HAL
open science

A finite volume scheme for boundary-driven convection-diffusion equations with relative entropy structure

Francis Filbet, Maxime Herda

To cite this version:

Francis Filbet, Maxime Herda. A finite volume scheme for boundary-driven convection-diffusion equations with relative entropy structure. Numerische Mathematik, 2017. hal-01326029v4

HAL Id: hal-01326029

https://hal.science/hal-01326029v4

Submitted on 19 Apr 2017

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.

A FINITE VOLUME SCHEME FOR BOUNDARY-DRIVEN CONVECTION-DIFFUSION EQUATIONS WITH RELATIVE ENTROPY STRUCTURE

FRANCIS FILBET AND MAXIME HERDA

Abstract. We propose a finite volume scheme for a class of nonlinear parabolic equations endowed with non-homogeneous Dirichlet boundary conditions and which admit relative entropy functionals. For this kind of models including porous media equations, Fokker-Planck equations for plasma physics or dumbbell models for polymer flows, it has been proved that the transient solution converges to a steady-state when time goes to infinity. The present scheme is built from a discretization of the steady equation and preserves steady-states and natural Lyapunov functionals which provide a satisfying long-time behavior. After proving well-posedness, stability, exponential return to equilibrium and convergence, we present several numerical results which confirm the accuracy and underline the efficiency to preserve large-time asymptotic.

Keywords. Finite volume methods, relative entropy, non-homogeneous Dirichlet boundary conditions, polymers, magnetized plasma, porous media

MSC2010 subject classifications. 65M08, 65M12, 76S05, 76X05, 82D60

Contents

1. Introduction 2
1.1. General Setting 2
1.2. Physical models 5
1.3. Outline and main results 6
2. Presentation of the numerical schemes 7
2.1. Mesh and time discretization 7
2.2. Discretization of the steady equation 8
2.3. Discretization of the evolution equation 9
2.4. Discrete relative ϕ-entropies and dissipations 10
3. Properties of the discrete flux 11
3.1. Preservation of the steady state 11
3.2. Non-negativity of dissipations 11
4. Analysis of the schemes 13
4.1. Implicit Euler 13
4.2. Explicit Euler 14
4.3. Long time behavior 16
5. Convergence 18
5.1. Hypotheses 18
5.2. Compactness 19
5.3. Convergence of the scheme 20
6. Numerical simulations 23
6.1. Implementation 23
6.2. Proof of concept 24

6.3. Fokker-Planck with magnetic field 25
6.4. Polymer flow in a dilute solution 29
6.5. Porous medium equation 32
7. Comments and conclusion 33
References 34

1. Introduction

1.1. **General Setting.** Let Ω be a polyhedral open bounded connected subset of $\mathbb{R}^d$ with boundary $\Gamma = \partial\Omega$. Let us introduce the steady advection field $\mathbf{E} : \Omega \to \mathbb{R}^d$ and $\eta : \mathbb{R} \to \mathbb{R}$ a strictly increasing smooth function onto $\mathbb{R}$ satisfying $\eta(0) = 0$. We consider the following nonlinear convection-diffusion equation with non-homogeneous Dirichlet boundary conditions

$$\left\{\begin{array}{ll} \dfrac{\partial f}{\partial t} + \nabla\cdot(\mathbf{E}(\mathbf{x})\,\eta(f) - \nabla\eta(f)) = 0 & \text{for } \mathbf{x}\in\Omega,\ t\geq 0,\\[2ex] f(t,\mathbf{x}) = f^b(\mathbf{x}) & \text{for } \mathbf{x}\in\Gamma,\ t\geq 0,\\[2ex] f(0,\mathbf{x}) = f^{\text{in}}(\mathbf{x}) & \text{for } \mathbf{x}\in\Omega. \end{array}\right. \tag{1}$$

In [7], T. Bodineau, C. Villani, C. Mouhot and J. Lebowitz showed that this equation admits a large class of Lyapunov functionals, that we denote, using their denomination, relative ϕ-entropies. Each functional is generated by a convex function ϕ satisfying the following properties.

Definition 1.1 (Entropy generating functions)**.** For any non-empty interval J of $\mathbb{R}$ containing 1, we say that $\phi \in \mathcal{C}^2(J,\mathbb{R}_+)$ is an *entropy generating function* or simply *entropy function* if it is strictly convex and satisfies $\phi(1) = 0$ and $\phi'(1) = 0$.

The entropies are defined relatively to a steady state of (1). Therefore, we assume that there exists f^∞ which satisfies

$$\left\{\begin{array}{ll} \nabla\cdot(\mathbf{E}\,\eta(f^\infty) - \nabla\eta(f^\infty)) = 0 & \text{in } \Omega,\\[2ex] f^\infty = f^b & \text{on } \Gamma. \end{array}\right. \tag{2}$$

Now we can define the relative ϕ-entropies and associated dissipations.

Definition 1.2 (Relative ϕ-entropy and dissipation)**.** For any entropy generating function ϕ, we denote by $\mathcal{H}_\phi$ the *relative ϕ-entropy* defined by

$$\mathcal{H}_\phi(t) = \int_\Omega \int_{f^\infty(\mathbf{x})}^{f(t,\mathbf{x})} \phi'\left(\frac{\eta(s)}{\eta(f^\infty(\mathbf{x}))}\right) \mathrm{d}s\,\mathrm{d}\mathbf{x},$$

and by $\mathcal{D}_\phi$ the relative ϕ-entropy dissipation defined by

$$\mathcal{D}_\phi(t) = \int_\Omega |\nabla h|^2\ \phi''(h)\,\eta(f^\infty)\,\mathrm{d}\mathbf{x},$$

where h is the ratio between the transient and stationary nonlinearities

$$h \ = \ \left\{\begin{array}{ll} \dfrac{\eta(f)}{\eta(f^\infty)} & \text{in} \quad \Omega\,,\\[2ex] 1 & \text{on} \quad \Gamma\,. \end{array}\right. \tag{3}$$

Let us note that for the linear problem, namely when η is the identity function, the relative ϕ-entropy rewrites

$$\mathcal{H}_\phi(t) = \int_\Omega \phi\left(\frac{f}{f^\infty}\right) f^\infty \, \mathrm{d}\mathbf{x}.$$

Typical examples of relative ϕ-entropies are the *physical relative entropy* and *p-entropies* (or *Tsallis relative entropies*) respectively generated, for $p \in (1, 2]$, by

$$\left\{ \begin{array}{rcl} \phi_1(x) & = & x\ln(x) - (x-1)\,, \\ \\ \phi_p(x) & = & \dfrac{x^p - px}{p-1} + 1\,. \end{array} \right. \tag{4}$$

One readily sees that, since η and ϕ' are increasing functions satisfying $\eta(0) = 0$ and $\phi'(1) = 0$, the relative ϕ-entropy is a non-negative quantity which cancels if and only if f and f^∞ coincide almost everywhere. The ϕ-entropies are not, in general, distances between the solution and the steady state. However Csiszar-Kullback type inequalities [16, 26, 30] yield a control of the L^1 distance between the solution and the equilibrium. Therefore if a relative ϕ-entropy goes to zero when time goes to infinity, the solution converges to equilibrium in a strong sense.

The following proposition was proved in [7, Theorem 1.4] and yields an entropy-entropy dissipation principle for Equation (1). It starts from a reformulation of (1) using the new unknown (3). It is easily derived using Leibniz product rule in Equation (1) together with (2) and (3) and reads

$$\left\{ \begin{array}{l} \dfrac{\partial f}{\partial t} + \nabla \cdot (\mathbf{U}^\infty h - \eta(f^\infty)\nabla h) = 0\,, \\ \\ \mathbf{U}^\infty = \mathbf{E}\,\eta(f^\infty) - \nabla\eta(f^\infty)\,, \\ \\ \nabla \cdot \mathbf{U}^\infty = 0\,. \end{array} \right. \tag{5}$$

Proposition 1.3. *Any L^∞ solution of* (1) *satisfies in the sense of distributions*

$$\frac{\mathrm{d}\mathcal{H}_\phi}{\mathrm{d}t} = -\mathcal{D}_\phi \leq 0, \tag{6}$$

for any entropy generating function ϕ.

The formal computations leading to (6) motivate our choices in the elaboration of the discrete scheme. Therefore, we recall the proof yielding the entropy equality.

Proof. First, we integrate (5) against $\phi'(h)$, integrate by parts and use the boundary conditions and the fact that $\phi'(1) = 0$ to get

$$\begin{array}{rcl} \dfrac{\mathrm{d}\mathcal{H}_\phi}{\mathrm{d}t} & = & \displaystyle\int_\Omega \phi'(h)\,\nabla \cdot (-\mathbf{U}^\infty h + \eta(f^\infty)\,\nabla h)\ \mathrm{d}\mathbf{x} \\ & = & \displaystyle\int_\Omega \mathbf{U}^\infty \cdot \nabla h\,\phi''(h)\ h\ \mathrm{d}\mathbf{x} - \int_\Omega |\nabla h|^2\ \phi''(h)\,\eta(f^\infty)\ \mathrm{d}\mathbf{x}. \end{array}$$

Let $\varphi : s \mapsto s\phi'(s) - \phi(s)$. It satisfies $\varphi'(s) = \phi''(s)\,s$ and $\varphi(1) = 0$. Hence, substituting it in the last expression yields

$$\frac{\mathrm{d}\mathcal{H}_\phi}{\mathrm{d}t} = \int_\Omega \mathbf{U}^\infty \cdot \nabla\varphi(h)\ \mathrm{d}\mathbf{x} - \mathcal{D}_\phi = -\mathcal{D}_\phi,$$

where we integrated the first term by parts, used the stationary equation (2) and the boundary conditions. □

There are two important facts that justify the use of (5) instead of (1) to derive the above entropy dissipation inequality. The rewriting transforms the advection field $\mathbf{E}$ on $\eta(f)$ into the incompressible field $\mathbf{U}^\infty$ on h so that the contribution of the convection can vanish when the time derivative of the relative entropy is computed. The underlying cancellations stems from the transformation of $\nabla h\, \phi''(h)\, h$ into a gradient thanks to φ and on the fact that f^∞ solves (2). The second reason is that considering the equation on h instead of f changes non-homogeneous Dirichlet boundary conditions into homogeneous ones on $h-1$. Together with properties of ϕ, it enables cancellations of boundary terms. In other words, relative ϕ-entropies are the correct functionals and (5) the right form of the equation to capture the boundary-driven dynamics.

The purpose of this work is the design and analysis of a finite volume scheme preserving the whole class of relative ϕ-entropy dissipation inequalities (6). This is done by discretizing the reformulated equation (5) in a way that enables the counterpart of computations of the proof of Proposition 1.3 to hold at the discrete level. Because of the reformulation involving the steady state, the scheme is based on a preliminary discretization of the steady state and flux.

Let us emphasize that $\mathbf{E}$ is a general field and need not to be either incompressible nor irrotational as for parabolic equations with a gradient flow structure [29]. Indeed, assuming some regularity on the advection field, one can apply the Hodge decomposition to get the existence of a potential $V : \Omega \to \mathbb{R}$ and $\mathbf{F} : \Omega \to \mathbb{R}^d$ such that

$$\mathbf{E} \;=\; -\nabla V + \mathbf{F}, \qquad \nabla \cdot \mathbf{F} \;=\; 0 \tag{7}$$

When $\mathbf{F} = \mathbf{0}$, there are many examples in the literature [23, 12, 4, 6, 10, 9, 13] of finite volume schemes preserving entropy dissipation properties. C. Chainais-Hillairet and F. Filbet studied in [12] a finite volume discretization for nonlinear drift-diffusion system and proved that the numerical solution converges to a steady-state when time goes to infinity. In [9], M. Burger, J. A. Carrillo and M. T. Wolfram proposed a mixed finite element method for nonlinear diffusion equations and proved convergence towards the steady-state in case of a nonlinear Fokker-Planck equation with uniformly convex potential. All these schemes exploit the gradient flow structure of the equation, which gives a natural entropy.

In the non-symmetric case $\mathbf{F} \neq \mathbf{0}$, the gradient structure cannot be exploited anymore, but as Proposition 1.3 shows, there is still a relative entropy structure, which may be investigated to prove convergence to a steady state. The relative entropy properties of Fokker-Planck type equations in the whole space are exhaustively studied in the famous paper [3] of A. Arnold, P. Markowich, G. Toscani, A. Unterreiter and specific properties of the non-symmetric equations have been investigated in [2, 1].

In bounded domains, entropy properties are often used in the context of no-flux boundary conditions or in the whole space, but few results concern Dirichlet boundary conditions. In [4], M. Bessemoulin-Chatard proposed an extension of the Scharfetter-Gummel for finite volume scheme for convection-diffusion equations with nonlinear diffusion and non-homogeneous and unsteady Dirichlet boundary conditions. While in the latter work the author presents a scheme with a satisfying long-time behavior for a similar class of models than those of the present paper, our strategy and objectives differ. Here we aim at preserving a whole class of relative entropies and build our scheme for the transient problem from a discretization of the stationary equation.

Let us precise that we can generalize our approach to the more general boundary conditions

$$\begin{cases} f = f^b \text{ on } \Gamma_D, \\ [\mathbf{E}\,\eta(f) - \nabla\eta(f)] \cdot \mathbf{n}(\mathbf{x}) = 0 \quad \text{on} \quad \Gamma_N, \end{cases}$$

with $\Gamma = \Gamma_D \cup \Gamma_N$. Our results hold in this setting with minor modifications but to avoid unnecessary technicalities in the notation and in the analysis we consider non-homogeneous Dirichlet conditions on the whole boundary. However, numerical results will be shown in both cases.

1.2. **Physical models.** Before describing our numerical scheme, let us present some physical models described by equation (1) for which the large-time asymptotic has been studied using entropy-entropy dissipation arguments. Some of these models are the homogeneous part of kinetic Fokker-Planck-type equations and this work constitutes a first step towards treating full kinetic models. In future work, we aim at adapting the strategy developed here to ensure the property of convergence to local equilibrium for the solutions of these equations.

1.2.1. **The Fokker-Planck equation with magnetic field.** A classical model of plasma physics describing the dynamic of charged particles evolving in an external electromagnetic field $(-\nabla_{\mathbf{x}}\phi, \mathbf{B})$ is given by the Vlasov-Fokker-Planck equation reading

$$\frac{\partial F}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}}F - \nabla_{\mathbf{x}}\phi\cdot\nabla_{\mathbf{v}}F + (\mathbf{v}\wedge\mathbf{B})\cdot\nabla_{\mathbf{v}}F = \nabla_{\mathbf{v}}\cdot(\mathbf{v}F + \nabla_{\mathbf{v}}F). \tag{8}$$

For more details on the model we refer to [8, 24]. In [8], Bouchut and Dolbeault proved that the solution of (8) in the whole phase space and without magnetic field converges to a global equilibrium. Their proof mainly relies on the decrease of the free energy functional, which corresponds to the physical relative entropy introduced in (4). The external magnetic field does not alter the relative entropy inequality. We refer to [24] for the corresponding computations. Here, we consider the phenomena happening in the velocity space which results in a Fokker-Planck equation with magnetic field, namely equation (1) with $\eta(s) = s$ and an advection field given by

$$\mathbf{E}(\mathbf{v}) = -\mathbf{v} + \mathbf{v}\wedge\mathbf{B}\,, \tag{9}$$

with constant magnetic field $\mathbf{B}$. In applications, the velocity variable $\mathbf{v}$ usually lives in $\mathbb{R}^3$. However when performing numerical simulations, one needs to restrict the velocity domain to a bounded set Ω. On the edge of this restricted domain Ω, we shall consider the following non-homogeneous Dirichlet boundary conditions

$$f(t,\mathbf{v}) = f^{\infty}(\mathbf{v}) \quad \forall \mathbf{v}\in\partial\Omega, \tag{10}$$

where f^{∞} is the local Maxwellian associated with (8) which writes

$$f^{\infty}(\mathbf{v}) = \frac{1}{(2\pi)^{3/2}} e^{-\frac{|\mathbf{v}|^2}{2}} \quad \forall \mathbf{v}\in\Omega, \tag{11}$$

and is a stationary state of (1) with (9)-(10). With the boundary conditions (10), one recovers the same stationary state as in the whole space while working in a bounded domain. As for the more complicated kinetic model (8), free energy (relative ϕ-entropy) decrease holds.

Our approach is particularly promising for this kind of problem when the solution develops some micro-instabilities around a steady state. In this situation, it is important that numerical artefacts do not generate spurious oscillations.

1.2.2. **The dumbbell model for the density of polymers in a dilute solution.** The following kinetic equation describes the evolution of the density $F \equiv F(t,\mathbf{x},\mathbf{k})$ of polymers at time t and position $\mathbf{x}$ diluted in a fluid flow of velocity $\mathbf{u}(\mathbf{x})$ from a mesoscopic point of view

$$\frac{\partial F}{\partial t} + \mathbf{u}\cdot\nabla_{\mathbf{x}}F = -\nabla_{\mathbf{k}}\cdot\left[\left(\nabla_{\mathbf{x}}\mathbf{u}\,\mathbf{k} - \frac{1}{2}\nabla_{\mathbf{k}}\Pi(\mathbf{k})\right)F - \frac{1}{2}\nabla_{\mathbf{k}}F\right]. \tag{12}$$

The polymers are pictured as two beads linked by a spring and the variable $\mathbf{k}$ stands for the vector indicating the length and orientation of the molecules. The potential Π is given by $\Pi(\mathbf{k}) = |\mathbf{k}|^2/2$ in the case of Hookean dumbbells or by $\Pi(\mathbf{k}) = -\ln(1-|\mathbf{k}|^2)/2$ in the case of Finite Extensible Nonlinear Elastic dumbbells. In the complete model, the velocity of the fluid $\mathbf{u}$ follows an incompressible Navier-Stokes equation featuring an additional force term modeling for the contribution of the polymers on the dynamic of the fluid which results in a nonlinear kinetic-fluid coupling. Here, we consider the simpler case where $u(\mathbf{x})$ is a given incompressible field. For more details on the modeling, we refer to [25] and references therein. We also refer to the paper [28] of Masmoudi that treats the well-posedness and provides additional information on the model.

Once again we aim at approximating numerically the "velocity" part of the kinetic equation (12) which rewrites as (1) with $\eta(s) = s$ and an advection field given by

$$\mathbf{E}(\mathbf{k}) = \mathbf{A}\,\mathbf{k} - \frac{1}{2}\nabla_{\mathbf{k}}\Pi(\mathbf{k}), \tag{13}$$

with constant matrix $\mathbf{A}$ satisfying $\mathrm{tr}(\mathbf{A}) = 0$. This matrix is the gradient of an incompressible velocity field at some space location. Natural boundary conditions for this model are given by null outward flux. In [25], the long-time behavior of (12) and of the latter reduced model are investigated using relative ϕ-entropies with ϕ given by (4).

1.2.3. **A nonlinear model, the porous medium equation.** The porous medium equation is a nonlinear PDE writing

$$\frac{\partial f}{\partial t} = \Delta f^m, \tag{14}$$

with $m > 1$. It can model many physical applications and generally describes processes involving fluid flow, heat transfer or diffusion. The typical example is the description of the flow of an isentropic gas through a porous medium. There is a huge literature on this equation and we refer to the book of Vásquez [31] for the detailed mathematical theory.

Here, equation (14) is set in a bounded domain Ω with non-homogeneous Dirichlet boundary conditions

$$f(t,\mathbf{x}) = f^b(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \partial\Omega,$$

such that it might be recast like (1) with a null advection field and with $\eta(s) = s^m$. Using their relative ϕ-entropy method, Bodineau, Mouhot, Villani and Lebowitz show exponential convergence to equilibrium for this nonlinear equation.

1.3. **Outline and main results.** The plan of the paper is as follows. In Section 2, we present the finite volume scheme and the discrete version of the relative ϕ-entropies. Then, in Section 3, we prove the main properties of our discrete flux , namely the preservation of steady states in Lemma 3.1 and the non-negativity of discrete ϕ entropy dissipations in Proposition 3.2. In Section 4, we analyze our fully-discrete schemes. In Theorem 4.2, we prove well-posedness, stability and decay of discrete ϕ-entropies for the time implicit version of our scheme. For the explicit version, we prove the same results in Theorem 4.4, under a parabolic Courant-Friedrichs-Lewy (CFL) condition. Besides, the long-time behavior of the discrete solution, for both schemes, is investigated in this section and we prove in Theorem 4.7 that discrete solutions return to equilibrium exponentially fast, with a rate that does not depend on the size of the discretization. In Section 5, we show convergence of the discrete solution to a solution of (5), when the size of the discretization goes to zero. Finally in Section 6, we end by providing numerical illustrations of the properties of our schemes on the models presented above.

2. Presentation of the numerical schemes

In this section, we introduce our finite volume schemes. In the following, T is a positive real number and Ω_T denotes the cylinder $[0,T)\times\Omega$. We start with some notations associated with the discretization of Ω_T.

2.1. **Mesh and time discretization.** An admissible mesh of Ω is defined by the triplet $(\mathcal{T},\mathcal{E},\mathcal{P})$. The set $\mathcal{T}$ is a finite family of nonempty connected open disjoint subsets $K\subset\Omega$ called control volumes or cells. The closure of the union of all control volumes is equal to $\bar\Omega$. The set $\mathcal{E}$ is a finite family of nonempty subsets $\bar\Omega$ called edges. Each edge is a subset of an affine hyperplane in $\mathbb{R}^{d-1}$. Moreover, for any control volume $K\in\mathcal{T}$ there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that the closure of the union of all the edges in $\mathcal{E}_K$ is equal to $\partial K=\bar K\setminus K$. We also define several subsets of $\mathcal{E}$. The family of interior edges $\mathcal{E}_{\rm int}$ is given by $\{\sigma\in\mathcal{E},\ \sigma\not\subseteq\Gamma\}$ and the family of exterior edges by $\mathcal{E}_{\rm ext}=\mathcal{E}\setminus\mathcal{E}_{\rm int}$. Similarly, for any control volume $K\in\mathcal{T}$, we define $\mathcal{E}_{{\rm int},K}=\mathcal{E}_{\rm int}\cap\mathcal{E}_K$ and $\mathcal{E}_{{\rm ext},K}=\mathcal{E}_{\rm ext}\cap\mathcal{E}_K$. We assume that for any edge σ, the number of control volumes sharing the edge σ is exactly 2 for interior edges and 1 for exterior edges. With these assumptions, every interior edge is shared by two control volumes, say K and L, so that we may use the notation $\sigma=K|L$ whenever $\sigma\in\mathcal{E}_{\rm int}$. The set $\mathcal{P}=\{\mathbf{x}_K\}_{K\in\mathcal{T}}$ is a finite family of points satisfying that for any control volume $K\in\mathcal{T}$, $\mathbf{x}_K\in K$. We introduce the transmissibility of the edge σ, given by

$$\tau_\sigma=\frac{m(\sigma)}{d_\sigma},$$

where

$$d_\sigma=\begin{cases} d(\mathbf{x}_K,\mathbf{x}_L), & \text{if } \sigma\in\mathcal{E}_{\rm int},\ \sigma=K|L\,,\\ d(\mathbf{x}_K,\sigma), & \text{if } \sigma\in\mathcal{E}_{{\rm ext},K}\,,\end{cases}$$

with $d(\cdot,\cdot)$ the euclidean distance in $\mathbb{R}^d$. The size of the mesh is defined by

$$\Delta x=\max_{K\in\mathcal{T}}\sup_{\mathbf{x},\mathbf{y}\in K} d(\mathbf{x},\mathbf{y})\,.$$

The Dirichlet condition on the boundary is given by $f^b\in L^\infty(\Gamma)$. Endowed with these boundary conditions, a discrete solution of the scheme at some fixed time is an element of the set

$$X_{f^b}=\left\{f\in\mathbb{R}^{\mathcal{T}}\times\mathbb{R}^{\mathcal{E}_{\rm ext}}:\ f_\sigma=\frac{1}{m(\sigma)}\int_\sigma f^b dm,\ \forall\sigma\in\mathcal{E}_{\rm ext}\right\}.$$

Remark 2.1. *The particular formula providing f_σ on the exterior edges is cosmetic here since by the reformulation of the equation, we will only use the constant boundary values of a reformulated unknown defined in* (15).

Let us mention that with a slight abuse we keep the same notation for discrete and continuous unknowns. Concerning the initial condition we assume that

(H1) $$f^{\rm in}\geq 0\quad\text{and}\quad K_{\rm in}:=\|f^{\rm in}\|_{L^\infty(\Omega)}<+\infty\,,$$

and the discrete initial condition is given by

$$f_K^{\rm in}=\frac{1}{m(K)}\int_K f^{\rm in}(\mathbf{x})\,\mathrm{d}\mathbf{x}\,,$$

for all $K\in\mathcal{T}$. For any function $\psi:\mathbb{R}\to\mathbb{R}$, and $f\in X_{f^b}$ we shall define the component-wise composition with the intuitive notation $\psi(f)=((\psi(f_K))_{K\in\mathcal{T}},(\psi(f_\sigma))_{\sigma\in\mathcal{E}_{\rm ext}})$.

We denote the time step by Δt and set $t^n=n\,\Delta t$. From this, a time discretization $[0,T)$ is given by the integer $N_T=\lfloor T/\Delta t\rfloor$ and the sequence $(t^n)_{0\leq n\leq N_T}$. A spacetime discretization

$\mathcal{D}$ of Ω_T is composed of an admissible mesh of Ω and the time discretization parameters Δt and N_T. Finally, the size of the discretization $\mathcal{D}$ is given by

$$\delta \,=\, \max(\Delta x,\, \Delta t).$$

2.2. **Discretization of the steady equation.** In order to build our numerical scheme for the reformulated equation (5), we need a discrete version of the steady state as well as corresponding discrete steady flux at interfaces. Hence we introduce an approximation of $\eta(f^\infty)$ on cells $(\eta(f^\infty)_K)_{K\in\mathcal{T}}$ and on edges $(\eta(f^\infty)_\sigma)_{\sigma\in\mathcal{E}}$. We suppose that there exists positive constants m_∞ and M_∞ that do not depend on the discretization $\mathcal{D}$ and such that for all $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}$

$$0 < m_\infty \,\leq\, \eta(f^\infty)_K,\ \eta(f^\infty)_\sigma \,\leq\, M_\infty\,. \tag{H2}$$

Remark 2.2. *Of course this hypothesis requires some kind of maximum principle holding at the continuous level. It is the case at least for divergence free fields* $\mathbf{E}$*, if* f^b *is bounded and bounded from below by a positive constant then by* [17, Section 6.4, Theorem 1].

Moreover for each cell $K \in \mathcal{T}$ and edge $\sigma \in \mathcal{E}_K$, we introduce a discrete flux $F^\infty_{K,\sigma}$ approximating $\int_\sigma \mathbf{U}^\infty(\mathbf{x}) \cdot \mathbf{n}_{K,\sigma}(\mathbf{x})\, \mathrm{d}\mathbf{x}$ along with the divergence free hypothesis

$$\sum_{\sigma\in\mathcal{E}_K} F^\infty_{K,\sigma} = 0\,, \tag{H3}$$

and the interior continuity condition

$$\forall \sigma = K|L \in \mathcal{E}_{\text{int}}\,, \quad F^\infty_{K,\sigma} = -F^\infty_{L,\sigma}. \tag{H4}$$

Let us emphasize that this last condition is required and satisfied for all the discrete flux we define in the following. From the flux we also introduce the discrete velocities

$$U^\infty_{K,\sigma} = \frac{1}{m(\sigma)} F^\infty_{K,\sigma}\,.$$

On one hand, for some models, the global equilibrium f^∞ may be known analytically. In this case, we may build a discrete approximation in X_{f^b} by a standard projection on the mesh and the numerical flux $F^\infty_{K,\sigma}$ may be computed exactly or approximated with a quadrature formula. In any case, hypotheses (H2)-(H4) must be satisfied. In Section 6.2 and Section 6.3, this strategy is used for computing f^∞_K and $F^\infty_{K,\sigma}$ on our test cases.

On the other hand, when the steady state is not known we apply a finite volume scheme to compute a numerical approximation. Our method does not impose any scheme for solving the stationary equation as long as assumptions (H2)-(H4) are satisfied. However to fix ideas let us provide an example here.

Example 2.3. *One can determine the discrete steady state by solving* (H3) *on each cell* $K \in \mathcal{T}$ *with the flux*

$$F^\infty_{K,\sigma} \;=\; \begin{cases} m(\sigma)\left[E^+_{K,\sigma}\ \eta(f^\infty)_K - E^-_{K,\sigma}\eta(f^\infty)_L \;-\; \dfrac{\eta(f^\infty)_L - \eta(f^\infty)_K}{d_\sigma}\right], & \text{if } \sigma = K|L\,, \\ m(\sigma)\left[E^+_{K,\sigma}\ \eta(f^\infty)_K - E^-_{K,\sigma}\eta(f^\infty)_\sigma \;-\; \dfrac{\eta(f^\infty)_\sigma - \eta(f^\infty)_K}{d_\sigma}\right], & \text{otherwise}\,, \end{cases}$$

where for any real number u*, we denote by* $u^+ = \max(u, 0)$ *and* $u^- = \max(-u, 0)$ *the positive and negative parts of* u*. The quantity* $E_{K,\sigma}$ *is a consistent approximation of* $\mathbf{E} \cdot \mathbf{n}_{K,\sigma}$ *on the edge* σ *such as* $\mathbf{E}(\mathbf{x}_\sigma) \cdot \mathbf{n}_{K,\sigma}$*, where* $\mathbf{x}_\sigma$ *is the center of mass of the edge* σ*. Then, for* $\sigma = K|L$*, the approximation of* $\eta(f^\infty)$ *on the edge* σ *can be given by* $\eta(f^\infty)_\sigma = (\eta(f^\infty)_K + \eta(f^\infty)_L)/2$ *and for each cell* $K \in \mathcal{T}$*,* $f^\infty_K = \eta^{-1}(\eta(f^\infty)_K)$*.*

In Section 6.4, we provide another example of resolution of the steady equation by a finite volume discretization.

2.3. **Discretization of the evolution equation.** Now we treat the time evolution problem and use the discrete steady state to build a numerical approximation of the reformulated equation (5). We start by introducing our finite volume scheme in a general implicit and explicit form. Then we define the flux.

2.3.1. *Fully discrete schemes.* In the continuous setting, we defined a new unknown to reformulate the convection-diffusion equation. Its discrete equivalent is still denoted h and belongs to the space $X_1^{N_T+1}$. It is defined for any $K \in \mathcal{T}$ and $n \in \{0, \dots, N_T\}$ by

$$h_K^n = \frac{\eta(f_K^n)}{\eta(f^\infty)_K}\,. \tag{15}$$

Let us insist on the fact that boundary conditions are contained in the definition of the approximation space so that here, $h_\sigma = 1$ for $\sigma \in \mathcal{E}_{\text{ext}}$. Solving the fully discrete implicit scheme consists in finding $f = (f_K^n)_{n\in\{0,\dots,N_T\}} \in X_{f^b}^{N_T+1}$ such that for all $K \in \mathcal{T}$

$$\begin{cases} m(K)\dfrac{f_K^{n+1} - f_K^n}{\Delta t} + \displaystyle\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(h^{n+1}) \;=\; 0, \quad \forall n \in \{0, \dots, N_T - 1\}\,, \\ f_K^0 \;=\; f_K^{\text{in}}\,. \end{cases} \tag{16}$$

In explicit form, it amounts to building f sequentially by

$$\forall K \in \mathcal{T}, \begin{cases} m(K)\dfrac{f_K^{n+1} - f_K^n}{\Delta t} + \displaystyle\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(h^n) \;=\; 0, \quad \forall n \in \{0, \dots, N_T - 1\} \\ f_K^0 = f_K^{\text{in}}, \end{cases} \tag{17}$$

In order to lighten the notation we sometimes write $F_{K,\sigma}(h^n) = F_{K,\sigma}^n$ in the following. For any discrete element $u \in X_{u^b}^{N_T+1}$, its reconstruction is given almost everywhere on Ω_T by

$$u_\delta(t,x) = \begin{cases} u_K^{n+1} & \text{if } x \in K \text{ and } t \in [t^n, t^{n+1})\,, \quad \text{(Implicit case)} \\ u_K^n & \text{if } x \in K \text{ and } t \in [t^n, t^{n+1})\,. \quad \text{(Explicit case)} \end{cases}$$

2.3.2. *Definition of the flux.* The flux is divided in two parts,

$$F_{K,\sigma} = F_{K,\sigma}^{\text{conv}} + F_{K,\sigma}^{\text{diss}}, \tag{18}$$

corresponding respectively to the discretization of the convective term $\int_\sigma h\,\mathbf{U}^\infty\cdot\mathbf{n}_{K,\sigma}(\mathbf{x})\,\mathrm{d}m$ and of the dissipative term $-\int_\sigma \eta(f^\infty)\,\nabla h\cdot\mathbf{n}_{K,\sigma}(\mathbf{x})\,\mathrm{d}m$. Accordingly, we call $F_{K,\sigma}^{\text{conv}}$ the convective flux and $F_{K,\sigma}^{\text{diss}}$ the dissipative flux. The former is discretized thanks to an monotone upstream discretization reading

$$F_{K,\sigma}^{\text{conv}}(h) = \begin{cases} m(\sigma)\left[U_{K,\sigma}^{\infty,+}\; g(h_K, h_L) - U_{K,\sigma}^{\infty,-} g(h_L, h_K)\,\right], & \text{if } \sigma = K|L, \\ m(\sigma)\left[U_{K,\sigma}^{\infty,+}\; g(h_K, h_\sigma) - U_{K,\sigma}^{\infty,-} g(h_\sigma, h_K)\,\right], & \text{otherwise}, \end{cases} \tag{19}$$

where $U_{K,\sigma}^{\infty,+} = \max(U_{K,\sigma}^\infty, 0)$ and $U_{K,\sigma}^{\infty,-} = \max(-U_{K,\sigma}^\infty, 0)$ denote the positive and negative parts of $U_{K,\sigma}^\infty$ respectively. The classical upwind flux, that we use for our numerical simulations is obtained by choosing $g(s,t) = s$. With this more general version, we aim at highlighting the importance of the monotony property of this flux concerning stability of the scheme and

decay of discrete relative ϕ-entropy. Accordingly, $g : \mathbb{R}^2 \to \mathbb{R}$ ought to satisfy the following assumptions.

$$\text{(H5)} \quad \begin{cases} g \text{ is locally Lipschitz-continuous,} \\ g \text{ is non-decreasing in the first variable and non-increasing in the second variable,} \\ g(s,s) = s, \text{ for all } s \in \mathbb{R}. \end{cases}$$

The first and last conditions ensure consistency of the approximation.

The dissipative flux is built on a two-point approximation of the derivative along the outward normal vector of each edge, namely

$$F^{\text{diss}}_{K,\sigma}(h) = -\tau_\sigma\, \eta(f^\infty)_\sigma\, D_{K,\sigma} h, \tag{20}$$

where the difference operator $D_{K,\sigma}$ is defined for any $K \in \mathcal{T}$, $\sigma \in \mathcal{E}_K$ and $u \in X_{u^\flat}$ by

$$D_{K,\sigma} u = \begin{cases} u_L - u_K & \text{if } \sigma \in \mathcal{E}_{\text{int}},\ \sigma = K|L, \\ u_\sigma - u_K & \text{if } \sigma \in \mathcal{E}_{\text{ext},K}. \end{cases}$$

For consistency of discrete gradients, we require an orthogonality condition for the mesh, namely

$$\forall\, \mathbf{x}, \mathbf{y} \in \sigma = K|L, \quad (\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x}_K - \mathbf{x}_L) = 0. \tag{H6}$$

Remark 2.4. *Hypothesis* (H6) *is necessary for establishing the convergence of the scheme in Section 5 and even if it is standard* [20]*, it remains a restrictive condition on the shape of the mesh. The necessity for this hypothesis stems from the fact that we choose a two-point flux* (20) *for the diffusion. This choice is motivated by the monotony properties of this flux which enable the decay of ϕ-entropies at the discrete level.*

Remark 2.5. *Even if the flux* (19) *and* (20) *seem fairly classical, one must not forget that they act on the unknown h while the discrete time derivative in* (16) *or* (17) *is on f. As in the continuous setting, this is the most important part of the strategy.*

2.4. **Discrete relative ϕ-entropies and dissipations.** For $f \in X_{f^\flat}$, the discrete equivalent of the relative ϕ-entropy in Definition 1.2 is given by

$$H_\phi(f) = \sum_{K \in \mathcal{T}} m(K) e_{\phi,K}(f), \tag{21}$$

where $e_\phi = (e_{\phi,K})_{K \in \mathcal{T}}$ is the local discrete relative ϕ-entropy writing, for $K \in \mathcal{T}$

$$e_{\phi,K}(f) = \int_{f_K^\infty}^{f_K} \phi'\left(\frac{\eta(s)}{\eta(f_K^\infty)}\right) \mathrm{d}s.$$

Contributions of the convective and diffusive part of the equation to the relative entropy variation are defined, for $h \in X_1$ by

$$C_\phi(h) = \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \phi'(h_K) F^{\text{conv}}_{K,\sigma}(h), \qquad D_\phi(h) = \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \phi'(h_K) F^{\text{diss}}_{K,\sigma}(h). \tag{22}$$

We write H^n_ϕ, C^n_ϕ and D^n_ϕ to denote respectively $H_\phi(f^n)$, $C_\phi(h^n)$ and $D_\phi(h^n)$. The precise relation between these quantities is derived in the proof of Theorem 4.2 and 4.4. For the moment, $C^n_\phi + D^n_\phi$ should be thought as the "time derivative" of H^n_ϕ. In Proposition 3.2 we show that D_ϕ is as expected consistent with its continuous analogue $\mathcal{D}_\phi$ and non-negative. Moreover, thanks to the monotonicity properties of the convective flux (19), we also prove that $C_\phi(f)$ remains non-negative, creating an additional numerical dissipation consistent with 0 as $\delta \to 0$.

3. Properties of the discrete flux

We present here, independently of time discretizations (16) or (17), the properties of the flux introduced in the previous section. In this section, f is an element of X_{f^b} and h the corresponding element of X_1 using relation (15).

3.1. Preservation of the steady state.

The next lemma show that the discrete steady state is a steady state of our schemes.

Lemma 3.1. *Under hypotheses* (H2)-(H3)*, if for all cells $K \in \mathcal{T}$, $\eta(f_K) = \eta(f^\infty)_K$, then*

$$\forall K \in \mathcal{T}, \quad \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}(h) \;=\; 0.$$

Proof. Just observe that every component of h equals 1. Hence, for any $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, $F^{\text{diss}}_{K,\sigma}(h) = 0$ and

$$\sum_{\sigma \in \mathcal{E}_K} F^{\text{conv}}_{K,\sigma}(h) = \sum_{\sigma \in \mathcal{E}_K} m(\sigma) \left(U^{\infty,+}_{K,\sigma} - U^{\infty,-}_{K,\sigma} \right) = \sum_{\sigma \in \mathcal{E}_K} F^{\infty}_{K,\sigma} = 0,$$

using (H3). □

3.2. Non-negativity of dissipations.

Proposition 3.2. *Let ϕ be any entropy generating function and $h \in X_1$. Under hypotheses* (H2)-(H5)*, the following results hold.*

(i) *The numerical dissipation is non-negative, namely,*

$$C_\phi(h) \;\geq\; 0\,.$$

(ii) *The physical dissipation rewrites*

$$\begin{aligned} D_\phi(h) \;\; = \;\; & \sum_{\substack{\sigma \in \mathcal{E}_{int} \\ \sigma = K|L}} \tau_\sigma \; D_{K,\sigma} h \; D_{K,\sigma}\phi'(h) \; \eta(f^\infty)_\sigma \\ & + \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{ext,K}} \tau_\sigma \; D_{K,\sigma} h \; D_{K,\sigma}\phi'(h) \; \eta(f^\infty)_\sigma \;\geq\; 0. \end{aligned} \tag{23}$$

In order to prove the non-negativity of the numerical dissipation we compare it to $C^{M_\phi}_\phi$ which is the numerical dissipation of a centered convective flux that we define hereafter as well as some complementary notation.

Definition 3.3. A function $M : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is called a *mean function* if it satisfies for all $s, t \in \mathbb{R}_+$,

(1) $M(s,t) = M(t,s)$,
(2) $M(s,s) = s$,
(3) If $s \leq t$, then $s \leq M(s,t) \leq t$.

We also define $M_\sigma : X_{u^b} \longrightarrow \mathbb{R}$ by $M_\sigma(u) = M(u_K, u_L)$ if $\sigma = K|L$ and $M_\sigma(u) = M(u_K, u_\sigma)$ otherwise. This is well defined thanks to the symmetry of M. For any such function M, we define the centered convective flux associated to M by

$$F^M_{K,\sigma}(h) = m(\sigma) \; U^\infty_{K,\sigma} \; M_\sigma(h)$$

and

$$C^M_\phi = \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \phi'(h_K) F^M_{K,\sigma}(h).$$

Finally for any entropy generating function ϕ, it is elementary to show that

$$M^{\phi}(s,t) = \frac{\varphi(s)-\varphi(t)}{\phi'(s)-\phi'(t)},$$

where $\varphi(s) = s\phi'(s) - \phi(s)$, defines a continuous mean function. We call it the ϕ-mean.

Remark 3.4. *Let us note that for the 2-entropy generating function $\phi_2(s) = (s-1)^2$, the corresponding ϕ_2-mean is the arithmetic average and therefore $F_{K,\sigma}^{M_{\phi_2}}$ is a centered approximation for the convective flux, namely for $\sigma \in \mathcal{E}_{int}$*

$$F_{K,\sigma}^{M_{\phi_2}}(h) = m(\sigma)\ U_{K,\sigma}^{\infty}\ \frac{h_K + h_L}{2}.$$

When choosing the generator of the physical entropy $\phi_1(s) = s\log(s) - s + 1$, the corresponding mean function is the logarithmic average reading $M_{\phi_1}(s,t) = (s-t)/(\log(s) - \log(t))$.

We are now ready to prove Proposition 3.2.

Proof of Proposition 3.2. In order to prove (23), we use (22) and (20) to get

$$D_{\phi}(f) = -\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\ D_{K,\sigma}h\ \phi'(h_K)\ \eta(f_\sigma^{\infty})$$

and the result stems from a discrete integration by parts.

Now, let us prove the non-negativity of C_ϕ. For M a mean function (see Definition 3.3), we perform a discrete integration by parts of $C_\phi - C_\phi^M$ which yields

$$C_\phi - C_\phi^M = -\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} (F_{K,\sigma}^{\rm conv} - F_{K,\sigma}^{M})D_{K,\sigma}(\phi'(h)) - \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{{\rm ext},K}} (F_{K,\sigma}^{\rm conv} - F_{K,\sigma}^{M})D_{K,\sigma}(\phi'(h))\,.$$

Now let us just remark that for any $\sigma \in \mathcal{E}_{\rm int}$ with $\sigma = K|L$,

$$\begin{aligned}
-(F_{K,\sigma}^{\rm conv} - F_{K,\sigma}^{M})D_{K,\sigma}(\phi'(h)) &= m(\sigma)U_{K,\sigma}^{\infty,+}\,(g(h_K,h_K) - g(h_K,h_L))\,(\phi'(h_L) - \phi'(h_K))\\
&+ m(\sigma)\,U_{K,\sigma}^{\infty,+}\,(M_\sigma(h) - h_K)\,(\phi'(h_L) - \phi'(h_K))\\
&+ m(\sigma)\,U_{K,\sigma}^{\infty,-}\,(h_L - M_\sigma(h))\,(\phi'(h_L) - \phi'(h_K))\\
&+ m(\sigma)\,U_{K,\sigma}^{\infty,-}\,(g(h_L,h_K) - g(h_L,h_L))\,(\phi'(h_L) - \phi'(h_K))\,,
\end{aligned}$$

where we used that $g(s,s) = s$. If $\sigma \in \mathcal{E}_{\rm ext}$, the same equation holds replacing h_L with h_σ. Therefore, since ϕ' and $g(s,\cdot)$ are monotonically non-decreasing functions and $M_\sigma(h)$ is always between h_K and h_L (*resp.* h_σ), the above quantity is non-negative. Hence, it proves that $C_\phi \geq C_\phi^M$.

Finally, a simple computation using two integrations by parts yields

$$\begin{aligned}
C_\phi^{M_\phi} &= -\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} F_{K,\sigma}^{M_\phi}D_{K,\sigma}(\phi'(h)) - \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{{\rm ext},K}} F_{K,\sigma}^{M_\phi}D_{K,\sigma}(\phi'(h))\,,\\
&= -\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} F_{K,\sigma}^{\infty}D_{K,\sigma}(\varphi(h)) - \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{{\rm ext},K}} F_{K,\sigma}^{\infty}D_{K,\sigma}(\varphi(h))\,,\\
&= \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{\infty}\varphi(h_K)\ =\ 0\,,
\end{aligned}$$

where we used (H3) in the last equality. Thus, $C_\phi \geq C_\phi^{M_\phi} = 0$. □

Remark 3.5. *Let us note that if we had used the fluxes $F^{M_\phi}_{K,\sigma}$ instead of $F^{conv}_{K,\sigma}$ in our scheme, then the contribution of the convection would be null in the discrete relative ϕ-entropy variation. However, the scheme would have been ϕ-dependent. With the upstream flux, we get the whole class of relative entropy inequalities at the cost of an additional numerical dissipation.*

4. Analysis of the schemes

4.1. **Implicit Euler.** Before stating our main result for the implicit scheme (16)-(20). Let us show that the control of a large class of relative ϕ-entropies yields L^∞ stability of the discrete solution. To establish this result with bounds that are independent of the discretization $\mathcal{D}$, L^∞ bound on the initial data, given by (H1) is mandatory.

Lemma 4.1. *Assume that for any entropy function ϕ, and $n \in \{0, \dots, N_T\}$*

$$H_\phi(h^n) \leq H_\phi(h^0)\,.$$

Then, under hypotheses (H1) *and* (H2)*, there exists a positive constant K_∞ depending only on K_{in}, η, m_∞ and M_∞ such that for all $(t,x) \in [0,T) \times \mathbb{R}^d$,*

$$0 \leq\ f_\delta(t,x),\ h_\delta(t,x)\ \leq\ K_\infty\,.$$

Proof. Let us define $\phi_{u_0}(\cdot) = (\cdot - u_0)^+$ if $u_0 > 1$ and $\phi_{u_0}(\cdot) = (\cdot - u_0)^-$ if $u_0 < 1$. It is differentiable on $\mathbb{R} \setminus \{u_0\}$ with derivative $\phi'_{u_0} = \mathbb{1}_{[u_0,+\infty)}$ if $u_0 > 1$ and $\phi'_{u_0} = -\mathbb{1}_{(-\infty,u_0]}$ if $u_0 < 1$, where $\mathbb{1}_A$ denotes the indicator function of the set A. The function ϕ_{u_0} is not an entropy function. However, if for any $n \in \mathbb{N}$ one has $H_{\phi_{u_0}}(h^n) = 0$ for $u_0 = 0$ and $u_0 = \eta(K_{\rm in})/m_\infty$, then $e_{\phi_{u_0},K}(h^n) = 0$ and $h^n_K \in [0, \eta(K_{\rm in})/m_\infty]$ for all cell K. Hence let us show that $H_{\phi_{u_0}}(h^n) = 0$ by an approximation argument.

Let $B : x \mapsto x/(\exp(x) - 1)$ be the Bernoulli function, which is a strictly convex $\mathcal{C}^2$ function. Then, for any $\varepsilon > 0$ and $u_0 \in \mathbb{R}$, one readily checks that

$$\phi_{\varepsilon,u_0} : u \longmapsto \phi_{\varepsilon,u_0}(u) = \varepsilon \left[B\left(\frac{u-u_0}{\varepsilon}\right) - B\left(\frac{1-u_0}{\varepsilon}\right)\right] + B'\left(\frac{1-u_0}{\varepsilon}\right)(1-u),$$

are entropy generating functions. When ε tends to 0, ϕ'_{ε,u_0} converges pointwise to ϕ'_{u_0} on $\mathbb{R} \setminus \{u_0\}$. Therefore, by dominated convergence, one has for any $n \in \{0, \dots, N_T\}$

$$0 \leq H^n_{\phi_{u_0}} \leq H^0_{\phi_{u_0}}\,.$$

Since, by (H2) and (H1), $H^0_{u_0} = 0$ for $u_0 = \eta(K_{\rm in})/m_\infty$ and $u_0 = 0$ then $H^n_{u_0} = 0$ for all $n \in \{0, \dots, N_T\}$. Hence we infer the uniform bounds on h_δ, and consequently on f_δ. □

Theorem 4.2 (Implicit Euler)**.** *Under hypotheses* (H1)-(H5) *the scheme* (16) *together with* (18)-(20) *satisfies the following properties.*

(i) *There exists a unique discrete solution $f \in X^{N_T+1}_{f^b}$;*

(ii) *there is a positive constant K_∞ depending only on K_{in}, η, m_∞ and M_∞ such that for all $(t,x) \in [0,T) \times \mathbb{R}^d$,*

$$0 \leq\ f_\delta(t,x),\ h_\delta(t,x)\ \leq\ K_\infty\,;$$

(iii) *the scheme preserves the steady state f^∞ and for any entropy function ϕ and $n \in \{0, \dots, N_T - 1\}$,*

$$\frac{H^{n+1}_\phi - H^n_\phi}{\Delta t} + D^{n+1}_\phi \leq 0 \quad \textit{and} \quad D^{n+1}_\phi \geq 0\,. \tag{24}$$

Proof. (i): The existence of a unique solution to the implicit scheme can be shown with a fixed point strategy close to that in [20, Remark 4.9] and we do not detail this part.
(iii): Let us derive the entropy inequality. The Taylor-Young theorem provides the existence of $\theta_K^{n,n+1} \in (\min(f_K^n, f_K^{n+1}),\ \max(f_K^n, f_K^{n+1}))$ such that

$$\begin{aligned} e_{\phi,K}^{n+1} - e_{\phi,K}^n &= \int_{f_K^n}^{f_K^{n+1}} \phi'\left(\frac{\eta(s)}{\eta(f_K^\infty)}\right) \mathrm{d}s \\ &= \phi'(h_K^{n+1})(f_K^{n+1} - f_K^n) - \frac{1}{2}\psi_K(\theta_K^{n,n+1})(f_K^{n+1} - f_K^n)^2 \\ &= -\frac{\Delta t}{m(K)}\phi'(h_K^{n+1}) \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} - \frac{1}{2}\psi_K(\theta_K^{n,n+1})(f_K^{n+1} - f_K^n)^2, \end{aligned}$$

where ψ_K is given by

$$\psi_K : x \mapsto \frac{\eta'(x)}{\eta(f_K^\infty)}\phi''\left(\frac{\eta(x)}{\eta(f_K^\infty)}\right). \tag{25}$$

Note that ψ_K is a positive function thanks to the positive monotony of η and ϕ'. From the definition of the dissipation in (22), it yields

$$\frac{H_\phi^{n+1} - H_\phi^n}{\Delta t} + D_\phi^{n+1} \le -C_\phi^{n+1} - \frac{1}{2\Delta t}\sum_{K\in\mathcal{T}} \psi_K(\theta_K^{n,n+1})(f_K^{n+1} - f_K^n)^2 m(K) \le 0,$$

where we applied Proposition 3.2 to control the first term of the right-hand side.
(ii): By summing (24) over n and using the positivity of D_ϕ^n, we obtain the boundedness of H_ϕ^n and we can use Lemma 4.1 to conclude. □

4.2. **Explicit Euler.** Before stating the main result on this scheme, let us introduce

$$a_{K,\sigma}^n = \begin{cases} d_\sigma\left(U_{K,\sigma}^{\infty,+}\ \dfrac{h_K^n - g(h_K^n, h_L^n)}{D_{K,\sigma}h^n} + U_{K,\sigma}^{\infty,-}\ \dfrac{g(h_L^n, h_K^n) - h_K^n}{D_{K,\sigma}h^n}\right), & \text{if } \sigma = K|L, \\ d_\sigma\left(U_{K,\sigma}^{\infty,+}\ \dfrac{h_K^n - g(h_K^n, h_\sigma)}{D_{K,\sigma}h^n} + U_{K,\sigma}^{\infty,-}\ \dfrac{g(h_\sigma, h_K^n) - h_K^n}{D_{K,\sigma}h^n}\right), & \text{otherwise,} \end{cases} \tag{26}$$

with the convention $a_{K,\sigma}^n = 0$ if $D_{K,\sigma}h^n = 0$. Then observe that we can use this $a_{K,\sigma}$ to reformulate the convective part of the scheme as a "diffusive term", thanks to the incompressibility of $U_{K,\sigma}^\infty$, which is a consequence of (H3). Indeed, for all $K \in \mathcal{T}$, we have

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\ a_{K,\sigma}\ D_{K,\sigma}h = \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{\mathrm{conv}} - h_K \sum_{\sigma\in\mathcal{E}_K} m(\sigma)U_{K,\sigma}^\infty = \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{\mathrm{conv}},$$

where we used (H3) in the last equality. Now we suppose that there is a positive constant V_∞ that does not depend on $\delta > 0$ and such that

$$\max_{K\in\mathcal{T}} \max_{\sigma\in\mathcal{E}_K} |U_{K,\sigma}^\infty| \le V_\infty\,. \tag{H7}$$

Thanks to the monotonicity and regularity properties of g from (H5), one has

$$0 \le a_{K,\sigma} \le C_g\ V_\infty\ \mathrm{diam}(\Omega),$$

where C_g is the Lipschitz constant of g on $[0,\ \eta(K_{\mathrm{in}})/m_\infty]^2$ and $\mathrm{diam}(\Omega)$ is the diameter of Ω.

Remark 4.3. *Hypothesis* (H7) *is the most restrictive of our assumptions since it implicitly demand uniform $W^{1,\infty}$ bound on the discrete steady state $\eta(f^\infty)$. A more natural and similar hypothesis is a uniform L^2 control on the velocity (instead of L^∞). It corresponds to discrete H^1 control on $\eta(f^\infty)$, uniformly in $\delta > 0$, which can be easily obtained for a finite volume discretization of an elliptic equation like* (2)*. For the convergence analysis in Section 5, this hypothesis follows from* (H10)*. Unfortunately, we need the stronger* (H7) *to obtain L^∞ stability for the explicit scheme.*

Theorem 4.4 (Explicit Euler)**.** *Let $f \in X_{f^b}^{N_T+1}$ be defined by the scheme* (17)-(20)*. Then, under hypotheses* (H1)-(H5), (H7) *the following properties hold.*

(i) *There exists a positive constant C_{L^∞} depending only on K_{in}, M_∞, V_∞, g, η and Ω such that under the CFL condition*

$$\max_{K\in\mathcal{T}} \frac{\Delta t}{m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \le C_{L^\infty},$$

there is a positive constant K_∞ depending only on K_{in}, η, m_∞ and M_∞ such that for all $(t,x) \in [0,T) \times \mathbb{R}^d$,

$$0 \le f_\delta(t,x),\ h_\delta(t,x) \le K_\infty\,.$$

(ii) *If Φ is a family of entropy functions with second derivate bounded between m_Φ and M_Φ, then there exists a positive constant C_{ent} depending only on on K_{in}, m_∞, M_∞, V_∞, g, η and Ω such that for every $\varepsilon \in (0,1)$, under the CFL condition*

$$\max_{K\in\mathcal{T}} \frac{\Delta t}{m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \le \min\left(C_{L^\infty},\, C_{ent}\,\frac{m_\Phi}{M_\Phi}\,\varepsilon\right) \tag{27}$$

for any $\phi \in \Phi$ and $n \in \{0,\dots,N_T-1\}$,

$$\frac{H_\phi^{n+1} - H_\phi^n}{\Delta t} + (1-\varepsilon) D_\phi^n \le 0 \quad \text{and} \quad D_\phi^n \ge 0\,. \tag{28}$$

Moreover the scheme preserves the stationary state f^∞.

Proof. (i): For explicit time discretization it is rather classical (see [20]) to use the convexity property of the scheme to show L^∞ stability. However due to the fact that flux are expressed in terms of h instead of f adds up some technicalities here. We have to proceed in two steps to show the heredity of the induction hypothesis $h^n \in J = [0, \eta(K_{\rm in})/m_\infty]$.

First note that for any $\delta \ge 0$, under the CFL condition

$$\max_{K\in\mathcal{T}} \frac{\Delta t}{m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \le C_1(\delta) := \frac{\delta}{K_{\rm in}\,(C_g\,V_\infty\,{\rm diam}(\Omega) + M_\infty)}\,,$$

one has $f_K^{n+1} \in I_\delta = [-\delta,\, K_{\rm in} + \delta]$ for all control volume $K \in \mathcal{T}$ since the scheme rewrites

$$f_K^{n+1} = f_K^n + \frac{\Delta t}{m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma (a_{K,\sigma} + \eta(f_K^\infty))\, D_{K,\sigma} h^n\,.$$

Let us define $M_\eta^\delta = \sup_{s\in I_\delta} \eta'(s)$. Now we show that under a possibly more restrictive CFL condition, h_K^n belongs to J. By the mean value theorem, there exists $g_K^n \in I_\delta$ such that

$$h_K^{n+1} - h_K^n = \frac{\eta'(g_K^n)}{\eta(f_K^\infty)}(f_K^{n+1} - f_K^n).$$

The scheme can then be rewritten as

$$\begin{aligned} h_K^{n+1} &= \left(1 - \frac{\eta'(g_K^n)\,\Delta t}{\eta(f_K^\infty)\,m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma(a_{K,\sigma} + \eta(f_K^\infty))\right) h_K^n \\ &+ \frac{\eta'(g_K^n)\,\Delta t}{\eta(f_K^\infty)\,m(K)} \sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int},K}\\ \sigma=K|L}} \tau_\sigma(a_{K,\sigma} + \eta(f_K^\infty))\, h_L^n \\ &+ \frac{\eta'(g_K^n)\,\Delta t}{\eta(f_K^\infty)\,m(K)} \sum_{\sigma\in\mathcal{E}_{\mathrm{ext},K}} \tau_\sigma(a_{K,\sigma} + \eta(f_K^\infty))\, h_\sigma^n\,. \end{aligned}$$

Under the CFL condition

$$\max_{K\in\mathcal{T}} \frac{\Delta t}{m(K)} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \le C_2(\delta) := \frac{m_\infty}{M_\eta^\delta\,(C_g\,V_\infty\,\mathrm{diam}(\Omega) + M_\infty)},$$

it provides h_K^{n+1} as a convex combination of elements of J and hence $\{h_K^{n+1}\}_{K\in\mathcal{T}} \subseteq J$. The CFL constant C_{L^∞} can then be taken as the supremum of $\min(C_1(\delta), C_2(\delta))$ when $\delta > 0$.

(ii): We proceed exactly as in the proof of Theorem 4.2 to get the existence of $\theta_K^{n,n+1}$ such that

$$\frac{H_\phi^{n+1} - H_\phi^n}{\Delta t} + D_\phi^n \le -C_\phi^n + \frac{1}{2\Delta t} \sum_{K\in\mathcal{T}} \psi_K(\theta_K^{n,n+1})(f_K^{n+1} - f_K^n)^2 m(K), \tag{29}$$

for ψ_K defined by (25). Note that the sign of the last term has changed compared to the implicit scheme. Using the scheme, the last term in (29) can be estimated with the Cauchy-Schwartz inequality as

$$\begin{aligned} &\frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \frac{1}{m(K)} \psi_K(\theta_K^{n,n+1}) \left(\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^n\right)^2 \le \\ &\qquad \frac{\Delta t\, M_\Phi\, M_\eta^0\,(C_g\,V_\infty\,\mathrm{diam}(\Omega) + M_\infty)^2}{m_\infty} \sum_{K\in\mathcal{T}} \frac{1}{m(K)} \left(\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\right) \left(\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\,(D_{K,\sigma}h^n)^2\right), \end{aligned}$$

Then, using that

$$D_\phi^n \ge \frac{m_\Phi\, m_\infty}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\ (D_{K,\sigma}h^n)^2\,,$$

it yields (28) provided that the CFL condition is satisfied with constant

$$C_{\mathrm{ent}} = \frac{m_\infty^2}{2\,M_\eta^0\,(C_g\,V_\infty\,\mathrm{diam}(\Omega) + M_\infty)^2}.$$

□

4.3. **Long time behavior.** In this section, we study the long-time behavior of the discrete solution in the linear case $\eta(s) = s$. We prove exponential decay to equilibrium with the following strategy. From a classical discrete Poincaré inequality, we establish a discrete Poincaré-Sobolev inequality for controlling H_{ϕ_2} by D_{ϕ_2}, where we recall that $\phi_2(s) = (s-1)^2$. Then, properties (24) and (28) provide exponential decay to equilibrium by a discrete Gronwall-type argument for H_{ϕ_2} and all relative ϕ-entropies controlled by the latter.

Remark 4.5. *The question of the existence of a general ϕ-Poincaré-Sobolev functional inequality for any entropy generating function ϕ, even in the continuous setting, goes way beyond the scope of this paper and we refer to* [7, 3] *for discussions on this matter.*

The 2-entropy and its dissipation are closely related to, respectively, the L^2 norm

$$\|u\|_{0,2} \;=\; \left(\sum_{K\in\mathcal{T}} m(K)\;|u_K|^2\right)^{1/2},$$

and the discrete H^1 semi-norm

$$|u|_{1,2,\mathcal{T}} \;=\; \left(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma\;|D_{K,\sigma}u|^2 + \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{{\rm ext},K}} \tau_\sigma\;|D_{K,\sigma}u|^2\right)^{1/2}, \tag{30}$$

for which M. Bessemoulin-Chatard, C. Chainais-Hillairet and F. Filbet proved the following discrete Poincaré inequality in [5, Theorem 6], based on older work that may be found in references therein. Before stating it, we need to introduce the following regularity constraint on the mesh. There is a positive constant ξ_1 that does not depend on $\mathcal{D}$ such that

$$\forall K\in\mathcal{T},\ \forall\sigma\in\mathcal{E}_K,\quad d_{K,\sigma}\geq\xi_1\,d_\sigma \tag{H8}$$

Proposition 4.6 ([5])**.** *Under hypothesis* (H8)*, there exists a constant C_P only depending on Ω such that for all $u\in X_0$, it holds*

$$\|u\|_{0,2} \;\leq\; \frac{C_P}{\xi_1^{1/2}}|u|_{1,2,\mathcal{T}}\,.$$

As a consequence we have the following chain of inequalities. For any $h\in X_1$

$$H_{\phi_2}(h)\;\leq\;M_\infty\,\|h-1\|_{0,2}^2\;\leq\;\frac{C_P^2\,M_\infty}{\xi_1}|h|_{1,2,\mathcal{T}}^2\;\leq\;\frac{C_P^2\,M_\infty}{\xi_1\,m_\infty}D_{\phi_2}(h)\,. \tag{31}$$

Theorem 4.7. *(Exponential return to equilibrium)*
We suppose that η is the identity function, and that the assumptions of Theorem 4.2 (respectively those of Theorem 4.4 as well as the CFL condition (27) *for $\Phi=\{\phi_2\}$) and* (H8) *are satisfied. In the implicit case, we also assume that Δt is smaller that a given constant k. Then, the solution f of the scheme* (16),(18)-(20) *(respectively* (18)-(20)*) is such that there are two positive constants, κ depending only on Ω, m_∞, M_∞, ξ_1, and C_{in} depending additionally on $H^0_{\phi_2}$ (and k in the implicit case) such that for all $n\in\{0,\dots,N_T-1\}$,*

$$H^n_{\phi_2}\leq C_{in}\;e^{-\kappa t^n}\,.$$

As a consequence, for all $n\in\{0,\dots,N_T-1\}$, it holds

$$\|f_\delta(t^n,\cdot)-f_\delta^\infty(\cdot)\|^2_{L^1(\Omega)}\leq C_{in}\;e^{-\kappa t^n}.$$

Proof. By combining (31) and (24) (respectively (28)), we obtain $H^{n+1}_{\phi_2}\leq(1+\kappa\Delta t)^{-1}\,H^n_{\phi_2}$ in the implicit case with $\kappa=\xi_1\,m_\infty/(C_P^2\,M_\infty)$ and $H^{n+1}_{\phi_2}\leq(1-\kappa\Delta t)\,H^n_{\phi_2}$ in the explicit case with $\kappa=(1-\varepsilon)\xi_1\,m_\infty/(C_P^2\,M_\infty)$ where ε is in the CFL condition (27). Hence we get exponential decay of the 2-entropy, with $C_{\rm in}=H^0_{\phi_2}$ in the explicit case and $C_{\rm in}=H^0_{\phi_2}\,\exp(\kappa^2k^2/2)$ in the implicit case. Finally, the estimate in L^1 is obtained from a Cauchy-Schwartz inequality

$$\|f_\delta(t^n)-f_\delta^\infty\|^2_{L^1(\Omega)}\leq\|f_\delta^\infty\|_{L^1(\Omega)}\|f_\delta(t^n)/\sqrt{f_\delta^\infty}-\sqrt{f_\delta^\infty}\|^2_{L^2(\Omega)}=\|f_\delta^\infty\|_{L^1(\Omega)}\,H^n_{\phi_2}\,,$$

and the decay of H_{ϕ_2}.

□

Remark 4.8. *The exponential decay holds for many other ϕ-entropies. For instance, let us restrict our class of entropy generating functions to that introduced by Arnold, Markowich, Toscani and Unterreiter in* [3]*, that is those satisfying $\phi \in \mathcal{C}^4(\mathbb{R}_+)$ and*

$$\left(\phi'''\right)^2 \;\leq\; \frac{1}{2}\phi''\;\phi^{IV}.$$

Mark that the physical and p-entropies are generated by these entropy functions. As a consequence of [3, Lemma 2.6]*, one has that any such ϕ is bounded from above by a quadratic entropy function, namely*

$$\frac{\phi(s)}{\phi''(1)} \;\leq\; \phi_2(s) = (s-1)^2.$$

Therefore, since the same inequality holds for the corresponding relative entropies, namely $H_\phi \leq \phi''(1) H_{\phi_2}$, the exponential decay holds for any relative ϕ-entropy in this class.

5. Convergence

In this section, we prove that when $\delta \to 0$, the sequence $(h_\delta)_{\delta>0}$ converges to a solution of (5) in the following sense.

Definition 5.1. We say that h is a solution of (5) starting at initial data f^{in} if

- $h - 1 \in L^2(0,T;\, H_0^1(\Omega))$
- $f = \eta^{-1}(\eta(f^\infty)h) \in L^1(\Omega_T)$
- For any test function $\psi \in \mathcal{C}_c^\infty(\Omega_T)$, it holds

$$-\iint_{\Omega_T} [f\,\partial_t\psi + (\mathbf{U}^\infty\, h - \eta(f^\infty)\,\nabla h)\cdot\nabla\psi] - \int_\Omega f^{\text{in}}\,\psi(0,\cdot) \;=\; 0\,.$$

Thanks to the reformulation of Equation (1) into (5), the only nonlinearity lies in the first term. In order to prove convergence of the scheme we need strong compactness to recover pointwise convergence and identify the limit of $(f_\delta)_\delta$ as $\eta^{-1}(\eta(f^\infty)h)$ with h the limit of $(h_\delta)_\delta$.

Except for this originality, the strategy is fairly standard. We first derive uniform-in-δ estimates on h_δ and its (discrete) gradient in $\mathbf{x}$ in order to get compactness in the space variable. Then thanks to the structure of the equation compactness in time is obtained from previous estimates. Using the consistency we shall then take limits in our scheme and recover a global weak solution in the sense of Definition 5.1.

5.1. Hypotheses.

We discuss specific hypotheses needed for the convergence result to hold. While, it was not crucial until now, the orthogonality condition (H6) is very important in this part, for consistency of discrete gradients. As in [14, 15, 4], the latter are introduced on a dual mesh. For $\sigma = K|L$, we define T_σ as the cell with vertices $\mathbf{x}_K$, $\mathbf{x}_L$ and those of σ. If $\sigma \in \mathcal{E}_{\text{ext}}$ then T_σ is the cell with vertices $\mathbf{x}_K$ and those of σ, where K is the only cell having $\sigma \in \mathcal{E}_K$. We refer to [15, Fig. 1] for a visualization. On this dual mesh we set

$$\nabla^\delta h_\delta(t,\mathbf{x}) = \begin{cases} \dfrac{m(\sigma)}{m(T_\sigma)}\, D_{K,\sigma} h^{n+1}\,\mathbf{n}_{K,\sigma}, & \text{if } \mathbf{x}\in T_\sigma \text{ and } t\in[t^n,t^{n+1}) \quad \text{(Implicit case)} \\ \dfrac{m(\sigma)}{m(T_\sigma)}\, D_{K,\sigma} h^{n}\,\mathbf{n}_{K,\sigma}, & \text{if } \mathbf{x}\in T_\sigma \text{ and } t\in[t^n,t^{n+1}) \quad \text{(Explicit case)} \end{cases}$$

Remark 5.2. *The discrete gradients are well-defined since for $u \in X_{u_b}$ and $\sigma = K|L$, the product $D_{K,\sigma}u\;\mathbf{n}_{K,\sigma}$ is independent of the cell K. For $\sigma \in \mathcal{E}_{ext}$, there is no ambiguity since only one cell is sharing the edge.*

In order to obtain convergence of these gradients we need the following regularity hypothesis on the mesh (see [15] and references therein for related comments). There exist a constant $\xi_2 > 1$ that does not depend on the discretization such that

$$\text{(H9)} \qquad \forall \sigma \in \mathcal{E}, \quad m(T_\sigma) \leq m(\sigma) d_\sigma \leq \xi_2 \, m(T_\sigma) \, .$$

Since our scheme is based on a discretization of the stationary equation (2), one needs the convergence of the latter to obtain that of the former. Therefore we assume that both the discrete steady state and the discrete velocity converge to their continuous counterparts. First, let us define them on the dual mesh, by

$$\eta(f^\infty)_\delta(\mathbf{x}) = \eta(f^\infty)_\sigma \quad \text{for} \quad \mathbf{x} \in T_\sigma \, ,$$

and

$$\mathbf{U}^\infty_\delta(\mathbf{x}) = \frac{m(\sigma) d_\sigma}{m(T_\sigma)} \, U^\infty_{K,\sigma} \, \mathbf{n}_{K,\sigma} \quad \text{for} \quad \mathbf{x} \in T_\sigma \, .$$

Then, when $\delta \to 0$, we assume that

$$\text{(H10)} \qquad \begin{cases} \eta(f^\infty)_\delta & \longrightarrow \quad \eta(f^\infty) \quad \text{strongly in } L^2(\Omega) \\ \mathbf{U}^\infty_\delta & \longrightarrow \quad \mathbf{U}^\infty \qquad \text{weakly in } L^2(\Omega) \end{cases}$$

Remark 5.3. *Hypothesis* (H10) *is very natural. Indeed, since $\eta(f^\infty)$ satisfies a linear non-degenerate second order elliptic equation, one expects to get uniform H^1 control on discrete solutions of any reasonable finite volume scheme.*

In the rest of Section 5, all results are stated for the implicit scheme. However, everything holds also for the explicit scheme with minor modifications provided that one adds hypothesis (H7) and the CFL condition (27) with $\Phi = \{\phi_2\}$ to each proposition.

5.2. **Compactness.** In the following proposition, we provide the uniform L^∞ and $L^2(0,T;H^1)$ estimates we need for compactness. The discrete $L^2(0,T;H^1)$ norm is defined by

$$\|h_\delta\|_{1,2,\mathcal{D}} = \left(\sum_{n=0}^{N_T-1} \Delta t \, |h^{n+1}|^2_{1,2,\mathcal{T}} \right)^{1/2}$$

for the implicit scheme.

Proposition 5.4. *(Uniform estimates) Under hypotheses* (H1)-(H6) *and* (H9)-(H10)*, there are positive constants K_∞ and $K_{1,2}$ depending only on K_{in}, η, m_∞ and M_∞ such that the solution of the scheme* (16) *together with* (18)-(20) *satisfies*

$$0 \leq f_\delta(t,\mathbf{x}), \, h_\delta(t,\mathbf{x}) \leq K_\infty \, ,$$

and

$$\|h_\delta\|_{1,2,\mathcal{D}} \leq K_{1,2}.$$

Proof. The first estimate is proved in Theorem 4.2 and 4.4. The second estimate is a consequence of the boundedness of the dissipation of 2-entropy and of the lower bound on the steady state since

$$m_\infty \, |h^n|^2_{1,2,\mathcal{T}} \leq D^n_{\phi_2} \, .$$

By summing (24) over n and inserting the above inequality one obtains the result. □

From the first result of Proposition 5.4 one obtains weak-$\star$ compactness in L^∞. Now, we need to gain weak compactness on the sequence of discrete gradients as well as strong compactness on the sequence of discrete solutions. The strong compactness is obtained by the Riesz-Fréchet-Kolmogorov criterion on spacetime translates. In order to recover right boundary conditions at the limit the criterion is stated on the whole space $\mathbb{R}^{d+1}$ for $\tilde{h}_\delta$ defined by

$$\tilde{h}_\delta(t,x) = \begin{cases} h_\delta(t,\mathbf{x}) - 1 & \text{if } (t,x) \in \Omega_T\,, \\ 0 & \text{if } (t,x) \in \mathbb{R}^{d+1} \setminus \overline{\Omega_T}\,. \end{cases}$$

The proof of the following lemma can be readily adapted from [20, Lemma 4.3 and 4.7].

Lemma 5.5. *Under hypotheses* (H1)-(H6) *and* (H9)-(H10)*, there is a positive constant K_2 depending only on K_∞, $K_{1,2}$, Ω, T and η such that the solution of the scheme* (16) *together with* (18)-(20) *satisfies for all $\mathbf{y} \in \mathbb{R}^d$*

$$\|\tilde{h}_\delta(\cdot,\cdot+\mathbf{y}) - \tilde{h}_\delta(\cdot,\cdot)\|^2_{L^2(\mathbb{R}^{d+1})} \;\le\; K_2\,|\mathbf{y}|\,(|\mathbf{y}|+\delta)\,,$$

and for all $\tau \in (0,T)$

$$\|\tilde{h}_\delta(\cdot+\tau,\cdot) - \tilde{h}_\delta(\cdot,\cdot)\|^2_{L^2((0,T-\tau)\times\mathbb{R}^d)} \;\le\; K_2\,|\tau|\,.$$

From the previous results we obtain the following convergences.

Proposition 5.6. *There is a function $h \in L^\infty(0,T;H^1(\Omega))$ such that as $\delta \to 0$ and up to the extraction of a subsequence,*

$$\begin{aligned} h_\delta &\longrightarrow h && \textit{strongly in } L^2(\Omega_T), \\ \nabla^\delta h_\delta &\longrightarrow \nabla h && \textit{weakly in } \left(L^2(\Omega_T)\right)^d, \\ f_\delta &\longrightarrow f && \textit{strongly in } L^1(\Omega_T), \end{aligned}$$

with $f(t,\mathbf{x}) = \eta^{-1}(\eta(f^\infty(t,\mathbf{x}))\,h(t,\mathbf{x}))$, $(t,\mathbf{x})-$almost everywhere in Ω_T. Moreover $h-1 \in L^\infty(0,T;H^1_0(\Omega))$.

Proof. The first result is a consequence of the Riesz-Fréchet-Kolmogorov L^p compactness criterion. It yields the convergence of $(\tilde{h}_\delta)_{\delta>0}$ to a function $\tilde{h}$ strongly in $L^2([0,T)\times\mathbb{R}^d)$. By taking limits in the first space translates estimate, we know that $\tilde{h} \in L^2(0,T;H^1(\mathbb{R}^d))$. Since by definition $\tilde{h}_\delta(t,\cdot) \equiv 0$ on $\mathbb{R}^d \setminus \overline{\Omega}$, we get that $\tilde{h} \in L^2(0,T;H^1_0(\Omega))$. Hence $(h_\delta)_\delta$ converges strongly in $L^2(\Omega_T)$ to $h = 1+\tilde{h}$ and up to the extraction of a sparser subsequence, it also converges almost everywhere. Since h_δ and $\eta(f^\infty)_\delta$ are uniformly bounded, η^{-1} is continuous and the cylinder Ω_T is bounded, one can apply the dominated convergence theorem to obtain the convergence of f_δ towards f. For the weak convergence of discrete gradients we refer to [14, Lemma 4.4] with minor modifications. □

5.3. **Convergence of the scheme.**

Theorem 5.7. *Under hypotheses* (H1)-(H6) *and* (H9)-(H10)*, the function h defined in Proposition 5.6 is a solution of Equation* (5) *in the sense of Definition 5.1.*

Let us mention that we follow hereafter the methods of proof from [14, 15].

Proof. Let us consider a test function $\psi \in \mathcal{C}^\infty_c([0,T)\times\Omega)$ and set $\psi^n_K = \psi(t^n,\mathbf{x}_K)$ for all $K \in \mathcal{T}$ and $n = 0,\dots,N_T$. We suppose that the mesh size δ is sufficiently small for the

inclusion $\mathrm{supp}(\psi) \subset [0,(N_T-1)\Delta t) \times \{\mathbf{x} \in \Omega,\ d(\mathbf{x},\Gamma) > \delta\}$ to hold. With this assumption sums on exterior edges disappear as one may remark in various terms hereafter. Let us define

$$\begin{cases} T_{10}(\delta) &= -\displaystyle\iint_{\Omega_T} f_\delta(t,\mathbf{x})\,\partial_t\psi(t,\mathbf{x})\ \mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_\Omega f_\delta^{\mathrm{in}}(\mathbf{x})\,\psi(0,\mathbf{x})\ \mathrm{d}\mathbf{x}, \\ T_{20}(\delta) &= \displaystyle\iint_{\Omega_T} \eta(f^\infty)_\delta\,\nabla^\delta h_\delta\cdot\nabla\psi\ \mathrm{d}\mathbf{x}\,\mathrm{d}t, \\ T_{30}(\delta) &= -\displaystyle\iint_{\Omega_T} h_\delta\,\mathbf{U}_\delta^\infty\cdot\nabla\psi\ \mathrm{d}\mathbf{x}\,\mathrm{d}t. \end{cases}$$

From the results of Proposition 5.6 and assumption (H10), it is clear that

$$T_{10}(\delta)+T_{20}(\delta)+T_{30}(\delta) \longrightarrow -\iint_{[0,T)\times\Omega} [f\,\partial_t\psi + (h\,\mathbf{U}^\infty - \eta(f^\infty)\,\nabla h)\cdot\nabla\psi] - \int_\Omega f^{\mathrm{in}}\,\psi(0,\cdot)$$

as $\delta \to 0$. Now let us show that it also converges to 0. By multiplying the scheme by $\Delta t\,\psi_K^n$ and summing over n and K we obtain

$$T_1(\delta)+T_2(\delta)+T_3(\delta) = 0,$$

with

$$\begin{cases} T_1(\delta) &= \displaystyle\sum_{n=0}^{N_T}\sum_{K\in\mathcal{T}} m(K)\,(f_K^{n+1}-f_K^n)\,\psi_K^n, \\ T_2(\delta) &= \displaystyle\sum_{n=0}^{N_T}\Delta t\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}} \tau_\sigma\,\eta(f^\infty)_\sigma\,(h_K^{n+1}-h_L^{n+1})\,\psi_K^n, \\ T_3(\delta) &= \displaystyle\sum_{n=0}^{N_T}\Delta t\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}} m(\sigma)\left[U_{K,\sigma}^{\infty,+}g(h_K^{n+1},h_L^{n+1}) - U_{K,\sigma}^{\infty,-}g(h_L^{n+1},h_K^{n+1})\right]\psi_K^n\,. \end{cases}$$

Let us show that each $T_i(\delta)$ gets asymptotically close to $T_{i0}(\delta)$, for $i=1$, 2 or 3, as δ goes to 0. After a discrete integration by parts, one gets

$$\begin{aligned} T_1(\delta) &= \sum_{n=0}^{N_T}\sum_{K\in\mathcal{T}} m(K)\,f_K^{n+1}\,(\psi_K^n-\psi_K^{n+1}) - \sum_{K\in\mathcal{T}} m(K)\,f_K^0\,\psi_K^0\,, \\ &= -\sum_{n=0}^{N_T}\sum_{K\in\mathcal{T}}\int_{t^n}^{t^{n+1}}\int_K f_K^{n+1}\,\partial_t\psi(t,\mathbf{x}_K)\ \mathrm{d}\mathbf{x}\,\mathrm{d}t - \sum_{K\in\mathcal{T}}\int_K f_K^0\,\psi(0,\mathbf{x}_K)\ \mathrm{d}\mathbf{x}\,, \end{aligned}$$

which yields

$$|T_1(\delta)-T_{10}(\delta)| \le \delta\ (T+1)\,m(\Omega)\,\|f_\delta\|_{L^\infty(\Omega_T)}\,\|\psi\|_{\mathcal{C}^2(\overline{\Omega_T})} \to 0\,,$$

as $\delta \to 0$ by Proposition 5.4. Concerning the diffusion term one has, integrating by parts

$$T_2(\delta) = \sum_{n=0}^{N_T}\Delta t\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}} \tau_\sigma\,\eta(f_\sigma^\infty)\,(h_K^{n+1}-h_L^{n+1})\,(\psi_K^n-\psi_L^n)$$

and

$$T_{20}(\delta) = \sum_{n=0}^{N_T}\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}} m(\sigma)\,\eta(f^\infty)_\sigma\,(h_L^{n+1}-h_K^{n+1})\int_{t^n}^{t^{n+1}}\frac{1}{m(T_\sigma)}\int_{T_\sigma}\nabla\psi\cdot\mathbf{n}_{K,\sigma}\ \mathrm{d}\mathbf{x}\,\mathrm{d}t.$$

Now just note that by the orthogonality hypothesis (H6) on the mesh and the regularity of ψ one has,

$$\left|\psi_L^n - \psi_K^n - \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \frac{1}{m(T_\sigma)}\int_{T_\sigma} \nabla\psi\cdot d_\sigma \mathbf{n}_{K,\sigma}\ \mathrm{d}\mathbf{x}\,\mathrm{d}t\right| \le \|\psi\|_{\mathcal{C}^2(\overline{\Omega_T})}\,\delta. \tag{32}$$

Therefore, with a Cauchy-Schwartz inequality and thanks to the regularity of the mesh one obtains

$$|T_2(\delta) - T_{20}(\delta)| \le \delta\ \sqrt{\xi_2\,T}\,\|\eta(f^\infty)_\delta\|_{L^2(\Omega)}\,\|h_\delta\|_{1,2,\mathcal{D}}\,\|\psi\|_{\mathcal{C}^2(\overline{\Omega_T})} \to 0\,,$$

as $\delta\to 0$, by Proposition 5.4 and assumption (H10). Finally let us deal with the convection term. As it is classical we transform the upwind form of $T_3(\delta)$ into the sum of a numerical diffusion and a centered flux yielding $T_3(\delta) = T_{31}(\delta) + T_{32}(\delta)$, with

$$\begin{aligned}
T_{31}(\delta) &= \frac{1}{2}\sum_{n=0}^{N_T}\Delta t\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\left|U_{K,\sigma}^\infty\right|\left(g(h_K^{n+1},h_L^{n+1}) - g(h_L^{n+1},h_K^{n+1})\right)\,\psi_K^n\,,\\
&= \frac{1}{2}\sum_{n=0}^{N_T}\Delta t\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\left|U_{K,\sigma}^\infty\right|\left(g(h_K^{n+1},h_L^{n+1}) - g(h_L^{n+1},h_K^{n+1})\right)\,(\psi_K^n - \psi_L^n)\,,
\end{aligned}$$

whereas $T_{32}(\delta)$ is

$$\begin{aligned}
T_{32}(\delta) &= \frac{1}{2}\sum_{n=0}^{N_T}\Delta t\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\,U_{K,\sigma}^\infty\,\left(g(h_K^{n+1},h_L^{n+1}) + g(h_L^{n+1},h_K^{n+1})\right)\,\psi_K^n\,,\\
&= \sum_{n=0}^{N_T}\Delta t\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\,U_{K,\sigma}^\infty\,g(h_K^{n+1},h_L^{n+1})\,(\psi_K^n - \psi_L^n)\,,\\
&= T_{321}(\delta)\ +\ T_{322}(\delta)\,,
\end{aligned}$$

with $T_{321}(\delta)$ and $T_{322}(\delta)$ given by

$$\begin{aligned}
T_{321}(\delta) &= \sum_{n=0}^{N_T}\Delta t\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\,U_{K,\sigma}^\infty\,(g(h_K^{n+1},h_L^{n+1}) - h_K^{n+1})\,(\psi_K^n - \psi_L^n)\,,\\
T_{322}(\delta) &= \sum_{n=0}^{N_T}\Delta t\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\,U_{K,\sigma}^\infty\,h_K^{n+1}\,(\psi_K^n - \psi_L^n)\,.
\end{aligned}$$

Similarly we introduce the decomposition $T_{30}(\delta) = T_{310}(\delta) + T_{320}(\delta)$ with

$$\begin{aligned}
T_{310}(\delta) &= -\sum_{n=0}^{N_T}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\int_{t^n}^{t^{n+1}}\int_{T_\sigma\cap L}(h_L^{n+1} - h_K^{n+1})\,\mathbf{U}_\delta^\infty\cdot\nabla\psi\ \mathrm{d}\mathbf{x}\,\mathrm{d}t\,,\\
T_{320}(\delta) &= -\sum_{n=0}^{N_T}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\int_{t^n}^{t^{n+1}}\int_{T_\sigma} h_K^{n+1}\,\mathbf{U}_\delta^\infty\cdot\nabla\psi\ \mathrm{d}\mathbf{x}\,\mathrm{d}t\,.
\end{aligned}$$

Let us prove that $T_{31}(\delta)$, $T_{310}(\delta)$, $T_{321}(\delta)$ and $T_{322}(\delta) - T_{320}(\delta)$ converge to zero when δ goes to zero. First, notice that using the Lipschitz continuity of g, the definition of $\mathbf{U}_\delta^\infty$ and eventually

the Cauchy-Schwartz inequality yields

$$|T_{321}(\delta)| + |T_{31}(\delta)| \;\le\; 2\,\delta\,\|g\|_{W^{1,\infty}([0,\,\|h_\delta\|_{L^\infty}]^2)}\;\|\psi\|_{\mathcal{C}^1(\overline{\Omega_T})}\,\|h_\delta\|_{1,2,\mathcal{D}}\,\|\mathbf{U}_\delta^\infty\|_{L^2(\Omega_T)}.$$

Then, similarly one obtains

$$|T_{310}(\delta)| \;\le\; \delta\,\|\psi\|_{\mathcal{C}^1(\overline{\Omega_T})}\,\|h_\delta\|_{1,2,\mathcal{D}}\,\|\mathbf{U}_\delta^\infty\|_{L^2(\Omega_T)},$$

and by Proposition 5.4 and hypotheses (H5) (H10) both right-hand sides go to 0 as δ goes to 0. Finally

$$T_{322}(\delta)-T_{320}(\delta) \;=\; \sum_{n=0}^{N_T}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} m(\sigma)\;\; U_{K,\sigma}^\infty\, h_K^{n+1}\int_{t^n}^{t^{n+1}}\left(\psi_K^n-\psi_L^n-\frac{1}{m(T_\sigma)}\int_{T_\sigma}\nabla\psi\cdot d_\sigma\mathbf{n}_{K,\sigma}\;\mathrm{d}\mathbf{x}\right)\mathrm{d}t,$$

and by using (32) we obtain,

$$|T_{322}(\delta)-T_{320}(\delta)| \;\le\; \delta\sqrt{T\,m(\Omega)}\,\|\psi\|_{\mathcal{C}^2(\overline{\Omega_T})}\,\|h_\delta\|_{L^\infty(\Omega_T)}\,\|\mathbf{U}_\delta^\infty\|_{L^2(\Omega_T)}.$$

Hence, $|T_i(\delta)-T_{i0}(\delta)|\to 0$ as $\delta\to 0$, for any $i\in\{1,2,3\}$. □

6. Numerical simulations

6.1. **Implementation.** Before presenting our numerical results let us state an important remark concerning the implementation of our scheme. By Theorem 4.7, we expect the solution h_δ of the scheme (16),(18)-(20) or (17)-(20) to converge to 1 when time goes to infinity. Due to floating point numbers repartition there can be non-negligible numerical errors in the computation of difference operators leading to saturation of ϕ-entropies. To avoid this issue, the scheme should be implemented in the following way. We introduce new unknowns defined by,

$$\tilde f_K^n \;=\; f_K^n - f_K^\infty, \qquad \tilde h_K^n \;=\; h_K^n - 1\,,$$

for $n\in\{0,\dots,N_T\}$ and $K\in\mathcal{T}$. One can readily check that in the upwind case $g(s,t)=s$, thanks to (H3), the schemes remain unchanged by replacing f by $\tilde f$ and h by $\tilde h$. Even for other g, this modified scheme is the discretization of

$$\frac{\partial}{\partial t}(f-f^\infty) \;+\; \nabla\cdot\left(\mathbf{U}^\infty\,(h-1) \;-\; \eta(f^\infty)\nabla(h-1)\right) \;=\; 0\,,$$

which is the same as (5) since f^∞ is steady and $\mathbf{U}^\infty$ is incompressible.

The new unknowns converge to 0 when time goes to infinity and thus differences are computed with a better precision. Hence one should solve the scheme (and compute ϕ-entropies, dissipations, *etc...*) on the new unknowns. Moreover boundary conditions on $\tilde h$ become homogeneous, which actually makes the implementation easier.

In the following test cases, the implicit scheme (16),(18)-(20) is used only for linear models in Section 6.2 - 6.4. The nonlinear model in Section 6.5 is solved with the explicit scheme (17)-(20). For the first test cases we use the linear solver SuperLU [27], which provides efficient results for large sparse and non-symmetric systems by performing a sparsity-preserving LU factorization. Since the advection field is steady in Equation (5), the resolution matrix can be factorized only once at the beginning of the simulation.

6.2. **Proof of concept.** In this part, we provide a numerical experiment showing the spatial accuracy of our scheme, especially in the long-time dynamics. It is performed on the following one dimensional toy model. The test case is the linear ($\eta(s) = s$) drift-diffusion equation (1) endowed with the (scalar) advection $E(x) = 1$ and set on the domain $\Omega = (0, 1)$. With the boundary conditions $f(t, 0) = 2$ and $f(t, 0) = 1 + \exp(1)$ the function

$$f(t, x) = 1 + \exp(x) + \exp\left(\frac{x}{2} - \left(\pi^2 + \frac{1}{4}\right) t\right) \sin(\pi x)$$

is the exact solution of (1) and converges to the stationary state

$$f^\infty(x) = 1 + \exp(x)\,,$$

as time goes to infinity.

In order to illustrate the advantage of our approach compared to the one consisting in a direct approximation of (1), we perform numerical simulations using our scheme (16), (18)-(20) and a classical finite volume discretization of (1) with upwind flux for the convective term and two-points approximation of the gradient for the diffusive term. Both schemes are implicit in time and we shall call the former "ϕ-entropic" and the latter "upwind".

The domain is discretized by the regular Cartesian mesh $\mathcal{T} = \{K_i := (x_i - \Delta x/2,\, x_i + \Delta x/2),\ i = 0, \ldots, N-1\}$ where $x_i = \Delta x/2 + i\Delta x$ and $\Delta x = 1/N$. Concerning the time discretization, the final time is $T = 5$ and we choose the small time step $\Delta t = 10^{-6}$ in order to minimize the error due to time discretization. For the implementation of our scheme we explicitly compute both the discrete steady solution and the steady flux.

We define the L^p error at time t between the reconstruction of the approximate solution f_N (corresponding to f_δ with previous notation) and the projection of the analytic solution on the mesh by

$$\epsilon_N^p(t) = \|\Pi_N f(t, \cdot) - f_N(t, \cdot)\|_{L^p(\Omega)}$$

where $\Pi_N f(t, x) = f_i^n$ if $(t, x) \in [t^n, t^{n+1}) \times K_i$ with f_i^n a numerical approximation of the average of $f(t^{n+1}, \cdot)$ on the cell K_i, computed with the trapezoidal rule. In Table 1, we measure, for both schemes and for different number of points N, the global error and experimental order of accuracy, respectively given by

$$e_N^p = \sup_{t \in [0,T)} \epsilon_N^p(t), \quad k_{2N}^p = |\log(e_{2N}) - \log(e_N)|/\log(2)\,.$$

N	Error e_N^1 ϕ-entropic	Order	Error e_N^1 Upwind	Order	Error e_N^∞ ϕ-entropic	Order	Error e_N^∞ Upwind	Order
20	$2.07.10^{-3}$		$4.28.10^{-3}$		$3.33.10^{-3}$		$7.38.10^{-3}$	
40	$1.21.10^{-3}$	0.77	$2.36.10^{-3}$	0.86	$1.93.10^{-3}$	0.79	$4.03.10^{-3}$	0.87
80	$6.45.10^{-4}$	0.91	$1.24.10^{-3}$	0.93	$1.02.10^{-3}$	0.91	$2.11.10^{-3}$	0.94
160	$3.30.10^{-4}$	0.97	$6.30.10^{-4}$	0.97	$5.22.10^{-4}$	0.97	$1.07.10^{-3}$	0.98
320	$1.64.10^{-4}$	1.01	$3.15.10^{-4}$	1.00	$2.59.10^{-4}$	1.01	$5.31.10^{-4}$	1.00
640	$7.87.10^{-5}$	1.06	$1.55.10^{-4}$	1.03	$1.26.10^{-4}$	1.06	$2.61.10^{-4}$	1.02
1280	$3.57.10^{-5}$	1.14	$7.38.10^{-5}$	1.07	$5.65.10^{-5}$	1.14	$1.25.10^{-4}$	1.06

TABLE 1. **Proof of concept.** Experimental spatial order of convergence in L^1 and L^∞.

Both schemes are first order accurate, but we observe that our ϕ-entropic scheme (16), (18)-(20) is performing better since the numerical error is smaller than the classical upwind scheme.

Furthermore in Figure 1, we observe that for large time the numerical error corresponding to the entropy preserving scheme (16),(18)-(20) decays to zero and the error becomes negligible compared to that of the upwind scheme.

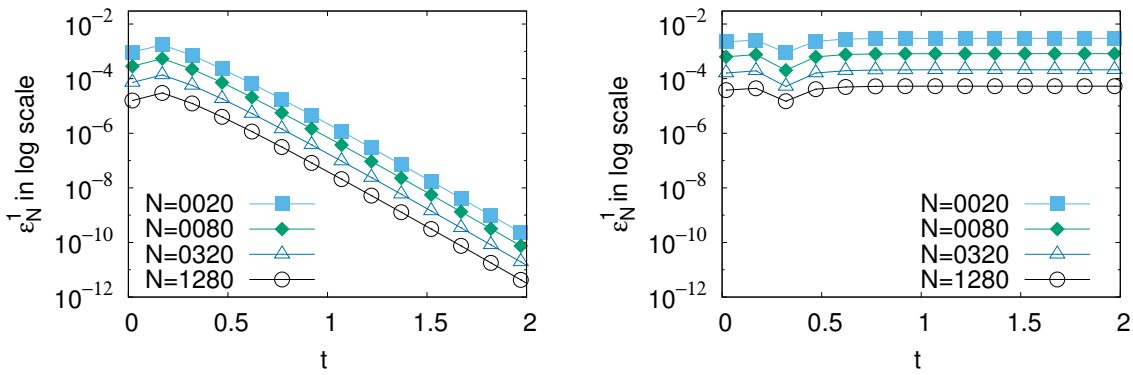


FIGURE 1. **Proof of concept.** Time evolution of the ϵ_N^1 error for (a) the ϕ-entropic scheme and (b) the classical upwind scheme.

The accurate long-time behavior is confirmed by Figure 2 where the time variation of the distance to the discrete solution is represented. While the upwind scheme saturates quickly, our scheme reproduces perfectly the exponential decay to 0 of the solution, even if $N = 40$, illustrating the result of Theorem 4.7.

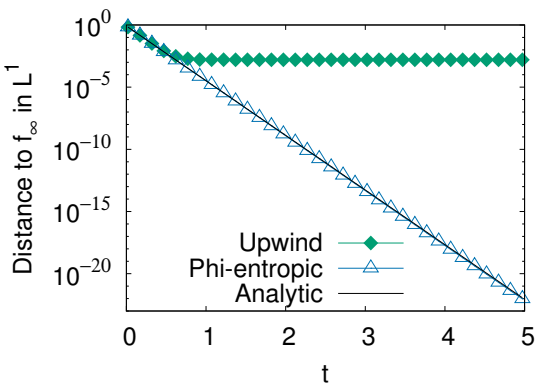


FIGURE 2. **Proof of concept.** Time evolution of the $L^1(\Omega)$ distance to the discrete steady state for $N = 40$.

6.3. **Fokker-Planck with magnetic field.** We now consider the two-dimensional version of homogeneous Fokker-Planck equation with an external magnetic field

$$\begin{cases} \dfrac{\partial f}{\partial t} + b\,\mathbf{v}^{\perp} \cdot \nabla_{\mathbf{v}} f = \nabla_{\mathbf{v}} \cdot (\mathbf{v} f + \nabla_{\mathbf{v}} f) \quad \text{in} \quad \mathbb{R}^+ \times \mathbb{R}^2\,, \\ f(t=0) = f_0 \quad \text{in} \quad \mathbb{R}^2\,. \end{cases}$$

The external magnetic field is along a third direction that is orthogonal to the plane under consideration and has amplitude $b = 4$. Compared to the 3D case, the vector $\mathbf{v}^\perp = (v_y, -v_x)$ replaces the cross product between $\mathbf{v}$ and the direction of the magnetic field. The initial datum f_0 is given by the sum of two Gaussian distributions

$$f_0(\mathbf{v}) = \frac{1}{2\pi}\left[\alpha \exp\left(-\frac{|\mathbf{v}-\mathbf{v}_1|^2}{2}\right) + (1-\alpha)\exp\left(-\frac{|\mathbf{v}-\mathbf{v}_2|^2}{2}\right)\right],$$

with $\alpha = 3/4$, $\mathbf{v}_1 = (-1, 2)$ and $\mathbf{v}_2 = (2, -1)$.

This equation is solved numerically in a bounded domain $\Omega = (-8, 8)^2$ on various regular Cartesian meshes from $N = 40^2$ to $N = 640^2$ points. We use our implicit scheme (16),(18)-(20) with a time step $\Delta t = 0.001$ until the final time $T = 10$. We choose non homogeneous Dirichlet boundary conditions $f^b = f^\infty$, where f^∞ is the steady state, that is, the Maxwellian distribution

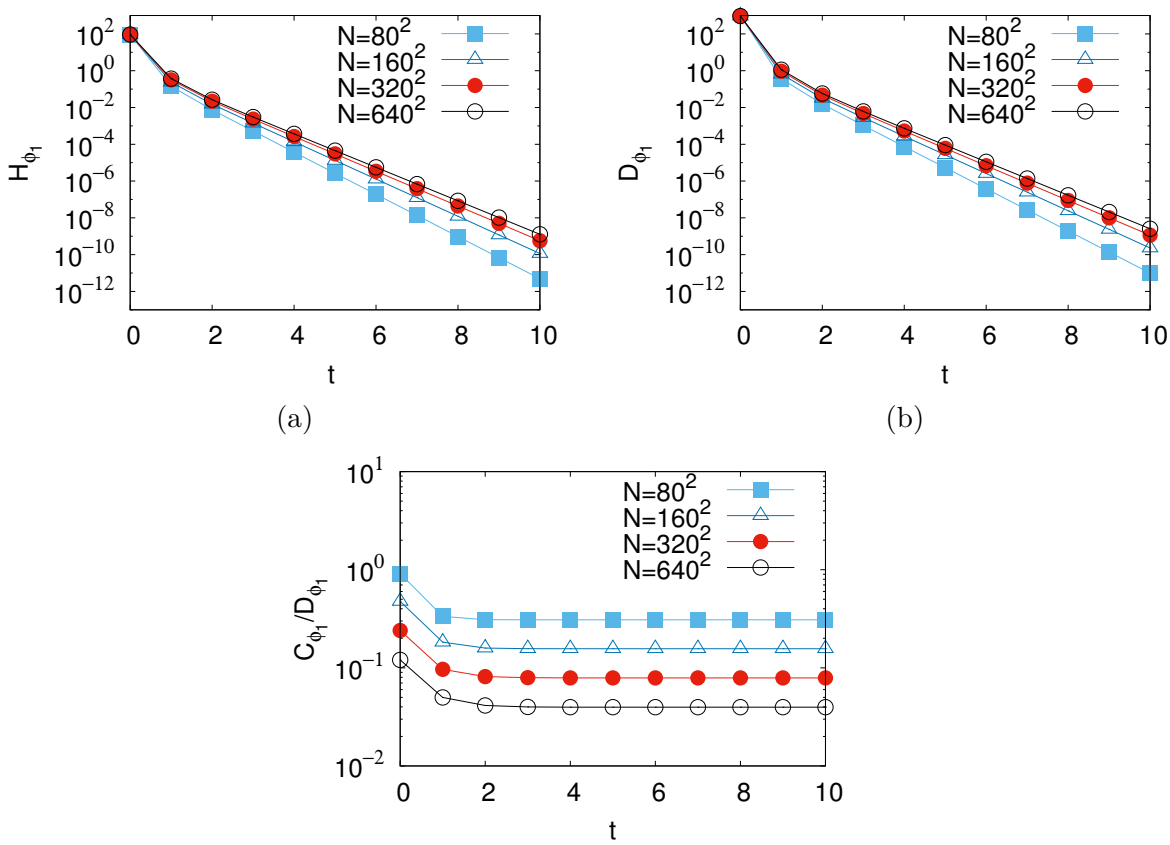
$$f^\infty(\mathbf{v}) = \frac{1}{2\pi}\exp\left(-\frac{|\mathbf{v}|^2}{2}\right).$$

Here the knowledge of the steady state f^∞ allows us to compute the steady flux $F^\infty_{K,\sigma}$ analytically. Indeed, since $\mathbf{v}f^\infty + \nabla_\mathbf{v} f^\infty = 0$, one only needs to evaluate

$$F^\infty_{K,\sigma} = b\int_\sigma \mathbf{n}_{K,\sigma}\cdot\mathbf{v}^\perp f^\infty = b\int_\sigma \mathbf{n}^\perp_{K,\sigma}\cdot\nabla_\mathbf{v} f^\infty,$$

which is, up to a multiplicative constant, the difference between f^∞ evaluated at each endpoint of σ.

We recall that the physical entropy generating function is $\phi_1 : s \mapsto s\log(s) - s + 1$. In Figure 3, we represent the time evolution of the entropy H_{ϕ_1}, the physical dissipation D_{ϕ_1} and the ratio between the numerical and physical dissipation C_{ϕ_1}/D_{ϕ_1} in log scale. On one hand, H_{ϕ_1} and D_{ϕ_1} decay exponentially and are well approximated when N is larger than 160^2. Figure 3 (a) and (b) also illustrate the convergence to equilibrium at exponential rate, when time goes to infinity. On the other hand, the numerical dissipation C_{ϕ_1} converges to zero when the space step goes to zero, but since the scheme is only first order accurate, it is relatively slow. Besides, the numerical dissipation C_{ϕ_1} also converges to zero at the same exponential rate than D_{ϕ_1}, hence it does not affect the accuracy on the decay rate for large time. Also note that for the chosen meshes the numerical dissipation is smaller than the physical dissipation. From these numerical experiments, we get some numerical evidence of the uniform accuracy of the scheme with respect to time.



(a) (b) (c)

FIGURE 3. **Fokker-Planck equation with magnetic field.** Time evolution of (a) the entropy H_{ϕ_1} (b) the physical dissipation D_{ϕ_1} and (c) the (normalized) numerical dissipation C_{ϕ_1}/H_{ϕ_1}.

Remark 6.1. *For the same mesh if we take a larger magnetic field, the numerical dissipation can become larger than the physical one. Even if both of them seem to converge to zero with the same decay rate, the dissipation is amplified.*

Finally, in Figures 4, we propose the time evolution of the distribution function at different times. The black lines represent the isovalues $f(t, \mathbf{v}) \equiv 1.10^{-5}$, 3.10^{-4}, 5.10^{-3}, $2.5.10^{-3}$, 7.10^{-2}, $1.2.10^{-1}$, $1.5.10^{-1}$ of the distribution function. We observe the effect of the magnetic field by the rotation of the two bumps and under the effect of the Fokker-Planck operator, the solution converges to a Maxwellian distribution.

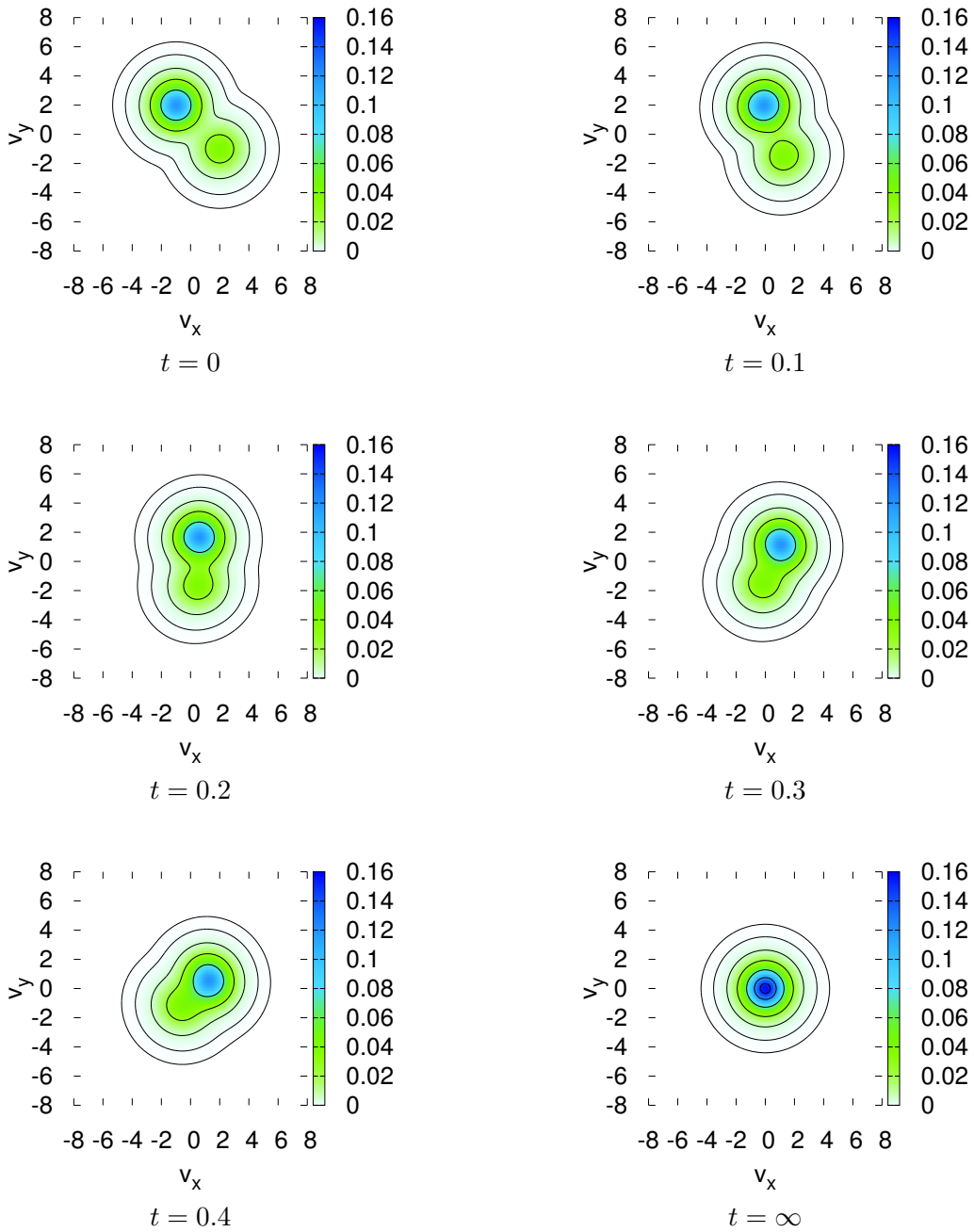


FIGURE 4. **Fokker-Planck equation with magnetic field.** Time evolution of the distribution on the fine mesh $N = 160^2$.

6.4. **Polymer flow in a dilute solution.** We investigate the numerical approximation of the Fokker-Planck part of the kinetic Fokker-Planck equation for polymers [28]

$$\begin{cases} \dfrac{\partial F}{\partial t} = -\nabla_{\mathbf{k}} \cdot \left[\left(\mathbf{A}\,\mathbf{k} - \dfrac{1}{2}\nabla_{\mathbf{k}}\Pi(\mathbf{k}) \right) F - \dfrac{1}{2}\nabla_{\mathbf{k}} F \right]. \\ \\ F(t=0) = F_0 \quad \text{in} \quad \Omega \subset \mathbb{R}^3, \end{cases}$$

where the matrix $\mathbf{A}$ represents the gradient of an external velocity field and is given by

$$\mathbf{A} = \begin{pmatrix} 1/4 & -1/2 & 0 \\ 1/2 & -1/4 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The domain is $\Omega = (-4,4)^3$ and we choose the Hookean model $\Pi(\mathbf{k}) = |\mathbf{k}|^2/2$. The initial datum F_0 is given by the sum of two Gaussian distributions

$$F_0(\mathbf{k}) = \frac{1}{2\,(2\pi)^{3/2}} \left[\exp\left(-\frac{|\mathbf{k}-\mathbf{k}_1|^2}{2} \right) + \exp\left(-\frac{|\mathbf{k}-\mathbf{k}_2|^2}{2} \right) \right],$$

with $\mathbf{k}_1 = (-3/2, 1, 0)$ and $\mathbf{k}_2 = (1, -3/2, 0)$. This equation is supplemented with homogeneous Neumann boundary conditions such that global mass is conserved. For numerical simulations we choose various meshes from $N = 24^3$ to 64^3 points with $\Delta t = 0.01$ using a time implicit scheme until $T = 5$. In this case, the steady state is not known, hence the steady equation is first solved numerically to compute a consistent approximation of the equilibrium $(f_K^\infty)_{K\in\mathcal{T}}$ and the stationary flux $F_{K,\sigma}^\infty$. If the matrix $\mathbf{A}$ were equal to zero we would expect the steady state to be close to a Gaussian $G(\mathbf{k}) = \exp(-\Pi(\mathbf{k}))$, at least far from edges since G do not satisfy the boundary conditions. Thus we can expect f^∞/G to be close to some constant in the domain, which seems easier to approximate numerically. Hence for all $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_{\text{int}}$, we define the quantities G_K and G_σ as the evaluation of G at the center of the respective cell or edge, as well as

$$h_K^\infty = \frac{f_K^\infty}{G_K}.$$

The scheme is solved on this new unknown h^∞. More precisely, it is given by (H3) with the flux

$$F_{K,\sigma}^\infty = m(\sigma)\, G_\sigma \left[A_{K,\sigma}^+ h_K^\infty - A_{K,\sigma}^- h_L^\infty - \frac{1}{2\,d_\sigma}\,(h_L^\infty - h_K^\infty) \right],$$

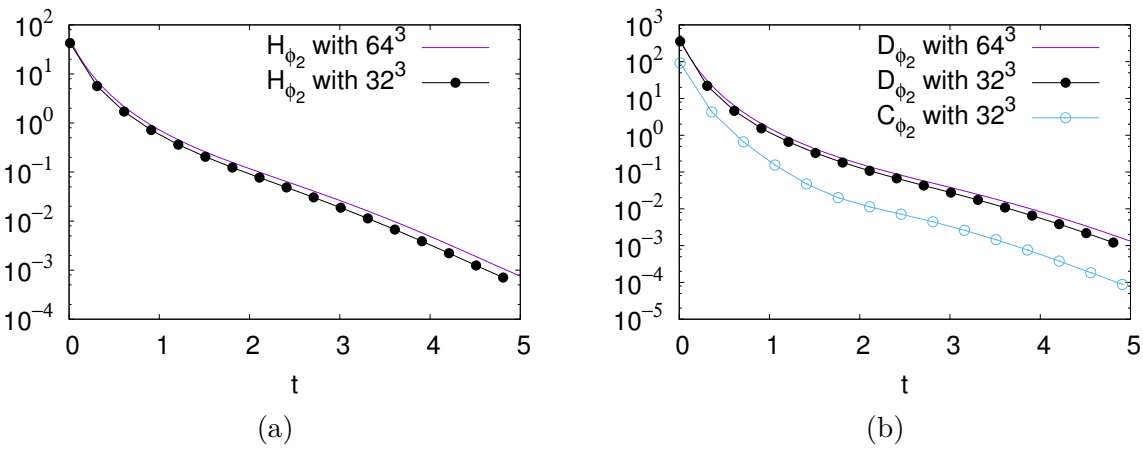
if $\sigma = K|L$ and $F_{K,\sigma}^\infty = 0$ if $\sigma \in \mathcal{E}_{\text{ext}}$. The quantity $A_{K,\sigma}$ is the evaluation of $(\mathbf{Ak}) \cdot \mathbf{n}_{K,\sigma}$ at the center of the edge σ. Besides, because of the conservative boundary conditions, we need to specify the mass of the steady state to be that of the initial data in order to get a unique solution to the scheme. Then we define the steady state on interior edges by $f_\sigma^\infty = (f_K^\infty + f_L^\infty)/2$ if $\sigma = K|L$. Observe that this scheme is consistent with the following equation

$$\nabla \cdot \left[G(\mathbf{k})((\mathbf{Ak})\, h^\infty - \frac{1}{2}\nabla h^\infty) \right] = 0,$$

which is only a reformulation of the original steady equation.

In Figure 5, we represent the time evolution of the entropy H_{ϕ_2}, its dissipation D_{ϕ_2} and the numerical dissipation due to the convective term C_{ϕ_2} in log scale. First when $N \geq 32^3$ points, the entropy and the physical dissipation are well approximated compared to the solution computed with a fine mesh $N \geq 64^3$. Once again both of them are decreasing function of time and converge to zero with an exponential decay rate. The numerical dissipation of the convective term C_{ϕ_2} is much smaller than the physical one and also converges to zero when

times goes to infinity exponentially fast, hence it does not affect the accuracy on the decay rate for large time.



(a) (b)

FIGURE 5. **Polymer flow in a dilute solution.** Time evolution of the 2-entropy H_{ϕ_2} and the corresponding physical dissipation and numerical dissipation (D_{ϕ_2}, C_{ϕ_2}) with $N = 32^3$ mesh points.

Finally, in Figure 6, we set forth the time evolution of the distribution function at different time. The first column represents an isovalue $f(t, \mathbf{k}) \equiv 0.02$ of the distribution function whereas the second column is a two dimensional projection in the plane $k_x - k_y$, the solution converges to the discrete steady state, which is consistent with the equilibrium.

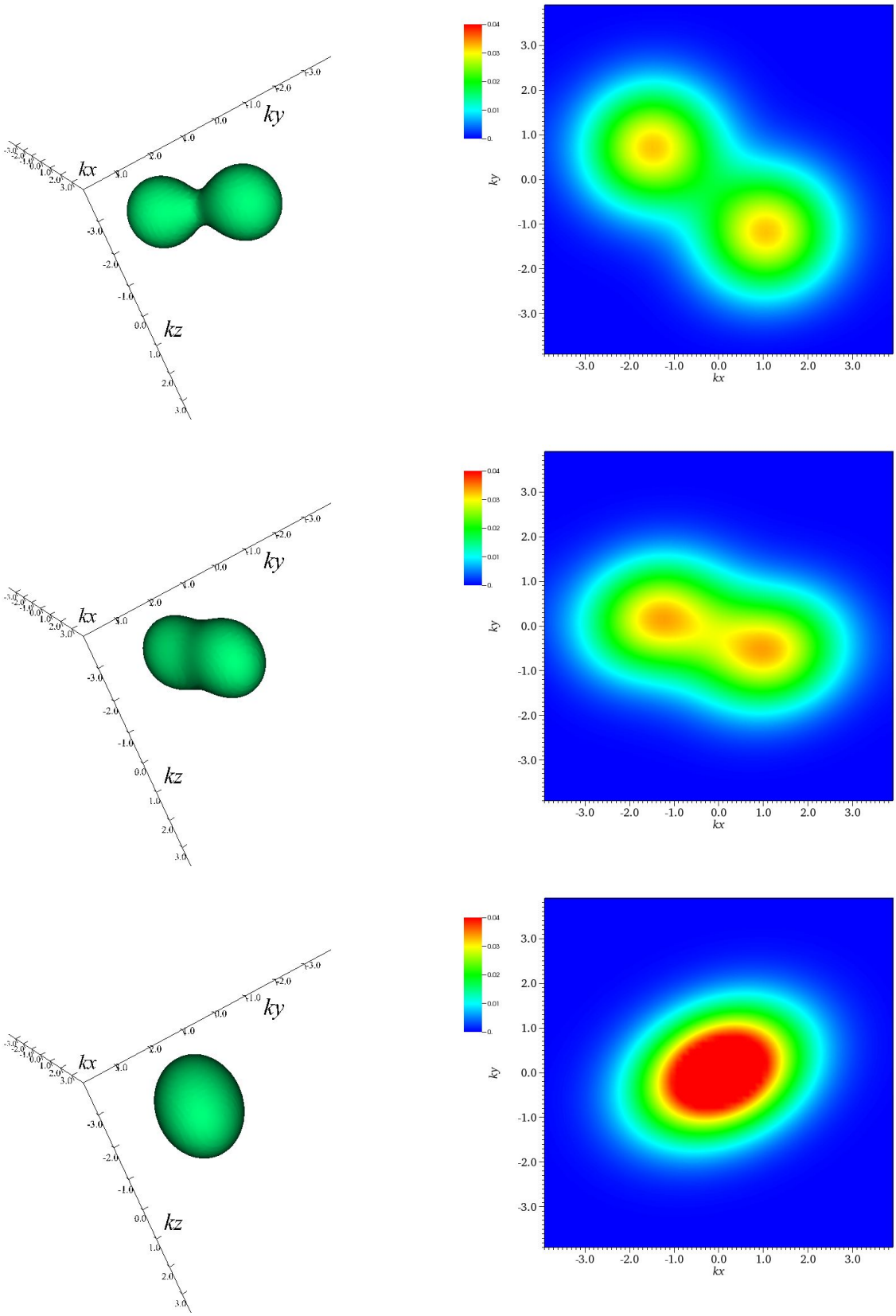


FIGURE 6. **Polymer flow in a dilute solution.** (a) one isovalue $F(t, \mathbf{k}) = 0.02$ (b) $k_x - k_y$ projection of the distribution in the $\mathbf{k}$ space at time $t = 0.2$, $t = 0.7$ and $t = 5$.

6.5. **Porous medium equation.** We finally study the numerical approximation of the porous medium equation

$$\begin{cases} \dfrac{\partial f}{\partial t} = \Delta f^m \,, \\ \\ f(t=0) = f_0 \quad \text{in}\, \Omega = (0,1) \times (-1,1)^2, \end{cases}$$

with $m = 2$ and $f_0 \equiv 0$ together with the non-homogeneous Dirichlet boundary conditions

$$f^b = \begin{cases} 2.5\,, & \text{if } x = 1 \text{ and } y^2 + z^2 \leq 1/8\,, \\ \\ 1\,, & \text{else.} \end{cases}$$

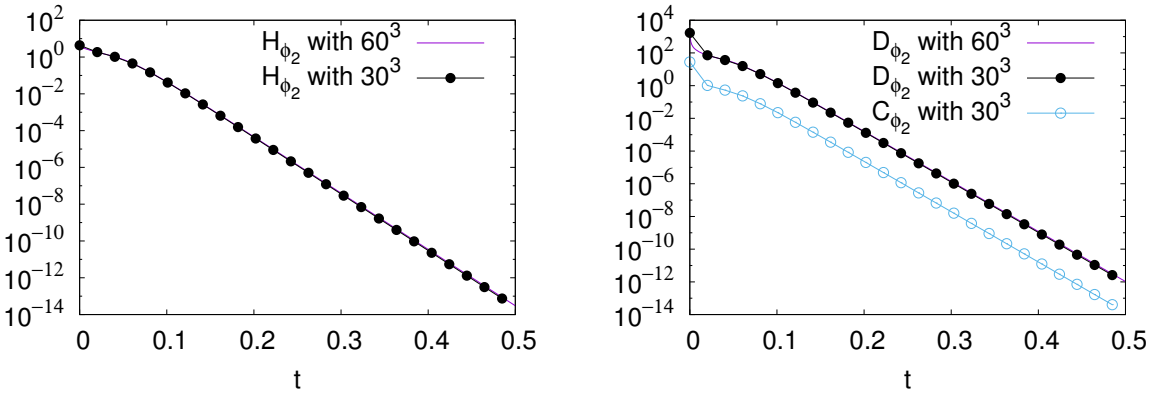
This model is nonlinear, with $\eta(s) = s^m$, and without convective terms. Moreover, since $\eta'(0) = 0$ the equation is degenerate. As a consequence, with our choice of initial data and boundary conditions, we expect the solution to be equal to zero on some subset of $\omega \subset \Omega$, having non-zero measure, for some positive time.

Since the steady state is not known, we first compute a numerical approximation $(f_K^\infty)_{K\in\mathcal{T}}$ and the corresponding stationary flux $F_{K,\sigma}^\infty$ using the scheme given in Example 2.3, which amounts to solving (H3) with

$$F_{K,\sigma}^\infty \;=\; -\tau_\sigma\, D_{K,\sigma}\eta(f^\infty)\,.$$

Then from the knowledge of $\eta(f^\infty) \in X_{\eta(f^b)}^{N_T+1}$, we set $f_K^\infty = \eta^{-1}(\eta(f^\infty)_K)$ on each cell $K \in \mathcal{T}$ and $\eta(f^\infty)_\sigma = (\eta(f^\infty)_K + \eta(f^\infty)_L)/2$ for interior edges $\sigma = K|L$.

For the numerical simulations we choose two Cartesian meshes with $N = 30^3$ and 60^3 points using the explicit scheme (17)-(20). Hence the time step now satisfies a CFL condition $\Delta t = O(\Delta x^2)$ (precisely defined in Theorem 4.4) which is satisfied for the two meshes with $\Delta t = 10^{-5}$. In Figure 7, we represent the time evolution of the relative entropy and the physical and numerical dissipation until final time $T = 0.5$. These results are in good agreement with those obtained using a finer mesh and the numerical dissipation is several orders of magnitude smaller than the physical dissipation.



(a) (b)

FIGURE 7. **Porous medium equation.** Time evolution of the 2-entropy H_{ϕ_2} and the corresponding physical dissipation and numerical dissipation (D_{ϕ_2}, C_{ϕ_2}) with $N = 30^3$ mesh points.

Finally in Figure 8 we represent the intersection of the graph of the discrete solution with the plane $z = 0$ at different time. The black line represents the isovalue 10^{-16} that surrounds the zone where the diffusion has not yet happened, namely where the distribution is null. This illustrates the good behavior of our scheme with respect to the degeneracy of the porous medium equation.

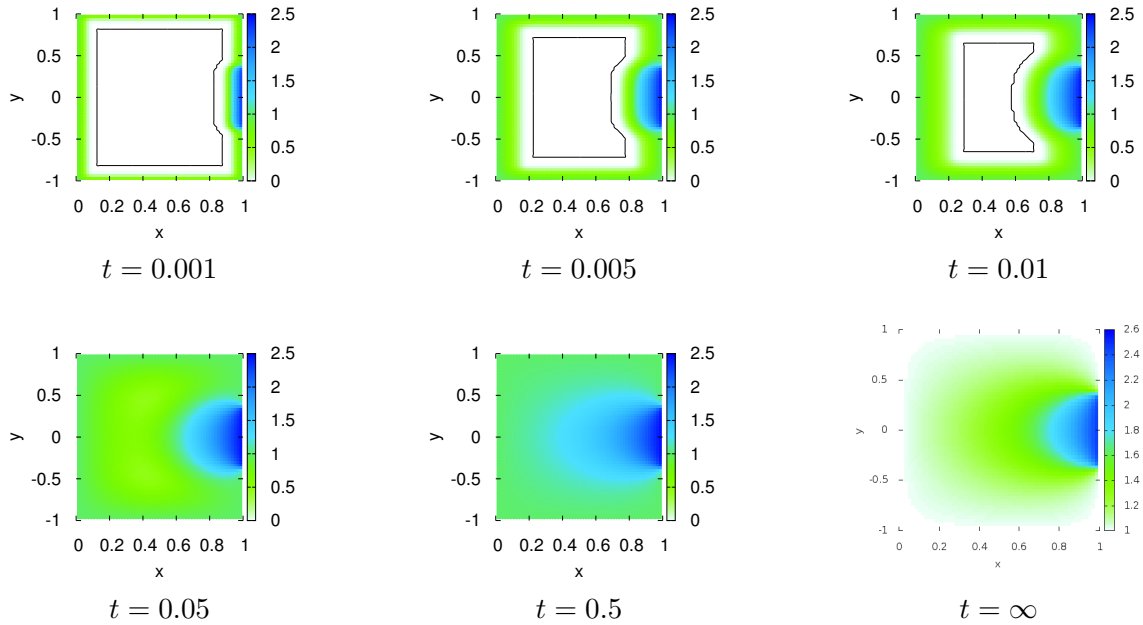


FIGURE 8. **Porous medium equation.** Time evolution of the distribution at $z = 0$ for $N = 60^3$. The black line is the isovalue $f(t, x, y, 0) \equiv 10^{-16}$.

7. Comments and conclusion

In this paper, we have built a scheme for boundary-driven convection-diffusion equations that preserves the relative ϕ-entropy structure of the model. After proving well-posedness, stability, exponential return to equilibrium and convergence, we gave several test cases that confirm the satisfying long time behavior of the numerical scheme in different settings (non-homogeneous Dirichlet/generalized Neumann boundary conditions, explicit and implicit time discretizations, linear and nonlinear models).

There are several directions that may be investigated for future work. The first objective is to generalize this scheme to anisotropic diffusions on possibly non-orthogonal meshes. It requires an adapted discretization of the gradient operator in every direction. There are several papers [21, 18, 19, 22, 11] of Eymard, Herbin, Gallouët, and Guichard and Cancès that are dealing with this type of problem. Their techniques are based on hybrid finite volume schemes for which the discrete gradient relies on the use of auxiliary unknowns located on edges between control volumes. However it is still unclear for the authors whether one can get the equivalent of the monotony properties of the 2-point flux used here, and that allows us to get the whole class of ϕ-entropy inequalities.

The spirit of our scheme is to start from a consistent discretization of the steady state and build the transient scheme upon the latter to ensure a satisfying behavior in the long-time

asymptotic. We would like to adapt this kind of strategy to other types of numerical scheme (Discontinuous Galerkin, Finite elements, *etc.*) and to different models.

Finally, as we saw in the introduction, many kinetic models (depending on space and velocity variables) such the Vlasov-Fokker-Planck equation or the full dumbbell model for polymers write as the sum of a transport in space and a (convection)-diffusion in velocity. The second part of these models is treated in the present paper. While the diffusion operator is not coercive in all the variables, thanks to the phase space mixing properties of the transport operator, this still leads to an entropy-diminishing behavior and a trend to a global equilibrium. This property is called hypocoercivity [32] and its preservation by numerical schemes has never been studied to our knowledge and this would be another interesting extension of this work.

Acknowledgements. The second author would like to thank Thierry Dumont for his kind help on sparse matrix routines. Both authors would like to acknowledge the anonymous reviewer for many constructive remarks that helped improving the quality of this paper.

References

[1] Franz Achleitner, Anton Arnold, and Dominik Stürzer. Large-time behavior in non-symmetric fokker-planck equations. 2015.

[2] Anton Arnold, Eric Carlen, and Qiangchang Ju. Large-time behavior of non-symmetric Fokker-Planck type equations. *Commun. Stoch. Anal.*, 2(1):153–175, 2008.

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

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

[5] Marianne Bessemoulin-Chatard, Claire Chainais-Hillairet, and Francis Filbet. On discrete functional inequalities for some finite volume schemes. *IMA J. Numer. Anal.*, 35(3):1125–1149, 2015.

[6] Marianne Bessemoulin-Chatard and Francis Filbet. A finite volume scheme for nonlinear degenerate parabolic equations. *SIAM J. Sci. Comput.*, 34(5):B559–B583, 2012.

[7] Thierry Bodineau, Joel Lebowitz, Clément Mouhot, and Cédric Villani. Lyapunov functionals for boundary-driven nonlinear drift-diffusion equations. *Nonlinearity*, 27(9):2111–2132, 2014.

[8] F. Bouchut and J. Dolbeault. On long time asymptotics of the Vlasov-Fokker-Planck equation and of the Vlasov-Poisson-Fokker-Planck system with Coulombic and Newtonian potentials. *Differential Integral Equations*, 8(3):487–514, 1995.

[9] Martin Burger, José A. Carrillo, and Marie-Therese Wolfram. A mixed finite element method for nonlinear diffusion equations. *Kinet. Relat. Models*, 3(1):59–83, 2010.

[10] Clément Cancès and Cindy Guichard. Numerical analysis of a robust entropy-diminishing Finite Volume scheme for parabolic equations with gradient structure. working paper or preprint, 2015.

[11] Clément Cancès and Cindy Guichard. Convergence of a nonlinear entropy diminishing control volume finite element scheme for solving anisotropic degenerate parabolic equations. *Math. Comp.*, 85(298):549–580, 2016.

[12] Claire Chainais-Hillairet and Francis Filbet. Asymptotic behaviour of a finite-volume scheme for the transient drift-diffusion model. *IMA J. Numer. Anal.*, 27(4):689–716, 2007.

[13] Claire Chainais-Hillairet, Ansgar Jüngel, and Stefan Schuchnigg. Entropy-dissipative discretization of nonlinear diffusion equations and discrete Beckner inequalities. *ESAIM Math. Model. Numer. Anal.*, 50(1):135–162, 2016.

[14] Claire Chainais-Hillairet, Jian-Guo Liu, and Yue-Jun Peng. Finite volume scheme for multi-dimensional drift-diffusion equations and convergence analysis. *M2AN Math. Model. Numer. Anal.*, 37(2):319–338, 2003.

[15] Claire Chainais-Hillairet and Yue-Jun Peng. Finite volume approximation for degenerate drift-diffusion system in several space dimensions. *Math. Models Methods Appl. Sci.*, 14(3):461–481, 2004.

[16] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[17] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[18] R. Eymard, T. Gallouët, and R. Herbin. A cell-centered finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension. *IMA J. Numer. Anal.*, 26(2):326–353, 2006.
[19] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4):1009–1043, 2010.
[20] Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.
[21] Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. A finite volume scheme for anisotropic diffusion problems. *C. R. Math. Acad. Sci. Paris*, 339(4):299–302, 2004.
[22] Robert Eymard, Cindy Guichard, and Raphaèle Herbin. Small-stencil 3D schemes for diffusive flows in porous media. *ESAIM Math. Model. Numer. Anal.*, 46(2):265–290, 2012.
[23] Francis Filbet and Chi-Wang Shu. Approximation of hyperbolic models for chemosensitive movement. *SIAM J. Sci. Comput.*, 27(3):850–872 (electronic), 2005.
[24] Maxime Herda. On massless electron limit for a multispecies kinetic system with external magnetic field. *J. Differential Equations*, 260(11):7861–7891, 2016.
[25] Benjamin Jourdain, Claude Le Bris, Tony Lelièvre, and Félix Otto. Long-time asymptotics of a multiscale model for polymeric fluid flows. *Arch. Ration. Mech. Anal.*, 181(1):97–148, 2006.
[26] Solomon Kullback. *Information theory and statistics*. John Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London, 1959.
[27] Xiaoye S. Li. An overview of SuperLU: algorithms, implementation, and user interface. *ACM Trans. Math. Software*, 31(3):302–325, 2005.
[28] Nader Masmoudi. Well-posedness for the FENE dumbbell model of polymeric flows. *Comm. Pure Appl. Math.*, 61(12):1685–1714, 2008.
[29] Mark A Peletier. Variational modelling: Energies, gradient flows, and large deviations. *arXiv preprint arXiv:1402.1990*, 2014.
[30] Andreas Unterreiter, Anton Arnold, Peter Markowich, and Giuseppe Toscani. On generalized Csiszár-Kullback inequalities. *Monatsh. Math.*, 131(3):235–253, 2000.
[31] Juan Luis Vázquez. *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, 2007. Mathematical theory.
[32] Cédric Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

Francis FILBET

Institut de Mathématiques de Toulouse,
Université Toulouse III & Institut Universitaire de France,
Bâtiment 1R3, 118, route de Narbonne
F-31062, Toulouse cedex 9 France

E-mail: francis.filbet@math.univ-toulouse.fr

Maxime HERDA

Institut Camille Jordan,
Université Claude Bernard Lyon 1, CNRS UMR 5208,
43 blvd. du 11 novembre 1918
F-69622, Villeurbanne cedex France

E-mail: herda@math.univ-lyon1.fr