

# NUMERICAL ANALYSIS OF A NONLINEAR FREE-ENERGY DIMINISHING DISCRETE DUALITY FINITE VOLUME SCHEME FOR CONVECTION DIFFUSION EQUATIONS

CLÉMENT CANCÈS, CLAIRE CHAINAIS-HILLAIRET, AND STELLA KRELL

**ABSTRACT.** We propose a nonlinear Discrete Duality Finite Volume scheme to approximate the solutions of drift diffusion equations. The scheme is built to preserve at the discrete level even on severely distorted meshes the energy / energy dissipation relation. This relation is of paramount importance to capture the long-time behavior of the problem in an accurate way. To enforce it, the linear convection diffusion equation is rewritten in a nonlinear form before being discretized. We establish the existence of positive solutions to the scheme. Based on compactness arguments, the convergence of the approximate solution towards a weak solution is established. Finally, we provide numerical evidences of the good behavior of the scheme when the discretization parameters tend to 0 and when time goes to infinity.

**Keywords.** Convection diffusion equation, Discrete Duality Finite Volumes, convergence analysis, discrete entropy method

**AMS subjects classification.** 65M08, 65M12, 35K20

## 1. INTRODUCTION

**1.1. Motivation.** The modeling of systems of interacting particles, like electrons in electronic devices, ions in plasmas, chemical species in biological membranes for instance, leads to systems of evolutive partial differential equations. The knowledge of the large time behavior of such systems is crucial for the understanding of the underlying physical phenomena. In many cases, the relaxation to an equilibrium configuration is based on the second law of thermodynamics and on the dissipation of some entropies.

Based on works in kinetic theory, mathematicians have intensively developed the entropy method for the study of the large time behavior of different systems of PDEs. Let us mention works on Boltzmann and Landau equations [43], on linear Fokker-Planck equations [15], on porous media equations [16], on reaction-diffusion systems [23, 24, 34], on drift-diffusion systems for semiconductor devices [31, 32, 30]. We also refer to the survey paper [5] and to the reference book [37]. Similar results were obtained based on the interpretation of PDE models as Wasserstein gradient flows [1]. We refer for instance to [36, 10] for linear Fokker Planck equations, to [40] for the porous medium equation, to [11] for granular media. This list is far from being exhaustive.

The knowledge of the large time behavior of such evolution equations, the existence of some entropies which are dissipated along time are structural features, as positivity of densities or conservation of mass, that should be preserved at the discrete level by numerical schemes. The question of the large time behavior of numerical schemes has been investigated for instance for coagulation-fragmentation models [29], for nonlinear diffusion equations [17, 38], for reaction-diffusion systems

---

The authors are supported by the Inria teams RAPSODI and COFFEE, the LabEx CEMPI (ANR-11-LABX-0007-01), the GEOPOR project (ANR-13-JS01-0007-01) and the MOONRISE project (ANR-14-CE23-0007).

[34, 35], for drift-diffusion systems [19, 7]. These last works show that the Scharfetter-Gummel numerical fluxes, first introduced in [42] for the approximation of convection-diffusion fluxes and widely used later for the simulation of semiconductor devices, preserve the thermal equilibrium. Their use in the numerical approximation of drift-diffusion systems ensure the exponential decay towards equilibrium of the numerical scheme. Unfortunately, such numerical fluxes can only be applied in two-points flux approximation finite volume schemes and therefore on restricted meshes. Moreover, they do not extend to anisotropic convection-diffusion equations.

Therefore, it seems crucial to propose new finite volume schemes which preserve the large-time behavior of anisotropic convection-diffusion equations and which apply on almost general meshes. In [14], the authors proposed and analyzed a VAG scheme satisfying these prescribed properties. In this work, we propose and study the convergence analysis of a nonlinear free-energy diminishing discrete duality finite volume scheme [12].

**1.2. Presentation of the continuous problem.** We focus on a very basic drift-diffusion equation with potential convection and anisotropy. Let  $\Omega$  be a polygonal connected open bounded subset of  $\mathbb{R}^2$  and let  $T > 0$  be a finite time horizon. The problem writes:

$$\begin{aligned} (1a) \quad & \partial_t u + \operatorname{div} \mathbf{J} = 0, \quad \text{in } Q_T = \Omega \times (0, T), \\ (1b) \quad & \mathbf{J} = -\mathbf{\Lambda} \nabla u - u \mathbf{\Lambda} \nabla V, \quad \text{in } Q_T, \\ (1c) \quad & \mathbf{J} \cdot \mathbf{n} = 0, \quad \text{on } \partial\Omega \times (0, T), \\ (1d) \quad & u(\cdot, 0) = u_0, \quad \text{in } \Omega, \end{aligned}$$

with  $\mathbf{n}$  the outward unit normal to  $\partial\Omega$  and the following assumptions on the data:

(A1) The initial data  $u_0$  is measurable, nonnegative and satisfies

$$(2) \quad \int_{\Omega} u_0 d\mathbf{x} > 0 \quad \text{and} \quad \int_{\Omega} H(u_0) d\mathbf{x} < \infty,$$

where  $H(s) = s \log s - s + 1$  for all  $s \geq 0$ .

(A2) The exterior potential  $V$  belongs to  $C^1(\overline{\Omega}, \mathbb{R})$ . Without loss of generality, we assume that  $V \geq 0$  in  $\overline{\Omega}$ .

(A3) The anisotropy tensor  $\mathbf{\Lambda}$  is supposed to be bounded (i.e.,  $\mathbf{\Lambda} \in L^\infty(\Omega)^{2 \times 2}$ ), symmetric (i.e.,  $\mathbf{\Lambda} = \mathbf{\Lambda}^T$  a.e. in  $\Omega$ ), and uniformly elliptic: there exist  $\lambda_m > 0$  and  $\lambda^M > 0$  such that

$$(3) \quad \lambda_m |\mathbf{v}|^2 \leq \mathbf{\Lambda}(\mathbf{x}) \mathbf{v} \cdot \mathbf{v} \leq \lambda^M |\mathbf{v}|^2, \quad \text{for all } \mathbf{v} \in \mathbb{R}^2 \text{ and almost all } \mathbf{x} \in \Omega.$$

The flux  $\mathbf{J}$  can be reformulated in the nonlinear form

$$\mathbf{J} = -u \mathbf{\Lambda} \nabla (\log u + V).$$

Testing equation (1a) by  $\log(u) + V$  leads to the so-called *energy/energy dissipation* relation (energy/dissipation for short)

$$(4) \quad \frac{d\mathbb{E}}{dt} + \mathbb{I} = 0,$$

where the free energy  $\mathbb{E}$  and the dissipation  $\mathbb{I}$  for (1) are respectively defined by

$$(5) \quad \mathbb{E}(t) = \int_{\Omega} (H(u) + Vu)(\mathbf{x}, t) d\mathbf{x},$$

$$(6) \quad \mathbb{I}(t) = \int_{\Omega} u \mathbf{\Lambda} \nabla (\log u + V) \cdot \nabla (\log u + V) d\mathbf{x}.$$

Since  $u$  is nonnegative, so does  $\mathbb{I}$  and the free energy  $\mathbb{E}$  is decaying with time. As highlighted for instance in [6] and [10], the solution  $u$  to (1) converges towards the steady-state

$$u_\infty = \left( \int_{\Omega} u_0 dx / \int_{\Omega} e^{-V} dx \right) e^{-V}$$

when time goes to infinity. In the case where  $\mathbf{\Lambda}$  does not depend on  $\mathbf{x}$  and where both  $\Omega$  and  $V$  are convex, this convergence is exponentially fast.

The energy/energy dissipation relation (4) provides a control on the Fisher information

$$(7) \quad \iint_{Q_T} u |\nabla \log(u)|^2 d\mathbf{x} dt = 4 \iint_{Q_T} |\nabla \sqrt{u}|^2 d\mathbf{x} dt \leq C.$$

Thus it is natural to seek the solution in the space

$$\left\{ u : Q_T \rightarrow \mathbb{R}_+ \mid \int_{\Omega} H(u(\mathbf{x}, \cdot)) d\mathbf{x} \in L^\infty(0, T) \text{ and } \sqrt{u} \in L^2(0, T; H^1(\Omega)) \right\}.$$

This motivates the following notion of weak solution.

**Definition 1.1.** A function  $u : Q_T \rightarrow \mathbb{R}_+$  is said to be a weak solution to the problem (1) if  $H(u) \in L^\infty(0, T; L^1(\Omega))$ ,  $\sqrt{u} \in L^2(0, T; H^1(\Omega))$ , and (1) is satisfied in the distributional sense, i.e., for all  $\varphi \in C_c^\infty(\bar{\Omega} \times [0, T))$ , there holds

$$(8) \quad \iint_{Q_T} u \partial_t \varphi d\mathbf{x} dt + \int_{\Omega} u_0 \varphi(\cdot, 0) d\mathbf{x} - \iint_{Q_T} (u \nabla V + \nabla u) \cdot \mathbf{\Lambda} \nabla \varphi d\mathbf{x} dt = 0.$$

**1.3. Outline of the paper.** In Section 2, we introduce the numerical scheme and state the main results of the paper: existence of a positive solution to the scheme and convergence of a sequence of approximate solutions towards a weak solution. The existence of a solution to the scheme is established in Section 3. It strongly relies on the conservation of mass at the discrete level and on a discrete counterpart of an energy/dissipation estimate. Section 4 is devoted to the proof of convergence of the scheme. The effective behavior of the numerical method is eventually discussed in Section 5. It is shown that the method is second order accurate w.r.t. space in  $L^2$  norm, whereas the approximate gradient super-converges with observed order 3/2. Moreover, the method exhibit a very accurate long-time behavior.

## 2. PRESENTATION OF THE SCHEME AND MAIN RESULTS

**2.1. Meshes and notations.** In order to define a DDFV scheme, as for instance in [25, 3], we need to introduce three different meshes – the primal mesh, the dual mesh and the diamond mesh – and some associated notations.

The primal mesh denoted  $\overline{\mathfrak{M}}$  is composed of the interior primal mesh  $\mathfrak{M}$  (a partition of  $\Omega$  with polygonal control volumes) and the set  $\partial\mathfrak{M}$  of boundary edges seen as degenerate control volumes. For all  $K \in \overline{\mathfrak{M}}$ , we define  $x_K$  the center of  $K$ . The family of centers is denoted by  $\mathfrak{X} = \{x_K, K \in \overline{\mathfrak{M}}\}$ .

Let  $\mathfrak{X}^*$  denote the set of the vertices of the primal control volumes in  $\overline{\mathfrak{M}}$ . Distinguishing the interior vertices from the vertices lying on the boundary, we split  $\mathfrak{X}^*$  into  $\mathfrak{X}^* = \mathfrak{X}_{int}^* \cup \mathfrak{X}_{ext}^*$ . To any point  $x_{K^*} \in \mathfrak{X}_{int}^*$ , we associate the polygon  $K^*$ , whose vertices are  $\{x_K \in \mathfrak{X} / x_{K^*} \in \bar{K}, K \in \mathfrak{M}\}$ . The set of these polygons defines the interior dual mesh denoted by  $\mathfrak{M}^*$ . To any point  $x_{K^*} \in \mathfrak{X}_{ext}^*$ , we then associate the polygon  $K^*$ , whose vertices are  $\{x_{K^*}\} \cup \{x_K \in \mathfrak{X} / x_{K^*} \in \bar{K}, K \in \mathfrak{M}\}$ . The set of these polygons is denoted by  $\partial\mathfrak{M}^*$  called the boundary dual mesh and the dual mesh is  $\mathfrak{M}^* \cup \partial\mathfrak{M}^*$ , denoted by  $\overline{\mathfrak{M}^*}$ .

For all neighboring primal cells  $K$  and  $L$ , we assume that  $\partial K \cap \partial L$  is a segment, corresponding to an edge of the mesh  $\mathfrak{M}$ , denoted by  $\sigma = K|L$ . Let  $\mathcal{E}$  be the set of such edges. We similarly define the set  $\mathcal{E}^*$  of the edges of the dual mesh. For each couple  $(\sigma, \sigma^*) \in \mathcal{E} \times \mathcal{E}^*$  such that  $\sigma = K|L = (x_{K^*}, x_{L^*})$  and  $\sigma^* = K^*|L^* = (x_K, x_L)$ , we define the quadrilateral diamond cell  $\mathcal{D}_{\sigma, \sigma^*}$  whose diagonals are  $\sigma$  and  $\sigma^*$ , as shown on Figure 1. If  $\sigma \in \mathcal{E} \cap \partial\Omega$ , we note that the diamond degenerates into a triangle. The set of the diamond cells defines the diamond mesh  $\mathfrak{D}$ . It is a partition of  $\Omega$ . We can rewrite  $\mathfrak{D} = \mathfrak{D}^{ext} \cup \mathfrak{D}^{int}$  where  $\mathfrak{D}^{ext}$  is the set of all the boundary diamonds and  $\mathfrak{D}^{int}$  the set of all the interior diamonds.

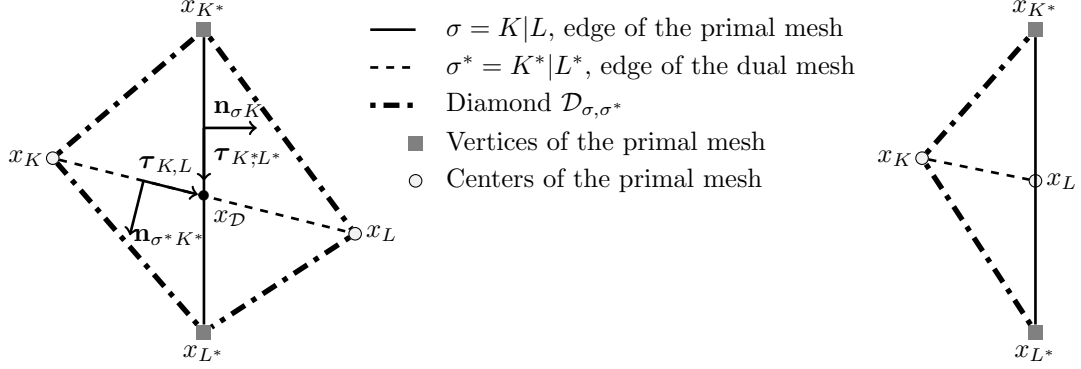


FIGURE 1. Definition of the diamonds  $\mathcal{D}_{\sigma, \sigma^*}$  and related notations.

Finally, the DDFV mesh is made of  $\mathcal{T} = (\overline{\mathfrak{M}}, \overline{\mathfrak{M}^*})$  and  $\mathfrak{D}$ . For each primal or dual cell  $M$  ( $M \in \mathfrak{M}$  or  $M \in \mathfrak{M}^*$ ), we define  $m_M$  the measure of  $M$ ,  $\mathcal{E}_M$  the set of the edges of  $M$  (it coincides with the edge  $\sigma = M$  if  $M \in \partial\mathfrak{M}$ ),  $\mathfrak{D}_M$  the set of diamonds  $\mathcal{D}_{\sigma, \sigma^*} \in \mathfrak{D}$  such that  $m(\mathcal{D}_{\sigma, \sigma^*} \cap M) > 0$ , and  $d_M$  the diameter of  $M$ .

For a diamond  $\mathcal{D}_{\sigma, \sigma^*}$ , whose vertices are  $(x_K, x_{K^*}, x_L, x_{L^*})$ , we define:  $x_D$  the center of the diamond cell  $\mathcal{D}$ :  $\{x_D\} = \sigma \cap \sigma^*$ ,  $m_\sigma$  the length of the primal edge  $\sigma$ ,  $m_{\sigma^*}$  the length of the dual edge  $\sigma^*$ ,  $d_D$  the diameter of  $\mathcal{D}$ ,  $\alpha_D$  the angle between  $(x_K, x_L)$  and  $(x_{K^*}, x_{L^*})$ . We will also use two direct basis  $(\tau_{K^*, L^*}, \mathbf{n}_{\sigma K})$  and  $(\mathbf{n}_{\sigma^* K^*}, \tau_{K, L})$ , where  $\mathbf{n}_{\sigma K}$  is the unit normal to  $\sigma$ , outward  $K$ ,  $\mathbf{n}_{\sigma^* K^*}$  is the unit normal to  $\sigma^*$ , outward  $K^*$ ,  $\tau_{K^*, L^*}$  is the unit tangent vector to  $\sigma$ , oriented from  $K^*$  to  $L^*$ ,  $\tau_{K, L}$  is the unit tangent vector to  $\sigma^*$ , oriented from  $K$  to  $L$ . Denoting by  $m_D$  the 2-dimensional Lebesgue measure of  $\mathcal{D}$ , one has

$$(9) \quad m_D = \frac{1}{2} m_\sigma m_{\sigma^*} \sin(\alpha_D), \quad \forall \mathcal{D} = \mathcal{D}_{\sigma, \sigma^*} \in \mathfrak{D}.$$

We define two local regularity factors  $\theta_D, \tilde{\theta}_D$  of the diamond cell  $\mathcal{D} = \mathcal{D}_{\sigma, \sigma^*} \in \mathfrak{D}$  by

$$(10) \quad \theta_D = \frac{1}{2 \sin(\alpha_D)} \left( \frac{m_\sigma}{m_{\sigma^*}} + \frac{m_{\sigma^*}}{m_\sigma} \right) \geq 1, \quad \tilde{\theta}_D = \max \left( \max_{K \in \mathfrak{M}_D} \frac{m_D}{m_{D \cap K}}; \max_{K^* \in \mathfrak{M}_D^*} \frac{m_D}{m_{D \cap K^*}} \right).$$

In what follows, we assume that there exists  $\theta^* \geq 1$  such that

$$(11) \quad 1 \leq \theta_D, \tilde{\theta}_D \leq \theta^*, \quad \forall \mathcal{D} \in \mathfrak{D}.$$

In particular, this implies that

$$(12) \quad \sin(\alpha_D) \geq \frac{1}{\theta^*}, \quad \forall \mathcal{D} \in \mathfrak{D}.$$

Moreover, owing to the definition of  $\tilde{\theta}_{\mathcal{D}}$  and to (11), one has

$$(13) \quad \sum_{\mathcal{D} \in \mathfrak{D}_K} m_{\mathcal{D}} \leq \theta^* m_K \quad \text{and} \quad \sum_{\mathcal{D} \in \mathfrak{D}_{K^*}} m_{\mathcal{D}} \leq \theta^* m_{K^*}$$

Finally, we define the size of the mesh:  $\text{size}(\mathcal{T}) = \max_{\mathcal{D} \in \mathfrak{D}} d_{\mathcal{D}}$ .

**2.2. Discrete unknowns and discrete operators.** We first define the different sets of discrete unknowns. As it is usual for DDFV methods, we need several types of degrees of freedom to represent scalar and vector fields in the discrete setting. We introduce  $\mathbb{R}^{\mathcal{T}}$  the linear space of scalar fields constant on the cells of  $\mathfrak{M}$  and  $\mathfrak{M}^*$ :

$$u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}} \iff u_{\mathcal{T}} = ((u_K)_{K \in \mathfrak{M}}, (u_{K^*})_{K^* \in \mathfrak{M}^*})$$

and  $(\mathbb{R}^2)^{\mathfrak{D}}$  the linear space of vector fields constant on the diamonds:

$$\xi_{\mathfrak{D}} \in (\mathbb{R}^2)^{\mathfrak{D}} \iff \xi_{\mathfrak{D}} = (\xi_{\mathcal{D}})_{\mathcal{D} \in \mathfrak{D}}.$$

Let us mention that we similarly denote by  $\mathbb{R}^{\mathfrak{D}}$  the set of scalar fields constant on the diamonds.

Then, we define the positive semi-definite bilinear form<sup>1</sup>  $\llbracket \cdot, \cdot \rrbracket_{\mathcal{T}}$  on  $\mathbb{R}^{\mathcal{T}}$  and the scalar product  $(\cdot, \cdot)_{\Lambda, \mathfrak{D}}$  on  $(\mathbb{R}^2)^{\mathfrak{D}}$  by

$$\begin{aligned} \llbracket v_{\mathcal{T}}, u_{\mathcal{T}} \rrbracket_{\mathcal{T}} &= \frac{1}{2} \left( \sum_{K \in \mathfrak{M}} m_K u_K v_K + \sum_{K^* \in \mathfrak{M}^*} m_{K^*} u_{K^*} v_{K^*} \right), \quad \forall u_{\mathcal{T}}, v_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}, \\ (\xi_{\mathfrak{D}}, \varphi_{\mathfrak{D}})_{\Lambda, \mathfrak{D}} &= \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} \xi_{\mathcal{D}} \cdot \Lambda^{\mathcal{D}} \varphi_{\mathcal{D}}, \quad \forall \xi_{\mathfrak{D}}, \varphi_{\mathfrak{D}} \in (\mathbb{R}^2)^{\mathfrak{D}}, \end{aligned}$$

where

$$\Lambda^{\mathcal{D}} = \frac{1}{m_{\mathcal{D}}} \int_{\mathcal{D}} \Lambda(x) dx, \quad \forall \mathcal{D} \in \mathfrak{D}.$$

We denote by  $\|\cdot\|_{\Lambda, \mathfrak{D}}$  the Euclidian norm associated to the scalar product  $(\cdot, \cdot)_{\Lambda, \mathfrak{D}}$ , i.e.,

$$\|\xi_{\mathfrak{D}}\|_{\Lambda, \mathfrak{D}}^2 = (\xi_{\mathfrak{D}}, \xi_{\mathfrak{D}})_{\Lambda, \mathfrak{D}}, \quad \forall \xi_{\mathfrak{D}} \in (\mathbb{R}^2)^{\mathfrak{D}}.$$

The DDFV method is based on the definitions of a discrete gradient and of a discrete divergence, which are linked by duality formula as shown in [25]. The discrete gradient has been introduced in [20] and developed in [25]. It is a mapping from  $\mathbb{R}^{\mathcal{T}}$  to  $(\mathbb{R}^2)^{\mathfrak{D}}$  defined by  $\nabla^{\mathfrak{D}} u_{\mathcal{T}} = (\nabla^{\mathcal{D}} u_{\mathcal{T}})_{\mathcal{D} \in \mathfrak{D}}$  for all  $u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ , where

$$\nabla^{\mathcal{D}} u_{\mathcal{T}} = \frac{1}{\sin(\alpha_{\mathcal{D}})} \left( \frac{u_L - u_K}{m_{\sigma^*}} \mathbf{n}_{\sigma K} + \frac{u_{L^*} - u_{K^*}}{m_{\sigma}} \mathbf{n}_{\sigma^* K^*} \right), \quad \forall \mathcal{D} \in \mathfrak{D}.$$

Using (9), the discrete gradient can be equivalently written:

$$\nabla^{\mathcal{D}} u_{\mathcal{T}} = \frac{1}{2m_{\mathcal{D}}} (m_{\sigma} (u_L - u_K) \mathbf{n}_{\sigma K} + m_{\sigma^*} (u_{L^*} - u_{K^*}) \mathbf{n}_{\sigma^* K^*}), \quad \forall \mathcal{D} \in \mathfrak{D}.$$

---

<sup>1</sup>Although it mimics the continuous  $L^2(\Omega)$  scalar product, the bilinear form  $\llbracket \cdot, \cdot \rrbracket_{\mathcal{T}}$  is not a scalar product since it does not involve the primal boundary edges  $\partial \mathfrak{M}$ . It is therefore not definite.

The discrete divergence has been introduced in [25]. It is a mapping  $\text{div}^\mathcal{T}$  from  $(\mathbb{R}^2)^\mathfrak{D}$  to  $\mathbb{R}^\mathcal{T}$  defined for all  $\xi_\mathfrak{D} \in (\mathbb{R}^2)^\mathfrak{D}$  by

$$\text{div}^\mathcal{T} \xi_\mathfrak{D} = \left( \text{div}^\mathfrak{M} \xi_\mathfrak{D}, \text{div}^{\partial\mathfrak{M}} \xi_\mathfrak{D}, \text{div}^{\mathfrak{M}^*} \xi_\mathfrak{D}, \text{div}^{\partial\mathfrak{M}^*} \xi_\mathfrak{D} \right),$$

with  $\text{div}^\mathfrak{M} \xi_\mathfrak{D} = (\text{div}_K \xi_\mathfrak{D})_{K \in \mathfrak{M}}$ ,  $\text{div}^{\partial\mathfrak{M}} \xi_\mathfrak{D} = 0$ ,  $\text{div}^{\mathfrak{M}^*} \xi_\mathfrak{D} = (\text{div}_{K^*} \xi_\mathfrak{D})_{K^* \in \mathfrak{M}^*}$  and  $\text{div}^{\partial\mathfrak{M}^*} \xi_\mathfrak{D} = (\text{div}_{K^*} \xi_\mathfrak{D})_{K^* \in \partial\mathfrak{M}^*}$  such that:

$$\forall K \in \mathfrak{M}, \text{div}_K \xi_\mathfrak{D} = \frac{1}{m_K} \sum_{\substack{\mathcal{D} \in \mathfrak{D}_K \\ \mathcal{D} = \mathcal{D}_{\sigma, \sigma^*}}} m_\sigma \xi_\mathcal{D} \cdot \mathbf{n}_{\sigma K},$$

and analogous definitions for  $\text{div}_{K^*} \xi_\mathfrak{D}$  for  $K^* \in \overline{\mathfrak{M}^*}$ .

In [2], the authors study the convergence of DDFV schemes for degenerate hyperbolic-parabolic problems. They show that a penalization operator is needed in order to establish the convergence proof. Indeed, this penalization operator ensures that the two components of a discrete function (reconstructions on the primal and dual meshes) converge to the same limit. For similar reasons (see Section 4), we consider the same penalization operator  $\mathcal{P}^\mathcal{T} : \mathbb{R}^\mathcal{T} \rightarrow \mathbb{R}^\mathcal{T}$  as in [2] and [18]. It is defined for all  $u_\mathcal{T} \in \mathbb{R}^\mathcal{T}$  by

$$\mathcal{P}^\mathcal{T} u_\mathcal{T} = \left( \mathcal{P}^\mathfrak{M} u_\mathcal{T}, \mathcal{P}^{\partial\mathfrak{M}} u_\mathcal{T}, \mathcal{P}^{\mathfrak{M}^*} u_\mathcal{T}, \mathcal{P}^{\partial\mathfrak{M}^*} u_\mathcal{T} \right),$$

with  $\mathcal{P}^\mathfrak{M} u_\mathcal{T} = (\mathcal{P}_K u_\mathcal{T})_{K \in \mathfrak{M}}$ ,  $\mathcal{P}^{\partial\mathfrak{M}} u_\mathcal{T} = 0$ ,  $\mathcal{P}^{\mathfrak{M}^*} u_\mathcal{T} = (\mathcal{P}_{K^*} u_\mathcal{T})_{K^* \in \mathfrak{M}^*}$  and  $\mathcal{P}^{\partial\mathfrak{M}^*} u_\mathcal{T} = (\mathcal{P}_{K^*} u_\mathcal{T})_{K^* \in \partial\mathfrak{M}^*}$  such that, for a given parameter  $\beta \in (0, 2)$ ,

$$\begin{aligned} \forall K \in \mathfrak{M}, \mathcal{P}_K u_\mathcal{T} &= \frac{1}{m_K} \frac{1}{\text{size}(\mathcal{T})^\beta} \sum_{K^* \in \overline{\mathfrak{M}^*}} m_{K \cap K^*} (u_K - u_{K^*}), \\ \forall K^* \in \overline{\mathfrak{M}^*}, \mathcal{P}_{K^*} u_\mathcal{T} &= \frac{1}{m_{K^*}} \frac{1}{\text{size}(\mathcal{T})^\beta} \sum_{K \in \mathfrak{M}} m_{K \cap K^*} (u_{K^*} - u_K). \end{aligned}$$

It clearly satisfies : for all  $u_\mathcal{T}, v_\mathcal{T} \in \mathbb{R}^\mathcal{T}$ ,

$$(14) \quad \llbracket \mathcal{P}^\mathcal{T} u_\mathcal{T}, v_\mathcal{T} \rrbracket_\mathcal{T} = \frac{1}{2} \frac{1}{\text{Size}(\mathcal{T})^\beta} \sum_{K^* \in \overline{\mathfrak{M}^*}} \sum_{K \in \mathfrak{M}} m_{K \cap K^*} (u_K - u_{K^*}) (v_K - v_{K^*}) \quad \text{with } \beta \in (0, 2).$$

Finally, we introduce a reconstruction operator on diamonds  $r^\mathfrak{D}$ . It is a mapping from  $\mathbb{R}^\mathcal{T}$  to  $\mathbb{R}^\mathfrak{D}$  defined for all  $u_\mathcal{T} \in \mathbb{R}^\mathcal{T}$  by  $r^\mathfrak{D}[u_\mathcal{T}] = (r^\mathcal{D}(u_\mathcal{T}))_{\mathcal{D} \in \mathfrak{D}}$ , where for  $\mathcal{D} \in \mathfrak{D}$ , whose vertices are  $x_K, x_L, x_{K^*}, x_{L^*}$ ,

$$(15) \quad r^\mathcal{D}(u_\mathcal{T}) = \frac{1}{4} (u_K + u_L + u_{K^*} + u_{L^*}).$$

We conclude this section with a remark on the particular structure of the scalar product of two discrete gradients  $(\nabla^\mathfrak{D} u_\mathcal{T}, \nabla^\mathfrak{D} v_\mathcal{T})_{\mathbf{A}, \mathfrak{D}}$  for  $u_\mathcal{T}, v_\mathcal{T} \in \mathbb{R}^\mathcal{T}$ . Indeed, for  $u_\mathcal{T} \in \mathbb{R}^\mathcal{T}$  and  $\mathcal{D} \in \mathfrak{D}$ , we define  $\delta^\mathcal{D} u_\mathcal{T}$  by

$$\delta^\mathcal{D} u_\mathcal{T} = \begin{pmatrix} u_K - u_L \\ u_{K^*} - u_{L^*} \end{pmatrix}.$$

Then, we can write

$$(\nabla^\mathfrak{D} u_\mathcal{T}, \nabla^\mathfrak{D} v_\mathcal{T})_{\mathbf{A}, \mathfrak{D}} = \sum_{\mathcal{D} \in \mathfrak{D}} \delta^\mathcal{D} u_\mathcal{T} \cdot \mathbb{A}^\mathcal{D} \delta^\mathcal{D} v_\mathcal{T},$$

where the local matrices  $\mathbb{A}^{\mathcal{D}}$  are defined by

$$(16) \quad \mathbb{A}^{\mathcal{D}} = \frac{1}{4m_{\mathcal{D}}} \begin{pmatrix} m_{\sigma}^2(\Lambda^{\mathcal{D}} \mathbf{n}_{K,\sigma} \cdot \mathbf{n}_{K,\sigma}) & m_{\sigma} m_{\sigma^*}(\Lambda^{\mathcal{D}} \mathbf{n}_{K,\sigma} \cdot \mathbf{n}_{K^*,\sigma^*}) \\ m_{\sigma} m_{\sigma^*}(\Lambda^{\mathcal{D}} \mathbf{n}_{K,\sigma} \cdot \mathbf{n}_{K^*,\sigma^*}) & m_{\sigma^*}^2(\Lambda^{\mathcal{D}} \mathbf{n}_{K^*,\sigma^*} \cdot \mathbf{n}_{K^*,\sigma^*}) \end{pmatrix} = \begin{pmatrix} A_{\sigma,\sigma}^{\mathcal{D}} & A_{\sigma,\sigma^*}^{\mathcal{D}} \\ A_{\sigma^*,\sigma}^{\mathcal{D}} & A_{\sigma^*,\sigma^*}^{\mathcal{D}} \end{pmatrix}.$$

It follows from elementary calculations left to the reader that the condition number of  $\mathbb{A}^{\mathcal{D}}$  with respect to the 2-norm can be bounded by

$$(17) \quad \text{Cond}_2(\mathbb{A}^{\mathcal{D}}) \leq \text{Cond}_2(\Lambda^{\mathcal{D}}) \left( \theta_{\mathcal{D}} + \sqrt{\theta_{\mathcal{D}}^2 - \frac{1}{\text{Cond}_2(\Lambda^{\mathcal{D}})}} \right)^2 < 4(\theta^*)^2 \frac{\lambda^M}{\lambda_m}, \quad \forall \mathcal{D} \in \mathfrak{D}.$$

**2.3. The nonlinear DDFV scheme.** Let  $N_T$  be a positive integer, we consider for simplicity the constant time step is given by  $\Delta t = T/N_T$ . For  $n \in \{0, \dots, N_T\}$ , we denote by  $t^n = n\Delta t$ . We first discretize the initial condition by taking the mean values of  $u_0$ , i.e.,

$$(18) \quad u_K^0 = \frac{1}{m_K} \int_K u_0 d\mathbf{x}, \quad u_{K^*}^0 = \frac{1}{m_{K^*}} \int_K u_0 d\mathbf{x}, \quad \forall K \in \mathfrak{M}, \forall K^* \in \overline{\mathfrak{M}}, \quad u_{\partial\mathfrak{M}}^0 = 0,$$

and the exterior potential  $V$  by taking its nodal values on the primal and dual cells, i.e.,

$$(19) \quad V_K = V(\mathbf{x}_K), \quad V_{K^*} = V(\mathbf{x}_{K^*}), \quad \forall K \in \mathfrak{M}, \forall K^* \in \overline{\mathfrak{M}}.$$

It defines in particular  $u_{\mathcal{T}}^0$  and  $V_{\mathcal{T}}$ .

The scheme requires a stabilization parameter denoted by  $\kappa > 0$ . It is a fixed parameter. Then, for all  $n \geq 0$ , we look for  $u_{\mathcal{T}}^{n+1} \in (\mathbb{R}_+^*)^{\mathcal{T}}$  solution to the following variational formulation:

$$(20a) \quad \left[ \frac{u_{\mathcal{T}}^{n+1} - u_{\mathcal{T}}^n}{\Delta t}, \psi_{\mathcal{T}} \right]_{\mathcal{T}} + T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}; g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}) + \kappa [\mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}]_{\mathcal{T}} = 0, \quad \forall \psi_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}},$$

$$(20b) \quad T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}; g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}) = \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} g_{\mathcal{T}}^{n+1} \cdot \mathbb{A}^{\mathcal{D}} \delta^{\mathcal{D}} \psi_{\mathcal{T}},$$

$$(20c) \quad g_{\mathcal{T}}^{n+1} = \log(u_{\mathcal{T}}^{n+1}) + V_{\mathcal{T}}.$$

Let us mention that, in view of its implementation, the scheme can be rewritten on each mesh as follows:

$$(21a) \quad \frac{u_{\mathfrak{M}}^{n+1} - u_{\mathfrak{M}}^n}{\Delta t} + \text{div}^{\mathfrak{M}}(J_{\mathfrak{D}}^{n+1}) + \kappa \mathcal{P}^{\mathfrak{M}} g_{\mathcal{T}}^{n+1} = 0,$$

$$(21b) \quad \frac{u_{\mathfrak{M}^*}^{n+1} - u_{\mathfrak{M}^*}^n}{\Delta t} + \text{div}^{\mathfrak{M}^*}(J_{\mathfrak{D}}^{n+1}) + \kappa \mathcal{P}^{\mathfrak{M}^*} g_{\mathcal{T}}^{n+1} = 0,$$

$$(21c) \quad \frac{u_{\partial\mathfrak{M}}^{n+1} - u_{\partial\mathfrak{M}}^n}{\Delta t} + \text{div}^{\partial\mathfrak{M}}(J_{\mathfrak{D}}^{n+1}) + \kappa \mathcal{P}^{\partial\mathfrak{M}} g_{\mathcal{T}}^{n+1} = 0,$$

$$(21d) \quad J_{\mathfrak{D}}^{n+1} = -r^{\mathfrak{D}}[u_{\mathcal{T}}^{n+1}] \Lambda^{\mathfrak{D}} \nabla^{\mathfrak{D}} g_{\mathcal{T}}^{n+1},$$

$$(21e) \quad m_{\sigma} J_{\mathcal{D}}^{n+1} \cdot \mathbf{n} = 0, \quad \forall \mathcal{D} = \mathcal{D}_{\sigma,\sigma^*} \in \mathfrak{D}_{ext}.$$

**2.4. Functional spaces.** For a given vector  $u_{\mathcal{T}}$  defined on a DDFV mesh  $\mathcal{T}$  of size  $h$ , one usually reconstructs three different approximate solutions :  $u_{h,\mathfrak{M}}$  is a piecewise constant reconstruction on the primal mesh,  $u_{h,\overline{\mathfrak{M}}^*}$  is a piecewise constant reconstruction on the dual mesh and  $u_h$  is the mean value of  $u_{h,\mathfrak{M}}$  and  $u_{h,\overline{\mathfrak{M}}^*}$ . They are defined by

$$u_{h,\mathfrak{M}} = \sum_{K \in \mathfrak{M}} u_K \mathbf{1}_K, \quad u_{h,\overline{\mathfrak{M}}^*} = \sum_{K^* \in \overline{\mathfrak{M}}^*} u_{K^*} \mathbf{1}_{K^*} \text{ and } u_h = \frac{1}{2}(u_{h,\mathfrak{M}} + u_{h,\overline{\mathfrak{M}}^*}).$$

Then, the set of the approximate solutions is denoted by  $H_{\mathcal{T}}$ :

$$(22) \quad H_{\mathcal{T}} = \left\{ u_h \in L^1(\Omega) \mid \exists u_{\mathcal{T}} = ((u_K)_{K \in \mathfrak{M}}, (u_{K^*})_{K^* \in \overline{\mathfrak{M}^*}}) \in \mathbb{R}^{\mathcal{T}} \right. \\ \left. \text{such that } u_h = \frac{1}{2} \sum_{K \in \mathfrak{M}} u_K \mathbf{1}_K + \frac{1}{2} \sum_{K^* \in \overline{\mathfrak{M}^*}} u_{K^*} \mathbf{1}_{K^*} \right\}.$$

In the sequel, we will also need some reconstruction of the approximate solutions on the diamond cells. Thanks to the reconstruction operator on diamonds  $r^{\mathfrak{D}}$ , we can define  $u_{\mathcal{D}} = r^{\mathcal{D}}(u_{\mathcal{T}})$  for all  $\mathcal{D} \in \mathfrak{D}$  for instance. Therefore, we can define a piecewise constant function on diamond cells  $u_{h,\mathfrak{D}}$  by  $u_{h,\mathfrak{D}} = \sum_{\mathcal{D} \in \mathfrak{D}} u_{\mathcal{D}} \mathbf{1}_{\mathcal{D}}$ . The set of such functions is denoted  $H_{\mathfrak{D}}$ .

For a function  $u_h \in H_{\mathcal{T}}$ , we define its approximate gradient  $\nabla^h u_h \in (H_{\mathfrak{D}})^2$  by

$$\nabla^h u_h = \sum_{\mathcal{D} \in \mathfrak{D}} \nabla^{\mathcal{D}} u_{\mathcal{T}} \mathbf{1}_{\mathcal{D}}.$$

As the problem (1) is an evolutive problem, the numerical scheme (20) defines  $u_{\mathcal{T}}^n \in \mathbb{R}^{\mathcal{T}}$  for all  $n \in \{0, \dots, N_T\}$ . We consider approximate solutions which are piecewise constant in time. Therefore, we define the space-time approximation spaces  $H_{\mathcal{T},\Delta t}$  and  $H_{\mathfrak{D},\Delta t}$  based respectively on  $H_{\mathcal{T}}$  and  $H_{\mathfrak{D}}$ :

$$H_{\mathcal{T},\Delta t} = \left\{ u_{h,\Delta t} \in L^1(Q_T) \mid u_{h,\Delta t}(\mathbf{x}, t) = u_h^n(\mathbf{x}) \quad \forall t \in [t_{n-1}, t_n), \text{ with } u_h^n \in H_{\mathcal{T}}, \quad \forall 1 \leq n \leq N_T \right\}, \\ H_{\mathfrak{D},\Delta t} = \left\{ u_{h,\Delta t,\mathfrak{D}} \in L^1(Q_T) \mid u_{h,\Delta t,\mathfrak{D}}(\mathbf{x}, t) = u_{h,\mathfrak{D}}^n(\mathbf{x}) \quad \forall t \in [t_{n-1}, t_n), \right. \\ \left. \text{with } u_{h,\mathfrak{D}}^n \in H_{\mathfrak{D}}, \quad \forall 1 \leq n \leq N_T \right\}.$$

We still keep the notation  $\nabla^h$  to define the approximate gradient of  $u_{h,\Delta t} \in H_{\mathcal{T},\Delta t}$ :

$$\nabla^h u_{h,\Delta t}(\mathbf{x}, t) = \nabla^h u_h^n(\mathbf{x}) \quad \forall t \in [t_{n-1}, t_n).$$

Therefore, for all  $u_{h,\Delta t} \in H_{\mathcal{T},\Delta t}$ , we have  $\nabla^h u_{h,\Delta t} \in (H_{\mathfrak{D},\Delta t})^2$ . Furthermore, we introduce the following reconstructions

$$(23a) \quad u_{h,\Delta t,\mathfrak{M}}(\mathbf{x}, t) = u_{h,\mathfrak{M}}^n(\mathbf{x}) = \sum_{K \in \mathfrak{M}} u_K^n \mathbf{1}_K(\mathbf{x}), \quad \forall t \in [t_{n-1}, t_n),$$

$$(23b) \quad u_{h,\Delta t,\overline{\mathfrak{M}^*}}(\mathbf{x}, t) = u_{h,\overline{\mathfrak{M}^*}}^n(\mathbf{x}) = \sum_{K^* \in \overline{\mathfrak{M}^*}} u_{K^*}^n \mathbf{1}_{K^*}(\mathbf{x}), \quad \forall t \in [t_{n-1}, t_n).$$

We may now introduce some norms on the functional spaces  $H_{\mathcal{T}}$  and  $H_{\mathcal{T},\Delta t}$ . For a discrete solution  $u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ , we define  $|u_{\mathcal{T}}|_{p,\mathcal{T}}$  for  $1 \leq p \leq \infty$  by

$$|u_{\mathcal{T}}|_{p,\mathcal{T}}^p = \left( \frac{1}{2} \sum_{K \in \mathfrak{M}} m_K |u_K|^p + \frac{1}{2} \sum_{K^* \in \overline{\mathfrak{M}^*}} m_{K^*} |u_{K^*}|^p \right)^{1/p} \\ |u_{\mathcal{T}}|_{\infty,\mathcal{T}} = \max \left( \max_{K \in \mathfrak{M}} |u_K|, \max_{K^* \in \overline{\mathfrak{M}^*}} |u_{K^*}| \right).$$

It permits to define discrete  $W^{1,p}$ -norms ( $1 \leq p \leq +\infty$ ) and a discrete  $W^{-1,1}$ -norm on  $H_{\mathcal{T}}$ . For all  $u_h \in H_{\mathcal{T}}$ , we set

$$\begin{aligned} \|u_h\|_{1,p,\mathcal{T}} &= \left( |u_{\mathcal{T}}|_{p,\mathcal{T}}^p + \|\nabla^h u_h\|_p^p \right)^{1/p}, \quad \forall 1 \leq p < +\infty, \\ \|u_h\|_{1,\infty,\mathcal{T}} &= |u_{\mathcal{T}}|_{\infty,\mathcal{T}} + \|\nabla^h u_h\|_{\infty}, \\ \|u_h\|_{1,\infty^*,\mathcal{T}} &= \|u_h\|_{1,\infty,\mathcal{T}} + \llbracket \mathcal{P}^{\mathcal{T}} u_{\mathcal{T}}, u_{\mathcal{T}} \rrbracket_{\mathcal{T}}^{\frac{1}{2}}, \\ \|u_h\|_{-1,1,\mathcal{T}} &= \max \left\{ \llbracket v_{\mathcal{T}}, u_{\mathcal{T}} \rrbracket_{\mathcal{T}}, \forall v_h \in H_{\mathcal{T}} \text{ verifying } \|v_h\|_{1,\infty^*,\mathcal{T}} \leq 1 \right\}. \end{aligned}$$

Let us just remark that, as  $\nabla^h u_h$  is a piecewise constant function on diamonds, we have:

$$\|\nabla^h u_h\|_p^p = \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} |\nabla^{\mathcal{D}} u_{\mathcal{T}}|^p \quad \forall 1 \leq p < +\infty \text{ and } \|\nabla^h u_h\|_{\infty} = \max_{\mathcal{D} \in \mathfrak{D}} |\nabla^{\mathcal{D}} u_{\mathcal{T}}|.$$

Then, we define some discrete  $L^q(0, T; W^{1,p}(\Omega))$  ( $1 \leq p, q < +\infty$ ),  $L^{\infty}(0, T; W^{1,\infty}(\Omega))$  and  $L^{\infty}(0, T; L^p(\Omega))$ -norms on  $H_{\mathcal{T},\Delta t}$ . For all  $u_{h,\Delta t} \in H_{\mathcal{T},\Delta t}$ , we set:

$$\begin{aligned} \|u_{h,\Delta t}\|_{q;1,p,\mathcal{T}} &= \left( \sum_{n=1}^{N_T} \Delta t \|u_h^n\|_{1,p,\mathcal{T}}^q \right)^{1/q}, \quad \forall 1 \leq p, q < +\infty, \\ \|u_{h,\Delta t}\|_{\infty;1,\infty,\mathcal{T}} &= \max_{n \in \{1, \dots, N_T\}} \|u_h^n\|_{1,\infty,\mathcal{T}}, \\ \|u_{h,\Delta t}\|_{\infty;0,p,\mathcal{T}} &= \max_{n \in \{1, \dots, N_T\}} \left( \frac{1}{2} \sum_{K \in \mathfrak{M}} m_K |u_K^n|^p + \frac{1}{2} \sum_{K^* \in \overline{\mathfrak{M}^*}} m_{K^*} |u_{K^*}^n|^p \right)^{1/p}, \quad \forall 1 \leq p < +\infty. \end{aligned}$$

**2.5. Main results.** The numerical analysis of the scheme strongly relies on a discrete version of the energy/energy dissipation relation (4). In order to make it explicit, let us introduce the discrete counterpart  $(\mathbb{E}_{\mathcal{T}}^n)_{n \geq 0}$  of the free energy  $\mathbb{E}$  defined by (5):

$$\mathbb{E}_{\mathcal{T}}^n = \llbracket H(u_{\mathcal{T}}^n), 1_{\mathcal{T}} \rrbracket_{\mathcal{T}} + \llbracket V_{\mathcal{T}}, u_{\mathcal{T}}^n \rrbracket_{\mathcal{T}}, \quad \forall n \geq 0,$$

and the discrete counterpart  $(\mathbb{I}_{\mathcal{T}}^n)_{n \geq 1}$  of the dissipation  $\mathbb{I}$  defined by (6):

$$(24) \quad \mathbb{I}_{\mathcal{T}}^n = T_{\mathfrak{D}}(u_{\mathcal{T}}^n; g_{\mathcal{T}}^n, g_{\mathcal{T}}^n), \quad \forall n \geq 1.$$

The first main result of our paper is the existence of a positive solution to the nonlinear scheme (20); it is stated in Theorem 2.1. The mesh is given and fulfills the very permissive requirements of Section 2.1. Our nonlinear scheme (20) yields a nonlinear system of algebraic equations. The fact that this system admits a solution is not obvious and is ensured by Theorem 2.1. The proof strongly relies on the fact that the scheme fulfills a discrete entropy/dissipation relation.

**Theorem 2.1** (Existence of a discrete solution). *For all  $n \geq 0$ , there exists a solution  $u_{\mathcal{T}}^{n+1} \in (\mathbb{R}_+^*)^{\mathcal{T}}$  to the nonlinear system (20) that satisfies the discrete entropy/entropy dissipation estimate*

$$(25) \quad \frac{\mathbb{E}_{\mathcal{T}}^{n+1} - \mathbb{E}_{\mathcal{T}}^n}{\Delta t} + \mathbb{I}_{\mathcal{T}}^{n+1} \leq 0, \quad \forall n \geq 0.$$

Once the existence of  $u_{\mathcal{T}}^{n+1}$  at hand for all  $n \geq 0$ , we can reconstruct the approximate solutions  $u_{h,\Delta t}$ ,  $u_{h,\Delta t,\mathfrak{M}}$ , and  $u_{h,\Delta t,\mathfrak{M}^*}$ . The convergence of these approximate solutions towards a weak

solution when the mesh size and the time step tend to 0 is then a very natural question. This question is addressed in Theorem 2.2.

In what follows,  $(\mathcal{T}_m)_{m \geq 1} = (\mathfrak{M}_m, \overline{\mathfrak{M}_m^*})_{m \geq 1}$  denotes a sequence of admissible discretization of  $\Omega$  and  $(\mathfrak{D}_m)_{m \geq 1}$  denotes the corresponding diamond mesh. We assume that the

$$(26) \quad \text{size}(\mathcal{T}_m) \xrightarrow{m \rightarrow \infty} 0, \quad \text{whereas} \quad \limsup_{m \rightarrow \infty} \max_{\mathcal{D} \in \mathfrak{D}_m} \max(\theta_{\mathcal{D}}, \tilde{\theta}_{\mathcal{D}}) \leq \theta^*,$$

the regularity factors  $\theta_{\mathcal{D}}$  and  $\tilde{\theta}_{\mathcal{D}}$  being defined by (10).

Concerning the time discretization, we consider a sequence  $(N_{T,m})_{m \geq 1}$  of positive integers tending to  $+\infty$ , and we denote by  $(\Delta t_m)_{m \geq 1} = \left(\frac{T}{N_{T,m}}\right)_{m \geq 1}$  the corresponding sequence of time steps. For technical reasons that will appear later on, and even though this condition does not seem to be mandatory from a practical point of view, we have to make the assumption that there exists some constant  $C_1 > 0$  such that

$$(27) \quad \Delta t_m \geq C_1 \text{size}(\mathcal{T}_m), \quad \forall m \geq 1.$$

The existence of a discrete solution  $u_{\mathcal{T}_m}^{n+1}$  to the scheme (20) for all  $n \in \{0, \dots, N_{T,m} - 1\}$  and all  $m \geq 1$  stated in Theorem 2.1 allows us to define the approximate solutions  $u_{h_m, \Delta t_m}$ ,  $u_{h_m, \Delta t_m, \mathfrak{M}_m}$ , and  $u_{h_m, \Delta t_m, \overline{\mathfrak{M}_m^*}}$  for all  $m \geq 1$ . The next theorem ensures that, up to a subsequence, the sequences of approximate solution converge towards a weak solution of the problem (1).

**Theorem 2.2** (Convergence towards a weak solution). *Assume that (26) and (27) holds. Then there exists a weak solution  $u$  in the sense of Definition 1.1 such that, up to a subsequence,*

$$u_{h_m, \Delta t_m, \mathfrak{M}_m} \xrightarrow{m \rightarrow \infty} u, \quad u_{h_m, \Delta t_m, \overline{\mathfrak{M}_m^*}} \xrightarrow{m \rightarrow \infty} u, \quad \text{and} \quad u_{h_m, \Delta t_m} \xrightarrow{m \rightarrow \infty} u \quad \text{in } L^p(0, T; L^1(\Omega))$$

for all  $p \in [1, \infty)$ .

Further convergence properties are established during the proof of Theorem 2.2. We don't make them explicit here in order to minimize the notations and to improve the readability of the paper. We refer to Section 4.1 for refined statements.

**Remark 2.3.** *The convergence of the scheme is only assessed up to a subsequence in Theorem 2.2. This comes from the fact that the uniqueness of weak solutions in the sense of Definition 1.1 is still an open problem even for initial data  $u_0$  belonging to  $L^2(\Omega)$ . However, we conjecture that for  $u_0$  being such that  $H(u_0)$  belongs to  $L^1(\Omega)$ , the weak solutions in the sense of Definition 1.1 are renormalized solutions (see for instance [9]). Uniqueness should follow, implying the convergence of the whole sequence.*

The main goal of the paper is to prove Theorems 2.1 and 2.2. The proof is articulated as follows: In Section 3, we derive some estimates on the discrete solution. These *a priori* estimates allow us to show that the nonlinear system originating from the (20) admits (at least) one solution, as claimed in Theorem 2.1. Most of the estimates derived in Section 3 are uniform w.r.t.  $m$ . This provides enough compactness on the approximate solutions to pass to the limit  $m \rightarrow \infty$  in Section 4.

### 3. ENERGY AND DISSIPATION ESTIMATES, EXISTENCE OF A SOLUTION TO THE SCHEME

**3.1. *a priori* estimates.** The first statement of this section is devoted to what we call the fundamental estimates, that are discrete counterparts of the conservation of mass and of the energy/dissipation relation (4). All the further *a priori* estimates on the discrete solution are based on these two estimates.

**Proposition 3.1** (fundamental estimates). *Let  $(u_{\mathcal{T}}^n)_{n \geq 1}$ , with  $u_{\mathcal{T}}^n \in (\mathbb{R}_+^*)^{\mathcal{T}}$  for all  $n \geq 0$ , be a solution to the scheme (20) corresponding to the initial data  $u_0$ . Then,*

(i) *the mass is conserved along time, i.e.,*

$$(28) \quad \int_{\Omega} u_h^n d\mathbf{x} = \llbracket u_{\mathcal{T}}^n, 1_{\mathcal{T}} \rrbracket_{\mathcal{T}} = \int_{\Omega} u_0 d\mathbf{x}, \quad \forall n \geq 0,$$

(ii) *the discrete free energy is dissipated along time, i.e.,*

$$(29) \quad \frac{\mathbb{E}_{\mathcal{T}}^{n+1} - \mathbb{E}_{\mathcal{T}}^n}{\Delta t} + \mathbb{I}_{\mathcal{T}}^{n+1} + \kappa \llbracket \mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^{n+1}, g_{\mathcal{T}}^{n+1} \rrbracket_{\mathcal{T}} \leq 0, \quad \forall n \geq 0.$$

Moreover, the discrete free energy is decaying along time and is bounded:

$$(30) \quad 0 \leq \mathbb{E}_{\mathcal{T}}^{n+1} \leq \mathbb{E}_{\mathcal{T}}^n \leq \mathbb{E}_{\mathcal{T}}^0 \leq \int_{\Omega} H(u_0) d\mathbf{x} + \|V\|_{\infty} \|u_0\|_{L^1(\Omega)}$$

and the “integrated over time” dissipation is also bounded:

$$(31) \quad 0 \leq \sum_{n=1}^{N_{\mathcal{T}}} \Delta t \mathbb{I}_{\mathcal{T}}^n \leq \sum_{n=1}^{N_{\mathcal{T}}} \Delta t (\mathbb{I}_{\mathcal{T}}^n + \kappa \llbracket \mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^n, g_{\mathcal{T}}^n \rrbracket_{\mathcal{T}}) \leq \int_{\Omega} H(u_0) d\mathbf{x} + \|V\|_{\infty} \|u_0\|_{L^1(\Omega)}.$$

*Proof.* Equation (28) is obtained directly by choosing  $\psi_{\mathcal{T}} = 1_{\mathcal{T}}$  in (20a). In order to get Estimate (29), it suffices to take  $\psi_{\mathcal{T}} = g_{\mathcal{T}}^{n+1}$  in (20a) and to remark that, because of the convexity of  $u \mapsto H(u) + uV$ , one has  $\llbracket u_{\mathcal{T}}^{n+1} - u_{\mathcal{T}}^n, g_{\mathcal{T}}^{n+1} \rrbracket_{\mathcal{T}} \geq \mathbb{E}_{\mathcal{T}}^{n+1} - \mathbb{E}_{\mathcal{T}}^n$ . Inequality (30) is just a consequence of (29), of the nonnegativity of the dissipation and of the penalization term and of Jensen’s inequality. By summation of (29) over  $n$ , we deduce (31).  $\square$

The goal of the remaining part of this section is to take advantage of the fundamental estimates of Proposition 3.1 to derive some further estimates to be used in the numerical analysis. As in the continuous framework, the energy/dissipation estimate (29) is used in order to estimate the discrete counterpart of the Fisher information. But in the discrete framework, the chain rule appearing in (7) does not longer hold, and we have to manipulate several objects related to the Fisher information. The last goal of this section is to prove that, for a fixed grid and a fixed time step, the discrete solutions  $(u_{\mathcal{T}}^n)_{n \geq 1}$  is uniformly bounded away from 0, cf. Lemma 3.5.

Define the discrete fields  $\xi_{\mathcal{T}}^n = \sqrt{u_{\mathcal{T}}^n}$  which play a key role as in the continuous level. In order to relate different discrete counterparts of the Fisher information, we first have to derive some properties on the local diffusion matrices  $\mathbb{A}^{\mathcal{D}}, \mathcal{D} \in \mathfrak{D}$ .

Let  $\mathcal{D} \in \mathfrak{D}$ , then we define the diagonal matrix  $\mathbb{B}^{\mathcal{D}}$  by

$$(32) \quad \mathbb{B}^{\mathcal{D}} = \begin{pmatrix} B_{\sigma}^{\mathcal{D}} & 0 \\ 0 & B_{\sigma^*}^{\mathcal{D}} \end{pmatrix} = \begin{pmatrix} |A_{\sigma, \sigma}^{\mathcal{D}}| + |A_{\sigma, \sigma^*}^{\mathcal{D}}| & 0 \\ 0 & |A_{\sigma^*, \sigma^*}^{\mathcal{D}}| + |A_{\sigma, \sigma^*}^{\mathcal{D}}| \end{pmatrix}.$$

For all  $\mathbf{w} = (w_{\sigma}, w_{\sigma^*})^T \in \mathbb{R}^2$  and all  $\mathcal{D} \in \mathfrak{D}$ , there holds

$$\mathbb{A}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w} \leq \mathbb{B}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w} \leq \|\mathbb{A}^{\mathcal{D}}\|_1 |\mathbf{w}|^2,$$

where  $\|\cdot\|_q$  is the usual matrix  $q$ -norm and  $|\cdot|$  is the Euclidian norm on  $\mathbb{R}^2$ . It follows from the equivalence of the matrix 1- and 2- norms on the finite dimensional space  $\mathbb{R}^{2 \times 2}$  that

$$\|\mathbb{A}^{\mathcal{D}}\|_1 |\mathbf{w}|^2 \leq \gamma \|\mathbb{A}^{\mathcal{D}}\|_2 |\mathbf{w}|^2 \leq \gamma \text{Cond}_2(\mathbb{A}^{\mathcal{D}}) \mathbb{A}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w}, \quad \forall \mathbf{w} \in \mathbb{R}^2$$

for some  $\gamma \geq 1$ . Therefore, in view of (17), we get the existence of  $C_2$  depending only on  $\theta^*$ ,  $\lambda_m$  and  $\lambda^M$  such that

$$(33) \quad \mathbb{A}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w} \leq \mathbb{B}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w} \leq C_2 \mathbb{A}^{\mathcal{D}} \mathbf{w} \cdot \mathbf{w}, \quad \forall \mathcal{D} \in \mathfrak{D}, \forall \mathbf{w} \in \mathbb{R}^2.$$

We introduce now a discrete counterpart of  $\int_{\Omega} u \Lambda \nabla \log u \cdot \nabla \log u$ , it is the quantity  $\widehat{\mathbb{I}}_{\mathcal{T}}^n$  defined by

$$(34) \quad \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1} = \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}) \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}), \quad n \geq 0.$$

Let us first relate this quantity to a discrete Fisher information.

**Lemma 3.2.** *For all  $n \geq 0$ , there holds*

$$\|\nabla^{\mathfrak{D}} \xi_{\mathcal{T}}^{n+1}\|_{\Lambda, \mathfrak{D}}^2 \leq \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1}.$$

*Proof.* Thanks to the first inequality of (33), one has

$$(35) \quad (\nabla^{\mathfrak{D}} \xi_{\mathcal{T}}^{n+1}, \nabla^{\mathfrak{D}} \xi_{\mathcal{T}}^{n+1})_{\Lambda, \mathfrak{D}} \leq \sum_{\mathcal{D} \in \mathfrak{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1}.$$

It results from the elementary inequality

$$\left| \sqrt{b} - \sqrt{a} \right| \leq \frac{\max(\sqrt{a}, \sqrt{b})}{2} |\log(b) - \log(a)|, \quad \forall (a, b) \in (\mathbb{R}_+^*)^2,$$

that for all  $\mathcal{D} \in \mathfrak{D}$ , one has

$$\begin{aligned} |\xi_K^{n+1} - \xi_L^{n+1}| &\leq \frac{\max(\xi_K^{n+1}, \xi_L^{n+1}, \xi_{K^*}^{n+1}, \xi_{L^*}^{n+1})}{2} |\log(u_K^{n+1}) - \log(u_L^{n+1})|, \\ |\xi_{K^*}^{n+1} - \xi_{L^*}^{n+1}| &\leq \frac{\max(\xi_K^{n+1}, \xi_L^{n+1}, \xi_{K^*}^{n+1}, \xi_{L^*}^{n+1})}{2} |\log(u_K^{n+1}) - \log(u_{L^*}^{n+1})|. \end{aligned}$$

This yields

$$\delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1} \leq \frac{\max(u_K^{n+1}, u_L^{n+1}, u_{K^*}^{n+1}, u_{L^*}^{n+1})}{4} \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}) \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}),$$

and since  $\max(u_K^{n+1}, u_L^{n+1}, u_{K^*}^{n+1}, u_{L^*}^{n+1}) \leq 4r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1})$ , one gets

$$(36) \quad \delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^{n+1} \leq r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}) \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}), \quad \forall \mathcal{D} \in \mathfrak{D}.$$

In order to conclude the proof of Lemma 3.2, it only remains to combine (35) and (36).  $\square$

We now want to get a bound on  $\widehat{\mathbb{I}}_{\mathcal{T}}^{n+1}$  in order to deduce some bound on  $\|\nabla^{\mathfrak{D}} \xi_{\mathcal{T}}^{n+1}\|_{\Lambda, \mathfrak{D}}$ . Therefore, we first need to establish an estimate on the discrete reconstruction by diamond  $r^{\mathfrak{D}}[u_{\mathcal{T}}^{n+1}]$ .

**Lemma 3.3.** *Let  $r^{\mathfrak{D}}[u_{\mathcal{T}}^{n+1}] \in \mathbb{R}^{\mathfrak{D}}$  be defined by (15). There exists  $C > 0$ , depending only on  $\Omega$ ,  $\lambda_m$ ,  $\lambda^M$  and  $\theta^*$  such that*

$$(37) \quad \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \leq C(1 + \text{size}(\mathcal{T})) \int_{\Omega} u_0 d\mathbf{x} + C \text{size}(\mathcal{T}) \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1}, \quad \forall n \geq 0.$$

*Proof.* The definition (15) implies that

$$(38) \quad \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) = (T^{n+1} + T^{n+1,*})/4,$$

where we have set

$$T^{n+1} = \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} (u_K^{n+1} + u_L^{n+1}) \quad \text{and} \quad T^{n+1,*} = \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} (u_{K^*}^{n+1} + u_{L^*}^{n+1}), \quad \forall n \geq 0.$$

The terms  $T^{n+1}$  and  $T^{n+1,*}$  can be rewritten

$$\begin{aligned} T^{n+1} &= \sum_{K \in \mathfrak{M}} u_K^{n+1} \sum_{\mathcal{D} \in \mathfrak{D}_K} m_{\mathcal{D}} + \sum_{L \in \partial \mathfrak{M}} u_L^{n+1} m_{\mathcal{D}_L} = T_{\mathfrak{M}}^{n+1} + T_{\partial \mathfrak{M}}^{n+1}, \\ T^{n+1,*} &= \sum_{K^* \in \overline{\mathfrak{M}^*}} u_{K^*}^{n+1} \sum_{\mathcal{D} \in \mathfrak{D}_{K^*}} m_{\mathcal{D}}, \end{aligned}$$

where  $\mathcal{D}_L$  denotes the unique diamond cell associated to the primal boundary cell  $L \in \partial \mathfrak{M}$ . The terms  $T_{\mathfrak{M}}^{n+1}$  and  $T^{n+1,*}$  can be estimated thanks to the regularity of the mesh (13) by

$$T_{\mathfrak{M}}^{n+1} + T^{n+1,*} \leq \theta^* \left( \sum_{K \in \mathfrak{M}} m_K u_K^{n+1} + \sum_{K^* \in \overline{\mathfrak{M}^*}} u_{K^*}^{n+1} m_{K^*} \right).$$

We deduce from (28) that

$$T_{\mathfrak{M}}^{n+1} + T^{n+1,*} \leq 2\theta^* \int_{\Omega} u_0 d\mathbf{x}.$$

Let us now focus on the term  $T_{\partial \mathfrak{M}}^{n+1}$ . The area of the diamond cell  $\mathcal{D}$  corresponding to a boundary edge  $\sigma \subset \partial \Omega$  (or equivalently to a primal boundary cell  $L \in \partial \mathfrak{M}$ ) can be estimated by

$$m_{\mathcal{D}} \leq \frac{1}{2} m_{\sigma} \text{size}(\mathcal{T}).$$

Therefore, we get that

$$T_{\partial \mathfrak{M}}^{n+1} \leq \frac{1}{2} \text{size}(\mathcal{T}) \|\gamma_{\partial \mathfrak{M}} u_{\mathcal{T}}^{n+1}\|_{L^1(\partial \Omega)} = \frac{1}{2} \text{size}(\mathcal{T}) \|\gamma_{\partial \mathfrak{M}} \xi_{\mathcal{T}}^{n+1}\|_{L^2(\partial \Omega)}^2,$$

where  $\gamma_{\partial \mathfrak{M}} u_{\mathcal{T}}(\mathbf{x}) = \sum_{L \in \partial \mathfrak{M}} u_L \mathbf{1}_L(\mathbf{x})$ ,  $\forall \mathbf{x} \in \partial \Omega$ . The trace inequality stated in Theorem A.1 gives with  $v_{\mathcal{T}} = \xi_{\mathcal{T}}^{n+1}$

$$T_{\partial \mathfrak{M}}^{n+1} \leq C \text{size}(\mathcal{T}) \left( |\xi_{\mathcal{T}}^{n+1}|_{2,\mathcal{T}}^2 + \|\nabla^h \xi_h^{n+1}\|_2^2 \right).$$

Thanks to (3) and the regularity of the mesh (10) and (11), there exists  $C > 0$  depending only on  $\lambda_m$ ,  $\lambda^M$  and  $\theta^*$  such that

$$(39) \quad \|\nabla^h \xi_h^{n+1}\|_2 \leq C \|\nabla^{\mathfrak{D}} \xi_{\mathcal{T}}^{n+1}\|_{\mathbf{A},\mathfrak{D}}$$

Since  $|\xi_{\mathcal{T}}^{n+1}|_{2,\mathcal{T}}^2 = |u_{\mathcal{T}}^{n+1}|_{1,\mathcal{T}}^2$ , Proposition 3.1 and Lemma 3.2 provide that

$$T_{\partial \mathfrak{M}}^{n+1} \leq C \text{size}(\mathcal{T}) \left( \int_{\Omega} u_0 d\mathbf{x} + \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1} \right).$$

□

Thanks to (37), it is now possible to relate  $\widehat{\mathbb{I}}_{\mathcal{T}}^{n+1}$  to  $\mathbb{I}_{\mathcal{T}}^{n+1}$ .

**Lemma 3.4.** *There exist  $C > 0$  and  $h^* > 0$  depending only on  $u_0$ ,  $V$ ,  $\lambda_m$ ,  $\lambda^M$  and  $\theta^*$  such that*

$$(40) \quad \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1} \leq C (1 + \mathbb{I}_{\mathcal{T}}^{n+1}), \quad \forall n \geq 0, \quad \text{if } \text{size}(\mathcal{T}) \leq h^*.$$

*Proof.* Bearing in mind the definition (24) of the dissipation  $\mathbb{I}_{\mathcal{T}}^{n+1}$ , we deduce thanks to (33) that

$$\mathbb{I}_{\mathcal{T}}^{n+1} \leq \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} g_{\mathcal{T}}^{n+1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} g_{\mathcal{T}}^{n+1} \leq C_2 \mathbb{I}_{\mathcal{T}}^{n+1}, \quad \forall n \geq 0.$$

But, as  $g_{\mathcal{T}}^{n+1} = \log u_{\mathcal{T}}^{n+1} + V_{\mathcal{T}}$ , the elementary inequality  $(a+b)^2 \leq 2(a^2 + b^2)$  implies that

$$(41) \quad \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1} \leq 2C_2 \mathbb{I}_{\mathcal{T}}^{n+1} + 2 \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} V_{\mathcal{T}} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} V_{\mathcal{T}}.$$

It follows from the regularity of  $V$  and from the regularity of the mesh (11) that

$$0 \leq \delta^{\mathcal{D}} V_{\mathcal{T}} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} V_{\mathcal{T}} \leq \|\nabla V\|_{\infty}^2 (B_{\sigma}^{\mathcal{D}} m_{\sigma}^* + B_{\sigma^*}^{\mathcal{D}} m_{\sigma}^2) \leq C \|\nabla V\|_{\infty}^2 m_{\mathcal{D}}, \quad \forall \mathcal{D} \in \mathfrak{D}$$

for some  $C$  depending only on  $\lambda^M$  and  $\theta^*$ . Therefore, we deduce from Lemma 3.3 that

$$(42) \quad \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} V_{\mathcal{T}} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} V_{\mathcal{T}} \leq C \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \leq C(1 + \text{size}(\mathcal{T})) \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1}.$$

We infer from (41) and (42) that if  $\text{size}(\mathcal{T})$  is small enough, (40) holds.  $\square$

We have at hand the necessary tool to address the uniform positivity of the solutions. The next lemma states that the discrete solutions remain bounded away from 0 by a small quantity  $\epsilon > 0$  depending on the data of the continuous problem and on the discretization parameters. This information is of great importance since, because of the singularity of the log, the nonlinear functional corresponding to the scheme is not continuous on the boundary of  $(\mathbb{R}_+^*)^T$ . The proof is inspired from the ones of [13, Lemma 3.10] and [14, Lemma 3.7]. We sketch it here to highlight how we overpass the difficulties related to the fact that there is a limited communication between the primal and dual meshes.

**Lemma 3.5.** *There exists  $\epsilon > 0$  depending on the data  $u_0$  and  $V$ , on the mesh  $\mathcal{T}$ , on the time step  $\Delta t$ , and on the stabilization parameter  $\kappa$ , such that*

$$(43) \quad u_{K^*}^{n+1} \geq \epsilon \quad \text{and} \quad u_{K^*}^n \geq \epsilon, \quad \forall K \in \overline{\mathfrak{M}}, \forall K^* \in \overline{\mathfrak{M}^*}, \forall n \geq 0.$$

*Proof.* In this proof as elsewhere in the paper, the generic constants  $C$  only depend on the data of the continuous problem and on the regularity bound  $\theta^*$  for the mesh. In order to highlight the dependency of a quantity with respect to the mesh or to the time step, we use subscripts. For instance  $C_{\Delta t}$  may depend on  $\Delta t$ , whereas  $C_{\mathcal{T}}$  may depend on the mesh and  $C_{\mathcal{T}, \Delta t}$  may depend on the mesh and on the time step.

Owing to (28), there exists  $M_0 \in \mathfrak{M} \cap \overline{\mathfrak{M}^*}$  such that

$$u_{M_0}^{n+1} \geq \frac{1}{m_{\Omega}} \int_{\Omega} u_0 d\mathbf{x} > 0,$$

implying in particular that

$$(44) \quad \log(u_{M_0}^{n+1}) \geq -C \quad \text{and} \quad r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \geq \frac{1}{4m_{\Omega}} \int_{\Omega} u_0 d\mathbf{x} > 0 \quad \text{for all } \mathcal{D} \in \mathfrak{D}_{M_0}.$$

On the other hand, it follows from (31) that  $\mathbb{I}_{\mathcal{T}}^{n+1} \leq C_{\Delta t}$ . Together with Estimate (40), this provides that

$$(45) \quad \widehat{\mathbb{I}}_{\mathcal{T}}^{n+1} \leq C_{\Delta t}.$$

Assume for instance that  $M_0 = K_0 \in \mathfrak{M}$  (the case  $M_0 \in \overline{\mathfrak{M}^*}$  is similar). Since  $B_{\sigma}^{\mathcal{D}} > \frac{1}{C}$  for all  $\mathcal{D} \in \mathfrak{D}$ , we deduce from (44)–(45) that  $\log(u_{K_1}^{n+1}) \geq -C_{\Delta t}$  for all neighboring cell  $K_1 \in \mathfrak{M}$  such that

$K_0|K_1 \in \mathcal{E}$  and for all  $L \in \partial\mathfrak{M} \cap \partial K_0$ . It follows from a simple induction based on the reproduction of this argument (see [13, Lemma 3.10] or [14, Lemma 3.7] for details) that

$$(46) \quad \log(u_K^{n+1}) \geq -C_{\Delta t, \mathcal{T}}, \quad \forall K \in \overline{\mathfrak{M}}.$$

Let  $K^* \in \overline{\mathfrak{M}^*}$ , then there exists  $K \in \mathfrak{M}$  such that  $m(K \cap K^*) > 0$ . Thanks to the penalization term in Estimate (29), one has that

$$(g_{K^*}^{n+1} - g_K^{n+1})^2 \leq C_{\mathcal{T}, \Delta t}.$$

The regularity of  $V$  implies that

$$(\log(u_{K^*}^{n+1}) - \log(u_K^{n+1}))^2 \leq C_{\mathcal{T}, \Delta t},$$

hence, using (46), one gets that

$$(47) \quad \log(u_{K^*}^{n+1}) \geq -C_{\Delta t, \mathcal{T}}, \quad \forall K^* \in \overline{\mathfrak{M}^*}.$$

The relation (43) follows from (46) and (47).  $\square$

**3.2. Existence of a solution to the scheme.** The numerical scheme (20) amounts at each time step  $n \geq 0$  to solve a nonlinear system  $\mathcal{F}^n(u_{\mathcal{T}}^{n+1}) = \mathbf{0}$ . The existence of a solution  $u_{\mathcal{T}}^{n+1}$  is therefore non trivial. It is established in the following proposition.

**Proposition 3.6.** *For all  $n \geq 0$ , there exists (at least) one solution  $u_{\mathcal{T}}^{n+1} \in (\mathbb{R}_+^*)^{\mathcal{T}}$  to the nonlinear system (20).*

The proof of Proposition 3.6 relies on a topological degree argument [39, 21, 27]. The key point is that, owing to Lemma 3.5, one can restrain our search for the solution on a compact subset of  $(\mathbb{R}_+^*)^{\mathcal{T}}$  on which the functional  $\mathcal{F}^n$  is (uniformly) continuous, making Leray-Schauder's theorem applicable. We do not detail the proof here since it is very close to the one of [14, Proposition 3.8].

#### 4. CONVERGENCE W.R.T. DISCRETIZATION PARAMETERS

**4.1. Compactness of the approximate solutions.** Thanks to Proposition 3.6, we have at hand discrete solutions  $(u_{\mathcal{T}}^{n+1})_{n \geq 0}$  corresponding to all the time steps, and thus the corresponding reconstructions  $u_{h, \Delta t} \in H_{\mathcal{T}, \Delta t}$ ,  $u_{h, \mathfrak{D}, \Delta t}$  as defined in Section 2.4. We also define  $\xi_{h, \Delta t} \in H_{\mathcal{T}, \Delta t}$  based on  $\xi_{\mathcal{T}}^n = \sqrt{u_{\mathcal{T}}^n}$  for all  $n \in \{0, \dots, N_T\}$ .

Thanks to the estimates established in Section 3.1, we can obtain some further estimates satisfied by the discrete reconstructions. These estimates, stated in Lemma 4.1, will then be used to deduce some compactness properties of sequences of approximate solutions.

**Lemma 4.1.** (i) *There exists  $C$  depending only on  $u_0$  and  $V$  such that*

$$(48) \quad \int_{\Omega} H(u_{h, \Delta t}(\mathbf{x}, t)) d\mathbf{x} \leq \frac{1}{2} \left( \int_{\Omega} H(u_{h, \Delta t, \mathfrak{M}}(\mathbf{x}, t)) d\mathbf{x} + \int_{\Omega} H(u_{h, \Delta t, \mathfrak{M}^*}(\mathbf{x}, t)) d\mathbf{x} \right) \leq C, \quad \forall t \geq 0.$$

(ii) *There exists  $C$  depending only on  $u_0$ ,  $V$ ,  $\lambda_m$ ,  $\lambda^M$  and  $\theta^*$  such that, for  $\text{size}(\mathcal{T})$  small enough, one has*

$$(49) \quad \sum_{n=1}^{N_T} \Delta t \widehat{\mathbb{I}}_{\mathcal{T}}^n \leq C(1+T) \quad \text{and} \quad \iint_{Q_T} |\nabla^h \xi_{h, \Delta t}|^2 d\mathbf{x} dt \leq C(1+T).$$

(iii) There exists  $C$  depending only on  $\theta^*$ ,  $u_0$ ,  $V$ ,  $C_1$ ,  $\lambda_m$  and  $\lambda^M$  such that

$$(50) \quad \int_{\Omega} u_{h,\mathfrak{D}}^n d\mathbf{x} \leq C(1+T), \quad \forall n \in \{1, \dots, N_T\}.$$

*Proof.* The first inequality in (48) is just a consequence of Jensen's inequality because  $H$  is a convex function. Moreover, since we assumed that  $V \geq 0$  and proved the positivity of the discrete solution, we have:

$$\begin{aligned} \int_{\Omega} H(u_{h,\Delta t}(\mathbf{x}, t)) &\leq \frac{1}{2} \left( \int_{\Omega} H(u_{h,\Delta t,\mathfrak{M}}(\mathbf{x}, t)) d\mathbf{x} + \int_{\Omega} H(u_{h,\Delta t,\mathfrak{M}^*}(\mathbf{x}, t)) d\mathbf{x} \right) \\ &\leq \mathbb{E}_{\mathcal{T}}^{n+1}, \quad \forall t \in (t^n, t^{n+1}], \end{aligned}$$

The last inequality in (48) is then a straightforward consequence of (30). The estimates in (49) are deduced from (31), (40) and Lemma 3.2.

It remains to prove (50). From Lemma 3.3, we have that

$$\int_{\Omega} u_{h,\mathfrak{D}}^n d\mathbf{x} \leq C \left( 1 + \text{size}(\mathcal{T}) \widehat{\mathbb{I}}_{\mathcal{T}}^n \right), \quad \forall n \geq 1.$$

We infer from the first inequality of (49) that  $\widehat{\mathbb{I}}_{\mathcal{T}}^n \leq C(1+T)/\Delta t$  and the assumption (27) implies that

$$\text{size}(\mathcal{T}) \widehat{\mathbb{I}}_{\mathcal{T}}^n \leq C(1+T), \quad \forall n \geq 1.$$

This concludes the proof of Lemma 4.1.  $\square$

In order to get the compactness of a sequence of approximate solutions, it is also crucial to establish a discrete counterpart of a  $L^1(0, T; W^{-1,1}(\Omega))$  estimate on the discrete time derivative. In what follows, we denote by

$$\partial_{t,\mathcal{T}} u_{h,\Delta t}^n = \left( \left( \frac{u_K^{n+1} - u_K^n}{\Delta t} \right)_{K \in \mathfrak{M}}, \left( \frac{u_{K^*}^{n+1} - u_{K^*}^n}{\Delta t} \right)_{K^* \in \mathfrak{M}^*} \right) \in \mathbb{R}^{\mathcal{T}}, \quad \forall n \geq 0.$$

**Lemma 4.2.** *There exists  $C$  depending only on  $V$ ,  $u_0$ ,  $T$ ,  $\kappa$ ,  $\theta^*$ ,  $C_1$ ,  $\lambda_m$  and  $\lambda^M$  such that*

$$\sum_{n=0}^{N_T-1} \Delta t \left\| \partial_{t,\mathcal{T}} u_{h,\Delta t}^n \right\|_{-1,1,\mathcal{T}} \leq C.$$

*Proof.* We proceed as in [18, Lemma 3.4]. It follows from (20a) that for all  $\psi_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ , one has

$$\llbracket \partial_{t,\mathcal{T}} u_{h,\Delta t}^n, \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}} = -T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}) - \kappa \llbracket \mathcal{P}(g_{\mathcal{T}}^{n+1}), \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}}, \quad \forall n \geq 0.$$

The application  $(g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}) \mapsto T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, g_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}) + \kappa \llbracket \mathcal{P}(g_{\mathcal{T}}^{n+1}), \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}}$  is a scalar product on  $\mathbb{R}^{\mathcal{T}}$ , then it follows from Cauchy-Schwarz inequality that

$$\llbracket \partial_{t,\mathcal{T}} u_{h,\Delta t}^n, \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}} \leq (\mathbb{I}_{\mathcal{T}}^{n+1} + \kappa \llbracket \mathcal{P}(g_{\mathcal{T}}^{n+1}), g_{\mathcal{T}}^{n+1} \rrbracket_{\mathcal{T}})^{1/2} (T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}, \psi_{\mathcal{T}}) + \kappa \llbracket \mathcal{P}(\psi_{\mathcal{T}}), \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}})^{1/2}.$$

We can estimate the term  $T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}, \psi_{\mathcal{T}})$  by

$$T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}, \psi_{\mathcal{T}}) = \int_{\Omega} u_{h,\mathfrak{D}}^{n+1} \Lambda_{h,\mathfrak{D}} \nabla^h \psi_h \cdot \nabla^h \psi_h d\mathbf{x} \leq \lambda^M \left\| u_{h,\mathfrak{D}}^{n+1} \right\|_{L^1(\Omega)} \left\| \nabla^h \psi_h \right\|_{\infty}^2.$$

Thanks to (50) and since  $u_{h,\mathfrak{D}}^{n+1} \geq 0$ , one gets that

$$T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}, \psi_{\mathcal{T}}) \leq C \left\| \nabla^h \psi_h \right\|_{\infty}^2, \quad \forall n \geq 0,$$

therefore

$$(T_{\mathcal{D}}(u_{\mathcal{T}}^{n+1}, \psi_{\mathcal{T}}, \psi_{\mathcal{T}}) + \kappa \llbracket \mathcal{P}(\psi_{\mathcal{T}}), \psi_{\mathcal{T}} \rrbracket_{\mathcal{T}})^{1/2} \leq C \|\psi_h\|_{1,\infty^*,\mathcal{T}}, \quad \forall \psi_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}.$$

Since  $\psi_{\mathcal{T}}$  can be chosen arbitrarily, this implies that

$$\|\partial_{t,\mathcal{T}} u_{h,\Delta t}^n\|_{-1,1,\mathcal{T}} \leq C (\mathbb{I}_{\mathcal{T}}^{n+1} + \kappa \llbracket \mathcal{P}(g_{\mathcal{T}}^{n+1}), g_{\mathcal{T}}^{n+1} \rrbracket_{\mathcal{T}})^{1/2}, \quad \forall n \in \{0, \dots, N_T - 1\}.$$

Thanks to Cauchy-Schwarz inequality once again, we obtain that

$$\begin{aligned} \sum_{n=0}^{N_T-1} \Delta t \|\partial_{t,\mathcal{T}} u_{h,\Delta t}^n\|_{-1,1,\mathcal{T}} &\leq \sqrt{T} \left( \sum_{n=0}^{N_T-1} \Delta t \|\partial_{t,\mathcal{T}} u_{h,\Delta t}^n\|_{-1,1,\mathcal{T}}^2 \right)^{1/2} \\ &\leq C \left( \sum_{n=1}^{N_T} \Delta t (\mathbb{I}_{\mathcal{T}}^n + \kappa \llbracket \mathcal{P}(g_{\mathcal{T}}^n), g_{\mathcal{T}}^n \rrbracket_{\mathcal{T}}) \right)^{1/2}. \end{aligned}$$

One concludes the proof by using (31).  $\square$

Let  $(\mathcal{T}_m)_{m \geq 1}$  be a sequence of meshes as in Section 2.1 such that  $\text{size}(\mathcal{T}_m)$  tends to 0 as  $m$  tends to  $\infty$ , and such that the regularity of the discretization  $\mathcal{T}_m$  is uniformly bounded w.r.t.  $m$ , i.e.,

$$1 \leq \theta_{\mathcal{D}}, \tilde{\theta}_{\mathcal{D}} \leq \theta^*, \quad \forall \mathcal{D} \in \mathfrak{D}_m, \forall m \geq 1.$$

Assume moreover that simultaneously the time step  $\Delta t_m$  tends to 0 as  $m$  tends to  $\infty$  while satisfying (27). The estimates stated in this section are uniform w.r.t.  $m$ . We deduce from Lemma 4.1 that the sequences  $(u_{h_m, \Delta t_m})_{m \geq 1}$ ,  $(u_{h_m, \Delta t_m, \mathfrak{M}})_{m \geq 1}$ , and  $(u_{h_m, \Delta t_m, \mathfrak{M}^*})_{m \geq 1}$  are equi-integrable in  $L^1(Q_T)$  while uniformly bounded in  $L^\infty(0, T; L^1(\Omega))$ , hence there exists  $u, u^{(1)}, u^{(2)} \in L^\infty(0, T; L^1(\Omega))$  such that, for all  $p \in [1, \infty)$ , the following convergence holds up to the extraction of an unlabeled subsequence:

$$(51a) \quad u_{h_m, \Delta t_m} \xrightarrow{m \rightarrow \infty} u \quad \text{weakly in } L^p(0, T; L^1(\Omega)),$$

$$(51b) \quad u_{h_m, \Delta t_m, \mathfrak{M}} \xrightarrow{m \rightarrow \infty} u^{(1)} \quad \text{weakly in } L^p(0, T; L^1(\Omega)).$$

$$(51c) \quad u_{h_m, \Delta t_m, \mathfrak{M}^*} \xrightarrow{m \rightarrow \infty} u^{(2)} \quad \text{weakly in } L^p(0, T; L^1(\Omega)).$$

Moreover, it follows from (28) and (49) that

$$(52) \quad \|\xi_{h, \Delta t}\|_{2;1,2,\mathcal{T}} \leq C.$$

Thus, similarly to [2, Proposition 5.3] (which is strongly related to [3, Lemma 3.8]), we get the existence of  $\xi \in L^2(0, T; H^1(\Omega))$  such that

$$(53) \quad \xi_{h_m, \Delta t_m} \xrightarrow{m \rightarrow \infty} \xi \quad \text{weakly in } L^2(Q_T), \quad \nabla^h \xi_{h_m, \Delta t_m} \xrightarrow{m \rightarrow \infty} \nabla \xi \quad \text{weakly in } L^2(Q_T)^2.$$

Up to now, we only have weak convergence results, which are not sufficient to pass to the limit in the scheme because of the nonlinearities. The purpose of the following proposition is to recover some strong convergence results. Our approach is based on the time-compactness toolbox presented in [4] (see also [33] for a closely related approach).

**Proposition 4.3.** *Up to the extraction of an unlabeled subsequence,*

$$(54a) \quad u_{h_m, \Delta t_m} \xrightarrow{m \rightarrow \infty} u \quad \text{strongly in } L^p(0, T; L^1(\Omega)),$$

$$(54b) \quad u_{h_m, \Delta t_m, \mathfrak{M}_m} \xrightarrow{m \rightarrow \infty} u \quad \text{strongly in } L^p(0, T; L^1(\Omega)),$$

$$(54c) \quad u_{h_m, \Delta t_m, \mathfrak{M}_m^*} \xrightarrow{m \rightarrow \infty} u \quad \text{strongly in } L^p(0, T; L^1(\Omega)),$$

$$(54d) \quad u_{h_m, \Delta t_m, \mathfrak{D}_m} \xrightarrow{m \rightarrow \infty} u \quad \text{strongly in } L^1(Q_T),$$

for all  $p \in [1, \infty)$ . Moreover, let  $\xi \in L^2(0, T, H^1(\Omega))$  be as in (53), then  $\xi = \sqrt{u}$ .

*Proof.* The proof is divided into three steps. We remove the subscripts  $m$  for the ease of reading.

*Step 1.* The goal of this part is to make use on both  $(u_{h, \Delta t, \mathfrak{M}})$  and  $(u_{h, \Delta t, \mathfrak{M}^*})$  of the time-compactness criterion of [4, Theorem 3.9].

It follows directly from their definitions that

$$u_{h, \Delta t, \mathfrak{M}} = (\xi_{h, \Delta t, \mathfrak{M}})^2, \quad u_{h, \Delta t, \mathfrak{M}^*} = (\xi_{h, \Delta t, \mathfrak{M}^*})^2.$$

Thanks to discrete Poincaré-Sobolev Inequality [8], it follows from (52) that

$$\|\xi_{h, \Delta t, \mathfrak{M}}\|_{L^2(0, T; L^p(\Omega))} \leq C_p, \quad \|\xi_{h, \Delta t, \mathfrak{M}^*}\|_{L^2(0, T; L^p(\Omega))} \leq C_p, \quad \forall p \in [1, +\infty),$$

for some  $C_p$  depending only on  $p$ , on  $\Omega$  and on the regularity  $\theta^*$  of the mesh and therefore

$$(55) \quad \|u_{h, \Delta t, \mathfrak{M}}\|_{L^1(0, T; L^p(\Omega))} \leq C_p, \quad \|u_{h, \Delta t, \mathfrak{M}^*}\|_{L^1(0, T; L^p(\Omega))} \leq C_p, \quad \forall p \in [1, \infty).$$

On the other hand, the mass conservation (28) and the positivity of the solutions yield

$$(56) \quad \|u_{h, \Delta t, \mathfrak{M}}\|_{L^\infty(0, T; L^1(\Omega))} \leq C, \quad \|u_{h, \Delta t, \mathfrak{M}^*}\|_{L^\infty(0, T; L^1(\Omega))} \leq C.$$

It results from (55), (56) and from Riesz-Thorin interpolation theorem that

$$(57) \quad \|u_{h, \Delta t, \mathfrak{M}}\|_{L^p(Q_T)} \leq C_p, \quad \|u_{h, \Delta t, \mathfrak{M}^*}\|_{L^p(Q_T)} \leq C_p, \quad \forall p \in [1, 2),$$

thus in particular for  $p = 3/2$ . The weak limits  $u^{(1)}$ ,  $u^{(2)}$  of  $u_{h, \Delta t, \mathfrak{M}}$  and  $u_{h, \Delta t, \mathfrak{M}^*}$  thus belong to  $L^{3/2}(Q_T)$ , and the weak limits of  $\xi_{h, \Delta t, \mathfrak{M}}$  and  $\xi_{h, \Delta t, \mathfrak{M}^*}$  (up to the extraction of an unlabeled subsequence), denoted by  $\xi^{(1)}$ ,  $\xi^{(2)}$ , belong to  $L^3(Q_T)$  of . The functions  $u_{h, \Delta t, \mathfrak{M}}$  and  $\xi_{h, \Delta t, \mathfrak{M}}$ , as well as  $u_{h, \Delta t, \mathfrak{M}^*}$  and  $\xi_{h, \Delta t, \mathfrak{M}^*}$ , are thus in duality.

We now want to apply [4, Theorem 3.9] in order to show that  $\xi^{(1)} = \sqrt{u^{(1)}}$  and  $\xi^{(2)} = \sqrt{u^{(2)}}$ . Therefore, we have to verify that the three assumptions **(A<sub>x</sub>1)**, **(A<sub>x</sub>2)** and **(A<sub>x</sub>3)** of [4] are satisfied.

- (i) As a direct consequence of the arguments developed in the proofs of [3, Lemma 3.8] or [18, Proposition 4.3] and of Poincaré Sobolev embedding [8], any sequence  $(v_{\mathcal{T}_m})_m$  such that  $\|v_h\|_{1,2,\mathcal{T}} \leq C$  is such that  $v_{h, \mathfrak{M}}$  and  $v_{h, \mathfrak{M}^*}$  converges strongly in  $L^3(\Omega)$  (up to the extraction of an unlabeled subsequence). Therefore Assumption **(A<sub>x</sub>1)** of [4] is fulfilled.
- (ii) The reconstructions  $v_{h, \mathfrak{M}}$  and  $v_{h, \mathfrak{M}^*}$  from  $v_{\mathcal{T}}$  are piecewise constant, the functions being equal to nodal values of  $v_{\mathcal{T}}$  a.e. in  $\Omega$ , then Assumption **(A<sub>x</sub>2)** of [4] is fulfilled by these reconstructions. Note here that this is not the case of the reconstruction  $v_h$ .

(iii) Let  $\varphi \in C^\infty(\bar{\Omega})$  and let us define  $\varphi_{\mathcal{T}}$  by

$$(58) \quad \begin{cases} \varphi_K = \frac{1}{m_K} \int_K \varphi d\mathbf{x} & \text{for } K \in \mathfrak{M}, \\ \varphi_L = \frac{1}{m_\sigma} \int_\sigma \varphi d\mathbf{x} & \text{for } L \equiv \sigma \in \partial\mathfrak{M}, \\ \varphi_{K^*} = \frac{1}{m_{K^*}} \int_{K^*} \varphi d\mathbf{x} & \text{for } K^* \in \overline{\mathfrak{M}^*}. \end{cases}$$

Following the proof of [18, Proposition 4.2], there exists  $C$  depending only on the regularity of the mesh  $\theta^*$  such that

$$(59) \quad \|\varphi_h\|_{1,\infty,\mathcal{T}} \leq C \|\nabla \varphi\|_\infty, \quad \forall \varphi \in C^\infty(\bar{\Omega}).$$

On the other hand, one can show that

$$(60) \quad \llbracket \mathcal{P}^{\mathcal{T}} \varphi_{\mathcal{T}}, \varphi_{\mathcal{T}} \rrbracket_{\mathcal{T}} \leq C \|\nabla \varphi\|_\infty^2$$

for some  $C$  depending only on the regularity  $\theta^*$  of the mesh. Therefore,

$$(61) \quad \|\varphi_h\|_{1,\infty,\mathcal{T}} \leq C \|\nabla \varphi\|_\infty, \quad \forall \varphi \in C^\infty(\bar{\Omega}).$$

Assumption **(A<sub>x</sub>3)** of [4] is thus fulfilled.

We can then make use of Lemma 4.2 and apply [4, Theorem 3.9] to claim that  $\xi^{(i)} = \sqrt{u^{(i)}}$  and that

$$(62) \quad \begin{cases} u_{h,\Delta t,\mathfrak{M}} \xrightarrow{m \rightarrow \infty} u^{(1)} \\ u_{h,\Delta t,\mathfrak{M}^*} \xrightarrow{m \rightarrow \infty} u^{(2)} \end{cases} \quad \text{a.e. in } Q_T.$$

Setting  $u = (u^{(1)} + u^{(2)})/2$ , we get that

$$(63) \quad u_{h,\Delta t} \xrightarrow{m \rightarrow \infty} u.$$

Moreover, as the sequences  $(u_{h,\Delta t})_m$ ,  $(u_{h,\Delta t,\mathfrak{M}})_m$ , and  $(u_{h,\Delta t,\mathfrak{M}^*})_m$  are uniformly equi-integrable in  $L^p(0, T, L^1(\Omega))$  and converge point-wise, thanks to (62)–(63), we can apply Vitali's convergence theorem to claim that the sequences converge strongly in  $L^p(0, T, L^1(\Omega))$ .

*Step 2.* Here, the goal is to show that the sequences  $(u_{h,\Delta t})_{m \geq 1}$ ,  $(u_{h,\Delta t,\mathfrak{M}})_{m \geq 1}$  and  $(u_{h,\Delta t,\mathfrak{M}^*})_{m \geq 1}$  share the same limit  $u$ , i.e.,  $u^{(1)} = u^{(2)} = u$ . As a consequence of (31), one has

$$\sum_{n=1}^{N_T} \Delta t \llbracket \mathcal{P}(g_{\mathcal{T}}^n), g_{\mathcal{T}}^n \rrbracket_{\mathcal{T}} \leq C,$$

hence

$$\|\log(u_{h,\Delta t,\mathfrak{M}}) + V_{h,\mathfrak{M}} - \log(u_{h,\Delta t,\mathfrak{M}^*}) - V_{h,\mathfrak{M}^*}\|_{L^2(Q_T)}^2 \leq C \text{size}(\mathcal{T})^\beta.$$

The regularity of the exterior potential  $V$  implies that

$$\|V_{h,\mathfrak{M}} - V_{h,\mathfrak{M}^*}\|_{L^2(Q_T)}^2 \leq C \text{size}(\mathcal{T})^2,$$

so that one gets that

$$\|\log(u_{h,\Delta t,\mathfrak{M}}) - \log(u_{h,\Delta t,\mathfrak{M}^*})\|_{L^2(Q_T)}^2 \leq C \text{size}(\mathcal{T})^\beta.$$

Up to a subsequence, this ensures that  $\log(u_{h,\Delta t,\mathfrak{M}}) - \log(u_{h,\Delta t,\mathfrak{M}^*})$  tends to 0 a.e. in  $Q_T$ . But owing to (62),  $\log(u_{h,\Delta t,\mathfrak{M}})$  tends to  $\log(u^{(1)})$  while  $\log(u_{h,\Delta t,\mathfrak{M}^*})$  tends to  $\log(u^{(2)})$ . Then  $\log(u^{(1)}) = \log(u^{(2)})$  a.e. in  $Q_T$ , thus  $u^{(1)} = u^{(2)} = u$ .

*Step 3.* In order to conclude the proof of Proposition 4.3, it remains to check that (54d) holds. To this end, we compute  $u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}$  on each quarter diamond to get

$$\|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)} \leq \frac{1}{4} \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} (|u_K^n - u_L^n| + |u_{K^*}^n - u_{L^*}^n|).$$

Using the identity  $|a - b| = (\sqrt{a} + \sqrt{b}) |\sqrt{a} - \sqrt{b}|$  for  $a, b \geq 0$ , one gets

$$\begin{aligned} \|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)} &\leq \frac{1}{4} \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} ((\xi_K^n + \xi_L^n) |\xi_K^n - \xi_L^n| + (\xi_{K^*}^n + \xi_{L^*}^n) |\xi_{K^*}^n - \xi_{L^*}^n|) \\ &\leq \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} r^{\mathcal{D}}(\xi_{\mathcal{T}}^n) (|\xi_K^n - \xi_L^n| + |\xi_{K^*}^n - \xi_{L^*}^n|). \end{aligned}$$

Owing to Cauchy-Schwarz inequality, there holds

$$\|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)} \leq \left( \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^n) \right)^{1/2} \left( \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} |\delta^{\mathcal{D}} \xi_{\mathcal{T}}^n|^2 \right)^{1/2}.$$

It is easy to verify that  $B_{\sigma}^{\mathcal{D}} \geq \frac{1}{\theta^*}$  and  $B_{\sigma^*}^{\mathcal{D}} \geq \frac{1}{\theta^*}$  (see (32) for their definition). Hence, using (50), it provides

$$\|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)} \leq C \text{size}(\mathcal{T}) \left( \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^n \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \xi_{\mathcal{T}}^n \right)^{1/2}.$$

Using (36) together with (49), we obtain that

$$\|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)} \leq C \text{size}(\mathcal{T}),$$

thus  $u_{h,\Delta t,\mathfrak{D}}$  also converges towards  $u$  in  $L^1(Q_T)$ . To get the convergence in  $L^p(0, T, L^1(\Omega))$  for any  $p \in [1, \infty)$ , it only remains to write

$$\begin{aligned} \|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^p(0,T,L^1(\Omega))} &\leq \|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^{\frac{p-1}{p}}(0,T,L^1(\Omega))}^{\frac{p-1}{p}} \|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)}^{1/p} \\ &\leq \left( \|u_{h,\Delta t,\mathfrak{D}}\|_{L^{\infty}(0,T,L^1(\Omega))} + \|u_{h,\Delta t}\|_{L^{\infty}(0,T,L^1(\Omega))} \right)^{\frac{p-1}{p}} \|u_{h,\Delta t,\mathfrak{D}} - u_{h,\Delta t}\|_{L^1(Q_T)}^{1/p} \end{aligned}$$

and to use (28) and (50).  $\square$

The purpose of the following statement is the uniform in time weak- $L^1$  in space convergence of the approximate solution towards  $u$ .

**Proposition 4.4.** *Up to the extraction of an additional subsequence,*

$$(64) \quad u_{h_m, \Delta t_m}(\cdot, t) \xrightarrow{m \rightarrow \infty} u(\cdot, t) \quad \text{in the } L^1(\Omega)\text{-weak sense for all } t \in [0, T].$$

Moreover, the limit function  $u$  satisfies

$$\sup_{t \in [0, T]} \int_{\Omega} H(u) d\mathbf{x} \leq C$$

for some  $C$  depending only  $u_0$  and  $V$ .

*Proof.* Let  $R > 0$  be arbitrary. As a consequence of the de La Vallée Poussin theorem [22] the space

$$E_R = \left\{ f : \Omega \rightarrow \mathbb{R}_+ \left| \int_{\Omega} f d\mathbf{x} = \int_{\Omega} u_0 d\mathbf{x} \text{ and } \int_{\Omega} H(f) dx \leq R \right. \right\}$$

is equi-integrable in  $L^1(\Omega)$ , thus it follows from Dunford-Pettis theorem that  $E_R$  is relatively compact for the weak- $L^1(\Omega)$  topology. Since the function  $f \mapsto \int_{\Omega} H(f) d\mathbf{x}$  is lower semi-continuous, any limit value for a sequence of  $E_R$  also belongs to  $E_R$ , hence  $E_R$  is closed, thus compact.

Since  $\Omega$  is bounded and because  $E_R$  is equi-integrable, the  $L^1(\Omega)$ -weak topology coincides with the topology corresponding to the narrow convergence of measures restricted to  $E_R$ . It can thus be endowed with the bounded-Lipschitz metric:

$$\text{dist}_{\text{BL}}(f_n, f) \xrightarrow{n \rightarrow \infty} 0 \quad \text{iff} \quad f_n \xrightarrow{n \rightarrow \infty} f \text{ weakly in } L^1(\Omega).$$

In the above formula,  $f$  and  $f_n$  ( $n \geq 1$ ) belong to  $E_R$ , and

$$\text{dist}_{\text{BL}}(f, g) = \sup_{\|\nabla \varphi\|_{\infty} \leq 1} \int_{\Omega} (f - g) \varphi d\mathbf{x}.$$

We refer for instance to [41, Theorem 5.9] for the equivalence of the topology induced by the bounded-Lipschitz distance with the one of narrow convergence of positive measures. The fact that this latter topology coincides with the weak- $L^1(\Omega)$  topology on  $E_R$  results from its equi-integrability.

As a consequence of Lemma 4.1, there exists  $R$  such that  $u_{h_m, \Delta t_m}(\cdot, t)$  belongs to  $E_R$  for all  $t \in [0, T]$  and all  $m \geq 1$ . Let  $\tau \in (0, T)$ , and let  $t \in (0, T - \tau)$  and let  $\varphi : \bar{\Omega} \rightarrow \mathbb{R}$  be Lipschitz continuous with  $\|\nabla \varphi\|_{\infty} \leq 1$ . Define  $\varphi_{\tau_m}$  as in (58), then (we remove the subscript  $m$  for legibility)

$$\int_{\Omega} (u_{h, \Delta t}(\mathbf{x}, t + \tau) - u_{h, \Delta t}(\mathbf{x}, t)) \varphi(\mathbf{x}) d\mathbf{x} = \left[ u_{\mathcal{T}}^{N^{(2)}} - u_{\mathcal{T}}^{N^{(1)}}, \varphi_{\mathcal{T}} \right]_{\mathcal{T}}$$

where  $N^{(1)}$  and  $N^{(2)}$  are the positive integers such that

$$\left( N^{(1)} - 1 \right) \Delta t < t \leq N^{(1)} \Delta t, \quad \left( N^{(2)} - 1 \right) \Delta t < t + \tau \leq N^{(2)} \Delta t.$$

Using the scheme (20a), we obtain that

$$\left[ u_{\mathcal{T}}^{N^{(2)}} - u_{\mathcal{T}}^{N^{(1)}}, \varphi_{\mathcal{T}} \right]_{\mathcal{T}} = \sum_{n=N^{(1)}+1}^{N^{(2)}} \Delta t \left( T_{\mathfrak{D}}(u_{\mathcal{T}}^n; g_{\mathcal{T}}^n, \varphi_{\mathcal{T}}) + \kappa \left[ \mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^n, \varphi_{\mathcal{T}} \right]_{\mathcal{T}} \right).$$

Cauchy-Schwarz inequality yields

$$\begin{aligned} \left[ u_{\mathcal{T}}^{N^{(2)}} - u_{\mathcal{T}}^{N^{(1)}}, \varphi_{\mathcal{T}} \right]_{\mathcal{T}} &\leq \left( \sum_{n=N^{(1)}+1}^{N^{(2)}} \Delta t \left( T_{\mathfrak{D}}(u_{\mathcal{T}}^n; g_{\mathcal{T}}^n, g_{\mathcal{T}}^n) + \kappa \left[ \mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^n, g_{\mathcal{T}}^n \right]_{\mathcal{T}} \right) \right)^{1/2} \\ &\quad \times \left( \sum_{n=N^{(1)}+1}^{N^{(2)}} \Delta t \left( T_{\mathfrak{D}}(u_{\mathcal{T}}^n; \varphi_{\mathcal{T}}, \varphi_{\mathcal{T}}) + \kappa \left[ \mathcal{P}^{\mathcal{T}} \varphi_{\mathcal{T}}, \varphi_{\mathcal{T}} \right]_{\mathcal{T}} \right) \right)^{1/2}. \end{aligned}$$

The first term in the right-hand side is uniformly bounded by  $\mathbb{E}(0)$  thanks to (31). On the other hand, Estimate (59) together with  $\|\nabla \varphi\|_{\infty} \leq 1$  provide that

$$T_{\mathfrak{D}}(u_{\mathcal{T}}^n; \varphi_{\mathcal{T}}, \varphi_{\mathcal{T}}) \leq C \int_{\Omega} u_{h, \mathfrak{D}}^n d\mathbf{x}, \quad \forall n \in \{1, \dots, N_T\}.$$

Owing to (50), we get that

$$T_{\mathfrak{D}}(u_{\mathcal{T}}^n; \varphi_{\mathcal{T}}, \varphi_{\mathcal{T}}) \leq C, \quad \forall n \in \{1, \dots, N_T\}.$$

Hereby, we deduce that

$$\left\| u_{\mathcal{T}}^{N^{(2)}} - u_{\mathcal{T}}^{N^{(1)}}, \varphi_{\mathcal{T}} \right\|_{\mathcal{T}} \leq C \sqrt{(N^{(2)} - N^{(1)})\Delta t} \leq C \sqrt{\tau + \Delta t},$$

and since  $\varphi$  was chosen arbitrarily in  $\{\varphi : \Omega \rightarrow \mathbb{R} \mid \|\nabla \varphi\|_{\infty} \leq 1\}$ , we obtain that

$$\text{dist}_{\text{BL}}(u_{h,\Delta t}(\cdot, t + \tau), u_{h,\Delta t}(\cdot, t)) \leq C \sqrt{\tau + \Delta t}, \quad \forall t \in [0, T - \tau].$$

We can apply the refined version of Arzelà-Ascoli theorem [1, Proposition 3.3.1] (see also [26, Theorem 4.26]) and claim that  $u_{h,\Delta t}$  converges uniformly towards  $u \in C([0, T]; E_R)$ ,  $E_R$  being endowed with the  $L^1(\Omega)$ -weak topology.  $\square$

**4.2. Identification of the limit.** The goal of this section is to show that the limit function  $u$  exhibited in Proposition 4.3 is a weak solution in the sense of Definition 1.1.

The second and last point to check to complete the proof of Theorem 2.2 is the fact that  $u$  is a solution to (1) in the distributional sense, i.e., that the weak formulation (8) is fulfilled. This is the purpose of the following statement.

**Proposition 4.5.** *Let  $u$  be as in Proposition 4.3, then  $u$  satisfies the weak formulation (8).*

*Proof.* Here again, we remove the subscript  $m$  when it appears us to be detrimental for the readability. Let  $\varphi \in C_c^\infty(\bar{\Omega} \times [0, T])$ , then denote by

$$\varphi_K^n = \varphi(\mathbf{x}_K, t^n), \quad \varphi_{K^*}^n = \varphi(\mathbf{x}_{K^*}, t^n), \quad \forall K \in \bar{\mathfrak{M}}, \forall K^* \in \bar{\mathfrak{M}}^*.$$

Choosing the corresponding  $\psi_{\mathcal{T}} = \varphi_{\mathcal{T}}^n$  in (20a), multiplying by  $\Delta t$  and summing over  $n \in \{0, \dots, N_{T-1}\}$  leads to

$$(65) \quad A_m + B_m + C_m = 0,$$

where

$$\begin{aligned} A_m &= \sum_{n=0}^{N_{T-1}} \llbracket u_{\mathcal{T}}^{n+1} - u_{\mathcal{T}}^n, \varphi_{\mathcal{T}}^n \rrbracket_{\mathcal{T}}, \\ B_m &= \sum_{n=0}^{N_{T-1}} \Delta t T_{\mathfrak{D}}(u_{\mathcal{T}}^{n+1}; g_{\mathcal{T}}^{n+1}, \varphi_{\mathcal{T}}^n), \\ C_m &= \kappa \sum_{n=0}^{N_{T-1}} \Delta t \llbracket \mathcal{P}^{\mathcal{T}} g_{\mathcal{T}}^{n+1}, \varphi_{\mathcal{T}}^n \rrbracket_{\mathcal{T}}. \end{aligned}$$

For the terms  $A_m$  and  $C_m$ , we can proceed as is [18] to get

$$(66) \quad A_m \xrightarrow{m \rightarrow \infty} - \iint_{Q_T} u \partial_t \varphi \, d\mathbf{x} dt - \int_{\Omega} u_0 \varphi(\cdot, 0) d\mathbf{x}, \quad \text{and} \quad C_m \xrightarrow{m \rightarrow \infty} 0.$$

Let us now detail the treatment of the term  $B_m$  and start by splitting it into

$$(67) \quad B_m = B_{1,m} + B_{2,m},$$

where

$$\begin{aligned} B_{1,m} &= \sum_{n=0}^{N_T-1} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} V_{\mathcal{T}} \cdot \mathbb{A}^{\mathcal{D}} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^n \\ &= \iint_{Q_T} u_{h,\Delta t,\mathfrak{D}}(\mathbf{x}, t) \mathbf{\Lambda}_{h,\mathfrak{D}} \nabla^h V_h(\mathbf{x}) \cdot \nabla^h \varphi_{h,\Delta t}(\mathbf{x}, t - \Delta t) d\mathbf{x} dt, \end{aligned}$$

and

$$B_{2,m} = \sum_{n=0}^{N_T-1} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} r^{\mathcal{D}}(u_{\mathcal{T}}^{n+1}) \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^{n+1}) \cdot \mathbb{A}^{\mathcal{D}} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^n.$$

Since  $V$  and  $\varphi$  are smooth functions, one has

$$\nabla^h V_h \xrightarrow{m \rightarrow \infty} \nabla V \quad \text{uniformly on } \Omega, \quad \nabla^h \varphi_{h,\Delta t}(\cdot, \cdot - \Delta t) \xrightarrow{m \rightarrow \infty} \nabla \varphi \quad \text{uniformly on } Q_T,$$

whereas  $\mathbf{\Lambda}_{h,\mathfrak{D}}$  converges a.e. towards  $\mathbf{\Lambda}$ . Then it follows from (54d) that

$$(68) \quad B_{1,m} \xrightarrow{m \rightarrow \infty} \iint_{Q_T} u \mathbf{\Lambda} \nabla V \cdot \nabla \varphi d\mathbf{x} dt.$$

The last term  $B_{2,m}$  is treated following the method proposed in [14] that consists in writing

$$B_{2,m} = B_{2,m}^{(1)} + B_{2,m}^{(2)},$$

with

$$\begin{aligned} B_{2,m}^{(1)} &= 2 \iint_{Q_T} \sqrt{u_{h,\Delta t,\mathfrak{D}}} \mathbf{\Lambda}_{h,\mathfrak{D}} \nabla^h \xi_{h,\Delta t} \cdot \nabla^h \varphi_{h,\Delta t}(\cdot, \cdot - \Delta t) d\mathbf{x} dt, \\ B_{2,m}^{(2)} &= \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} \sqrt{r^{\mathcal{D}}(u_{\mathcal{T}}^n)} \delta^{\mathcal{D}} \log(u_{\mathcal{T}}^n) \cdot \\ &\quad \begin{pmatrix} \xi_{KL}^n - \sqrt{r^{\mathcal{D}}(u_{\mathcal{T}}^n)} & 0 \\ 0 & \xi_{K^*L^*}^n - \sqrt{r^{\mathcal{D}}(u_{\mathcal{T}}^n)} \end{pmatrix} \mathbb{A}^{\mathcal{D}} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^{n-1}, \end{aligned}$$

where we have set

$$(69) \quad \xi_{KL}^n = \begin{cases} 2 \frac{\xi_K^n - \xi_L^n}{\log(u_K^n) - \log(u_L^n)} & \text{if } u_K^n \neq u_L^n, \\ \xi_K^n & \text{if } u_K^n = u_L^n, \end{cases} \quad \xi_{K^*L^*}^n = \begin{cases} 2 \frac{\xi_{K^*}^n - \xi_{L^*}^n}{\log(u_{K^*}^n) - \log(u_{L^*}^n)} & \text{if } u_{K^*}^n \neq u_{L^*}^n, \\ \xi_{K^*}^n & \text{if } u_{K^*}^n = u_{L^*}^n. \end{cases}$$

We know that, up to a subsequence,  $\sqrt{u_{h,\Delta t,\mathfrak{D}}}$  converges strongly in  $L^2(Q_T)$  towards  $\sqrt{u}$ , whereas  $\nabla^h \xi_{h,\Delta t}$  converges weakly in  $L^2(Q_T)^2$  towards  $\nabla \sqrt{u}$ , and  $\nabla^h \varphi_{h,\Delta t}(\cdot, \cdot - \Delta t)$  converges uniformly towards  $\nabla \varphi$ . Thus we can pass to the limit in  $B_{2,m}^{(1)}$  and obtain that

$$(70) \quad B_{2,m}^{(1)} \xrightarrow{m \rightarrow \infty} 2 \iint_{Q_T} \sqrt{u} \mathbf{\Lambda} \nabla \sqrt{u} \cdot \nabla \varphi d\mathbf{x} dt = \iint_{Q_T} \mathbf{\Lambda} \nabla u \cdot \nabla \varphi d\mathbf{x} dt.$$

In order to show that  $B_{2,m}^{(2)}$  tends to 0, we need a few preliminaries. Owing to the definition (69) of  $\xi_{KL}^n$  and  $\xi_{K^*L^*}^n$ , one always has

$$\min(\xi_K^n, \xi_L^n) \leq \xi_{KL}^n \leq \max(\xi_K^n, \xi_L^n), \quad \min(\xi_{K^*}^n, \xi_{L^*}^n) \leq \xi_{K^*L^*}^n \leq \max(\xi_{K^*}^n, \xi_{L^*}^n),$$

so that, denoting by  $\tilde{\xi}_{h,\Delta t,\mathfrak{D}}$  and  $\tilde{\xi}_{h,\Delta t,\mathfrak{D}}^*$  the functions defined almost everywhere by

$$\tilde{\xi}_{h,\Delta t,\mathfrak{D}}(\mathbf{x}, t) = \xi_{KL}^n \quad \text{and} \quad \tilde{\xi}_{h,\Delta t,\mathfrak{D}}^*(\mathbf{x}, t) = \xi_{K^*L^*}^n \quad \text{if } (\mathbf{x}, t) \in \mathcal{D} \times (t_{n-1}, t_n],$$

one obtains that

$$\|\xi_{h,\Delta t,\mathfrak{M}} - \tilde{\xi}_{h,\Delta t,\mathfrak{D}}\|_{L^2(Q_T)}^2 \leq \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} m_{\mathcal{D}} |\xi_K^n - \xi_L^n|^2 \leq C \text{size}(\mathcal{T})^2 \|\nabla^h \xi_{h,\Delta t}\|_{L^2(Q_T)}^2$$

for some  $C$  depending only on  $\theta^*$  and  $\mathbf{A}$ . Similarly, one has

$$\|\xi_{h,\Delta t,\mathfrak{M}^*} - \tilde{\xi}_{h,\Delta t,\mathfrak{D}}^*\|_{L^2(Q_T)} \leq C \text{size}(\mathcal{T}) \|\nabla^h \xi_{h,\Delta t}\|_{L^2(Q_T)}.$$

Bearing in mind that  $\nabla^h \xi_{h,\Delta t}$  is uniformly bounded in  $L^2(Q_T)^2$  w.r.t.  $m$ , this ensures in particular that

$$(71) \quad \tilde{\xi}_{h,\Delta t,\mathfrak{D}} \xrightarrow{m \rightarrow \infty} \xi = \sqrt{u} \quad \text{and} \quad \tilde{\xi}_{h,\Delta t,\mathfrak{D}}^* \xrightarrow{m \rightarrow \infty} \xi = \sqrt{u} \quad \text{in } L^2(Q_T).$$

As a consequence of (54d), the function  $\sqrt{u_{h,\Delta t,\mathfrak{D}}}$  also converges towards  $\xi$  in  $L^2(Q_T)$ .

We now have at hand all the necessary material to study  $B_{2,m}^{(2)}$ . It results from Cauchy-Schwarz inequality that

$$\begin{aligned} B_{2,m}^{(2)} &\leq \left( \sum_{n=1}^{N_T} \Delta t \hat{\mathbb{I}}_{\mathcal{T}}^n \right)^{1/2} \\ &\times \left( \sum_{n=1}^{N_T} \Delta t \sum_{\mathcal{D} \in \mathfrak{D}} \begin{pmatrix} |\xi_{KL}^n - \sqrt{r^{\mathcal{D}}(u_{\mathcal{T}}^n)}|^2 & 0 \\ 0 & |\xi_{K^*L^*}^n - \sqrt{r^{\mathcal{D}}(u_{\mathcal{T}}^n)}|^2 \end{pmatrix} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^{n-1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^{n-1} \right)^{1/2}. \end{aligned}$$

Thanks to the regularity of the mesh and of  $\varphi$ , one has

$$\delta^{\mathcal{D}} \varphi_{\mathcal{T}}^{n-1} \cdot \mathbb{B}^{\mathcal{D}} \delta^{\mathcal{D}} \varphi_{\mathcal{T}}^{n-1} \leq C m_{\mathcal{D}}$$

for some  $C > 0$  depending only on  $\theta^*$  and on  $\|\nabla \varphi\|_{\infty}$ , whereas Lemma 3.4 and (31) ensure that

$$\sum_{n=1}^{N_T} \Delta t \hat{\mathbb{I}}_{\mathcal{T}}^n \leq C.$$

Hence, we get

$$B_{2,m}^{(2)} \leq C \left( \|\tilde{\xi}_{h,\Delta t,\mathfrak{D}} - \sqrt{u_{h,\Delta t,\mathfrak{D}}}\|_{L^2(Q_T)} + \|\tilde{\xi}_{h,\Delta t,\mathfrak{D}}^* - \sqrt{u_{h,\Delta t,\mathfrak{D}}}\|_{L^2(Q_T)} \right).$$

Using (71) together with the fact that  $\sqrt{u_{h,\Delta t,\mathfrak{D}}}$  also converges towards  $\xi$  in  $L^2(Q_T)$ , we get that

$$(72) \quad B_{2,m}^{(1)} \xrightarrow{m \rightarrow \infty} 0.$$

We conclude the proof by putting together the statements (66), (67), (68), (70) and (72) in (65).  $\square$

## 5. NUMERICAL EXPERIMENTS

**5.1. About the practical implementation.** The nonlinear system (20) is solved thanks to Newton's method. In order to avoid the singularity of the log near 0, the sequence  $(u_{\mathcal{T}}^{n+1,i})_{i \geq 0}$  to compute  $u_{\mathcal{T}}^{n+1}$  from the previous state  $(u_{\mathcal{T}}^n)_{i \geq 0}$  is initialized by  $u_{\mathcal{T}}^{n+1,0} = \max(u_{\mathcal{T}}^n, 10^{-12})$ . In practice, we observe that the threshold criterion is not used. As a stopping criterion, we require the  $\ell^1$ -norm of the residual to be smaller than  $10^{-10}$ .

**5.2. Convergence w.r.t. to the discretization parameters.** We test our method on a test case inspired from the one in [14]. We set  $\Omega = (0, 1)^2$ , and  $V(x_1, x_2) = -x_2$ . The exact solution  $u_{\text{ex}}$  is then defined by

$$u_{\text{ex}}((x_1, x_2), t) = e^{-\alpha t + \frac{x_2}{2}} \left( \pi \cos(\pi x_2) + \frac{1}{2} \sin(\pi x_2) \right) + \pi e^{(x_2 - \frac{1}{2})}$$

with  $\alpha = \pi^2 + \frac{1}{4}$ . We choose  $u_0 = u_{\text{ex}}(\cdot, 0)$ . Note that  $u_0$  vanishes on  $\{x_2 = 1\}$ .

In order to illustrate the convergence and the robustness of our method, we study its convergence on two sequences of meshes. The first sequence of primal meshes is made of successively refined Kershaw meshes. The second sequence of primal meshes is the so-called quadrangle meshes `mesh_quad_i` of the FVCA8 benchmark on incompressible flows. One mesh of each sequence is depicted in Figure 2. In the refinement procedure, the time step is divided by 4 when the mesh size is divided by 2.

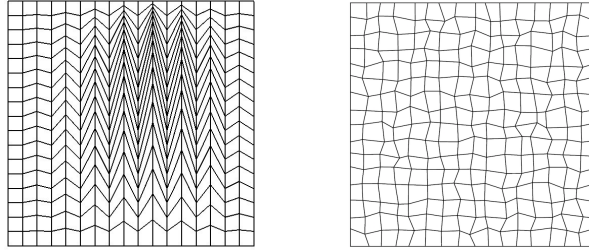


FIGURE 2. Left: First Kershaw mesh. Right: Third quadrangle mesh.

We have introduced a penalization operator in order to prove that reconstruction on the primal mesh  $u_{h,\mathfrak{M},\Delta t}$  and the reconstruction on the dual mesh  $u_{h,\mathfrak{M}^*,\Delta t}$  converge to the same limit. In Table 1, we compute `normU` the  $L^2(\Omega \times (0, T))^2$  norm of the difference between the two different reconstructions and `ordU` the corresponding convergence order for different values of  $\kappa$  the penalization parameter. We numerically observe the same result : the two reconstructions converge to the same limit even if  $\kappa$  is zero.

M	dt	$\kappa = 0$		$\kappa = 10^{-1}$	
		normU	ordU	normU	ordU
1	4.032E-03	1.798E-01	—	1.796E-01	—
2	1.008E-03	9.316E-02	0.95	9.313E-02	0.94
3	2.520E-04	4.717E-02	1.03	4.716E-02	1.03
4	6.300E-05	2.361E-02	1.05	2.361E-02	1.05
5	1.575E-05	1.135E-02	0.87	1.135E-02	0.87

TABLE 1. Numerical results on the Quadrangle mesh family, final time T=0.25.

In the following of this section, the penalization parameter  $\kappa$  is set to zero. In Tables 2 and 3, the quantities `erru` and `errgu` respectively denote the  $L^\infty((0, T); L^2(\Omega))$  error on the solution and the  $L^2(\Omega \times (0, T))^2$  error on the gradient, whereas `ordu` and `ordgu` are the corresponding convergence orders. It appears that the method is slightly more than second order accurate w.r.t. space.

The maximal (resp. mean) number of Newton iterations by time step is denoted by  $N_{\max}$  (resp.  $N_{\text{mean}}$ ). We observe that the needed number of Newton iterations starts from a reasonably small value and falls down to 1 after a small number of time steps. Therefore, our method does not imply an important extra computational cost when compared to linear methods. Eventually, we can check that the minimal value  $\min u_{\mathcal{T}}^n$  remains strictly greater than 0, as proved in Lemma 3.5.

M	dt	errgu	ordgu	erru	ordu	$N_{\max}$	$N_{\text{mean}}$	Min $u^n$
1	2.0E-03	6.693E-02	—	7.254E-03	—	9	2.15	1.010E-01
2	5.0E-04	2.353E-02	1.54	1.751E-03	2.09	8	2.02	2.582E-02
3	1.25E-04	1.235E-02	1.61	7.237E-04	2.20	7	1.49	6.488E-03
4	3.125E-05	7.819E-03	1.60	3.962E-04	2.11	7	1.07	1.628E-03
5	3.125E-05	5.507E-03	1.58	2.556E-04	1.98	7	1.04	1.628E-03

TABLE 2. Numerical results on the Kershaw mesh family, final time T=0.25.

M	dt	errgu	ordgu	erru	ordu	$N_{\max}$	$N_{\text{mean}}$	Min $u^n$
1	4.032E-03	1.754E-01	—	2.149E-02	—	9	2.26	1.803E-01
2	1.008E-03	5.933E-02	1.56	5.055E-03	2.08	9	2.04	5.079E-02
3	2.520E-04	2.294E-02	1.44	1.299E-03	2.06	8	1.96	1.352E-02
4	6.300E-05	8.631E-03	1.48	3.256E-04	2.09	8	1.22	3.349E-03
5	1.250E-05	2.715E-03	1.37	7.702E-05	1.70	7	1.01	8.695E-04

TABLE 3. Numerical results on the Quadrangle mesh family, final time T=0.25.

**5.3. Long time behavior.** In this section, the penalisation parameter  $\kappa$  is set to zero. The discrete stationary solution  $u_{\mathcal{T}}^{\infty}$  is defined by  $u_K^{\infty} = \rho e^{-V(x_K)}$  and  $u_{K^*}^{\infty} = \rho^* e^{-V(x_{K^*})}$  for  $K \in \overline{\mathfrak{M}}$  and  $K^* \in \overline{\mathfrak{M}}^*$ , the quantities  $\rho$  and  $\rho^*$  being fixed so that  $\sum_{K \in \mathfrak{M}} u_K^{\infty} m_K = \sum_{K \in \overline{\mathfrak{M}}^*} u_{K^*}^{\infty} m_{K^*} = \int_{\Omega} u_0(x) dx$ . In order to give an evidence of the good large-time behavior of our scheme, we plot in Figure 3 the evolution of the relative energy

$$\mathbb{E}_{\mathcal{T}}^n - \mathbb{E}_{\mathcal{T}}^{\infty} = \left\| u_{\mathcal{T}}^n \log \left( \frac{u_{\mathcal{T}}^n}{u_{\mathcal{T}}^{\infty}} \right) - u_{\mathcal{T}}^n + u_{\mathcal{T}}^{\infty}, 1_{\mathcal{T}} \right\|_{\mathcal{T}}, \quad n \geq 0$$

computed on the Kershaw meshes and on the quadrangle meshes. We observe the exponential decay of the relative energy, recovering on general grids the behavior of the Scharfetter-Gummel scheme [19].

#### APPENDIX A. A TRACE INEQUALITY

First, to a given vector  $u_{\mathcal{T}} = ((u_K)_{K \in \overline{\mathfrak{M}}}, (u_{K^*})_{K^* \in \overline{\mathfrak{M}}^*}) \in \mathbb{R}^{\mathcal{T}}$  defined on a DDFV mesh  $\mathcal{T}$ , we associate its *primal trace*  $\gamma_{\partial \mathfrak{M}} u_{\mathcal{T}}$  on  $\partial \Omega$  defined by

$$\gamma_{\partial \mathfrak{M}} u_{\mathcal{T}}(\mathbf{x}) = \sum_{L \in \partial \mathfrak{M}} u_L \mathbf{1}_L(\mathbf{x}), \quad \forall \mathbf{x} \in \partial \Omega.$$

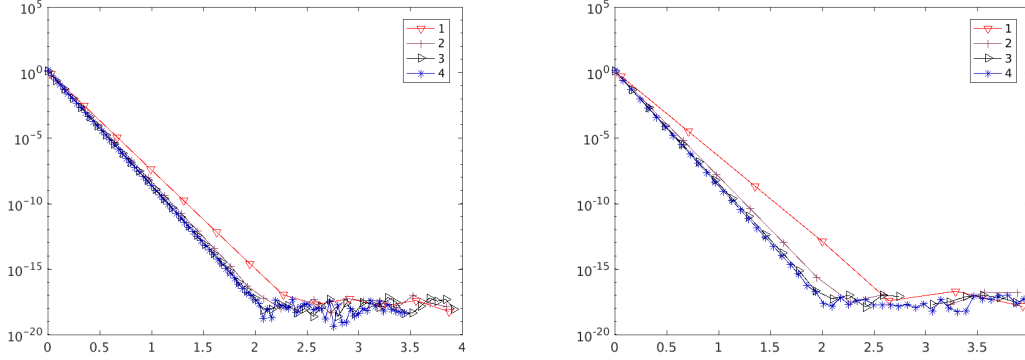


FIGURE 3. Discrete relative energy  $\mathbb{E}_{\mathcal{T}}^n - \mathbb{E}_{\mathcal{T}}^{\infty}$  as a function of  $n\Delta t$  computed on the first four Kershaw meshes (on the left) and on the first four quadrangle meshes (on the right).

**Theorem A.1** (Trace inequality). *Let  $\Omega$  be a convex polygonal domain of  $\mathbb{R}^2$  and  $\mathcal{T}$  a DDFV mesh of this domain. There exist  $C > 0$ , depending only on  $\Omega$  and  $\theta^*$ , such that  $\forall u_{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ :*

$$(73) \quad \|\gamma_{\partial\mathfrak{M}} u_{\mathcal{T}}\|_{2,\partial\Omega} \leq C (|u_{\mathcal{T}}|_{2,\mathcal{T}} + \|\nabla^h u_h\|_2).$$

*Proof.* The calculations are similar to those followed in [28, Lemma 10.5] and in [18, Theorem 7.1] for  $L^1$ -norm. The difference comes from the fact that here we define  $u_{\partial\mathfrak{M}}$  using the boundary primal mesh instead of the interior primal mesh. Adapting the proof of [18, Theorem 7.1] in the  $L^2$ -norm, we get for  $K \in \mathfrak{M}$  such that  $\bar{K} \cap \partial\Omega \neq \emptyset$  the inequality

$$\sum_{\mathcal{D} \in \mathfrak{D}_{ext}} m_{\sigma} |u_K|^2 \leq C (|u_{\mathcal{T}}|_{2,\mathcal{T}}^2 + \|\nabla^h u_h\|_2^2).$$

It implies

$$\begin{aligned} \|\gamma_{\partial\mathfrak{M}} u_{\mathcal{T}}\|_{2,\partial\Omega}^2 &= \sum_{\mathcal{D} \in \mathfrak{D}_{ext}} m_{\sigma} |u_L - u_K + u_K|^2 \\ &\leq 2 \sum_{\mathcal{D} \in \mathfrak{D}_{ext}} m_{\sigma} |u_L - u_K|^2 + 2 \sum_{\mathcal{D} \in \mathfrak{D}_{ext}} m_{\sigma} |u_K|^2 \\ &\leq 2 \sum_{\mathcal{D} \in \mathfrak{D}_{ext}} m_{\sigma} |u_L - u_K|^2 + C (|u_{\mathcal{T}}|_{2,\mathcal{T}}^2 + \|\nabla^h u_h\|_2^2). \end{aligned}$$

Using the fact that  $u_L - u_K = m_{\sigma^*} (\nabla^{\mathcal{D}} u_{\mathcal{T}}) \cdot \tau_{K,L}$ , we conclude

$$\|\gamma_{\partial\mathfrak{M}} u_{\mathcal{T}}\|_{2,\partial\Omega}^2 \leq C (|u_{\mathcal{T}}|_{2,\mathcal{T}}^2 + (1 + \text{size}(\mathcal{T})) \|\nabla^h u_h\|_2^2).$$

□

## REFERENCES

- [1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [2] B. Andreianov, M. Bendahmane, and K. H. Karlsen. Discrete duality finite volume schemes for doubly nonlinear degenerate hyperbolic-parabolic equations. *J. Hyperbolic Differ. Equ.*, 7(1):1–67, 2010.

- [3] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions-type elliptic problems on general 2D meshes. *Numer. Methods Partial Differential Equations*, 23(1):145–195, 2007.
- [4] B. Andreianov, C. Cancès, and A. Moussa. A nonlinear time compactness result and applications to discretization of degenerate parabolic-elliptic PDEs. HAL: hal-01142499, 2015.
- [5] A. Arnold, J. A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P. A. Markowich, G. Toscani, and C. Villani. Entropies and equilibria of many-particle systems: an essay on recent research. *Monatsh. Math.*, 142(1-2):35–43, 2004.
- [6] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.
- [7] M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift-diffusion systems. *Journal of Numerical Mathematics*, 2016.
- [8] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. *IMA J. Numer. Anal.*, 35:1125–1149, 2015.
- [9] D. Blanchard and A. Porretta. Stefan problems with nonlinear diffusion and convection. *J. Differential Equations*, 210(2):383–428, 2005.
- [10] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.
- [11] F. Bolley, I. Gentil, and A. Guillin. Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.*, 208(2):429–445, 2013.
- [12] C. Cancès, C. Chainais-Hillairet, and S. Krell. A nonlinear discrete duality finite volume scheme for convection-diffusion equations. In C. Cancès & P. Omnes, editor, *Finite Volumes for Complex Applications VIII*, Proceedings in Mathematics and Statistics. Springer, 2017.
- [13] C. Cancès and C. 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.
- [14] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Found Comput Math*, 2016.
- [15] J. A. Carrillo and G. Toscani. Exponential convergence toward equilibrium for homogeneous Fokker-Planck-type equations. *Math. Methods Appl. Sci.*, 21(13):1269–1286, 1998.
- [16] J. A. Carrillo and G. Toscani. Asymptotic  $L^1$ -decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, 49(1):113–142, 2000.
- [17] C. Chainais-Hillairet, A. Jüngel, and S. Schuchnigg. Entropy-dissipative discretization of nonlinear diffusion equations and discrete Beckner inequalities. *ESAIM Math. Model. Numer. Anal.*, 50(1):135–162, 2016.
- [18] C. Chainais-Hillairet, S. Krell, and A. Mouton. Convergence analysis of a DDFV scheme for a system describing miscible fluid flows in porous media. *Numer. Methods Partial Differential Equations*, 31(3):723–760, 2015.
- [19] M. Chatard. Asymptotic Behavior of the Scharfetter–Gummel Scheme for the Drift-Diffusion Model. In *FVCA VI*. Springer Berlin Heidelberg, 2011.
- [20] Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two-dimensional convection-diffusion problem. *M2AN Math. Model. Numer. Anal.*, 33(3):493–516, 1999.
- [21] K. Deimling. *Nonlinear functional analysis*. Springer-Verlag, Berlin, 1985.
- [22] C. Dellacherie and P.-A. Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam-New York, 1978.
- [23] L. Desvillettes and K. Fellner. Exponential decay toward equilibrium via entropy methods for reaction-diffusion equations. *J. Math. Anal. Appl.*, 319(1):157–176, 2006.
- [24] L. Desvillettes and K. Fellner. Duality and entropy methods for reversible reaction-diffusion equations with degenerate diffusion. *Math. Methods Appl. Sci.*, 38(16):3432–3443, 2015.
- [25] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.*, 39(6):1203–1249, 2005.
- [26] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. The gradient discretisation method . working paper or preprint, November 2016.
- [27] R. Eymard, T. Gallouët, M. Ghilani, and R. Herbin. Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by finite volume schemes. *IMA J. Numer. Anal.*, 18(4):563–594, 1998.
- [28] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. Ciarlet, P. G. (ed.) et al., in *Handbook of numerical analysis*. North-Holland, Amsterdam, pp. 713–1020, 2000.

- [29] F. Filbet. An asymptotically stable scheme for diffusive coagulation-fragmentation models. *Commun. Math. Sci.*, 6(2):257–280, 2008.
- [30] H. Gajewski and K. Gärtner. On the discretization of van Roosbroeck’s equations with magnetic field. *Z. Angew. Math. Mech.*, 76(5):247–264, 1996.
- [31] H. Gajewski and K. Gröger. On the basic equations for carrier transport in semiconductors. *J. Math. Anal. Appl.*, 113(1):12–35, 1986.
- [32] H. Gajewski and K. Gröger. Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics. *Math. Nachr.*, 140:7–36, 1989.
- [33] T. Gallouët. Some discrete functional analysis tools. In C. Cancès & P. Omnes, editor, *Finite Volumes for Complex Applications VIII*, Proceedings in Mathematics and Statistics. Springer, 2017.
- [34] A. Glitzky. Exponential decay of the free energy for discretized electro-reaction-diffusion systems. *Nonlinearity*, 21(9):1989–2009, 2008.
- [35] A. Glitzky. Uniform exponential decay of the free energy for Voronoi finite volume discretized reaction-diffusion systems. *Math. Nachr.*, 284(17-18):2159–2174, 2011.
- [36] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [37] A. Jüngel. *Entropy methods for diffusive partial differential equations*. SpringerBriefs in Mathematics. Springer, [Cham], 2016.
- [38] A. Jüngel and S. Schuchnigg. Entropy-dissipating semi-discrete Runge-Kutta schemes for nonlinear diffusion equations. *Commun. Math. Sci.*, 15(1):27–53, 2017.
- [39] J. Leray and J. Schauder. Topologie et équations fonctionnelles. *Ann. Sci. École Norm. Sup. (3)*, 51:45–78, 1934.
- [40] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [41] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications 87. Birkhäuser Basel, 1 edition, 2015.
- [42] D.L. Scharfetter and H.K. Gummel. Large signal analysis of a silicon Read diode. *IEEE Trans. Elec. Dev.*, 16:64–77, 1969.
- [43] G. Toscani and C. Villani. On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds. *J. Statist. Phys.*, 98(5-6):1279–1309, 2000.

CLÉMENT CANCÈS ([clement.cances@inria.fr](mailto:clement.cances@inria.fr)). TEAM RAPSODI, INRIA LILLE – NORD EUROPE, 40 AV. HALLEY, F-59650 VILLENEUVE D’ASCQ, FRANCE.

CLAIRE CHAINAIS-HILLAIRET ([Claire.Chainais@math.univ-lille1.fr](mailto:Claire.Chainais@math.univ-lille1.fr)), UNIV. LILLE, CNRS, UMR 8524-LABORATOIRE PAUL PAINLEVÉ. F-59000 LILLE, FRANCE.

STELLA KRELL ([stella.krell@unice.fr](mailto:stella.krell@unice.fr)). UNIVERSITÉ DE NICE, CNRS, UMR7351-LABORATOIRE J.-A. DIEUDONNÉ. F-06100 NICE, FRANCE.