gradient tends toward 1 for regular triangular meshes and locally refined nonconforming rectangular meshes.

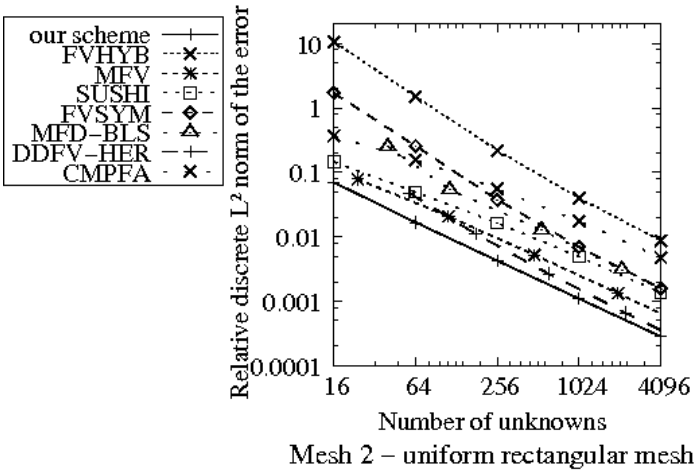


Mesh 3 – Locally refined nonconforming rectangular mesh



Mesh 1 – regular triangular mesh

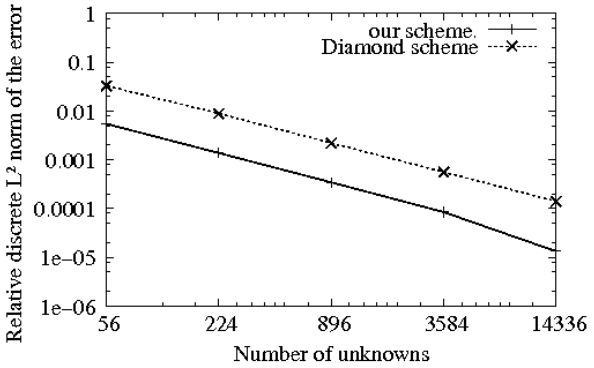
In test 2D.2, for uniform rectangular meshes where Λ is an heterogeneous tensor, we obtain an order of convergence near to 2 in the L^2 norm and the order of convergence of the gradient tends toward 1.



Mesh 2 – uniform rectangular mesh

From the graphs which describe the number of unknowns and relative discrete L^2 norm of the error, we see that the errors of the FECC scheme are less important than the errors of [6], [9], [23], [44], [48], [32], [40], [33] and the orders for the gradient of the scheme are close to those of the Galerkin finite element method (see [9]).

In test 2D.3, for regular triangular meshes where Λ is discontinuous, we obtain an order of convergence near to 2 in the L^2 norm and the order of convergence for the gradient tends toward 1. With the same number of unknowns, the errors of the scheme in the L^2 norm are less than the errors of [16].



Mesh 1 – regular triangular mesh

e) In test 2D.4, for regular triangular meshes where Λ is an heterogeneous tensor with a strong anisotropy, we obtain an order of convergence in the L^2 norm near to 2 and the order of convergence for the gradient also tends toward 1.

The participating schemes are used to compare with the FECC scheme:

Cell centered schemes:

- CMPFA: *Compact-stencil MPFA method for heterogeneous highly anisotropic second-order elliptic problems*, [44].

- FVHYB: *A symmetric finite volume scheme for anisotropic heterogeneous second-order elliptic problems*, [6].
- FVSYM: *Numerical results with two cell-centered finite volume schemes for heterogeneous anisotropic diffusion operators*, [48, 47].
- SUSHI: *A scheme using stabilization and Hybrid Interfaces for anisotropic heterogeneous diffusion problems*, [23].

Discrete duality finite volume schemes:

- DDFV-HER: *Numerical experiments with the DDFV method*, [32].

Finite elements schemes:

- FEP1: *A Galerkin finite element solution*, [9].

Mixed or hybrid methods:

- MFD-BLS: *Mimetic finite difference method*, [40].
- MFV: *Use of mixed finite volume method*, [33].

7.3 3D numerical tests.

We use numerical tests in the 3D benchmark on discretization schemes FVCA 6.

7.3.1 Notations in 3D numerical tests.

We use the following notations in all tests of the section 7.3:

- The relative L^2 norm of the gradient of the error is given by:

$$ergrad = \left(\frac{\sum_{T \in \mathcal{M}^{**}} m(T) |\nabla_T u - \nabla_T P(u_{\rm ana})|^2}{\sum_{T \in \mathcal{M}^{**}} m(T) |\nabla_T P(u_{\rm ana})|^2} \right)^{\frac{1}{2}},$$

  where $\mathcal{M}^{**}$ is the set of all tetrahedral elements of the third meshes, $m(T) > 0$ is the measure of the tetrahedron T.

- The convergence rates are defined by: for $i \geq 2$,

$$\begin{aligned} ratiol2(i) &= -3 \frac{\log(erl2(i)/erl2(i-1))}{\log(nu(i)/nu(i-1))}, \\ ratiograd(i) &= -3 \frac{\log(ergrad(i)/ergrad(i-1))}{\log(nu(i)/nu(i-1))}. \end{aligned}$$

7.3.2 3D numerical results.

Test 3D. 1: Flow on random meshes.

The anisotropy tensor is

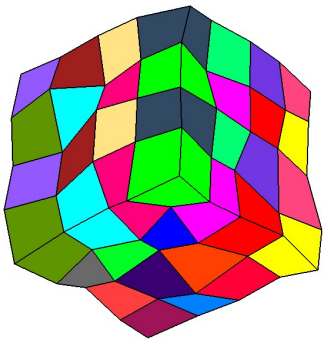
$$\Lambda(x,y,z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10^3 \end{pmatrix}$$

The analytical solution is

$$u_{\text{ana}}(x,y,z) = \sin(2\pi x)\sin(2\pi y)\sin(2\pi z)$$

with min = 0, max = 1.

We consider on the random meshes



random mesh.

We get the following numerical results:

nunkw	umin	umax	erl2	ratiol2	ergrad	ratiograd
64	-0.928	0.740	1.957E-01		2.935E-01	
512	-1.037	1.013	1.004E-01	0.963	2.578E-01	0.186
4096	-1.015	1.006	3.403E-02	1.561	1.501E-01	0.771

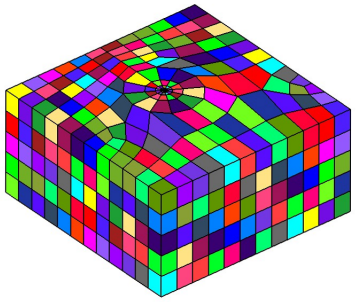
Test 3D. 2: Flow around a well.

The tensor is given by

$$\Lambda(x,y,z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}$$

with min = 0, max = 5.415.

The domain Ω and the exact solution are detailed in [4]. Moreover, the domain Ω is partitioned by the well meshes:



well mesh.

We obtain the following numerical results:

nunkw	umin	umax	erl2	ratiol2
890	0.390	5.316	6.089E-03	
2232	0.231	5.327	2.624E-03	2.745
5016	0.151	5.328	1.173E-03	2.982
11220	0.116	5.330	6.004E-04	2.495

Test 3D. 3: Discontinuous anisotropy.

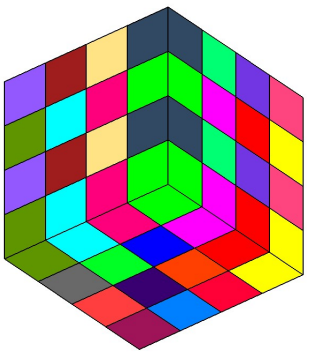
We consider the discontinuous anisotropic permeability, as follows:

$$\Lambda(x,y,z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{if } x \leq 0.5, \qquad \Lambda(x,y,z) = \begin{pmatrix} 1000 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{if } x > 0.5.$$

The analytical solution is

$$\begin{cases} u_{\text{ana}}(x,y,z) = \cos(\pi x)\sin(\pi y)\sin(\pi z) & \text{if } x \leq 0.5. \\ u_{\text{ana}}(x,y,z) = 0.001\cos(\pi x)\sin(\pi y)\sin(\pi z) & \text{if } x > 0.5. \end{cases}$$

The cube meshes are used:



cube mesh.

The numerical results of Test 3D. 3 are shown in the following three cases:

- **Case 1**: The source term is computed by $\int_\Omega f(x).P(p_K)(x)dx$ for all $K \in \mathcal{M}$ and $\int_\Omega f(x).P(p_{K^*})(x)dx$ for all $K^* \in \mathcal{M}^*$.
- **Case 2**: The source term is computed by $\int_K f(x)dx$ for all $K \in \mathcal{M}$.
- **Case 3**: We use harmonic averaging points and the edge unknowns u_σ in the equation $(37) - (40)$ [26].

	Case 1		Case 2		Case 3	
nunkw	erl2	ratiol2	erl2	ratiol2	erl2	ratiol2
64	4.614E-02		2.702E-01		6.413E-02	
512	1.163E-02	1.98	6.313E-02	2.09	1.579E-02	2.02
4096	2.914E-03	1.99	1.541E-02	2.03	3.883E-03	2.02
16000	1.180E-03	1.99	6.101E-03	2.04	1.558E-03	2.01

7.3.3 Comments about 3D numerical results.

We use the following classical schemes to compare with the 3D FECC scheme:

Cell-centered schemes:

- LS-FVM: *The cell-centered finite volume method using least squares vertex reconstruction (diamond scheme)*, by Y. Coudiére and G. Manzini [15].

Finite element schemes:

- MELODIE: *A linear finite element solver*, by H. Amor, M. Bourgeois, and G. Mathieu [8].

Gradient schemes:

- SUSHI: *The SUSHI scheme*, by R. Eymard, T. Gallouët and R. Herbin [24].
- VAG: *The VAG scheme*, by R. Eymard, C. Guichard and R. Herbin [25].

Nonlinear schemes:

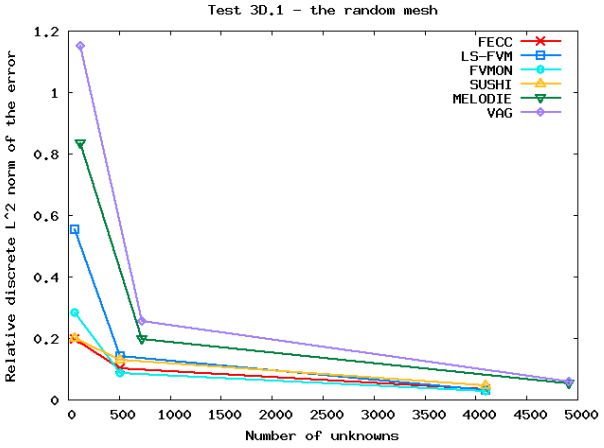
- FVMON: *A monotone nonlinear finite volume method for diffusion equations on polyhedral meshes*, by A. Danolov and Y. Vassilevski [17].

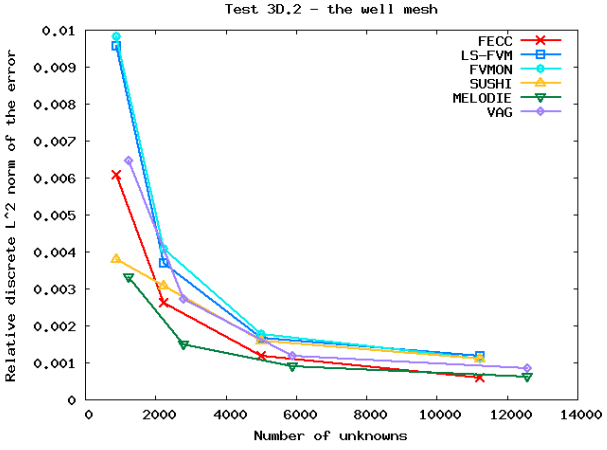
In test 3D.1, for the random meshes, the order of convergence in the L^2 norm and the order of convergence of the gradient of the FECC scheme increase, which is similar to the

numerical results of the 3D Sushi scheme in the same test.

In test 3D.2, for the well meshes, the order of convergence in the L^2 norm is greater than 2.

For test 3D.1 and test 3D.2, we obtain the following figures:





In test 3D.3, for the cube meshes and the discontinuous anisotropic tensor, we compare with the three different source terms. These methods are often used to compute the discrete source terms in classical finite volume and finite element schemes. We see that, in the three cases, the orders of convergence in the L^2 norm are close to 2.

7.4 3D numerical tests of the non-linear FECC schemes for Discrete Maximum Principle.

7.4.1 Notations in tests for Discrete Maximum Principle.

The following notations are used to present the numerical results:

- numkw: number of unknowns.
- $\text{umin}_{\mathcal{A}_i^{\mathcal{D}}}$: minimum value of the approximate solution,
- $\text{umax}_{\mathcal{A}_i^{\mathcal{D}}}$: maximum value of the approximate solution,
- $\text{erl2}_{\mathcal{A}_i^{\mathcal{D}}}$: the relative L^2 norm of the error,
- $\text{ergrad}_{\mathcal{A}_i^{\mathcal{D}}}$: the relative L^2 norm of the error,
- $\text{ratiol2}_{\mathcal{A}_i^{\mathcal{D}}}$: the convergence rate,

where the values $i = 1, 2$ correspond to the initial and the second FECC scheme.

We define, for the first non-linear FECC scheme (NLFECC1),

- $\text{umin}_{\mathcal{S}_j^{\mathcal{D}}}$: minimum value of the approximate solution,
- $\text{umax}_{\mathcal{S}_j^{\mathcal{D}}}$: maximum value of the approximate solution,
- $\text{erl2}_{\mathcal{S}_j^{\mathcal{D}}}$: the relative L^2 norm of the error,
- $\text{ratiol2}_{\mathcal{S}_j^{\mathcal{D}}}$: the convergence rate,
- $\text{nit}_{\mathcal{S}_j^{\mathcal{D}}}$: number of iterations needed to compute the approximate solution of $\mathcal{S}^{\mathcal{D}}$,
- $\left(\frac{\mathcal{A}_{K_0}}{|K_0|}\right)_{\mathcal{S}_j^{\mathcal{D}}} = \max\left\{\frac{|\mathcal{A}_K(u)|}{|K|}, \forall K \in \mathcal{M}\right\}$,

where the values $j = \overline{1,3}$ correspond to the values of $\eta = 0.25,\ 0.5,\ 1,\ 2$.

For the second non-linear FECC scheme (NLFECC2), we use the notations $\text{umin}_{\mathcal{S}_j^{\mathcal{D},\mathcal{D}^*}}$, $\text{umax}_{\mathcal{S}_j^{\mathcal{D},\mathcal{D}^*}}$, $\text{erl2}_{\mathcal{S}_j^{\mathcal{D},\mathcal{D}^*}}$, $\text{ratiol2}_{\mathcal{S}_j^{\mathcal{D},\mathcal{D}^*}}$ and $\text{nit}_{\mathcal{S}_j^{\mathcal{D},\mathcal{D}^*}}$.

The iterative algorithm of the NLFECC1 scheme is presented, as follows:

With a fixed point iteration i, we fix u in $\beta_{K,Z}(u)$ by the value of solution u^i.

Step 1: For all $K^* \in \mathcal{M}^*$, we construct the system of linear combinations depending on $\left((u_{K^*}^{i+1})_{K^* \in \mathcal{M}_{int}^*}, (u_K^{i+1})_{K \in \mathcal{M}}\right)$ with the following linear equation:

$$-\mathcal{A}_{K^*}^{\mathcal{D},\mathcal{D}^*}(u^{i+1}) + \sum_{Z \in V(K^*)} \beta_{K^*,Z}(u^i)(u_{K^*}^{i+1} - u_Z^{i+1}) = \int_\Omega f(x).P(p_{K^*})(x)dx.$$

We note that the set $V(K^*)$ does not contain $L^* \in \mathcal{M}^*$ such that $L^* \neq K^*$. Hence we compute $u_{K^*}^{i+1}$ by linear combinations depending on $(u_K^{i+1})_{K \in \mathcal{M}}$.

Step 2: For all $K \in \mathcal{M}$, we also have the system of linear combinations depending on $\big((u_{K^*}^{i+1})_{K^* \in \mathcal{M}_{int}^*}, (u_K^{i+1})_{K \in \mathcal{M}}\big)$ by the following linear equation:

$$-\mathcal{A}_K^{\mathcal{D},\mathcal{D}^*}(u^{i+1}) + \sum_{Z \in V(K)} \beta_{K,Z}(u^i)(u_K^{i+1} - u_Z^{i+1}) = \int_\Omega f(x).P(p_K)(x)dx.$$

By *Step* 1, we transform $u_{K^*}^{i+1}$ for all $K^* \in \mathcal{M}_{int}^*$ into the linear combinations depending on $(u_K^{i+1})_{K \in \mathcal{M}}$. Therefore, we can rewrite the system of linear equations in *Step* 2 by another system of linear equations only depending on $(u_K^{i+1})_{K \in \mathcal{M}}$.

In the iterative algorithm of the NLFECC2 scheme, we also fix $u = u^i$ in $\beta_{K,L}(u)$, where u^i is the value of the solution, i is a fixed point iteration. This iterative algorithm is written by:

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

The two algorithms are stopped by the criterion $\frac{||u^{i+1}-u^i||}{||u^i||} \leq 10^{-4}$.

7.4.2 3D numerical results of the non-linear FECC schemes.

Test 1: Stationary analytical solution

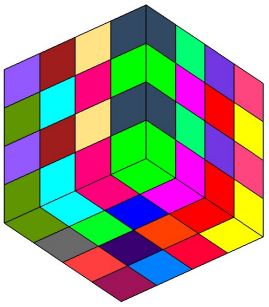
The strong anisotropic permeability is

$$\Lambda(x,y,z) = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The analytical solution is given by:

$$u(x,y,z) = \sin\left(\frac{\pi}{2}x\right).\sin\left(\frac{\pi}{2}y\right).\sin\left(\frac{\pi}{2}z\right).$$

The domain Ω is partitioned by the cube meshes:



the unit cube.

In test 1, we consider accuracy, the order of convergence in the L^2 norm and existence of the discrete maximum principle for the FECC schemes and the non-linear FECC schemes.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	8.023E-3	1.238E-3	1.748E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.25)$	9.334E-3	1.407E-3	1.917E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.5)$	1.108E-2	1.630E-3	2.129E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}(\eta = 1)$	1.490E-2	2.199E-3	2.666E-4
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.9622	0.9907	0.9976
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.9443	0.9844	0.9958
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.9330	0.9805	0.9946
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.9193	0.9757	0.9931
$\text{erl2}_{\mathcal{A}_1^{\mathcal{D}}}$	1.569E-2	4.274E-3	1.043E-3
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	9.403E-3	7.585E-3	4.216E-3
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	2.706E-2	1.651E-2	8.502E-3
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	5.476E-2	3.189E-2	1.612E-2
$\text{ratiol2}_{\mathcal{A}_1^{\mathcal{D}}}$		1.88	2.03
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$		0.31	0.84
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$		0.71	0.96
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$		0.78	0.98
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	6	6	5
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	8	7	7
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	11	12	15

Table 1.1, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the first correction** $\beta^{\mathcal{D},\mathcal{D}^*}$.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	8.023E-3	1.238E-3	1.748E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}(\eta=0.25)}$	8.917E-3	1.331E-3	1.822E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}(\eta=0.5)}$	1.001E-2	1.440E-3	1.910E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}(\eta=1)}$	1.248E-2	1.698E-3	2.139E-4
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.9622	0.9907	0.9976
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.9580	0.9902	0.9976
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.9541	0.9897	0.9975
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.9474	0.9888	0.9975
$\text{erl2}_{\mathcal{A}_1^{\mathcal{D}}}$	1.569E-2	4.274E-3	1.043E-3
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	8.432E-3	2.419E-3	6.383E-4
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	9.980E-3	3.085E-3	7.821E-4
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	2.486E-2	7.753E-3	1.880E-3
$\text{ratiol2}_{\mathcal{A}_1^{\mathcal{D}}}$		1.88	2.03
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$		1.80	1.92
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$		1.69	1.98
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$		1.68	2.04
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	5	4	4
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	6	5	5
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	7	8	8

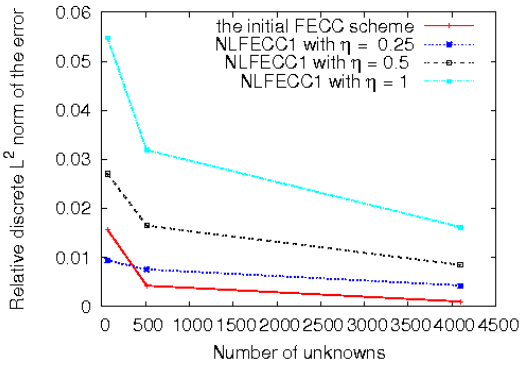
Table 1.2, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the second correction** $\overline{\beta}^{\mathcal{D},\mathcal{D}^*}$.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	8.023E-3	1.238E-3	1.748E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.25)$	9.365E-3	1.383E-3	1.865E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.5)$	1.105E-2	1.558E-3	2.001E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}(\eta = 1)$	1.460E-2	1.988E-3	2.441E-4
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.9622	0.9907	0.9976
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.9549	0.9898	0.9976
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.9488	0.9889	0.9975
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.9391	0.9873	0.9973
$\text{erl2}_{\mathcal{A}_1^{\mathcal{D}}}$	1.569E-2	4.274E-3	1.043E-3
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	7.243E-3	2.359E-3	6.479E-4
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	1.776E-2	6.010E-3	1.565E-3
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	4.065E-2	1.409E-2	3.697E-3
$\text{ratiol2}_{\mathcal{A}_1^{\mathcal{D}}}$		1.88	2.03
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$		1.62	1.86
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$		1.56	1.94
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$		1.53	1.93
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	5	5	5
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	7	6	7
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	9	11	13

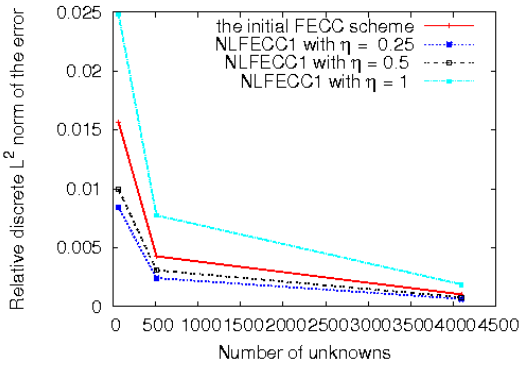
Table 1.3, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the third correction** $\widetilde{\beta}^{\mathcal{D},\mathcal{D}^*}$.

Comments about Table 1.1, 1.2 and 1.3: for the strong anisotropic tensor and the cube meshes, we obtain a second order in the L^2 norm for the initial FECC scheme. With the NLFECC1 schemes constructed from the first corrections, their orders of convergence in the L^2 norm are near to 1 with $\eta = 0.25, 0.5, 1$. On the other hand, when we consider the NLFECC1 schemes with the second and the third corrections, the orders of convergence in the L^2 norm are close to 2 with $\eta = 0.25, 0.5, 1$. All these schemes satisfy the discrete maximum principle.

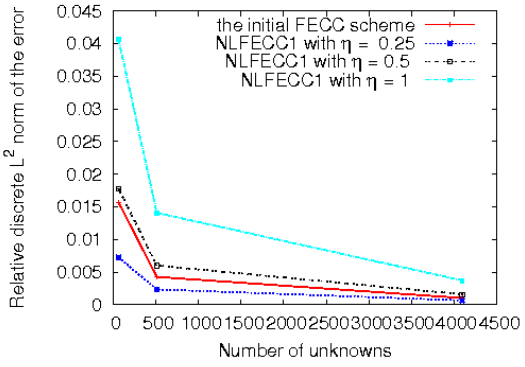
In the following figures, we show the comparison between the initial FECC scheme and the non-linear schemes:



the first correction.



the second correction.



the third correction.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_2^{\mathcal{D}}}$	8.501E-3	1.278E-3	1.777E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D}}}(\eta = 0.25)$	1.091E-2	1.545E-3	2.019E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D}}}(\eta = 0.5)$	1.364E-2	1.870E-3	2.313E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D}}}(\eta = 1)$	1.828E-2	2.548E-3	2.962E-4
$\text{umax}_{\mathcal{A}_2^{\mathcal{D}}}$	1.0172	1.0096	1.0033
$\text{umax}_{\mathcal{S}_1^{\mathcal{D}}}$	0.9850	0.9981	0.9998
$\text{umax}_{\mathcal{S}_2^{\mathcal{D}}}$	0.9653	0.9913	0.9977
$\text{umax}_{\mathcal{S}_3^{\mathcal{D}}}$	0.9432	0.9837	0.9955
$\text{erl2}_{\mathcal{A}_2^{\mathcal{D}}}$	9.634E-2	4.453E-2	1.848E-2
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D}}}$	6.508E-2	2.945E-2	1.294E-2
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D}}}$	5.292E-2	2.507E-2	1.202E-2
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D}}}$	5.909E-2	3.311E-2	1.730E-2
$\text{ratiol2}_{\mathcal{A}_2^{\mathcal{D}}}$		1.11	1.27
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D}}}$		1.14	1.18
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D}}}$		1.08	1.06
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D}}}$		0.84	0.94
$\text{nit}_{\mathcal{S}_1^{\mathcal{D}}}$	6	6	6
$\text{nit}_{\mathcal{S}_2^{\mathcal{D}}}$	8	8	9
$\text{nit}_{\mathcal{S}_3^{\mathcal{D}}}$	12	16	60

Table 1.4, **the NLFECC2 schemes** $S^{\mathcal{D}}$ with **the first correction** $\beta^{\mathcal{D}}$.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_2^{\mathcal{D}}}$	8.501E-3	1.278E-3	1.777E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D}}}(\eta = 0.25)$	1.041E-2	1.457E-3	1.910E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D}}}(\eta = 0.5)$	1.243E-2	1.651E-3	2.059E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D}}}(\eta = 1)$	1.610E-2	2.041E-3	2.411E-4
$\text{umax}_{\mathcal{A}_2^{\mathcal{D}}}$	1.0172	1.0096	1.0033
$\text{umax}_{\mathcal{S}_1^{\mathcal{D}}}$	1.0084	1.0085	1.0032
$\text{umax}_{\mathcal{S}_2^{\mathcal{D}}}$	1.0009	1.0075	1.0031
$\text{umax}_{\mathcal{S}_3^{\mathcal{D}}}$	0.9883	1.0056	1.0029
$\text{erl2}_{\mathcal{A}_2^{\mathcal{D}}}$	9.634E-2	4.453E-2	1.848E-2
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D}}}$	8.115E-2	4.040E-2	1.767E-2
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D}}}$	6.990E-2	3.715E-2	1.698E-2
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D}}}$	5.782E-2	3.306E-2	1.595E-2
$\text{ratiol2}_{\mathcal{A}_2^{\mathcal{D}}}$		1.11	1.27
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D}}}$		1.01	1.19
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D}}}$		0.91	1.13
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D}}}$		0.81	1.05
$\text{nit}_{\mathcal{S}_1^{\mathcal{D}}}$	5	5	5
$\text{nit}_{\mathcal{S}_2^{\mathcal{D}}}$	6	6	6
$\text{nit}_{\mathcal{S}_3^{\mathcal{D}}}$	7	9	11

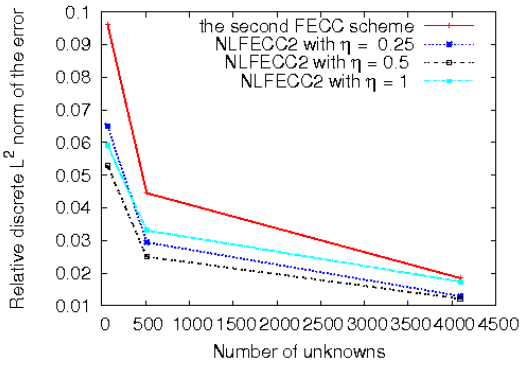
Table 1.5, **the NLFECC2 schemes** $S^{\mathcal{D}}$ with **the second correction** $\overline{\beta}^{\mathcal{D}}$.

numkw	64	512	4096
$\text{umin}_{\mathcal{A}_2^{\mathcal{D}}}$	8.501E-3	1.278E-3	1.777E-4
$\text{umin}_{\mathcal{S}_1^{\mathcal{D}}}(\eta = 0.25)$	1.119E-2	1.544E-3	1.982E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D}}}(\eta = 0.5)$	1.397E-2	1.828E-3	2.204E-4
$\text{umin}_{\mathcal{S}_3^{\mathcal{D}}}(\eta = 1)$	1.832E-2	2.361E-3	2.862E-4
$\text{umax}_{\mathcal{A}_2^{\mathcal{D}}}$	1.0172	1.0096	1.0033
$\text{umax}_{\mathcal{S}_1^{\mathcal{D}}}$	1.0026	1.0076	1.0031
$\text{umax}_{\mathcal{S}_2^{\mathcal{D}}}$	0.9911	1.0058	1.0029
$\text{umax}_{\mathcal{S}_3^{\mathcal{D}}}$	0.9737	1.0027	1.0025
$\text{erl2}_{\mathcal{A}_2^{\mathcal{D}}}$	9.634E-2	4.453E-2	1.848E-2
$\text{erl2}_{\mathcal{S}_1^{\mathcal{D}}}$	7.363E-2	3.774E-2	1.704E-2
$\text{erl2}_{\mathcal{S}_2^{\mathcal{D}}}$	6.041E-2	3.335E-2	1.596E-2
$\text{erl2}_{\mathcal{S}_3^{\mathcal{D}}}$	5.605E-2	3.057E-2	1.587E-2
$\text{ratiol2}_{\mathcal{A}_2^{\mathcal{D}}}$		1.11	1.27
$\text{ratiol2}_{\mathcal{S}_1^{\mathcal{D}}}$		0.96	1.15
$\text{ratiol2}_{\mathcal{S}_2^{\mathcal{D}}}$		0.85	1.06
$\text{ratiol2}_{\mathcal{S}_3^{\mathcal{D}}}$		0.87	0.94
$\text{nit}_{\mathcal{S}_1^{\mathcal{D}}}$	5	5	6
$\text{nit}_{\mathcal{S}_2^{\mathcal{D}}}$	6	7	8
$\text{nit}_{\mathcal{S}_3^{\mathcal{D}}}$	8	19	38

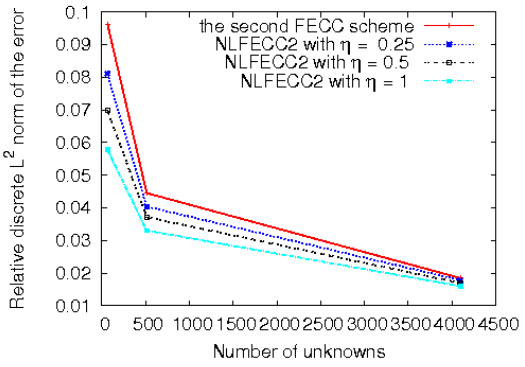
Table 1.6, **the NLFECC2 schemes** $S^{\mathcal{D}}$ with **the third correction** $\widetilde{\beta}^{\mathcal{D}}$.

Comments about Table 1.4, 1.5 and 1.6: for the strong anisotropic tensor and the cube meshes, the order of convergence in the L^2 norm for the second FECC scheme is not second order, because we do not use enough meshes. However, we observe the orders for all the NLFECC2 schemes are close to 1 with $\eta = 0.25, 0.5, 1$. All these schemes also satisfy the discrete maximum principle.

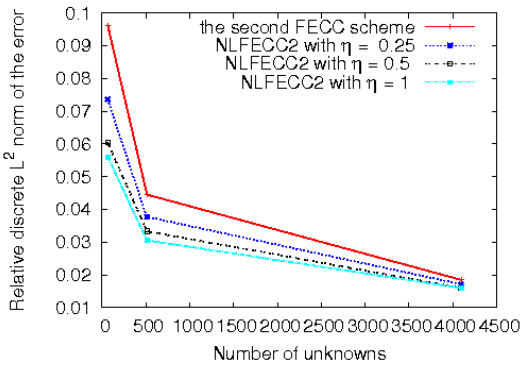
In the following figures, we show the comparison between the second FECC scheme and the non-linear schemes:



the first correction.



the second correction.



the third correction.

Test 2: Stationary non analytical solution

The permeability is

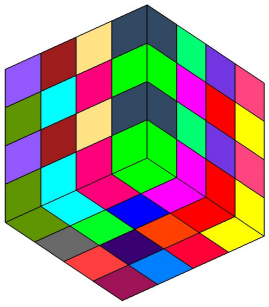
$$\Lambda(x,y,z) = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The analytical solution is equal to 0 on the boundary $\partial\Omega$.

The source term is a discontinuous function on Ω :

$$f(x,y,z) = \begin{cases} 1000 & \text{if } (x,y,z) \in (0,5;0,75) \times (0,5;0,75) \times (0,5;0,75). \\ 0 & \text{if } \Omega/(0,5;0,75) \times (0,5;0,75) \times (0,5;0,75). \end{cases}$$

The primary meshes are the cube meshes:

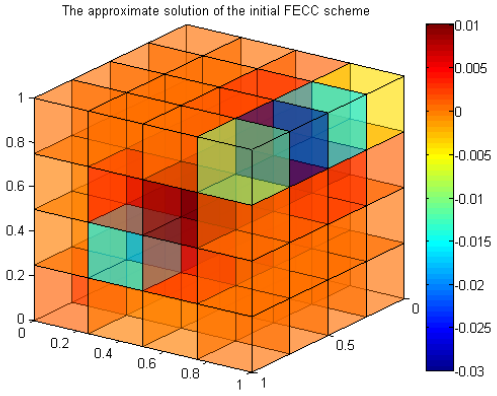


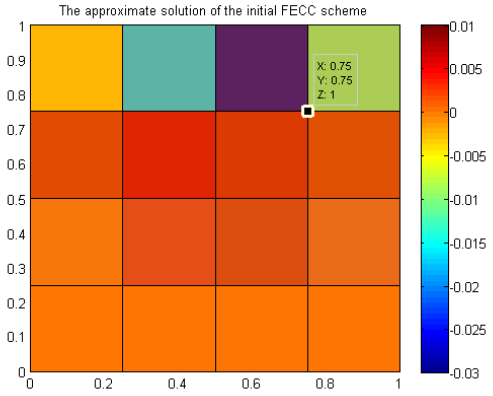
the unit cube.

Test 2 is used to evaluate the respect of the discrete maximum principle. We obtain the following numerical results:

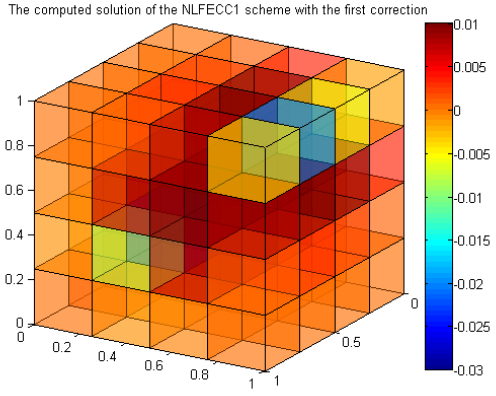
numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	-4.733E-2	-3.435E-2	-1.101E-2
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}(\eta=0.25)}$	-2.621E-2	-9.054E-3	-5.247E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}(\eta=0.5)}$	-4.950E-3	-6.112E-4	-3.355E-5
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}(\eta=1)}$	3.923E-4	1.518E-5	2.919E-8
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.5031	0.4220	0.4651
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.3956	0.3519	0.3909
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.3040	0.2879	0.3398
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.1798	0.2028	0.2701
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	11	9	8
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	19	14	11
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	24	37	23

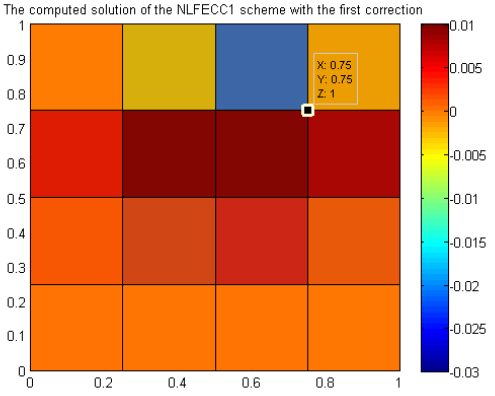
Table 2.1, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the first correction** $\beta^{\mathcal{D},\mathcal{D}^*}$.



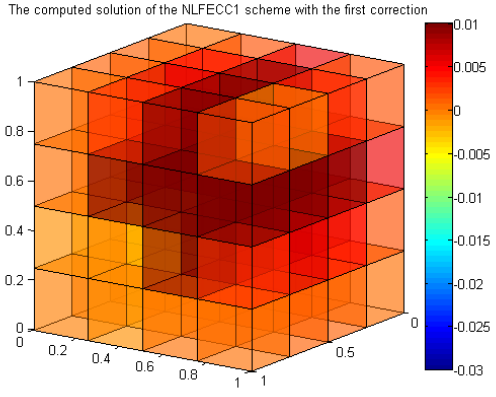


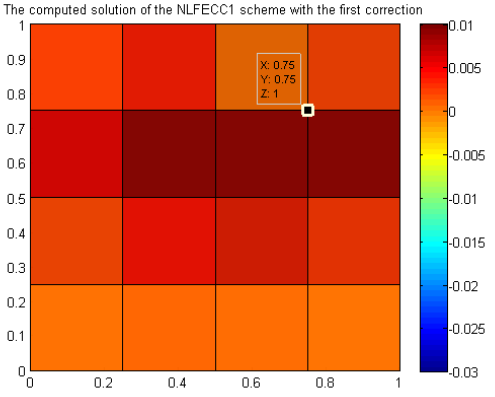
For $\eta = 0.25$,



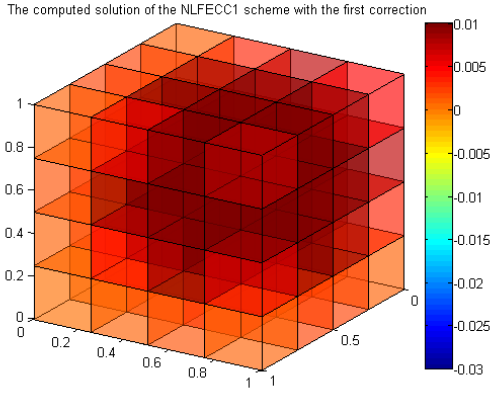


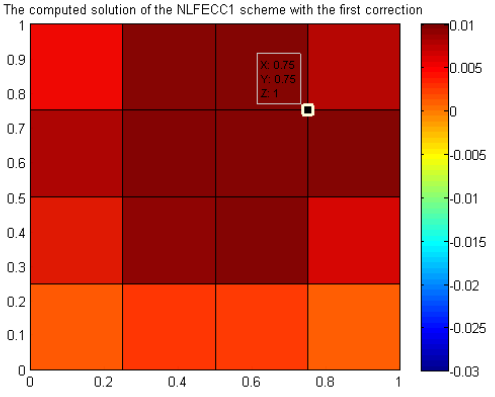
For $\eta = 0.5$,





For $\eta = 1$,

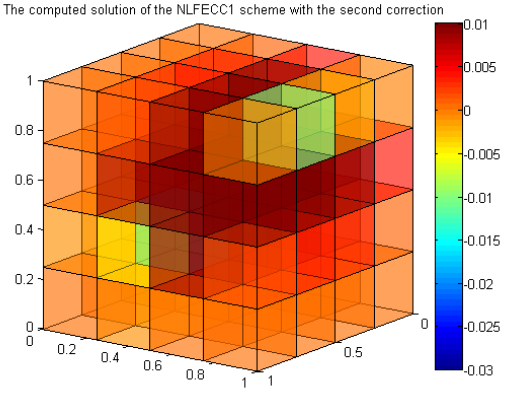


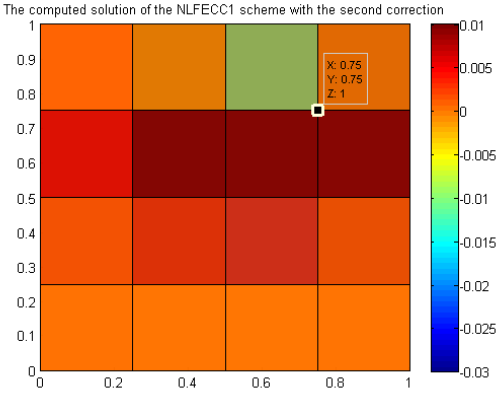


numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	-4.733E-2	-3.435E-2	-1.101E-2
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}(\eta=0.25)}$	-1.637E-2	-1.669E-2	-7.162E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}(\eta=0.5)}$	4.676E-5	8.092E-8	-9.163E-7
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}(\eta=1)}$	4.780E-4	6.297E-6	1.913E-9
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.5031	0.4220	0.4651
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.3579	0.3823	0.4493
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.2638	0.3321	0.4334
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.1763	0.2709	0.4037
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	12	10	7
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	16	14	9
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	16	22	17

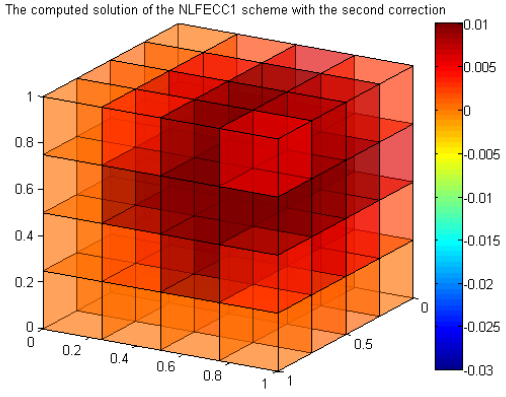
Table 2.2, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the second correction** $\overline{\beta}^{\mathcal{D},\mathcal{D}^*}$.

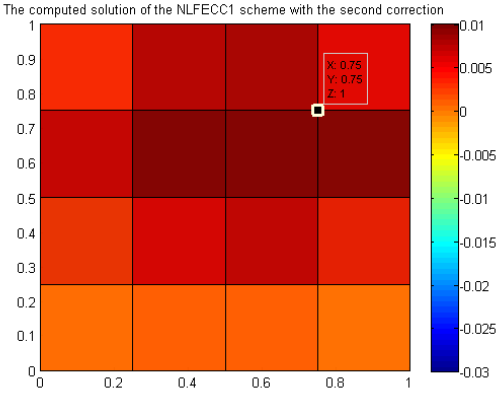
For $\eta = 0.25$,



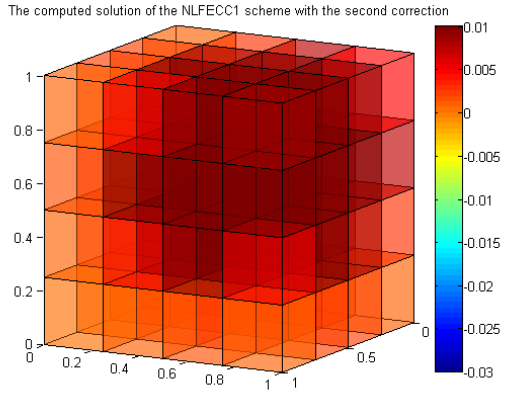


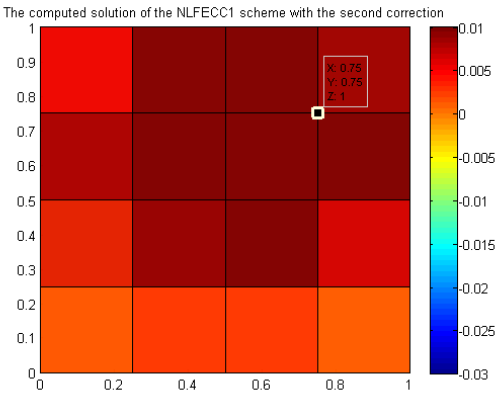
For $\eta = 0.5$,





For $\eta = 1$,

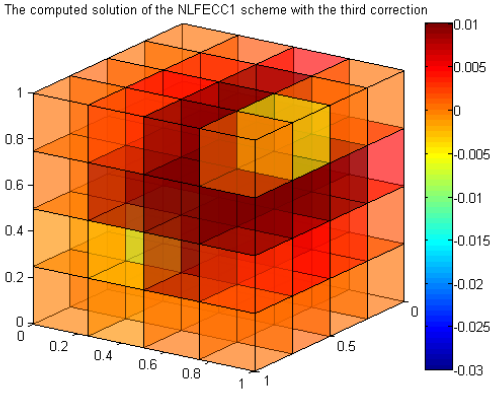


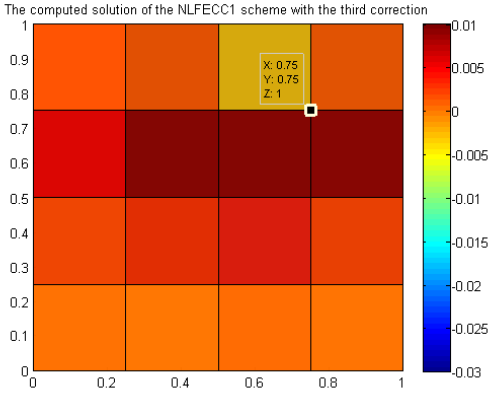


numkw	64	512	4096
$\text{umin}_{\mathcal{A}_1^{\mathcal{D}}}$	-4.733E-2	-3.435E-2	-1.101E-2
$\text{umin}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.25)$	-9.517E-3	-6.994E-3	–4.698E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}(\eta = 0.5)$	1.190E-4	6.376E-7	2.371E-11
$\text{umin}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}(\eta = 1)$	6.931E-4	2.086E-5	2.653E-8
$\text{umax}_{\mathcal{A}_1^{\mathcal{D}}}$	0.5031	0.4220	0.4651
$\text{umax}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	0.3348	0.3593	0.4375
$\text{umax}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	0.2350	0.3010	0.4117
$\text{umax}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	0.1439	0.2284	0.3677
$\text{nit}_{\mathcal{S}_1^{\mathcal{D},\mathcal{D}^*}}$	14	11	7
$\text{nit}_{\mathcal{S}_2^{\mathcal{D},\mathcal{D}^*}}$	19	15	11
$\text{nit}_{\mathcal{S}_3^{\mathcal{D},\mathcal{D}^*}}$	17	37	28

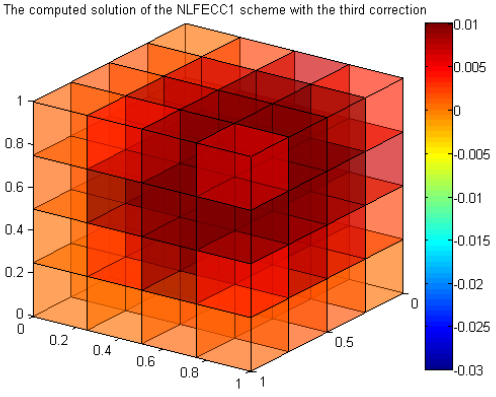
Table 2.3, **the NLFECC1 schemes** $S^{\mathcal{D},\mathcal{D}^*}$ with **the third correction** $\widetilde{\beta}^{\mathcal{D},\mathcal{D}^*}$.

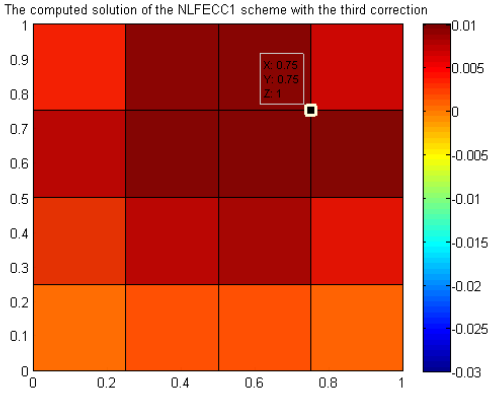
For $\eta = 0.25$,



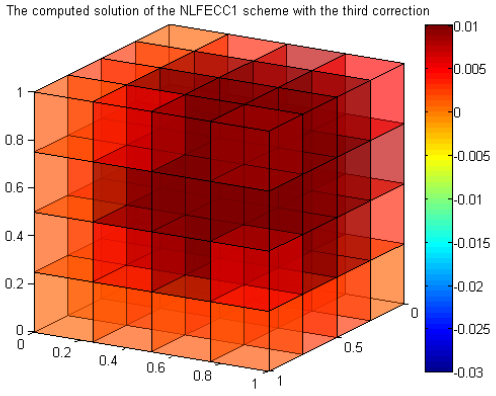


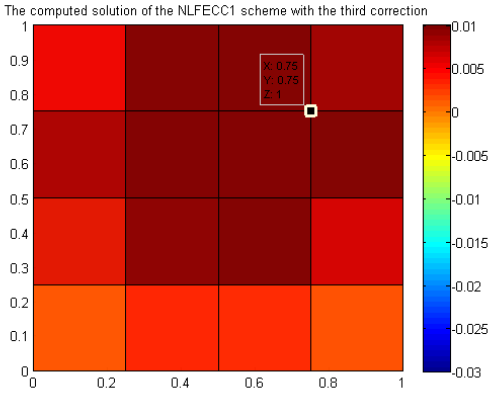
For $\eta = 0.5$,





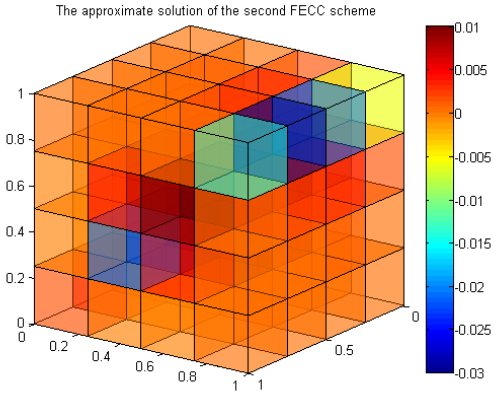
For $\eta = 1$,

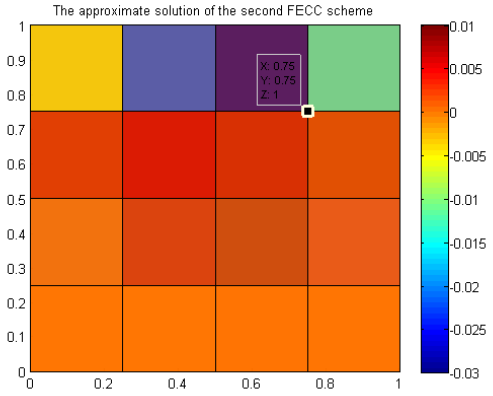




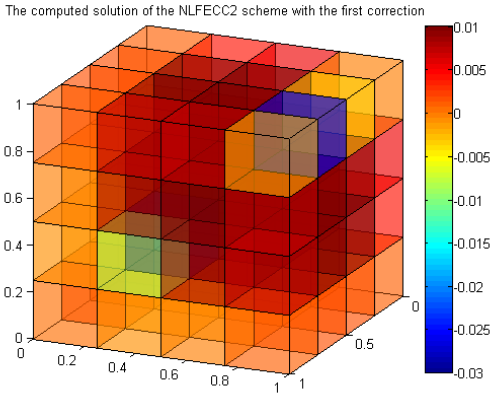
numkw	64	512	4096
$\text{umin}_{\mathcal{A}_2^{\mathcal{D}}}$	-6.311E-2	-4.287E-2	-1.301E-2
$\text{umin}_{\mathcal{S}_1^{\mathcal{D}}}(\eta = 0.25)$	-3.341E-2	-6.052E-3	-3.034E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D}}}(\eta = 0.5)$	-2.243E-3	-1.661E-5	-4.651E-7
$\text{umin}_{\mathcal{S}_3^{\mathcal{D}}}(\eta = 1)$	1.107E-3	2.816E-5	5.755E-8
$\text{umax}_{\mathcal{A}_2^{\mathcal{D}}}$	0.6708	0.5584	0.5027
$\text{umax}_{\mathcal{S}_1^{\mathcal{D}}}$	0.5156	0.4322	0.4168
$\text{umax}_{\mathcal{S}_2^{\mathcal{D}}}$	0.3715	0.3363	0.3539
$\text{umax}_{\mathcal{S}_3^{\mathcal{D}}}$	0.1965	0.2166	0.2701
$\text{nit}_{\mathcal{S}_1^{\mathcal{D}}}$	12	10	9
$\text{nit}_{\mathcal{S}_2^{\mathcal{D}}}$	20	15	13
$\text{nit}_{\mathcal{S}_3^{\mathcal{D}}}$	20	29	68

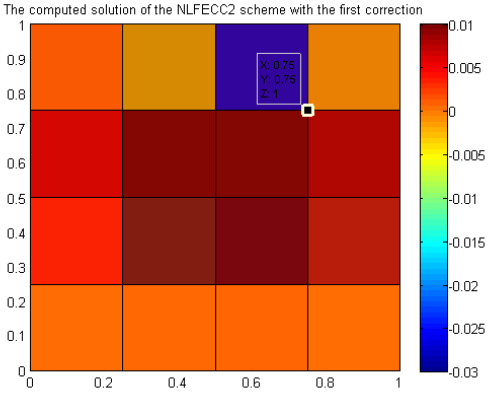
Table 2.4,**the NLFECC2 schemes** with **the first correction** $\overline{\beta}^{\mathcal{D}}$.



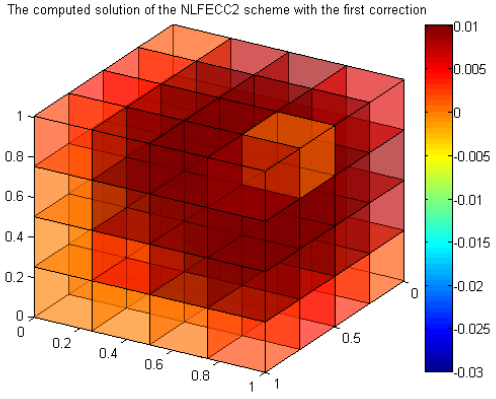


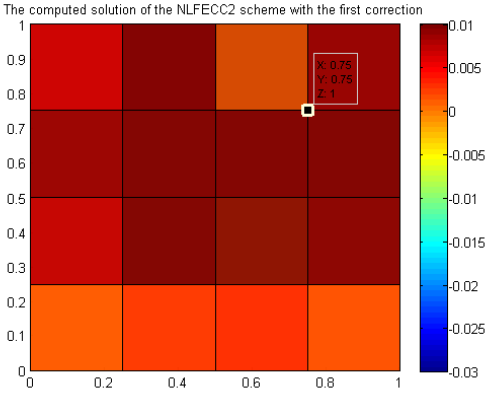
For $\eta = 0.25$,



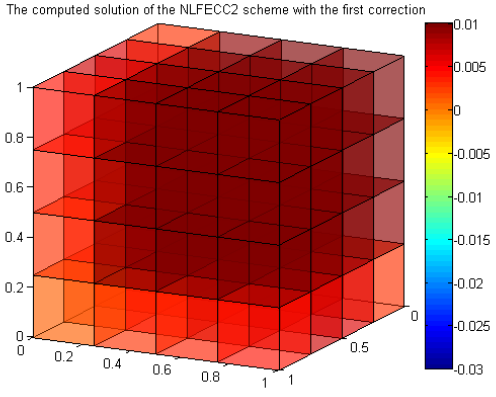


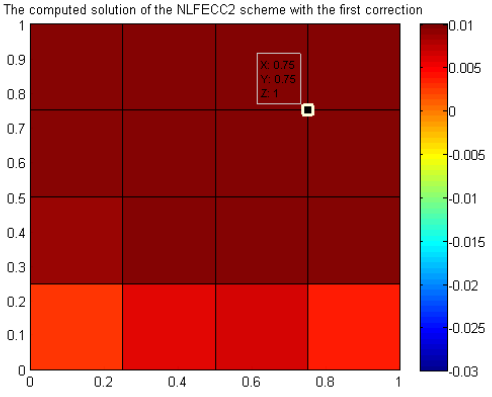
For $\eta = 0.5$,





For $\eta = 1$,

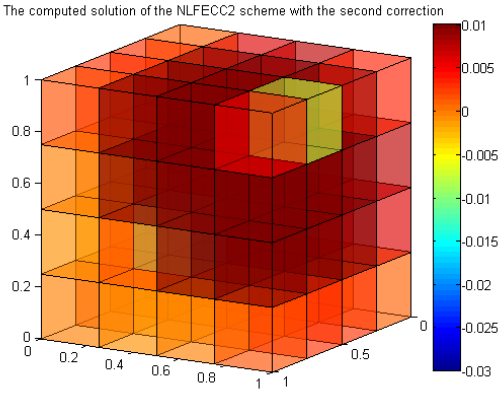


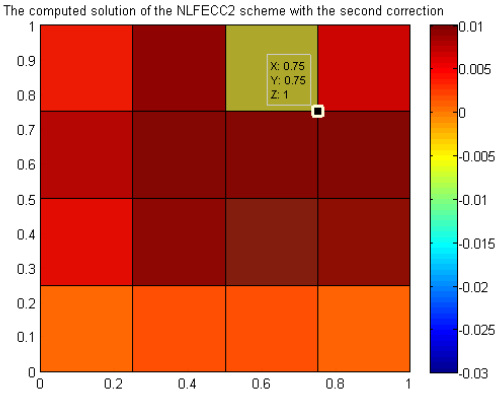


numkw	64	512	4096
$\text{umin}_{\mathcal{A}_2^{\mathcal{D}}}$	-6.311E-2	-4.287E-2	-1.301E-2
$\text{umin}_{\mathcal{S}_1^{\mathcal{D}}}(\eta = 0.25)$	-1.086E-2	-2.809E-3	-3.263E-4
$\text{umin}_{\mathcal{S}_2^{\mathcal{D}}}(\eta = 0.5)$	2.792E-4	9.661E-7	4.000E-11
$\text{umin}_{\mathcal{S}_3^{\mathcal{D}}}(\eta = 1)$	1.251E-3	1.835E-5	1.239E-8
$\text{umax}_{\mathcal{A}_2^{\mathcal{D}}}$	0.6708	0.5584	0.5027
$\text{umax}_{\mathcal{S}_1^{\mathcal{D}}}$	0.4238	0.4416	0.4784
$\text{umax}_{\mathcal{S}_2^{\mathcal{D}}}$	0.2855	0.3618	0.4517
$\text{umax}_{\mathcal{S}_3^{\mathcal{D}}}$	0.1725	0.2679	0.4041
$\text{nit}_{\mathcal{S}_1^{\mathcal{D}}}$	13	9	7
$\text{nit}_{\mathcal{S}_2^{\mathcal{D}}}$	14	14	11
$\text{nit}_{\mathcal{S}_3^{\mathcal{D}}}$	10	33	25

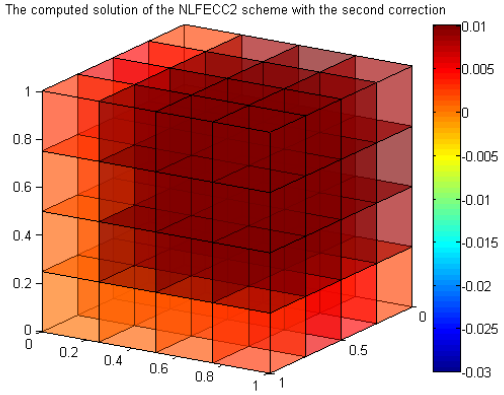
Table 2.5, **the NLFECC2 schemes** with **the second correction** $\overline{\beta}^{\mathcal{D}}$.

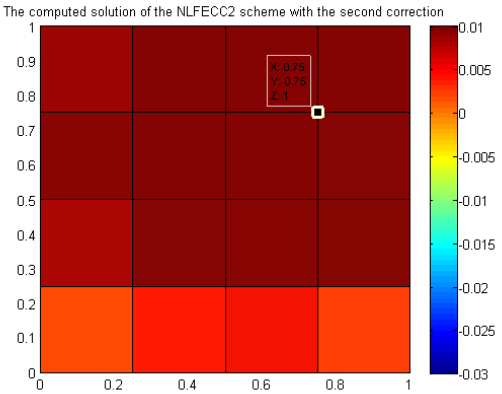
For $\eta = 0.25$,



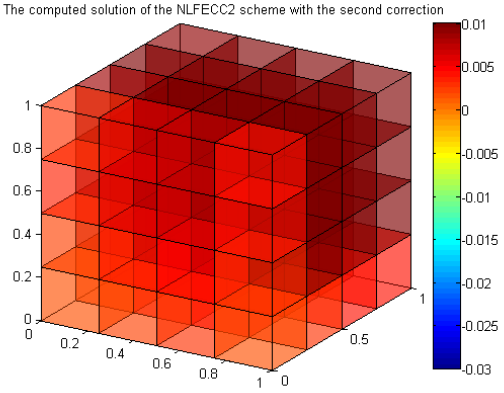


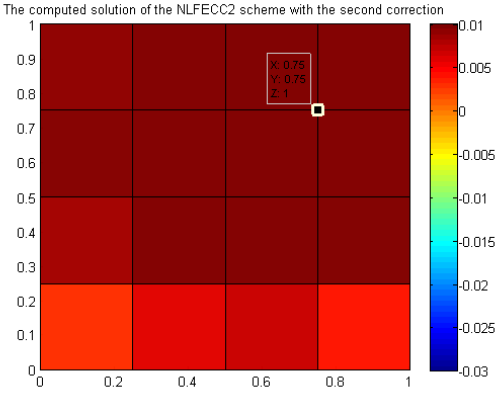
For $\eta = 0.5$,





For $\eta = 1$,





In Test 2, although we do not know the analytical solution, this solution is equal or greater than 0 because of the positive source term and the discrete maximum principle (DMP).

Comments about Table 2.1, 2.2, 2.4: for the cube meshes, the strong anisotropic tensor and the discontinuous source term, the FECC schemes and the non-linear schemes with $\eta = 0.25,\ 0.5$ violate DMP. The non-linear schemes with $\eta = 1$ satisfy DMP. However, number of iterations needed to compute the approximate solution of the non-linear schemes with $\eta = 1$ (nit) are greater than nit of the other non-linear schemes.

Comments about Table 2.3, 2.5: for the cube meshes, the strong anisotropic tensor and the discontinuous source term, the FECC schemes and the non-linear schemes with $\eta = 0.25$ do not satisfy DMP. On the other hand, with $\eta = 0.5,\ 1$, the non-linear schemes satisfy DMP.

Chapter 8

Conclusion and Perspectives

Contents

8.1 Conclusion

The objective of this work was to represent a new cell-centered scheme and study its main properties for heterogeneous anisotropic diffusion problem (2.1) on general meshes. Its name is the Finite Element Cell-Centered scheme (the FECC scheme). The FECC scheme has the following main characteristics:

- Its ideas are based on the standard finite element method.
- It uses a particular definition of dual meshes.
- In heterogeneous and homogeneous anisotropic cases, it is locally conservative.
- It is a cell-centered scheme, its stencil is equal or less than nine on quadrangular meshes and twenty seven on hexahedral meshes.
- It is exact on cell-wise affine solutions for cell-wise constant diffusion tensors.
- In general cases, with light assumption (hypothesis 2.1 and 3.1), the matrix which is associated to our scheme is symmetric and positive definite on general meshes.
- It takes into account nonconforming meshes. This is helpful to use Adaptive Mesh Refinement to locally increase the precision.
- It is convergent for discontinuous tensors which are piecewise Lipschitz-continuous.
- It is very precise in L^2 norm in comparison with classical finite volume schemes (FVCA 5 & 6 tests).

The other objectives were to study the non-linear FECC schemes with non-linear corrections to satisfy the maximum principle. It is well-known that classical finite volume and finite element schemes fail to satisfy the maximum principle for distorted meshes or for high anisotropy ratio of diffusion tensors [19], [45], [34], [41]. Moreover, in [45], the authors proved that it is impossible to construct nine-point methods which unconditionally satisfy the monotonicity criteria when the discretization satisfies local conservation and exact reproduction of linear potential fields. In these papers, they propose some conditions to satisfy the maximum principle. However, these conditions are difficult to satisfy, especially in three dimensions. The FECC scheme also violates the discrete maximum principle. Hence, we used non-linear corrections to correct the initial and the second FECC schemes. These non-linear schemes are named the NLFECC1 and the NLFECC2 schemes satisfying the maximum principle and preserving the main properties: symmetry, positive-definiteness, coercivity, a priori estimate, existence of solution to the schemes and the condition for convergence toward the solution of (2.2) when the size of the meshes tends to 0. Besides, it is easy for us to implement, because we can use the computed data of the linear system associated to the FECC schemes. In the section 7.4, we showed 3D numerical results of the non-linear schemes and comparisons between the FECC schemes and these schemes.

8.2 Perspectives

A natural extension could be to use the FECC scheme in a convection diffusion dispersion equation. If one needs to satisfy the minimum and the maximum principles, the NLFECC1 and the NLFECC2 schemes will be coupled to a classical cell-centered convective scheme. Another application could be a coupling with a chemical model, where it is crucial to obtain a solution satisfying the physical bounds.

Résumé

Nous présentons de nouveaux schémas numériques pour l'approximation de problèmes de diffusion hétérogène et anisotrope sur des maillages généraux. Sous des hypothèses correspondant aux cas industriels, nous montrons qu'un premier schéma, qui est centré sur les mailles, possède un petit stencil et converge dans le cas de tenseurs discontinus. La preuve de la convergence repose sur des propriétés de consistance des gradients discrets issus du schéma. Dans une seconde partie, nous proposons des méthodes de correction non linéaire du schéma initial pour obtenir le principe du maximum.

L'efficacité de ces schémas est étudiée sur des tests numériques ayant fait l'objet de bancs d'essais d'une grande variété de schémas de volumes finis. Les comparaisons avec les schémas volumes finis classiques montrent l'apport de ces schémas en termes de précision. Nous montrons ainsi le bon comportement de ces schémas sur des maillages déformés, et le maintien de la précision des schémas non-linéaires, alors que les oscillations ont été supprimées.

Mots clés : Diffusion hétérogène anisotrope, maillages généraux, volumes finis, schéma centré sur les mailles, principe du maximum.

Abstract

We present a new scheme for the discretization of heterogeneous anisotropic diffusion problems on general meshes. With light assumptions, we show that the algorithm can be written as a cell-centered scheme with a small stencil and that it is convergent for discontinuous tensors. The key point of the proof consists in showing both the strong and the weak consistency of the method. Besides, we study non-linear corrections to correct the FECC scheme, in order to satisfy the discrete maximum principle (DMP).

The efficiency of the scheme is demonstrated through numerical tests of the 5th & 6th International Symposium on Finite Volumes for Complex Applications - FVCA 5 & 6. Moreover, the comparison with classical finite volume schemes emphasizes the precision of the method. We also show the good behaviour of the algorithm for nonconforming meshes. In addition, we give some numerical tests to check the existence for the non-linear FECC schemes.

Keywords : Heterogeneous anisotropic diffusion, general grids, finite volumes, finite elements, cell-centered scheme, discrete maximum principle.

Bibliography

[1] I. Aavatsmark. An introduction to multipoint flux approximations for quadrilateral grids. *Comput. Geosci.*, 6(3-4):405–432, 2002.

[2] I. Aavatsmark, T. Barkve, ∅. B∅e, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods. *SIAM Journal on SC.Comp*, 19:1700–1716, 1998.

[3] I. Aavatsmark, T. Barkve, ∅. B∅e, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. Part II: Discussion and numerical results. *SIAM Journal on SC.Comp*, 19:1717–1736, 1998.

[4] I. Aavatsmark and R. Klausen. Well index in reservoir simulation for slanted and slightly curved wells in 3D grid. *SPE Journal*, 8:41–48, 2008.

[5] L. Agelas, R. Eymard, and R. Herbin. A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media. *Comptes rendus de l'Académie des Sciences Mathématique*, 347:673–676, 2009.

[6] L. Agelas and D.A. Di Pietro. A symmetric finite volume scheme for anisotropic heterogeneous second-order elliptic problems. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 705–716, 2008.

[7] L. Agelas, D.A. Di Pietro, and J. Droniou. The G method for heterogeneous anisotropic diffusion on general meshes. *M2AN Math. Model.Numer.Anal.*, 44:597–625, 2010.

[8] H. Amor, M. Bourgeois, and G. Mathieu. A linear finite element solver. *Finite Volumes for Complex Applications VI (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, 2:131–135, 2011.

[9] G. Anasanay-Alex, B. Piar, and D. Vola. A Galerkin finite element solution. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 717–733, 2008.

[10] A. Bourgeat, M. Kern, S. Schumacher, and J. Talandier. The couplex test cases: Nuclear Waste Disposal Simulation. *In Proc. of the 5th Internat. Conf. on Supercomputing in Nuclear Applications*, 2003.

[11] E. Burman and A. Ern. Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes. *Comptes rendus de l'Académie des Sciences Mathématique*, 338:641–646, 2001.

[12] M. Butez, G. Bordier, X. Vitart, and I. Hablot. Nuclear waste management and processing: between legacy and anticipation. *CLEFS CEA*, No. 53:26–31, 2005.

[13] C. Cances, M. Cathala, and C. Le Potier. Monotone coercive cell-centered finite volume schemes for anisotropic diffusion equations. *http://hal.archives-ouvertes.fr/hal-00643838/*.

[14] P. G. Ciarlet and P. A. Raviart. Maximum principle and uniform convergence for the finite element method. *Comput. Methods Appl. Mech. Engrg*, 2:17–31, 1973.

[15] Y. Coudière and G. Manzini. The cell-centered finite volume method using least squares vertex reconstruction('Diamond Scheme'). *Finite Volumes for Complex Applications VI (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, 2:185–192, 2011.

[16] Y. Coudière, J.P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two dimensional convection diffusion problem. *M2AN*, 33:493–516, 1999.

[17] A. Danilov and Y. Vassilevski. A monotone nonlinear finite volume method for diffusion equations on polyhedral meshes. *Finite Volumes for Complex Applications VI (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, 2:213–222, 2011.

[18] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN*, 39:1203–1249, 2005.

[19] A. Draganescu, T. F. Dupont, and L. R. Scott. Failure of the discrete maximum priciple for an elliptic finite problem. *Math. Comput*, 74(249):1–23, 2004.

[20] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to Mimetic Finite Difference, Hybrid Finite Volume and Mixed Finite Volume Methods. *Math. Model. Methods. Appl. Sci (M3AS)*, 20:265–295, 2010.

[21] J. Droniou and C. Le Potier. Construction and convergence study of schemes preserving the elliptic local maximum principle. *SIAM J.Numer.Anal*, 49:459–490, 2011.

[22] R. Eymard, T. Gallouët, and R. Herbin. The finite volume method. *Handbook of Numerical Analysis, Ph. Ciarlet, J.L. Lions eds, North Holland*, pages 715–1022, 2000.

[23] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general non-conforming meshes. Sushi: A Scheme Using Stabilization and Hybrid Interfaces. *IMA Journal of Numerical Analysis*, 30:1009–1034, 2010.

[24] R. Eymard, T. Gallouët, and R. Herbin. The Sushi scheme. *Finite Volumes for Complex Applications VI (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, 2:205–212, 2011.

[25] R. Eymard, C. Guichard, and R. Herbin. The VAG scheme. *Finite Volumes for Complex Applications VI (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, 2:213–222, 2011.

[26] R. Eymard, C. Guichard, and R. Herbin. Small-stencil 3D schemes for diffusive flows in porous media. *M2AN*, 46:265–290, 2012.

[27] I. Faille. A control volume method to solve an elliptic equation on a two-dimensional irregular mesh. *Comput. Methods Appl. Mech. Engrg*, 100(2):275–290, 1992.

[28] Zhang Fuzhen. The Schur complement and its applications. *Springer*, 2005.

[29] A. Genty and 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*, 88:65–85, 2011.

[30] G.Yuan and Z. Sheng. Monotone finite volume schemes for diffusion equations on polygonal meshes. *J. Comput. Physics*, 227:6288–6312, 2008.

[31] F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys*, 160:481–499, 2000.

[32] F. Hermeline. Numerical experiments with the DDFV method. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 851–864, 2008.

[33] C. Chainais Hillairet, J. Droniou, and R. Eymard. Use of Mixed Finite Volume Method. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 751–760, 2008.

[34] H. Hoteit, R. Mose, B. Phillippe, and Ph. Ackerer. The maximum principle violations of the mixed-hybrid finite-element method applied to diffusion equations. *Numer. Meth. Eng*, 55(12):1373–1390, 2002.

[35] I. Kapyrin. A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes. *Dokl. Math*, 76:734–738, 2007.

[36] L. F. Konikow. Numerical models of groundwater flow and transport. In: Manual on Mathematical Models in Isotope Hydrogeology. *International Atomic Energy Agency Rept.*, IAEA-TECDOC-910, Vienna, Austria:59–112, 1996.

[37] L. F. Konikow and T. E. Reilly. Groundwater modelling. In: The Handbook of Groundwater Engineering (J. W. Delleur, ed.). *CRC Press, Boca Raton 20:*, 1998.

[38] S. Korotov, M. Krízek, and P. Neittaanmäki. Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle. *Mathematics of Computation*, 70(233):107–119, 2000.

[39] R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski. Finite volume methods for convection-diffusion problems. *SIAM Journal on Numerical Analysis*, 33, No.1:31–55, 1996.

[40] K. Lipnikov. Mimetic finite difference method. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 801–814, 2008.

[41] K. Lipnikov and D. Svyatskiy G. Manzini. Analysis of the monotonicity conditions in the mimetic finite difference method for elliptic problems. *J. Comput. Phys*, 230:2620–2642, 2011.

[42] K. Lipnikov, M. Shashkov, D. Svyatskiy, and Yu. Vassilevski. Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes. *J. Comput. Phys*, 227:492–512, 2007.

[43] K. Lipnikov, D. Svyatskiy, and Yu. Vassilevski. Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes. *J. Comput. Phys*, 228:703–716, 2009.

[44] S. Mundal, D.A. Di Pietro, and I. Aavatsmark. Compact-stencil MPFA method for heterogeneous highly anisotropic second-order elliptic problems. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 905–918, 2008.

[45] J. M. Nordbotten, I. Aavatsmark, and G. T. Eigestad. Monotonicity of control volume methods. *Numer. Math*, 106:255–288, 2007.

[46] C. Le Potier. Schéma volumes finis monotone pour des opérateurs de diffusion fortement anisotropes sur des maillages de triangles non structurés. *C. R. Acad. Sci. Paris Ser. I*, 341:787–792, 2005.

[47] 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*, 340:921–926, 2005.

[48] C. Le Potier. Numerical results with two cell-centered finite volume schemes for heterogeneous anisotropic diffusion operators. *Finite Volumes for Complex Applications V (R. Eymard and J.M. Hérard, eds.), John Wiley & Sons*, pages 825–842, 2008.

[49] C. Le Potier. A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators. *IJFV*, 6:1–20, 2009.

[50] C. Le Potier. Un schéma linéaire vérifiant le principle du maximum pour des opérateurs de diffusion très anisotropes sur des maillages déformés. *C. R. Acad. Sci. Paris Ser. I*, 347:105–110, 2009.

[51] C. Le Potier. A nonlinear correction and maximum principle for diffusion operators discretized using cell-centered finite volume schemes. *C. R. Acad. Sci. Paris Ser. I*, 348:691–695, 2010.

[52] C. Le Potier and O. Thanh Hai. A cell-centered scheme for heterogeneous anisotropic diffusion problems on general meshes. *IJFV*, 8:1–40, 2012.

[53] C. Le Potier and A. Mahamane. A nonlinear correction and maximum principle for diffusion operators with hybrid schemes. *C. R. Acad. Sci. Paris, Ser. I*, 350:101–106, 2012.

[54] R. Herbin R. Eymard and M. Rhoudaf. Approximation of the biharmonic problem using P1 finite elements. *J. of Num. Math.*, 19(1):1–26, 2011.

[55] Z. Sheng and G. Yuan. The finite volume scheme preserving extremum principle for diffusion problems on polygonal meshes. *J. Comput. Physics*, 230:2588–2604, 2011.

[56] Shuyu Sun and Jiangguo Liu. A locally conservative finite element method based on piecewise constant enrichment of the continuous Galerkin method. *J. Comput. Phys*, 228:703–716, 2009.

Résumé:

Nous présentons de nouveaux schémas numériques pour l'approximation de problèmes de diffusion hétérogène et anisotrope sur des maillages généraux. Sous des hypothèses correspondant aux cas industriels, nous montrons qu'un premier schéma, qui est centré sur les mailles, possède un petit stencil et converge dans le cas de tenseurs discontinus. La preuve de la convergence repose sur des propriétés de consistance des gradients discrets issus du schéma. Dans une seconde partie, nous proposons des méthodes de correction non linéaire du schéma initial pour obtenir le principe du maximum.

L'efficacité de ces schémas est étudiée sur des tests numériques ayant fait l'objet de bancs d'essais d'une grande variété de schémas de volumes finis. Les comparaisons avec les schémas volumes finis classiques montrent l'apport de ces schémas en termes de précision. Nous montrons ainsi le bon comportement de ces schémas sur des maillages déformés, et le maintien de la précision des schémas non-linéaires, alors que les oscillations ont été supprimées.

Mots clés :

Diffusion hétérogène anisotrope, maillages généraux, volumes finis, schéma centré sur les mailles, principe du maximum.

Abstract :

We present a new scheme for the discretization of heterogeneous anisotropic diffusion problems on general meshes. With light assumptions, we show that the algorithm can be written as a cell-centered scheme with a small stencil and that it is convergent for discontinuous tensors. The key point of the proof consists in showing both the strong and the weak consistency of the method. Besides, we study non-linear corrections to correct the FECC scheme, in order to satisfy the discrete maximum principle (DMP).

The efficiency of the scheme is demonstrated through numerical tests of the 5th & 6th International Symposium on Finite Volumes for Complex Applications - FVCA 5 & 6. Moreover, the comparison with classical finite volume schemes emphasizes the precision of the method. We also show the good behaviour of the algorithm for nonconforming meshes. In addition, we give some numerical tests to check the existence for the non-linear FECC schemes.

Keywords :

Heterogeneous anisotropic diffusion, general grids, finite volumes, finite elements, cell-centered scheme, discrete maximum principle.