Abstract. Cross-diffusion systems are systems of nonlinear parabolic partial differential equations that are used to describe dynamical processes in several application, including chemical concentrations and cell biology. We present a space-time approach to the proof of existence of bounded weak solutions of cross-diffusion systems, making use of the system entropy to examine long-term behavior and to show that the solution is nonnegative, even when a maximum principle is not available. This approach naturally gives rise to a novel space-time Galerkin method for the numerical approximation of cross-diffusion systems that conserves their entropy structure. We prove existence and convergence of the discrete solutions, and present numerical results for the porous medium, the Fisher-KPP, and the Maxwell-Stefan problem.

Key words. Space-time Galerkin method, entropy method, strongly coupled parabolic systems, global-in-time existence, bounded weak solutions, space-time finite elements

AMS subject classifications. 35K51, 35K55, 35Q92, 65M60, 41A10

1. Introduction. In this paper we develop a new space-time approach to the celebrated boundedness by entropy method by Ansgar Jünzel [26]. For a textbook version see [27]; see also [11,31].

Cross-diffusion systems are systems of nonlinear parabolic partial differential equations that are commonly used to describe dynamical processes appearing in modeling, for example, population dynamics, ion transport through nanopores, tumor growth models, and multicomponent gas mixtures. The challenge in the analysis of these systems is that the diffusion matrix is not necessarily symmetric nor positive semi-definite, and thus no maximum principle is available. Following [26], the remedy is to make use of the entropy structure of the system. Introducing the entropy function, a transformation of the solution, allows us to examine long-term behavior and show that the solution is nonnegative and bounded. Here, we present a space-time approach to the proof of existence of bounded weak solutions of cross-diffusion systems. The main tool will be the method of compensated compactness, which is a special technique of applying the classical div-curl lemma [48]. The key difference to the existing literature is that we do not make use of time-stepping, but instead consider time and space altogether. This naturally leads to a novel space-time Galerkin method for the numerical approximation of cross-diffusion systems. The space-time approach entails test and trail spaces, as well as the mesh, where time is included as additional dimension. This provides an easy way to increase the approximation degree simultaneously in space and time, and makes space-time $hp$-refinement possible. In a schematic way, our overall approach consists of the following four steps:

1. space-time variational formulation,
2. transformation to entropy variables,
3. regularization with a space-time $H^1$ inner product,
4. Galerkin discretization.

Existing numerical schemes for cross-diffusion systems rely on time-stepping methods. An entropy/energy conserving time-stepping algorithm for thermomechanical problems was developed in [41] being of second order in time. In [32], assuming existence of sufficient regular strong solutions on some time interval $[0,T]$ of a scalar diffusion equation, Runge-Kutta methods were studied using maximal regularity. Although maximal regularity also applies to a certain type of cross-diffusion systems [42], Runge-Kutta methods were only applied to very restrictive classes; an example (semi-discrete Runge-Kutta scheme) can be found in [29]. In [25], an entropy diminishing/mass conserving fully discrete variational formulation for a cross-diffusion system was presented. An alternative discretization for cross-diffusion systems based on the change to entropy-
variables has been proposed in [16], where a dissipation-preserving approximation by Galerkin methods in space and discontinuous Galerkin methods in time has been studied.

Maxwell-Stefan systems, see [37, 46], describe multicomponent diffusive fluxes in non-dilute solutions or gas mixtures, and are a prime example for the cross-diffusion systems considered here. The first result on global solutions for the Maxwell-Stefan equations close to the equilibrium is given in [23]. The global existence of solutions close to equilibrium and the large time convergence to this equilibrium can be found in [21, Chapter 9], [22, 24], and [42, Chapter 12]. The proof of existence of local classical solutions to the Maxwell-Stefan equations can be found in [6]. For a textbook on this topic, see [42]. The fact that the Maxwell-Stefan equations satisfy the assumptions made in this paper, see (H1)-(H3) below, is due to [30], where the entropy structure of the Maxwell-Stefan system was used to prove the existence of globally bounded weak solutions. An entropy structure was also identified for a generalized Maxwell-Stefan system coupled to the Poisson equation in [28], where the existence of global weak solutions was proven as well. The unconditional convergence to the unique equilibrium for given mass was shown in [24,36] without reaction terms. Those results were extended to also include reaction terms using mass-action kinetics in [13], whenever a detailed balance equilibrium exists. The heat equation can be recovered from the Maxwell-Stefan equation as a relaxation limit [43].

As of yet, numerical schemes for the Maxwell-Stefan equations commonly employ time-stepping. A finite differences approximation can be found in [34, 35]. Fast solvers for explicit finite-difference schemes were studied in [20]. A posteriori estimates for finite elements in the stationary case are given in [10]. In [40], a mass conserving finite volume scheme was presented. Existence of solutions for a mixed finite element scheme under some restrictions on the coefficients was proven in [38]. The scheme of [14] was proven to also conserve the \( L^\infty \) bound by making use of a maximum principle. A scheme using finite elements in space and implicit Euler in time was used to approximate a Poisson-Maxwell-Stefan system in [28]. That scheme, which is based on a formulation in entropy variables, admits solutions that conserve the mass as well as the entropy structure. As a by product, the solution satisfy an \( L^\infty \) bound. Another scheme that is mass conserving and conserves the \( L^\infty \) bounds of the solutions was presented in [8].

On simultaneous space-time finite element approaches for parabolic problems, there is a rich literature on the linear case, focusing on the heat equation; see, e.g. the recent overview [33, Ch. 7]. We point out that due to the different orders of derivatives present, conforming discretizations are typically based on a Petrov-Galerkin approach; see e.g. [2, 4, 45, 47, 50]. For nonlinear parabolic equations, the existing literature on space-time methods is much sparser. The adaptive finite element scheme introduced in [17] for linear parabolic problems was extended in [18] to the scalar version of the nonlinear reaction-diffusion equation treated in this paper. A space-time discontinuous Galerkin method for scalar nonlinear convection and diffusion was introduced in [49]. A space-time method for nonlinear PDEs using adaptive wavelets was introduced in [1].

The structure of this paper is as follows. In section 2, we state the problem and make the necessary assumptions for the existence of an entropy function. In section 3, we present the space-time Galerkin method on a regularized formulation of the problem in the entropy variable unknown, and state our two main results in Proposition 3.3 and Proposition 3.4, namely existence and convergence of discrete solutions, respectively. Existence of discrete solutions is proven in subsection 3.1. The proof of convergence will be split into two parts, first showing convergence with respect to mesh size in subsection 3.2, then proving convergence as the regularization parameter goes to zero in subsection 3.3. In subsection 3.4, we are then able to prove existence of a weak solution of the continuous problem. Numerical tests for the porous medium, the Fisher-KPP, and the Maxwell-Stefan problem are presented in section 4. All numerical results\(^1\) were obtained using the finite element software NGSolve, see [44]. Additionally, in section 5, we reformulate the Maxwell-Stefan system with implicitly given currents in terms of the concentrations, and test it numerically.

\(^1\)The code is available online at https://github.com/PaulSt/CrossDiff
2. General setting. Let \( \Omega \subset \mathbb{R}^d \) be a bounded domain, and \( \rho_0 \in L^\infty(\Omega)^N, N \geq 1 \), a vector-valued function. We consider the following nonlinear reaction-diffusion system in the vector-valued unknown \( \rho(t) = (\rho_1, \ldots, \rho_N)(\cdot, t) : \Omega \to \mathbb{R}^N \):

\[
\begin{aligned}
&\frac{\partial \rho}{\partial t} - \nabla \cdot (A(\rho) \nabla \rho) = f(\rho) \quad \text{in } \Omega, \ t > 0, \\
&\quad (A(\rho) \nabla \rho) \cdot \nu = 0 \quad \text{on } \partial \Omega, \ t > 0, \\
&\quad \rho(0) = \rho_0 \quad \text{in } \Omega.
\end{aligned}
\]

Here, \( A(\rho) \in \mathbb{R}^{N \times N} \) is the diffusion matrix, \( f(\rho) : \mathbb{R}^N \to \mathbb{R}^N \) represents the reactions, and \( \nu \) is the outward pointing unit normal vector at \( \partial \Omega \); moreover, for \( 1 \leq i \leq N \),

\[
(\nabla \cdot (A(\rho) \nabla \rho))_i = \sum_{\mu=1}^d \sum_{j=1}^N \frac{\partial}{\partial x_\mu} \left( A_{ij}(\rho) \frac{\partial \rho_j}{\partial x_\mu} \right), \quad ((A(\rho) \nabla \rho) \cdot \nu)_i = \sum_{\mu=1}^d \sum_{j=1}^N A_{ij}(\rho) \frac{\partial \rho_j}{\partial x_\mu} \nu_\mu.
\]

We make the following hypotheses, which are slightly stronger assumptions compared to those made by A. Jüngel in [26].

(H1) \( A \in C^0(\overline{\Omega}; \mathbb{R}^{N \times N}) \) and \( f \in C^0(\overline{\Omega}; \mathbb{R}^N) \), for a bounded domain \( \Omega \subset (0, \infty)^N \).

(H2) There exists a convex function \( s \in C^2([0, \infty)) \cap C^0([0, \infty)) \), with \( s' : \mathbb{R}^N \to \mathbb{R}^N \) invertible and \( u := (s')^{-1} \in C^1(\mathbb{R}^N, \mathbb{D}) \), such that the following two conditions are satisfied:

(H2a) There exists a constant \( \gamma > 0 \) such that

\[
z \cdot s''(\rho) A(\rho) z \geq \gamma |z|^2 \quad \forall z \in \mathbb{R}^N, \rho \in \mathbb{D}.
\]

Note that \( s''(\rho) \) is matrix-valued, with \( (s''(\rho))_{k\ell} = \frac{\partial^2}{\partial \rho_k \partial \rho_\ell} s(\rho) \).

(H2b) There exists a constant \( C_f \geq 0 \) such that

\[
f(\rho) \cdot s'(\rho) \leq C_f \quad \forall \rho \in \mathbb{D}.
\]

Additionally, we make the following assumption on \( \rho_0 \):

(H3) The initial datum \( \rho_0 : \Omega \to \mathbb{D} \) is measurable.

A discussion on when it is possible to find a convex function \( s \) such that (H2) is satisfied for cross-diffusion equations can be found in [12] (see [12, Lemma 22]).

Let \( T > 0 \). A weak formulation of (2.1) reads as follows: Find \( \rho \in L^2(0, T; H^1(\Omega)^N) \cap H^1(0, T; (H^1(\Omega)^N)^N) \) such that

\[
\int_0^T \langle \phi, \partial_t \rho \rangle dt + \sum_{i,j=1}^N \int_0^T \int_\Omega \nabla \phi_i \cdot A_{ij}(\rho) \nabla \rho_j dx dt = \int_0^T \int_\Omega \phi \cdot f(\rho) dx dt
\]

for all \( \phi \in L^2(0, T; H^1(\Omega)^N) \), with \( \rho(0) = \rho_0 \), where \( \langle \cdot, \cdot \rangle \) denotes the duality product between \( H^1(\Omega)^N \) and \((H^1(\Omega)^N)^N\).

By introducing the so-called entropy variable \( w \), which satisfies \( \rho = u(w) \), problem (2.1), as well as (2.2), can be rewritten in terms of the unknown \( w \).

Remark 2.1. In [26], a more degenerate version of (H2a) is permitted where, in a nutshell, the coercivity inequality is replaced by \( z \cdot s''(\rho) A(\rho) z \geq \gamma \sum_i \rho_i^{2(m-1)} z_i^2 \) for some \( m \). In that context, an \( L^2(0, T; H^2(\Omega)) \) estimate for \( \rho_i \) might be out of reach. Instead, the system is rewritten there by using \( \rho_i^{m-1} \) as an \( L^2(0, T; H^1(\Omega)) \) function. Moreover, also entropy densities \( s \), which are not bounded, such as \( s(u) = u - \log u \), are allowed in [26]. As a consequence, a different version of the hypothesis (H2b) is considered there. We believe that our ansatz can be extended to that scenario applying the ideas from [26]. This will mainly affect the entropy inequalities and the proof of Proposition 3.8. However, we chose to use our simplified assumptions that already cover a large class of parabolic systems. By this, we try to keep the idea of our proof as fundamental as possible.
3. Space-time Galerkin method. Let the time $T \in (0, \infty)$ be fixed. We denote by $Q_T = (0, T) \times \Omega$ the space-time cylinder for a domain $\Omega \subset \mathbb{R}^d$, $d \geq 1$. We derive our method in four steps.

Step 1 (space-time variational formulation). The first step is to perform integration by parts in the time variable in (2.2), and to use the embedding

\[(3.1) \quad C([0, T]; L^2(\Omega)^N) \subset L^2(0, T; H^1(\Omega)^N) \cap H^1(0, T; (H^1(\Omega)^N)^N),\]

which can be proved exactly as in [19, Chapter 5.9, Theorem 3]. Then, we define the following variational formulation of (2.1).

**Definition 3.1 (space-time variational formulation/weak solution to (2.1)).** Find $\rho \in L^2(0, T; H^1(\Omega)^N) \cap H^1(0, T; (H^1(\Omega)^N)^N)$ such that

\[(3.2) \quad \int_\Omega \phi(T) \cdot \rho(T)dx - \int_0^T \int_\Omega \partial_t \phi \cdot \rho dx dt + \sum_{i,j=1}^N \int_0^T \int_\Omega \nabla \phi_i \cdot A_{ij}(\rho) \nabla \rho_j dx dt = \int_0^T \int_\Omega \phi(0) \cdot \rho_0 dx dt \]

for all $\phi \in H^1(Q_T)^N$.

The following lemma, which will be proven in section 3.4 below (see Remark 3.11), establishes that the variational formulation in Definition 3.1 is actually equivalent to the one in (2.2).

**Lemma 3.2.** Let $T > 0$. A function $\rho \in L^2(0, T; H^1(\Omega)^N) \cap H^1(0, T; (H^1(\Omega)^N)^N)$ satisfies (3.2) for all $\phi \in H^1(Q_T)^N$ if and only if it satisfies (2.2) for all $\phi \in L^2(0, T; H^1(\Omega)^N)$, with $\rho(0) = \rho_0$.

Step 2 (transformation to entropy variables). In the second step, we express $\rho$ in formulation (3.2) as a function of the entropy variable $w$, namely $\rho = u(w)$. The resulting space-time variational formulation is the following: Find $w \in H^1(Q_T)^N$ such that

\[(3.3) \quad \int_\Omega \phi(T) \cdot u(w)(T)dx - \int_0^T \int_\Omega \partial_t \phi \cdot u(w) dx dt + \sum_{i,j=1}^N \int_0^T \int_\Omega \nabla \phi_i \cdot A_{ij}(u(w)) \nabla (u(w))_j dx dt = \int_0^T \int_\Omega \phi(0) \cdot \rho_0 dx dt \]

for all $\phi \in H^1(Q_T)^N$. Here, we use the notation $\phi(t) := \text{tr}(\phi)(t, \cdot)$, where tr denotes the trace operator $H^1(Q_T)^N \to L^2([0, T] \times \Omega)^N$.

Step 3 (regularization). Then, we introduce the following regularized problem with regularization parameter $\varepsilon$: Find $w^\varepsilon \in H^1(Q_T)^N$ such that

\[(3.4) \quad \varepsilon(\phi, w^\varepsilon)_{H^1(Q_T)^N} + a(u(w^\varepsilon), \phi) + \sum_{i,j=1}^N \int_0^T \int_\Omega \nabla \phi_i \cdot A_{ij}(u(w^\varepsilon)) \nabla (u(w^\varepsilon))_j dx dt = \int_0^T \int_\Omega \phi(0) \cdot \rho_0 dx dt \]

for all $\phi \in H^1(Q_T)^N$. The regularization term is given by the scaled $H^1(Q_T)^N$ inner product defined as

\[(f, g)_{H^1(Q_T)^N} := \sum_{i=1}^N \left( \int_{Q_T} f_i g_i dx dt + \int_{Q_T} \nabla f_i \cdot \nabla g_i dx dt + \varepsilon \int_{Q_T} \partial_t f_i \partial_t g_i dx dt \right), \]
for \( f,g \in H^1(\Omega)^N \). Its associated norm will be denoted by \( \| \cdot \|_{H^1(\Omega)^N} \).

**Step 4 (Galerkin discretization).** Finally, we discretize equation (3.4). Let \( \{V_h\}_{h>0} \) be a family of finite dimensional spaces, parametrized by \( h > 0 \), such that, for every \( h, V_h \subset H^1(\Omega)^N \). We make the following approximability assumption on the family of spaces \( (3.6) \)

\[
\text{(H4) For all } v \in H^1(\Omega)^N,
\lim_{h \to 0} \inf_{v_h \in V_h} \| v - v_h \|_{H^1(\Omega)^N} = 0.
\]

Therefore, we consider the following space-time Galerkin scheme in the entropy variable unknown: Find \( w_h^\varepsilon \in V_h \) such that

\[
(3.5) \quad \varepsilon(\phi, w_h^\varepsilon)_{H^1(\Omega)^N} + a(u(w_h^\varepsilon), \phi) + \sum_{i,j=1}^N \int_0^T \int_{\Omega} \nabla \phi_i \cdot A_{ij}(u(w_h^\varepsilon)) \nabla (u(w_h^\varepsilon))_j dx dt
\]

\[
= \int_0^T \int_{\Omega} \phi \cdot f(u(w_h^\varepsilon)) dx dt + \int_{\Omega} \phi(0) \cdot \rho_0 dx
\]

for all \( \phi \in V_h \).

The first term in (3.5) can be interpreted as a stabilization term for the Galerkin scheme, with parameter \( \varepsilon > 0 \). This is used to obtain a control of the entropy variable. Note that, due to the nonlinearity of \( u \), we expect that \( u(w_h^\varepsilon) \notin V_h \).

The following two propositions constitute the main result of this paper. Proposition 3.3 establishes that the Galerkin problem (3.5) admits solutions \( w_h^\varepsilon \) and these solutions satisfy an entropy estimate. Then, this is exploited in Proposition 3.4 to obtain existence of weak solutions \( \rho \) to the continuous problem (2.1), together with related entropy estimates. Propositions 3.3 and 3.4 will be proven in section 3.1 and section 3.4, respectively. Here and in the following, \(|\Omega|\) denotes the volume of \( \Omega \), and \( \gamma \) and \( C_f \) are as in Assumption (H2).

**Proposition 3.3 (Existence of discrete solutions).** There exists a solution \( w_h^\varepsilon \in V_h \) of method (3.5). Moreover, every solution \( w_h^\varepsilon \in V_h \) of (3.5), for \( \varepsilon, h > 0 \), satisfies the entropy estimate

\[
(3.6) \quad \varepsilon \| w_h^\varepsilon \|_{H^1(\Omega)^N}^2 + \int_{\Omega} s(u(w_h^\varepsilon(T))) dx + \gamma \int_{\Omega} \| \nabla u(w_h^\varepsilon) \|^2 dx dt \leq \int_{\Omega} s(\rho_0) dx + C_f |\Omega| T.
\]

**Proposition 3.4 (Convergence).** Let \( w_h^\varepsilon \in V_h \) be a solution of (3.5) for \( \varepsilon, h > 0 \). Then there exist a weak solution

\[
\rho \in L^2(0,T; H^1(\Omega)^N) \cap H^1(0,T; (H^1(\Omega)^N)^N) \cap L^\infty((0,T) \times \Omega)^N
\]

of (2.1) and sequences \( h_i, \varepsilon_i \to 0 \), as \( i \to \infty \), such that

\[
u(w_h^{\varepsilon_i}) \to \rho \quad \text{in } L^r(\Omega)^N, \text{ as } i \to \infty
\]

for all \( r \in [1, \infty) \). Moreover, \( \rho \) satisfies the entropy estimate

\[
(3.7) \quad \int_{\Omega} s(\rho(\tau)) dx + \gamma \int_0^T \int_{\Omega} |\nabla \rho|^2 dx dt \leq \int_{\Omega} s(\rho_0) dx + C_f |\Omega| T
\]

for all \( \tau \in (0,T] \).

**Remark 3.5 (Non-closed systems).** The general setting given by (2.1) describes closed systems, i.e., without influx or outflux at \( \partial \Omega \). These systems are of particular interest, as they obey the second law of thermodynamics proving the decay of the entropy. However, in some cases (e.g. the lung model presented in section 5.3), one is interested in a non-closed subsystem involving, for instance, inhomogeneous Dirichlet boundary conditions on some part \( \Gamma_D \subset \partial \Omega \) of the boundary.
If $\rho = g$ is prescribed on $\Gamma_D$, for a given $g \in H^1(\Omega)^N$ taking values in $D$, we assume that the approximation spaces $V_h$ are affine subspaces of $s'(g) + H^1_0(Q_T)^N$, where $H^1_0(Q_T)^N$ is the closure in $H^1(Q_T)^N$ of the space of $C^\infty(Q_T)^N$ functions vanishing at $\Gamma_D \times (0, T)$. Unfortunately, we can neither guarantee the discrete entropy estimate (3.6) nor its continuous version (3.7). Instead, we have to work with the relative entropy

$$s^\infty(\rho | g) := s(\rho) - s(g) - s'(g) \cdot (\rho - g),$$

which is still convex in the variable $\rho$. We conjecture that, under the assumption $s'(g) \in L^\infty(\Omega)$, one can prove corresponding versions of Proposition 3.3 and Proposition 3.4 by employing an estimate of the relative entropy of the form

$$\int _\Omega s^\infty(\rho | g)dx + \gamma \int _0^T \int _\Omega |\nabla \rho|^2 dxdt \leq \int _\Omega s^\infty(\rho_0 | g)dx + (C_f|\Omega| + C_{A,g,f}) \tau$$

for all $\tau \in (0, T)$ and some $C_{A,g,f}$ only depending on $\gamma$, $\|g\|_{H^1(\Omega)}$ and $\|s'(g)\|_{L^\infty(\Omega)}$, as well as on $\|A\|_{L^\infty(D)}$ and $\|f\|_{L^\infty(D)}$. This estimate can be derived from (2.2), with test function $\phi = s'(\rho) - s'(g)$. However, the proof of convergence of the discrete scheme (3.5) remains an open problem and is under ongoing investigation.

### 3.1. Existence of a solution of the numerical scheme.

**Proof of Proposition 3.3.** The idea is to use the Leray-Schauder fixed-point theorem for the mapping $\Phi : V_h \to V_h$, $v \mapsto w_h^\sigma$, where $w_h^\sigma$ denotes the unique solution of (3.5) with all occurrences of $u(w_h^\sigma)$ replaced by $u(v)$. Since $A, f, u$ are continuous, so is $\Phi$. Since $V_h$ has finite dimension, $\Phi$ is also compact. Then by the Leray-Schauder fixed-point theorem, we obtain that $\Phi$ admits a fixed-point if we can show that the set

$$\{ w \in V_h : w = \sigma \Phi(w), \ \sigma \in [0, 1] \}$$

is bounded.

Let $w_h^\sigma = \sigma \Phi(w_h^\sigma)$ for $\sigma \in (0, 1]$ and choose $\phi := w_h^\sigma$. Then (3.5) entails

$$\frac{\varepsilon}{\sigma} \|w_h^\sigma\|_{H^1(Q_T)^N}^2 + \int _\Omega w_h^\sigma(T) \cdot u(w_h^\sigma(T))dx - \int _0^T \int _\Omega \partial_t w_h^\sigma \cdot u(w_h^\sigma)dxdt + \sum _{i,j=1}^N \int _0^T \int _\Omega \nabla(w_h^\sigma)_i \cdot A_{ij}(u(w_h^\sigma))\nabla(u(w_h^\sigma))_j dxdt = \int _0^T \int _\Omega w_h^\sigma \cdot f(u(w_h^\sigma))dxdt + \int _\Omega w_h^\sigma(0) \cdot \rho_0 dx.$$

Using that $\partial_t s(u(w_h^\sigma)) = s'(u(w_h^\sigma)) \cdot \partial_t (u(w_h^\sigma)) = w_h^\sigma \cdot \partial_t (u(w_h^\sigma))$, we have

$$\partial_t w_h^\sigma \cdot u(w_h^\sigma) = \partial_t (w_h^\sigma \cdot u(w_h^\sigma)) - w_h^\sigma \cdot \partial_t (u(w_h^\sigma)) = \partial_t (w_h^\sigma \cdot u(w_h^\sigma) - s(u(w_h^\sigma))).$$

Thus, by the fundamental theorem of calculus,

$$\int _\Omega w_h^\sigma(T) \cdot u(w_h^\sigma(T))dx - \int _\Omega w_h^\sigma(0) \cdot \rho_0 dx - \int _0^T \int _\Omega \partial_t w_h^\sigma \cdot u(w_h^\sigma)dxdt = - \int _\Omega (s(u(w_h^\sigma(0))) + w_h^\sigma(0) \cdot (\rho_0 - u(w_h^\sigma(0))))dx + \int _\Omega s(u(w_h^\sigma(T)))dx.$$

Note that, by definition, $s'(u(w_h^\sigma)) = w_h^\sigma$. The convexity of $s$ then implies that

$$s(u(w_h^\sigma(0))) + w_h^\sigma(0) \cdot (\rho_0 - u(w_h^\sigma(0))) = s(u(w_h^\sigma(0))) + s'(u(w_h^\sigma(0))) \cdot (\rho_0 - u(w_h^\sigma(0))) \leq s(\rho_0)$$

and hence,

$$\int _\Omega w_h^\sigma(T) \cdot u(w_h^\sigma(T))dx - \int _\Omega w_h^\sigma(0) \cdot \rho_0 dx - \int _0^T \int _\Omega \partial_t w_h^\sigma \cdot u(w_h^\sigma)dxdt \geq \int _\Omega s(u(w_h^\sigma(T)))dx - \int _\Omega s(\rho_0)dx.$$
The next step is to use (H2a) in combination with \( w^\varepsilon_h = s'(u(w^\varepsilon_h)) \), which yields that
\[
\sum_{i,j=1}^{N} \nabla (w^\varepsilon_h)_i \cdot A_{ij}(u(w^\varepsilon_h)) \nabla (u(w^\varepsilon_h))_j = \sum_{i,j=1}^{N} \nabla (s'(u(w^\varepsilon_h)))_i \cdot A_{ij}(u(w^\varepsilon_h)) \nabla (u(w^\varepsilon_h))_j
\]
\[
= \sum_{i,j,k=1}^{N} \nabla (u(w^\varepsilon_h))_k \cdot (s''(u(w^\varepsilon_h)))_{ki} A_{ij}(u(w^\varepsilon_h)) \nabla (u(w^\varepsilon_h))_j \geq \gamma |\nabla u(w^\varepsilon_h)|^2,
\]
where \( |\nabla u(w^\varepsilon_h)|^2 := \sum_{\ell=1}^{d} |\partial u_{\ell}(w^\varepsilon_h)|^2 \). Moreover, due to (H2b) and \( w^\varepsilon_h = s'(u(w^\varepsilon_h)) \), we have
\[
w^\varepsilon_h \cdot f(u(w^\varepsilon_h)) = s'(u(w^\varepsilon_h)) \cdot f(u(w^\varepsilon_h)) \leq C_f.
\]
Therefore, we can conclude the entropy estimate
\[
\frac{\varepsilon}{\sigma} \|w^\varepsilon_h\|_{H^1(Q_T)}^2 + \int_{\Omega} s(\rho^\varepsilon(T))dx + \gamma \int_{Q_T} |\nabla \rho^\varepsilon|^2dxdt \leq \int_{\Omega} s(\rho_0)dx + C_f |\Omega| T.
\]
Hence, \( \|w^\varepsilon_h\|_{H^1(Q_T)}^2 \) is uniformly bounded, because \( \sigma \leq 1 \). Thus, the Leray-Schauder theorem is applicable and yields that \( \Phi \) has a fixed point, and therefore the scheme (3.5) admits a solution. Using these calculations for \( \sigma = 1 \), it follows that every solution has to satisfy the entropy inequality (3.6). \( \quad \Box \)

### 3.2. Convergence of the numerical scheme as \( h \to 0 \)

We will show that, for a fixed \( \varepsilon > 0 \), the numerical scheme (3.5) converges as \( h \to 0 \).

**Proposition 3.6** (Convergence of the scheme for fixed \( \varepsilon > 0 \)). There exists \( w^\varepsilon \in H^1(Q_T)^N \) with \( \rho^\varepsilon := u(w^\varepsilon) \in L^2(0,T,H^1(\Omega)^N) \), and a sequence \( h_\ell \to 0 \) such that
\[
\rho^\varepsilon_{h_\ell} := u(w^\varepsilon_{h_\ell}) \to \rho^\varepsilon \text{ strongly in } L^r(Q_T) \text{ for all } r \in [1,\infty).
\]
Moreover, \( w^\varepsilon \) solves (3.4) and satisfies the entropy estimate
\[
\frac{\varepsilon}{\sigma} \|w^\varepsilon\|_{H^1(Q_T)}^2 + \int_{\Omega} s(\rho^\varepsilon(T))dx + \gamma \int_{Q_T} |\nabla \rho^\varepsilon|^2dxdt \leq \int_{\Omega} s(\rho_0)dx + C_f |\Omega| T. \tag{3.8}
\]
Proof. The first part of the assertion follows from the fact that \( w^\varepsilon_{h_\ell} \) is uniformly bounded in \( H^1(Q_T)^N \), which yields that there exists \( w^\varepsilon \in H^1(Q_T)^N \) and subsequence \( h_\ell \to 0 \) such that \( w^\varepsilon_{h_\ell} \to w^\varepsilon \) in \( H^1(Q_T)^N \), due to the Banach-Alaoglu theorem, and \( w^\varepsilon_{h_\ell} \to w^\varepsilon \) in \( L^2(Q_T)^N \), due to Rellich’s theorem. In particular, we can choose this subsequence in such a way that \( w^\varepsilon_{h_\ell} \) converges a.e. to \( w^\varepsilon \). As \( u \) is bounded (see Assumption (H2)), the dominated convergence theorem entails the strong convergence of \( \rho^\varepsilon_{h_\ell} \equiv u(w^\varepsilon_{h_\ell}) \to u(w^\varepsilon) =: \rho^\varepsilon \) in \( L^r(Q_T)^N \) for all \( r \in [1,\infty) \). Combining this with the entropy estimate (3.6), there exists another subsequence (which we do not relabel) such that \( \rho^\varepsilon_{h_\ell} \to \rho^\varepsilon \) weakly in \( L^2(0,T,H^1(\Omega)^N) \).

Finally, owing to assumption (H4), for every \( \phi \in H^1(\Omega)^N \), there exists \( \phi_{h_\ell} \in V_{h_\ell} \) such that \( \phi_{h_\ell} \to \phi \) in \( H^1(\Omega)^N \). Using \( \phi_{h_\ell} \) as a test function in (3.4), we obtain (3.4) in the limit \( h_\ell \to 0 \), as each integral in (3.5) converges separately. The entropy inequality (3.8) is a consequence of Fatou’s lemma and the weak lower semicontinuity of the norm. \( \quad \Box \)

**Corollary 3.7.** Let \( \tau,\delta \geq 0 \) be such that \( \tau + \delta \leq T \). It holds true that
\[
\frac{\varepsilon}{\sigma} \|w^\varepsilon\|_{H^1(Q_T)}^2 + \frac{1}{\delta} \int_{\tau}^{\tau+\delta} \int_{\Omega} s(\rho^\varepsilon)dxdt + \gamma \left( 1 + \frac{\sqrt{\varepsilon}}{\delta} \right) \int_{\tau}^{\tau+\delta} \int_{\Omega} |\nabla \rho^\varepsilon|^2dxdt \leq \left( 1 + \frac{\sqrt{\varepsilon}}{\delta} \right) \int_{\Omega} s(\rho_0)dx + C_f |\Omega| \left( \tau + \frac{\delta}{2} + \frac{\sqrt{\varepsilon}}{\delta} T \right),
\]
where \( \rho^\varepsilon := u(w^\varepsilon) \) and \( Q_T := (0,\tau) \times \Omega \).
Thus, \(w^e \psi \in H^1(Q_T)^N\). Similarly as in the proof of Proposition 3.3, we use \(\rho^e := u(w^e)\) and
\[
\partial_t (\psi w^e) \cdot \rho^e = \partial_t (\psi w^e \cdot \rho^e) - \psi w^e \cdot \partial_t \rho^e = \partial_t (\psi w^e \cdot \rho^e - \psi s(\rho^e)) + \partial_t \psi s(\rho^e)
\]
and, since \(\psi(T) = 0\) and \(\psi(0) = 1\),
\[
\int_{Q_T} \partial_t (w^e \psi) \cdot \rho^e \, dx \, dt + \int_{Q_T} w^e(0) \cdot \rho_0 \, dx = \int_{Q_T} \partial_t \psi s(\rho^e) \, dx \, dt + \int_{Q_T} (s(\rho^e(0)) + w^e(0) \cdot (\rho_0 - \rho^e(0))) \, dx.
\]
Thus, using the definition of \(\psi\), and treating the last term of the previous equation as in the proof of Proposition 3.3, we get
\[
\int_{Q_T} \partial_t (w^e \psi) \cdot \rho^e \, dx \, dt + \int_{Q_T} \psi(0) w^e(0) \cdot \rho_0 \, dx + \frac{1}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} s(\rho^e) \, dx \, dt \leq \int_{Q_T} s(\rho_0) \, dx.
\]
From (3.4) tested with \(\phi = \psi w^e\) and the previous inequality, we get
\[
\varepsilon(\psi w^e, w^e)_{H^1_\rho(Q_T)^N} + \frac{1}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} s(\rho^e) \, dx \, dt + \frac{1}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} |\nabla \rho|^2 \, dx \, dt \leq \int_{Q_T} s(\rho_0) \, dx + C_f |\Omega| (\tau + \delta/2).
\]
Finally, we can estimate the first term as
\[
\varepsilon(\psi w^e, w^e)_{H^1_\rho(Q_T)^N} = \varepsilon \int_{Q_T} \int_{Q_T} (\psi |w^e|^2 + \psi |\nabla \psi|^2 + \varepsilon \partial_t (\psi w^e) \cdot \partial_t w^e) \, dx \, dt
\]
\[
= \varepsilon \int_{Q_T} \int_{Q_T} (\psi |w^e|^2 + \psi |\nabla \psi|^2 + \varepsilon \partial_t |w^e|^2 + \varepsilon \partial_t \psi w^e \cdot \partial_t w^e) \, dx \, dt
\]
\[
\geq \varepsilon \|w^e\|_{H^1_\rho(Q_T)^N}^2 - \frac{\varepsilon}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} |w^e|^2 \, dx \, dt.
\]
Using the Cauchy-Schwarz inequality and the definition of the \(H^1_\delta\) norm yields
\[
\frac{\varepsilon^2}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} |w^e|^2 \, dx \, dt \leq \frac{\varepsilon^2}{\delta} \|w^e\|_{L^2((\tau, \tau + \delta) \times \Omega)^N}^2 \|\partial_t w^e\|_{L^2((\tau, \tau + \delta) \times \Omega)^N} \leq \frac{\varepsilon^3}{\delta} \|w^e\|_{H^1_\rho((\tau, \tau + \delta) \times \Omega)^N}^2,
\]
and therefore
\[
\varepsilon \|w^e\|_{H^1_\rho(Q_T)^N} + \frac{1}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} s(\rho^e) \, dx \, dt + \frac{1}{\delta} \int_{\tau}^{\tau + \delta} \int_{Q_T} |\nabla \rho|^2 \, dx \, dt
\]
\[
\leq \frac{\varepsilon^3}{\delta} \|w^e\|_{H^1_\rho((\tau, \tau + \delta) \times \Omega)^N}^2 + \int_{\Omega} s(\rho_0) \, dx + C_f |\Omega| (\tau + \delta/2).
\]
Note that we cannot estimate the first term on the right-hand side by the first term on the left-hand side, because the domain of the norms are disjoint. Fortunately, we have the entropy estimate (3.8), which we add \( \sqrt{\varepsilon / \delta} \) times to this inequality to get

\[
\varepsilon \| w^\varepsilon \|_{H^1(\Omega)^N} + \frac{1}{\delta} \int_\Omega \int_T^{T+\delta} s(\rho^\varepsilon)dxdt + \frac{\varepsilon}{\delta} \int_\Omega \int_T^{T+\delta} s(\rho^\varepsilon(T))dxdt + \gamma(1 + \frac{\varepsilon}{\delta}) \int_0^T \int_\Omega |\nabla \rho^\varepsilon|^2dxdt \\
\leq (1 + \frac{\varepsilon}{\delta}) \int_0^T \int_\Omega s(\rho_0)dx + C_f|\Omega| \left( \tau + \frac{\delta}{2} + \frac{\varepsilon}{\delta} T \right).
\]

which, since \( s(\rho^\varepsilon(T)) \geq 0 \), implies the assertion.

3.3. Limit of \( \varepsilon \to 0 \). We consider the limiting problem

\[
(3.10) \quad -\int_\Omega \phi(0) \cdot \rho_0 dx - \int_0^T \int_\Omega \partial_t \phi \cdot \rho dxdt + \sum_{i,j=1}^N \int_0^T \int_\Omega \nabla \phi_i \cdot A_{ij}(\rho) \nabla \rho_j dxdt = \int_0^T \int_\Omega \phi \cdot f(\rho)dxdt
\]

for all \( \phi \in (H^1(\Omega)^N) \) with \( \phi(T) = 0 \). As above, we use the notation \( \phi(t) := \text{tr}(\phi)(t, \cdot) \), where \( \text{tr} \) denotes the trace operator \( \text{tr} : H^1(\Omega)^N \to L^2(\{0, T\} \times \Omega)^N \).

**Proposition 3.8.** Let \( \tau, \delta \geq 0 \) such that \( \tau + \delta \leq T \). Set \( \rho^\varepsilon := u(w^\varepsilon) \). Then there exist \( \rho \in L^2(0, T; H^1(\Omega)^N) \) with \( \rho(t, x) \in \mathcal{D} \) for a.e. \( (t, x) \in Q_T \) being a solution of (3.10) and a subsequence \( \varepsilon_j \to 0 \) such that

\[
\rho^\varepsilon \rightharpoonup \rho \quad \text{in every } L^r(\Omega)^N, r \in [1, \infty), \text{ as } \varepsilon_j \to 0.
\]

Moreover, \( \rho \) satisfies the entropy inequality

\[
\frac{1}{\delta} \int_\Omega \int_T^{T+\delta} s(\rho)dxdt + \gamma \int_0^T \int_\Omega |\nabla \rho|^2dxdt \leq \int_\Omega s(\rho_0)dx + C_f|\Omega| (\tau + \delta/2).
\]

In the proof of Proposition 3.8, the key ingredient to prove strong convergence of (at least a subsequence of) \( \rho^\varepsilon \) will be the idea of compensated compactness, which is a special technique applying the classical div-curl lemma; see, e.g. [48, Lemma 7.2].

**Lemma 3.9** (div-curl lemma). Let \( \alpha, \alpha^\ell \in L^2(Q_T)^{1+d} \) and \( \beta, \beta^\ell \in L^2(Q_T)^{1+d} \). Then

\[
\alpha^\ell \rightharpoonup \alpha \quad \text{in } L^2(Q_T)^{1+d} \text{ as } \ell \to +\infty, \quad \text{and} \quad \langle \text{div}(t, x) \alpha^\ell \rangle_{t \in \mathbb{N}} \text{ is bounded in } L^2(Q_T),
\]

\[
\beta^\ell \rightharpoonup \beta \quad \text{in } L^2(Q_T)^{1+d} \text{ as } \ell \to +\infty, \quad \text{and} \quad \langle \text{curl}(t, x) \beta^\ell \rangle_{t \in \mathbb{N}} \text{ is bounded in } L^2(Q_T)^{(1+d) \times (1+d)}
\]

implies that

\[
\alpha^\ell \cdot \beta^\ell \rightharpoonup \alpha \cdot \beta \quad \text{in } \mathcal{D}'(Q_T) \quad \text{as } \ell \to +\infty,
\]

where \( \mathcal{D}'(Q_T) \) denotes the dual space of \( \mathcal{D}(Q_T) := C_c^\infty(Q_T) \).

**Proof of Proposition 3.8.** Let \( w^\varepsilon, \rho^\varepsilon := u(w^\varepsilon) \) denote the solution of (3.4) satisfying the entropy inequality (3.8). For any fixed \( i, j = 1, \ldots, N \), we define the vector-valued functions with \((1 + d)\) components

\[
\alpha^\varepsilon = \begin{pmatrix} \rho_i^\varepsilon - \varepsilon^2 \partial_t w_i^\varepsilon \\ J_i^\varepsilon - \varepsilon \nabla w_i^\varepsilon \end{pmatrix} \quad \text{and} \quad \beta^\varepsilon := \begin{pmatrix} \rho_i^\varepsilon \\ 0 \end{pmatrix}, \quad \text{where} \ J_i^\varepsilon = -\sum_{j=1}^N A(\rho^\varepsilon)_{ij} \nabla \rho_j^\varepsilon.
\]

Note that, by assumption, \( \mathcal{D} \) is bounded and so is \( \rho^\varepsilon = u(w^\varepsilon) \). Thus, thanks to the entropy estimate (3.8), \( \alpha^\varepsilon, \beta^\varepsilon \) are bounded uniformly in \( L^2(Q_T)^{1+d} \) w.r.t. \( \varepsilon \in (0, 1) \). By the Banach-Alaoglu theorem, there exist \( \alpha, \beta \in L^2(Q_T)^{1+d} \) and a subsequence \( \varepsilon_{\ell} \to 0 \) such that

\[
\alpha^\varepsilon \rightharpoonup \alpha, \quad \beta^\varepsilon \rightharpoonup \beta \quad \text{in } L^2(Q_T)^{1+d} \quad \text{as } \varepsilon_{\ell} \to 0.
\]
Clearly, $\beta$ has the form $(\rho_1,0)$ for some $\rho_1 \in L^2(Q_T)$. Due to the entropy estimate (3.8), $\sqrt{\varepsilon} \| w^\varepsilon \|_{H^1_0(Q_T)}$ is bounded. Hence, $\beta_0^\varepsilon - \alpha_0^\varepsilon = \varepsilon^2 \partial_i w_i^\varepsilon \to 0$ in $L^2(Q_T)$ as $\varepsilon \to 0$, implying that $\rho_i := \beta_0 = \alpha_0$ and $\alpha \cdot \beta = \rho_i^2$, where in this context the index 0 denotes the first component of the $(1+d)$-dimensional vector. Moreover, one can easily show that

$$\| \text{curl}_{(t,x)} \beta^\varepsilon \|_{L^2(Q_T)^{(1+d)\times(1+d)}} \leq C \| \nabla \rho_i^\varepsilon \|_{L^2(Q_T)^d}$$

for some $C > 0$. Again by the entropy estimate (3.8), this implies that $\text{curl}_{(t,x)} \beta^\varepsilon$ is uniformly bounded\(^2\) in $L^2(Q_T)^{(1+d)\times(1+d)}$ w.r.t. $\varepsilon \in (0,1)$. In order to apply the div-curl lemma, it remains to prove that the space-time divergence of $\alpha^\varepsilon$ is bounded. For this, we require the equation for $\rho_i^\varepsilon$ in the interior of $Q_T$, i.e., from equation (3.4),

$$\varepsilon \int_{Q_T} \psi w_i^\varepsilon dx dt + \varepsilon^2 \int_{Q_T} \partial_i \psi \partial_i w_i^\varepsilon dx dt + \varepsilon \int_{Q_T} \nabla \psi \cdot \nabla w_i^\varepsilon dx dt - \int_{Q_T} \partial_t \psi \rho_i^\varepsilon dx dt$$

$$+ \sum_{j=1}^N \int_{Q_T} \nabla \psi \cdot A_{ij} (\rho^\varepsilon) \nabla \rho_j^\varepsilon dx dt = \int_{Q_T} \psi f_i (\rho^\varepsilon) dx dt$$

for all $\psi \in H^1_0(Q_T)$. We can rewrite this by using the weak space-time divergence of $\alpha^\varepsilon$ as

$$- \int_{Q_T} \nabla (t,x) \psi \cdot \alpha^\varepsilon dx dt = \int_{Q_T} \partial_t \psi (\varepsilon^2 \partial_i w_i^\varepsilon - \rho_i^\varepsilon) dx dt$$

$$+ \int_{Q_T} \nabla \psi \cdot \left( \varepsilon \nabla w_i^\varepsilon + \sum_{j=1}^N A_{ij} (\rho^\varepsilon) \nabla \rho_j^\varepsilon \right) dx dt$$

$$= \int_{Q_T} \psi f_i (\rho^\varepsilon) dx dt - \varepsilon \int_{Q_T} \psi w_i^\varepsilon dx dt$$

for all $\psi \in H^1_0(Q_T)$. We observe that the right-hand side defines a bounded operator on $L^2(Q_T)$ due to the entropy estimate (3.8) and the fact that $f_i$ is uniformly bounded as a continuous function defined on a compact set (see (H2)). This yields that $\text{div}_{(t,x)} \alpha^\varepsilon$ is uniformly bounded in $L^2(Q_T)$. Therefore, we can apply the div-curl lemma and obtain that

$$(\rho_i^\varepsilon - \varepsilon^2 \partial_i w_i^\varepsilon) \rho_i^\varepsilon = \alpha^\varepsilon \cdot \beta^\varepsilon \to \alpha \cdot \beta = \rho_i^2$$

in $\mathcal{D}'(Q_T)$ as $\varepsilon \to 0$.

Using that $\rho_i^\varepsilon \to \rho_i$ and $\varepsilon^2 \partial_i w_i^\varepsilon \to 0$ in $L^2(Q_T)$, we obtain that

$$\int_{Q_T} (\rho_i^\varepsilon)^2 \phi^2 dx dt \to \int_{Q_T} \rho_i^2 \phi^2 dx dt \quad \text{as } \varepsilon \to 0$$

for all $\phi \in C_c^\infty(Q_T)$. Hence, $\phi \rho_i^\varepsilon \to \phi \rho_i$ in $L^2(Q_T)$ for all $\phi \in C_c^\infty(Q_T)$. In particular, there exists a subsequence not being relabeled such that $\rho_i^\varepsilon \to \rho_i$ a.e. in $Q_T$. For almost every $(t,x) \in Q_T$, we know that $\rho_i^\varepsilon (t,x) \in \mathcal{D}$ and that $\mathcal{D}$ is bounded. Thus, we can apply the dominated convergence theorem, which yields that

$$\rho_i^\varepsilon \to \rho_i \quad \text{in every } L^r(Q_T), r \in [1, \infty), \text{ as } \varepsilon \to 0,$$

and that $\rho(t,x) \in \mathcal{D}$ for almost every $(t,x) \in Q_T$.

Moreover, the entropy inequality (3.8) also states that $\nabla \rho_i^\varepsilon$ is bounded in $L^2(Q_T)^d$ independently of $\varepsilon$. Since $|\rho^\varepsilon| = |u(w^\varepsilon)| = |(s')^{-1}(w^\varepsilon)| \leq \sup_{t \in [0,T]} |u|^2$, according to (H2), then, using

\(^2\)The fact that the $L^2$ norm of $\nabla \rho_i^\varepsilon$ is uniformly bounded is ultimately a consequence of our hypothesis (H2a), i.e., the matrix $s''(\rho)A(\rho)$ being coercive. Using instead the original assumptions made by A. Jüngel in [26], one can only assure that $\nabla \rho_i^\varepsilon$ is bounded in $L^2$ for some $m$. However, one can circumvent this issue by defining $\beta^\varepsilon := (\rho_i^m,0)^T$ in this case.
again (3.8), we obtain
\[ \|\rho_i^\varepsilon\|_{L^2(0, T; H^1(\Omega))}^2 = \int_{Q_T} (\rho_i^\varepsilon)^2 dx dt + \int_{Q_T} |\nabla \rho_i^\varepsilon|^2 dx dt \]
\[ \leq |\Omega| T \|\rho_i^\varepsilon\|_{L^\infty(Q_T)}^2 + \frac{1}{\gamma} \left( \int_{\Omega} s(\rho_0) dx + C_f |\Omega| T \right) \]
\[ \leq \frac{1}{\gamma} \int_{\Omega} s(\rho_0) dx + \left( \sup_{v \in D} |\nabla v|^2 + \frac{C_f}{\gamma} \right) |\Omega| T, \]

namely, \( \rho_i^\varepsilon \) is bounded in \( L^2(0, T; H^1(\Omega)) \) independent on \( \varepsilon \). Taking yet another subsequence, which we do not relabel, we can see that there exists \( \rho_i \in L^2(0, T; H^1(\Omega)) \) such that \( \rho_i^\varepsilon \rightharpoonup \rho_i \) in \( L^2(0, T; H^1(\Omega)) \). In particular, \( \nabla \rho_i^\varepsilon \rightharpoonup \nabla \rho_i \) in \( L^2(Q_T)^d \). We already have seen that \( \varepsilon \|w^\varepsilon\|_{H^1(Q_T)} \) is bounded, then \( \varepsilon \|w^\varepsilon\|_{H^1(Q_T)} \to 0 \) implying \( \varepsilon(w^\varepsilon, \phi)_{H^1(Q_T)} \to 0 \).

Now, we prove that \( \rho \) is solution to the limiting problem (3.10). Let \( \phi \in H^1(Q_T) \) with trace \( \phi(T) = 0 \). Using that \( A \) is bounded, according to (H1), the dominated convergence theorem yields
\[ \int_{Q_T} |\nabla \phi| A_{ij}(\rho^\varepsilon) |A_{ij}(\rho)| \, dx dt \to \int_{Q_T} |\nabla \phi| A_{ij}(\rho) \, dx dt \quad \text{as} \quad \varepsilon \to 0. \]

In particular, \( \nabla \phi A_{ij}(\rho^\varepsilon) \) converges strongly in \( L^2(Q_T)^d \). For each \( i = 1, \ldots, N \), we test the equation for \( \rho_i^\varepsilon \) (see (3.4)) with functions \( \phi \in H^1(Q_T) \) with trace \( \phi(T) = 0 \), take the limit for \( \varepsilon = \varepsilon_\ell \to 0 \), and obtain
\[ - \int_{\Omega} \phi(0) \rho_i^\varepsilon \, dx - \int_0^T \int_{\Omega} \partial_t \phi \rho_i \, dx dt + \sum_{j=1}^N \int_0^T \int_{\Omega} \nabla \phi \cdot A_{ij}(\rho) \nabla \rho_j \, dx dt \]
\[ = \int_0^T \int_{\Omega} \phi f_i(\rho) \, dx dt \]

for all \( i = 1, \ldots, N \).

Finally, recall that \( \rho^\varepsilon \) satisfies the entropy estimate (3.9) from Corollary 3.7. Thus, we obtain the entropy inequality (3.11) as a direct consequence of the lower weak continuity of the \( L^2 \) norm and the Fatou lemma.

### 3.4. Existence of a weak solution.

In this section, we prove that problem (2.1) possesses a weak solution \( \rho \) in the sense of Definition 3.1. Moreover, we prove the equivalence stated in Lemma 3.2 between the weak formulation (3.2) in Definition 3.1 and the weak formulation (2.2).

**Proposition 3.10.** Let \( \rho \) be given by Proposition 3.8. Then \( \rho \in H^1(0, T; (H^1(\Omega))^N) \) and \( \rho \in C^0([0, T]; L^2(\Omega)) \) with \( \rho(0) = \rho_0 \). Moreover, it satisfies the entropy inequality
\[ \int_{\Omega} s(\rho(\tau)) \, dx + \gamma \int_0^\tau \int_{\Omega} |\nabla \rho|^2 \, dx dt \leq \int_{\Omega} s(\rho_0) \, dx + C_f |\Omega| \tau. \]

for almost all \( \tau \in (0, T) \).

**Proof.** Using the equation (the equation (3.10)), we obtain that
\[ \left| \int_{Q_T} \partial_t \phi \rho \, dx dt \right| \leq \sum_{j=1}^N \int_{Q_T} |\nabla \phi| A_{ij}(\rho) \nabla \rho_j \, dx dt + \int_{Q_T} |\phi| |f_i(\rho)| \, dx dt + \int_{\Omega} |\phi(0)| \rho_{0,i} \, dx \]
\[ \leq C \rho \|\phi\|_{L^2(0, T; H^1(\Omega))}, \]

since \( \rho \in L^\infty(Q_T) \cap L^2(0, T; H^1(\Omega)). \) This implies that, for each \( i = 1, \ldots, N \), \( \rho_i \) has a weak time derivative satisfying \( \partial_t \rho_i \in L^2(0, T; H^1(\Omega)) \). Then the embedding \( H^1(0, T; H^1(\Omega)) \cap L^2(0, T; H^1(\Omega)) \subset C^0([0, T]; L^2(\Omega)) \), entails that every \( \rho_i \) is continuous in time, and so is \( \rho \). We obtain the desired entropy estimate as a limit \( \delta \to 0 \) of (3.11).
It remains to show that \( \rho(0) = \rho_0 \) in \( L^2(\Omega)^N \). For this, let \( \psi \in H^1(\Omega)^N \) and, for \( \tau \in (0, T) \), define

\[
\phi_\tau(t, \cdot) := \begin{cases} 
(1 - \frac{\tau}{\tau}) \psi(\cdot) & \text{in } \Omega \times [0, \tau], \\
0 & \text{in } \Omega \times (\tau, +\infty).
\end{cases}
\]

We easily see that \( \phi_\tau \to 0 \) in \( L^2(0, T; H^1(\Omega)^N) \) as \( \tau \to 0 \). Then, from equation (3.10) tested with \( \phi_\tau \), we get that, for all \( \psi \in H^1(\Omega)^N \),

\[
\int_\Omega \left( \frac{1}{\tau} \int_0^\tau \rho dt - \rho_0 \right) \psi dx \to 0 \quad \text{as } \tau \to 0.
\]

Finally, the continuity implies that \( \lim_{\tau \to 0} \frac{1}{\tau} \int_0^\tau \rho dt = \rho(0) \), which entails \( \rho(0) = \rho_0 \).

**Remark 3.11.** Using the last part of the proof of Proposition 3.10, we can easily show that any solution \( \rho \) of (3.2) satisfies \( \rho(0) = \rho_0 \). Therefore, the proof of Lemma 3.2 is a straightforward application of the integration by parts formula and of the embedding (3.1).

**Corollary 3.12.** Let \( \rho \) be given by Proposition 3.8. Then \( \rho \) is a solution of (2.2).

**Proof.** Thanks to Proposition 3.10, we know that \( \rho \) possesses enough regularity such that we can integrate in (3.10) w.r.t. \( t \), which yields (2.2) for all \( \phi \in H^1(Q_T)^N \) with \( \phi(T) = 0 \). Using a density argument yields the assertion.

The proof of Proposition 3.4 is now straightforward.

**Proof of Proposition 3.4.** We only have to collect the previous results to obtain the proposition using a diagonal sequence argument.

### 4. Applications and numerical tests

In this section, we apply the general setting of section 2 and numerically test the space-time Galerkin method of section 3 by considering four problems: the (linear) heat equation (section 4.1), the porous medium equation (section 4.2), the Fisher-KPP equation (section 4.3), and the Maxwell-Stefan system in the case of \( N = 2 \) species (section 4.4). For the Maxwell-Stefan system, the discussion of the general setting and of an alternative space-time Galerkin method for the case of \( N > 2 \) is postponed to section 5. We remark that we apply this nonlinear setting to the linear heat equation for validation purposes and, in particular, in order to stress its unconditional stability on a simple test problem.

In all cases, we consider the entropy density \( s : \mathcal{D} \to [0, +\infty) \) defined by

\[
s(\rho) = \sum_{j=1}^N \rho_j \log \rho_j + \left( 1 - \sum_{j=1}^N \rho_j \right) \log \left( 1 - \sum_{j=1}^N \rho_j \right) + \log(N + 1),
\]

where \( \mathcal{D} := \{ \rho \in (0, 1)^N : \sum_{i=1}^N \rho_i < 1 \} \). We have

\[
(s'(\rho))_\ell = \frac{\rho_\ell}{1 - \sum_{j=1}^N \rho_j} \quad \text{and} \quad (s''(\rho))_{\ell\ell} = \frac{\delta_{\ell\ell}}{1 - \sum_{j=1}^N \rho_j} - \frac{1}{1 - \sum_{j=1}^N \rho_j}.
\]

Then \( s \in C^2(\mathcal{D}, [0, \infty)) \cap C^0(\overline{\mathcal{D}}) \) and is convex. Moreover, \( u : \mathbb{R}^N \to \mathcal{D} \) defined as

\[
u_\ell(w) = \frac{e^{w_\ell}}{1 + \sum_{i=1}^N e^{w_i}} \quad \text{for } \ell = 1, \ldots, N,
\]

a choice first used in [9] to investigate the case \( N = 2 \), is in \( C^1(\mathbb{R}^N, \mathcal{D}) \), and is the inverse of \( s' \). Thus, the preamble of assumption (H2) is satisfied.

In the numerical experiments below, we use continuous space-time finite element discretization spaces. On the space-time cylinder \( Q_T = \Omega \times (0, T) \), with \( \Omega \) bounded interval \( (d = 1) \) or Lipschitz polytope \( (d > 1) \), we consider families of shape-regular simplicial or Cartesian meshes \( \{ T_h \}_{h>0} \).
The parameter $h$ denotes the mesh granularity, namely $T_h = \{K_i, i = 1, \ldots, N_h\}$, $h_K := \text{diam}(K)$, and $h := \max_{K \in T_h} h_K$.

As discretization spaces, we choose $\{V_h\}_{h > 0} = \{V^p_h, p \in \mathbb{N}\}_{h > 0}$, with
\[
V^p_h = \{v \in C^0(\mathcal{Q}_T)^N : v_K \in \mathcal{P}_p(K)^N \ \forall K \in T_h\},
\]
where $\mathcal{P}_p(K)$ denotes the space of polynomial functions on $K$ of degree at most $p$, if $K$ is a simplex, or of degree at most $p$ in each variable, if $K$ is a cuboid. Therefore, the approximability assumption (H4) in the first part of section 3 is satisfied.

Defining $B : \mathbb{R}^N \to \mathbb{R}^{N \times N}$ as
\[
B(w) = A(u(w))w'(w),
\]
the space-time Galerkin method (3.5) can be rewritten more explicitly in terms of the entropy variable unknown as follows:
\[
\begin{aligned}
\varepsilon \langle \phi, w_h^N \rangle_{(\mathcal{H}^1)^N(\mathcal{Q}_T)} + \int_\Omega \phi(T) \cdot u(w_h^N(T))dx - \int_\Omega \phi(0) \cdot \rho_0dx - \int_{\mathcal{Q}_T} \partial_t \phi \cdot u(w_h^N)dxdt \\
+ \sum_{i,j=1}^N \int_{\mathcal{Q}_T} \nabla \phi_i \cdot B_{ij}(w_h^N) \nabla (w_h^N_j)dxdt = \int_{\mathcal{Q}_T} \phi \cdot f(u(w_h^N))dxdt
\end{aligned}
\]
for all $\phi \in V_h^p$.

Throughout this section, we measure the absolute numerical error defined by $||\rho - u(w_h^N)||_{L^2(\mathcal{Q}_T)}$.

**Remark 4.1** (On the choice of $\varepsilon$). For the analysis the regularization term with the factor $\varepsilon > 0$ is crucial, as it delivers essential bounds on the entropy variable $w$. In the numerical examples we will specify the choice of $\varepsilon$ for every example, and show on several occasions that it is possible to choose $\varepsilon = 0$. However, it is crucial to note that in these examples the entropy variable representing the exact solution has ‘nice’ bounds. Conversely, when the solution approaches the singularities of the entropy, it is required to choose $\varepsilon > 0$ large enough for the solver to converge. In general, it is best to choose $\varepsilon$ as small as possible, as we observe in the second example in subsection 4.2.

### 4.1. Heat equation

We apply our general approach to the linear heat equation:
\[
\begin{cases}
\partial_t \rho = \Delta \rho & \text{in } \Omega, \ t > 0, \\
\partial_n \rho = 0 & \text{on } \partial \Omega, \ t > 0, \\
\rho(0) = \rho_0 & \text{in } \Omega.
\end{cases}
\]

This corresponds to problem (2.1) with $N = 1$, $A \equiv 1$, and $f \equiv 0$. Furthermore, $\mathcal{D} = (0, 1)$ and the entropy density $s : \mathcal{D} \to [0, +\infty)$ is given by
\[
s(\rho) = \rho \log \rho + (1 - \rho) \log(1 - \rho) + \log(2),
\]
and thus $s'(\rho) = \log \frac{\rho}{1 - \rho}$, and $s''(\rho) = \frac{1}{(1 - \rho)^2}$.

For this choice of $A(\rho)$ and $f(\rho)$, assumption (H1) is obviously satisfied, and assumptions (H2a) and (H2b) are fulfilled with $\gamma = 4$ and $C_f = 0$.

For the numerical tests, we take $\Omega = (0, 1)^2$ and $\rho_0(\mathbf{x}) = 0.5 \cos(\pi x_1) \cos(\pi x_2) + 0.5$, so that the problem has the analytical solution given by
\[
\rho(t, \mathbf{x}) = 0.5 \exp(-8\pi^2 t/\tau) \cos(2\pi x_1) \cos(2\pi x_2) + 0.5,
\]
where we use $\tau = 7$ to rescale the time. The solution is shifted and scaled in order to avoid the singularities of $s'$ at 0 and 1. Without this rescaling, the system matrix is highly ill-conditioned,
which prohibits optimal convergence rates. We solve (4.3), setting $\varepsilon = 0$ and solving the nonlinearity by Newton’s method. We use unstructured space-time simplicial meshes. The Newton method converges in 6 steps, for all considered values of $h$ and $p$. We measure the $L^2$ error on the whole space-time domain. In Figure 1, the convergence rates of the $h$- and the $p$-version of the method are shown. We observe optimal rates, exponential in $p$ and of order $p + 1$ in $h$. In the case of $p = 4$, we observe a preasymptotic region for very large mesh sizes; the exact rates are shown in Table 1.

![Fig. 1. Convergence rates for the space-time Galerkin approximation towards the exact solution of the heat equation, in polynomial degree $p$ (left), and mesh size $h$ (middle). On the right we are plotting the entropy on a logarithmic scale, showing exponential convergence.](image)

Table 1

| $h$   | $p = 3$       |       | $p = 4$       |       |
|-------|---------------|-------|---------------|-------|
|       | $h$ | error | rate | $h$ | error | rate |
| $2^{-1}$ | 2.3 | $10^{-3}$ | 0   | $2^{-1}$ | 4.8 | $10^{-4}$ | 0   |
| $2^{-2}$ | 3.1 | $10^{-4}$ | 2.9 | $2^{-2}$ | 3.0 | $10^{-5}$ | 4.0 |
| $2^{-3}$ | 2.4 | $10^{-5}$ | 3.7 | $2^{-3}$ | 2.3 | $10^{-6}$ | 3.7 |
| $2^{-4}$ | 2.1 | $10^{-6}$ | 3.5 | $2^{-4}$ | 5.9 | $10^{-8}$ | 5.3 |
| $2^{-5}$ | 1.3 | $10^{-7}$ | 4.0 | $2^{-5}$ | 1.8 | $10^{-9}$ | 5.0 |
| $2^{-6}$ | 8.4 | $10^{-9}$ | 4.0 | $2^{-6}$ | 5.7 | $10^{-11}$ | 5.0 |

![Fig. 2. Comparison of a mesh made from time slabs (left) and an adapted space-time mesh (middle). The convergence of the two methods with respect to the number of degrees of freedom is shown on the right for $p = 1$.](image)

In Figure 2 we highlight another feature of the space-time approach, namely the ability to use adapted mesh in space and time. As a comparison, we use a mesh made of time slabs, a mesh structure similar to what would result from classic time-stepping methods. The timeslab height is given by $h_t \approx h_x^2$, with $h_x$ being the mesh size of the spatial mesh. For the space-time adapted mesh, we start with a unstructured simplicial mesh of size $h = 0.2$ and then apply...
adaptive refinement. For this simple example, we use a flux based error estimator and Dörfler marking with $\theta = 0.5$. We observe the same rate of convergence on both meshes, however using the space-time adapted mesh allows us to obtain a given accuracy with fewer degrees of freedom.

### 4.2. The porous medium equation

Let $m > 1$. The porous medium equation is given by

$$
\begin{align*}
\partial_t \rho &= \Delta \rho^m \quad \text{in } \Omega, t > 0, \\
\partial_n (\rho^m) &= 0 \quad \text{on } \partial \Omega, t > 0, \\
\rho(0) &= \rho_0 \quad \text{in } \Omega.
\end{align*}
$$

We can write it in the form of (2.1) for $N = 1$, $A(\rho) = m \rho^{m-1}$, and $f \equiv 0$. The entropy density is the same as for the heat equation.

**Proposition 4.2.** Assumptions (H1) and (H2) are satisfied for $m \in (1, 2]$.

**Proof.** For $\mathcal{D} = (0,1)$ and $m > 1$, $A(\rho) = m \rho^{m-1}$ is in $C^0(\mathcal{D})$, thus (H1) is satisfied. As (H2b) is obvious, we only need to prove that (H2a) is satisfied, namely that $s''(\rho)A(\rho) \geq \gamma$ for some $\gamma > 0$ and all $\rho \in \mathcal{D}$. Thus let $\rho \in (0,1) = \mathcal{D}$. Then, whenever $m \in (1, 2]$,

$$
s''(\rho)A(\rho) = \frac{m \rho^{m-1}}{\rho(1-\rho)} = \frac{m}{\rho^2(1-\rho)^{m-1}} \geq m \equiv \gamma.
$$

We test the space-time Galerkin method for this problem with initial conditions and Neumann boundary conditions chosen such that

$$
\rho(x,t) = \left[ \frac{(m-1)(x-\alpha)^2}{2m(m+1)(\beta-t)} \right]^{\frac{1}{m-1}}
$$

is the exact solution, with $\alpha$ and $\beta$ real parameters, on $\Omega = (0,1)$. We consider the case $m = 2$, $\alpha = 2$, $\beta = 5$ on unstructured simplicial space-time meshes.

In Figure 3, we show the convergence rates of the scheme. Regardless of the nonlinearity, we match the convergence rates of the heat equation, i.e. exponential in $p$ and of order $p + 1$ in $h$. The convergence rates in terms of $h$ are also considered for different values of $\varepsilon$. We observe that $\varepsilon$ introduces a lower bound on the error. Therefore, choosing it as small as possible, such that the solver still converges, gives the best results. On the other hand, in the next example, we can see that for certain solutions, that produce a very ill-conditioned system, we must choose $\varepsilon$ fairly large.

![Graph showing convergence rates](image)

**Figure 3.** Convergence rates towards the exact solution of the porous medium equation. Convergence in terms of the polynomial degree $p$ for different mesh sizes is shown on the left. We consider convergence in mesh size $h$ for different values of the polynomial order $p$ and fixed $\varepsilon = 0$ in the middle, and for fixed $p = 4$ and different values of regularization parameter $\varepsilon$ on the right.

In contrast to the heat equation, the power law in the porous medium equation introduces a finite propagation speed of the solution. This is best observed by the interesting behavior of
certain initial conditions that induce a waiting time. That is, the solution keeps a fixed support until the waiting time is reached. On $\Omega = (0, \pi)$, the initial condition given by

$$
\rho_0(x) = \begin{cases} 
\sin^{2/(m-1)}(x) & \text{if } 0 \leq x \leq \pi, \\
0 & \text{otherwise,}
\end{cases}
$$

produces this behavior. It is shown in [39] that the corresponding solution has a waiting time of

$$
t^* = \frac{m-1}{2m(m+1)}. 
$$

As we choose $m = 2$, here $t^* = 0.083$. We modify the initial condition to

$$
\rho_0(x) = 10^{-16} 
$$

for $x \notin [0, \pi]$ to avoid ill-conditioning. Furthermore, to ensure convergence of the Newton method used as a nonlinear solver, we had to choose $\varepsilon = 10^{-6}$, making use of the regularization term. We solve on a Cartesian space-time mesh until final time $T = 0.2$, with spatial mesh size $h_s = 0.05$, and temporal mesh size $h_t = h_s/2$, and fix $p = 5$. The results are shown in Figure 4. Looking at snapshots of the numerical solution we can observe that it keeps a compact support set. In Figure 4, on the right, we plot the value of the solution on the left interface against time, marking the expected waiting time $t^*$ with the vertical line.

![Figure 4](image-url)

**Fig. 4.** Snapshots of the solution of the porous medium equation emitting a waiting time, at different times (left) and the value at the left interface (right).

### 4.3. The Fisher-KPP equation.

We consider the Fisher-KPP equation

$$
\begin{aligned}
\partial_t \rho &= A \Delta \rho + \rho(1 - \rho) & \text{in } \Omega, \ t > 0, \\
A \partial_n \rho &= 0 & \text{on } \partial \Omega, \ t > 0, \\
\rho(0) &= \rho_0 & \text{in } \Omega,
\end{aligned}
$$

with $A > 0$ now constant. This agrees with formulation (2.1), with $N = 1$, $A(\rho) = A$, and $f(\rho) = \rho(1 - \rho)$. We set again $\mathcal{D} := (0, 1)$. Assumptions (H1) and (H2a) are clearly satisfied. Choosing an entropy density such that assumption (H2b) is satisfied with $C_f = 0$ allows for the right-hand side of the entropy estimate (3.6) to be independent of time. Motivated by this, we now investigate the rescaled entropy density $s : \mathcal{D} \to (0, +\infty)$ given by

$$
s(\rho) = \rho \log \rho + (n - \rho) \log(n - \rho),
$$

with $n$ to be chosen. Note that $f(\rho) > 0$ for $\rho \in (0, 1)$, and $n/\rho - 1 > 1$ if and only if $\rho < n/2$. Thus,

$$
f(\rho)s'(\rho) = \rho(1 - \rho) \log \frac{\rho}{n - \rho} = -\rho(1 - \rho) \log \left( \frac{n}{\rho} - 1 \right) \leq 0
$$

for all for $\rho \in (0, 1)$ if and only if $n \geq 2$. We choose $n = 2$ so that the hypothesis (H2b) is fulfilled with $C_f = 0$. 
We start again by investigate convergence towards a smooth solution. We choose $\Omega = (0, 1)$, and initial conditions and Neumann boundary conditions such that

$$
\rho(x,t) = \frac{1}{\left[1 + \exp\left(-\frac{5}{6}t + \frac{1}{\sqrt{6}}x\right)\right]^2}
$$

is the exact solution for $A = 1$. We set $\varepsilon = 0$ and solve on unstructured simplicial space-time meshes. The results are presented in Figure 5. We observe again optimal convergence rates in both $p$ and $h$, namely exponential in $p$ and of order $p + 1$ in $h$.

![Figure 5](image)

**Fig. 5.** Convergence rates in polynomial degree $p$ (left) and mesh size $h$ for the exact solution of the Fisher-KPP equation.

Next, we aim to reproduce the experiments presented in [5], considering an initial condition with a jump, given by $\rho_0(x) = 1$ if $0 < x < 1/2$ and 0 elsewhere, with diffusion coefficient $A = 10^{-4}$. We solve using $p = 3$ on a Cartesian mesh with $h_s = 0.025$, $h_t = 0.4$ up to $T = 8$. We choose $\varepsilon = 10^{-8}$ to avoid ill-conditioning in the solver.

![Figure 6](image)

**Fig. 6.** Snapshots of the numerical solution for the Fisher-KPP (left) and different choices of the entropy (right). The choices are as follows: Entropy 1 is the one used in [5], Entropy 2 is given by (4.4) with $n = 2$, and Entropy 3 is (4.4) with $n = 2.1$.

Snapshots of the numerical solution are taken every 1.3 seconds, the results are shown in Figure 6 on the left. In Figure 6 on the right, we consider different choices for the entropy up to $T = 15$. Note that at the point in time the solution has already converged to $\rho \equiv 1$. The choice for the entropy density in [5] was $\rho \log(\rho) - \rho + 1$. This choice is not covered by our assumptions, however, using it produces the correct results, as conjectured in Remark 2.1. We compare it to the entropy in (4.4) for different values of $n$ in Figure 6. For the choice of $n = 2$, we recover a similar behavior of the entropy, namely, a region with slow decay followed by an exponential decay. As the solution converges to 1 it can easily be seen that for $n > 2$ the entropy does not converge to
zero exponentially, as exemplified by the choice of $n = 2.1$ in the figure, since it is not the correct relative entropy with respect to the equilibrium.

### 4.4. The three-component Maxwell-Stefan system.

The Maxwell-Stefan system for $N = 2$ can be written as

\[
\begin{align*}
\partial_t \rho_i &= \nabla \cdot \left( \sum_{j=1}^{2} A_{ij}(\rho_1, \rho_2) \nabla \rho_j \right) \quad \text{in } \Omega, \ t > 0, \\
\sum_{j=1}^{2} A_{ij}(\rho_1, \rho_2) \partial_\nu \rho_j &= 0 \quad \text{on } \partial \Omega, \ t > 0, \\
\rho_i(0) &= (\rho_0)_i \quad \text{in } \Omega
\end{align*}
\]

for $i = 1, 2$, with

\begin{equation}
A(\rho_1, \rho_2) = \frac{1}{\delta(\rho_1, \rho_2)} \begin{pmatrix}
d_1 + (d_3 - d_1)\rho_1 \\
(d_3 - d_1)\rho_2 \\
(d_3 - d_2)\rho_1
\end{pmatrix}
\end{equation}

and

\[
\delta(\rho_1, \rho_2) = d_1 d_2 (1 - \rho_1 - \rho_2) + d_2 d_3 \rho_1 + d_3 d_1 \rho_2.
\]

The unknowns $\rho_1$ and $\rho_2$ represent the concentrations of the first two gases ($\rho_3 = 1 - (\rho_1 + \rho_2)$); the parameters $d_1$, $d_2$, and $d_3$ are related to the binary diffusion coefficients of the three gases. In section 5 below, we derive this form of the Maxwell-Stefan system, prove that it fits our framework, and discuss the case $N > 2$.

---

**Fig. 7.** The mesh used for the Duncan-Toor example, depicting the carbon dioxide content after about ten hours.

**Fig. 8.** Comparison of the mole fractions in the left side of the device.

In [7, Sec. 2] numerical results were presented for the three component gas diffusion experiment originally performed by Duncan and Toor in [15]. The setting is the following. Consider two spherical bulbs of volume 77.99 cm$^3$ (radius 26.49 mm) and 78.63 cm$^3$ (radius 26.58 mm), respectively, which are connected by a capillary tube of length 85.9 mm and diameter 2.08 mm, with a valve in the middle. We consider the Maxwell-Stefan equations with $N = 2$, corresponding to the gas mixture composed of hydrogen ($\rho_1$), nitrogen ($\rho_2$), and carbon dioxide ($\rho_3$). We consider the following initial gas mixture in the left- and right-hand side of the device:

- **Left:** $(\rho_0)_1 = 0.000, \ (\rho_0)_2 = 0.501, \ (\rho_0)_3 = 0.499$
- **Right:** $(\rho_0)_1 = 0.501, \ (\rho_0)_2 = 0.499, \ (\rho_0)_3 = 0.000.$
For these gases, the coefficients $d_1$, $d_2$, and $d_3$ are given in terms of the binary diffusion coefficients (see section 5.1 below) as follows:

$$d_1^{-1} = D_{13} = 68.0 \text{ mm}^2 \text{s}^{-1}, \quad d_2^{-1} = D_{23} = 16.8 \text{ mm}^2 \text{s}^{-1}, \quad d_3^{-1} = D_{12} = 83.3 \text{ mm}^2 \text{s}^{-1}.$$ 

As in [7], we can reduce the domain to two dimensions, as the device and initial conditions are axially symmetric and the flux vector has no angular component. In Figure 7, the computational domain is shown. We choose the spatial mesh size $h_s = 2.08 \text{ mm}$, equal to the diameter of the tube. The size of the Cartesian product mesh in time is chosen as 20.8 s. We solve iteratively on these slabs, restarting the computations with the previous solution as initial condition. We fix $p = 2$ and $\varepsilon = 10^{-10}$.

The results are shown in Figure 8. We recover the same behavior shown in [7]. Both hydrogen and carbon dioxide converge monotonically to the expected equilibrium. Nitrogen shows the peculiar behavior known from the experiment.

![Figure 9](image.png)

**Fig. 9.** Entropy and entropy dissipation (in absolutes) for the Duncan-Toor example, both showing exponential convergence to the equilibrium.

In Figure 9, we show the relative entropy and its dissipation, i.e. the time derivative of the entropy, both converge exponentially. This can be expected, as this behavior has been proven for similar types of equations, see for example [3, eq. (1.16)] where it is formally stated that the entropy dissipation converges exponentially for the Fokker-Planck equation by the use of the Bakry-Emery method.

5. The Maxwell-Stefan system revisited. In this section, we derive the formulation of the Maxwell-Stefan system as that used in section 4.4, and show that it fits into the general framework of section 2 (section 5.1). For the case $N > 2$, in which an explicit representation of the currents may not be easily derived, we introduce and analyze an alternative space-time Galerkin method, which is based on a formulation that is implicit for the currents (section 5.2).

Let $\rho_0 \in L^\infty(\Omega)^{N+1}$ such that $\rho_0 \geq 0$ and $\sum_{i=1}^{N+1} \rho_0 = 1$. The Maxwell-Stefan equations are given by the continuity equations

$$\begin{cases}
\partial_t \rho_i + \nabla \cdot J_i = 0 & \text{in } (0, T) \times \Omega, \\
\nu \cdot J_i = 0 & \text{on } (0, T) \times \partial \Omega, \\
\rho_i(0) = (\rho_0)_i & \text{in } \Omega
\end{cases}
$$

for $i = 1, \ldots, N+1$, where the currents $J_i$ are implicitly given by

$$\nabla \rho_i = \sum_{j=1}^{N+1} \frac{\rho_i J_j - \rho_j J_i}{D_{ij}}$$

for some $D_{ij} = D_{ji} > 0$.

5.1. Explicit formula for the currents. In this section, we establish an explicit representation of the currents, which allows us to derive the formulation of the Maxwell-Stefan system in the concentration variable unknowns. We follow [6] (see also [30]).
Let $M_{ij}(\rho) := D_{ij}^{-1} \rho_i - \delta_{ij} \sum_{k=1}^{N+1} D_{ik}^{-1} \rho_k$, $i, j = 1, \ldots, N + 1$. Thus,

$$\nabla \rho_i = \sum_{j=1}^{N+1} M_{ij}(\rho) J_j.$$

Using $\rho_i \geq 0$ and $D_{ij} = D_{ji} > 0$, it is easy to see that $M(\rho)$ is quasi-positive ($M_{ij}(\rho) \geq 0$ for $i \neq j$). Moreover, provided that $\rho_i > 0$ for all $1 \leq i \leq N + 1$, $M(\rho)$ is irreducible. Direct calculations show that

$$\text{Ker} \ M(\rho) \supseteq \text{span} \{\rho\} \quad \text{and} \quad \text{Im} \ M(\rho) \subseteq \left\{ v : \sum_{i=1}^{N+1} v_i = 0 \right\}.$$

Moreover, $R^{-1} M(\rho) R$, with $R = \text{diag}(\rho_1^{1/2}, \ldots, \rho_{N+1}^{1/2})$, is symmetric, thus all the eigenvalues of $M(\rho)$ are real. By the Perron-Frobenius theory for quasi-positive, irreducible matrices, one deduces that the eigenvalue zero has multiplicity one (we refer to [6] or [30] for details). We deduce

$$(5.3) \quad \text{Ker} \ M(\rho) = \text{span} \{\rho\} \quad \text{and} \quad \text{Im} \ M(\rho) = \left\{ v : \sum_{i=1}^{N+1} v_i = 0 \right\}.$$ 

As $M(\rho)$ is not invertible, we have to restrict ourselves to a subspace of all possible currents $J$ in order to obtain an explicit formula for $J$. For this, we make the assumption that the total current

$$J_{\text{tot}} := \sum_{i=1}^{N+1} J_i$$

vanishes. Then by summing in (5.1) over all $i = 1, \ldots, N + 1$, we see that

$$\rho_{\text{tot}} = \sum_{i=1}^{N+1} \rho_i$$

is constant in time, and hence $\rho_{\text{tot}} = \sum_{i=1}^{N+1} (\rho_0)_i = 1$. Using this, we can rewrite the implicit formulation of the currents as

$$(5.4) \quad \nabla \rho_i = \frac{\rho_i \left( - \sum_{j=1}^{N} J_j \right) - \left( 1 - \sum_{j=1}^{N} \rho_j \right) J_i}{D_{i(N+1)}} + \sum_{j=1}^{N} \frac{\rho_i J_j - \rho_j J_i}{D_{ij}}.$$

As before, we can define a matrix

$$(5.5) \quad M_{ij}(\rho) := \frac{\rho_i}{D_{ij}} - \frac{\rho_i}{D_{i(N+1)}} - \delta_{ij} \left( \sum_{k=1}^{N} \rho_k D_{ik}^{-1} + \frac{1}{D_{i(N+1)}} \sum_{i=1}^{N} \rho_i \right), \quad i, j = 1, \ldots, N.$$

From (5.3), the matrix $M(\rho)$ has full rank, and hence it is invertible. We have

$$J_i = -\sum_{j=1}^{N} A_{ij}(\rho) \nabla \rho_j \quad \text{with} \quad A(\rho) := -M(\rho)^{-1}.$$

**Remark 5.1.** The matrix $M(\rho)$ is actually independent from the diagonal elements $D_{ii}$.

**Proposition 5.2.** Let $s$ be as in (4.1), and let $M$ be given by (5.5). Then, the matrix-valued function $A(\rho) := -M(\rho)^{-1}$ fulfills (H1) and (H2a).
Proof. Let \( A(\rho) = -\mathcal{M}^{-1}(\rho) \). The fact that \( \mathcal{M} \) is smooth directly implies that \( A \) is smooth. Similarly as in the proof of [30, Lemma 3.2], one can show that

\[
\sum_{i=1}^{d} \partial_i w \cdot A(u(w))s''(u(w))^{-1} \partial_i w \geq \gamma |\nabla u(w)|^2
\]

for some \( \gamma > 0 \) and all smooth \( w \).

In order to prove \( \text{(H2a)} \), we have to show that

\[
z \cdot s''(\rho)A(\rho)z \geq \gamma |z|^2 \quad \text{for all } z \in \mathbb{R}^N, \rho \in \mathcal{D}.
\]

Let \( \rho \in \mathcal{D}, x_0 \in \Omega \), and \( z \in \mathbb{R}^N \). We define the following vector-valued function of \( x \):

\[
w(x) := s'(\rho) + s''(\rho)z(x - x_0) \cdot \hat{e}_1,
\]

where \( \hat{e}_1 \) denotes the unit vector \((1, 0, \ldots, 0) \in \mathbb{R}^d \). We have

\[
\partial_i w(x_0) = \delta_{i1} s''(\rho)z
\]

and, for \( u = (s')^{-1} \),

\[
\partial_i u(w(x_0)) = u'(w(x_0))\partial_i w(x_0) = u'(w(x_0))\delta_{i1} s''(\rho)z = u'(w(x_0))\delta_{i1} s''(u(w(x_0)))z = \delta_{i1} z.
\]

This, together with (5.6), implies that

\[
z \cdot s''(\rho)A(\rho)z = (s''(\rho)z) \cdot A(\rho)s''(\rho)^{-1} (s''(\rho)z)
\]

\[
= \sum_{i=1}^{d} \partial_i w(x_0) \cdot A(u(w(x_0)))s''(u(w(x_0)))^{-1} \partial_i w(x_0)
\]

\[
\geq \gamma |\nabla u(w(x_0))|^2 = \gamma |z|^2,
\]

which proves the assertion. \( \square \)

For \( N = 1 \), the matrix \( \mathcal{M}(\rho) \) is actually a scalar, which is given by

\[
\mathcal{M}(\rho) = -\frac{\rho_1}{D_{12}} - \frac{1 - \rho_1}{D_{12}} = -\frac{1}{D_{12}}.
\]

Hence, \( J_1 = D_{12} \nabla \rho_1 \). Therefore, in this case the Maxwell-Stefan system reduces to the heat equation.

For three species/gases \( (N = 2) \), we have

\[
\mathcal{M}(\rho_1, \rho_2) = \left( \frac{\rho_1}{D_{11}} - \frac{\rho_1}{D_{13}} - \frac{\rho_1}{D_{12}} - \frac{\rho_2}{D_{22}} - \frac{1 - \rho_1 - \rho_2}{D_{13}} + \frac{\rho_1}{D_{13}} - \frac{\rho_1}{D_{12}} - \frac{\rho_2}{D_{23}} - \frac{1 - \rho_1 - \rho_2}{D_{23}} \right)
\]

Let

\[
d_1 := \frac{1}{D_{13}}, \quad d_2 := \frac{1}{D_{23}}, \quad d_3 := \frac{1}{D_{12}},
\]

and recall that \( D_{21} = D_{12} \). One can verify that

\[
\delta(\rho_1, \rho_2) := \det \mathcal{M}(\rho_1, \rho_2) = d_1 d_2 (1 - \rho_1 - \rho_2) + d_2 d_3 \rho_1 + d_3 d_1 \rho_2 \neq 0.
\]

Let \( A(\rho) \) denote the inverse of \(-\mathcal{M}(\rho)\). We can rewrite the Maxwell-Stefan equations as the system in section 4.4.
5.2. Implicit formulation for the currents. In subsection 5.1, we have seen that the Maxwell-Stefan system (5.1)-(5.2), can be written in the form (2.1), with \( f = 0 \) and \( A(\rho) \) being given by the inverse of \(-\mathcal{M}(\rho)\) for

\[
\mathcal{M}_{ij}(\rho) := \frac{\rho_i}{D_{ij}} - \frac{\rho_i}{D_{(N+1)i}} - \delta_{ij} \left( \sum_{k=1}^{N} \frac{\rho_k}{D_{ik}} + \frac{1 - \sum_{l=1}^{N} \rho_l}{D_{(N+1)i(N+1)}} \right), \quad i, j = 1, \ldots, N.
\]

Moreover, we have computed \( A(\rho) \) explicitly for \( N = 1 \) and \( N = 2 \). However, for large \( N \), it is more complicated to find the explicit formulation for \( A(\rho) \). In any case we do not expect a simple formulation in these cases. Therefore, this section provides a space-time Galerkin scheme, which avoids the explicit computation of the inverse of \( \mathcal{M} \).

Let \( q, p \in \mathbb{N} \). We consider the following problem:

\[
(5.7) \quad \text{Find } w_h^\varepsilon \in V_h^p, J^\mu \in V_h^q, \mu = 1, \ldots, d, \text{ such that }
0 = \varepsilon(\phi^0, w_h^\varepsilon)_{H^1_0(\Omega_T)} + \int_{\Omega_T} \phi^0(T) \cdot u(w_h^\varepsilon(T))dx - \int_{\Omega} \phi^0(0) \cdot \rho_0 dx - \int_{\Omega_T} \partial_t \phi^0 \cdot u(w_h^\varepsilon)dx dt
- \sum_{\mu=1}^{d} \left( \int_{\Omega_T} \partial_{x_{\mu}} \phi^0 \cdot J^\mu dx dt + \int_{\Omega_T} \phi^\mu \cdot (\partial_{x_{\mu}} w_h^\varepsilon - s''(u(w_h^\varepsilon))\mathcal{M}(u(w_h^\varepsilon))J^\mu) dx dt \right)
\forall \phi^0 \in V_h^p, \phi^\mu \in V_h^q, \mu = 1, \ldots, d.
\]

Proposition 5.3. Assume that \( \rho_0 : \Omega \to \mathcal{D} \) is measurable. Then there exists a solution \( w_h^\varepsilon \in V_h^p, J^\mu \in V_h^q, \mu = 1, \ldots, d \) of the method (5.7).

For the proof of Proposition 5.3, we need the following lemma.

Lemma 5.4. If \( w_h^\varepsilon \in V_h^p, J^\mu \in V_h^q, \mu = 1, \ldots, d \), solves (5.7), then

\[
\varepsilon \|w_h^\varepsilon\|^2_{H^1_0(\Omega_T)} + \int_{\Omega_T} s(u(w_h^\varepsilon(T)))dx + \gamma \sum_{\mu=1}^{d} \int_{\Omega_T} |\mathcal{M}(u(w_h^\varepsilon))J^\mu|^2 dx dt \leq \int_{\Omega} s(\rho_0)dx.
\]

Proof. We can use \( \phi^0 = w_h^\varepsilon \) and \( \phi^\mu = 0 \) for \( \mu = 1, \ldots, d \) as test functions and, similarly to the proof of Proposition 3.3, we obtain that

\[
\varepsilon \|w_h^\varepsilon\|^2_{H^1_0(\Omega_T)} + \int_{\Omega_T} s(u(w_h^\varepsilon(T)))dx - \sum_{\mu=1}^{d} \int_{\Omega_T} J^\mu \cdot \partial_{x_{\mu}} w_h^\varepsilon dx dt \leq \int_{\Omega} s(\rho_0)dx.
\]

The next step is to use the test functions \( \phi^0 = 0 \) and \( \phi^\mu = J^\mu \) for \( \mu = 1, \ldots, d \) to obtain

\[
\frac{d}{\sum_{\mu=1}^{d}} \int_{\Omega_T} J^\mu \cdot \partial_{x_{\mu}} w_h^\varepsilon dx dt = \frac{d}{\sum_{\mu=1}^{d}} \int_{\Omega_T} J^\mu \cdot s''(u(w_h^\varepsilon))\mathcal{M}(u(w_h^\varepsilon))J^\mu dx dt.
\]

According to assumption (H2a), we know that \( s''(v)A(v) \) is positive semi-definite and satisfies

\[
z \cdot s''(v)A(v)z \geq \gamma |z|^2 \quad \text{for all } z \in \mathbb{R}^N, v \in \mathcal{D}.
\]

Choosing \( v = u(w_h^\varepsilon) \), \( z := \mathcal{M}(u(w_h^\varepsilon))J^\mu \), we see that

\[
\gamma |\mathcal{M}(u(w_h^\varepsilon))J^\mu|^2 \leq J^\mu \cdot \mathcal{M}(v)s''(v)A(v)\mathcal{M}(v)J^\mu = -J^\mu \cdot \mathcal{M}(v)s''(v)J^\mu,
\]

where in the last step we have used that \( A(v) \) is the inverse of \(-\mathcal{M}(v)\). Thus, we conclude that

\[
\varepsilon \|w_h^\varepsilon\|^2_{H^1_0(\Omega_T)} + \int_{\Omega_T} s(u(w_h^\varepsilon(T)))dx + \gamma \sum_{\mu=1}^{d} \int_{\Omega_T} |\mathcal{M}(u(w_h^\varepsilon))J^\mu|^2 dx dt \leq \int_{\Omega} s(\rho_0)dx.
\]
Proof of Proposition 5.3. The idea of the proof is to proceed similarly to the proof of Proposition 3.3. We define the mapping
\[ \Phi : V^p_h \times (V^q_h)^d \to V^p_h \times (V^q_h)^d, \ (v, J^1, \ldots, J^d) \mapsto (w, J^1, \ldots, J^d), \]
where \( w \) is (uniquely) defined via the equation
\[ 0 = \varepsilon \phi^0 (v, w)_{H^2(Q_T)} + \int_{Q_T} \phi^0 (v(T), v(T)) dx - \int_{Q_T} \phi^0 (0, \rho^0) dx - \int_{Q_T} \partial_t \phi^0 \cdot u(v) dx dt - \sum_{\mu=1}^d \int_{Q_T} \partial_{x_\mu} \phi^0 \cdot J^\mu dx dt \quad \text{for all } \phi^0 \in V^p_h, \]
and \( J^\mu \) denotes the unique solution (see below for a justification) of
\[ (5.8) \quad \int_{Q_T} \phi^\mu \cdot \partial_x v dx dt = \int_{Q_T} \phi^\mu \cdot s''(u(v)) M(u(v)) J^\mu dx dt \quad \text{for all } \phi^\mu \in V^p_h. \]

Note that the mapping \( \Phi \) is well-defined, as (5.8) admits a unique solution for given \( v \in V^p_h \) according to the Lemma of Lax-Milgram: we see that \( \partial_x v \in L^2(Q_T)^N \) and the matrix \( -s''(u(v)) M(u(v)) \in L^\infty(Q_T)^{N \times N} \) is positive definite, because for all \( z \in \mathbb{R}^N \)
\[ z \cdot (-s''(u(v)) M(u(v))) z = A(u(v)) y \cdot s''(u(v)) y \]
\[ = y \cdot s''(u(v)) A(u(v)) y \]
\[ \geq \gamma |y|^2 = \frac{\gamma}{\|A(u(v))\|^2} \|A(u(v))\|^2 |y|^2 \]
\[ \geq \frac{\gamma}{\|A(u(v))\|^2} \|A(u(v))\|^2 |y|^2 = \frac{\gamma}{\|A(u(v))\|^2} |z|^2 \]
for \( y := A(u(v))^{-1} z = -M(u(v)) z \). Moreover, the mapping \( \Phi \) is continuous since \( A \) and \( u \) are continuous. Then by the Leray-Schauder fixed-point theorem, we obtain that \( \Phi \) admits a fixed-point if we can show that the set
\[ \{(w, J^1, \ldots, J^d) \in V_h \times (V^q_h)^d : (w, J^1, \ldots, J^d) = \sigma \Phi(w, J^1, \ldots, J^d), \sigma \in [0, 1]\} \]
is bounded. Let \( (w, J^1, \ldots, J^d) = \sigma \Phi(w, J^1, \ldots, J^d) \) for \( \sigma \in (0, 1] \). Similarly to Lemma 5.4, we can prove the entropy estimate
\[ \frac{\varepsilon}{\sigma} \|w\|^2_{H^2(Q_T)} + \int_{Q_T} s(u(w(T))) dx + \gamma \sum_{\mu=1}^d \int_{Q_T} |M(u(w)) J^\mu|^2 dx dt \leq \int_{Q_T} s(\rho) dx. \]
Using that \( \sigma \in (0, 1] \) is bounded from above yields a uniform bound on \( w \) in \( V^q_h \) and on \( M(u(w)) J^\mu \) in \( L^2(Q_T)^N \). As \( V^q_h \) is finite dimensional, we directly obtain that \( \|w\|_{L^\infty(Q_T)^N} \) is uniformly bounded. Thus,
\[ \|J^\mu\|_{L^2(Q_T)^N} \leq \|A(u(w))\|_{L^\infty(Q_T)^{N \times N}} \|M(u(w)) J^\mu\|_{L^2(Q_T)^N} \]
is also uniformly bounded. As all norms are equivalent on \( V^q_h \), this directly implies that \( J^\mu \) is uniformly bounded in \( V^p_h \). Thus, the Leray-Schauder theorem is applicable and yields that \( \Phi \) has a fixed-point, and therefore the scheme (5.7) admits a solution. \( \Box \)

**Proposition 5.5.** Let \( \rho_0 : \Omega \to \overline{D} \) be measurable and \( w^\varepsilon_h \in V^p_h, J^{\varepsilon, \mu}_h \in V^q_h, \mu = 1, \ldots, d, \) be a solution of (5.7) for \( \varepsilon, h > 0 \). Then there exist a solution \( \rho \) of (3.3) and sequences \( h_i, \varepsilon_i \to 0, \) as \( i \to \infty \), such that
\[ u(w^\varepsilon_h) \to \rho \quad \text{in } L^r(Q_T), \text{ as } i \to \infty. \]
for all \( r \in [1, \infty) \). Moreover, \( \rho \) satisfies the entropy estimate

\[
\int_{\Omega} s(\rho(\tau))d\tau + \gamma \int_{0}^{T} \int_{\Omega} |\nabla \rho|^2 d\tau dt \leq \int_{\Omega} s(\rho_0)d\tau
\]

for all \( \tau \in (0, T) \), where \(|\Omega|\) is the volume of \( \Omega \).

**Proof.** The proof is analogue to the proof of Proposition 3.4. We only need to replace Proposition 3.6 by Lemma 5.6 below.

**Lemma 5.6 (Convergence of the scheme for fixed \( \varepsilon > 0 \)).** Let \( w_h \in V_h^\mu, J_h^\mu \in V_h^\rho, \mu = 1, \ldots, d \) be a solution of (5.7), with fixed \( \varepsilon > 0 \). Then there exists \( \rho \in H^1(Q_T)^N \) with \( \rho(t, x) \in \overline{\mathcal{D}} \) for a.e. \( (t, x) \in Q_T \) and \( s'(\rho) \in H^1(Q_T)^N \), and a sequence \( h_t \to 0 \) such that

\[
\rho_h := u(w_{h_t}) \to \rho \quad \text{and} \quad w_{h_t} \to s'(\rho)
\]

strongly in \( L^2(Q_T) \) and weakly in \( H^1_\varepsilon(Q_T) \). Moreover, \( \rho \) solves (3.4) and satisfies the entropy estimate (3.8) for \( w = s'(\rho) \).

**Proof.** The fact that \( w_h \) is uniformly bounded in \( H^1_\varepsilon(Q_T)^N \) yields that there exists \( w \in H^1_\varepsilon(Q_T)^N \) and subsequence \( h_t \to 0 \) such that \( w_{h_t} \to w \) in \( H^1_\varepsilon(Q_T)^N \), due to the Banach-Alaoglu theorem, and \( w_{h_t} \to w \) in \( L^2(Q_T)^N \) due to Rellich’s theorem. As \( u \) is bounded, the dominated convergence theorem entails the convergence for \( \rho_{h_t} \equiv u(w_{h_t}) \) to \( \rho := u(w) \) along another subsequence (which we do not relabel).

For the second part, we note that, due to the Banach-Alaoglu theorem and the boundedness of \( \mathcal{M}(u(w))J_h^\mu \) in \( L^2(Q_T)^N \), we know that there exist \( \xi^\mu \in L^2(Q_T)^N \) such that, for a subsequence (not being relabeled),

\[
\mathcal{M}(u(w_h))J_h^\mu \rightharpoonup \xi^\mu \quad \text{weakly in } L^2(Q_T)^N.
\]

In particular,

\[
J_h^\mu = -A(u(w_h))\mathcal{M}(u(w_h))J_h^\mu \rightharpoonup -A(\rho)\xi^\mu =: J^\mu \quad \text{weakly in } L^r(Q_T)^N
\]

for every \( r \in [1, 2) \). Finally, for every \( \phi^\mu \in H^1_\varepsilon(Q_T)^N, \mu = 0, \ldots, d \), there exist \( \phi_{h_t}^\mu \in V_h^\mu \cap V_h^\rho \) such that \( \phi_{h_t}^\mu \to \phi^\mu \) in \( H^1_\varepsilon(Q_T)^N \). Using \( \phi_{h_t}^\mu \) as a test function in (5.7), in the limit \( h_t \to 0 \), we obtain

\[
0 = \varepsilon(\phi^0, w)\varphi^1(Q_T) + \int_{\Omega} \phi^0(T) \cdot u(w(T)) dx - \int_{\Omega} \phi^0(0) \cdot \rho_0 dx - \int_{Q_T} \partial_t \phi^0 \cdot u(w) dx dt
\]

\[
- \sum_{\mu=1}^{d} \left( \int_{Q_T} \partial_{x_{\mu}} \phi^0 \cdot J^\mu dx dt + \int_{Q_T} \phi^\mu \cdot (\partial_{x_{\mu}} w - s''(u(w))\mathcal{M}(u(w))J^\mu) dx dt \right),
\]

as each integral in (5.7) converges separately. In particular, by the fundamental lemma of calculus of variations, we see that \( \partial_{x_{\mu}} w = s''(u(w))\mathcal{M}(u(w))J^\mu \) and equivalently

\[
J^\mu = \mathcal{M}(u(w))^{-1}s''(u(w))^{-1}\partial_{x_{\mu}} w = -A(u(w))u'(w)\partial_{x_{\mu}} w = -A(u(w))\partial_{x_{\mu}} u(w),
\]

which implies that \( \rho \) solves (3.4). Finally, the entropy inequality is a consequence of Fatou’s lemma.

**5.3. Numerical Tests.** We again turn to [7, Sec. 3] for numerical results we can compare our method to. This time, we consider a model for the lung. The computational domain resembles a branch of the tree structure found in the bottom of the lung. The domain, depicted in Figure 10, consists of the inflow, \( \Gamma_1 \), on top, the outflow, \( \Gamma_2 \), located on the bottom of the two branches, and the alveoli, \( \Gamma_3 \), located in the middle of each of the branches. The remaining boundary \( \Gamma_4 \) is a wall where nothing goes in or out. Opposed to the domain presented in the reference, we consider the branches of the lung to be symmetrical and perpendicular to each other. The paper does not mention the angle between the branches used there. Also the size of the alveoli is left
unspecified in the paper. Here, we split the boundary of the branches into three equal parts, with the alveoli (Γ₃) in the middle. On Γ₁, Γ₂, Γ₃ we impose Dirichlet boundary conditions to model the gas exchange with the other parts of the lung. On the wall, Γ₄, we take homogeneous Neumann boundary conditions.

We make use of the implicit formulation (5.7) to find the numerical solution. To incorporate the Dirichlet boundary condition, we use Nitsche’s method and add to (5.7) the following terms:

\[
\sum_{\mu=1}^{d} \int_{(0,T) \times \Gamma_D} J^\mu \nu^\mu \cdot \phi^0 + \int_{(0,T) \times \Gamma_D} (u(w) - \rho_D) \cdot \phi^\mu \nu^\mu + \int_{(0,T) \times \Gamma_D} \eta h_s^{-1} (u(w) - \rho_D) \cdot \phi^0
\]

for a parameter \(\eta > 0\), \(h_s\) being the spatial mesh size, on the Dirichlet boundary \(\Gamma_D\). In the examples below, we use \(\eta = 1\). The first term comes from the integration by parts. The second and third terms are productive zeros that weakly enforce the Dirichlet boundary condition, and are chosen such that they agree with Nitsche’s method for the heat equation in the degenerative case.

### 5.3.1. Diffusion of air.

In the following example, compare [7, Sec. 3.4], we choose alveolar air as initial condition and as the Dirichlet data on the outflow and alveoli. On the inflow boundary we choose humidified air as Dirichlet data. See Table 2 for the gas components of the different types of air, and Table 3 for the diffusion coefficients.

#### Table 2

**Components of the different gas mixtures.**

|         | Humidified air | Alveolar air | Alveolar heliox |
|---------|----------------|--------------|-----------------|
| Nitrogen| 0.7409         | 0.7490       | 0.0000          |
| Oxygen  | 0.1967         | 0.1360       | 0.1360          |
| Carbon dioxide | 0.0004        | 0.0530       | 0.0530          |
| Water   | 0.0620         | 0.0620       | 0.0620          |
| Helium  | 0.0000         | 0.0000       | 0.7490          |

#### Table 3

**Diffusion coefficients of the different gases.**

|         | Oxygen | Carbon dioxide | Water | Helium |
|---------|--------|----------------|-------|--------|
| Nitrogen| 21.87  | 16.63          | 23.15 | 74.07  |
| Oxygen  | 16.40  | 22.85          | 79.07 |        |
| Carbon dioxide | 16.02 | 63.45          |       |        |
| Water   |        |                |       | 90.59  |
Since there is no helium present we can reduce the number of species involved, setting $N = 3$. For the numerical calculations we choose spatial mesh size $h_s = 0.3$ and measure the value of the gas every 0.001 seconds. The discrete system is not ill-conditioned and we are able to choose $\varepsilon = 0$. In Figure 11 we show the numerical results for Oxygen and Carbon dioxide as the other gases stay (almost) constant. Both converge to their equilibrium value. Comparing the results to [7], we can see that the equilibrium value slightly differs, which is likely due to the symmetry of the domain and size of the alveoli.

![Figure 11](image1)

**Fig. 11.** Numerical results of the mole fractions Oxygen and Carbon dioxide inside the lung for air mixture.

5.3.2. Diffusion of air/heliox. Next, we try to reproduce the results form [7, Sec. 3.5]. We consider alveolar heliox as initial condition. As the Dirichlet data on the outflow and alveoli, we also choose alveolar heliox, whereas we put humidified air on the inflow. The discrete system is very ill-conditioned due to the gas components taking zero values. In order for the solver to converge, we had to choose $\varepsilon = 10^{-3}$. Furthermore, to avoid the singularity of the entropy density, we adjust the helium content in air and the nitrogen content in heliox to be $10^{-7}$, subtracting the same amount of water, in order to keep them summing to one. Note that this is not unreasonable, for example, the correct amount of helium in air is about $5.3 \cdot 10^{-7}$. With these adjustments, the solver converges. The numerical results are shown in Figure 12. Both oxygen and carbon dioxide levels rise above the values in provided gas mixtures, before they start to decrease towards the equilibrium value. This is the expected behavior. However, the maximum values reached here are slightly lower than the ones found in [7]. This can be attributed to the perturbations of the zero concentrations and, as already seen, to the approximation of the geometry.

![Figure 12](image2)

**Fig. 12.** Numerical results of the mole fractions Oxygen and Carbon dioxide inside the lung for air/heliox mixture.
6. Conclusions. We have presented and analyzed a continuous space-time Galerkin method for cross-diffusion systems in entropy variable formulation, proving existence and convergence of discrete solutions, as well as existence of a weak solution of the continuous problem using the space-time approach. As opposed to time-stepping schemes, this approach provides an easy way to increase the approximation order simultaneously in space and time, makes it space-time approach. As opposed to time-stepping schemes, this approach provides an easy way to increase the approximation order simultaneously in space and time, makes it possible for cross-diffusion systems in entropy variable formulation, proving existence and convergence of discrete solutions, as well as existence of a weak solution of the continuous problem using the space-time approach. When using a tensor-product mesh.

Acknowledgments. All authors have been supported by the Austrian Science Fund (FWF) through the project F 65. I. Perugia and P. Stocker have also been supported by the FWF through the projects P 29197-N32 and W1245, respectively.

REFERENCES

[1] J. M. Alam, N. K.-R. Kevlahan, and O. V. Vasilyev, Simultaneous space-time adaptive wavelet solution of nonlinear parabolic differential equations, J. Comput. Phys., 214 (2006), pp. 829–857, https://doi.org/10.1016/j.jcp.2005.10.009.

[2] R. Andreev, Stability of sparse space-time finite element discretizations of linear parabolic evolution equations,IMA J. Numer. Anal., 33 (2013), pp. 242–260, https://doi.org/10.1093/imanum/dr3014.

[3] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, Comm. Partial Differential Equations, 26 (2001), pp. 43–100, https://doi.org/10.1081/PDE-100022246.

[4] A. K. Aziz and P. Monk, Continuous finite elements in space and time for the heat equation, Math. Comp., 52 (1989), pp. 255–274, https://doi.org/10.2307/2008467.

[5] F. Bonizzoni, M. Braukhoff, A. Jüngel, and I. Perugia, A structure-preserving discontinuous Galerkin scheme for the Fisher-KPP equation, Numer. Math., 146 (2020), pp. 209–226, https://doi.org/10.1007/s00211-020-01136-w.

[6] D. Bothe, On the Maxwell-Stefan approach to multicomponent diffusion, in Parabolic problems, vol. 80 of Progr. Nonlinear Differential Equations Appl., Birkhäuser/Springer Basel AG, Basel, 2011, pp. 81–93, https://doi.org/10.1007/978-3-0348-0075-4_5.

[7] L. Boudin, D. Götz, and B. Grec, Diffusion models of multicomponent mixtures in the lung, ESAIM: Proceedings, 30 (2010), pp. 91–104, https://doi.org/10.1051/proc/2010008, https://hal.archives-ouvertes.fr/hal-00455656.

[8] L. Boudin, B. Grec, and F. Salvarani, A mathematical and numerical analysis of the Maxwell-Stefan diffusion equations, Discrete Contin. Dyn. Syst. Ser. B, 17 (2012), pp. 1427–1440, https://doi.org/10.1093/dcdsdb/2012.17.1427.

[9] M. Burger, M. Di Francesco, J.-F. Pietschmann, and B. Schlake, Nonlinear cross-diffusion with size exclusion, SIAM J. Math. Anal., 42 (2010), pp. 2842–2871, https://doi.org/10.1137/100783674.

[10] B. Carnes and G. F. Carey, Local boundary value problems for the error in FE approximation of non-linear diffusion systems, Internat. J. Numer. Methods Engrg., 73 (2008), pp. 665–684, https://doi.org/10.1002/nme.2103.

[11] X. Chen and A. Jüngel, Analysis of an incompressible Navier-Stokes-Maxwell-Stefan system, Comm. Math. Phys., 340 (2015), pp. 471–497, https://doi.org/10.1007/s00220-015-2472-z.

[12] X. Chen and A. Jüngel, When do cross-diffusion systems have an entropy structure?, J. Diff. Eq., 278 (2021), pp. 60–72, https://doi.org/10.1016/j.jde.2020.12.037.

[13] E. S. Daus, A. Jüngel, and B. Q. Tang, Exponential time decay of solutions to reaction-cross-diffusion systems of Maxwell-Stefan type, Arch. Ration. Mech. Anal., 235 (2020), pp. 1059–1104, https://doi.org/10.1007/s00205-019-1439-9.

[14] K. Dieter-Kissling, H. Marschall, and D. Bothe, Numerical method for coupled interfacial surfactant transport on dynamic surface meshes of general topology, Comput. & Fluids, 109 (2015), pp. 168–184, https://doi.org/10.1016/j.compfluid.2014.12.017.

[15] J. B. Duncan and H. L. Toor, An experimental study of three component gas diffusion, AIChE Journal, 8 (1962), pp. 38–41, https://doi.org/10.1002/aic.690080112.

[16] H. Egger, Structure preserving approximation of dissipative evolution problems, Numer. Math., 143 (2019), pp. 85–106, https://doi.org/10.1007/s00211-019-01050-w.

[17] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems. I. A linear model problem, SIAM J. Numer. Anal., 28 (1991), pp. 43–77, https://doi.org/10.1137/0728003.
[18] K. Eriksson and C. Johnson, *Adaptive finite element methods for parabolic problems. IV. Nonlinear problems*, SIAM J. Numer. Anal., 32 (1995), pp. 1729–1749, https://doi.org/10.1137/0732078.

[19] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2010, https://doi.org/10.1090/gsm/019.

[20] J. Geiser, *Iterative solvers for the Maxwell-Stefan diffusion equations: methods and applications in plasma and particle transport*, Coget Math., 2 (2015), pp. Art. ID 1092913, 16, https://doi.org/10.1080/23311855.2015.1092913.

[21] V. Giovangigli and M. Massot, *Asymptotic stability of equilibrium states for multicomponent reactive flows*, Math. Models Methods Appl. Sci., 8 (1998), pp. 251–297, https://doi.org/10.1142/S0218202598000123.

[22] V. Giovangigli and M. Massot, *The local Cauchy problem for multicomponent reactive flows in full vibrational non-equilibrium*, Math. Methods Appl. Sci., 21 (1998), pp. 1415–1439, https://doi.org/10.1002/(SICI)1099-1476(199810)21:15<1415::AID-MMA230>3.0.CO;2-D.

[23] M. Herberg, M. Meyries, J. Prüss, and M. Wilke, *Reaction-diffusion systems of Maxwell-Stefan type with reversible mass-action kinetics*, Nonlinear Anal., 159 (2017), pp. 264–284, https://doi.org/10.1016/j.na.2016.07.010.

[24] O. Junge, D. Matthies, and H. Osberger, *A fully discrete variational scheme for solving nonlinear Fokker-Planck equations in multiple space dimensions*, SIAM J. Numer. Anal., 55 (2017), pp. 419–443, https://doi.org/10.1137/16M1056560.

[25] A. Jüngel, *The boundedness-by-entropy method for cross-diffusion systems*, Nonlinearity, 28 (2015), pp. 1963–2001, https://doi.org/10.1088/0951-7715/28/6/1963.

[26] A. Jüngel, *Entropy methods for diffusive partial differential equations*, SpringerBriefs in Mathematics, Springer, [Cham], 2016, https://doi.org/10.1007/978-3-319-34219-1.

[27] A. Jüngel and O. Leingang, *Convergence of an implicit Euler Galerkin scheme for Poisson-Maxwell-Stefan systems*, Adv. Comput. Math., 45 (2019), pp. 1469–1498, https://doi.org/10.1007/s10444-019-09674-0.

[28] A. Jüngel and S. Schuchnigg, *Entropy-dissipating semi-discrete Runge-Kutta schemes for nonlinear diffusion equations*, Commun. Math. Sci., 15 (2017), pp. 27–53, https://doi.org/10.4310/CMS.v15.n1.a2.

[29] A. Jüngel and I. V. Stelzer, *Existence analysis of Maxwell-Stefan systems for multicomponent mixtures*, SIAM J. Math. Anal., 45 (2013), pp. 2421–2440, https://doi.org/10.1137/120898164.

[30] A. Jüngel, *Cross-diffusion systems with entropy structure*, in Proceedings of Equadiff 2017 Conference, 2017, pp. 181–190.

[31] P. C. Kunstmann, B. Li, and C. Luchs, *Runge-Kutta time discretization of nonlinear parabolic equations studied via discrete maximal parabolic regularity*, Found. Comput. Math., 18 (2018), pp. 1109–1130, https://doi.org/10.1007/s10208-017-9364-x.

[32] U. Langer and O. Steinbach, eds., *Space-Time Methods. Applications to Partial Differential Equations*, vol. 25 of Radon Series on Computational and Applied Mathematics, De Gruyter, 2019, https://doi.org/10.1515/9783110548488.

[33] E. Leonardi and C. Angeli, *On the Maxwell-Stefan approach to diffusion: A general resolution in the transient regime for one-dimensional systems*, J. Phys. Chem. B, 114 (2010), pp. 151–164.

[34] J.-B. W. P. Loos, P. J. T. Verheijen, and J. A. Moulijn, *Multicomponent flow modeling*, Elsevier, 2016.

[35] M. Marion and R. Temam, *Iterative solvers for the Maxwell-Stefan diffusion equations: methods and applications in plasma and particle transport*, Cogent Math., 2 (2015), pp. Art. ID 1092913, 16, https://doi.org/10.1080/23311855.2015.1092913.
Wiss. Wien, 63 (1871).

[47] O. Steinbach, Space-time finite element methods for parabolic problems, Comput. Methods Appl. Math., 15 (2015), pp. 551–566, https://doi.org/10.1515/cmam-2015-0026.

[48] L. Tartar, The general theory of homogenization - A personalized introduction, vol. 7 of Lecture Notes of the Unione Matematica Italiana, Springer-Verlag, Berlin; UMI, Bologna, 2009, https://doi.org/10.1007/978-3-642-05195-1.

[49] J. Česenek and M. Feistauer, Theory of the space-time discontinuous Galerkin method for nonstationary parabolic problems with nonlinear convection and diffusion, SIAM J. Numer. Anal., 50 (2012), pp. 1181–1206, https://doi.org/10.1137/110828903.

[50] M. Zank, An exact realization of a modified Hilbert transformation for space-time methods for parabolic evolution equations, Comput. Methods Appl. Math., 21 (2021), pp. 479–496, https://doi.org/10.1515/cmam-2020-0026.