Analysis of Accurate and Stable Nonlinear Finite Volume Scheme for Anisotropic Diffusion Equations with Drift on Simplicial Meshes

El Houssaine Quenjel

Analysis of accurate and stable nonlinear finite volume scheme for anisotropic diffusion equations with drift on simplicial meshes

El Houssaine QUENJEL

Université Côte d'Azur, LJAD, CNRS UMR 7351, and COFFEE team, INRIA Sophia Antipolis Méediterranée, Parc Valrose 06108 Nice Cedex 02, France. quenjel@unice.fr

August 1, 2021

Abstract

This work addresses the development and analysis of a second-order accurate finite volume scheme for parabolic equations with anisotropy on general simplicial meshes. The discretization involves only vertex unknowns without processing additional ones. The scheme construction makes use of a nonlinear transformation of the linear elliptic term. Two propositions are mainly presented for the approximation of the mobility function at the interfaces. The existence of positive solutions for the discrete system is guaranteed thanks to the proved a priori estimates. The energy dissipation of the scheme is moreover ensured. The convergence of the approach is established. Numerical tests are given to show the efficiency, accuracy and robustness of the proposed approach, with respect to the anisotropy, while a particular emphasis is set on the effects of the approximate mobility. They also confirm the obtained theoretical results, especially the decay of the free energy when time grows.

1 Introduction

Anisotropic convection-diffusion equations are essential in modeling many real-world problems arising from complex reservoir models [19], chemotaxis processes [31] and systems of semiconductor devices [4]. In such an application the numerical approximation of the model should ideally preserve the physical properties of the continuous problem. For instance, in the context of porous media flows the permeability can be strongly anisotropic and highly heterogeneous. Then, in these situations, the approximate solution accounting for the saturation or the concentration must be nonnegative without spurious oscillations in order to respect some standard laws of physics (e.g. Fick's law). This holds the name of a discrete maximum principle. Moreover, in the numerical modeling of semiconductors, the control of some dissipated entropies is very useful to understand the large time behavior of the physics being modeled. Hence, developing consistent, accurate and cheap numerical methods of type finite volumes on general meshes respecting the underlined properties is of great importance.

In view of its fundamental role, extensive research has been devoted to the discrete maximum principle for anisotropic linear diffusion problems. The pioneer two-point flux approximation (TPFA) scheme fulfills this property by construction [24]. In the presence of convective effects, it is recommended to employ the Scharfetter-Gummel approach of the TPFA scheme [16, 41] instead of the upwind version for better accuracy and stability. Unfortunately, these methods necessitates a stringent orthogonality condition on the mesh together with a scalar isotropic tensor. From a general viewpoint, it is well understood that no linear finite volume method for the linear diffusion preserves unconditionally the physical ranges of the approximate solution [22, 39] without additional assumptions. A tough attention was then given to nonlinear versions [10, 26, 34, 35, 38, 42]. Most of these schemes make use of a convex splitting of the flux in the co-normal direction. Other unknowns, placed at the vertices or edges etc, are required to define accurately each component and they can be eliminated after that. Although they are accurate of second-order, these kinds of methods do not fulfill the coercivity of the scheme without further assumptions on the mesh and the anisotropic ratio. To ensure the latter, a general approach was addressed in [11]. It is applied to centered scheme or merely built on coercive centered finite volume methods. Yet, consistency or convergence is valid when a numerical assumption holds. Recently, a few convergent positive schemes were designed and analyzed for degenerate parabolic equations with anisotropy, see for instance [13, 28, 27, 40]. The methodology used in these references is accurate of first order for the diffusion since the idea is based on an upwinding strategy of Godunov type. In the work [14], the authors proposed and investigated an accurate vertex approximation gradient methodology. It uses vertex and cell unknowns where the latter are eliminated at the solver level. Also, the implementation of this methodology shows that the iterative solver is extremely sensible to highly anisotropic ratios of the permeability when the discrete mobility function is given by a centered arithmetic mean. Recall that the advantage of the methods [12, 14] is that they can deal with more general grids and heterogeneities.

On the other hand, the considered model problem is dissipated and possesses an energy functional of Lyapunov type. The control of the latter allows to obtain relevant information on the asymptotic behavior of the solutions such as the exponential decay to the equilibrium [32, 33]. In the discrete setting, the large time behavior has been addressed in [7, 5, 18, 25] using a Two-Point finite volume discretizations together with the Scharfetter-Gummel strategy. The exponential decay to equilibrium states has been shown in [36] using jump processes. Establishing finite volume methods coupling between the energy diminishing, optimal accuracy, as well as positivity for such a model under general assumptions on the data and on the mesh is quite challenging. Only few papers have been devoted to such a topic in the context of multi-point framework [12, 14].

In this paper, we propose a vertex-centered finite volume scheme on general simplicial meshes applied to a basic anisotropic linear diffusion equation with drift, due to its practical importance. The main objective of this contribution consists in extending the ideas of the scheme presented in [40]. Indeed, even if the discretization elaborated in [40] enjoys some advantages, it however suffers from several limitations. For instance, we reveal :

- the numerical convergence is only first order in space;
- it does not allow the control of the free energy;
- it is not suitable to preserve the long time behavior of the solutions since the methodology is based on the upwinding, see [6] for deep details.

In the present contribution, we aim to dispense with the previous items weakening the use of the scheme of [40] in studying the long time behavior. We below recall and state the assets of the new scheme.

- Unknowns are only placed at the vertices, which turns out the computational cost cheaper;
- accuracy of second order;
- the control of the free energy;
- preserving the asymptotic of solutions as fast as possible (numerical exponential decay);
- handling highly anisotropic and heterogeneous domains;
- honoring the physical bounds (positivity) on the approximate solution;
- theoretical convergence towards the continuous solution of the mathematical problem.

The task is first achieved by formulating the elliptic term into a very specific nonlinear manner allowing the dissipation of the entropy function. Next, we use the spatial discretization elaborated in [40]. Moreover, in [40], we employed an upwind approximation of the mobility-like function. In the current paper, we take advantage of a centered scheme to avoid the reduction in the numerical convergence rate. Two propositions on the centered approximation can be made. First, one has the possibility to take the arithmetic mean value as done for instance in [13]. Second, one considers a fractional logarithmic average. Both of them respect the above prescribed physical properties of the solutions and allow the convergence of the numerical scheme from a theoretical point of view. In practice, the first choice does not support highly anisotropic tensors whereas the behavior of the logarithmic centered choice is independent of anisotropy.

The content of this paper is outlined as follows. We next state the mathematical model we are interested in as well as the definition of weak solutions. Section 2 describes the finite volume discretization, namely the spatial and temporal meshes, the reconstruction operators and the discrete functional spaces. We next present the proposed numerical scheme under its pointwise and concise forms. In Section 3, we establish several properties of the scheme, especially the energy estimates, estimations on the free energy of the system and existence of discrete solutions. In Section 4, we study the convergence of these solutions towards a weak solution by means of a recent compactness argument. Finally, Section 5 shows numerical assessment of errors produced by our method and their associate convergence rates. The approach robustness with respect to the tensor anisotropy is also tested. It is illustrated that the proposed methodology is free diminishing in the sense that the energy dissipates in the course of time.

Let Ω be a bounded connected open polyhedral subset of $\mathbb{R}^d$ ($d \in \mathbb{N}^*$) with a Lipschitz boundary Γ. Let us set $Q_{t_f} = \Omega \times (0, t_f)$ where $t_f > 0$ is the final time. In this work, we are interested in the model example

$$s_t - \operatorname{div}\Big(\mathbb{D}\nabla s + s\mathbb{D}\nabla G(x)\Big) = 0 \quad \text{in} \quad Q_{t_f}, \tag{1.1}$$

with the no-flux boundary condition

$$\Big(\mathbb{D}\nabla s + s\mathbb{D}\nabla G(x)\Big) \cdot \mathbf{n} = 0 \quad \text{on} \quad \Gamma \times (0, t_f), \tag{1.2}$$

and the initial state of the solution

$$s(0, \cdot) = s^0(x) \quad \text{in} \quad \Omega. \tag{1.3}$$

The approach we adapt here is based on an equivalent formulation of the potential. Therefore, the equation (1.1) rewrites

$$s_t - \operatorname{div}\Big(s\mathbb{D}\nabla(\log(s) + G(x))\Big) = 0 \quad \text{in} \quad Q_{t_f}. \tag{1.4}$$

Now, let us state the main assumptions on the physical data of the problem. They are mandatory to give a proper sense to the sought solutions.

($\mathbf{H}_1$) The matrix $\mathbb{D} \in L^\infty(\Omega)^{d\times d}$ is symmetric and uniformly elliptic in the sense that there exist α_0 and α_1 such that

$$\alpha_0 |u|^2 \leq \mathbb{D}(x)u \cdot u \leq \alpha_1 |u|^2 \quad \forall \, a.e. \; x \in \Omega \text{ and } \forall u \in \mathbb{R}^d.$$

($\mathbf{H}_2$) The exterior potential G is a nonnegative function of $\mathcal{C}^1(\overline{\Omega}, \mathbb{R})$.

($\mathbf{H}_3$) The initial datum s^0 is nonnegative and belongs to $L^1(\Omega)$. We define the convex mapping $\mathcal{H} : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ by

$$\mathcal{H}(v) = v \log(v) - v + 1. \tag{1.5}$$

We require that this function possesses a finite entropy condition on s^0 i.e.

$$\int_\Omega \mathcal{H}(s^0(x)) \, \mathrm{d}x < +\infty, \quad \text{and} \quad \int_\Omega s^0(x) \, \mathrm{d}x > 0. \tag{1.6}$$

The following statement surveys the definition of a weak solution.

Definition 1.1. *Under Assumptions **($\boldsymbol{H}_1$)-($\boldsymbol{H}_3$)**, a measurable function $s : Q_{t_f} \longrightarrow (0, +\infty)$ is said to be a weak solution to the problem* (1.1)-(1.3) *if the following items are fulfilled.*

(i) $\mathcal{H}(s) \in L^\infty(0, t_f; L^1(\Omega))$ *and* $\xi(s) := \sqrt{s} \in L^2(0, t_f; H^1(\Omega))$.

(ii) The integral identity holds

$$-\int_{Q_{t_f}} s\psi_t \, \mathrm{d}x \, \mathrm{d}t - \int_\Omega s^0 \psi(\cdot, 0) \, \mathrm{d}x + \int_{Q_{tf}} \mathbb{D}\Big(\nabla s + s\nabla G(x)\Big) \cdot \nabla \psi \, \mathrm{d}x \, \mathrm{d}t = 0, \;\; \forall \psi \in \mathcal{C}_c^\infty(\overline{\Omega} \times [0, t_f)). \tag{1.7}$$

The convergence (up to a subsequence) of the proposed numerical scheme (2.10)-(2.12) of the current work yields directly the existence of a weak solution to the model (1.1)-(1.3). More details are given in Propositions 4.1-4.2. A same strategy was employed in [12] in the context of discrete duality finite volume method. Even though no-flux Neumann boundary conditions are imposed in our case, one can mimic the compactness approach developed in [1.7. Existence Theorem.][1] to provide another alternative for the existence result. In this reference, the authors treated a general parabolic model with mixed Dirichlet-Neumann boundary conditions. Note that the current assumptions on the data (e.g. **($\mathbf{H}_3$)**) and the definition of the weak solution (e.g. item (i) of Definition 1.1) are different from the ones prescribed in [1]. As highlighted in [12], the uniqueness question of such a weak solution is still an open problem for $s_0 \in L^2(\Omega)$. In the case where $\mathcal{H}(s^0) \in L^1(\Omega)$, the authors conjectured that the weak solutions in the sense of Definition 1.1 are renormalized ones, which yields the uniqueness. As a consequence, one could recover the convergence of the whole sequence of solutions.

We define the free energy $\mathfrak{E}$ and the dissipation $\mathfrak{I}$ for the evolution equation (1.1) by

$$\mathfrak{E}(t) = \int_\Omega \mathcal{H}(s) + sG \, \mathrm{d}x, \quad \mathfrak{I}(t) = \int_\Omega s\mathbb{D}\nabla(\log(s) + G) \cdot \nabla(\log(s) + G) \, \mathrm{d}x.$$

Formally, taking $\psi = \log(s) + G$ in the weak formulation (1.7) implies the energy-dissipation relationship

$$\frac{\mathrm{d}\mathfrak{E}}{\mathrm{d}t} + \mathfrak{I} = 0.$$

Using the fact that $\mathfrak{I} \geq 0$ we get

$$\frac{\mathrm{d}\mathfrak{E}}{\mathrm{d}t} = -\mathfrak{I} \leq 0. \tag{1.8}$$

Thereby, the free energy is diminishing in time. This structural feature was intensively employed to study the large time behavior of solutions thanks to entropy's methods. For more details on this strategy, we refer to the article [15]. In the recent years, a particular attention was paid to setting the entropy principles from a numerical point of view. For instance, using TPFA scheme, the numerical investigation of asymptotic of solutions to convection-diffusion equations based on the discrete version of the entropy methodology can be found in [5, 17, 29, 25, 18]. We also mention the work [12] where a discrete duality finite volume scheme enjoying (1.8) was studied. We will see later on that our methodology respects the decay of the free energy even on anisotropic situations.

2 Finite volume discretization

In this section, we set the discrete tools that are necessary to define and study the proposed numerical scheme.

2.1 Meshes and reconstruction operators

Spatial discretization : the domain Ω is discretized with two different partitions. The first one consists in an admissible finite element triangulation referred to as the primal mesh. The second one or the dual mesh is built around the vertices of the elements. While the presentation of the terminologies below can be extended to the 3D setting, we rather expose the 2D version in order to keep the ease readability of the discrete framework.

Let $\mathcal{T}$ be a conforming triangulation of Ω composed of triangles satisfying $\bigcup_{T\in\mathcal{T}} \overline{T} = \overline{\Omega}$. We define x_T the mass center and h_T the diameter of the simplex $T \in \mathcal{T}$. Also, let ρ_T be the largest diameter of the inscribed ball in T. The size and regularity of the primal mesh are respectively given by

$$h_{\mathcal{T}} = \max_{T\in\mathcal{T}} h_T, \quad \text{and} \quad \Theta_{\mathcal{T}} = \max_{T\in\mathcal{T}} \frac{h_T}{\rho_T}.$$

When $h_{\mathcal{T}}$ tends to zero, we assume that the refined meshes $\mathcal{T}_h$ of the same triangulation have a uniformly bounded regularity $\Theta_{\mathcal{T}_h}$ [20].

We denote $\mathcal{V}$ the set of vertices of the mesh, and $\mathcal{V}_T$ the nodes of the simplex $T \in \mathcal{T}$. We can label each vertex of T thanks to the permutation τ acting on $\mathcal{V}_T$ as depicted on the left side of Figure 1. We denote $\mathcal{T}_i$ the set of elements having in common the vertex $i \in \mathcal{V}$. We assign a unique dual cell or control volume K_i centered at i for any given node $i \in \mathcal{V}$. A simple way to construct K_i is to connect the center of any triangle $T \in \mathcal{T}_i$ with the midpoint of edges of T sharing the same vertex i. Given two finite volumes K_i and $K_{\tau(i)}$ we set $\overline{\sigma^T_{i\tau(i)}} = \overline{K_i} \cap \overline{K_{\tau(i)}} \cap \overline{T}$. The notation $\left|\sigma^T_{i\tau(i)}\right|$ represents the $(d-1)$-dimensional measure of $\sigma^T_{i\tau(i)}$. The unit normal on this interface oriented from K_i to $K_{\tau(i)}$ is denoted by $\mathbf{n}_{\sigma^T_{i\tau(i)}}$. The set $\mathcal{E}^T_{K_i}$ refers to all the edges of K_i contained in the triangle $T \in \mathcal{T}_i$. We designate by x_i the center of the control volume K_i, and by $|K_i|$ the d-dimensional measure of K_i. The set of the control volumes K_i will be referred to as $\mathcal{M}$. Note that x_i denotes also the coordinate of the node i.

Remark 2.1. *In case of $d \geq 3$ one always can label the faces sharing a vertex $i \in \mathcal{V}_T$ for some element T thanks to a suitable permutation τ. A possible way to do so is given as follows. First, we fix the combination*

$$\tau^*(i) = \tau(i)\tau\circ\tau(i)\cdots\tau^d(i), \quad \tau^d = \underbrace{\tau\circ\cdots\circ\tau}_{d \text{ times}},$$

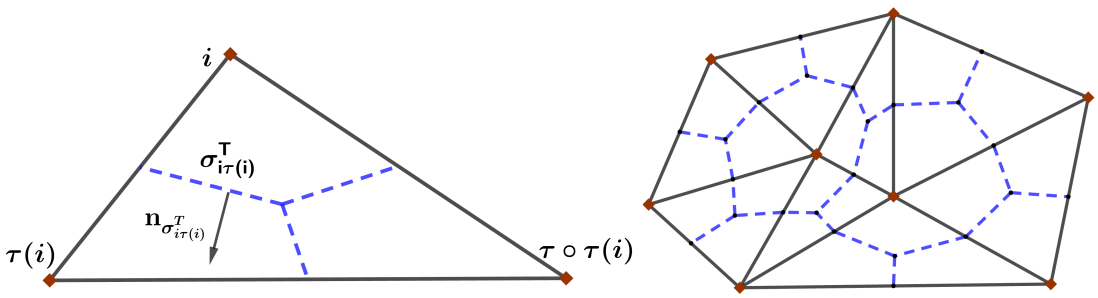

Figure 1: The spatial discretization of $T \in \mathcal{T}$ (left) and illustration of the dual mesh (right) given by dashed segments.

such that $\tau^{d+1}(i) = i$. Then we apply τ d-times on $i\tau^*(i)$ to obtain all faces sharing the node i. In this case, we would use the notation $\sigma^T_{i\tau^*(i)}$ instead of $\sigma^T_{i\tau(i)}$ for the dual interfaces.

Temporal discretization : the time interval $(0, t_f)$ is discretized thanks to a uniform subdivision given by the increasing sequence $(t^k)_{k=0,\cdots,N}$ such that $t^0 = 0$ and $t^N = t_f$. Then, the time step is fixed to $\triangle t = t_f/N$. This uniform stepping is only taken to avoid heavy notations. The proofs carried out in this paper can be broadened to nonuniform time steps without any technical issues.

Reconstruction operators : first, let us define $X_{\mathcal{T}}$ the standard $\mathbb{P}_1$ finite elements space on $\mathcal{T}$. We consider $(\varphi_i)_{i\in\mathcal{V}}$ the classical basis of $X_{\mathcal{T}}$ satisfying $\varphi_i(x_j) = 1$ if $j = i$ and $\varphi_i(x_j) = 0$ else. We define the reconstruction operator $\pi_{\mathcal{T}} : \mathbb{R}^{\#\mathcal{V}} \longrightarrow X_{\mathcal{T}}$ by setting

$$\pi_{\mathcal{T}} u = \sum_{i\in\mathcal{V}} u_i \varphi_i, \quad \forall u \in \mathbb{R}^{\#\mathcal{V}}.$$

Similarly, let $X_{\mathcal{M}}$ be the space of piecewise constant functions on the dual mesh $\mathcal{M}$. We define the piecewise constant operator $\pi_{\mathcal{M}} : \mathbb{R}^{\#\mathcal{V}} \longrightarrow X_{\mathcal{M}}$ such that

$$\pi_{\mathcal{M}} u = \sum_{i\in\mathcal{V}} u_i \mathbf{1}_{K_i}, \quad \forall u \in \mathbb{R}^{\#\mathcal{V}}.$$

If g is a nonlinear function from $\mathbb{R}$ into $\mathbb{R}$. We adopt the convention making $g(u) = (g(u_i))_{i\in\mathcal{V}}$ for every $u \in \mathbb{R}^{\#\mathcal{V}}$. Then, one has

$$\pi_{\mathcal{T}} g(u) = \sum_{i\in\mathcal{V}} g(u_i)\varphi_i, \quad \text{and} \quad \pi_{\mathcal{M}} g(u) = \sum_{i\in\mathcal{V}} g(u_i)\mathbf{1}_{K_i}, \quad \forall u \in \mathbb{R}^{\#\mathcal{V}}.$$

Let $\{\pi_{\mathcal{T}} u^k\}_{k=0,\cdots,N}$ be a sequence of functions of $X_{\mathcal{T}}$ and let $\{\pi_{\mathcal{M}} u^k\}_{k=0,\cdots,N}$ be a family of functions of the trial space $X_{\mathcal{M}}$. We define the time-dependent reconstruction functions $\pi_{\mathcal{T},\triangle t}, \pi_{\mathcal{M},\triangle t}$ respectively by

$$\pi_{\mathcal{T},\triangle t} u(\cdot, t) = \pi_{\mathcal{T}} u^k, \text{ and } \pi_{\mathcal{M},\triangle t} u(\cdot, t) = \pi_{\mathcal{M}} u^k, \ \forall t \in (t^{k-1}, t^k], \ \forall k = 1, \cdots, N. \tag{2.1}$$

This conduct us to define the time-dependent discrete gradient $\nabla\pi_{\mathcal{T},\triangle t} u$ by

$$\nabla\pi_{\mathcal{T},\triangle t} u(\cdot, t) = \nabla\pi_{\mathcal{T}} u^k, \quad \forall t \in (t^{k-1}, t^k], \ \forall k = 1, \cdots, N.$$

The following result establishes the link between some reconstruction functions and the discrete gradient.

Lemma 2.1. *(see [40, Lemma 2.4]) Let $\pi_{\mathcal{T}} u$ be in $X_{\mathcal{T}}$. We consider the piecewise constant functions $\overline{\pi_{\mathcal{T}} u}$ and $\underline{\pi_{\mathcal{T}} u}$ defined respectively by*

$$\overline{\pi_{\mathcal{T}} u}(x) = \max_{y \in T} \pi_{\mathcal{T}} u(y) \text{ and } \underline{\pi_{\mathcal{T}} u}(x) = \min_{y \in T} \pi_{\mathcal{T}} u(y), \quad \forall x \in T,\ \forall T \in \mathcal{T}.$$

Then, one gets

$$\|\pi_{\mathcal{T}} \overline{u} - \pi_{\mathcal{T}} \underline{u}\|_{L^2(\Omega)} \leq \#\mathcal{V}_T h_{\mathcal{T}} \|\nabla \pi_{\mathcal{T}} u\|_{L^2(\Omega)^d}. \tag{2.2}$$

As a consequence

$$\|\pi_{\mathcal{M}} u - \pi_{\mathcal{T}} u\|_{L^2(\Omega)} \leq \#\mathcal{V}_T h_{\mathcal{T}} \|\nabla \pi_{\mathcal{T}} u\|_{L^2(\Omega)^d}. \tag{2.3}$$

2.2 Description of the numerical scheme

For the sake of simplicity we set $\Upsilon(s) = \log(s) + G$ throughout the paper. Without loss of generality, we assume that the tensor $\mathbb{D}$ is constant per triangles. Before presenting the proposed numerical scheme, we would like to survey how we derive the main discrete terms. Let us select a control volume K_i for some $i \in \mathcal{V}$. We integrate on K_i with $\partial K_i = \bigcup_{T \in \mathcal{T}_i} \bigcup \sigma^T_{i\tau(i)}$ and make use of Green's formula to get

$$-\int_{\partial K_i} s\mathbb{D}\nabla\Upsilon(s) \cdot \mathbf{n}\, \mathrm{d}\sigma = -\sum_{T \in \mathcal{T}_i} \sum_{\sigma^T_{i\tau(i)} \in \mathcal{E}^T_{K_i}} \int_{\sigma^T_{i\tau(i)}} s\nabla\Upsilon(s) \cdot \mathbb{D}\mathbf{n}_{\sigma^T_{i\tau(i)}}\, \mathrm{d}\sigma. \tag{2.4}$$

Being a finite volume method, our construction is based on the discretization of the local continuous fluxes. Following [40] we propose the following centered scheme to the integral flux across the interface $\sigma^T_{i\tau(i)}$

$$-\int_{\sigma^T_{i\tau(i)}} s\nabla\Upsilon(s) \cdot \mathbb{D}\mathbf{n}_{\sigma^T_{i\tau(i)}}\, \mathrm{d}\sigma \approx \sqrt{s_{i\tau(i)}}\mathcal{G}^T_{i\tau(i)}(s_{\mathcal{T}}), \tag{2.5}$$

where $s_{i\tau(i)}$ is given by combining the two different means

$$s^c_{i\tau(i)} = \frac{s_i + s_{\tau(i)}}{2}, \quad s^\ell_{i\tau(i)} = \begin{cases} \left(2\dfrac{\sqrt{s}_i - \sqrt{s}_{\tau(i)}}{\log(s_i) - \log(s_{\tau(i)})}\right)^2, & \text{if } s_i \neq s_{\tau(i)}, \\ s_i, & \text{otherwise} \end{cases}, \tag{2.6}$$

in the sense that

$$s_{i\tau(i)} = (1-\lambda)s^\ell_{i\tau(i)} + \lambda s^c_{i\tau(i)}, \tag{2.7}$$

where λ is a parameter belonging to $(0, 1]$. The fact that $\lambda > 0$ is required to reinforce the existence of $\lambda^\star > 0$ such that

$$s_{i\tau(i)} \geq \lambda^\star \max\big(s_i, s_{\tau(i)}\big). \tag{2.8}$$

In our case $\lambda^\star = \lambda/2$. This inequality is a key ingredient to prove the existence of positive solution to the proposed finite volume scheme. From a numerical viewpoint, we can consider a very small λ in order to show the efficiency of $s_{i\tau(i)} \approx s^\ell_{i\tau(i)}$ compared to $s^c_{i\tau(i)}$.

The quantity $\mathcal{G}^T_{i\tau(i)}(s_{\mathcal{T}})$ is a numerical flux-like function that approximates $\sqrt{s}\nabla\Upsilon(s)$ along the co-normal vector $\mathbb{D}\mathbf{n}_{\sigma^T_{i\tau(i)}}$

$$\mathcal{G}^T_{i\tau(i)}(s_{\mathcal{T}}) = \sum_{j \in \mathcal{V}_T} a^T_{ij} \sqrt{s_{j\tau(j)}} \Big(\Upsilon(s_j) - \Upsilon(s_{\tau(j)})\Big), \tag{2.9}$$

where the weight a_{ij}^T is given by [40]

$$a_{ij}^T = \frac{1}{|T|}\left|\sigma_{i\tau(i)}^T\right|\left|\sigma_{j\tau(j)}^T\right|\mathbb{D}\mathbf{n}_{\sigma_{i\tau(i)}^T}\cdot\mathbf{n}_{\sigma_{j\tau(j)}^T}.$$

By construction, notice that the numerical flux-like function is conservative. Indeed, the flux from $\tau(i)$ to i in the triangle T is given by $\mathcal{G}_{\tau(i)i}(s_\mathcal{T})$. Replacing i by $\tau(i)$ in (2.9), with $\tau\circ\tau(i)=i$, implies

$$\mathcal{G}_{\tau(i)i}(s_\mathcal{T}) = \sum_{j\in\mathcal{V}_T} a_{\tau(i)j}^T\sqrt{s_{j\tau(j)}}\Big(\Upsilon(s_j)-\Upsilon(s_{\tau(j)})\Big).$$

Now, observe that

$$a_{\tau(i)j}^T = \frac{1}{|T|}\left|\sigma_{i\tau(i)}^T\right|\left|\sigma_{j\tau(j)}^T\right|\mathbb{D}\mathbf{n}_{\tau(i)i}\cdot\mathbf{n}_{\sigma_{j\tau(j)}^T} = -\frac{1}{|T|}\left|\sigma_{i\tau(i)}^T\right|\left|\sigma_{j\tau(j)}^T\right|\mathbb{D}\mathbf{n}_{\tau(i)i}\cdot\mathbf{n}_{\sigma_{j\tau(j)}^T} = -a_{ij}^T.$$

This entails that $\mathcal{G}_{\tau(i)i}(s_\mathcal{T}) = -\mathcal{G}_{i\tau(i)}(s_\mathcal{T})$, which ensures the discrete local flux conservation.

To avoid a CFL constraint on the time steps in case of strong anisotropic ratios of the tensor $\mathbb{D}$ we consider a fully implicit (backward Euler) discretization in time. Then, the numerical scheme consists in finding a sequence $(s_i^k)_i$ solving the following algebraic set of equations at each time iteration $k\in\{1,\cdots,N\}$,

$$s_i^0 = \frac{1}{|K_i|}\int_{K_i} s^0(x)\,\mathrm{d}x,\quad \forall i\in\mathcal{V}, \tag{2.10}$$

$$|K_i|\,\frac{s_i^k - s_i^{k-1}}{\triangle t} + \sum_{T\in\mathcal{T}_i}\sum_{\sigma_{i\tau(i)}^T\in\mathcal{E}_{K_i}^T}\sqrt{s_{i\tau(i)}^k}\,\mathcal{G}_{i\tau(i)}^T(s_\mathcal{T}^k) = 0,\quad \forall i\in\mathcal{V}, \tag{2.11}$$

$$s_i^k > 0,\quad \forall i\in\mathcal{V}. \tag{2.12}$$

The last constraint gives sense to sought discrete solutions due to the use of the log function.

Remark 2.2. *The nonlinear scheme takes into account a zero-flux boundary condition. The discretization can be applied to homogeneous Neumann and (positive) Dirichlet boundary conditions for instance. For this purpose, one first needs to specify the set of Dirichlet boundary vertices $\mathcal{V}^D$, see [40]. Then, the corresponding numerical scheme* (2.10)-(2.12) *remains the same for any degree of freedom $i\in\mathcal{V}\backslash\mathcal{V}^D$ and we impose the considered (positive) Dirichlet function for $i\in\mathcal{V}^D$. The scope of this paper (positivity, a priori estimates, convergence) holds true for these kinds of boundary conditions.*

The above system can be rewritten under a compact form. This formulation is of chief importance since it allows to straightforwardly conduct the analysis of the resulting finite volume scheme in a rigorous setting. To this purpose, we introduce the local diagonal matrix $(\mathcal{S}_{ij}^T)_{1\le i,j\le\#\mathcal{V}_T}$ whose entries are

$$\mathcal{S}_{ij}^T = \begin{cases}\sqrt{s_{i\tau(i)}^k} & \text{if } j=\tau(i)\\ 0 & \text{if } j\neq\tau(i)\end{cases}.$$

For every $u\in\mathbb{R}^{\#\mathcal{V}_T}$, we denote $\delta_T u$ the vector of $\mathbb{R}^{\#\mathcal{V}_T}$ such that the i^{th} component is $(\delta_T u)_i = u_i - u_{\tau(i)}$. We next define the discrete L^2-norm on $\mathbb{R}^{\#\mathcal{V}}$ by

$$[\![u_\mathcal{T}, v_\mathcal{T}]\!]_\mathcal{T} = \sum_{i=1}^{\#\mathcal{V}} |K_i|\,u_i v_i.$$

Therefore, the equivalent form of the numerical scheme is obtained by performing a discrete integration by parts in space. It reads

$$\left[\!\!\left[\frac{s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}}{\triangle t}, v_{\mathcal{T}}\right]\!\!\right]_{\mathcal{T}} + \sum_{T\in\mathcal{T}} \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T v_{\mathcal{T}} = 0, \quad \forall v_{\mathcal{T}} \in X_{\mathcal{T}},\ \forall k = 1, \cdots, N. \tag{2.13}$$

The symmetric matrix $\mathcal{A}^T$ is defined by $\mathcal{A}^T = (a_{ij}^T)_{1\le i,j\le \#\mathcal{V}_T}$. By taking $v_{\mathcal{T}} = \mathbf{1}_{\mathcal{T}}$ in (2.13) we recover the discrete conservation of mass claiming

$$\left[\!\!\left[s_{\mathcal{T}}^k, \mathbf{1}_{\mathcal{T}}\right]\!\!\right]_{\mathcal{T}} = \int_\Omega \pi_{\mathcal{M}} s^k \,\mathrm{d}x = \int_\Omega s^0 \,\mathrm{d}x, \quad \forall k = 1, \cdots, N. \tag{2.14}$$

The following lemma is one of the keystone results for the convergence analysis of our scheme. It shows in particular that the matrix $\mathcal{A}^T$ is definite-positive.

Lemma 2.2. *There exist two positive constant μ_1 and μ_2 that depend only on α_0, α_1 and the mesh regularity so that for all $T \in \mathcal{T}$, one has*

$$\mu_1 \frac{|T|}{h_T^2} \delta_T u \cdot \delta_T u \le \delta_T u \cdot \mathcal{A}^T \delta_T u \le \mu_2 \frac{|T|}{h_T^2} \delta_T u \cdot \delta_T u, \quad \forall u \in \mathbb{R}^{\#\mathcal{V}_T}. \tag{2.15}$$

Proof. The proof is carried out in [40, Lemma 2.3]. It extends the ideas provided in [9, 14]. □

Remark 2.3. *From a theoretical viewpoint, $s_{i\tau(i)}^k$ can take other possible mean values. Indeed, the proposed scheme and the analysis are valid for any average $s_{i\tau(i)}^k$ satisfying*

$$s_{i\tau(i)}^{\ell,k} < s_{i\tau(i)}^k \le s_{i\tau(i)}^{c,k}, \quad \textit{and} \quad s_{i\tau(i)}^k \ge \lambda^\star \max\left(s_i^k, s_{\tau(i)}^k\right),$$

for some $\lambda^\star > 0$. An example fulfilling these equalities is given by [37]

$$s_{i\tau(i)}^k = \left(\frac{(s_i^k)^r + (s_{\tau(i)}^k)^r}{2}\right)^{1/r}, \quad \frac{1}{6} \le r \le 1.$$

In the sequel, unless specified, we denote by C generic positive constants that depend only on the physical data described in **($\mathbf{H}_1$)-($\mathbf{H}_3$)** and possibly on the mesh regularity.

3 A priori estimates and existence of discrete solutions

A priori estimates serve for multiple purposes. They ensure the stability of the proposed approach, namely the coercivity of the scheme and the decay of the free energy. The coercivity means that we can get a control on the approximate gradient of $\xi(s)$. This property allows in particular to prove the existence of the finite volume scheme. Therefore, one can derive a positive lower bound on any discrete solution. Contrary to [12] our energy estimates, namely the next result, are not based on the energy-dissipation estimations.

Proposition 3.1. *Let $(s_i^k)_{i\in\mathcal{V}}$ be a sequence such that the finite volume scheme* (2.10)-(2.12) *holds. Then, there exist positive constants C_1, C_2, and C_3, depending only on the physical data and the regularity of the mesh such that*

$$\sum_{T\in\mathcal{T}} |T| \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \le C_1, \tag{3.1}$$

$$\sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \Big(\log(s_i^k) - \log(s_{\tau(i)}^k) \Big)^2 \le C_2, \tag{3.2}$$

$$\sum_{k=1}^{N} \triangle t \left\| \nabla \pi_{\mathcal{T}} \xi(s^k) \right\|_{L^2(\Omega)^d} \le C_3. \tag{3.3}$$

Proof. First, we observe that $s_{i\tau(i)}^k \le s_i^k + s_{\tau(i)}^k$ and $|T| \le C\,|T \cap K_i|$ for all $i \in \mathcal{V}_T$. Summing by dual cells and using the mass conservation property (2.14) we directly check

$$\sum_{T\in\mathcal{T}} |T| \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \le 2C \sum_{i\in\mathcal{V}} |K_i|\, s_i^k \le 2C \left\| s^0 \right\|_{L^1(\Omega)} = C_1.$$

Next, take $v_{\mathcal{T}} = \log(s_{\mathcal{T}}^k)$ in the formulation (2.13) and sum over $k = 1, \cdots, N$. This gives

$$E_1 + E_2 + E_3 = 0, \tag{3.4}$$

where each term writes

$$\begin{aligned}
E_1 &= \sum_{k=1}^{N} \left[\!\!\left[s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}, \log(s_{\mathcal{T}}^k) \right]\!\!\right]_{\mathcal{T}}, \\
E_2 &= \sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \mathcal{S}^T \delta_T \log(s_{\mathcal{T}}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \log(s_{\mathcal{T}}^k), \\
E_3 &= \sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \mathcal{S}^T \delta_T G_{\mathcal{T}} \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \log(s_{\mathcal{T}}^k).
\end{aligned}$$

Thanks to the convexity of the function $\mathcal{H}$ defined in (1.5) we obtain

$$\left[\!\!\left[\mathcal{H}(s_{\mathcal{T}}^N) - \mathcal{H}(s_{\mathcal{T}}^0), \mathbf{1}_{\mathcal{T}} \right]\!\!\right]_{\mathcal{T}} \le E_1. \tag{3.5}$$

According to the Cauchy-Schwarz inequality, the fact that $\left|G(x_i) - G(x_{\tau(i)})\right| \le \|\nabla G\|_\infty h_T$ for all $i \in \mathcal{V}_T$, Lemma 2.2 and the estimation (3.1) we get

$$\begin{aligned}
|E_3| &\le (E_2)^{1/2} \left(\sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \mathcal{S}^T \delta_T G_{\mathcal{T}} \cdot \mathcal{A}^T \mathcal{S}^T \delta_T G_{\mathcal{T}} \right)^{1/2} \\
&\le C (E_2)^{1/2} \left(\sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \Big(G(x_i) - G(x_{\tau(i)}) \Big)^2 \right)^{1/2} \\
&\le C \|\nabla G\|_\infty (E_2)^{1/2} \left(\sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} |T| \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \right)^{1/2} \le C (E_2)^{1/2}.
\end{aligned}$$

By virtue of Young's inequality we infer

$$|E_3| \leq C + \frac{1}{2}E_2. \tag{3.6}$$

In addition, we combine (3.4)–(3.6) and use the finite entropy condition (1.6) to ensure the uniform bound $E_2 \leq C$. Owing to (2.15), it can be checked that

$$\sum_{k=1}^{N} \triangle t \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \Big(\log(s_i^k) - \log(s_{\tau(i)}^k) \Big)^2 \leq C_2.$$

Applying once more Lemma 2.2 yields

$$\Big\| \nabla \pi_{\mathcal{T}} \xi(s^k) \Big\|_{L^2(\Omega)^d}^2 \leq \mu_1 \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} \Big(\xi(s_i^k) - \xi(s_{\tau(i)}^k) \Big)^2.$$

By definition of $s_{i\tau(i)}^{\ell,k}$ (2.6) we automatically obtain

$$\begin{aligned} \Big\| \nabla \pi_{\mathcal{T}} \xi(s^k) \Big\|_{L^2(\Omega)^d}^2 &\leq \frac{\mu_1}{4} \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^{\ell,k} \Big(\log(s_i^k) - \log(s_{\tau(i)}^k) \Big)^2 \\ &\leq \frac{\mu_1}{4} \sum_{T\in\mathcal{T}} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^{k} \Big(\log(s_i^k) - \log(s_{\tau(i)}^k) \Big)^2. \end{aligned}$$

We therefore derive the estimation (3.3). □

Let us elaborate the discrete counterpart of the energy-dissipation relationship (1.8). To this end, we define the approximate free energy by

$$\mathfrak{E}_h^k = \Big[\!\Big[\mathcal{H}(s_{\mathcal{T}}^k), \mathbf{1}_{\mathcal{T}} \Big]\!\Big]_{\mathcal{T}} + \Big[\!\Big[G_{\mathcal{T}}, s_{\mathcal{T}}^k \Big]\!\Big]_{\mathcal{T}} \geq 0, \quad \forall k = 0, \cdots, N. \tag{3.7}$$

The approximate dissipation is given by

$$\mathfrak{J}_h^k = \sum_{T\in\mathcal{T}} \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k). \tag{3.8}$$

Proposition 3.2. *The finite volume scheme is free diminishing* (2.10)-(2.12) *in the sense that*

$$\frac{\mathfrak{E}_h^k - \mathfrak{E}_h^{k-1}}{\triangle t} + \mathfrak{J}_h^k \leq 0, \quad \forall k = 1, \cdots, N. \tag{3.9}$$

Consequently, the discrete free energy is bounded and is decreasing in time

$$0 \leq \mathfrak{E}_h^k \leq \mathfrak{E}_h^{k-1} \leq \big\| \mathcal{H}(s^0) \big\|_{L^1(\Omega)} + \|G\|_\infty \big\| s^0 \big\|_{L^1(\Omega)}, \quad \forall k = 1, \cdots, N. \tag{3.10}$$

Furthermore, the time integral of the discrete dissipation is uniformly bounded

$$0 \leq \sum_{k=1}^{N} \triangle t \mathfrak{J}_h^k \leq \big\| \mathcal{H}(s^0) \big\|_{L^1(\Omega)} + \|G\|_\infty \big\| s^0 \big\|_{L^1(\Omega)}. \tag{3.11}$$

Finally, there holds

$$\| \pi_{\mathcal{M},\triangle t} \mathcal{H}(s) \|_{L^\infty(0,t_f;L^1(\Omega))} \leq \big\| \mathcal{H}(s^0) \big\|_{L^1(\Omega)} + \|G\|_\infty \big\| s^0 \big\|_{L^1(\Omega)}. \tag{3.12}$$

Proof. We proceed as in the proof of [12, Porposition 3.1]. By selecting $v_{\mathcal{T}} = \Upsilon(s_{\mathcal{T}}^k)$ in the formulation (2.13) we see

$$\Big[\!\Big[s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}, \Upsilon(s_{\mathcal{T}}^k) \Big]\!\Big]_{\mathcal{T}} + \triangle t \mathfrak{J}_h^k = 0, \quad \forall k = 1, \cdots, N.$$

Thanks to the convexity of the function $\mathcal{H}(s) + sG$ and the nonnegativity of $\mathfrak{J}_h^k$ we infer

$$\mathfrak{E}_h^k - \mathfrak{E}_h^{k-1} \leq \mathfrak{E}_h^k - \mathfrak{E}_h^{k-1} + \triangle t \mathfrak{J}_h^k \leq \Big[\!\Big[s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}, \Upsilon(s_{\mathcal{T}}^k) \Big]\!\Big]_{\mathcal{T}} + \triangle t \mathfrak{J}_h^k = 0.$$

This proves in particular (3.9). We now apply the Jensen inequality to obtain $|K_i|\, \mathcal{H}(s_i^0) \leq \int_{K_i} \mathcal{H}(s^0)\, \mathrm{d}x$ for all $i \in \mathcal{V}$. Whence, we show (3.10). The inequality (3.11) is a direct consequence of (3.9)-(3.10). The last inequality (3.12) results from (3.7), (3.10) and Assumption **($\mathbf{H}_2$)**. This finishes up the proof. □

Lemma 3.1. *Let $(s_i^k)_{i\in\mathcal{V}}$ be a family fulfilling the numerical scheme* (2.10)-(2.12)*. Then, there exists a positive constant $\eta_{h,\triangle t} > 0$ depending on the mesh parameters such that*

$$s_i^k \geq \eta_{h,\triangle t}, \quad \forall i \in \mathcal{V}, \quad \forall k = 1, \cdots, N. \tag{3.13}$$

Proof. By the mass conservation equality (2.14) and the assumption (1.6), one can find some $i_0 \in \mathcal{V}$ so that one has

$$0 < \frac{1}{|\Omega|} \int_\Omega s^0(x)\, \mathrm{d}x \leq s_{i_0}^k, \quad \text{and} \quad \log(s_{i_0}^k) \geq -C'.$$

As a consequence of (2.7) -(2.8), we get

$$0 < \underline{s_{i_0}^0} := \frac{\lambda^\star}{|\Omega|} \int_\Omega s^0(x)\, \mathrm{d}x \leq \lambda^\star s_{i_0}^k \leq s_{i_0 j}^k, \qquad \forall j \in \mathcal{V}_T \setminus \{i_0\}.$$

Owing to (3.2) we claim for all $T \in \mathcal{T}_{i_0}$, the set of elements sharing the vertex i_0, that

$$\underline{s_{i_0}^0} \Big(\log(s_{i_0}^k) - \log(s_j^k) \Big)^2 \leq s_{i_0 j}^k \Big(\log(s_{i_0}^k) - \log(s_j^k) \Big)^2 \leq C'_{h,\triangle t}, \quad \forall j \in \mathcal{V}_T \setminus \{i_0\}. \tag{3.14}$$

The role of $\lambda^\star$ is to obtain $\underline{s_{i_0}^0} > 0$ independently of s_j^k so that one can make use of [13, Lemma 3.10]. As a result, it is possible to estimate $\log(s_j^k)$ in a straightforward way from (3.14). Note that the inequality (3.14) does not hold true if $s_{i_0 j}^k = s_{i_0 j}^{\ell,k}$ since $s_{i_0 j}^{\ell,k}$ is not far away from 0 with a positive constant independent of s_j^k. This is the reason why we perturb $s_{i_0 j}^k$ to include a very small $s_{i_0 j}^{c,k}$ satisfying the first inequality of (3.14).

We deduce from (3.14) that, $\log(s_j^k) \geq -C''_{h,\triangle t}, \forall j \in \mathcal{V}_T \setminus \{i_0\},\ \forall T \in \mathcal{T}_{i_0}$. It is now sufficient to employ a similar procedure on the next vertex j and so on. This practical argument is therefore applied by induction to the set of mesh vertices, which is finite, see [13, Lemma 3.10]. Accordingly, one ends up with (3.13). □

We now state without proof the existence of discrete positive solutions at each time level. The proof is based on the topological degree argument [21]. The main idea consists in choosing a suitable homotopy function yielding a monotone scheme whose existence is shown by mimicking the guidelines of [23]. Additionally, the energy estimate (3.3) and the positivity estimate (3.13) allow to prove the uniform continuity of the residual function, coming from the finite volume scheme (2.10)-(2.11), on a compact subset of $(\mathbb{R}_*^+)^{\#\mathcal{V}}$. Then, no solution could exist on the boundary of this compact which directly implies the existence of at least one positive solution to the scheme in question. More details can be consulted in [14].

Proposition 3.3. *For every $k = 1, \cdots, N$, there exists at least a positive solution $(s_i^k)_{i\in\mathcal{V}}$ to the finite volume scheme* (2.10)-(2.12)*.*

4 Convergence of the finite volume scheme

In this section we are concerned with the convergence proof of the numerical scheme. To this end, we should establish some compactness arguments. A possible way to do that is to derive uniform estimations on the space and time translations [24, Lemma 4.6]. Instead of using this traditional strategy, we here apply the practical time-compactness criterion recently elaborated in [2]. This ensures the existence of a convergent subsequence of discrete solutions. Furthermore, the limit of the latter has to be identified to a weak solution. We first require the following results.

Lemma 4.1. *(see [24, Lemma 4.2], [27, Lemma 8.2]) For a given $a \in \mathbb{R}^d$, let us set $Q_{t_f,a} = \Omega_a \times (0,t_f)$ with $\Omega_a = \{x \in \Omega / [x,a] \subset \Omega\}$. Then, there exists a positive constant C independent of the discretization parameters such that the following integral satisfies*

$$\int_{Q_{t_f,a}} |\pi_{\mathcal{M},\triangle t} u(x+a,t) - \pi_{\mathcal{M},\triangle t} u(x,t)| \,\mathrm{d}x\,\mathrm{d}t \leq C\,|a| \Big(\sum_{k=1}^{N} \triangle t \left\| \nabla \pi_{\mathcal{T}} u^k \right\|^2_{L^2(\Omega)^d} \Big)^{\frac{1}{2}}, \tag{4.1}$$

for all $\pi_{\mathcal{M},\triangle t} u$ defined by (2.1).

Let us now state and prove an estimate on the dual norm of the discrete time derivative. Let $\psi \in \mathcal{C}_c^\infty(\overline{\Omega} \times [0,t_f))$. In the sequel, we denote $\psi_{\mathcal{T}}^k$ the vector of $\mathbb{R}^{\#\mathcal{V}}$ such that $(\psi_{\mathcal{T}}^k)_i = \psi(x_i, t^k)$ for all $i = 1, \cdots, \#\mathcal{V}$.

Lemma 4.2. *There exists a positive constant that depends only on the physical data and the regularity of the mesh such that*

$$\sum_{k=1}^{N} \left[\!\left[s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}, \psi_{\mathcal{T}}^k \right]\!\right]_{\mathcal{T}} \leq C \left\| \nabla \psi \right\|_\infty, \quad \forall \psi \in \mathcal{C}_c^\infty(\overline{\Omega} \times [0,t_f)). \tag{4.2}$$

Proof. We consider $v_{\mathcal{T}} = \psi_{\mathcal{T}}^k$ in the concise form of the scheme (2.13). Then, we obtain the splitting

$$X + Y = 0,$$

where

$$X = \sum_{k=1}^{N} \left[\!\left[s_{\mathcal{T}}^k - s_{\mathcal{T}}^{k-1}, \psi_{\mathcal{T}}^k \right]\!\right]_{\mathcal{T}},$$
$$Y = \sum_{k=1}^{N} \triangle t \sum_{T \in \mathcal{T}} \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \psi_{\mathcal{T}}^k.$$

Applying the Cauchy-Schwarz inequality on the last term, making use of 2.15, employing the smoothness of the test function ψ and (3.1) yield

$$\begin{aligned} |Y| &\leq \left(\sum_{k=1}^{N} \triangle t \sum_{T \in \mathcal{T}} \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \Upsilon(s_{\mathcal{T}}^k) \right)^{1/2} \left(\sum_{k=1}^{N} \triangle t \sum_{T \in \mathcal{T}} \mathcal{S}^T \delta_T \psi_{\mathcal{T}}^k \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \psi_{\mathcal{T}}^k \right)^{1/2} \\ &\leq C \left\| \nabla \psi \right\|_\infty \left(\sum_{T \in \mathcal{T}} |T| \sum_{i \in \mathcal{V}_T} s_{i\tau(i)}^k \right)^{1/2} \leq C \left\| \nabla \psi \right\|_\infty. \end{aligned}$$

This concludes the proof of the required inequality. □

Proposition 4.1. *We assume that* ***(H$_1$)-(H$_3$)*** *are fulfilled. Let $(\mathcal{T}_m)_m$ be a sequence of refined meshes to $\mathcal{T}$ such that $h_{\mathcal{T}_m}, \triangle t_m$ go to 0 as m tends to infinity. Then, there exists a measurable function $s : Q_{t_f} \longrightarrow (0, +\infty)$ so that the following convergences hold up to a subsequence*

$$\pi_{\mathcal{M}_m,\triangle t_m} s,\ \pi_{\mathcal{T}_m,\triangle t_m} s \longrightarrow s \qquad \text{a.e. in } Q_{t_f},\ \text{and strongly in } L^1(Q_{t_f}), \tag{4.3}$$

$$\nabla \pi_{\mathcal{T}_m,\triangle t_m} \xi(s) \longrightarrow \nabla \xi(s) \qquad \text{weakly in } L^2(0, t_f; H^1(\Omega)). \tag{4.4}$$

Proof. According to the energy estimate (3.3) we have the uniform boundedness in $L^2(Q_{t_f})^d$ of the sequence $(\nabla \pi_{\mathcal{T}_m,\triangle t_m} \xi(s))_m$. Then, there exists a function $\zeta \in L^2(Q_{t_f})^d$ such that

$$\nabla \pi_{\mathcal{T}_m,\triangle t_m} \xi(s) \longrightarrow \zeta, \quad \text{weakly in } L^2(Q_{t_f})^d, \quad as \quad m \longrightarrow +\infty. \tag{4.5}$$

Next, the mass conservation property gives a uniform estimation on the sequence $(\pi_{\mathcal{M}_m,\triangle t_m} \xi(s))_m$ in $L^2(Q_{t_f})$ where we recall that $\xi(s) = \sqrt{s}$. Thanks to Lemma 2.1, we infer that $(\pi_{\mathcal{T}_m,\triangle t_m} \xi(s))_m$ is uniformly bounded in the same space. Now, by virtue of Lemma 4.1 and Lemma 4.2 we apply the compactness argument [2, Theorem 3.9] to ensure the existence of a subsequence converging a.e. to a measurable function $s : Q_{t_f} \longrightarrow (0, +\infty)$, i.e.

$$\pi_{\mathcal{M}_m,\triangle t_m} s \longrightarrow s \quad \text{a.e. in } Q_{t_f}, \quad as \quad m \longrightarrow +\infty.$$

We recall that

$$\left\| \mathcal{H}(\pi_{\mathcal{M}_m,\triangle t_m} s) \right\|_{L^1(Q_{t_f})} \leq C.$$

Then, the sequence $(\pi_{\mathcal{M}_m,\triangle t_m} s)$ is equi-integrable thanks to De La Vallée Poussin criterion [8]. Consequently, Vitali's convergence theorem claims the strong convergence of the above sequence i.e.

$$\pi_{\mathcal{M}_m,\triangle t_m} s \longrightarrow s \quad \text{strongly in } L^1(Q_{t_f}), \quad as \quad m \longrightarrow +\infty.$$

In view of Lemma 2.1, we deduce

$$\pi_{\mathcal{T}_m,\triangle t_m} s = \pi_{\mathcal{T}_m,\triangle t_m} \xi^2(s) \longrightarrow s \quad \text{strongly in } L^1(Q_{t_f}) \text{ and a.e. in } Q_{t_f}, \quad as \quad m \longrightarrow +\infty.$$

To conclude the proof, it suffices to employ the standard identification of the limit process to check that $\zeta = \nabla \xi(s)$ in the sense of distributions. □

Now, we are in a position to identify the limit function of the previous result as a weak solution.

Proposition 4.2. *Keeping the statement of Proposition 4.1, the obtained function s is indeed a weak solution to the continuous model* (1.1)-(1.3) *in the sense of Definition 1.1.*

Proof. Following [12, Proposition 4.4], it can be checked that $\mathcal{H}(s)$ belongs to $L^\infty(0, t_f; L^1(\Omega))$. Let us now take $v^k_{\mathcal{T}_m} = \psi^k_{\mathcal{T}_m}$ in the discrete weak formulation (2.13) and sum on $k = 1, \cdots, N$. Then, the latter summation is decomposed as follows

$$X_m + Y_m + Z_m = 0. \tag{4.6}$$

$$X_m = \sum_{k=1}^{N} \left[\!\left[s_{\mathcal{T}_m}^k - s_{\mathcal{T}_m}^{k-1}, \psi_{\mathcal{T}_m}^k \right]\!\right]_{\mathcal{T}_m},$$

$$Y_m = \sum_{k=1}^{N} \triangle t_m \sum_{T\in\mathcal{T}_m} \mathcal{S}^T \delta_T \log(s_{\mathcal{T}_m}^k) \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \psi_{\mathcal{T}_m}^k,$$

$$Z_m = \sum_{k=1}^{N} \triangle t_m \sum_{T\in\mathcal{T}_m} \mathcal{S}^T \delta_T G_{\mathcal{T}_m} \cdot \mathcal{A}^T \mathcal{S}^T \delta_T \psi_{\mathcal{T}_m}^k.$$

Notice that $\psi_{\mathcal{T}_m}^N = 0_{\mathbb{R}^{\#\mathcal{V}_m}}$. Performing a discrete integration by parts in time implies

$$\begin{aligned} X_m &= -\left[\!\left[s_{\mathcal{T}_m}^0, \psi_{\mathcal{T}_m}^0 \right]\!\right]_{\mathcal{T}_m} - \sum_{k=1}^{N} \triangle t_m \left[\!\left[s_{\mathcal{T}_m}^{k-1}, \frac{\psi_{\mathcal{T}_m}^k - \psi_{\mathcal{T}_m}^{k-1}}{\triangle t_m} \right]\!\right]_{\mathcal{T}_m}, \\ &= -\int_{Q_{t_f}} \pi_{\mathcal{M}_m,\triangle t_m} s(\cdot, -\triangle t_m)\; D\pi_{\mathcal{M}_m,\triangle t_m}\psi \,\mathrm{d}x\,\mathrm{d}t - \int_\Omega s^0\; \pi_{\mathcal{M}_m,\triangle t_m}\psi(x,0)\,\mathrm{d}x. \end{aligned}$$

where $D\pi_{\mathcal{M}_m,\triangle t_m}\psi$ is the discrete time derivative given by

$$D\pi_{\mathcal{M}_m,\triangle t_m}\psi(x,t) = \frac{\psi_i^k - \psi_i^{k-1}}{\triangle t_m}, \quad \forall (x,t) \in K_i \times (t^{k-1}, t^k],\ \forall i \in \mathcal{V}, \forall k \geq 1.$$

It follows from the smoothness of the test function ψ that $(D\pi_{\mathcal{M}_m,\triangle t_m}\psi)_m$ (resp. $(\pi_{\mathcal{M}_m,\triangle t_m}\psi(\cdot,0))_m$) converges uniformly towards ψ_t (resp. $\psi(\cdot,0)$). Thanks to (4.3) we get

$$X_m = -\int_{Q_{t_f}} s\;\psi \,\mathrm{d}x\,\mathrm{d}t - \int_\Omega s^0\psi(x,0)\,\mathrm{d}x, \quad as \quad m \longrightarrow +\infty.$$

Let us now study the convergence of the diffusive part Y_m. To this purpose, we first consider

$$\begin{aligned} Y_m^* &= 2\sum_{k=1}^{N} \triangle t_m \sum_{T\in\mathcal{T}_m} \xi(s_{T,min}^k)\delta_T\xi(s_{\mathcal{T}_m}^k) \cdot \mathcal{A}^T \delta_T \psi_{\mathcal{T}_m}^k, \\ &= 2\int_{Q_{t_f}} \widehat{\pi_{\mathcal{T}_m,\triangle t_m}\xi}(s) \nabla \pi_{\mathcal{T}_m,\triangle t_m}\xi(s) \cdot \nabla \pi_{\mathcal{T}_m,\triangle t_m}\psi \,\mathrm{d}x\,\mathrm{d}t, \end{aligned}$$

where $s_{T,min}^k = \min_{i\in\mathcal{V}_T}\{s_i^k\}$ and we have defined

$$\widehat{\pi_{\mathcal{T}_m,\triangle t_m}\xi}(s) = \xi(s_{T,min}^k), \quad \forall (x,t) \in T \times (t^{k-1}, t^k],\ \forall T \in \mathcal{T}_m, \forall k \geq 1.$$

Using Lemma 2.1 we estimate

$$\begin{aligned} \left\| \widehat{\pi_{\mathcal{T}_m,\triangle t_m}\xi}(s) - \pi_{\mathcal{T}_m,\triangle t_m}\xi(s) \right\|_{L^2(Q_{t_f})} &\leq \left\| \overline{\pi_{\mathcal{T}_m,\triangle t_m}\xi(s)} - \underline{\pi_{\mathcal{T}_m,\triangle t_m}\xi(s)} \right\|_{L^2(Q_{t_f})} \\ &\leq h_{\mathcal{T}_m} C \left(\sum_{k=1}^{N} \triangle t_m \left\| \nabla \pi_{\mathcal{T}_m}\xi(s^k) \right\|_{L^2(\Omega)^d} \right)^{\frac{1}{2}} \\ &\leq C h_{\mathcal{T}_m} \longrightarrow 0, \quad as \quad m \longrightarrow +\infty. \end{aligned}$$

We deduce in particular that

$$\widehat{\pi_{\mathcal{T}_m,\triangle t_m}\xi}(s) \longrightarrow \xi(s) \quad \text{strongly in } L^2(Q_{t_f}), \quad as \quad m \longrightarrow +\infty. \tag{4.7}$$

Thereby, we can pass to the limit in Y_m^* to obtain

$$Y_m^* \longrightarrow Y = 2\int_{Q_{t_f}} \sqrt{s}\nabla\sqrt{s}\cdot\nabla\psi \,\mathrm{d}x\,\mathrm{d}t = \int_{Q_{t_f}} \nabla s\cdot\nabla\psi \,\mathrm{d}x\,\mathrm{d}t, \quad as \quad m \longrightarrow +\infty.$$

We furthermore define the local diagonal matrix $(\mathfrak{S}_{ij}^T)_{1\leq i,j\leq \#\mathcal{V}_T}$ such that

$$\mathfrak{S}_{ij}^T = \begin{cases} \sqrt{s_{T,min}^k} & \text{if } j=\tau(i) \\ 0 & \text{if } j\neq\tau(i) \end{cases}.$$

We denote $\mathcal{S}^{T,2}$ the diagonal matrix corresponding to the second choice of (2.6), namely the logarithmic mean. Note that Y_m^* can be rewritten under the form

$$Y_m^* = \sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \mathcal{S}^{T,2}\delta_T \log(s_{\mathcal{T}_m}^k)\cdot\mathcal{A}^T\mathfrak{S}^T\delta_T\psi_{\mathcal{T}_m}^k = \sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \delta_T\xi(s_{\mathcal{T}_m}^k)\cdot\mathcal{A}^T\mathfrak{S}^T\delta_T\psi_{\mathcal{T}_m}^k.$$

Let us moreover set

$$Y_m^{**} = \sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \mathcal{S}^T\delta_T \log(s_{\mathcal{T}_m}^k)\cdot\mathcal{A}^T\mathfrak{S}^T\delta_T\psi_{\mathcal{T}_m}^k.$$

Observe that $s_{T,min}^k \leq s_{i\tau(i)}^k$, for every $i\in\mathcal{V}_T$. Also, the regularity of ψ gives $|\psi_i^k - \psi_{\tau(i)}^k| \leq \|\nabla\psi\|_\infty h_T$ for all $i\in\mathcal{V}_T$ and $k\geq 0$. We subtract the last two identities. We utilize once more the Cauchy-Schwarz inequality, the result (2.15), Proposition 3.1, and Lemma2.1 to find that

$$\begin{aligned}
|Y_m^* - Y_m^{**}| &= \left| \sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \mathfrak{S}^T\delta_T \log(s_{\mathcal{T}_m}^k)\cdot(\mathcal{S}^{T,2}-\mathcal{S}^T)\mathcal{A}^T\delta_T\psi_{\mathcal{T}_m}^k \right| \\
&\leq C\,\|\nabla\psi\|_\infty \left(\sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \frac{|T|}{h_T^2} \sum_{i\in\mathcal{V}_T} s_{i\tau(i)}^k \Big(\log(s_i^k) - \log(s_{\tau(i)}^k) \Big)^2 \right)^{\frac{1}{2}} \\
&\qquad \times \left(\sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} |T| \left\| \mathcal{S}^{T,2} - \mathcal{S}^T \right\|_\infty^2 \right)^{\frac{1}{2}} \\
&\leq C_\psi \left\| \overline{\pi_{\mathcal{T}_m,\triangle t_m}\xi(s)} - \underline{\pi_{\mathcal{T}_m,\triangle t_m}\xi(s)} \right\|_{L^2(Q_{t_f})} \\
&\leq C_\psi h_{\mathcal{T}_m}, \quad as \quad m\longrightarrow+\infty.
\end{aligned}$$

Similarly, one shows that $|Y_m - Y_m^{**}| \longrightarrow 0$, *as* $m\longrightarrow+\infty$. To conclude, it remains to treat the convective potential term Z_m. Mimicking an analogous reasoning we consider

$$\begin{aligned}
Z_m^* &= \sum_{k=1}^N \triangle t_m \sum_{T\in\mathcal{T}_m} \mathfrak{S}^T\delta_T G_{\mathcal{T}_m}\cdot\mathcal{A}^T\mathfrak{S}^T\delta_T\psi_{\mathcal{T}_m}^k \\
&= \int_{Q_{t_f}} \widehat{\pi_{\mathcal{T}_m,\triangle t_m}\xi}(s)^2 \,\nabla\pi_{\mathcal{T}_m}G\cdot\nabla\pi_{\mathcal{T}_m,\triangle t_m}\psi \,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

The regularity of the potential G entails the uniform convergence of the sequence $(\nabla \pi_{\mathcal{T}_m} G)_m$ towards ∇G in $L^2(Q_{t_f})^d$. Thanks to the strong convergence (4.7) we get

$$Z_m^* \longrightarrow Z = \int_{Q_{t_f}} s \nabla G \cdot \nabla \psi \, \mathrm{d}x \, \mathrm{d}t, \quad as \quad m \longrightarrow +\infty.$$

We continue in the same fashion by setting

$$Z_m^{**} = \sum_{k=1}^{N} \triangle t_m \sum_{T \in \mathcal{T}_m} \mathcal{S}^T \delta_T G_{\mathcal{T}_m} \cdot \mathcal{A}^T \mathfrak{S}^T \delta_T \psi_{\mathcal{T}_m}^k.$$

Thanks to the smoothness of G we have $\left|G(x_i) - G(x_{\tau(i)})\right| \leq \|\nabla G\|_\infty h_T$ for all $i \in \mathcal{V}_T$ and $k \geq 0$. We follow the previous arguments together with (3.1) to discover

$$\begin{aligned}
|Z_m^* - Z_m^{**}| &= \left| \sum_{k=1}^{N} \triangle t_m \sum_{T \in \mathcal{T}_m} (\mathfrak{S}^T - \mathcal{S}^T) \delta_T G_{\mathcal{T}_m} \cdot \mathcal{A}^T \mathfrak{S}^T \delta_T \psi_{\mathcal{T}_m}^k \right| \\
&\leq C \|\nabla \psi\|_\infty \left(\sum_{k=1}^{N} \triangle t_m \sum_{T \in \mathcal{T}_m} |T| \left\| \mathfrak{S}^T - \mathcal{S}^T \right\|_\infty^2 \right)^{\frac{1}{2}} \\
&\qquad \times \left(\sum_{k=1}^{N} \triangle t_m \sum_{T \in \mathcal{T}_m} \frac{|T|}{h_T^2} \sum_{i \in \mathcal{V}_T} s_{T,min}^k \Big(G(x_i) - G(x_{\tau(i)}) \Big)^2 \right)^{\frac{1}{2}} \\
&\leq C_\psi \|\nabla G\|_\infty \left\| \overline{\pi_{\mathcal{T}_m, \triangle t_m} \xi(s)} - \underline{\pi_{\mathcal{T}_m, \triangle t_m} \xi(s)} \right\|_{L^2(Q_{t_f})} \left(\sum_{k=1}^{N} \triangle t_m \sum_{T \in \mathcal{T}_m} |T| \sum_{i \in \mathcal{V}_T} s_{i\tau(i)}^k \right)^{\frac{1}{2}} \\
&\leq C_{\psi, G} h_{\mathcal{T}_m}, \quad as \quad m \longrightarrow +\infty.
\end{aligned}$$

We adapt the above procedure and the same steps to establish that $|Z_m - Z_m^{**}| \longrightarrow 0$, $as\ m \longrightarrow +\infty$. Hence, the proof is complete. □

5 Numerical tests

The goal of this section is to validate the efficiency and the robustness of our methodology through several examples including anisotropic and heterogeneous tensors. We solve the algebraic system (2.10)-(2.11) issuing from the numerical scheme by the Newton method that we implemented in Matlab. The Newton method computes a family of iterates$(s_{\mathcal{T}}^{k-1,\eta})_{\eta \geq 0}$. To avoid the singularity of the logarithm, we reinforce the initial guess to be positive as follows $s_{\mathcal{T}} = \max(s_{\mathcal{T}}, 10^{-10})$. In the case of its convergence, it tends to the sought solution $s_{\mathcal{T}}^k$. Fixing the maximum number of iterations as 20, the stopping criterion is made on the difference of the successive iterates in the discrete ℓ^2-norm i.e., the procedure stops when it yields

$$\left\| s_{\mathcal{T}}^{k,\eta+1} - s_{\mathcal{T}}^{k,\eta} \right\|_{\ell^2} \leq \varepsilon,$$

where we set $\varepsilon = 10^{-10}$.

Here we consider $\Omega = (0,1)^2$ the computational domain. It is covered by a family of refined triangular meshes coming from benchmarking problems [30]. We refer to Figure 2 for a depiction

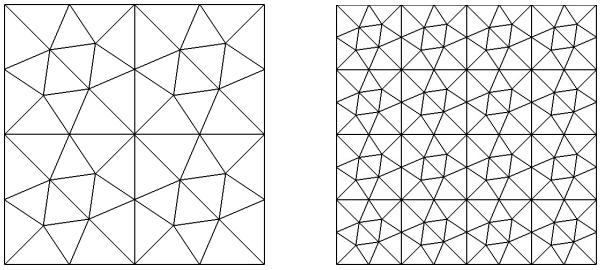

Figure 2: First two meshes of the considered triangulation.

of this triangulation. In order to identify in a straightforward way the analytical solution to the continuous problem, we take a diagonal tensor $\mathbb{D}$

$$\mathbb{D}(x,y) = \begin{pmatrix} \mathbb{D}_{xx}(x,y) & 0 \\ 0 & \mathbb{D}_{yy}(x,y) \end{pmatrix}.$$

To evaluate the convergence speed of our approach, we calculate the numerical errors between the computed solution and the analytical one using the following norms

$$\|\pi_{\mathcal{M},\triangle t} s - s_e\|_r = \|\pi_{\mathcal{M},\triangle t} s - s_e\|_{L^r(Q_{t_f})}, \quad r = 2, \infty.$$

For clarity, in all the tests below, the logarithmic-mean (log-mean) corresponds to $s_{i\tau(i)}$ defined in (2.7) when the parameter λ is assigned to the machine precision. That is, $\lambda = 10^{-16}$ so that $s_{i\tau(i)}$ and $s^{\ell}_{i\tau(i)}$ can be numerically identical. The arithmetic-mean is obtained for $\lambda = 1$.

5.1 Accuracy of the scheme

In this subsection we test the efficiency and the robustness of our scheme using a known exact solution to the continuous problem. We first consider a constant tensor $\mathbb{D}$. Then, the analytical solution reads

$$s_e(x,y,t) = \big(1 + \cos(\pi y) e^{-ct}\big) e^x,$$

where the final time is fixed to $t_f = 0.2$ and $c = \pi^2 \mathbb{D}_{yy}$. The initial condition is then $s^0 = s_e(\cdot, 0)$. This state degenerates on the segment $[0,1] \times \{1\}$. The exterior potential is set to $G(x,y) = -x$. The second eigenvalue of $\mathbb{D}$ is taken as $\mathbb{D}_{yy}(x,y) = 1$. We also compare both formulas of the mobility function given in (2.6) in terms of the anisotropy ratio of the tensor $\mathbb{D}$. The first expression i.e. $s^{c}_{i\tau(i)}$ will referred to as arithmetic-mean while the second one i.e. $s^{\ell}_{i\tau(i)}$ will be called log-mean. In this example the time step is proportional to the size of the mesh i.e. $\triangle t / h^2 = 0.2$. A summary of the time-space discretization is given in Table 1.

Mesh	$\#\mathcal{V}$	$h_{\mathcal{T}}$	$\triangle t$
$\mathbf{Tri}_1$	37	0.250	0.125 E-01
$\mathbf{Tri}_2$	129	0.125	0.312 E-02
$\mathbf{Tri}_3$	481	0.063	0.793 E-03
$\mathbf{Tri}_4$	1857	0.031	0.192 E-03
$\mathbf{Tri}_5$	7297	0.016	0.512 E-04

Table 1: Discretization data used to assess the accuracy of the scheme.

We report in the following tables the obtained results. Each table displays, the errors as well as their convergence rates, the minimum (s_{min}) of the computed solution $\pi_{\mathcal{M}} s^k$ for $k \geq 1$, and the number of iterations of the Newton solver, that we denote by Niter.

In Tables 2-7, we observe that the second-order accuracy of the method is achieved for several anisotropic ratios less than 10 even on coarse meshes. This fact holds for both the log-mean and the arithmetic mean of the mobility-like function. We also record that the solution remains positive as established in Lemma 3.1. When the tensor becomes strongly anisotropic $\mathbb{D}_{xx} \geq 100$, the underlined features are preserved for the log-mean case as provided in Tables 8-9. There is no huge difference in the number of Newton's iterations in the weak anisotropic case for the both averages. However, the arithmetic-mean choice turns out to break down the iterative solver. In order to converge, it necessitates a stringent constraint on the time step, that is $\triangle t/h^2$ should be less than $1/\mathbb{D}_{xx}$, which allows extremely small time steps and makes the computational cost exorbitant.

The Newton solver corresponding to the arithmetic-mean choice fails to find the discrete solution when the anisotropic ratio is important. It produces spurious oscillations on the first iterates. This behavior is not recorded on the logarithmic-mean version of the solver. Reducing reasonably the time step or cutting the oscillations do not help to overpass the observed issue. In Figure 3, we plot the ℓ^2-norm of the Newton residual function in terms of $\lambda \in (0, 1)$ at the first time iteration ($t^1 = \triangle t = 0.0031$) on the second mesh and for $\mathbb{D}_{xx} = 100, 1000$. We observe that the norm augments as this parameter increases, meaning that we get close to the arithmetic-centered scheme. Through several experiments we also noticed that the norm of the residual function blows up when $\lambda > 0.7$.

Mesh	$\|\pi_{\mathcal{M},\triangle t} s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t} s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.284 E-01	-	0.242 E-00	-	0.16 E-00	36
$\mathbf{Tri}_2$	0.709 E-02	2.004	0.687 E-01	1.817	0.44 E-01	107
$\mathbf{Tri}_3$	0.181 E-02	1.991	0.183 E-01	1.926	0.11 E-01	385
$\mathbf{Tri}_4$	0.441 E-03	1.989	0.464 E-02	1.939	0.27 E-02	1284
$\mathbf{Tri}_5$	0.116 E-03	2.012	0.126 E-02	1.971	0.73 E-03	4038

Table 2: Numerical errors with the log-mean choice of the mobility and $\mathbb{D}_{xx} = 0.1$.

Mesh	$\|\pi_{\mathcal{M},\triangle t} s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t} s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.157 E-01	-	0.159 E-00	-	0.11 E-00	45
$\mathbf{Tri}_2$	0.432 E-02	1.861	0.467 E-01	1.769	0.31 E-01	116
$\mathbf{Tri}_3$	0.113 E-02	1.953	0.126 E-01	1.901	0.81 E-02	395
$\mathbf{Tri}_4$	0.278 E-03	1.979	0.325 E-02	1.919	0.19 E-02	1292
$\mathbf{Tri}_5$	0.735 E-04	2.013	0.902 E-03	1.939	0.52 E-03	4027

Table 3: Numerical errors with the arithmetic-mean choice of the mobility and $\mathbb{D}_{xx} = 0.1$.

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.282 E-01	-	0.243 E-00	-	0.16 E-00	34
$\mathbf{Tri}_2$	0.700 E-02	2.012	0.668 E-01	1.863	0.43 E-01	105
$\mathbf{Tri}_3$	0.178 E-02	1.994	0.174 E-01	1.962	0.11 E-01	385
$\mathbf{Tri}_4$	0.434 E-03	1.992	0.429 E-02	1.973	0.27 E-02	1302
$\mathbf{Tri}_5$	0.115 E-03	2.009	0.113 E-02	2.015	0.72 E-03	4045

Table 4: Numerical errors with the log-mean choice of the mobility and $\mathbb{D}_{xx} = 1$.

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.148 E-01	-	0.143 E-00	-	0.112 E-00	45
$\mathbf{Tri}_2$	0.409 E-02	1.860	0.412 E-01	1.802	0.308 E-01	116
$\mathbf{Tri}_3$	0.107 E-02	1.956	0.108 E-01	1.950	0.798 E-02	395
$\mathbf{Tri}_4$	0.262 E-03	1.982	0.269 E-02	1.965	0.194 E-02	1309
$\mathbf{Tri}_5$	0.696 E-04	2.008	0.709 E-03	2.016	0.515 E-03	4037

Table 5: Numerical errors with the arithmetic-mean choice of the mobility and $\mathbb{D}_{xx} = 1$.

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.269 E-01	-	0.241 E-00	-	0.145 E-00	37
$\mathbf{Tri}_2$	0.660 E-02	2.026	0.660 E-01	1.869	0.401 E-01	108
$\mathbf{Tri}_3$	0.167 E-02	1.998	0.170 E-01	1.978	0.106 E-01	389
$\mathbf{Tri}_4$	0.407 E-03	1.996	0.428 E-02	1.945	0.262 E-02	1421
$\mathbf{Tri}_5$	0.108 E-03	2.004	0.112 E-02	2.021	0.706 E-03	4140

Table 6: Numerical errors with the log-mean choice of the mobility and $\mathbb{D}_{xx} = 10$.

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.120 E-01	-	0.132 E-00	-	0.108 E-00	46
$\mathbf{Tri}_2$	0.342 E-02	1.820	0.379 E-01	1.802	0.299 E-01	117
$\mathbf{Tri}_3$	0.901 E-03	1.947	0.980 E-02	1.975	0.775 E-02	397
$\mathbf{Tri}_4$	0.220 E-03	1.983	0.253 E-02	1.910	0.188 E-02	1429
$\mathbf{Tri}_5$	0.587 E-04	2.001	0.660 E-03	2.031	0.502 E-03	4132

Table 7: Numerical errors with the arithmetic-mean choice of the mobility and $\mathbb{D}_{xx} = 10$.

5.2 Large time behavior

This example focuses on the asymptotic behavior of our numerical scheme when time becomes large. It is inspired from [12]. We keep the same data of the first test except $\mathbb{D}$ and t_f. According

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.315 E-01	-	0.380 E-00	-	0.121 E-00	42
$\mathbf{Tri}_2$	0.753 E-02	2.065	0.102 E-00	1.893	0.317 E-01	116
$\mathbf{Tri}_3$	0.186 E-02	2.039	0.269 E-01	1.952	0.855 E-02	394
$\mathbf{Tri}_4$	0.454 E-03	1.989	0.684 E-02	1.930	0.216 E-02	1573
$\mathbf{Tri}_5$	0.117 E-03	2.042	0.180 E-02	2.014	0.596 E-03	4363

Table 8: Numerical errors with the log-mean choice of the mobility and $\mathbb{D}_{xx} = 100$.

Mesh	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_2$	Rate	$\|\pi_{\mathcal{M},\triangle t}s - s_e\|_\infty$	Rate	s_{min}	Niter
$\mathbf{Tri}_1$	0.403 E-01	-	0.494 E-00	-	0.113 E-00	45
$\mathbf{Tri}_2$	0.937 E-02	2.107	0.128 E-00	1.944	0.265 E-01	124
$\mathbf{Tri}_3$	0.227 E-02	2.067	0.337 E-01	1.951	0.676 E-02	401
$\mathbf{Tri}_4$	0.559 E-03	1.978	0.921 E-02	1.830	0.168 E-02	1578
$\mathbf{Tri}_5$	0.141 E-03	2.079	0.249 E-02	1.978	0.482 E-03	4461

Table 9: Numerical errors with the log-mean choice of the mobility and $\mathbb{D}_{xx} = 1000$.

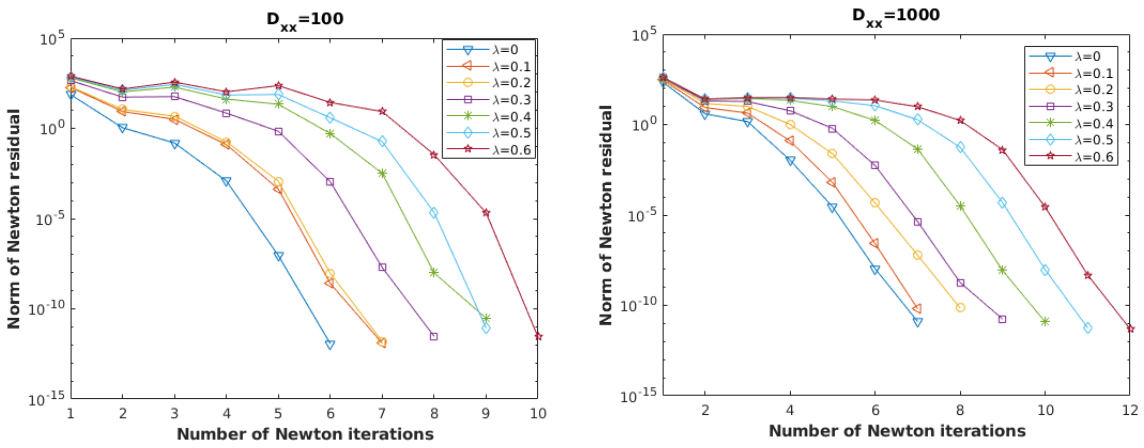

Figure 3: Impact of the parameter λ on the ℓ^2-norm of the Newton residual function for $\mathbb{D}_{xx} = 100$ (left) and $\mathbb{D}_{xx} = 1000$ (right) using the second mesh at the first time iteration $t^1 = \triangle t = 0.0031$.

to [3], the exact solution converges to the following stationary-state as the time tends to infinity

$$s^\infty = \Big(\int_\Omega s^0 \, \mathrm{d}x / \int_\Omega e^{-G} \, \mathrm{d}x \Big) e^{-G}.$$

Therefore, the discrete counterpart of the steady-state solution $s_\mathcal{T}^\infty = (s_i^\infty)_{i \in \mathcal{V}}$ reads

$$s_i^\infty = \Big(\int_\Omega s^0 \, \mathrm{d}x / \sum_{j \in \mathcal{V}} |K_j| \, e^{-G(x_j, y_j)} \Big) e^{-G(x_i, y_i)}.$$

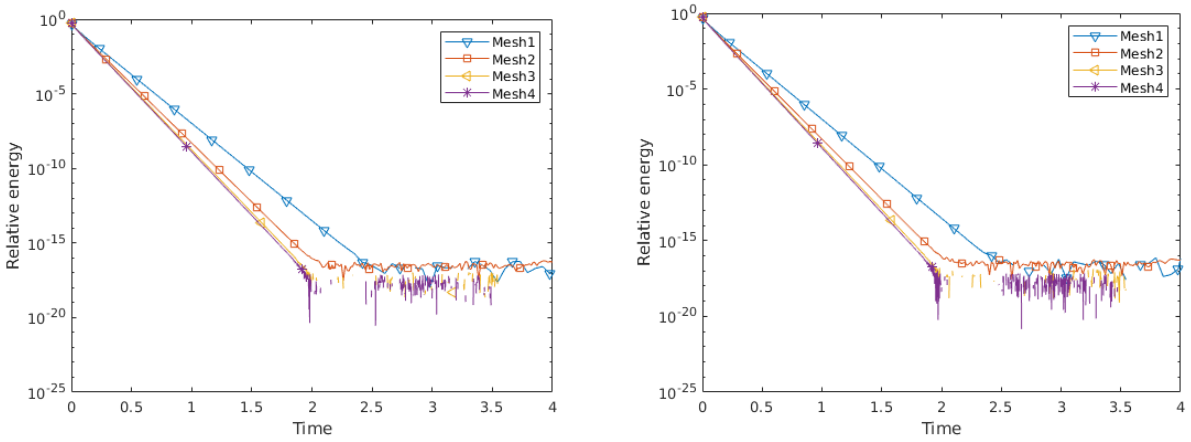

Figure 4: Relative free energy with the log-mean choice of the mobility (left) and the arithmetic-mean (right) for $\mathbb{D}_{xx} = \mathbb{D}_{yy} = 1$, and $t_f = 4$ on the first four meshes.

To illustrate the long time behavior of the proposed approach we look at the evolution of the discrete relative energy given as follows

$$\mathfrak{E}_h^k - \mathfrak{E}_h^\infty = \left[\!\left[s_{\mathcal{T}}^k \log(s_{\mathcal{T}}^k / s_{\mathcal{T}}^\infty) - s_{\mathcal{T}}^k + s_{\mathcal{T}}^\infty, \mathbf{1}_{\mathcal{T}} \right]\!\right]_{\mathcal{T}}, \quad \forall k = 0, \cdots, N.$$

We represent on Figure 4-5 this quantity as a function of time for the log-mean and arithmetic-mean approximations of mobility in the isotropic and the anisotropic cases. The first plot addresses the isotropic case with $\mathbb{D}_{xx} = \mathbb{D}_{yy} = 1$ and $t_f = 4$. It shows that the relative free energy is rapidly diminishing towards 0. The second plot concerns the anisotropic case $\mathbb{D}_{xx} = 1$ and $\mathbb{D}_{yy} = 0.05$ with $t_f = 55$. We also observe the exponential decreasing of the free energy towards its stationary state. This requires more time, which is due to the impact of the tensor anisotropy.

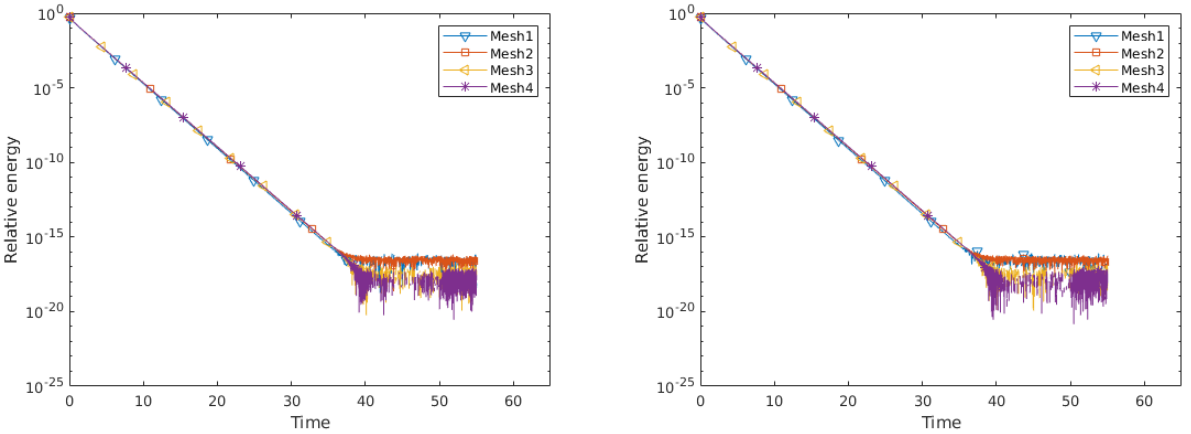

Figure 5: Relative free energy with the log-mean choice of the mobility (left) and the arithmetic-mean (right) for $\mathbb{D}_{xx} = 1$, $\mathbb{D}_{yy} = 0.05$, and $t_f = 55$ on the first four meshes.

5.3 Heterogeneous five-spot problem

The objective of this last test is to exhibit the applicability of our scheme to a situation where the anisotropy is dependent on the position. This example is termed the five-spot problem in the context of porous media flows. Let us set $\Omega = \Omega_1 \cup \Omega_2$ such that $\Omega_1 = \{(x, y) \in \Omega / x < 0.5\}$ and $\Omega_2 = \Omega \setminus \overline{\Omega}_1$. Let us next consider the heterogeneous tensor

$$\mathbb{D}(x, y) = \begin{cases} \begin{pmatrix} 0.5 & 0 \\ 0 & 0.1 \end{pmatrix} & \text{if } (x, y) \in \Omega_1 \\ \\ \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix} & \text{if } (x, y) \in \Omega_2 \end{cases}.$$

The domain is initially empty, that is $s^0 = 0$. The potential is chosen as $G(x, y) = x$. We impose a Dirichlet boundary condition $s_{|\partial\Omega^D} = 1$ on the left portion $\partial\Omega^D = \{x = 0\} \times \{0 \leq y \leq 0.5\}$ and an outflow condition on the right zone located at $\{x = 1\} \times \{0.5 \leq y \leq 1\}$. The remainder border is assumed impervious. For the simulation, we use the third mesh of the family depicted in Figure 2. We consider the fixed time step $\triangle t = 0.00089$. The final simulation time is set to $t_f = 0.1$. The obtained results are plotted on Figure 6-7. The first figure shows the solution computed accurately using the log-mean approximation of the mobility while the second one exhibits the solution of the scheme using the arithmetic mean. Both approaches produce the expected behavior. It is observed that the medium Ω_1 is more permeable than Ω_2, which is due the the contrast of the intrinsic permeabilities. We moreover remark that our scheme generates no undershoots. These facts are finally confirmed by the cross section of the solution at the point $(0, 0.4)$ illustrated in Figure 8. To sum up, both strategies behave similarly as the plotted solutions show, since the tensor is weakly homogeneous in each subdomain and the ratio is small.

References

[1] H. W. Alt and S. Luckhaus. Quasilinear elliptic-parabolic differential equations. *Mathematische Zeitschrift*, 183(3):311–341, 1983.

[2] 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.

[3] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Communications in Partial Differential Equations*, 26(1-2):43–100, 2001.

[4] R. E. Bank, D. J. Rose, and W. Fichtner. Numerical methods for semiconductor device simulation. *SIAM Journal on Scientific and Statistical Computing*, 4(3):416–435, 1983.

[5] M. Bessemoulin-Chatard. A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme. *Numerische Mathematik*, 121(4):637–670, 2012.

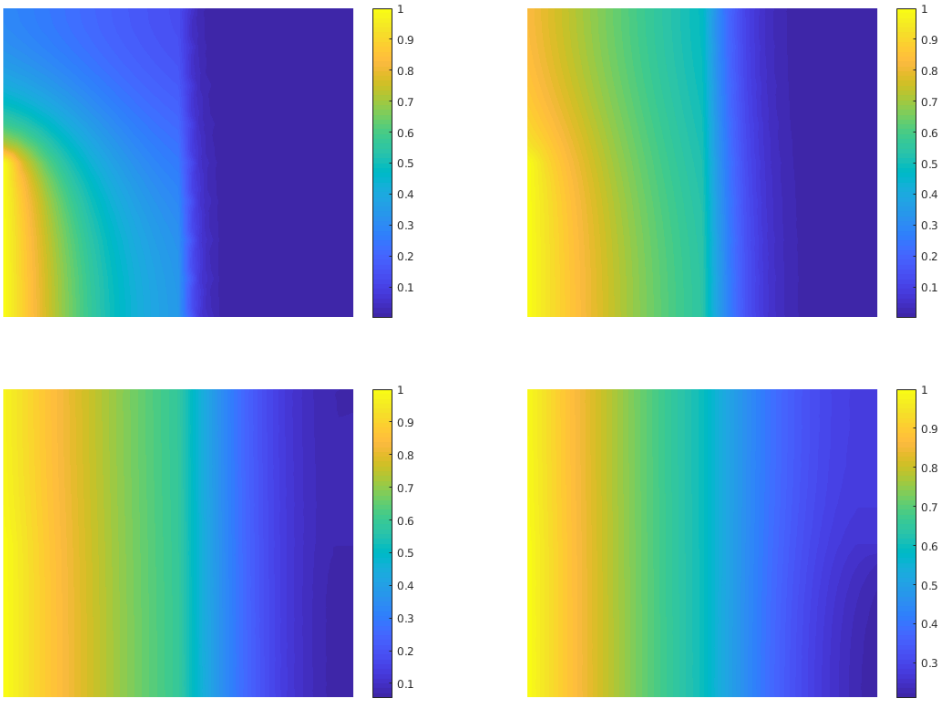

Figure 6: From left-up to right-bottom, the log-mean approximate solution for different simulation times $t \in \{0.25, 1, 4, 10\}$.

[6] M. Bessemoulin-Chatard. *Développement et analyse de schémas volumes finis motivés par la présentation de comportements asymptotiques. Application à des modèles issus de la physique et de la biologie*. PhD thesis, 2012.

[7] M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems. *Journal of Numerical Mathematics*, 25(3):147–168, 2017.

[8] V. I. Bogachev. *Measure theory*, volume 1. Springer Science & Business Media, 2007.

[9] 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.

[10] J.-S. Camier and F. Hermeline. A monotone nonlinear finite volume method for approximating diffusion operators on general meshes. *International Journal for Numerical Methods in Engineering*, 107(6):496–519, 2016.

[11] C. Cancès, M. Cathala, and C. Le Potier. Monotone corrections for generic cell-centered finite volume approximations of anisotropic diffusion equations. *Numerische Mathematik*, 125(3):387–417, 2013.

[12] C. Cancès, C. Chainais-Hillairet, and S. Krell. Numerical analysis of a nonlinear free-energy diminishing Discrete Duality Finite Volume scheme for convection diffusion equations. *Computational Methods in Applied Mathematics*, 18(3):407–432, 2018.

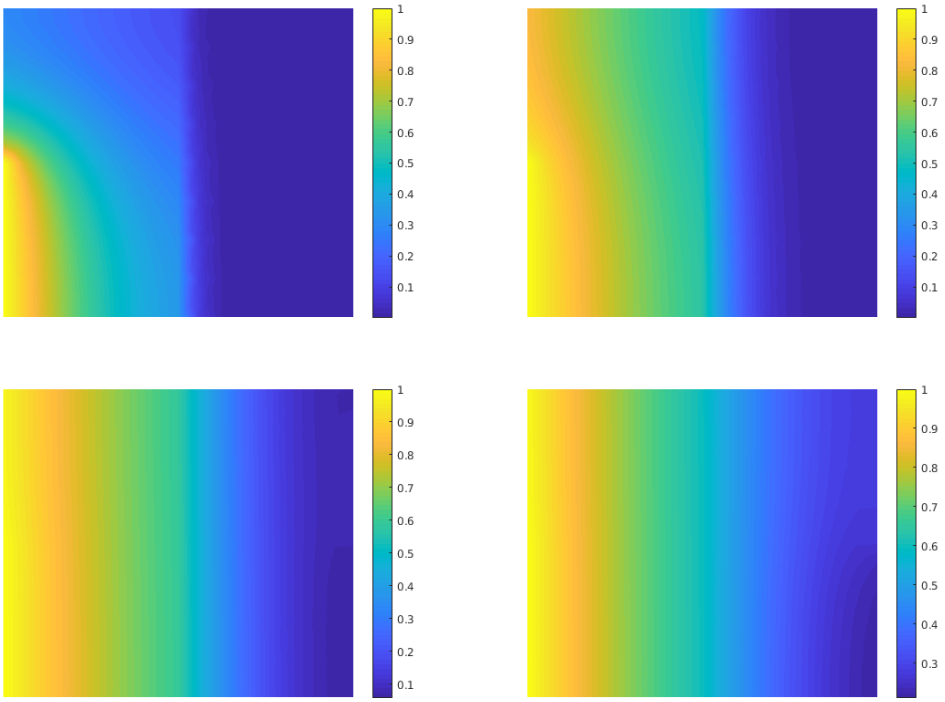

Figure 7: From left-up to right-bottom, the arithmetic-mean approximate solution for different simulation times $t \in \{0.25, 1, 4, 10\}$.

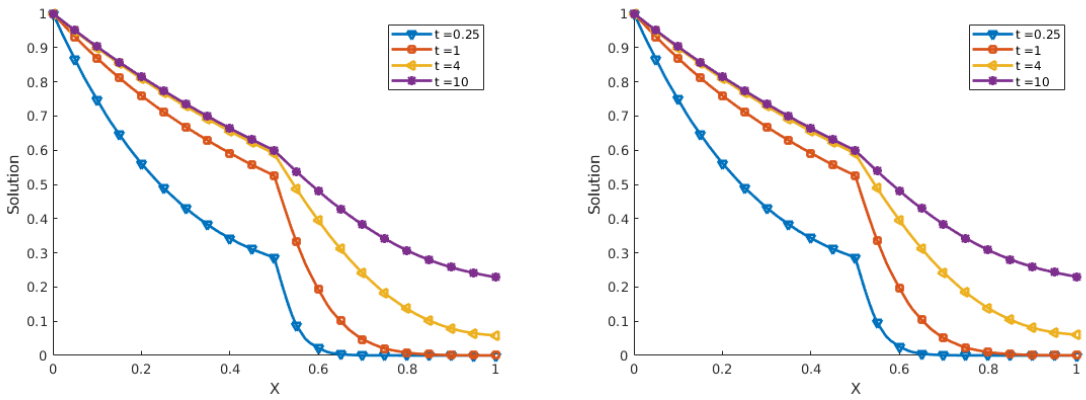

Figure 8: Log-mean (left) and arithmetic-mean (right) cross section of the computed solution at the point $(0, 0.4)$ for $t \in \{0.25, 1, 4, 10\}$.

[13] 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.

[14] 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.

[15] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter. Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Monatshefte für Mathematik*, 133(1):1–82, 2001.

[16] C. Chainais-Hillairet and J. Droniou. Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions. *IMA journal of numerical analysis*, 31(1):61–85, 2009.

[17] C. Chainais-Hillairet and F. Filbet. Asymptotic behaviour of a finite-volume scheme for the transient drift-diffusion model. *IMA Journal of Numerical Analysis*, 27(4):689–716, 2007.

[18] C. Chainais-Hillairet and M. Herda. Large-time behaviour of a family of finite volume schemes for boundary-driven convection–diffusion equations. *IMA Journal of Numerical Analysis*, 40(4):2473–2504, 2020.

[19] 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.

[20] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40. SIAM, 2002.

[21] K. Deimling. *Nonlinear Functional Analysis*. Springer-Verlag, Berlin, 1985.

[22] J. Droniou. Finite volume schemes for diffusion equations: introduction to and review of modern methods. *Mathematical Models and Methods in Applied Sciences*, 24(08):1575–1619, 2014.

[23] R. Eymard, T. Gallouët, M. Ghilani, and R. Herbin. Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by finite volume schemes. *IMA Journal of Numerical Analysis*, 18(4):563–594, 1998.

[24] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of Numerical Analysis*, volume 7, pages 713–1018. Elsevier, 2000.

[25] F. Filbet and M. Herda. A finite volume scheme for boundary-driven convection–diffusion equations with relative entropy structure. *Numerische Mathematik*, 137(3):535–577, 2017.

[26] Z. Gao and J. Wu. A second-order positivity-preserving finite volume scheme for diffusion equations on general meshes. *SIAM Journal on Scientific Computing*, 37(1):A420–A438, 2015.

[27] 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.

[28] 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.

[29] A. Glitzky. Exponential decay of the free energy for discretized electro-reaction–diffusion systems. *Nonlinearity*, 21(9):1989, 2008.

[30] 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.

[31] D. Horstmann. From 1970 until present: the Keller–Segel model in chemotaxis and its consequences. *I. Jahresberichte DMV*, 105(3):103–165, 2003.

[32] A. Jüngel. On the existence and uniqueness of transient solutions of a degenerate nonlinear drift-diffusion model for semiconductors. *Mathematical Models and Methods in Applied Sciences*, 4(05):677–703, 1994.

[33] A. Jüngel. Qualitative behavior of solutions of a degenerate nonlinear drift-diffusion model for semiconductors. *Mathematical Models and Methods in Applied Sciences*, 5(04):497–518, 1995.

[34] I. Kapyrin. A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes. In *Doklady Mathematics*, volume 76, pages 734–738. Springer, 2007.

[35] C. Le Potier. Finite volume scheme satisfying maximum and minimum principles for anisotropic diffusion operators. *Finite volumes for complex applications V*, pages 103–118, 2008.

[36] L. Li and J.-G. Liu. Large time behaviors of upwind schemes and B-schemes for Fokker-Planck equations on $\mathbb{R}$ by jump processes. *Mathematics of Computation*, 89:2283–2320, 2020.

[37] T.-P. Lin. The power mean and the logarithmic mean. *The American Mathematical Monthly*, 81(8):879–883, 1974.

[38] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes. *Journal of Computational Physics*, 228(3):703–716, 2009.

[39] J. M. Nordbotten, I. Aavatsmark, and G. Eigestad. Monotonicity of control volume methods. *Numerische Mathematik*, 106(2):255–288, 2007.

[40] 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.

[41] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode oscillator. *IEEE Transactions on electron devices*, 16(1):64–77, 1969.

[42] Z. Sheng and G. Yuan. The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes. *Journal of Computational Physics*, 230(7):2588–2604, 2011.