A CONVERGENT ENTROPY-DISSIPATING BDF2 FINITE-VOLUME SCHEME FOR A POPULATION CROSS-DIFFUSION SYSTEM

ANSGAR JÜNGEL AND MARTIN VETTER

Abstract. A second-order backward differentiation formula (BDF2) finite-volume discretization for a nonlinear cross-diffusion system arising in population dynamics is studied. The numerical scheme preserves the Rao entropy structure and conserves the mass. The existence and uniqueness of discrete solutions and their large-time behavior as well as the convergence of the scheme are proved. The proofs are based on the G-stability of the BDF2 scheme, which provides an inequality for the quadratic Rao entropy and hence suitable a priori estimates. The novelty is the extension of this inequality to the system case. Some numerical experiments in one and two space dimensions underline the theoretical results.

1. Introduction

The design of structure-preserving finite-volume schemes for parabolic equations is fundamental to describe accurately the behavior of the numerical solutions to these equations. In the literature, usually implicit Euler time discretization are used to derive such schemes; see, e.g., [2, 3, 6, 9, 26]. However, implicit Euler schemes are only first order accurate in time, while finite-volume implementations often show second-order accuracy in space [9, 27] (also see [17] for an analytical result). In order to match the convergence rates in space and time, there is the need to design second-order time approximations, which lead to structure-preserving and convergent schemes. Some works suggest higher-order time discretizations (e.g. [8, 15, 19, 24, 29]), but they are only concerned with semidiscrete equations or different numerical methods, or they do not contain any numerical analysis. In this paper, we propose a second-order BDF two-point flux approximation finite-volume scheme, which conserves the mass and dissipates the Rao entropy, for a nonlinear cross-diffusion system arising in population dynamics. The quadratic structure of the Rao entropy allows us to extend the G-stability theory of Dahlquist to the system case, leading to existence, uniqueness, and convergence results.

Date: January 9, 2023.

2000 *Mathematics Subject Classification.* 65L06, 65M08, 65M12, 35Q92, 92D25.

Key words and phrases. Cross-diffusion equations, Rao entropy, discrete entropy dissipation, linear multistep method, finite-volume method, population dynamics.

The authors acknowledge partial support from the Austrian Science Fund (FWF), grants P33010 and F65. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, ERC Advanced Grant no. 101018153.

The dynamics of the population density $u_i(x,t)$ of the ith species is modeled by the cross-diffusion equation

$$(1)\qquad \partial_t u_i = \operatorname{div}(\gamma\nabla u_i + u_i\nabla p_i(u)), \quad p_i(u) := \sum_{j=1}^n a_{ij}u_j \quad \text{in } \Omega,\ t>0,\ i=1,\ldots,n,$$

where $\Omega\subset\mathbb{R}^d$ ($d\geq 1$) is a bounded domain and $u=(u_1,\ldots,u_n)$. This model was derived rigorously from a moderately interacting stochastic particle system in a mean-field-type limit [11]. The parameter $\gamma>0$ is related to the stochastic diffusion of the particle system, and $a_{ij}\in\mathbb{R}$ describes the strength of the repulsive or attractive interaction between the ith and the jth species. We impose initial and no-flux boundary conditions,

$$(2)\qquad u_i(0)=u_i^0 \quad \text{in } \Omega, \quad \nabla u_i\cdot\nu = 0 \quad \text{on } \partial\Omega,\ t>0,\ i=1,\ldots,n,$$

where ν is the exterior unit normal vector to $\partial\Omega$. In the absence of the diffusion parameter γ, (1) can be interpreted as a mass conservation equation with the partial velocity $\nabla p_i(u)$, which is determined according to Darcy's law by the partial pressure $p_i(u)$. System (1) in one space dimension for two species, $\gamma=0$, and $\det(a_{ij})=0$ was first studied in [4], proving the global existence of segregated solutions (i.e., the supports of u_1 and u_2 do not intersect for all times if this holds true initially). This result was generalized to arbitrary space dimensions in [5], still for two species. For an arbitrary number of species, the existence of global weak solutions to (1)–(2) was shown in [25, Appendix B] if $\det(a_{ij})>0$ and the existence of local strong solutions was proved in [18] if $\det(a_{ij})=0$.

The matrix $A=(a_{ij})\in\mathbb{R}^{n\times n}$ does not need to be symmetric nor positive definite so that the diffusion matrix associated to (1) is generally neither symmetric nor positive definite too. A minimal requirement for local solvability at the linear level is the parabolicity in the sense of Petrovskii, which is satisfied if all eigenvalues of A have a positive real part [1]. Global solvability is guaranteed under the detailed-balance condition, i.e., there exist $\pi_1,\ldots,\pi_n>0$ such that $\pi_i a_{ij}=\pi_j a_{ji}$ for all $i\neq j$ [25, Theorem 17]. This condition also appears in the theory of time-continuous Markov chains generated by A, and $(\pi_1,\ldots,\pi_n)$ is the associated invariant measure. We assume this condition throughout this paper. It implies that $\widetilde{u}_i := \pi_i u_i$ solves the system

$$\partial_t\widetilde{u}_i = \operatorname{div}\left(\widetilde{u}_i\sum_{j=1}^n\frac{a_{ij}}{\pi_j}\nabla\widetilde{u}_j\right),$$

with a symmetric positive definite matrix (a_{ij}/π_j). Consequently, we may assume, without loss of generality, that the matrix A in (1) is already symmetric and positive definite.

Due to the nonlinear cross-diffusion structure, the analysis of (1) is highly nontrivial. The key idea of the analysis is to exploit the entropy structure of (1). This means that there exist Lyapunov functionals, called entropies, that are nonincreasing in time along solutions to (1)–(2) and that provide gradient estimates. In the present situation, these functionals are given by the Boltzmann (or Shannon) entropy H_B and the Rao entropy

H_R,

$$H_B(u)=\sum_{i=1}^n\int_\Omega u_i(\log u_i-1)\mathrm{d}x,\quad H_R(u)=\frac{1}{2}\int_\Omega u^TAu\mathrm{d}x,$$

giving formally the entropy equalities

$$\frac{\mathrm{d}H_B}{\mathrm{d}t}+\int_\Omega\bigg(4\gamma\sum_{i=1}^n|\nabla\sqrt{u_i}|^2+\sum_{i,j=1}^n a_{ij}\nabla u_i\cdot\nabla u_j\bigg)\mathrm{d}x=0, \tag{3}$$

$$\frac{\mathrm{d}H_R}{\mathrm{d}t}+\int_\Omega\bigg(\gamma\sum_{i,j=1}^n a_{ij}\nabla u_i\cdot\nabla u_j+\sum_{i=1}^n u_i|\nabla p_i(u)|^2\bigg)\mathrm{d}x=0, \tag{4}$$

and thus providing gradient bounds for u_i. The Boltzmann entropy is related to the thermodynamic entropy of the system, while the Rao entropy measures the functional diversity of the species [30].

Since the Boltzmann entropy H_R is convex, the implicit Euler scheme preserves the entropy inequality (3) (see, e.g., [27] for a related system). The logarithmic structure of H_R seems to prevent entropy stability in higher-order schemes like BDF or Crank–Nicolson approximations [22]. However, thanks to the quadratic structure of the Rao entropy H_R, we are able to prove stability of H_R for the BFD2 approximation. To explain the idea, let $\mathcal{T}$ be a triangulation of Ω into control volumes $K\subset\Omega$ with measure $\mathrm{m}(K)$ and let Δt be the time step size. Furthermore, let $u_{i,K}^k$ be an approximation of $u_i(x_K,t_k)$, where $x_K\in K$ and $t_k=k\Delta t$. We write the BDF2 discretization of (1) as

$$\frac{\mathrm{m}(K)}{\Delta t}\bigg(\frac{3}{2}u_{i,K}^k-2u_{i,K}^k+\frac{1}{2}u_{i,K}^{k-2}\bigg)+\sum_{\sigma\in\mathcal{E}_K}\mathcal{F}_{i,K,\sigma}^k=0, \tag{5}$$

where $\mathcal{E}_K$ is the set of the edges (or faces) of K and $\mathcal{F}_{i,K,\sigma}^k$ is the numerical flux, defined in (18) below. The usual idea to derive a priori bounds is to choose the test function $u_{i,K}^k$ in (5) and to use the inequality

$$\bigg(\frac{3}{2}u_{i,K}^k-2u_{i,K}^k+\frac{1}{2}u_{i,K}^{k-2}\bigg)u_{i,K}^k\ge h_0(u_{i,K}^k,u_{i,K}^{k-1})-h_0(u_{i,K}^{k-1},u_{i,K}^{k-2}), \tag{6}$$

where

$$h_0(a,b)=\frac{1}{4}\big(5a^2-4ab+b^2\big)=\frac{1}{4}\begin{pmatrix}a\\b\end{pmatrix}^T\begin{pmatrix}5&-2\\-2&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix},\quad a,b\in\mathbb{R},$$

is a positive definite quadratic form. Assuming that $\mathcal{F}_{i,K,\sigma}^k u_{i,K}^k$ can be bounded from below, this gives a priori bounds for $(u_{i,K}^k)^2$. Inequality (6) can be explained in the framework of Dahlquist's G-stability theory [23].

In our case, we need the test function $p_i(u_K^k)$ to derive the discrete analog of (4). Then the question is whether there exists a functional $h(u,v)$ such that

$$\sum_{i=1}^n\bigg(\frac{3}{2}u_{i,K}^k-2u_{i,K}^k+\frac{1}{2}u_{i,K}^{k-2}\bigg)p_i(u_K^k)\ge h(u_K^k,u_K^{k-1})-h(u_K^{k-1},u_K^{k-2}). \tag{7}$$

Note that we need to sum over all species in this inequality. The main novelty of this paper is the observation that the scalar inequality (6) can be extended to inequality (7) for vectors u, $v \in \mathbb{R}^n$. Indeed, we show in Lemma 7 that (7) holds for

$$h(u,v) = \frac{1}{4}(5u^TAu - 4u^TAv + v^TAv) = \frac{1}{4}\begin{pmatrix} u \\ v \end{pmatrix}^T \begin{pmatrix} 5A & -2A \\ -2A & A \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \tag{8}$$

with $u, v \in \mathbb{R}^n$. Introducing the discrete Rao entropy by $H(u,v) = \sum_{K\in\mathcal{T}} \mathrm{m}(K)h(u,v)$ for piecewise constant functions u and v, this yields the BDF2 analog of the Rao entropy inequality

$$H(u^k, u^{k-1}) + c\Delta t|u^k|^2_{1,2,\mathcal{T}} \le H(u^{k-1}, u^{k-2}) \quad \text{for } k \ge 2,$$

where $|\cdot|_{1,2,\mathcal{T}}$ is the discrete $H^1(\Omega)$ norm, defined in Section 2.3, and $c > 0$ depends on the smallest eigenvalue of A and on γ. This inequality is the key for proving our main results:

- Existence and uniqueness of discrete solutions: There exists a solution u_i^k to the BDF2 finite-volume scheme (5), which conserves the mass $\sum_{K\in\mathcal{T}} \mathrm{m}(K)u_{i,K}^k$ of the ith species and dissipates the discrete Rao entropy. Moreover, the solution is unique if $\Delta t/(\Delta x)^{d+2}$ is sufficiently small, where Δx is the size of the mesh (Theorem 3). This unusual quotient comes from an inverse inequality needed to bound higher-order norms.
- Large-time behavior: The discrete solution u_i^k converges for large times $k \to \infty$ to the constant steady state $\bar{u}_i = \mathrm{m}(\Omega)^{-1}\int_\Omega u_i^0 \mathrm{d}x$ with a quasi-explicit exponential rate (Theorem 4). The proof uses the well-established relative entropy (or energy) method, but the two-step scheme requires an iteration of this argument.
- Convergence of the discrete scheme: The fully discrete solution converges to a solution to the semidiscrete problem if $\Delta x \to 0$, and the semidiscrete solution converges to a weak (nonnegative) solution to (1)–(2) as $\Delta t \to 0$ (up to subsequences; see Theorem 5).
- Convergence rate: If the solution to (1)–(2) is sufficiently smooth, the semidiscrete solution converges with order two, as expected for the BDF2 scheme (Theorem 6).

The paper is organized as follows. The numerical scheme and our main results are detailed in Section 2. In Section 3, we prove the existence and uniqueness of a discrete solution, while its large-time behavior is analyzed in Section 4. Section 5 is devoted to the convergence of the full scheme, and the second-order convergence in time is verified in Section 6. Finally, we present in Section 7 some numerical examples in one and two space dimensions.

2. Numerical scheme and main results

We need some simple auxiliary results and some notation before formulating the numerical scheme and the main results.

2.1. Some linear algebra.

We denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^n$. Given a symmetric positive matrix $A \in \mathbb{R}^{n\times n}$, we introduce the weighted norm $|u|_A^2 := u^TAu$ and the weighted inner product $(u,v)_A := u^TAv$ for $u,v \in \mathbb{R}^n$. With this notation, the discrete Rao entropy density can be written as

$$h(u,v) = \frac{1}{4}(5|u|_A^2 - 4(u,v)_A + |v|_A^2) \quad \text{for } u,v \in \mathbb{R}^n. \tag{9}$$

Denoting by $\lambda_m > 0$ the smallest and by $\lambda_M > 0$ the largest eigenvalue of A, it holds that

$$\lambda_m|u|^2 \le |u|_A^2 \le \lambda_M|u|^2 \quad \text{for } u \in \mathbb{R}^n. \tag{10}$$

Let $\lambda_1, \ldots, \lambda_n > 0$ be the eigenvalues of A. Then the eigenvalues of the matrix in (8) equal $(3 \pm \sqrt{8})\lambda_i$ for $i = 1, \ldots, n$. This shows that for $u,v \in \mathbb{R}^n$,

$$\begin{aligned} \frac{1}{4}(3-\sqrt{8})(|u|_A^2 + |v|_A^2) \le h(u,v) \le \frac{1}{4}(3+\sqrt{8})(|u|_A^2 + |v|_A^2), \\ \frac{1}{4}(3-\sqrt{8})\lambda_m(|u|^2 + |v|^2) \le h(u,v) \le \frac{1}{4}(3+\sqrt{8})\lambda_M(|u|^2 + |v|^2). \end{aligned} \tag{11}$$

2.2. **Spatial domain and mesh.** Let $d \ge 1$ and let $\Omega \subset \mathbb{R}^d$ be a bounded polygonal (if $d = 2$) or polyhedral (if $d \ge 3$) domain. We associate to this domain an admissible mesh, given by (i) a family $\mathcal{T}$ of open polygonal or polyhedral control volumes, which are also called cells, (ii) a family $\mathcal{E}$ of edges (or faces if $d \ge 3$), and (iii) a family of points $(x_K)_{K\in\mathcal{T}}$ associated to the control volumes and satisfying [21, Definition 9.1]. This definition implies that the straight line $\overline{x_Kx_L}$ between two centers of neighboring cells is orthogonal to the edge (or face) $\sigma = K|L$ between two cells. For instance, triangular meshes with acute angles, Delaunay meshes, rectangular meshes, and Voronoï meshes satisfy this condition [21, Example 9.2]. The size of the mesh is given by $\Delta x = \max_{K\in\mathcal{T}} \operatorname{diam}(K)$. The family of edges $\mathcal{E}$ is assumed to consist of interior edges $\mathcal{E}_{\rm int}$ satisfying $\sigma \subset \Omega$ and boundary edges $\sigma \in \mathcal{E}_{\rm ext}$ satisfying $\sigma \subset \partial\Omega$. For a given $K \in \mathcal{T}$, $\mathcal{E}_K$ denotes the set of edges of K with $\mathcal{E}_K = \mathcal{E}_{{\rm int},K} \cup \mathcal{E}_{{\rm ext},K}$. For any $\sigma \in \mathcal{E}$, there exists at least one cell $K \in \mathcal{T}$ such that $\sigma \in \mathcal{E}_K$.

For given $\sigma \in \mathcal{E}$, we define the distance

$$\mathrm{d}_\sigma = \begin{cases} \mathrm{d}(x_K, x_L) & \text{if } \sigma = K|L \in \mathcal{E}_{{\rm int},K}, \\ \mathrm{d}(x_K, \sigma) & \text{if } \sigma \in \mathcal{E}_{{\rm ext},K}, \end{cases}$$

where d is the Euclidean distance in $\mathbb{R}^d$, and the transmissibility coefficient

$$\tau_\sigma = \frac{\mathrm{m}(\sigma)}{\mathrm{d}_\sigma}, \tag{12}$$

where $\mathrm{m}(\sigma)$ denotes the $(d-1)$-dimensional Lebesgue measure of σ. We suppose the following mesh regularity condition: There exists $0 < \zeta \le 1/2$ such that for all $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$,

$$\mathrm{d}(x_K, \sigma) \ge \zeta \mathrm{d}_\sigma. \tag{13}$$

This is equivalent to

$$\eta \le \frac{\mathrm{d}(x_K,\sigma)}{\mathrm{d}(x_L,\sigma)} \le \frac{1}{\eta} \quad \text{for all } \sigma = K|L,$$

where $\eta = \zeta/(1-\zeta) \in (0,1]$. The statement follows by observing that $\mathrm{d}(x_K,\sigma)+\mathrm{d}(x_L,\sigma) = \mathrm{d}(x_K,x_L)$ holds, which is a consequence of the orthogonality of $\sigma = K|L$ and $\overline{x_K x_L}$. Hence, the mesh regularity (13) means that the mesh is locally quasi-uniform. A consequence of the mesh regularity is the following estimate

$$\sum_{\sigma\in\mathcal{E}_{\mathrm{int},K}} \mathrm{m}(\sigma)\mathrm{d}_\sigma \le \frac{1}{\zeta}\sum_{\sigma\in\mathcal{E}_{\mathrm{int},K}} \mathrm{m}(\sigma)\mathrm{d}(x_K,\sigma) = \frac{d}{\zeta}\mathrm{m}(K) \quad \text{for } K\in\mathcal{T}, \tag{14}$$

where we used in the last step the formula for the volume of a (hyper-)pyramid.

2.3. **Function spaces.** Given a triangulation $\mathcal{T}$, let $T>0$, $N_T\in\mathbb{N}$ and introduce the time step size $\Delta t = T/N_T$ and the time steps $t_k = k\Delta t$ for $k=0,\ldots,N_T$. We set $\Omega_T = \Omega\times(0,T)$. The space of piecewise constant functions is defined by

$$V_\mathcal{T} = \bigg\{ v:\Omega\to\mathbb{R} : \exists (v_K)_{K\in\mathcal{T}}\subset\mathbb{R},\ v(x) = \sum_{K\in\mathcal{T}} v_K \mathbf{1}_K(x)\bigg\},$$

where $\mathbf{1}_K$ is the indicator function on K. To define a norm on this space, we define for $K\in\mathcal{T}$, $\sigma\in\mathcal{E}_K$,

$$v_{K,\sigma} = \begin{cases} v_L & \text{if } \sigma = K|L\in\mathcal{E}_{\mathrm{int},K}, \\ v_K & \text{if } \sigma\in\mathcal{E}_{\mathrm{ext},K}, \end{cases} \quad \mathrm{D}_{K,\sigma}v := v_{K,\sigma}-v_K, \quad \mathrm{D}_\sigma v := |\mathrm{D}_{K,\sigma}v|.$$

Let $1\le q<\infty$ and $v\in V_\mathcal{T}$. The discrete $W^{1,q}(\Omega)$ norm on $V_\mathcal{T}$ is given by

$$\begin{aligned} &\|v\|_{1,q,\mathcal{T}} = \big(\|v\|_{0,q,\mathcal{T}}^q + |v|_{1,q,\mathcal{T}}^q\big)^{1/q}, \quad \text{where} \\ &\|v\|_{0,q,\mathcal{T}}^q = \sum_{K\in\mathcal{T}} \mathrm{m}(K)|v_K|^q, \quad |v|_{1,q,\mathcal{T}}^q = \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \mathrm{m}(\sigma)\mathrm{d}_\sigma\bigg|\frac{\mathrm{D}_\sigma v}{\mathrm{d}_\sigma}\bigg|^q \quad \text{for } v\in V_\mathcal{T}. \end{aligned}$$

When $q=\infty$, we define $|v|_{1,\infty,\mathcal{T}} = \max_{\sigma\in\mathcal{E}_{\mathrm{int}}}|\mathrm{D}_\sigma v|/\mathrm{d}_\sigma$. If $v=(v_1,\ldots,v_n)\in V_\mathcal{T}^n$ is a vector-valued function, we write for notational convenience

$$\|v\|_{0,q,\mathcal{T}} = \sum_{i=1}^n \|v_i\|_{0,q,\mathcal{T}}, \quad \|\nabla v\|_{0,q,\mathcal{T}} = \sum_{i=1}^n \|\nabla v_i\|_{0,q,\mathcal{T}}.$$

We associate to the discrete $W^{1,q}$ norm a dual norm with respect to the L^2 inner product:

$$\|v\|_{-1,q,\mathcal{T}} = \sup\bigg\{\int_\Omega vw\mathrm{d}x : w\in V_\mathcal{T},\ \|w\|_{1,q,\mathcal{T}} = 1\bigg\}.$$

Finally, we introduce the space $V_{\mathcal{T},\Delta t}$ of piecewise constant functions with values in $V_\mathcal{T}$,

$$V_{\mathcal{T},\Delta t} = \bigg\{ v:\Omega_T\to\mathbb{R} : \exists(v^k)_{k=1,\ldots,N_T}\subset V_\mathcal{T},\ v(x,t) = \sum_{k=1}^{N_T} v^k(x)\mathbf{1}_{[t_{k-1},t_k)}(t)\bigg\},$$

equipped with the $L^2(0,T;H^1(\Omega))$ norm

$$\|v\|_{L^2(0,T;H^1(\Omega))} = \left(\sum_{k=1}^{N_T} \Delta t \|v^k\|_{1,2,\mathcal{T}}^2\right)^{1/2} \quad \text{for all } v \in V_{\mathcal{T},\Delta t}.$$

2.4. **Discrete gradient.** The discrete gradient is defined on a dual mesh. For this, we define the cell $T_{K,\sigma}$ of the dual mesh for $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$:

- "Diamond": Let $\sigma = K|L \in \mathcal{E}_{\text{int},K}$. Then $T_{K,\sigma}$ is that cell whose vertices are given by x_K, x_L, and the end points of the edge σ. In higher dimensions, they might be (double) (hyper-)pyramids.
- "Triangle": Let $\sigma \in \mathcal{E}_{\text{ext},K}$. Then $T_{K,\sigma}$ is that cell whose vertices are given by x_K and the end points of the edge σ.

The union of all "diamonds" and "triangles" $T_{K,\sigma}$ equals the domain Ω (up to a set of measure zero). The property that the straight line $\overline{x_K x_L}$ is orthogonal to the edge $\sigma = K|L$ implies that

$$\mathrm{m}(\sigma)\mathrm{d}(x_K, x_L) = d\,\mathrm{m}(T_{K,\sigma}) \quad \text{for all } \sigma = K|L \in \mathcal{E}_{\text{int}}.$$

The approximate gradient of $v \in V_{\mathcal{T},\Delta t}$ is then defined by

$$\nabla^{\mathcal{T}} v(x,t) = \frac{\mathrm{m}(\sigma)}{\mathrm{m}(T_{K,\sigma})}\mathrm{D}_{K,\sigma}(v^k)\nu_{K,\sigma} \quad \text{for } x \in T_{K,\sigma},\ t \in (t_{k-1}, t_k],$$

where $\nu_{K,\sigma}$ is the unit vector that is normal to σ and points outwards of K.

2.5. **Numerical scheme.** The initial functions are approximated by their $L^2(\Omega)$-orthogonal projection on $V_{\mathcal{T}}$:

$$u_{i,K}^0 = \frac{1}{\mathrm{m}(K)}\int_K u_i^0(x)\mathrm{d}x \quad \text{for all } K \in \mathcal{T},\ i = 0,\ldots,n. \tag{15}$$

Let $u_K^{k-1} = (u_{1,K}^{k-1},\ldots,u_{n,K}^{k-1})$ for $K \in \mathcal{T}$ be given. Since the BDF2 scheme is a two-step method, we need a first time step which is computed from the implicit Euler method. The following time steps are determined from the BDF2 method. The finite-volume scheme reads as

$$\frac{\mathrm{m}(K)}{\Delta t}(u_{i,K}^1 - u_{i,K}^0) + \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{i,K,\sigma}^1 = 0, \tag{16}$$

$$\frac{\mathrm{m}(K)}{\Delta t}\left(\frac{3}{2}u_{i,K}^k - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\right) + \sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{i,K,\sigma}^k = 0, \quad k \geq 2, \tag{17}$$

for $i = 1,\ldots,n$, $K \in \mathcal{T}$, and the numerical fluxes are given by

$$\mathcal{F}_{i,K,\sigma}^k = -\tau_\sigma\big(\gamma \mathrm{D}_{K,\sigma}u_i^k + (u_{i,\sigma}^k)^+ \mathrm{D}_{K,\sigma}p_i(u^k)\big), \tag{18}$$

where τ_σ is defined in (12) and $z^+ = \max\{0,z\}$ denotes the positive part of $z \in \mathbb{R}$. Finally, the so-called mobility is given by

$$u_{i,\sigma}^k = M(u_{i,K}^k, u_{i,L}^k) \quad \text{for } \sigma = K|L, \quad u_{i,\sigma}^k = 0 \quad \text{else}, \tag{19}$$

where M is a general mean function satisfying

(i) $M : [0,\infty)^2 \to [0,\infty)$ is Lipschitz continuous, satisfies $M(u,u) = u$ (consistency), and has linear growth in the sense $M(u,v) \le |u| + |v|$ for $u,v \ge 0$.

(ii) There exists $C > 0$ such that $|M(u,v) - u| \le C|u-v|$ for all $u,v \ge 0$.

Examples for M are $M(u,v) = (u+v)/2$ or $M(u,v) = \max\{u,v\}$. Note that we do not need logarithmic mean functions like in [27], since we do not use the chain rule in the cross-diffusion part, so that we can use simpler expressions.

Remark 1 (Nonnegativity)**.** We truncate the mobility by $(u^k_{i,\sigma})^+$ in the numerical flux (18) to ensure the discrete Rao entropy inequality (see (21) below). Indeed, when testing (17) with $p_i(u^k)$, we need that the sum $\sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma (u^k_{i,\sigma})^+ |\mathrm{D}_{K,\sigma} p_i(u^k)|^2$ is nonnegative. Unfortunately, the quadratic Rao entropy does not allow us to prove the nonnegativity of the discrete solution, and standard maximum principle arguments do not apply here, so that the truncation cannot be removed. A positivity-preserving BDF2 finite-difference scheme was proposed in [13], but the proof relies on discrete $L^\infty(\Omega)$ bounds for u^{k-1}, which are not available in our case. Also the Shannon entropy does not help (as in [27]), since it is not compatible with the BDF2 discretization. Indeed, when we wish to derive a discrete analog of (3), we need a finite continuous functional $h(u,v)$ satisfying $h(u,u) = \sum_{i=1}^n u_i(\log u_i - 1)$ (consistency condition) such that

$$\sum_{i=1}^n \left(\frac{3}{2}u^k_{i,K} - 2u^{k-1}_{i,K} + \frac{1}{2}u^{k-2}_{i,K}\right)\log u^k_{i,K} \ge h(u^k_K, u^{k-1}_K) - h(u^{k-1}_K, u^{k-2}_K).$$

If $u^k_{i,K} = u^{k-1}_{i,K} \to 0$ and $u^{k-2}_{i,K} > 0$ for all $i \in \{1,\ldots,n\}$, the previous inequality converges to $-\infty \ge -h(0, u^{k-2}_K)$, which is absurd. At least, we obtain nonnegative solutions in the limit $(\Delta x, \Delta t) \to 0$; see Theorem 5 below.

Remark 2 (Discrete integration by parts)**.** The fluxes $\mathcal{F}^k_{i,K,\sigma}$ are consistent approximations of the exact fluxes through the edges if we impose the conservation $\mathcal{F}_{i,K,\sigma} + \mathcal{F}_{i,L,\sigma} = 0$ for all edges $\sigma = K|L$, requiring that they vanish on the Neumann boundary edges, i.e., $\mathcal{F}_{i,K,\sigma} = 0$ for all $\sigma \in \mathcal{E}_{{\rm ext},K}$. In particular, for $v = (v_K) \in V_{\mathcal{T}}$, the following discrete integration-by-parts formulas hold:

$$\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \mathcal{F}_{i,K,\sigma} v_K = -\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \mathcal{F}_{i,K,\sigma}\mathrm{D}_{K,\sigma} v, \quad \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma(\mathrm{D}_{K,\sigma} v) v_K = -|v|^2_{1,2,\mathcal{T}}. \tag{20}$$

2.6. **Main results.** We impose the following hypotheses.

(H1) Data: $\Omega \subset \mathbb{R}^d$ with $d \ge 1$ is a bounded polygonal ($d = 2$) or polyhedral ($d \ge 3$) domain, $T > 0$, and $u^0 \in L^2(\Omega;\mathbb{R}^n)$. We set $\Omega_T = \Omega \times (0,T)$.

(H2) Discretization: $\mathcal{T}$ is an admissible discretization of Ω satisfying (13) and $t_k = k\Delta t$ for $k = 1,\ldots,N_T$.

(H3) Coefficients: Let $\gamma > 0$, and $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ is symmetric and positive definite. Let $\lambda_m > 0$ and $\lambda_M > 0$ be the smallest and largest eigenvalue of A, respectively.

The positivity of γ is not needed for the existence analysis but for the convergence result, where we need higher-order integrability that is deduced via the discrete Gagliardo–Nirenberg inequality from the gradient bound. As mentioned in the introduction, the symmetry and positive definiteness of A can be replaced by the positivity of the real parts of the eigenvalues of A and the detailed-balance condition.

Recall the discrete BDF2 Rao entropy (see (9))

$$H(u,v)=\sum_{K\in\mathcal{T}}\mathrm{m}(K)h(u_K,v_K)=\frac{1}{4}\sum_{K\in\mathcal{T}}\mathrm{m}(K)\big(5|u_K|_A^2-4(u_K,v_K)_A+|v_K|_A^2\big)$$

for $u,v\in V_{\mathcal{T}}$. If $u=v$, this expression reduces to the usual discrete Rao entropy, used for the implicit Euler scheme, $H(u):=H(u,u)=\frac{1}{2}\sum_{K\in\mathcal{T}}\mathrm{m}(K)|u_K|_A^2$. Our first result is the existence of a discrete solution.

Theorem 3 (Existence and uniqueness of discrete solutions)**.** *Let Hypotheses (H1)–(H3) hold, let $k\in\mathbb{N}$, and let $u^{k-1}\in V_{\mathcal{T}}^n$ be given. Then there exists a solution $u^k=(u_1^k,\ldots,u_n^k)\in V_{\mathcal{T}}^n$ to scheme* (15)–(19) *satisfying the discrete entropy inequality*

$$H(u^k,u^{k-1})+\gamma\Delta t|A^{1/2}u^k|_{1,2,\mathcal{T}}^2\le H(u^{k-1},u^{k-2})\quad\text{for } k\ge 2, \tag{21}$$

$$H(u^1)+\gamma\Delta t|A^{1/2}u^1|_{1,2,\mathcal{T}}^2\le H(u^0), \tag{22}$$

and the scheme preserves the mass, $\sum_{K\in\mathcal{T}}\mathrm{m}(K)u_{i,K}^k=\int_\Omega u_i^0(x)\mathrm{d}x$ for $i=1,\ldots,n$, $k\ge 1$. These results also hold if $\gamma=0$. Furthermore, the solution is unique if $\gamma>0$, $\min_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{d}_\sigma\ge\xi\Delta x$ for some $\xi>0$, and

$$\frac{\Delta t}{(\Delta x)^{d+2}}<\frac{C(d,\xi,\zeta)\gamma\lambda_m^2}{\lambda_M^2L^2H(u^0)},$$

where ζ is defined in (13) *and L is the Lipschitz constant of the mean function M, defined in* (19)*.*

The existence of a discrete solution is proved by a fixed-point argument using the Brouwer degree theorem. Uniform estimates are obtained from the discrete Rao entropy inequality (21), where the BDF2 time approximation is estimated according to (7). This inequality, which is the key of our analysis, is proved in Lemma 7.

The uniqueness of solutions is proved by using the relative entropy method, which is equivalent to the energy method in the present case, since the Rao entropy is quadratic. In other words, we use the test function $p_i(u^k)-p_i(v^k)$ in the difference of the equations (17) satisfied by two discrete solutions u^k and v^k. The cross-diffusion part contains cubic expressions, which turn into quadratic ones if $|A^{1/2}v^k|_{1,\infty,\mathcal{T}}$ is bounded (similar as in [12]). By an inverse inequality, this norm is bounded, up to some factor, by $(\Delta x)^{-d/2-1}\|A^{1/2}v^k\|_{0,2,\mathcal{T}}$, and $\|A^{1/2}v^k\|_{0,2,\mathcal{T}}$ is bounded because of (21)–(22). The remaining quadratic expression is estimated by using the gradient bounds (which requires $\gamma>0$) and the discrete $L^2(\Omega)$ bound coming from the time discretization (and introducing the factor Δt). The condition $\min_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{d}_\sigma\le\xi\Delta x$ is discussed in Remark 8.

For the next result, we set $\bar{u}_i=\mathrm{m}(\Omega)^{-1}\int_\Omega u_i^0\mathrm{d}x$ and recall the discrete Poincaré–Wirtinger inequality $\|v-\bar{v}\|_{0,2,\mathcal{T}}\le C_P\zeta^{-1/2}|v|_{1,2,\mathcal{T}}$ for $v\in V_{\mathcal{T}}$ [7, Theorem 3.6]. Then, in view of (10),

$$\|A^{1/2}(v-\bar{v})\|_{0,2,\mathcal{T}} \le C_P\left(\frac{\lambda_M}{\lambda_m\zeta}\right)^{1/2}|A^{1/2}v|_{1,2,\mathcal{T}} \quad \text{for } v\in V_{\mathcal{T}}. \tag{23}$$

Theorem 4 (Large-time behavior)**.** *Let u^k be a solution to scheme* (15)–(19)*. Then, for $k\ge 2$,*

$$\|A^{1/2}(u^k-\bar{u})\|_{0,2,\mathcal{T}} \le \sqrt{2}\|A^{1/2}(u^0-\bar{u})\|_{0,2,\mathcal{T}}(1+\kappa\Delta t)^{-(k-2)/4},$$

where $\kappa = 4\gamma\lambda_m\zeta/((3+\sqrt{8})C_P^2\lambda_M)$ and $C_P>0$ is the constant of the Poincaré–Wirtinger inequality (23)*.*

The theorem states that u^k converges exponentially fast to the constant steady state $\bar{u}$. Indeed, setting $\lambda_{\Delta t} := \log(1+\kappa\Delta t)/(\Delta t) \nearrow \kappa$ as $\Delta t\to 0$, we have

$$\|A^{1/2}(u^k-\bar{u})\|_{0,2,\mathcal{T}} \le \sqrt{2}\|A^{1/2}(u^0-\bar{u})\|_{0,2,\mathcal{T}}\exp(-\lambda_{\Delta t}t_k), \quad k\ge 2.$$

The proof of Theorem 4 is based on the discrete entropy inequality for the discrete relative Rao entropy $H(u^k-\bar{u},u^{k-1}-\bar{u})$, similar to (21). Indeed, by the discrete Poincaré–Wirtinger inequality, the discrete gradient term is bounded from below by the discrete $L^2(\Omega)$ norm of $u^k-\bar{u}$. As $H(u^k-\bar{u},u^{k-1}-\bar{u})$ can be estimated in terms of the discrete $L^2(\Omega)$ norms of $u^k-\bar{u}$ and $u^{k-1}-\bar{u}$, we need to iterate the entropy inequality a second time. Then, using (11), we arrive at the inequality

$$H(u^k-\bar{u},u^{k-1}-\bar{u}) \le (1+\kappa\Delta t)^{-1}H(u^{k-2}-\bar{u},u^{k-3}-\bar{u}),$$

and solving this recursion shows the result.

The numerial convergence of the scheme is proved in two steps. First, we show that the fully discrete solution $u_m^k\in V_{\mathcal{T}}^n$, indexed with the space grid size $\Delta x_m\to 0$ as $m\to\infty$, converges, up to a subsequence, to a solution $u^k\in H^1(\Omega)$ to the semidiscrete system

$$\begin{aligned}\frac{1}{\Delta t}(u_i^1-u_i^0) &= \operatorname{div}(\gamma_i\nabla u_i^1 + u_i^1\nabla p_i(u^1)),\\ \frac{1}{\Delta t}\left(\frac{3}{2}u_i^k - 2u_i^{k-1} + \frac{1}{2}u_i^{k-2}\right) &= \operatorname{div}(\gamma_i\nabla u_i^k + (u_i^k)^+\nabla p_i(u^k)) \quad \text{in } \Omega,\end{aligned}$$

with no-flux boundary conditions $\nabla u_i^k\cdot\nu = 0$ on $\partial\Omega$, $i=1,\ldots,n$. Second, we prove that a subsequence of the sequence of semidiscrete solutions converges to a weak solution to (1)–(2) as $\Delta t\to 0$. Both steps may be summarized as follows (the precise convergence statements can be found in Propositions 9 and 11).

Theorem 5 (Convergence of the scheme)**.** *Let Hypotheses (H1)–(H3) hold and let $(\mathcal{T}_m)_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of Ω satisfying* (13) *uniformly in m and $\Delta x_m\to 0$, $\Delta t_m\to 0$ as $m\to\infty$. Then the solution (u_m) to* (15)–(19)*, constructed in Theorem 3, converges, up to a subsequence, as $m\to\infty$ to a function $u=(u_1,\ldots,u_n)$ satisfying $u_i\ge 0$ in Ω_T for $i=1,\ldots,n$, $u_i\in L^2(0,T;H^1(\Omega))$, $\partial_t u_i\in L^{2d+4}(0,T;W^{1,2d+4}(\Omega)')$, and u is a weak solution to* (1)–(2)*.*

The proof is based on suitable estimates uniform with respect to Δx_m and Δt_m, derived from the entropy inequality (21). For the limit $\Delta x_m \to 0$, we follow the strategy of [10]. The compactness argument is different, since we still keep the time discretization. The limit $\Delta t_m \to 0$ is based on a higher-order integrability property derived from the Gagliardo–Nirenberg inequality and on the Aubin–Lions compactness lemma in the version of [16].

We need the condition $\gamma > 0$ since the application of the discrete Gagliardo–Nirenberg inequality requires discrete gradient bounds. However, the term involving $p_i(u)$ only provides a bound for the discrete kinetic energy $\sum_{\sigma\in\mathcal{E}_{\rm int}} \tau_\sigma (u^k_{i,\sigma})^+ |{\rm D}_{K,\sigma} p_i(u^k)|^2$, from which we are unable to conclude gradient bounds. For the Euler scheme, this issue can be overcome by using the Boltzmann entropy inequality, which provides bounds in $L^2(0,T;H^1(\Omega))$ and $L^\infty(0,T;L^1(\Omega))$ (see (3)), and consequently in $L^{2+2/d}(\Omega_T)$, which is the required higher-order integrability bound. As mentioned in the introduction, this entropy is not compatible with the BDF2 discretization. Therefore, the restriction $\gamma > 0$ seems to be unavoidable with our approach.

Finally, we verify that the convergence of the semidiscrete system is of second order.

Theorem 6 (Second-order convergence)**.** *Let u^k be a solution to* (31) *and assume that the solution to* (1)–(2) *satisfies $u \in C^3([0,T];L^2(\Omega)) \cap L^\infty(0,T;W^{1,\infty}(\Omega))$. Furthermore, let $\varepsilon > 0$ be arbitrary and assume that*

$$\Delta t < \frac{4(3-\sqrt{8})\gamma\lambda_m}{\lambda_M^3 \|\nabla u\|^2_{L^\infty(\Omega_T)} + 4\gamma\lambda_m\varepsilon}.$$

Then there exists $C(\varepsilon) > 0$, which is of order $\varepsilon^{-1/2}$ as $\varepsilon \to 0$ but independent of Δt, such that

$$\max_{k=1,\ldots,N_T} \|A^{1/2}(u_i^k - u_i(t_k))\|_{L^2(\Omega)} \le C(\varepsilon)(\Delta t)^2 \quad \textit{for } i = 1,\ldots,n.$$

We allow for the parameter $\varepsilon > 0$ to minimize the time step size constraint; however, optimizing this constraint gives large constants $C(\varepsilon)$. The theorem is proved by analyzing the relative entropy $H(u(t_k) - u^k, u(t_{k-1}) - u^{k-1})$, using a Taylor expansion for u_i up to order $(\Delta t)^3$ (which requires a bound for $\partial_t^3 u_i$), and iterating the entropy inequality once more. The resulting recursive inequality for the relative entropy can be solved, leading to the desired second-order bound.

3. Proof of Theorem 3

First, we make precise inequality (7). Recall definition (8) of $h(u,v)$ and let $H(u,v) = \sum_{K\in\mathcal{T}} {\rm m}(K) h(u,v)$ be the discrete Rao entropy.

Lemma 7 (BDF2 inequality)**.** *It holds for $u,v,w \in \mathbb{R}^n$ that*

$$\left(\frac{3}{2}u - 2v + \frac{1}{2}w\right)^T Au = h(u,v) - h(v,w) + \frac{1}{4}|u - 2v + w|_A^2.$$

In particular, for u^k, u^{k-1}, $u^{k-2} \in V_{\mathcal{T}}^n$,

$$\sum_{i,j=1}^n \sum_{K\in\mathcal{T}} \mathrm{m}(K)\bigg(\frac{3}{2}u_{i,K}^k - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\bigg)a_{ij}u_{j,K}^k \geq H(u^k,u^{k-1}) - H(u^{k-1},u^{k-2}).$$

Proof. The proof follows by a direct computation. □

3.1. **Definition and continuity of the fixed-point operator.** We assume that $k \geq 2$, since the existence of a solution $u^1 \in V_{\mathcal{T}}^n$ to the Euler scheme (17) satisfying (22) follows from [26, Theorem 1]. Let $u^{k-1} \in V_{\mathcal{T}}^n$ be given and let $R > 0$, $\delta > 0$. We set

$$Z_R = \big\{w = (w_1,\ldots,w_n) \in V_{\mathcal{T}}^n : \|w_i\|_{1,2,\mathcal{T}} < R \text{ for } i = 1,\ldots,n\big\},$$

and let $w \in Z_R$. We consider the linear regularized problem

$$\varepsilon\bigg(\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma \mathrm{D}_{K,\sigma}w_i^\varepsilon - \mathrm{m}(K)w_{i,K}^\varepsilon\bigg) = \frac{\mathrm{m}(K)}{\Delta t}\bigg(\frac{3}{2}w_{i,K} - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\bigg) + \sum_{\sigma\in\mathcal{E}_K}\mathcal{F}_{i,K,\sigma}^+(w) \tag{24}$$

for $i = 1,\ldots,n$, $K \in \mathcal{T}$, where

$$\mathcal{F}_{i,K,\sigma}^+(w) = -\tau_\sigma\big(\gamma \mathrm{D}_{K,\sigma}w_i + w_{i,\sigma}^+ \mathrm{D}_{K,\sigma}p_i(w)\big).$$

The ε-regularization guarantees the coercivity of the associated bilinear form, while the truncation $w_{i,\sigma}^+$ is needed to obtain the nonnegativity of the entropy dissipation (see the estimate of I_6 below).

We claim that (24) has a unique solution $w^\varepsilon \in V_{\mathcal{T}}^n$. Indeed, since the mapping $g(w^\varepsilon) = \varepsilon(\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma \mathrm{D}_{K,\sigma}w_i^\varepsilon - \mathrm{m}(K)w_{i,K}^\varepsilon)$ is linear and acting on finite-dimensional spaces, we only need to verify its injectivity. Let w^ε be in the kernel of this mapping. Multiplying $g(w^\varepsilon) = 0$ by $w_{i,K}^\varepsilon$, summing over $K \in \mathcal{T}$, and using the discrete integration-by-parts formula (20) gives

$$0 = \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma(\mathrm{D}_{K,\sigma}w_i^\varepsilon)w_{i,K}^\varepsilon - \sum_{K\in\mathcal{T}}\mathrm{m}(K)(w_{i,K}^\varepsilon)^2 = -\|w_i^\varepsilon\|_{1,2,\mathcal{T}}^2.$$

This yields $w^\varepsilon = 0$ and proves the claim.

Next, we show that the fixed-point mapping $F : Z_R \to V_{\mathcal{T}}^n$, $F(w) = w^\varepsilon$, is continuous. For this, we multiply (24) by $-w_{i,K}^\varepsilon$, sum over $K \in \mathcal{T}$, and use discrete integration by parts and the Cauchy–Schwarz inequality:

$$\begin{aligned}\varepsilon\|w_i^\varepsilon\|_{1,2,\mathcal{T}}^2 &= -\frac{1}{\Delta t}\sum_{K\in\mathcal{T}}\mathrm{m}(K)\bigg(\frac{3}{2}w_{i,K} - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\bigg)w_{i,K}^\varepsilon + \sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\mathcal{F}_{i,K,\sigma}^+(w)\mathrm{D}_{K,\sigma}w_i^\varepsilon \\ &\leq \frac{1}{\Delta t}\bigg\|\frac{3}{2}w_i - 2u_i^{k-1} + \frac{1}{2}u_i^{k-2}\bigg\|_{0,2,\mathcal{T}}\|w_i^\varepsilon\|_{0,2,\mathcal{T}} + \gamma|w_i|_{1,2,\mathcal{T}}|w_i^\varepsilon|_{1,2,\mathcal{T}} \\ &\quad - \sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma(w_{i,\sigma})^+\mathrm{D}_{K,\sigma}p_i(w)\mathrm{D}_{K,\sigma}w_i^\varepsilon.\end{aligned} \tag{25}$$

For the last term, we use the Cauchy–Schwarz inequality and the fact that any norm on $V_{\mathcal{T}}^n$ is equivalent:

$$
\begin{aligned}
-\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma(w_{i,\sigma})^+\mathrm{D}_{K,\sigma}p_i(w)\mathrm{D}_{K,\sigma}w_i^\varepsilon &= -\sum_{j=1}^n\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma a_{ij}w_{i,\sigma}^+\mathrm{D}_{K,\sigma}w_i\mathrm{D}_{K,\sigma}w_i^\varepsilon \\
&\le \sum_{j=1}^n\bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma|\mathrm{D}_\sigma w_i^\varepsilon|^2\bigg)^{1/2}\bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma a_{ij}^2(w_{i,\sigma}^+)^2|\mathrm{D}_\sigma w_j|^2\bigg)^{1/2} \\
&\le C(A)\|w\|_{0,\infty,\mathcal{T}}\sum_{j=1}^n|w_i^\varepsilon|_{1,2,\mathcal{T}}|w_j|_{1,2,\mathcal{T}} \le C(A,R)\|w_i^\varepsilon\|_{1,2,\mathcal{T}},
\end{aligned}
$$

where we took into account the linear growth of $w_{i,\sigma}^+$ with respect to $w_{i,K}$ and $w_{i,L}$ (see (19)) and the definition of Z_R. Inserting these estimates into (25) and dividing by $\|w_i^\varepsilon\|_{1,2,\mathcal{T}}$, it follows that $\varepsilon\|w_i^\varepsilon\|_{1,2,\mathcal{T}} \le C(A,R)$.

This bound allows us to verify the continuity of F. Indeed, let $w^\ell \to w$ as $\ell\to\infty$ and set $w^{\varepsilon,\ell} = F(w^\ell)$. Then $(w^{\varepsilon,\ell})_{\ell\in\mathbb{N}}$ is uniformly bounded in the discrete $H^1(\Omega)$ norm. Therefore, there exists a subsequence, which is not relabeled, such that $w^{\varepsilon,\ell}\to w^\varepsilon$ as $\ell\to\infty$. Passing to the limit $\ell\to\infty$ in scheme (24), we see that w^ε is a solution of the scheme and consequently $w^\varepsilon = F(w)$. Since the solution to the linear scheme (24) is unique, the entire sequence $(w^{\varepsilon,\ell})_{\ell\in\mathbb{N}}$ converges to w^ε, which shows the continuity of F.

3.2. **Existence of a fixed point.** According to the Brouwer degree fixed-point theorem, it is sufficient to show that for all $(w^\varepsilon,\rho)\in\overline{Z}_R\times[0,1]$ such that $w^\varepsilon = \rho F(w^\varepsilon)$, it holds that $w^\varepsilon\notin\partial Z_R$ or, equivalently, $\|w^\varepsilon\|_{1,2,\mathcal{T}} < R$. We claim that this is true for sufficiently large $R>0$. Indeed, let w^ε be such a fixed point. It satisfies

$$
\begin{aligned}
&\varepsilon\bigg(\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\mathrm{D}_{K,\sigma}w_i^\varepsilon - \mathrm{m}(K)w_{i,K}^\varepsilon\bigg) \\
&\qquad = \frac{\rho}{\Delta t}\,\mathrm{m}(K)\bigg(\frac{3}{2}w_{i,K}^\varepsilon - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\bigg) + \rho\sum_{\sigma\in\mathcal{E}_K}\mathcal{F}_{i,K,\sigma}^+(w^\varepsilon).
\end{aligned}
$$

We multiply this equation by $-(\Delta t)p_i(w^\varepsilon)$ and sum over $i=1,\ldots,n$, $K\in\mathcal{T}$. Then $0 = I_1+I_2+I_3$, where

$$
\begin{aligned}
I_1 &= -\varepsilon\Delta t\sum_{i,j=1}^n\sum_{K\in\mathcal{T}}\bigg(\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\mathrm{D}_{K,\sigma}w_i^\varepsilon - \mathrm{m}(K)w_{i,K}^\varepsilon\bigg)a_{ij}w_{j,K}^\varepsilon, \\
I_2 &= \rho\sum_{i,j=1}^n\sum_{K\in\mathcal{T}}a_{ij}\bigg(\frac{3}{2}w_{i,K}^\varepsilon - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\bigg)w_{j,K}^\varepsilon,
\end{aligned}
$$

$$I_3 = -\rho\Delta t \sum_{i=1}^n \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \mathcal{F}^+_{i,K,\sigma}(w^\varepsilon){\rm D}_{K,\sigma}p_i(w^\varepsilon).$$

By discrete integration by parts,

$$\begin{aligned} I_1 &= \varepsilon\Delta t \sum_{i,j=1}^n \bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma a_{ij}{\rm D}_{K,\sigma}w_i^\varepsilon {\rm D}_{K,\sigma}w_j^\varepsilon + \sum_{K\in\mathcal{T}} {\rm m}(K)a_{ij}w^\varepsilon_{i,K}w^\varepsilon_{j,K}\bigg) \\ &\geq \varepsilon\lambda_m\Delta t(|w^\varepsilon|^2_{1,2,\mathcal{T}} + \|w^\varepsilon\|^2_{0,2,\mathcal{T}}) = \varepsilon\lambda_m\Delta t\|w^\varepsilon\|^2_{1,2,\mathcal{T}}, \end{aligned}$$

and by Lemma 7,

$$I_2 \geq H(w^\varepsilon, u^{k-1}) - H(u^{k-1}, u^{k-2}).$$

For the third term, we obtain

$$\begin{aligned} I_3 &= \rho\Delta t \sum_{i,j=1}^n \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma\gamma a_{ij}{\rm D}_{K,\sigma}w_i^\varepsilon{\rm D}_{K,\sigma}w_j^\varepsilon + \rho\Delta t\sum_{i=1}^n \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma(w^\varepsilon_{i,\sigma})^+\bigg(\sum_{j=1}^n a_{ij}{\rm D}_{K,\sigma}w_j^\varepsilon\bigg)^2 \\ &\geq \gamma\rho\Delta t \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma|A^{1/2}{\rm D}_{K,\sigma}w^\varepsilon|^2 = \gamma\rho\Delta t|A^{1/2}w^\varepsilon|^2_{1,2,\mathcal{T}}. \end{aligned}$$

Collecting these estimates gives

$$\varepsilon\Delta t\|w^\varepsilon\|^2_{1,2,\mathcal{T}} + H(w^\varepsilon, u^{k-1}) + \gamma\Delta t\rho|A^{1/2}w^\varepsilon|^2_{1,2,\mathcal{T}} \leq H(u^{k-1}, u^{k-2}). \tag{26}$$

Setting $R = (\varepsilon\Delta t)^{-1/2}H(u^{k-1}, u^{k-2})^{1/2} + 1$, we infer that $\|w^\varepsilon\|^2_{1,2,\mathcal{T}} \leq (R-1)^2 < R^2$ and thus $w^\varepsilon \notin \partial Z_R$, which shows the claim. Hence, there exists a fixed point w^ε to F, which is a solution to

$$\varepsilon\bigg(\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma{\rm D}_{K,\sigma}w_i^\varepsilon - {\rm m}(K)w^\varepsilon_{i,K}\bigg) = \frac{{\rm m}(K)}{\Delta t}\bigg(\frac{3}{2}w^\varepsilon_{i,K} - 2u_{i,K} + \frac{1}{2}u^{k-2}_{i,K}\bigg) + \sum_{\sigma\in\mathcal{E}_K}\mathcal{F}_{i,K,\sigma}(w^\varepsilon). \tag{27}$$

3.3. **Limit $\varepsilon \to 0$.** The solution w^ε to (27) satisfies the regularized entropy inequality (26) with $\rho = 1$, and the right-hand side is independent of ε and M. It follows from the Bolzano–Weierstraß theorem that there exists a subsequence of w^ε, which is not relabeled, such that $w^\varepsilon \to w$ as $\varepsilon \to 0$. In particular, $\varepsilon^{1/2}w^\varepsilon \to 0$. Since the problem is finite dimensional, we can pass to the limit $\varepsilon \to 0$ in (27). Consequently, $u^k := w$ is a solution to (17)–(19). The same limit in (26) with $\rho = 1$ leads to the discrete entropy inequality of Theorem 3, which finishes the proof.

3.4. **Uniqueness of solutions.** Let $u^k, v^k \in V^n_{\mathcal{T}}$ be two solutions to (15)–(19) with the same initial data $u^0 = v^0$. We take the difference of the equations satisfied by u^k and v^k, multiply the resulting equation by $p_i(u^k_K) - p_i(v^k_K) = \sum_{j=1}^n a_{ij}(u^k_{j,K} - v^k_{j,K})$, sum over

$i=1,\ldots,n$, $K\in\mathcal{T}$, and use discrete integration by parts. This leads to $0=I_4+I_5+I_6$, where

$$\begin{aligned} I_4 &= \frac{3}{2\Delta t}\sum_{i,j=1}^n\sum_{K\in\mathcal{T}}\mathrm{m}(K)a_{ij}(u_{i,K}^k-v_{i,K}^k)(u_{j,K}^k-v_{j,K}^k) \\ I_5 &= \sum_{i,j=1}^n\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma\gamma a_{ij}\mathrm{D}_{K,\sigma}(u_i^k-v_i^k)\mathrm{D}_{K,\sigma}(u_j^k-v_j^k) \\ I_6 &= \sum_{i,j,\ell=1}^n\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma a_{ij}\big((u_{i,\sigma}^k)^+\mathrm{D}_{K,\sigma}u_j^k-(v_{i,\sigma}^k)^+\mathrm{D}_{K,\sigma}v_j^k\big)a_{i\ell}\mathrm{D}_{K,\sigma}(u_\ell^k-v_\ell^k). \end{aligned}$$

By the definition of the weighted norm, $I_4=(3/(2\Delta t))\|A^{1/2}(u^k-v^k)\|_{0,2,\mathcal{T}}^2$. Furthermore,

$$I_5\ge\gamma\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma|(A^{1/2}\mathrm{D}_{K,\sigma}(u^k-v^k))|^2=\gamma|A^{1/2}(u^k-v^k)|_{1,2,\mathcal{T}}^2.$$

We add and subtract the term $(u_{i,\sigma}^k)^+\mathrm{D}_{K,\sigma}v_j^k$ in I_6 and apply the Cauchy–Schwarz inequality:

$$\begin{aligned} I_6 &= \sum_{i,j,\ell=1}^n\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma a_{ij}a_{i\ell}\Big((u_{i,\sigma}^k)^+\mathrm{D}_{K,\sigma}(u_j^k-v_j^k) \\ &\quad+\big((u_{i,\sigma}^k)^+-(v_{i,\sigma}^k)^+\big)\mathrm{D}_{K,\sigma}v_j^k\Big)\mathrm{D}_{K,\sigma}(u_\ell^k-v_\ell^k) \\ &= \sum_{i=1}^n\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma(u_{i,\sigma}^k)^+\bigg(\sum_{j=1}^n a_{ij}\mathrm{D}_{K,\sigma}(u_j^k-v_j^k)\bigg)\bigg(\sum_{\ell=1}^n a_{i\ell}\mathrm{D}_{K,\sigma}(u_\ell^k-v_\ell^k)\bigg) \\ &\quad-\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma(A\mathrm{D}_{K,\sigma}v^k)^T\big[\operatorname{diag}\big((u_{i,\sigma}^k)^+-(v_{i,\sigma}^k)^+\big)A^{1/2}\big](A^{1/2}\mathrm{D}_{K,\sigma}(u^k-v^k)) \\ &\ge-\bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma|A\mathrm{D}_{K,\sigma}v^k|^2\big|\operatorname{diag}\big((u_{i,\sigma}^k)^+-(v_{i,\sigma}^k)^+\big)A^{1/2}\big|^2\bigg)^{1/2} \\ &\quad\times\bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=K|L}}\tau_\sigma|A^{1/2}\mathrm{D}_{K,\sigma}(u^k-v^k)|^2\bigg)^{1/2}, \end{aligned}$$

where $\operatorname{diag}((u_{i,\sigma}^k)^+-(v_{i,\sigma}^k)^+)$ denotes the diagonal matrix with the entries $(u_{i,\sigma}^k)^+-(v_{i,\sigma}^k)^+$ for $i=1,\ldots,n$. Together with

$$|A\mathrm{D}_{K,\sigma}v^k|\le|A^{1/2}||A^{1/2}\mathrm{D}_{K,\sigma}v^k|\le\lambda_M^{1/2}|\mathrm{D}_{K,\sigma}v^k|_A\quad\text{and}$$

$$\begin{aligned}\big|\operatorname{diag}\big((u_{i,\sigma}^k)^+ - (v_{i,\sigma}^k)^+\big)A^{1/2}\big| &\le \big|\operatorname{diag}\big((u_{i,\sigma}^k)^+ - (v_{i,\sigma}^k)^+\big)\big||A^{1/2}| \\ &\le \lambda_M^{1/2}\max_{i=1,\ldots,n}|(u_{i,\sigma}^k)^+ - (v_{i,\sigma}^k)^+| \le \lambda_M^{1/2}\max_{i=1,\ldots,n}|u_{i,\sigma}^k - v_{i,\sigma}^k|,\end{aligned}$$

we find that

$$I_6 \ge -\lambda_M|A^{1/2}v^k|_{1,\infty,\mathcal{T}}\max_{i=1,\ldots,n}\bigg(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\mathrm{m}(\sigma)\mathrm{d}_\sigma|u_{i,\sigma}^k - v_{i,\sigma}^k|^2\bigg)^{1/2}|A^{1/2}(u^k - v^k)|_{1,2,\mathcal{T}}.$$

It remains to estimate the term involving the difference $|u_{i,\sigma}^k - v_{i,\sigma}^k|$. By the Lipschitz continuity of the mean function $M(u_{i,K}^k, u_{i,L}^k) = u_{i,\sigma}^k$ with Lipschitz constant $L > 0$ and the mesh regularity (14),

$$\begin{aligned}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\mathrm{m}(\sigma)\mathrm{d}_\sigma|u_{i,\sigma}^k - v_{i,\sigma}^k|^2 &= \frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{{\rm int},K}}\mathrm{m}(\sigma)\mathrm{d}_\sigma|u_{i,\sigma}^k - v_{i,\sigma}^k|^2 \\ &\le \frac{L^2}{2}\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\mathrm{m}(\sigma)\mathrm{d}_\sigma\big(|u_{i,K}^k - v_{i,K}^k| + |u_{i,L}^k - v_{i,L}^k|\big)^2 \\ &\le 2L^2\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\mathrm{m}(\sigma)\mathrm{d}_\sigma|u_{i,K}^k - v_{i,K}^k|^2 \le \frac{2dL^2}{\zeta}\sum_{K\in\mathcal{T}}\mathrm{m}(K)|u_{i,K}^k - v_{i,K}^k|^2 \\ &= \frac{2dL^2}{\zeta}\|u^k - v^k\|_{0,2,\mathcal{T}}^2 \le \frac{2dL^2}{\lambda_m\zeta}\|A^{1/2}(u^k - v^k)\|_{0,2,\mathcal{T}}^2,\end{aligned}$$

This shows that

$$I_6 \ge -\frac{\lambda_M L}{\lambda_m^{1/2}}\bigg(\frac{2d}{\zeta}\bigg)^{1/2}|A^{1/2}v^k|_{1,\infty,\mathcal{T}}\|A^{1/2}(u^k - v^k)\|_{0,2,\mathcal{T}}|A^{1/2}(u^k - v^k)|_{1,2,\mathcal{T}}.$$

Collecting the estimates for I_4, I_5, and I_6 and using Young's inequality gives

$$\begin{aligned}(28)\qquad &\frac{3}{2\Delta t}\|A^{1/2}(u^k - v^k)\|_{0,2,\mathcal{T}}^2 + \gamma|A^{1/2}(u^k - v^k)|_{1,2,\mathcal{T}}^2 \\ &\le \frac{\lambda_M L}{\lambda_m^{1/2}}\bigg(\frac{2d}{\zeta}\bigg)^{1/2}|A^{1/2}v^k|_{1,\infty,\mathcal{T}}\|A^{1/2}(u^k - v^k)\|_{0,2,\mathcal{T}}|A^{1/2}(u^k - v^k)|_{1,2,\mathcal{T}} \\ &\le \frac{3}{2\Delta t}\|A^{1/2}(u^k - v^k)\|_{0,2,\mathcal{T}}^2 + \frac{\Delta t}{3}\frac{d\lambda_M^2L^2}{\lambda_m\zeta}|A^{1/2}v^k|_{1,\infty,\mathcal{T}}^2|A^{1/2}(u^k - v^k)|_{1,2,\mathcal{T}}^2.\end{aligned}$$

Now, the inverse inequality $|A^{1/2}v^k|_{1,\infty,\mathcal{T}} \le C'(d)(\Delta x)^{-d/2}\zeta^{-1/2}|A^{1/2}v^k|_{1,2,\mathcal{T}}$ [14, Prop. 3.10] and condition $\mathrm{d}_\sigma \ge \xi\Delta x$ imply that

$$|A^{1/2}v^k|_{1,\infty,\mathcal{T}}^2 \le \frac{C'(d)^2}{(\Delta x)^d\zeta}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\frac{\mathrm{m}(\sigma)}{\mathrm{d}_\sigma}|\mathrm{D}_{K,\sigma}(A^{1/2}v^k)|^2$$

$$\leq \frac{C'(d)^2}{(\Delta x)^d\zeta}\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\frac{\mathrm{m}(\sigma)\mathrm{d}_\sigma}{\xi^2(\Delta x)^2}|\mathrm{D}_{K,\sigma}(A^{1/2}v^k)|^2.$$

It follows from (14) and $|\mathrm{D}_{K,\sigma}(A^{1/2}v^k)|^2 \leq 2(|v_K^k|_A^2 + |v_L^k|_A^2)$ that

$$\begin{aligned}|A^{1/2}v^k|_{1,\infty,\mathcal{T}}^2 &\leq \frac{C'(d)^2}{(\Delta x)^{d+2}\xi^2\zeta}\sum_{K\in\mathcal{T}}\frac{d}{\zeta}\,\mathrm{m}(K)|\mathrm{D}_{K,\sigma}(A^{1/2}v^k)|^2 \\ &\leq \frac{2dC'(d)^2}{(\Delta x)^{d+2}(\xi\zeta)^2}\sum_{K\in\mathcal{T}}\mathrm{m}(K)|A^{1/2}v_K^k|^2 = \frac{C(d,\xi,\zeta)}{(\Delta x)^{d+2}}\|A^{1/2}v^k\|_{0,2,\mathcal{T}}^2.\end{aligned}$$

Using this inequality as well as the bound

$$(3-\sqrt{8})\|A^{1/2}v^k\|_{0,2,\mathcal{T}}^2 \leq 4H(v^1,v^0) \leq 2(3+\sqrt{8})(H(v^1)+H(v^0)) \leq 4(3+\sqrt{8})H(u^0),$$

obtained from (21)–(22), we deduce from (28), for another constant $C(d,\xi,\zeta)$ that

$$\gamma|A^{1/2}(u^k-v^k)|_{1,2,\mathcal{T}}^2 \leq C(d,\xi,\zeta)\frac{\lambda_M^2L^2}{\lambda_m}\frac{\Delta t}{(\Delta x)^{d+2}}H(u^0)|A^{1/2}(u^k-v^k)|_{1,2,\mathcal{T}}^2.$$

Then our smallness condition on $\Delta t/(\Delta x)^{d+2}$ implies that $|A^{1/2}(u^k-v^k)|_{1,2,\mathcal{T}} = 0$ and consequently, $u^k = v^k$, finishing the proof.

Remark 8. The quasi-uniform condition $\min_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{d}_\sigma \geq \xi\Delta x > 0$ implies condition (23) in [20], since the mesh regularity (13) gives $\min_{K\in\mathcal{T}}\min_{\sigma\in\mathcal{E}_K}\mathrm{d}(x_K,\sigma)/\operatorname{diam}(K) \geq \zeta\mathrm{d}_\sigma/\Delta x \geq \zeta\xi > 0$. It also implies the mesh regularity condition $\operatorname{diam}(K)/\mathrm{d}(x_K,\sigma) \leq \xi_0$ in [20, (9)], since, because of (13) again, $\operatorname{diam}(K)/\mathrm{d}(x_K,\sigma) \leq \Delta x/(\zeta\mathrm{d}_\sigma) \leq 1/(\xi\zeta) =: \xi_0$. It can be seen by considering quadratic cells that the quasi-uniform condition $\min_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{d}_\sigma \geq \xi\Delta x > 0$ generally does not imply the mesh regularity condition (13) and vice versa, so both conditions are independent from each other.

4. Proof of Theorem 4

We infer from mass conservation, $\sum_{K\in\mathcal{T}}\mathrm{m}(K)u_{i,K}^k = \sum_{K\in\mathcal{T}}\mathrm{m}(K)u_{i,K}^0 = \mathrm{m}(\Omega)\bar u_i$, that

$$\begin{aligned}H(u^k-\bar u, u^{k-1}-\bar u) &= H(u^k,u^{k-1}) + \frac12\sum_{K\in\mathcal{T}}\mathrm{m}(K)\big(|\bar u|_A^2 - 3(u_K^k,\bar u)_A + (u_K^{k-1},\bar u)_A\big) \\ &= H(u^k,u^{k-1}) - \frac12\,\mathrm{m}(\Omega)|\bar u|_A^2.\end{aligned}$$

Then the entropy inequality (21) shows that

$$H(u^k-\bar u,u^{k-1}-\bar u) + \gamma\Delta t|A^{1/2}u^k|_{1,2,\mathcal{T}}^2 \leq H(u^{k-1}-\bar u,u^{k-2}-\bar u) \tag{29}$$

for $k\geq 2$. Another iteration gives, for $k\geq 3$,

$$H(u^k-\bar u,u^{k-1}-\bar u) + \gamma\Delta t\big(|A^{1/2}u^k|_{1,2,\mathcal{T}}^2 + |A^{1/2}u^{k-1}|_{1,2,\mathcal{T}}^2\big) \leq H(u^{k-2}-\bar u,u^{k-3}-\bar u).$$

Hence, taking into account the discrete Poincaré–Wirtinger inequality (23),

$$\begin{aligned} H(u^k-\bar u,u^{k-1}-\bar u)&+\frac{\gamma\lambda_m\zeta}{C_P^2\lambda_M}\Delta t\big(\|A^{1/2}(u^k-\bar u)\|_{0,2,\mathcal T}^2+\|A^{1/2}(u^{k-1}-\bar u)\|_{0,2,\mathcal T}^2\big)\\ &\le H(u^{k-2}-\bar u,u^{k-3}-\bar u),\end{aligned}$$

and the norm equivalence (10),

$$H(u^k-\bar u,u^{k-1}-\bar u)+\frac{4\gamma\lambda_m\zeta\Delta t}{(3+\sqrt{8})C_P^2\lambda_M}H(u^k-\bar u,u^{k-1}-\bar u)\le H(u^{k-2}-\bar u,u^{k-3}-\bar u).$$

This can be written as

$$H(u^k-\bar u,u^{k-1}-\bar u)\le(1+\kappa\Delta t)^{-1}H(u^{k-2}-\bar u,u^{k-3}-\bar u),$$

where $\kappa=4\gamma\lambda_m\zeta/((3+\sqrt{8})C_P^2\lambda_M)$. Depending on whether k is odd or even, we resolve this iteration as follows:

$$\begin{aligned} H(u^{2\ell+1}-\bar u,u^{2\ell}-\bar u)&\le(1+\kappa\Delta t)^{-\ell}H(u^1-\bar u,u^0-\bar u),\\ H(u^{2\ell+2}-\bar u,u^{2\ell+1}-\bar u)&\le(1+\kappa\Delta t)^{-\ell}H(u^2-\bar u,u^1-\bar u)\\ &\le(1+\kappa\Delta t)^{-\ell}H(u^1-\bar u,u^0-\bar u),\end{aligned}$$

where we used (29) in the last step. As in both cases $\ell\ge(k-2)/2$, we conclude that

$$H(u^k-\bar u,u^{k-1}-\bar u)\le(1+\kappa\Delta t)^{-(k-2)/2}H(u^1-\bar u,u^0-\bar u). \tag{30}$$

We want to express this inequality in terms of the $\|A^{1/2}(\cdot)\|_{0,2,\mathcal T}$ norm. We observe that, by Young's inequality, $\|A^{1/2}(u^k-\bar u)\|_{0,2,\mathcal T}^2\le 4H(u^k-\bar u,u^{k-1}-\bar u)$ and, in view of (22),

$$\begin{aligned} H(u^1-\bar u,u^0-\bar u)&=H(u^1)-H(\bar u)\le H(u^0)-H(\bar u)\\ &=H(u^0-\bar u)=\frac12\|A^{1/2}(u^0-\bar u)\|_{0,2,\mathcal T}^2.\end{aligned}$$

Then we deduce from (30) that

$$\begin{aligned} \|A^{1/2}(u^k-\bar u)\|_{0,2,\mathcal T}^2&\le 4H(u^k-\bar u,u^{k-1}-\bar u)\le 4(1+\kappa\Delta t)^{-(k-2)/2}H(u^1-\bar u,u^0-\bar u)\\ &\le 2(1+\kappa\Delta t)^{-(k-2)/2}\|A^{1/2}(u^0-\bar u)\|_{0,2,\mathcal T}^2,\end{aligned}$$

which concludes the proof.

5. Proof of Theorem 5

We split the proof into two parts. We first prove the convergence in the space variable and then the convergence in the time variable. An alternative is to show the convergence in both variables simultaneously; see, e.g., [27].

5.1. **Convergence in space.** We show the following result for $\Delta x \to 0$.

Proposition 9 (Convergence in space)**.** *Let the assumptions of Theorem 5 hold and let* (u_m^k) *be the sequence of solutions to* (15)–(19) *constructed in Theorem 3 associated to an admissible mesh* $\mathcal{T}_m$ *with mesh size* Δx_m *for* $m \in \mathbb{N}$ *satisfying* $\Delta x_m \to 0$ *as* $m \to \infty$*. Then there exists a subsequence which is not relabeled such that* $u_{i,m}^k \to u_i^k$ *strongly in* $L^2(\Omega)$ *as* $m \to \infty$ *and* u_i^k *solves for all* $\phi_i \in W^{1,\max\{2,d\}}(\Omega)$, $i = 1, \ldots, n$,

$$\frac{1}{\Delta t}\int_\Omega \left(\frac{3}{2}u_i^k - 2u_i^{k-1} + \frac{1}{2}u_i^{k-2}\right)\phi_i \mathrm{d}x + \int_\Omega \big(\gamma \nabla u_i^k + (u_i^k)^+ \nabla p_i(u^k)\big)\cdot \nabla \phi_i \mathrm{d}x = 0. \tag{31}$$

Proof. For fixed Δt, the discrete entropy inequality in Theorem 3 provides a uniform bound for $\|u_m^k\|_{1,2,\mathcal{T}_m}$. Then, by the discrete Rellich–Kondrachov compactness theorem [21, Lemma 5.6], there exists a subsequence of $(u_m^k) = (u_{1,m}^k, \ldots, u_{n,m}^k)$, which is not relabeled, such that $u_m^k \to u^k$ strongly in $L^2(\Omega)$ as $m \to \infty$. Moreover, the sequence of discrete gradients $(\nabla^m u_m^k)$ converges weakly in $L^2(\Omega)$ to some function which can be identified by ∇u^k; see [10, Lemma 4.4]. Let $\phi_i \in C^2(\overline{\Omega})$ and set $\phi_{i,K} := \phi_i(x_K)$ for $K \in \mathcal{T}$. Then the limit $\Delta x_m \to 0$ in the BDF2 approximation becomes

$$\frac{1}{\Delta t}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\frac{3}{2}u_{i,K}^k - 2u_{i,K}^{k-1} + \frac{1}{2}u_{i,K}^{k-2}\right)\phi_{i,K} \to \frac{1}{\Delta t}\int_\Omega \left(\frac{3}{2}u_i^k - 2u_i^{k-1} + \frac{1}{2}u_i^k\right)\phi_i \mathrm{d}x.$$

Next, we set $F^m = F_1^m + F_2^m + F_3^m$, where

$$\begin{aligned}
F_1^m &= -\gamma \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \mathrm{D}_{K,\sigma} u_{i,m}^k \phi_{i,K}, \\
F_2^m &= -\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma (u_{i,m,K}^k)^+ \mathrm{D}_{K,\sigma} p_i(u_m^k)\phi_{i,K}, \\
F_3^m &= -\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \big((u_{i,m,\sigma}^k)^+ - (u_{i,m,K}^k)^+\big) \mathrm{D}_{K,\sigma} p_i(u_m^k)\phi_{i,K}.
\end{aligned}$$

We introduce the intermediate integral $F_0^m = F_{01}^m + F_{02}^m$, where

$$F_{01}^m = \gamma \int_\Omega \nabla^m u_{m,i}^k \cdot \nabla\phi_i \mathrm{d}x, \quad F_{02}^m = \int_\Omega (u_{i,m}^k)^+ \nabla^m p_i(u_m^k)\cdot\nabla\phi_i \mathrm{d}x.$$

It follows from the weak convergence of the discrete gradients and the strong convergence in $L^2(\Omega)$ that $F_0^m \to F$ as $m \to \infty$, where

$$F = \gamma \int_\Omega \nabla u_i^k \cdot \nabla\phi_i \mathrm{d}x + \int_\Omega (u_i^k)^+ \nabla p_i(u^k)\cdot\nabla\phi_i \mathrm{d}x.$$

Thus, if we can show that $F_0^m - F^m \to 0$, then $|F^m - F| \le |F^m - F_0^m| + |F_0^m - F| \to 0$, proving the claim.

By discrete integration by parts and the definition of the discrete gradient,

$$F_1^m = \gamma \sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}} \\ \sigma=K|L}} \tau_\sigma \mathrm{D}_{K,\sigma} u_{i,m}^k \mathrm{D}_{K,\sigma}\phi_i,$$

$$F_{01}^m = \gamma \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{\mathrm{m}(\sigma)}{\mathrm{m}(T_{K,\sigma})}\mathrm{D}_{K,\sigma}u_{i,m}^k \int_{T_{K,\sigma}} \nabla\phi_i\cdot\nu_{K,\sigma}\mathrm{d}x.$$

Using the Taylor expansion (here we need $\phi_i\in C^2(\overline{\Omega})$)

$$\frac{\mathrm{D}_{K,\sigma}\phi_i}{\mathrm{d}_\sigma} = \frac{\phi_{i,L}-\phi_{i,K}}{\mathrm{d}(x_K,x_L)} = \nabla\phi_i\cdot\nu_{K,\sigma} + \mathcal{O}(\Delta x_m) \quad \text{for } \sigma=K|L,$$

where we have taken into account the property $x_K - x_L = \mathrm{d}(x_K,x_L)\nu_{K,\sigma}$, we obtain

$$\begin{aligned} |F_{01}^m - F_1^m| &\le \gamma \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \mathrm{m}(\sigma)|\mathrm{D}_{K,\sigma}u_{i,m}^k|\left|\frac{1}{\mathrm{m}(T_{K,\sigma})}\int_{T_{K,\sigma}}\nabla\phi_i\cdot\nu_{K,\sigma}\mathrm{d}x - \frac{\mathrm{D}_{K,\sigma}\phi_i}{\mathrm{d}_\sigma}\right| \\ &\le C\gamma\Delta x_m \sum_{\sigma\in\mathcal{E}_{\rm int}} \mathrm{m}(\sigma)|\mathrm{D}_\sigma u_{i,m}^k|, \end{aligned}$$

where $C>0$ depends on the L^∞ norm of $\mathrm{D}^2\phi_i$. We apply the Cauchy–Schwarz inequality and use the mesh property (14) to find that

$$\begin{aligned} |F_{01}^m - F_1^m| &\le C\gamma\Delta x_m\bigg(\sum_{\sigma\in\mathcal{E}_{\rm int}}\frac{\mathrm{m}(\sigma)}{\mathrm{d}_\sigma}|\mathrm{D}_\sigma u_{i,m}^k|^2\bigg)^{1/2}\bigg(\sum_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{m}(\sigma)\mathrm{d}_\sigma\bigg)^{1/2} \\ &\le C\gamma\Delta x_m|u_{i,m}^k|_{1,2,\mathcal{T}_m}\bigg(\frac{d}{\zeta}\mathrm{m}(\Omega)\bigg)^{1/2} \to 0 \quad \text{as } m\to\infty. \end{aligned}$$

Similar arguments lead to

$$\begin{aligned} |F_{02}^m - F_2^m| &\le C\Delta x_m \sum_{K\in\mathcal{T}_m}\sum_{\sigma\in\mathcal{E}_{{\rm int},K}} \mathrm{m}(\sigma)(u_{i,m,K}^k)^+|\mathrm{D}_{K,\sigma}p_i(u_m^k)| \\ &\le C\Delta x_m\bigg(\sum_{K\in\mathcal{T}_m}|(u_{i,m,K}^k)^+|^2\sum_{\sigma\in\mathcal{E}_{{\rm int},K}}\mathrm{m}(\sigma)\mathrm{d}_\sigma\bigg)^{1/2}|p_i(u_m^k)|_{1,2,\mathcal{T}_m} \\ &\le C\Delta x_m\bigg(\frac{d}{\zeta}\sum_{K\in\mathcal{T}_m}\mathrm{m}(K)|(u_{i,m,K}^k)^+|^2\bigg)^{1/2}|p_i(u_m^k)|_{1,2,\mathcal{T}_m} \\ &\le C(\zeta)\Delta x_m\|u_{i,m}^k\|_{0,2,\mathcal{T}_m}|p_i(u_m^k)|_{1,2,\mathcal{T}_m}. \end{aligned}$$

The right-hand side converges to zero since

$$|p_i(u_m^k)|_{1,2,\mathcal{T}_m}^2 = \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma\bigg(\sum_{j=1}^n a_{ij}\mathrm{D}_{K,\sigma}u_{j,m}^k\bigg)^2 \le C(A)|u_m^k|_{1,2,\mathcal{T}_m}^2 \le C.$$

Finally, using $|\mathrm{D}_{K,\sigma}\phi_i|\le C(\phi_i)\Delta x_m$ and property (ii) of the mean function,

$$|F_3^m| \le \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}}\tau_\sigma|u_{i,m,\sigma}^k - u_{i,m,K}^k||\mathrm{D}_{K,\sigma}p_i(u_m^k)||\mathrm{D}_{K,\sigma}\phi_i|$$

$$\begin{aligned}&\le C(\phi_i)\Delta x_m \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \tau_\sigma |\mathrm{D}_\sigma u_{i,m}^k||\mathrm{D}_{K,\sigma}p_i(u_m^k)| \\ &\le C(\phi_i,A)\Delta x_m\bigg(\sum_{\sigma\in\mathcal{E}}\tau_\sigma|\mathrm{D}_\sigma u_{i,m}^k|^2\bigg)^{1/2}\bigg(\sum_{j=1}^n\sum_{\sigma\in\mathcal{E}}\tau_\sigma|\mathrm{D}_\sigma u_{j,m}^k|^2\bigg)^{1/2}\to 0.\end{aligned}$$

This shows that $F_0^m - F \to 0$ as $m\to\infty$, concluding the proof. □

5.2. **Convergence in time.** We wish to perform the limit $\Delta t\to 0$ in (31). For this, we need an estimate in a better space than $L^2(\Omega_T)$, provided by the following lemma.

Lemma 10 (Higher-order integrability). *Let $(u^{(\tau)})$ be a family of solutions to* (31) *associated to the time step size $\tau := \Delta t$, constructed in Proposition 9. Then there exists $C>0$ independent of τ such that*

$$\|u^{(\tau)}\|_{L^p(\Omega_T)} \le C \quad \textit{for } p = 2+4/d.$$

Proof. The lemma follows from the discrete entropy inequalities (21)–(22) and the Gagliardo–Nirenberg inequality. Indeed, we infer from the entropy inequalities after summation over $k=2,\ldots,N_T$ that

$$\|u^{(\tau)}\|_{L^\infty(0,T;L^2(\Omega))} + \|u^{(\tau)}\|_{L^2(0,T;H^1(\Omega))} \le C.$$

Then it follows from the Gagliardo–Nirenberg inequality with $\theta = d/2 - d/p$ that

$$\begin{aligned}\|u^{(\tau)}\|^p_{L^p(0,T;L^p(\Omega))} &\le C\int_0^T \|u^{(\tau)}\|^{p\theta}_{H^1(\Omega)}\|u^{(\tau)}\|^{p(1-\theta)}_{L^2(\Omega)}\mathrm{d}t \\ &\le C\|u^{(\tau)}\|^{p(1-\theta)}_{L^\infty(0,T;L^2(\Omega))}\int_0^T\|u^{(\tau)}\|^2_{H^1(\Omega)}\mathrm{d}t \le C,\end{aligned}$$

since $p\theta = 2$. This finishes the proof. □

Proposition 11 (Convergence in time). *Let $(u^{(\tau)})$ be a family of solutions to* (31) *with $\tau=\Delta t$. Then $u^{(\tau)}$ converges to a weak solution u to* (1)–(2) *satisfying*

$$u_i \in L^2(0,T;H^1(\Omega))\cap L^\infty(0,T;L^2(\Omega)), \quad \partial_t u_i \in L^r(0,T;W^{1,2d+4}(\Omega)'),$$

where $r = (2d+4)/(2d+3) > 1$.

Proof. We estimate the discrete time derivative $\mathrm{D}_\tau u_i^{(\tau)}(t) := \frac32 u_i^k - 2u_i^{k-1} + \frac12 u_i^{k-2}$ for $t\in[k\tau,(k+1)\tau)$ for $k\ge 2$. Let $\phi_i \in L^{2d+4}(0,T;W^{1,2d+4}(\Omega))$. Then

$$\begin{aligned}&\frac1\tau\int_{2\tau}^T \big|\langle \mathrm{D}_\tau u_i^{(\tau)},\phi_i\rangle_{W^{1,d+2}(\Omega)'}\big|^r\mathrm{d}t \\ &\quad\le \gamma^r C\int_{2\tau}^T\int_\Omega |\nabla u_i^{(\tau)}\cdot\nabla\phi_i|^r\mathrm{d}x\mathrm{d}t + C\int_{2\tau}^T\int_\Omega |(u_i^{(\tau)})^+\nabla p_i(u^{(\tau)})\cdot\nabla\phi_i|^r\mathrm{d}x\mathrm{d}t \\ &\quad\le \gamma^r C\|\nabla u_i^{(\tau)}\|^r_{L^2(\Omega_T)}\|\nabla\phi_i\|^r_{L^{2d+4}(\Omega_T)} \\ &\qquad + C\|u_i^{(\tau)}\|^r_{L^{(2d+4)/d}(\Omega_T)}\|\nabla p_i(u^{(\tau)})\|^r_{L^2(\Omega_T)}\|\nabla\phi_i\|^r_{L^{2d+4}(\Omega_T)}\end{aligned}$$

$$\le C\|\phi_i\|^r_{L^{2d+4}(0,T;W^{1,2d+4}(\Omega))},$$

where we used the fact that $p_i(u^{(\tau)})$ is a linear combination of all $u_j^{(\tau)}$ for $j=1,\ldots,n$. This implies the bound $\tau^{-1}\|\mathrm{D}_\tau u_i^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}\le C$.

Let $\pi_\tau u^{(\tau)}(t)=u^{(\tau)}(t-\tau)$ be a shift operator. We relate the implicit Euler scheme and the BDF2 scheme by

$$u_i^k-u_i^{k-1}=\frac{2}{3}\bigg(\frac{3}{2}u_i^k-2u_i^{k-1}+\frac{1}{2}u_i^{k-2}\bigg)+\frac{1}{3}(u_i^{k-1}-u_i^{k-2}).$$

Then

$$\begin{aligned}\|u^{(\tau)}-\pi_\tau u^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}&=\Big\|\frac{2}{3}\mathrm{D}_\tau u^{(\tau)}+\frac{1}{3}\pi_\tau(u^{(\tau)}-\pi_\tau u^{(\tau)})\Big\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}\\ &\le\frac{2}{3}\|\mathrm{D}_\tau u^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}+\frac{1}{3}\|u^{(\tau)}-\pi_\tau u^{(\tau)}\|_{L^r(\tau,T-\tau;W^{1,2d+4}(\Omega)')}.\end{aligned}$$

Adding $\|u^{(\tau)}-\pi_\tau u^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}\le C_1$ from the first Euler step (proved in a similar way as above) to the left-hand side and absorbing the last term on the right-hand side by the left-hand side, we find that

$$\frac{2}{3\tau}\|u^{(\tau)}-\pi_\tau u^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}\le\frac{2}{3\tau}\|\mathrm{D}_\tau u^{(\tau)}\|_{L^r(2\tau,T;W^{1,2d+4}(\Omega)')}\le C.$$

Together with the uniform $L^2(0,T;H^1(\Omega))$ bound for $u^{(\tau)}$, we can apply the Aubin–Lions compactness lemma in the version of [16] to conclude that, up to a subsequence, as $\tau\to 0$,

$$u^{(\tau)}\to u\quad\text{strongly in }L^2(\Omega_T).$$

In view of the higher-order estimate of Lemma 10, this convergence also holds in $L^q(\Omega_T)$ for all $q<2+4/d$. Furthermore, again up to a subsequence,

$$\mathrm{D}_\tau u^{(\tau)}\rightharpoonup\partial_t u\quad\text{weakly in }L^r(2\tau,T;W^{1,2d+4}(\Omega)').$$

These convergences are sufficient to pass to the limit $\tau\to 0$ in (31) for test functions $\phi_i\in L^{2d+4}(2\tau,T;W^{1,2d+4}(\Omega)')$. □

6. Second-order convergence

As in the previous section, we set $\mathrm{D}_{\Delta t}u_i^k=\frac{3}{2}u_i^k-2u_i^{k-1}+\frac{1}{2}u_i^{k-2}$ and write (31) as

$$\frac{1}{\Delta t}\int_\Omega \mathrm{D}_{\Delta t}u_i^k\phi_i\mathrm{d}x+\int_\Omega\big(\gamma\nabla u_i^k+(u_i^k)^+\nabla p_i(u^k)\big)\cdot\nabla\phi_i\mathrm{d}x=0. \tag{32}$$

A Taylor expansion shows that, for some $\xi_k\in(0,T)$,

$$\mathrm{D}_{\Delta t}u_i(t_k):=\frac{3}{2}u_i(t_k)-2u_i(t_{k-1})+\frac{1}{2}u_i(t_{k-2})=(\Delta t)\partial_t u_i(t_k)-\frac{(\Delta t)^3}{3}\partial_t^3u_i(\xi_k).$$

Then, using a test function $\phi_i\in H^1(\Omega)$ in (1),

$$\frac{1}{\Delta t}\int_\Omega \mathrm{D}_{\Delta t}u_i(t_k)\phi_i\mathrm{d}x+\int_\Omega(\gamma\nabla u_i+u_i\nabla p_i(u))(t_k)\cdot\nabla\phi_i\mathrm{d}x=\frac{(\Delta t)^2}{3}\int_\Omega\partial_t^3u_i(\xi_k)\phi_i\mathrm{d}x. \tag{33}$$

We take the difference of (32) and (33), choose the test function $\phi_i = p_i(u(t_k)) - p_i(u^k) = (A(u(t_k) - u^k))_i$, and sum over $i = 1, \ldots, n$:

$$\begin{aligned}
&\frac{1}{\Delta t}\int_\Omega \mathrm{D}_{\Delta t}(u(t_k) - u^k)^T A(u(t_k) - u^k)\mathrm{d}x = I_7 + I_8, \quad \text{where,} \qquad (34)\\
&I_7 = -\sum_{i=1}^n \int_\Omega \big[\gamma\nabla(u_i(t_k) - u_i^k) + u_i(t_k)\nabla p_i(u(t_k)) - (u_i^k)^+\nabla p_i(u^k)\big]\\
&\qquad \times\nabla(A(u(t_k) - u^k))_i\mathrm{d}x,\\
&I_8 = \frac{(\Delta t)^2}{3}\sum_{i=1}^n \int_\Omega \partial_t^3 u_i(\xi_k)(A(u(t_k) - u^k))_i\mathrm{d}x.
\end{aligned}$$

Set $v^k := u(t_k) - u_i^k$. It follows from the BDF2 inequality in Lemma 7, applied to the left-hand side, that

$$\frac{1}{\Delta t}\int_\Omega \mathrm{D}_{\Delta t}(u(t_k) - u^k)^T A(u(t_k) - u^k)\mathrm{d}x \geq \frac{1}{\Delta t}\big(H(v^k, v^{k-1}) - H(v^{k-1}, v^{k-2})\big).$$

For the terms I_7 and I_8, we use the definition $p_i(u^k) = (Au^k)_i$, the Lipschitz continuity of $z \mapsto z^+$, the nonnegativity of u_i, and Young's inequality:

$$\begin{aligned}
I_7 &= -\sum_{i=1}^n \int_\Omega \gamma\nabla(A^{1/2}v^k)_i \cdot \nabla(A^{1/2}v^k)_i\mathrm{d}x\\
&\quad -\sum_{i=1}^n \int_\Omega \big((u_i(t_k) - (u_i^k)^+)\nabla(Au(t_k))_i + (u_i^k)^+\nabla(A(u(t_k) - u^k))_i\big) \cdot \nabla(Av^k)_i\mathrm{d}x\\
&\leq -\gamma\|\nabla(A^{1/2}v^k)\|_{L^2(\Omega)}^2 + \lambda_m^{-1/2}\|A^{1/2}v^k\|_{L^2(\Omega)}\lambda_M^{3/2}\|\nabla u(t_k)\|_{L^\infty(\Omega)}\|\nabla(A^{1/2}v^k)\|_{L^2(\Omega)}\\
&\quad -\sum_{i=1}^n \int_\Omega (u_i^k)^+|\nabla(Av^k)_i|^2\mathrm{d}x \leq \frac{\lambda_M^3}{4\gamma\lambda_m}\|\nabla u\|_{L^\infty(\Omega_T)}^2\|A^{1/2}v^k\|_{L^2(\Omega)}^2 \quad \text{and}\\
I_8 &\leq \frac{(\Delta t)^2}{3\lambda_m^{1/2}}\|\partial_t^3 u\|_{L^\infty(0,T;L^2(\Omega))}\|A^{1/2}v^k\|_{L^2(\Omega)}.
\end{aligned}$$

Summarizing, we obtain from (34)

$$\begin{aligned}
&H(v^k, v^{k-1}) - H(v^{k-1}, v^{k-2}) \leq C_1\Delta t\|A^{1/2}v^k\|_{L^2(\Omega)}^2 + C_2(\Delta t)^3\|A^{1/2}v^k\|_{L^2(\Omega)}, \qquad (35)\\
&\quad \text{where} \quad C_1 = \frac{\lambda_M^3}{4\gamma\lambda_m}\|\nabla u\|_{L^\infty(\Omega_T)}^2, \quad C_2 = \frac{1}{3\lambda_m^{1/2}}\|\partial_t^3 u\|_{L^\infty(0,T;L^2(\Omega))}.
\end{aligned}$$

We iterate this inequality once more and use the inequality $a + b \leq \sqrt{2(a^2 + b^2)}$ as well as the norm equivalence (11):

$$\begin{aligned}
H(v^k, v^{k-1}) - H(v^{k-2}, v^{k-3}) &\leq C_1\Delta t\big(\|A^{1/2}v^k\|_{L^2(\Omega)}^2 + \|A^{1/2}v^{k-1}\|_{L^2(\Omega)}^2\big)\\
&\quad + C_2(\Delta t)^3\big(\|A^{1/2}v^k\|_{L^2(\Omega)} + \|A^{1/2}v^{k-1}\|_{L^2(\Omega)}\big)\\
&\leq C_1\Delta t\big(\|A^{1/2}v^k\|_{L^2(\Omega)}^2 + \|A^{1/2}v^{k-1}\|_{L^2(\Omega)}^2\big)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sqrt{2}C_2(\Delta t)^3\big(\|A^{1/2}v^k\|^2_{L^2(\Omega)} + \|A^{1/2}v^{k-1}\|^2_{L^2(\Omega)}\big)^{1/2} \\
&\le \frac{4C_1\Delta t}{3-\sqrt{8}}H(v^k,v^{k-1}) + \frac{4\sqrt{2}C_2(\Delta t)^3}{3-\sqrt{8}}H(v^k,v^{k-1})^{1/2}.
\end{aligned}
$$

We apply Young's inequality for $\varepsilon > 0$:

$$
\left(1 - \frac{4(C_1+\varepsilon)}{3-\sqrt{8}}\Delta t\right)H(v^k,v^{k-1}) \le H(v^{k-2},v^{k-3}) + \frac{2C_2^2(\Delta t)^5}{(3-\sqrt{8})\varepsilon},
$$

and assume that $\Delta t < (3-\sqrt{8})/(4(C_1+\varepsilon))$. This recursion is of the form $a_k \le ba_{k-2} + bc(\Delta t)^5$, where $a_k = H(v^k,v^{k-1})$ and

$$
b = \left(1 - \frac{4(C_1+\varepsilon)}{3-\sqrt{8}}\Delta t\right)^{-1}, \quad c = \frac{2C_2^2(\Delta t)^5}{(3-\sqrt{8})\varepsilon},
$$

and it can be resolved explicitly depending on whether k is odd or even:

$$
a_{2\ell+1} \le b^\ell a_1 + c(\Delta t)^5\sum_{j=0}^{\ell-1}b^j, \quad a_{2\ell+2} \le b^\ell a_2 + c(\Delta t)^5\sum_{j=0}^{\ell-1}b^j.
$$

The sum can be estimated according to

$$
\sum_{j=0}^{\ell-1}b^j = \frac{b^\ell-1}{b-1} \le \left(1 - \frac{4\Delta t}{3-\sqrt{8}}(C_1+\varepsilon)\right)^{-\ell+1}\frac{3-\sqrt{8}}{4\Delta t(C_1+\varepsilon)}.
$$

Since $\ell = t_\ell/\Delta t \le T/\Delta t$, the bracket approximates the exponential function and can be bounded by a constant depending only on $C_1+\varepsilon$ and T. This shows that there exist constants K_1, $K_2 > 0$ such that

$$
\begin{aligned}
H(v^{2\ell+1},v^{2\ell}) &\le K_1(C_1,\varepsilon,T)H(v^1,v^0) + K_2(C_1,C_2,\varepsilon^{-1},T)(\Delta t)^4, \\
H(v^{2\ell+2},v^{2\ell+1}) &\le K_1(C_1,\varepsilon,T)H(v^2,v^1) + K_2(C_1,C_2,\varepsilon^{-1},T)(\Delta t)^4.
\end{aligned}
$$

Going back to inequality (35) for $k = 2$, we can argue in a similar way as before that $H(v^2,v^1)$ is bounded by $K_3H(v^1,v^0) + K_4(\Delta t)^5$ for some constants K_3, $K_4 > 0$, which are independent of Δt. Furthermore, since $v^0 = 0$, we have $H(v^1,v^0) = (5/4)\|A^{1/2}(u(t_1) - u^1)\|_{L^2(\Omega)} \le K_5(\Delta t)^4$ for some $K_5 > 0$ independent of Δt. This shows that $H(v^k,v^{k-1}) \le K_6(\Delta t)^4$, where K_6 depends on C_1, C_2, ε^{-1}, and T. Taking the square root and using (11) shows the result.

7. Numerical examples

The finite-volume scheme (15)–(19) is implemented in Matlab, using the mobility $M(u,v) = \frac{1}{2}(u+v)$. As the numerical scheme is implicit, we have solved the nonlinear system of equations at each time step by using the Matlab routine `fsolve`, based on Newton's method with trust regions. The optimality tolerance was chosen as 10^{-14}.

7.1. **First example: one-dimensional domain, three species.** We choose the domain $\Omega = (0,1)$, the parameter $\gamma = 1/2$, as well as the positive definite matrix A and the initial data u^0 according to

$$A = \begin{pmatrix} 2 & 1 & 1/2 \\ 1 & 3 & 3/2 \\ 1/2 & 3/2 & 1 \end{pmatrix}, \quad u^0(x) = \begin{pmatrix} \cos(\pi x) + 2 \\ 2 - \cos(2\pi x) \\ 2 \end{pmatrix}.$$

The numerical parameters are $\Delta x = 1/12\,800$ and $\Delta t = 1/128$. The numerical solution is illustrated in Figure 1 at various times. All components converge to the constant steady state $\bar{u} = 2$. Interestingly, although initially equal to the steady state, the density u_3 becomes nonconstant for positive times before it tends to the constant steady state for large times. Such a phenomenon is sometimes called uphill diffusion, which typically appears in thermodynamic multicomponent systems due to cross diffusion [28].

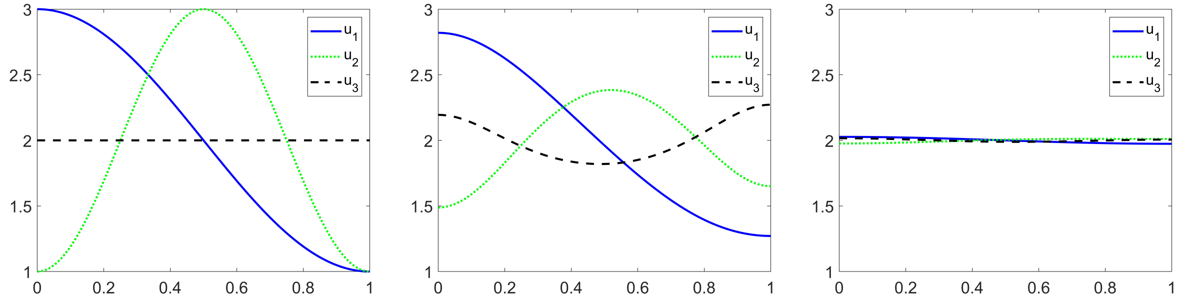

FIGURE 1. Densities $u_1(t)$ (darker blue line), $u_2(t)$ (lighter green line), $u_3(t)$ (dashed black line) at times $t = 0, 0.01, 0.1$ (from left to right) versus space.

7.2. **Second example: two-dimensional domain, two species.** We take $\Omega = (0,1)^2$, $\Delta x = \sqrt{2} \cdot 2^{-5} \approx 0.0044$, $\Delta t = 1/256$, $\gamma = 1/2$, and

$$A = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}, \quad u^0(x) = \begin{pmatrix} 1_{(0,1/2)^2}(x) \\ 1_{(1/2,1)^2}(x) \end{pmatrix}.$$

Figure 2 shows the evolution of $u = (u_1, u_2)$ at various times. Although being discontinuous and segregated initially, the solution becomes smooth and mixes the densities for positive times. This is not surprising, as full segregation (i.e., the supports of u_1 and u_2 do not intersect) is expected only when $\gamma = 0$ and $\det A = 0$. The numerical scheme preserved the nonnegativity in all our experiments, even for the initial data of this example. The numerical solutions are the same with or without the cutoff used in (18).

7.3. **Third example: exponential time decay.** We choose the one-dimensional domain $\Omega = (0,1)$, $\gamma = 0.1$, $\Delta x = 2^{-7}$, $\Delta t = (10 \cdot 2^7)^{-1}$, and

$$A = \begin{pmatrix} \beta & 2 \\ 2 & 1 \end{pmatrix}, \quad u^0(x) = \begin{pmatrix} 2 - \cos(\pi x) \\ 2 + \cos(\pi x) \end{pmatrix},$$

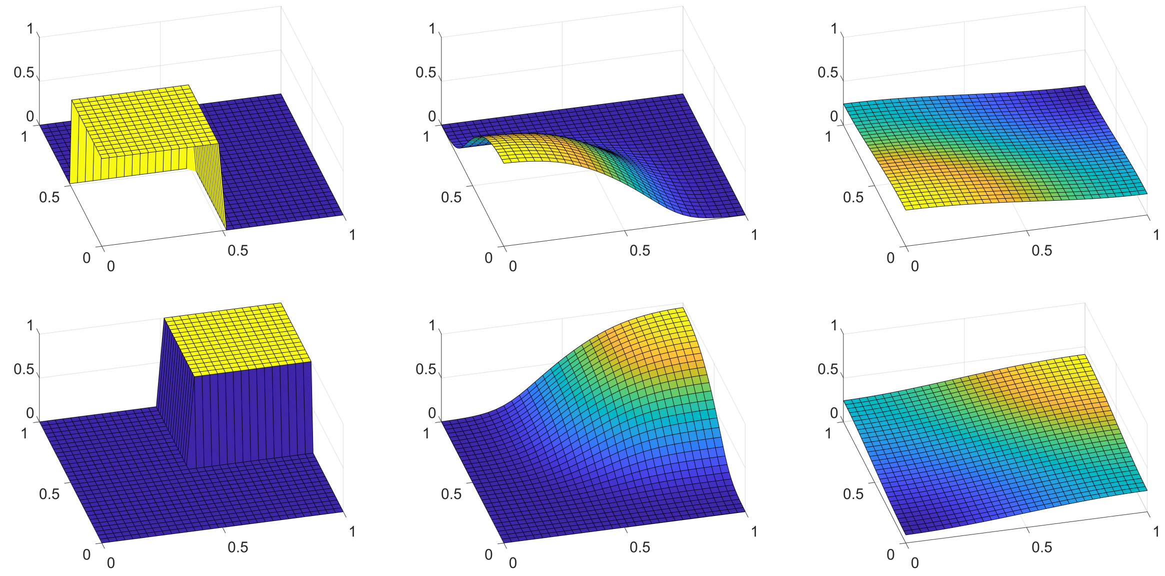

FIGURE 2. Density $u_1(t)$ (upper row) and $u_2(t)$ (lower row) at times $t = 0, 0.02, 0.2$ (from left to right) versus space.

where $\beta > 4$. The distance $\|A^{1/2}(u^k - \bar{u})\|_{L^2(\Omega)}$ presented in Figure 3 for $\beta = 5$ and $\beta = 4.01$ shows that the time decay behaves exponentially, as predicted by Theorem 4. The decay rates (excluding the initial decay) are -4.37 for $\beta = 5$ and -1.03 for $\beta = 4.01$, and they decrease for smaller values of $\det A$. We have also observed an exponential decay when $\gamma = 0$ with smaller decay rates.

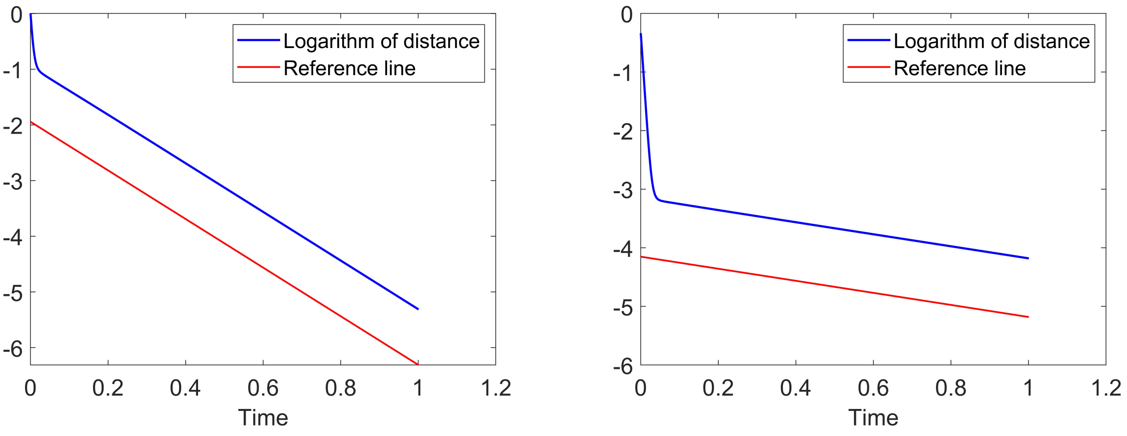

FIGURE 3. Semilogarithmic plot for $\|A^{1/2}(u^k - \bar{u})\|_{L^2(\Omega)}$ versus time t_k.

7.4. **Fourth example: Convergence rate in time.** We choose the values for A and u^0 as in the previous example as well as $\gamma = 0$, $\Delta x = 2^{-9}$, and $\Delta t = (10 \cdot 2^p)^{-1}$ with $p = 1, \ldots, 8$.

The reference solution u_{ref} is computed with the time step size $\Delta t = (10 \cdot 2^9)^{-1}$. As expected, the convergence rate at time $T = 0.02$, shown in Figure 4 for two different values of β, is about two, even in the case $\det A = 0$.

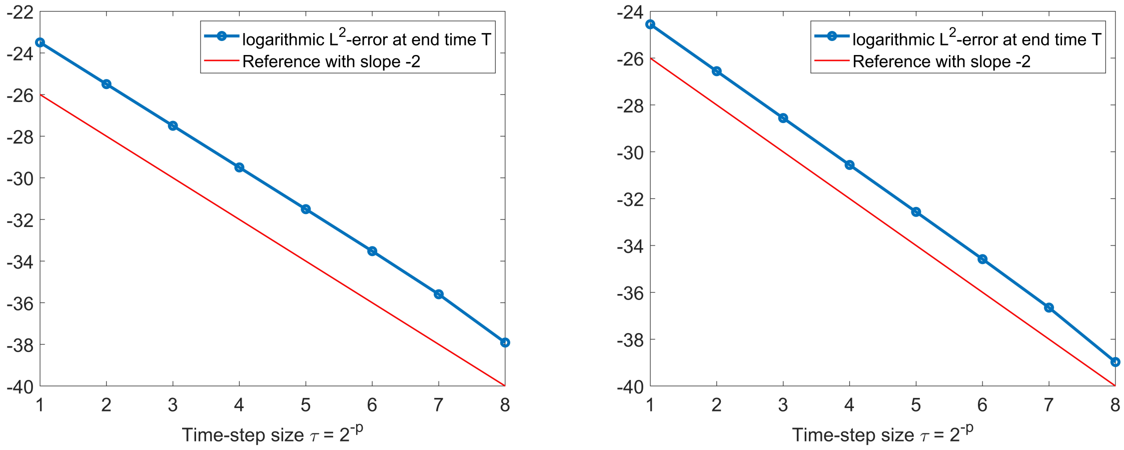

FIGURE 4. Discrete $L^2(\Omega)$ error $\|A^{1/2}(u^{(\Delta t)} - u_{\mathrm{ref}})(T)\|_{L^2(\Omega)}$ versus time step size $\Delta t = (10 \cdot 2^p)^{-1}$ for $p = 1, \ldots, 8$ for $\beta = 5$ (left) and $\beta = 4$ (right).

REFERENCES

[1] H. Amann. Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. In: H. J. Schmeisser and H. Triebel (eds.), Funct. Spaces Differ. Op. Nonlin. Anal., pp. 9–126. Teubner, Wiesbaden, 1993.

[2] B. Andreianov, M. Bendahmane, and R. Ruiz-Baier. Analysis of a finite-volume method for a cross-diffusion model in population dynamics. *Math. Models Meth. Appl. Sci.* 21 (2011), 307–344.

[3] R. Bailo, J. A. Carrillo, and J. Hu. Fully discrete positivity-preserving and energy-dissipating schemes for aggregation-diffusion equations with a gradient-flow structure. *Commun. Math. Sci.* 18 (2020), 1259–1303.

[4] M. Bertsch, M. Gurtin, D. Hilhorst, and L. Peletier. On interacting populations that disperse to avoid crowding: preservation of segregation. *J. Math. Biol.* 23 (1985), 1–13.

[5] M. Bertsch, D. Hilhorst, H. Izuhara, and M. Mimura. A nonlinear parabolic-hyperbolic system for contact inhibition of cell-growth. *Differ. Eqs. Appl.* 4 (2012), 137–157.

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

[7] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. *IMA J. Numer. Anal.* 35 (2015), 1125–1149.

[8] C. Calgaro and M. Ezzoug. L^∞-stability of the IMEX-BDF2 finite volume scheme for convection-diffusion equation. In: C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII*, pp. 245–253. Springer, Cham, 2017.

[9] C. Cancès and B. Gaudeul. A convergent entropy diminishing finite volume scheme for a cross-diffusion system. *SIAM J. Numer. Anal.* 58 (2020), 2684–2710.

[10] C. Chainais-Hillairet, J.-G. Liu, and Y.-J. Peng. Finite volume scheme for multi-dimensional drift-diffusion equations and convergence analysis. *ESAIM Math. Model. Numer. Anal.* 37 (2003), 319–338.

[11] L. Chen, E. Daus, and A. Jüngel. Rigorous mean-field limit and cross diffusion. *Z. Angew. Math. Phys.* 70 (2019), no. 122, 21 pages.

[12] X. Chen and A. Jüngel. Weak-strong uniqueness of renormalized solutions to reaction-cross-diffusion systems. *Math. Models Meth. Appl. Sci.* 29 (2019), 237–270.

[13] W. Chen, C. Wang, X. Wang, and S. Wise. Positivity-preserving, energy stable numerical schemes for the Cahn–Hilliard equation with logarithmic potential. *J. Comput. Phys.* X 3 (2019), no. 1000031, 29 pages.

[14] W. Dahmen, B. Faermann, I. Graham, W. Hackbusch, and S. Sauter. Inverse inequalities on non-quasi-uniform meshes and applications to the Mortar element method. *Math. Comput.* 73 (2004), 1107–1138.

[15] L. Dong, C. Wang, H. Zhang, and Z. Zhang. A positivity-preserving second-order BDF scheme for the Cahn–Hilliard equation with variable interfacial parameters. *Commun. Comput. Phys.* 28 (2020), 967–998.

[16] M. Dreher and A. Jüngel. Compact families of piecewise constant functions in $L^p(0,T;B)$. *Nonlin. Anal.* 75 (2012), 3072–3077.

[17] J. Droniou and N. Nataraj. Improved L^2 estimate for gradient schemes and super-convergence of the TPFA finite volume scheme. *IMA J. Numer. Anal.* 38 (2018), 1254–1293.

[18] P.-E. Druet, K. Hopf, and A. Jüngel. Hyperbolic-parabolic normal form and local classical solutions for cross-diffusion systems with incomplete diffusion. Submitted for publication, 2022. arXiv:2210.17244.

[19] E. Emmrich. Two-step BDF time discretization of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations. *Comput. Meth. Appl. Math.* 9 (2009), 37–62.

[20] R. Eymard, T. Gallouët, and R. Herbin. Convergence of finite volume schemes for semilinear convection diffusion equations. *Numer. Math.* 82 (1999), 91–116.

[21] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In: P. G. Ciarlet and J.-L. Lions (eds.). *Handbook of Numerical Analysis* 7 (2000), 713–1018.

[22] Y. Gu and J. Shen. Bound preserving and energy dissipative schemes for porous medium equation. *J. Comput. Phys.* 410 (2020), no. 109378, 21 pages.

[23] A. Hill. Global dissipativity for A-stable methods. *SIAM J. Numer. Anal.* 34 (1997), 119–142.

[24] A. Jüngel and J.-P. Milišič. Entropy dissipative one-leg multistep time approximations of nonlinear diffusive equations. *Numer. Meth. Part. Diff. Eqs.* 31 (2015), 1119–1149.

[25] A. Jüngel, S. Portisch, and A. Zurek. Nonlocal cross-diffusion systems for multi-species populations and networks. *Nonlin. Anal.* 219 (2022), no. 112800, 26 pages.

[26] A. Jüngel and A. Zurek. A finite-volume scheme for a cross-diffusion model arising from interacting many-particle population systems. In: R. Klöfkorn, E. Keilegavlen, F. Radu, and J. Fuhrmann (eds.), *Finite Volumes for Complex Applications IX*, pp. 223–231. Springer, Cham, 2020.

[27] A. Jüngel and A. Zurek. A discrete boundedness-by-entropy method for finite-volume approximations of cross-diffusion systems. To appear in *IMA J. Math. Anal.*, 2022. https://doi.org/10.1093/imanum/drab101.

[28] R. Krishna. Uphill diffusion in multicomponent mixtures. *Chem. Soc. Rev.* 44 (2015), 2812–2836.

[29] D. Matthes and S. Plazetta. A variational formulation of the BDF2 method for metric gradient flows. *ESAIM Math. Model. Numer. Anal.* 53 (2019), 145–172.

[30] C. Rao. Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* 21 (1982), 24–43.

Institute of Analysis and Scientific Computing, Technische Universität Wien, Wiedner Hauptstrasse 8–10, 1040 Wien, Austria
Email address: juengel@tuwien.ac.at

Institute of Analysis and Scientific Computing, Technische Universität Wien, Wiedner Hauptstrasse 8–10, 1040 Wien, Austria
Email address: martin.vetter@tuwien.ac.at