



HAL
open science

Weighted positive nonlinear finite volume method for dominated anisotropic diffusive equations

Cindy Guichard, El Houssaine Quenjel

► **To cite this version:**

Cindy Guichard, El Houssaine Quenjel. Weighted positive nonlinear finite volume method for dominated anisotropic diffusive equations. *Advances in Computational Mathematics*, 2022, 48 (81). hal-03414953v1

HAL Id: hal-03414953

<https://hal.science/hal-03414953v1>

Submitted on 4 Nov 2021 (v1), last revised 15 Dec 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Weighted positive nonlinear finite volume method for dominated anisotropic diffusive equations

Cindy Guichard¹ and El Houssaine Quenjel²

¹Sorbonne Université, CNRS, Université de Paris, Inria, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France. cindy.guichard@sorbonne-universite.fr

²Chair of Biotechnology, LGPM, CentraleSupélec, CEBB, 3 rue des Rouges Terres, 51110 Pomacle, France. el-houssaine.quenjel@centralesupelec.fr

November 4, 2021

Abstract

In this paper we propose a new positive finite volume scheme for degenerate parabolic equations with strongly anisotropic diffusion tensors. The key idea is to approximate the fluxes thanks to a weighted-centered scheme depending on the sign of the stiffness coefficients. More specifically, we employ a centered discretization for the mobility-like function when the transmissibilities are positive and a weighted harmonic scheme in regions where the negative transmissibilities occurs. This technique prevents the formation of undershoots which entails the positivity-preserving of the approach. Also, the scheme construction retains the main elements, namely the coercivity and compactness estimates, allowing the existence and the convergence of the nonlinear finite volume scheme under general assumptions on the physical inputs, the nonlinearities and the mesh. The numerical implementation of the methodology shows that the lower bound on the discrete solution is respected and optimal convergence rates are recovered on strongly anisotropic various test-cases.

1 Introduction

Unsteady degenerate diffusion equations with highly anisotropic tensors arise naturally in many models of porous media flows [12, 29], heat transfer [23] and chemotaxis [21]. The establishment of an accurate finite volume approximation to their positive solutions on generic meshes and highly anisotropic diffusion tensors is a challenging topic.

Some positive approaches have been already proposed in the literature for degenerate parabolic equations. We first mention the work of Cancès et al. [8] where a Godunov scheme was constructed and studied in the case of a nonlinear diffusion model. The main idea is to correct the oscillatory term using a first order upwind scheme with respect to the sign of transmissibilities. As a result, the method ensures the positivity of the approximated solution. It has been extended to a chemotaxis system [22] and to a compressible two-phase flows problem [18]. A slightly improvement was proposed in [19] where the upwinding is taken place only if necessary. However, under some circumstances, both methodologies lack robustness since they induce an important numerical diffusion as the anisotropic ratio becomes important. An alternative formulation was conceived in [25] to limit the impact of the artificial diffusion, but the accuracy remains of first order. Alternatively, accurate schemes based on logarithmic formulations of the potential function were suggested in [9, 27]. The advantage is to tackle quite general meshes and fully tensors. They reinforce the positivity of the solution by construction, but the numerical solver may encounter convergence problems for highly anisotropic diffusion matrices. Furthermore, their extension to complex problems such as porous media flows is not clear. A similar remark was underlined in the recent contribution [26].

The first objective of this work is to devise and investigate an improved positive finite volume discretization for diffusion equations by extending the ideas of [8, 19]. In the present paper, the employed technique is generic on finite volume schemes that could be written using the two points like structure [11, 16, 28]. Here, it is applied in the context of the control volume finite element approximation [1, 17]. The method gives stiffness coefficients with a possible negative sign when the mesh presents some obtuse angles or when the tensor is not scalar. This may yield nonphysical oscillations on the numerical solution. To circumvent this problem, the main contribution is to combine a centered scheme to an upwind one through a weighted harmonic mean. The key idea is to take advantage of the high accuracy offered by the centered scheme while keeping the stability provided by the upstream scheme. The second aim is to examine the robustness of the developed scheme from a practical perspective and compare its behavior to some versions of the literature. It is shown that the novel scheme provides competitive results in terms of stability and accuracy.

This paper follows the subsequent outline. In Section 2, we state the mathematical problem we are interested in and set up the main hypotheses on the physical parameters and on the nonlinearities. We also define the concept of the weak solution to which the numerical scheme must converge. Section 3 specifies the discrete setting consisting of the primal and dual meshes together with basic notations. In Section 4 we introduce a generic nonlinear finite volume discretization encompassing many schemes stemming from the approximation of the nonlinear diffusion function. Several propositions are suggested based on two essential elements, notably accuracy and positivity. A weighted scheme with a better compromise between these two properties is derived. In Section 5, it is established that this novel scheme enables the positivity of the solution and the derivation of the energy estimates. The existence of discrete solutions is also studied. The convergence of the proposed finite volume scheme requires compactness in time and in space together with the passage to the limit. Details on such a procedure are omitted since they greatly resemble to what is already done in the literature. Finally, In Section 6 we perform various numerical experiments putting in practice our approach. Compared to some other schemes, they highlight the ability of the weighted scheme to simultaneously guarantee the solution positivity and retain optimal convergence rates.

2 Model problem

The focus of this work consists in developing and analyzing an accurate positive finite volume method for solving the nonlinear parabolic equation

$$u_t - \operatorname{div} \alpha(u) \mathbb{A} \nabla \beta(u) = q(u), \quad \text{in } \Omega_{t_f}, \quad (2.1)$$

where $\Omega_{t_f} = \Omega \times (0, t_f)$, Ω is a bounded connected open subset of \mathbb{R}^d ($d \geq 2$) and $t_f > 0$ is the final time. The main unknown is $u(x, t)$ which refers to the diffusion of a given quantity (e.g. density, saturation, concentration, etc) or the propagation of heat within Ω . The function α measures the diffusion behavior of the quantity of interest and β is the potential function. The matrix \mathbb{A} accounts for anisotropy. Sources and sinks are represented by q . Set $\partial\Omega = \partial\Omega^N \cup \partial\Omega^D$. We also consider the mixed boundary conditions

$$\alpha(u) \mathbb{A} \nabla \beta(u) \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega^N \times (0, t_f), \quad u = 0 \quad \text{on } \partial\Omega^D \times (0, t_f), \quad (2.2)$$

where \mathbf{n} denotes the outward unit normal to $\partial\Omega^N$, as well as the initial condition

$$u|_{t=0} = u^0 \quad \text{in } \Omega. \quad (2.3)$$

The main assumptions are given below.

(**H**₁) The mobility function $\alpha : [0, +\infty) \rightarrow \mathbb{R}^+$ is continuous, nondecreasing and degenerate at 0 i.e. $\alpha(0) = 0$. It is extended by 0 on $(-\infty, 0)$. The potential β is continuous and (strictly) increasing on \mathbb{R}^+ . Assume that $z \rightarrow \sqrt{\alpha(z)}\beta'(z)$ is locally integrable on \mathbb{R}^+ . Define Kirchhoff's transform $\xi(s) = \int_0^s \sqrt{\alpha(z)}\beta'(z) dz$. We assume that ξ satisfies the inequality $\xi(s) \geq C_\xi \beta(s) - C'_\xi$, for all $s \geq 0$. It is also supposed to verify

$$\exists \varepsilon > 0, \quad s^{1+\varepsilon} \leq C''_\xi \left(1 + \xi(s)^2\right), \quad \forall s \geq 0, \quad (2.4)$$

where C_ξ, C'_ξ, C''_ξ are some positive constants.

(H₂) We suppose that $u^0 \geq 0$ and $u^0, \Upsilon(u^0)$ are belonging to $L^1(\Omega)$ where Υ is a positive primitive to the function β .

(H₃) The diffusion tensor $\mathbb{A}(\cdot)$ is a symmetric matrix belonging to $L^\infty(\Omega)^{d \times d}$ and obeys the ellipticity condition :

$$\mathbb{A}_0 |v|^2 \leq \mathbb{A}(\cdot)v \cdot v \leq \mathbb{A}_1 |v|^2, \quad \forall v \in \mathbb{R}^d,$$

where $\mathbb{A}_0, \mathbb{A}_1$ are given positive constants.

(H₄) The right hand side term q is a bounded continuous function from $[0, +\infty)$ into \mathbb{R}^+ . There exists $C_q > 0$ such that $q(s) \leq C_q, \forall s \geq 0$.

Remark 2.1. *The framework of Assumption (H₁) includes general nonlinearities of β and α . The main aim of the inequality (2.4) is to handle the convergence analysis of the accumulation term following the guidelines of [9, 25]. The examples taken in the numerical section meet the underlined requirements (e.g. $\alpha(u) = \beta(u) = \sqrt{2u}$).*

It is now possible to properly define the kind of the solution we are going to approximate. First, consider

$$H_D^1(\Omega) = \{u \in H^1(\Omega) / u = 0 \text{ on } \partial\Omega^D\}.$$

Definition 2.1. *Under Assumptions (H₁)-(H₄), a weak solution to the mathematical problem (2.1)-(2.3) is a measurable function $u : \Omega_{t_f} \rightarrow \mathbb{R}$ such that the following requirements are fulfilled:*

(i) $u \geq 0, \xi(u) \in L^2(0, t_f; H_D^1(\Omega))$,

(ii) *the weak formulation*

$$-\int_{\Omega_{t_f}} u \psi_t \, dx \, dt - \int_{\Omega} u^0 \psi(\cdot, 0) \, dx + \int_{\Omega_{t_f}} \sqrt{\alpha(u)} \mathbb{A} \nabla \xi(u) \cdot \nabla \psi \, dx \, dt = \int_{\Omega_{t_f}} q(u) \psi \, dx \, dt,$$

holds true for all $\psi \in C_c^\infty(\bar{\Omega} \times [0, t_f])$ with $\psi = 0$ on $\partial\Omega^D \times (0, t_f)$.

It was already established that the problem (2.1)-(2.3) has a unique weak solution in the sense of Definition 2.1. The proof of the existence result can be documented in [2]. The uniqueness question was treated in [24].

3 Mesh in space and in time

This section is devoted to the discretization of Ω_{t_f} and to the description of discrete unknowns and functions.

Two meshes are built, a primal conforming finite element mesh and a dual mesh. This primal mesh denoted by \mathcal{M} is a partition of Ω with nonempty convex disjoint simplices (triangles if $d = 2$, tetrahedra if $d = 3$) such that $\cup_{T \in \mathcal{M}} \bar{T} = \bar{\Omega}$. The vertices set of the collection \mathcal{M} is denoted by \mathcal{V} . Let us consider \mathcal{V}^* as the set of the degrees of freedom located at the domain interior and on the Neumann boundary $\partial\Omega^N$. They are ordered from 1 to $N_{\mathcal{V}^*} = \#\mathcal{V}^*$. The Dirichlet boundary vertices are labeled with numbers from $N_{\mathcal{V}^*} + 1$ to $N_{\mathcal{V}} = \#\mathcal{V}$.

The dual mesh denoted by \mathcal{D} is made of the control volumes that are constructed and centered around the vertices. We detail the construction in the case $d = 2$, similar steps are used for $d > 2$. Let us denote by \mathcal{M}_i the set of triangles sharing the vertex indexed by i , for all $i = 1 \dots N_{\mathcal{V}}$. To each vertex identified by i we link a unique cell K_i defined by connecting the barycenter of every $T \in \mathcal{M}_i$ to the edges of the same T intersecting in i . We designate by \mathcal{F}^T the set containing dual edges included in the triangle T . We set $\mathcal{F}_i^T = \{\nu \in \mathcal{F}^T \mid \nu \subset \partial K_i\}$ whenever $T \in \mathcal{M}_i$. Each interface $\nu_{ij} \in \mathcal{F}_i^T$ is characterized by its length $|\nu_{ij}|$ and the unit normal \mathbf{n}_{ij} oriented from K_i to K_j . The notation $|K_i|$ defines the volume of cell K_i . The center of K_i is denoted by x_i . Fig. 1 illustrates the example of primal mesh together with its corresponding dual mesh.

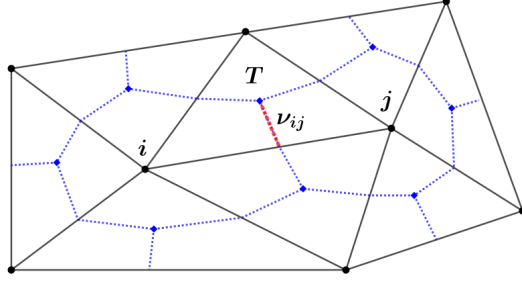


Figure 1: A primal triangulation with the associated dual mesh.

Let us denote by h_T the diameter of the triangle T . The size of the mesh is given by $h_{\mathcal{M}} = \max_{T \in \mathcal{M}} h_T$. Let ρ_T be the radius of the biggest ball included in T . The mesh regularity $\theta_{\mathcal{M}}$ is defined as

$$\theta_{\mathcal{M}} = \max_{T \in \mathcal{M}} \frac{h_T}{\rho_T}.$$

If $(\mathcal{M}_\ell)_{\ell \in \mathbb{N}}$ is a sequence of triangular meshes such that $h_{\mathcal{M}_\ell} \rightarrow 0$ as $\ell \rightarrow +\infty$, we assume that $\theta_{\mathcal{M}_\ell}$ is dominated by some θ_0 i.e. $\theta_{\mathcal{M}_\ell} \leq \theta_0$ for all $\ell \in \mathbb{N}$.

Set $\llbracket a, b \rrbracket = [a, b] \cap \mathbb{N}$. Let us fix $I_{\mathcal{V}^*} = \llbracket 1, N_{\mathcal{V}^*} \rrbracket$ and $I_D = \llbracket N_{\mathcal{V}^*} + 1, N_{\mathcal{V}} \rrbracket$. Consider the following space of \mathbb{P}_1 finite elements on the mesh \mathcal{M}

$$X_{\mathcal{M}} = \{v_h \in C^0(\bar{\Omega}) \mid v_h|_T \text{ is affine}, \forall T \in \mathcal{M}, \quad v_h(x_i) = 0, \quad \forall i \in I_D\} \subset H_0^1(\Omega).$$

This space admits a basis of shape functions denoted by $(\phi_i)_{i \in I_{\mathcal{V}^*}}$ where $\phi_i(x_j)$ equates 1 if $i = j$ and 0 otherwise. Given $(v_i)_{i \in I_{\mathcal{V}^*}}$, we construct the function v_h as the element of $X_{\mathcal{M}}$ and its corresponding discrete gradient defined as follows

$$v_h = \sum_{i=1}^{N_{\mathcal{V}^*}} v_i \phi_i, \quad \nabla v_h = \sum_{i=1}^{N_{\mathcal{V}^*}} v_i \nabla \phi_i.$$

If g is a nonlinear function from $\mathbb{R} \rightarrow \mathbb{R}$ we make the convention $g((v_i)_{i \in I_{\mathcal{V}^*}}) = (g(v_i))_{i \in I_{\mathcal{V}^*}}$. Then $g(v_h)$ is defined in the same way as v_h , it suffices to replace v_i by $g(v_i)$ in the above definition. It is similar for $\nabla_h g(v_h)$. We underline that $X_{\mathcal{M}}$ can be equipped by the energy norm $\|\nabla_h v_h\|_{L^2(\Omega)^d}$.

It is also possible to define a piecewise constant (finite volume) function using the dual mesh such that $v_{\mathcal{D}} = \sum_{i=1}^{N_{\mathcal{V}^*}} v_i \mathbf{1}_{K_i}$, where $\mathbf{1}_{K_i}$ is the indicator function on K_i .

The transmissibility coefficients are defined by

$$\lambda_{ij}^T = - \int_T \mathbb{A} \nabla \phi_i \cdot \nabla \phi_j \, dx.$$

The sign of λ_{ij}^T is of crucial importance in the stability analysis of the numerical scheme. Let us underline that the physical transmissibility occurs when $\lambda_{ij}^T \geq 0$. For instance, this situation holds true if \mathbb{A} is scalar and all the triangle angles of the triangulation satisfy Delaunay's condition [13]. On the other hand, when some λ_{ij}^T are negative, the maximum principle for parabolic equations, stating the positivity of the solution, is destroyed at the discrete level. Recovering this primordial property without diminishing accuracy is indeed a challenging task.

Using the properties of the shape functions [1, 8], it can be checked that

$$\int_{\Omega} \mathbb{A} \nabla u_h \cdot \nabla v_h \, dx = \sum_{T \in \mathcal{M}} \sum_{v_{ij} \in \mathcal{F}^T} \lambda_{ij}^T (u_i - u_j)(v_i - v_j).$$

Finally, the temporal mesh can be spanned by an increasing sequence $(t^\tau)_{\tau \in [0, n_f]}$ such that

$$t^0 = 0 < t^1 < \dots < t^{n_f} = t_f.$$

This subdivision is supposed to be uniform i.e. the points t^τ with $\tau \in \llbracket 0, n_f \rrbracket$ are equidistant. We will denote by $\delta\tau = t^\tau - t^{\tau-1}$ the time step. The developed scheme and its analysis remain valid in the case of adaptive time steps. We do not handle the variable time stepping to alleviate the notations.

Given now the time-dependent vector $(v_i^\tau)_{i \in I_{\mathcal{V}^*}, \tau \in \llbracket 0, n_f \rrbracket}$, we define

$$v_{h, \delta\tau}(\cdot, t) = v_h^\tau(\cdot), \quad \nabla_h v_{h, \delta\tau}(\cdot, t) = \nabla_h v_h^\tau(\cdot), \quad v_{\mathcal{D}, \delta\tau}(\cdot, t) = v_{\mathcal{D}}^\tau(\cdot), \quad \forall t \in (t^\tau, t^{\tau+1}].$$

4 Generic nonlinear finite volume scheme

The principle of the finite volume methodology is to discretize the integral formulation of (2.1) over the control volumes K_i . Using Green's formula, the task reduces to design appropriate flux approximation across interfaces of every adjacent cells. This strategy allows to propose the subsequent generic numerical scheme.

Firstly, the discrete initial condition is given by the integral mean

$$u_i^0 = \frac{1}{|K_i|} \int_{K_i} u^0(x) dx, \quad \forall i \in I_{\mathcal{V}^*}, \quad u_i^0 = 0, \quad \forall i \in I_D. \quad (4.1)$$

Secondly, we consider an implicit Euler scheme in time to preserve the unconditional stability of the proposed finite volume method in time. Then, our approach consists in computing $\{u_i^\tau\}$ by solving the following discrete nonlinear equations at each time level $\tau \in \llbracket 1, n_f \rrbracket$:

$$u_i^\tau - u_i^{\tau-1} + \frac{\delta\tau}{|K_i|} \sum_{T \in \mathcal{M}_i} \sum_{\nu_{ij} \in \mathcal{F}_i^T} \lambda_{ij}^T \alpha_{ij}^\tau (\beta(u_i^\tau) - \beta(u_j^\tau)) = \delta\tau q(u_i^\tau), \quad \forall i \in I_{\mathcal{V}^*}. \quad (4.2)$$

The Dirichlet boundary condition imposes

$$u_i^\tau = 0, \quad \forall i \in I_D. \quad (4.3)$$

The formula of α_{ij}^τ involves only the unknowns located at the vertices of the element T . This dependency is not highlighted for legibility. We are subsequently discussing some expressions of α_{ij}^τ based on two crucial quests, namely : positivity and accuracy. The pros and cons of each choice are also addressed.

4.1 Centered scheme

There are many ways to define a centered approximation for α_{ij}^τ . A standard possibility is to consider the arithmetic scheme written as

$$\alpha_{ij}^{\tau, C} = \frac{\alpha(u_i^\tau) + \alpha(u_j^\tau)}{2}. \quad (4.4)$$

The centered choice is highly recommended when all the coefficients λ_{ij}^T are nonnegative since the numerical accuracy is optimal, meaning that, quadratic (numerical) rate is generally achieved for the diffusion when the solution is smooth enough. It has been applied to many industrial and practical problems [1, 3, 16, 15, 5]. On the other hand, it is well established that the centered approximation in finite volumes violates the discrete maximum principle on coarse meshes as long as some λ_{ij}^T become negative. We will also see in the numerical section that this centered scheme lead to a divergent behavior of the nonlinear solver in the case of strong anisotropic ratios. This has led to the development of the following alternative.

4.2 Godunov scheme

The Godunov scheme, for nonlinear elliptic terms [8], consists in upwinding the diffusion coefficient α_{ij}^τ with respect to the sign of the transmissibility λ_{ij}^T as follows

$$\alpha_{ij}^{\tau, G} = \begin{cases} \max_{s \in I_{ij}^\tau} \alpha(s) & \text{if } \lambda_{ij}^T \geq 0 \\ \min_{s \in I_{ij}^\tau} \alpha(s) & \text{otherwise} \end{cases}, \quad (4.5)$$

where $I_{ij}^\tau = [\min(u_i^\tau, u_j^\tau), \max(u_i^\tau, u_j^\tau)]$. It induces the numerical diffusion allowing to prevent the formation of possible nonphysical oscillations and then turns the scheme to be positivity-preserving. The expected accuracy is of first order because of the upstreaming. For this reason, it has been applied to advanced models like Keller-Segel system in [10] and to compressible two-phase flow problem in [18]. In case of strong anisotropy, the produced numerical viscosity becomes important and slows drastically the convergence rate. To tackle this issue, a complete modification of the scheme has been proposed in [25]. An extension of the latter method to the context of the VAG (Vertex Approximate Gradient) discretization has been successfully used in [7]. The key idea is to introduce an upwinding with respect to flux-like function. On the other hand, there is still an increasing need to improve the accuracy of the corrected scheme in order to recover the optimal accuracy offered by the centered approach. Before elaborating our new approach to achieve this goal, we would like to indicate the first core hint towards our strategy in the following subsection.

4.3 Sub-upwinding scheme

The main idea of the sub-upwinding approach, already proposed in [19], is to apply the upwind scheme only when it is necessary. In other words, the centered scheme is maintained for $\lambda_{ij}^T \geq 0$ whereas it switches into the Godunov one when $\lambda_{ij}^T < 0$. Then, the chosen diffusion coefficient reads

$$\alpha_{ij}^{\tau,S} = \begin{cases} \alpha_{ij}^{\tau,C} & \text{if } \lambda_{ij}^T \geq 0 \\ \alpha_{ij}^{\tau,G} & \text{if } \lambda_{ij}^T < 0 \end{cases}. \quad (4.6)$$

Also, this strategy yields solutions honoring their physical bounds with an accuracy of first order. Even if the idea seems relevant, the scheme suffers from the numerical diffusion for highly anisotropic ratios. To overcome this challenge, it is mandatory to work on and develop another substitute for $\alpha_{ij}^{\tau,S}$. This is the object of the following discussion.

4.4 Weighted centered-upwinding scheme

In this paper, the main idea is to combine the formal second order accuracy of the centered scheme and the strength of the upwind discretization allowing to preserve positivity. More importantly, the chosen value of the diffusion coefficient should stay closer to a centered approximation when no undershoots occur and eliminate the problematic terms otherwise. A better compromise can be achieved using the following weighted harmonic average

$$\alpha_{ij}^{\tau,W} = \begin{cases} \alpha_{ij}^{\tau,T} & \text{if } \lambda_{ij}^T \geq 0 \\ \frac{(1+\gamma)\alpha_{ij}^{\tau,\min}\alpha_{ij}^{\tau,T}}{\gamma\alpha_{ij}^{\tau,T} + \alpha_{ij}^{\tau,\min}} & \text{if } \lambda_{ij}^T < 0 \end{cases}, \quad (4.7)$$

where

$$\alpha_{ij}^{\tau,T} = \frac{1}{\#\mathcal{V}_T} \sum_{\kappa=1}^{\#\mathcal{V}_T} \alpha(u_\kappa^\tau), \quad \text{and} \quad \alpha_{ij}^{\tau,\min} = \min_{\kappa \in \mathcal{V}_T} \alpha(u_\kappa^\tau).$$

In the case where $\alpha_{ij}^{\tau,T} = 0$, we simply consider $\alpha_{ij}^{\tau,W} = 0$. The parameter γ serves to reduce the influence of the numerical diffusion. It must be small enough so that $\alpha_{ij}^{\tau,W} \approx \alpha_{ij}^{\tau,T}$ holds in the favorable situation ($\alpha_{ij}^{\tau,\min} \neq 0$). Numerically, we can take $\gamma = 10^{-6}$ for instance. Notice that, if $\alpha_{ij}^{\tau,\min} \neq 0$, one has

$$\lim_{\gamma \rightarrow 0} \alpha_{ij}^{\tau,W}(\gamma) = \alpha_{ij}^{\tau,T}.$$

In the numerical section we will provide a full comparison of the aforementioned schemes. We also investigate the impact of the parameter γ , especially when the anisotropic contrast is large. Notice that

γ must be $(0, 1]$ so that one has

$$\alpha_{ij}^{\tau, \min} \leq \alpha_{ij}^{\tau, W} \leq \alpha_{ij}^{\tau, T}.$$

This is an important inequality for the convergence analysis following the arguments of [8].

Remark 4.1. *The function α is assumed to be nondecreasing to simplify the presentation of the scheme and its analysis. This might not cover a wide range of nonlinearities encountered in complex problems such as the capillary terms arising in multi-phase porous media flows [6, 7, 12]. To overcome this issue, it is possible to design an adequate finite volume scheme for general nonlinearities α . Indeed, this requires the introduction of the full Kirchhoff transform which is formally defined as*

$$\zeta(s) = \int_0^s \alpha(z) \beta'(z) \, dz.$$

Then, the equation (2.1) is rewritten as follows [27]

$$u_t - \operatorname{div} \varrho(u) \mathbb{A} \nabla \varrho(u) = q(u),$$

where $\varrho(u) = \sqrt{2\zeta(u)}$ which is increasing. Hence, the corresponding finite volume scheme becomes

$$u_i^\tau - u_i^{\tau-1} + \frac{\delta\tau}{|K_i|} \sum_{T \in \mathcal{M}_i} \sum_{\nu_{ij} \in \mathcal{F}_i^T} \lambda_{ij}^T \varrho_{ij}^\tau \left(\varrho(u_i^\tau) - \varrho(u_j^\tau) \right) = \delta\tau q(u_i^\tau), \quad \forall i \in I_{\mathcal{V}^*}.$$

Up to a slight modification of Assumptions (\mathbf{H}_1) - (\mathbf{H}_2) , the analysis of the next section remains similar and valid except for the energy estimates. More specifically, one needs to select the test function as $\zeta(u)$ in order to derive uniform estimation on the discrete gradient of $\zeta(u)$.

In practice, it is not clear how to explicitly compute ζ and therefore ϱ . A possible way is to use an accurate quadrature rule which does not increase the computational cost. For instance, one can make use of Simpson's rule to evaluate ζ .

5 Analysis of the positive finite volume scheme

In this section, we are only concerned with the novel case $\alpha_{ij}^\tau := \alpha_{ij}^{\tau, W}$. The objective is to show that it still yields a stable, coercive and convergent scheme.

In the rest of the paper, C denotes generic various constants which depend only on the prescribed physical data on possibly and on the mesh regularity, but not on the time and space steps.

5.1 Positivity and energy estimates

To begin with, we prove that the solution remains in its physical interval by construction.

Proposition 5.1. *The nonlinear numerical scheme (4.2)-(4.3) preserves the positivity. In other words, there holds*

$$u_i^\tau \geq 0, \quad \forall i \in I_{\mathcal{V}^*}, \quad \forall \tau \in \llbracket 0, n_f \rrbracket. \quad (5.1)$$

Proof. An inductive argument on τ is used. Take τ such that $1 \leq \tau \leq n_f$ and assume that the property (5.1) holds for $\tau - 1$. On the other hand, we proceed by contradiction in space. For this, select $i_0 \in \llbracket 1, N_{\mathcal{V}^*} \rrbracket$ and set $u_{i_0}^\tau = \min_{1 \leq i \leq N_{\mathcal{V}^*}} u_i^\tau$. Assume that $u_{i_0}^\tau < 0$. Next, multiply the numerical scheme line associated to the index i_0 by $u_{i_0}^\tau$ to obtain

$$(u_{i_0}^\tau - u_{i_0}^{\tau-1}) u_{i_0}^\tau + \frac{\delta\tau}{|K_{i_0}|} \sum_{T \in \mathcal{M}_{i_0}} \sum_{\nu_{i_0j} \in \mathcal{F}_{i_0}^T} \lambda_{i_0j}^T \alpha_{i_0j}^\tau \left(\beta(u_{i_0}^\tau) - \beta(u_j^\tau) \right) u_{i_0}^\tau - \delta\tau q(u_{i_0}^\tau) u_{i_0}^\tau = 0.$$

Observe that $\alpha_{i_0j}^\tau = 0$ for $\lambda_{i_0j}^T < 0$ since $u_{i_0}^\tau < 0$. This is due to the crucial choice (4.7) and the fact that α is extended by 0 on $(-\infty, 0)$. Recall that the source function being nonnegative implies that the last term in the previous identity is also nonnegative. Accordingly, one gets

$$(u_{i_0}^\tau)^2 - u_{i_0}^{\tau-1} u_{i_0}^\tau + \frac{\delta\tau}{|K_{i_0}|} \sum_{T \in \mathcal{M}_{i_0}} \sum_{\nu_{i_0j} \in \mathcal{F}_{i_0}^T} (\lambda_{i_0j}^T)^+ \alpha_{i_0j}^\tau (\beta(u_{i_0}^\tau) - \beta(u_j^\tau)) u_{i_0}^\tau - \delta\tau q(u_{i_0}^\tau) u_{i_0}^\tau = 0.$$

All the terms of the above equality are therefore nonnegative. This forces $u_{i_0}^\tau = 0$, which is in contradiction with the hypothesis $u_{i_0}^\tau < 0$. In conclusion, the scheme does not allow any solution to be lower than 0. \square

We next require the following result accounting for the equivalence of some important discrete norms. Its proof can be documented in [6, 8].

Lemma 5.1. *There exists C_e depending only on $\mathbb{A}_0, \mathbb{A}_1$ and the mesh regularity such that*

$$\sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| (v_i - v_j)^2 \leq C_e \sum_{\nu_{ij} \in \mathcal{F}^T} \lambda_{ij}^T (v_i - v_j)^2, \quad \forall T \in \mathcal{M}. \quad (5.2)$$

The second theoretical claim states that the proposed positive finite volume methodology maintains some fundamental a priori estimates referred to as coercivity. We investigate this in the following result.

Proposition 5.2. *There exists a positive constant C depending only on the data and on the geometrical regularity θ_0 such that the following energy estimates hold true*

$$\sum_{\tau=1}^{n_f} \delta\tau \|\nabla_h \xi(u_h^\tau)\|_{L^2(\Omega)^d}^2 \leq C, \quad (5.3)$$

$$\sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| \alpha_{ij}^\tau (\beta(u_i^\tau) - \beta(u_j^\tau))^2 \leq C. \quad (5.4)$$

Proof. Multiplying the discrete equation (4.2) by $|K_i| \beta(u_i^\tau)$, summing on all $i \in \mathcal{V}^*$ and τ gives

$$\mathfrak{S}_1 + \mathfrak{S}_2 = \mathfrak{S}_3,$$

where

$$\begin{aligned} \mathfrak{S}_1 &= \sum_{\tau=1}^{n_f} \sum_{i \in \mathcal{V}^*} |K_i| (u_i^\tau - u_i^{\tau-1}) \beta(u_i^\tau), \\ \mathfrak{S}_2 &= \sum_{\tau=1}^{n_f} \delta\tau \sum_{i \in \mathcal{V}^*} \sum_{T \in \mathcal{M}_i} \sum_{\nu_{ij} \in \mathcal{F}_i^T} \lambda_{ij}^T \alpha_{ij}^\tau (\beta(u_i^\tau) - \beta(u_j^\tau)) \beta(u_i^\tau), \\ \mathfrak{S}_3 &= \sum_{\tau=1}^{n_f} \delta\tau \sum_{i \in \mathcal{V}^*} |K_i| q(u_i^\tau) \beta(u_i^\tau). \end{aligned}$$

By the convexity of Υ , the accumulation term is greater than a telescopic series yielding

$$\mathfrak{S}_1 \geq \sum_{i \in \mathcal{V}^*} |K_i| (\Upsilon(u_i^{n_f}) - \Upsilon(u_i^0)) \geq -\|\Upsilon(u^0)\|_{L^1(\Omega)},$$

thanks to Jensen's inequality. Owing to the discrete integration-by-parts in space, we can reorder the summations of \mathfrak{S}_2 by elements and dual edges as follows

$$\mathfrak{S}_2 = \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} \lambda_{ij}^T \alpha_{ij}^\tau (\beta(u_i^\tau) - \beta(u_j^\tau))^2.$$

As a consequence of the definition of α_{ij}^τ given in (4.7), one has

$$\alpha_{ij}^\tau = \alpha_{ij}^{\tau,T}, \quad \forall \lambda_{ij}^T \geq 0, \quad \text{and} \quad \alpha_{ij}^\tau \leq \alpha_{ij}^{\tau,T}, \quad \forall \lambda_{ij}^T < 0. \quad (5.5)$$

Applying Lemma 5.1 implies

$$\begin{aligned} \mathfrak{S}_2 &\geq \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \alpha_{ij}^{\tau,T} \sum_{\nu_{ij} \in \mathcal{F}^T} \lambda_{ij}^T \left(\beta(u_i^\tau) - \beta(u_j^\tau) \right)^2 \\ &\geq \frac{1}{C_e} \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \alpha_{ij}^{\tau,T} \sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| \left(\beta(u_i^\tau) - \beta(u_j^\tau) \right)^2. \end{aligned}$$

Let us set

$$\tilde{\alpha}_{ij}^\tau = \begin{cases} \frac{\xi(u_i^\tau) - \xi(u_j^\tau)}{\beta(u_i^\tau) - \beta(u_j^\tau)} & \text{if } u_i^\tau \neq u_j^\tau \\ \sqrt{\alpha(u_i^\tau)} & \text{else} \end{cases}.$$

Notice that

$$\alpha_{ij}^{\tau,T} \geq \frac{1}{3} (\tilde{\alpha}_{ij}^\tau)^2, \quad \forall \nu_{ij} \in \mathcal{F}^T.$$

Using this inequality together with the ellipticity of \mathbb{A} we find

$$\mathfrak{S}_2 \geq \frac{1}{3C_e} \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} \lambda_{ij}^T (\xi(u_i^\tau) - \xi(u_j^\tau))^2 \geq \frac{\mathbb{A}_0}{3C_e} \sum_{\tau=1}^{n_f} \delta\tau \|\nabla_h \xi(u_h^\tau)\|_{L^2(\Omega)^d}^2.$$

Bear in mind that $\xi(s) \geq C_\xi \beta(s) - C'_\xi$. Next, introduce the fact that q is bounded to deduce

$$\begin{aligned} \mathfrak{S}_3 &\leq \left(\sum_{\tau=1}^{n_f} \delta\tau \sum_{i \in \mathcal{V}^*} |K_i| q(u_i^\tau)^2 \right)^{1/2} \left(\sum_{\tau=1}^{n_f} \delta\tau \sum_{i \in \mathcal{V}^*} |K_i| \beta(u_i^\tau)^2 \right)^{1/2} \\ &\leq C + \frac{\mathbb{A}_0}{6C_e C_p} \sum_{\tau=1}^{n_f} \delta\tau \sum_{i \in \mathcal{V}^*} |K_i| \xi(u_i^\tau)^2 \leq C + \frac{\mathbb{A}_0}{6C_e} \sum_{\tau=1}^{n_f} \delta\tau \|\nabla_h \xi(u_h^\tau)\|_{L^2(\Omega)^d}^2, \end{aligned}$$

where we employed the Cauchy-Schwarz, Young and Poincaré inequalities to acquire the last result. In conclusion, we arrive at the estimate

$$\begin{aligned} \frac{\mathbb{A}_0}{6C_e} \sum_{\tau=1}^{n_f} \delta\tau \|\nabla_h \xi(u_h^\tau)\|_{L^2(\Omega)^d}^2 &\leq \frac{1}{C_e} \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| \alpha_{ij}^{\tau,T} \left(\beta(u_i^\tau) - \beta(u_j^\tau) \right)^2 \\ &\leq \mathfrak{S}_2 \leq C + \|\Upsilon(u^0)\|_{L^1(\Omega)}. \end{aligned}$$

This allows to see that (5.3) is satisfied. We also infer from (5.5) that

$$\begin{aligned} &\sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| \alpha_{ij}^{\tau,T} \left(\beta(u_i^\tau) - \beta(u_j^\tau) \right)^2 \\ &\leq \sum_{\tau=1}^{n_f} \delta\tau \sum_{T \in \mathcal{M}} \sum_{\nu_{ij} \in \mathcal{F}^T} |\lambda_{ij}^T| \alpha_{ij}^{\tau,T} \left(\beta(u_i^\tau) - \beta(u_j^\tau) \right)^2 \leq C, \end{aligned}$$

which ensures the validity of (5.4). The proof is complete. \square

5.2 Existence result

In the following proposition we show that the nonlinear algebraic system resulting from the numerical scheme is solvable.

Proposition 5.3. *The nonlinear finite volume scheme (4.2)-(4.3) has a nonnegative solution.*

Proof. The proof is based on an inductive argument together with a fixed point method. Define the continuous vector field $\Gamma : \mathbb{R}^{N_{\mathcal{V}^*}} \rightarrow \mathbb{R}^{N_{\mathcal{V}^*}}$ such that, for all $i \in \llbracket 1, N_{\mathcal{V}^*} \rrbracket$:

$$\left(\Gamma(u_{\mathcal{V}^*}^\tau)\right)_i = u_i^\tau - u_i^{\tau-1} + \frac{\delta\tau}{|K_i|} \sum_{T \in \mathcal{M}_i} \sum_{\nu_{ij} \in \mathcal{F}_i^T} \lambda_{ij}^T \alpha_{ij}^\tau \left(\beta(u_i^\tau) - \beta(u_j^\tau)\right) - \delta\tau q(u_i^\tau).$$

The vector $u_{\mathcal{V}^*}^{\tau-1}$ is assumed to be known and we are looking for $u_{\mathcal{V}^*}^\tau$. If $j \in \llbracket N_{\mathcal{V}^*} + 1, N_{\mathcal{V}} \rrbracket$, the boundary condition $u_j^\tau = 0$ is imposed. Consider $\Phi : \mathbb{R}^{N_{\mathcal{V}^*}} \rightarrow \mathbb{R}^{N_{\mathcal{V}^*}}$ the homeomorphism mapping $u_{\mathcal{V}^*}^\tau$ to $\beta^{-1}(u_{\mathcal{V}^*}^\tau)$. It is well defined since β is continuous and strictly increasing. Let us set $\mathcal{B} = \Gamma \circ \Phi$ and $z_{\mathcal{V}^*}^\tau = \beta(u_{\mathcal{V}^*}^\tau)$.

So, the numerical scheme admits a solution if and only if the equation $\mathcal{B}(z_{\mathcal{V}^*}^\tau) = \Gamma(u_{\mathcal{V}^*}^\tau) = 0$ is solvable. To achieve this aim, we make use of the monotony criterion [14, Lemma of page 493]. In other words, it is sufficient to establish that

$$\mathcal{B}(z_{\mathcal{V}^*}^\tau) \cdot z_{\mathcal{V}^*}^\tau \geq 0, \quad \forall \|z_{\mathcal{V}^*}^\tau\|_{\mathbb{R}^{N_{\mathcal{V}^*}}} = r, \quad (5.6)$$

for some $r > 0$ to be determined. To this end, we reproduce the proof of Proposition 5.2 entailing

$$\begin{aligned} \mathcal{B}(z_{\mathcal{V}^*}^\tau) \cdot z_{\mathcal{V}^*}^\tau &= \Gamma(u_{\mathcal{V}^*}^\tau) \cdot \beta(u_{\mathcal{V}^*}^\tau) \\ &\geq \frac{1}{\delta\tau} \sum_{i \in \mathcal{V}^*} |K_i| (\Upsilon(u_i^\tau) - \Upsilon(u_i^{\tau-1})) + C_1 \|\nabla_h \xi(u_h^\tau)\|_{L^2(\Omega)^d}^2 - C_2. \end{aligned}$$

Next, introduce Poincaré's inequality, the fact that $\xi(s) \geq C_\xi \beta(s) - C'_\xi$, and the equivalence of discrete norms on $\mathbb{R}^{N_{\mathcal{V}^*}}$ to deduce that

$$\mathcal{B}(z_{\mathcal{V}^*}^\tau) \cdot z_{\mathcal{V}^*}^\tau \geq -\frac{1}{\delta\tau} \sum_{i \in \mathcal{V}^*} |K_i| \Upsilon(u_i^{\tau-1}) + C_{N_{\mathcal{V}^*}} \|z_{\mathcal{V}^*}^\tau\|_{\mathbb{R}^{N_{\mathcal{V}^*}}}^2 - C_3.$$

Then, in order to meet the requirement (5.6), the sought $r = \|z_{\mathcal{V}^*}^\tau\|_{\mathbb{R}^{N_{\mathcal{V}^*}}}$ should satisfy

$$r^2 > \frac{1}{C_{N_{\mathcal{V}^*}}} \left(C_3 + \frac{1}{\delta\tau} \sum_{i \in \mathcal{V}^*} |K_i| \Upsilon(u_i^{\tau-1}) \right),$$

which is always possible since $C_3, C_{N_{\mathcal{V}^*}}, \delta\tau$ and $u_i^{\tau-1}$ are already known. This enables the existence of at least one solution to the nonlinear finite volume scheme (4.2)-(4.3). Whence, the proof is concluded. \square

5.3 Convergence

The convergence proof of the proposed numerical scheme is an adaptation of the ones elaborated in [10, 25, 27]. The following Lemma is essential, it allows to apply the compactness criterion in time [4].

Proposition 5.4. *Consider $\varphi \in \mathcal{C}_c^\infty(\Omega)$. Define $\varphi_i = \int_{K_i} \varphi(x) dx / |K_i|$. There exists C depending only on the data such that*

$$\sum_{\tau=1}^{n_f} \sum_{i \in \mathcal{V}^*} |K_i| (u_i^\tau - u_i^{\tau-1}) \varphi_i \leq C \|\nabla \varphi\|_{L^\infty(\Omega)^d}.$$

The following last result states that the considered scheme is again convergent.

Proposition 5.5. *Let $(\mathcal{M}_\ell)_\ell$ be a sequence of the discretizations to Ω_{t_f} where their mesh regularity is bounded. Let $u_{h,\delta\tau}$ be the corresponding sequence of discrete solutions to the finite volume scheme. Up to a subsequence, there holds*

$$u_{\mathcal{D}_\ell, \delta\tau_\ell}, u_{h_\ell, \delta\tau_\ell} \longrightarrow u \quad \text{a.e. in } \Omega_{t_f} \text{ and strongly in } L^1(\Omega_{t_f}), \quad (5.7)$$

$$\nabla_{h_\ell} \xi(u_{h_\ell, \delta\tau_\ell}) \longrightarrow \nabla \xi(u) \quad \text{weakly } L^2(\Omega_{t_f})^d. \quad (5.8)$$

Furthermore, u is a weak solution to the continuous problem (2.1)-(2.3) in the sense of Definition 2.1.

6 Numerical validation

This section aims to validate the theoretical findings with $d = 2$, especially the good efficiency and the robustness of our novel approach using numerical evidences.

The domain of study Ω is the unit square i.e. $\Omega = (0, 1) \times (0, 1)$. It is discretized using a non-structured series of refined triangulations of the benchmark FVCA5 [20] on the anisotropic diffusion problems. The first two elements of this sequence are depicted on Fig. 2.

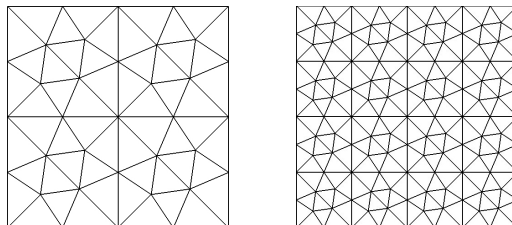


Figure 2: Illustration of the first two elements of the meshed domain.

We consider a diagonal anisotropic tensor under the form

$$\mathbb{A} = \begin{pmatrix} l_x & 0 \\ 0 & l_y \end{pmatrix}.$$

This choice is made in order to check analytical solutions in a straightforward way.

At each time step, the finite volume scheme (4.1)-(4.2) is formulated in a nonlinear algebraic set of equations solved thanks to the Newton-Raphson method. The embedded linear system is solved owing to a direct solver. The Newton algorithm is convergent if the ℓ^2 -norm on the relative residual is lower than 10^{-6} . To deal with the different issues that can be resulting from the degeneracy and the strong anisotropy we employ an adaptive time stepping. Indeed, given initial and thresholding time steps denoted by $\delta\tau_{\text{ini}}$ and $\delta\tau_{\text{max}}$ respectively, we compute

$$\delta\tau^{\tau+1} = \max(\delta\tau_{\text{max}}, 1.2\delta\tau^\tau), \quad \forall \tau \geq 1,$$

in case of the Newton algorithm converges with a number of iterations less than 25 iterations. Otherwise, the time iteration is recalculated by setting $\delta\tau^{\tau+1} = \delta\tau^\tau/2$. We denote by “#Dt chop” the number of time step chops and by “#Newton” the total number of successful Newton iterations. These indicators are considered to quantify the behavior of the nonlinear solver. Note that a maximum number of “#Dt chop” is fixed to 100 in case of the scheme diverges. In our numerical tests, we have fixed $\delta\tau_{\text{ini}} = \delta\tau_{\text{max}} \approx 0.16h^2$.

We will evaluate the accuracy of the finite volume scheme. Then, the errors are measured in the usual discrete norms

$$\text{err}_{L^r} = \|u_{\text{ex}} - u_{\mathcal{M}, \delta\tau}\|_{L^r(\Omega_{t_f})}, \quad r = 1, 2, \infty.$$

All the obtained numerical results are computed using an in-house code developed in Fortran.

6.1 Heat problem

In this subsection, we show that our framework can provide a good approximation to the well-known heat equation. In order to solve the standard linear heat equation we perform a nonlinear transformation. To this purpose, we consider the nonlinearities

$$\alpha(u) = \beta(u) = \sqrt{2u}, \quad q(u) = 0.$$

In this case, the equation (2.1) leads to $u_t - \operatorname{div} \mathbb{A} \nabla u = 0$. To compute the numerical convergence, we use, as in [8], the following analytical solution,

$$\tilde{u}((x, y), \tau) = \frac{1 + \cos(\pi x) \exp(-l_x \pi^2 \tau)}{2},$$

for $(x, y) \in \Omega$, $\tau \in (0, t_f)$ where the final time is set to $t_f = 0.2$. Zero-flux boundary conditions are imposed on $\partial\Omega \times (0, t_f)$. The tensor \mathbb{A} is anisotropic with $l_x = 1$ and l_y is to be chosen below.

6.1.1 Influence of the parameter γ

The aim here is to study numerically the influence of $\gamma \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ in the new weighted centered-upwinding scheme (4.7). We impose the anisotropy at $l_y = 100$. The numerical behavior of the scheme is listed in the Tables 1-5. We can observe that, except for $\gamma = 10^{-2}$, the results are close and order 2 is found as expected.

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{\min}	u_{\max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.758E-02	-	0.283E-02	-	0.521E-01	-	0.037	0.999	0	77
0.125	0.00256	0.00256	0.223E-02	1.765	0.767E-03	1.882	0.157E-01	1.728	0.009	1.001	0	252
0.062	0.00064	0.00064	0.673E-03	1.728	0.232E-03	1.722	0.453E-02	1.796	0.002	1.000	0	953
0.031	0.00016	0.00016	0.236E-03	1.513	0.859E-04	1.437	0.162E-02	1.481	0.001	1.000	0	3763
0.016	0.00004	0.00004	0.976E-04	1.273	0.370E-04	1.216	0.612E-03	1.406	0.000	1.000	0	11655

Table 1: Heat problem, $l_y = 100$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-2}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{\min}	u_{\max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.729E-02	-	0.273E-02	-	0.534E-01	-	0.033	1.000	0	77
0.125	0.00256	0.00256	0.198E-02	1.882	0.704E-03	1.955	0.158E-01	1.752	0.008	1.001	0	252
0.062	0.00064	0.00064	0.505E-03	1.970	0.178E-03	1.985	0.427E-02	1.893	0.002	1.000	0	952
0.031	0.00016	0.00016	0.129E-03	1.965	0.453E-04	1.974	0.110E-02	1.956	0.000	1.000	0	3762
0.016	0.00004	0.00004	0.342E-04	1.919	0.118E-04	1.941	0.281E-03	1.968	0.000	1.000	0	11657

Table 2: Heat problem, $l_y = 100$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-3}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{\min}	u_{\max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.727E-02	-	0.272E-02	-	0.535E-01	-	0.032	1.000	0	75
0.125	0.00256	0.00256	0.196E-02	1.890	0.699E-03	1.961	0.159E-01	1.755	0.008	1.001	0	250
0.062	0.00064	0.00064	0.495E-03	1.985	0.175E-03	1.999	0.427E-02	1.894	0.002	1.000	0	950
0.031	0.00016	0.00016	0.124E-03	1.996	0.437E-04	2.000	0.110E-02	1.956	0.000	1.000	0	3761
0.016	0.00004	0.00004	0.311E-04	1.995	0.110E-04	1.995	0.281E-03	1.968	0.000	1.000	0	11657

Table 3: Heat problem, $l_y = 100$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-4}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.726E-02	-	0.272E-02	-	0.535E-01	-	0.032	1.000	0	73
0.125	0.00256	0.00256	0.196E-02	1.891	0.698E-03	1.962	0.159E-01	1.755	0.008	1.001	0	249
0.062	0.00064	0.00064	0.494E-03	1.986	0.174E-03	2.001	0.427E-02	1.894	0.002	1.000	0	949
0.031	0.00016	0.00016	0.124E-03	1.999	0.435E-04	2.002	0.110E-02	1.956	0.000	1.000	0	3760
0.016	0.00004	0.00004	0.309E-04	2.000	0.109E-04	2.000	0.281E-03	1.968	0.000	1.000	0	11656

Table 4: Heat problem, $l_y = 100$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-5}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.726E-02	-	0.272E-02	-	0.535E-01	-	0.032	1.000	0	72
0.125	0.00256	0.00256	0.196E-02	1.891	0.698E-03	1.962	0.159E-01	1.755	0.008	1.001	0	248
0.062	0.00064	0.00064	0.494E-03	1.986	0.174E-03	2.001	0.427E-02	1.894	0.002	1.000	0	948
0.031	0.00016	0.00016	0.124E-03	1.999	0.435E-04	2.003	0.110E-02	1.956	0.000	1.000	0	3759
0.016	0.00004	0.00004	0.309E-04	2.000	0.109E-04	2.001	0.281E-03	1.968	0.000	1.000	0	11656

Table 5: Heat problem, $l_y = 100$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$

6.1.2 Comparison of the schemes

The aim here is to compare the schemes defined by (4.4), (4.5), (4.6) and (4.7) for $\gamma = 10^{-6}$. We impose the anisotropy as $l_y = 1000$. But the centered scheme (4.4) does not converge (the Newton's algorithm diverges) on this case even if we divide by 10 the $\delta\tau_{\text{init}}$. The numerical results are listed in Tables 6-8. We can underline clearly that the new scheme (4.7) has the smallest errors and the fastest convergence rates (close to 2). Its number of Newton iterations is slightly bigger but only on the finest mesh. For all the schemes, no time step chops have been recorded. Furthermore, the positivity of the solution is fulfilled as already proved in Lemma 5.1. This is no longer the case of the linear schemes for the heat equation where oscillations can be observed using strongly anisotropic tensors, see for instance [8, 9].

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.104E-01	-	0.387E-02	-	0.751E-01	-	0.027	1.027	0	82
0.125	0.00256	0.00256	0.272E-02	1.932	0.963E-03	2.006	0.204E-01	1.881	0.006	1.008	0	258
0.062	0.00064	0.00064	0.683E-03	1.997	0.239E-03	2.009	0.534E-02	1.933	0.001	1.002	0	957
0.031	0.00016	0.00016	0.170E-03	2.002	0.596E-04	2.005	0.136E-02	1.975	0.000	1.001	0	3767
0.016	0.00004	0.00004	0.426E-04	2.000	0.149E-04	2.001	0.342E-03	1.988	0.000	1.000	0	15015

Table 6: Heat problem, $l_y = 1000$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.615E-01	-	0.221E-01	-	0.294E+00	-	0.288	0.736	0	80
0.125	0.00256	0.00256	0.603E-01	0.028	0.221E-01	-0.005	0.293E+00	0.005	0.141	0.925	0	255
0.062	0.00064	0.00064	0.537E-01	0.167	0.200E-01	0.145	0.256E+00	0.194	0.047	0.998	0	953
0.031	0.00016	0.00016	0.442E-01	0.282	0.167E-01	0.261	0.204E+00	0.331	0.013	1.000	0	3763
0.016	0.00004	0.00004	0.334E-01	0.402	0.128E-01	0.383	0.150E+00	0.439	0.003	1.000	0	12044

Table 7: Heat problem, $l_y = 1000$ - Sub-upwinding scheme (4.6)

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.643E-01	-	0.229E-01	-	0.317E+00	-	0.319	0.699	0	80
0.125	0.00256	0.00256	0.642E-01	0.002	0.234E-01	-0.031	0.322E+00	-0.023	0.171	0.888	0	255
0.062	0.00064	0.00064	0.590E-01	0.121	0.218E-01	0.100	0.289E+00	0.153	0.061	0.994	0	954
0.031	0.00016	0.00016	0.507E-01	0.220	0.190E-01	0.199	0.239E+00	0.273	0.017	1.000	0	3764
0.016	0.00004	0.00004	0.404E-01	0.326	0.154E-01	0.306	0.185E+00	0.375	0.004	1.000	0	11736

Table 8: Heat problem, $l_y = 1000$ - Godunov scheme (4.5)

6.1.3 Influence of the anisotropy.

To conclude this first test case, we focus here on the influence of the anisotropy on the weighted centered-upwinding scheme (4.7). We take $\gamma = 10^{-6}$ and $l_y \in \{0.1, 10, 100, 1000\}$. The values $l_y = 100$ and $l_y = 1000$ have been already done, see Tables 5 and 6. See Table 9 for $l_y = 0.1$ and Table 10 for $l_y = 10$. These Tables exhibit the ability of the new scheme to successfully deal with several ratios of anisotropy without a huge impact on its numerical behavior (convergence rate and computational cost).

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.334E-02	-	0.125E-02	-	0.165E-01	-	0.001	0.999	6	115
0.125	0.00256	0.00256	0.986E-03	1.762	0.373E-03	1.738	0.474E-02	1.797	0.000	1.000	5	279
0.062	0.00064	0.00064	0.259E-03	1.929	0.986E-04	1.921	0.137E-02	1.794	0.000	1.000	4	774
0.031	0.00016	0.00016	0.656E-04	1.981	0.251E-04	1.976	0.373E-03	1.876	0.000	1.000	3	2604
0.016	0.00004	0.00004	0.165E-04	1.995	0.629E-05	1.994	0.999E-04	1.898	0.000	1.000	2	10076

Table 9: Heat problem, $l_y = 0.1$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.300E-02	-	0.941E-03	-	0.237E-01	-	0.039	0.969	0	65
0.125	0.00256	0.00256	0.787E-03	1.931	0.267E-03	1.818	0.650E-02	1.864	0.010	0.992	0	242
0.062	0.00064	0.00064	0.199E-03	1.981	0.685E-04	1.961	0.169E-02	1.948	0.003	0.998	0	942
0.031	0.00016	0.00016	0.500E-04	1.996	0.172E-04	1.990	0.437E-03	1.949	0.001	0.999	0	3056
0.016	0.00004	0.00004	0.125E-04	1.999	0.432E-05	1.998	0.113E-03	1.949	0.000	1.000	0	10377

Table 10: Heat problem, $l_y = 10$ - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$

6.2 Porous medium equation

In this subsection, as in [9], we consider the case of the anisotropic porous medium equation

$$u_t - \operatorname{div} \mathbb{A} \nabla(u^2) = 0,$$

which can be obtained from (2.1) by taking,

$$\alpha(u) = 2u, \quad \beta(u) = u, \quad q(u) = 0.$$

The problem is closed with Dirichlet boundary conditions on $\partial\Omega \times (0, t_f)$. Two analytical solutions have been tested. Once again, we stress that the centered scheme (4.4) diverges on the both cases.

6.2.1 One-dimensional analytical solution

The numerical convergence of the three schemes has first been compared thanks to the following 1d-analytical solution

$$\hat{u}((x, y), \tau) = \max(2l_x\tau - x, 0),$$

for $(x, y) \in \Omega$, $\tau \in (0, t_f)$ where the final time is set to $t_f = 0.25$. The anisotropic tensor is given by $l_x = 1$ and $l_y = 100$. The results are listed in Tables 11-13. The lack of regularity of the exact solution

impacts clearly the numerical behavior of the solver. By the way, in comparison with the two other schemes, the main advantages consist in the fact that the new scheme exhibits the smallest errors and its convergence rates are more than twice bigger. Its drawback is to produce several cuts of the time step and more Newton iterations, which relatively involves an expensive computational cost.

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.475E-02	-	0.787E-03	-	0.557E-01	-	0.000	0.500	6	186
0.125	0.00256	0.00256	0.202E-02	1.235	0.290E-03	1.443	0.309E-01	0.851	0.000	0.500	16	618
0.062	0.00064	0.00064	0.890E-03	1.180	0.111E-03	1.378	0.162E-01	0.928	0.000	0.500	25	1786
0.031	0.00016	0.00016	0.374E-03	1.252	0.370E-04	1.592	0.916E-02	0.825	0.000	0.500	43	5641
0.016	0.00004	0.00004	0.148E-03	1.337	0.111E-04	1.733	0.470E-02	0.964	0.000	0.500	70	19654

Table 11: PME 1d - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-3}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.162E-01	-	0.465E-02	-	0.107E+00	-	0.000	0.500	0	103
0.125	0.00256	0.00256	0.113E-01	0.519	0.312E-02	0.577	0.877E-01	0.287	0.000	0.500	0	359
0.062	0.00064	0.00064	0.798E-02	0.504	0.200E-02	0.643	0.679E-01	0.369	0.000	0.500	0	1206
0.031	0.00016	0.00016	0.557E-02	0.519	0.124E-02	0.690	0.527E-01	0.365	0.000	0.500	0	4704
0.016	0.00004	0.00004	0.382E-02	0.543	0.748E-03	0.726	0.406E-01	0.376	0.000	0.500	0	18762

Table 12: PME 1d - Sub-upwinding scheme (4.6)

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.159E-01	-	0.469E-02	-	0.110E+00	-	0.000	0.500	0	102
0.125	0.00256	0.00256	0.122E-01	0.385	0.350E-02	0.423	0.933E-01	0.243	0.000	0.500	0	350
0.062	0.00064	0.00064	0.918E-02	0.409	0.243E-02	0.527	0.751E-01	0.314	0.000	0.500	0	1186
0.031	0.00016	0.00016	0.672E-02	0.452	0.160E-02	0.604	0.600E-01	0.325	0.000	0.500	0	4694
0.016	0.00004	0.00004	0.477E-02	0.493	0.101E-02	0.664	0.472E-01	0.344	0.000	0.500	0	18750

Table 13: PME 1d - Godunov scheme (4.5)

6.2.2 Two-dimensional analytical solution

The numerical convergence of the three schemes has secondly been compared thanks to the following 2d-analytical solution

$$\hat{u}((x, y), \tau) = \frac{1}{1 - \tau} \left(\frac{1}{16l_x} (x - 0.5)^2 + \frac{1}{16l_y} (y - 0.5)^2 \right).$$

for $(x, y) \in \Omega$, $\tau \in (0, t_f)$ where the final time is set to $t_f = 0.2$. The anisotropic tensor is given by $l_x = 0.01$ and $l_y = 10$. This solution is more regular than the previous one. The results are gathered in Tables 14-16. The new scheme is, by far, the more accurate and its convergence rates tend to 2 but it comes with an additional cost in terms of time, and Newton iterations.

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.203E-01	-	0.570E-02	-	0.107E+00	-	0.000	1.955	22	285
0.125	0.00256	0.00256	0.107E-01	0.927	0.379E-02	0.589	0.540E-01	0.979	0.000	1.955	27	523
0.062	0.00064	0.00064	0.399E-02	1.419	0.154E-02	1.302	0.178E-01	1.599	0.000	1.955	49	1281
0.031	0.00016	0.00016	0.117E-02	1.772	0.479E-03	1.684	0.512E-02	1.801	0.000	1.955	70	3178
0.016	0.00004	0.00004	0.312E-03	1.906	0.133E-03	1.852	0.137E-02	1.906	0.000	1.955	52	10480

Table 14: PME 2d - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.110E+00	-	0.398E-01	-	0.423E+00	-	0.002	1.955	0	67
0.125	0.00256	0.00256	0.699E-01	0.652	0.268E-01	0.568	0.284E+00	0.574	0.001	1.955	0	246
0.062	0.00064	0.00064	0.428E-01	0.708	0.165E-01	0.700	0.187E+00	0.600	0.000	1.955	0	790
0.031	0.00016	0.00016	0.249E-01	0.779	0.951E-02	0.796	0.117E+00	0.684	0.000	1.955	0	2989
0.016	0.00004	0.00004	0.139E-01	0.843	0.522E-02	0.866	0.718E-01	0.700	0.000	1.955	0	10096

Table 15: PME 2d - Sub-upwinding scheme (4.6)

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.113E+00	-	0.408E-01	-	0.450E+00	-	0.002	1.955	0	65
0.125	0.00256	0.00256	0.810E-01	0.478	0.310E-01	0.395	0.331E+00	0.445	0.002	1.955	0	245
0.062	0.00064	0.00064	0.544E-01	0.574	0.210E-01	0.561	0.231E+00	0.515	0.000	1.955	0	761
0.031	0.00016	0.00016	0.340E-01	0.680	0.130E-01	0.692	0.151E+00	0.619	0.000	1.955	0	2917
0.016	0.00004	0.00004	0.199E-01	0.771	0.752E-02	0.791	0.960E-01	0.650	0.000	1.955	0	10178

Table 16: PME 2d - Godunov scheme (4.5)

6.3 Case of convection and diffusion

We add a nonlinear convection term to the nonlinear diffusion model problem. In this way, the equation (2.1) becomes

$$u_t + \text{div } c(u)\mathbf{v} - \text{div } \alpha(u)\mathbb{A}\nabla\beta(u) = q(u).$$

We assume that $\text{div } \mathbf{v} = 0$. In order to study the numerical convergence behaviors of the studied schemes, we present an original test case. We consider the following data

$$\alpha(u) = 2u, \quad \beta(u) = u, \quad q(u) = 0, \quad \mathbb{A} = \begin{pmatrix} 1 & 0 \\ 0 & l_y \end{pmatrix}, \quad c(u) = u^{3/2}, \quad \mathbf{v}(\tau) = \begin{pmatrix} -1 \\ \sqrt{9(1-\tau)} \\ 0 \end{pmatrix}.$$

The value of l_y will be specified in the numerical tests. A 1d-analytical solution can be exhibited

$$\bar{u}((x, y), \tau) = \frac{x^2}{9(1-\tau)},$$

for $(x, y) \in \Omega$, $\tau \in (0, t_f)$ where the final time is set to $t_f = 0.2$. The problem is closed with Dirichlet boundary conditions. At the discrete level, we approach the convection term thanks to the upwind scheme or the centered scheme.

6.3.1 Isotropic case

To begin we can notice that, in this isotropic case, the Sub-upwinding scheme (4.6) and the Centered scheme (4.4) are equivalents since the transmissibility coefficients are nonnegative. We compare the behavior of the several schemes introduced in this work for the diffusion term with, for the convection term, firstly the upwind scheme (see Tables 17-19), and secondly the centered scheme (see Tables 20-22). We can notice that, when the upwind scheme is used for the convection, all the schemes are of order 1, which is consistent with the upwinding. But the Godunov scheme produces significantly larger errors. On the other hand, when the centered scheme is used for the convection, all the schemes tend to the order 2, except the Godunov scheme. And we can observe that the new scheme produces the smallest errors without additional numerical cost.

- Upwinding convection

h	$\delta\tau_{\text{mit}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.185E-03	-	0.574E-04	-	0.107E-02	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.833E-04	1.150	0.274E-04	1.064	0.575E-03	0.891	0.000	0.139	0	236
0.062	0.00064	0.00064	0.392E-04	1.088	0.132E-04	1.060	0.261E-03	1.138	0.000	0.139	0	631
0.031	0.00016	0.00016	0.190E-04	1.047	0.639E-05	1.042	0.127E-03	1.038	0.000	0.139	0	2500
0.016	0.00004	0.00004	0.933E-05	1.024	0.314E-05	1.024	0.623E-04	1.031	0.000	0.139	0	10000

Table 17: Isotropic convection diffusion. Upwinding convection - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$, for diffusion

h	$\delta\tau_{\text{mit}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.240E-03	-	0.692E-04	-	0.159E-02	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.104E-03	1.206	0.322E-04	1.103	0.726E-03	1.131	0.000	0.139	0	236
0.062	0.00064	0.00064	0.442E-04	1.236	0.146E-04	1.137	0.311E-03	1.223	0.000	0.139	0	644
0.031	0.00016	0.00016	0.200E-04	1.145	0.678E-05	1.111	0.139E-03	1.166	0.000	0.139	0	2507
0.016	0.00004	0.00004	0.953E-05	1.067	0.324E-05	1.066	0.653E-04	1.088	0.000	0.139	0	10000

Table 18: Isotropic convection diffusion. Upwinding convection - Sub-upwinding scheme (4.6) for diffusion

h	$\delta\tau_{\text{mit}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.570E-03	-	0.178E-03	-	0.348E-02	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.279E-03	1.028	0.921E-04	0.948	0.192E-02	0.861	0.000	0.139	0	236
0.062	0.00064	0.00064	0.137E-03	1.027	0.461E-04	0.998	0.930E-03	1.044	0.000	0.139	0	644
0.031	0.00016	0.00016	0.679E-04	1.015	0.229E-04	1.009	0.457E-03	1.024	0.000	0.139	0	2507
0.016	0.00004	0.00004	0.338E-04	1.007	0.114E-04	1.008	0.226E-03	1.014	0.000	0.139	0	10000

Table 19: Isotropic convection diffusion. Upwinding convection - Godunov scheme (4.5) for diffusion

• Centered convection

h	$\delta\tau_{\text{mit}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.425E-04	-	0.129E-04	-	0.261E-03	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.118E-04	1.852	0.374E-05	1.789	0.832E-04	1.647	0.000	0.139	0	236
0.062	0.00064	0.00064	0.315E-05	1.902	0.106E-05	1.822	0.208E-04	1.999	0.000	0.139	0	629
0.031	0.00016	0.00016	0.815E-06	1.951	0.286E-06	1.886	0.529E-05	1.976	0.000	0.139	0	2500
0.016	0.00004	0.00004	0.207E-06	1.976	0.743E-07	1.942	0.133E-05	1.997	0.000	0.139	0	10000

Table 20: Isotropic convection diffusion. Centered convection - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$, for diffusion

h	$\delta\tau_{\text{mit}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.126E-03	-	0.351E-04	-	0.912E-03	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.531E-04	1.245	0.174E-04	1.013	0.354E-03	1.363	0.000	0.139	0	236
0.062	0.00064	0.00064	0.166E-04	1.679	0.604E-05	1.527	0.101E-03	1.805	0.000	0.139	0	645
0.031	0.00016	0.00016	0.455E-05	1.866	0.175E-05	1.788	0.268E-04	1.920	0.000	0.139	0	2506
0.016	0.00004	0.00004	0.118E-05	1.940	0.468E-06	1.902	0.685E-05	1.969	0.000	0.139	0	10000

Table 21: Isotropic convection diffusion. Centered convection - Sub-upwinding scheme (4.6) for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.442E-03	-	0.137E-03	-	0.276E-02	-	0.000	0.139	0	60
0.125	0.00256	0.00256	0.212E-03	1.063	0.694E-04	0.978	0.147E-02	0.909	0.000	0.139	0	236
0.062	0.00064	0.00064	0.102E-03	1.056	0.342E-04	1.019	0.698E-03	1.074	0.000	0.139	0	644
0.031	0.00016	0.00016	0.498E-04	1.031	0.168E-04	1.023	0.337E-03	1.049	0.000	0.139	0	2507
0.016	0.00004	0.00004	0.247E-04	1.014	0.833E-05	1.016	0.166E-03	1.024	0.000	0.139	0	10000

Table 22: Isotropic convection diffusion. Centered convection - Godunov scheme (4.5) for diffusion

6.3.2 Anisotropic case, $l_y = 1000$

We now introduce anisotropy by taking $l_y = 1000$. We still compare the behavior of all the schemes introduced in this work for the diffusion term with, for the convection term, firstly the upwind scheme (see Tables 23-26), and secondly the centered scheme (see Tables 27-30). When the upwind scheme is used for the convection, the new Weighted centered-upwinding scheme manages to capture the well behavior of the Centered scheme without additional cost in terms of time and Newton iterations. Moreover when the centered scheme is used for the convection, the Weighted centered-upwinding scheme has succeeded to produce smallest errors than the Centered scheme. On both cases the Sub-upwinding and the Godunov schemes are less efficient.

• Upwinding convection

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.938E-03	-	0.327E-03	-	0.386E-02	-	0.000	0.139	0	62
0.125	0.00256	0.00256	0.601E-03	0.642	0.221E-03	0.567	0.313E-02	0.303	0.000	0.139	0	241
0.062	0.00064	0.00064	0.365E-03	0.719	0.133E-03	0.732	0.214E-02	0.548	0.000	0.139	0	939
0.031	0.00016	0.00016	0.213E-03	0.777	0.756E-04	0.815	0.137E-02	0.642	0.000	0.139	0	3750
0.016	0.00004	0.00004	0.121E-03	0.819	0.415E-04	0.867	0.874E-03	0.649	0.000	0.139	0	10034

Table 23: Anisotropic convection diffusion, $l_y = 1000$. Upwinding convection - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$, for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.240E-02	-	0.855E-03	-	0.993E-02	-	0.000	0.139	0	64
0.125	0.00256	0.00256	0.145E-02	0.722	0.549E-03	0.638	0.680E-02	0.545	0.000	0.139	0	242
0.062	0.00064	0.00064	0.844E-03	0.784	0.321E-03	0.777	0.410E-02	0.732	0.000	0.139	0	939
0.031	0.00016	0.00016	0.483E-03	0.806	0.180E-03	0.831	0.259E-02	0.660	0.000	0.139	0	3750
0.016	0.00004	0.00004	0.273E-03	0.820	0.994E-04	0.860	0.163E-02	0.666	0.000	0.139	0	10272

Table 24: Anisotropic convection diffusion, $l_y = 1000$. Upwinding convection - Sub-upwinding scheme (4.6) for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err $_{L^2}$	rate	err $_{L^1}$	rate	err $_{L^\infty}$	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.249E-02	-	0.888E-03	-	0.101E-01	-	0.000	0.139	0	63
0.125	0.00256	0.00256	0.163E-02	0.609	0.618E-03	0.524	0.747E-02	0.431	0.000	0.139	0	241
0.062	0.00064	0.00064	0.100E-02	0.706	0.382E-03	0.693	0.476E-02	0.652	0.000	0.139	0	939
0.031	0.00016	0.00016	0.591E-03	0.758	0.223E-03	0.776	0.306E-02	0.639	0.000	0.139	0	3750
0.016	0.00004	0.00004	0.342E-03	0.791	0.126E-03	0.824	0.193E-02	0.661	0.000	0.139	0	10328

Table 25: Anisotropic convection diffusion, $l_y = 1000$. Upwinding convection - Godunov scheme (4.5) for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.153E-02	-	0.536E-03	-	0.619E-02	-	0.000	0.139	0	64
0.125	0.00256	0.00256	0.795E-03	0.942	0.290E-03	0.884	0.375E-02	0.724	0.000	0.139	0	243
0.062	0.00064	0.00064	0.424E-03	0.906	0.155E-03	0.910	0.239E-02	0.647	0.000	0.139	0	939
0.031	0.00016	0.00016	0.229E-03	0.886	0.818E-04	0.918	0.145E-02	0.723	0.000	0.139	0	3750
0.016	0.00004	0.00004	0.125E-03	0.876	0.431E-04	0.924	0.903E-03	0.682	0.000	0.139	0	10041

Table 26: Anisotropic convection diffusion, $l_y = 1000$. Upwinding convection - Centered scheme (4.4) for diffusion

• Centered convection

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.111E-03	-	0.361E-04	-	0.503E-03	-	0.000	0.139	0	61
0.125	0.00256	0.00256	0.514E-04	1.117	0.191E-04	0.920	0.241E-03	1.059	0.000	0.139	0	237
0.062	0.00064	0.00064	0.193E-04	1.411	0.743E-05	1.359	0.116E-03	1.056	0.000	0.139	0	939
0.031	0.00016	0.00016	0.667E-05	1.535	0.257E-05	1.532	0.649E-04	0.840	0.000	0.139	0	3339
0.016	0.00004	0.00004	0.223E-05	1.579	0.775E-06	1.729	0.333E-04	0.962	0.000	0.139	0	10005

Table 27: Anisotropic convection diffusion, $l_y = 1000$. Centered convection - Weighted centered-upwinding scheme (4.7), $\gamma = 10^{-6}$, for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.185E-02	-	0.664E-03	-	0.793E-02	-	0.000	0.139	0	64
0.125	0.00256	0.00256	0.105E-02	0.825	0.396E-03	0.746	0.503E-02	0.656	0.000	0.139	0	242
0.062	0.00064	0.00064	0.586E-03	0.837	0.222E-03	0.837	0.300E-02	0.744	0.000	0.139	0	939
0.031	0.00016	0.00016	0.328E-03	0.836	0.121E-03	0.868	0.187E-02	0.686	0.000	0.139	0	3510
0.016	0.00004	0.00004	0.183E-03	0.840	0.658E-04	0.884	0.116E-02	0.693	0.000	0.139	0	10099

Table 28: Anisotropic convection diffusion, $l_y = 1000$. Centered convection - Sub-upwinding scheme (4.6) for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.206E-02	-	0.741E-03	-	0.841E-02	-	0.000	0.139	0	63
0.125	0.00256	0.00256	0.130E-02	0.672	0.492E-03	0.592	0.600E-02	0.486	0.000	0.139	0	241
0.062	0.00064	0.00064	0.776E-03	0.739	0.296E-03	0.732	0.375E-02	0.680	0.000	0.139	0	939
0.031	0.00016	0.00016	0.453E-03	0.777	0.170E-03	0.800	0.240E-02	0.642	0.000	0.139	0	3460
0.016	0.00004	0.00004	0.260E-03	0.802	0.950E-04	0.840	0.153E-02	0.654	0.000	0.139	0	10232

Table 29: Anisotropic convection diffusion, $l_y = 1000$. Centered convection - Godunov scheme (4.5) for diffusion

h	$\delta\tau_{\text{init}}$	$\delta\tau_{\text{max}}$	err_{L^2}	rate	err_{L^1}	rate	err_{L^∞}	rate	u_{min}	u_{max}	#Dt chop	#Newton
0.250	0.01024	0.01024	0.100E-02	-	0.358E-03	-	0.423E-02	-	0.000	0.139	0	64
0.125	0.00256	0.00256	0.467E-03	1.101	0.181E-03	0.983	0.210E-02	1.006	0.000	0.139	0	243
0.062	0.00064	0.00064	0.166E-03	1.491	0.667E-04	1.441	0.792E-03	1.410	0.000	0.139	0	939
0.031	0.00016	0.00016	0.475E-04	1.805	0.198E-04	1.749	0.247E-03	1.683	0.000	0.139	0	3430
0.016	0.00004	0.00004	0.125E-04	1.924	0.536E-05	1.887	0.654E-04	1.914	0.000	0.139	0	10052

Table 30: Anisotropic convection diffusion, $l_y = 1000$. Centered convection - Centered scheme (4.4) for diffusion

7 Conclusion

In this paper, we developed a new nonlinear finite volume discretization to approximate the positive solution to a degenerate parabolic equation on simplicial meshes. The diffusion tensor can be highly heterogeneous and strongly anisotropic. The original idea lies in the flux approximation. For the diffusive function, it uses a weighted harmonic mean instead of the upwinding already used in the literature of schemes with a two-point flux structure. A parameter is introduced to control the effects of the artificial diffusion introduced by the method. As a result, the positivity of the solution holds by construction. Moreover, the scheme satisfies stability results, namely the energy estimates allowing to prove the existence of numerical solutions. Adapting standard compactness arguments, the convergence of the finite volume scheme is proved. In the last section, we have presented relevant numerical tests focusing on the robustness and the efficiency of the weighted centered scheme compared to the centered, the sub-upwinding and the Godunov versions both in the presence of pure diffusion and also with an additional convection term.

Acknowledgment: E.H. Quenjel would like to thank the Département de la Marne, Greater Reims, Région Grand Est for their financial support of the Chair of Biotechnology of CentraleSupélec.

References

- [1] M. Afif and B. Amaziane. Convergence of finite volume schemes for a degenerate convection–diffusion equation arising in flow in porous media. *Computer Methods in Applied Mechanics and Engineering*, 191(46):5265–5286, 2002.
- [2] H. W. Alt and S. Luckhaus. Quasilinear elliptic-parabolic differential equations. *Mathematische Zeitschrift*, 183(3):311–341, 1983.
- [3] B. Andreianov, M. Bendahmane, and M. Saad. Finite volume methods for degenerate chemotaxis model. *Journal of Computational and Applied Mathematics*, 235(14):4015–4031, 2011.
- [4] B. Andreianov, C. Cancès, and A. Moussa. A nonlinear time compactness result and applications to discretization of degenerate parabolic–elliptic PDEs. *Journal of Functional Analysis*, 273(12):3633–3670, 2017.
- [5] M. Bendahmane, Z. Khalil, and M. Saad. Convergence of a finite volume scheme for gas–water flow in a multi-dimensional porous medium. *Mathematical Models and Methods in Applied Sciences*, 24(01):145–185, 2014.
- [6] K. Brenner and R. Masson. Convergence of a vertex centred discretization of two-phase Darcy flows on general meshes. *International Journal on Finite Volumes*, 10:1–37, 2013.
- [7] K. Brenner, R. Masson, and E. H. Quenjel. Vertex Approximate Gradient Discretization preserving positivity for two-phase Darcy flows in heterogeneous porous media. *Journal of Computational Physics*, 409:109357, 2020.
- [8] C. Cancès and C. Guichard. Convergence of a nonlinear entropy diminishing control volume finite element scheme for solving anisotropic degenerate parabolic equations. *Mathematics of Computation*, 85(298):549–580, 2016.
- [9] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Foundations of Computational Mathematics*, 17(6):1525–1584, 2017.
- [10] C. Cancès, M. Ibrahim, and M. Saad. Positive nonlinear CVFE scheme for degenerate anisotropic Keller–Segel system. *The SMAI Journal of Computational Mathematics*, 3:1–28, 2017.
- [11] G. Chamoun, M. Saad, and R. Talhouk. Monotone combined edge finite volume–finite element scheme for anisotropic Keller–Segel model. *Numerical Methods for Partial Differential Equations*, 30(3):1030–1065, 2014.

- [12] G. Chavent and J. Jaffré. *Mathematical models and finite elements for reservoir simulation: single phase, multiphase and multicomponent flows through porous media*, volume 17. North-Holland, Amsterdam, Stud. Math. Appl. edition, 1986.
- [13] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159. Springer Science & Business Media, 2013.
- [14] L. C. Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2010.
- [15] R. Eymard, T. Gallouët, D. Hilhorst, and Y. N. Slimane. Finite volumes and nonlinear diffusion equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 32(6):747–761, 1998.
- [16] R. Eymard, D. Hilhorst, and M. Vohralík. A combined finite volume–nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems. *Numerische Mathematik*, 105(1):73–131, 2006.
- [17] P. A. Forsyth. A control volume finite element approach to napl groundwater contamination. *SIAM Journal on Scientific and Statistical Computing*, 12(5):1029–1057, 1991.
- [18] M. Ghilani, E. H. Quenjel, and M. Saad. Positive control volume finite element scheme for a degenerate compressible two-phase flow in anisotropic porous media. *Computational Geosciences*, 23(1):55–79, 2019.
- [19] M. Ghilani, E. H. Quenjel, and M. Saad. Positivity-preserving finite volume scheme for compressible two-phase flows in anisotropic porous media: The densities are depending on the physical pressures. *Journal of Computational Physics*, 407:109233, 2020.
- [20] R. Herbin and F. Hubert. Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In R. Eymard and J.-M. Herard, editors, *Finite Volumes for Complex Applications V*, pages 659–692. Wiley, 2008.
- [21] D. Horstmann. From 1970 until present: the Keller-Segel model in chemotaxis and its consequences. I, Jahresber. *Deutsch. Math.-Verein.*, 105:103–165, 2003.
- [22] M. Ibrahim, E. H. Quenjel, and M. Saad. Positive nonlinear DDFV scheme for a degenerate parabolic system describing chemotaxis. *Computers & Mathematics with Applications*, 80(12):2972–3003, 2020.
- [23] M. Kaviany. *Principles of heat transfer in porous media*. Springer Science & Business Media, 2012.
- [24] F. Otto. L^1 -contraction and uniqueness for quasilinear elliptic–parabolic equations. *Journal of differential equations*, 131(1):20–38, 1996.
- [25] E. H. Quenjel. Enhanced positive vertex-centered finite volume scheme for anisotropic convection-diffusion equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 54(2):591–618, 2020.
- [26] E. H. Quenjel. Analysis of accurate and stable nonlinear finite volume scheme for anisotropic diffusion equations with drift on simplicial meshes. *Journal of Scientific Computing*, 88(3):1–26, 2021.
- [27] E. H. Quenjel. Nonlinear finite volume discretization for transient diffusion problems on general meshes. *Applied Numerical Mathematics*, 161:148–168, 2021.
- [28] E. H. Quenjel, M. Saad, M. Ghilani, and M. Bessemoulin-Chatard. Convergence of a positive nonlinear DDFV scheme for degenerate parabolic equations. *Calcolo*, 57(19), 2020.
- [29] J. L. Vázquez. *The porous medium equation: mathematical theory*. Oxford University Press on Demand, 2007.